Designation: D 6518 – 00

# Standard Practice for Bias Testing a Mechanical Coal Sampling System[1]

This standard is issued under the fixed designation D 6518; 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 (ε) indicates an editorial change since the last revision or reapproval.

## 1. Scope

1.1 This practice presents sample collection and statistical evaluation procedures for testing mechanical sampling systems, subsystems, and individual system components for bias. It is the responsibility of the user of this practice to select the appropriate procedure for a specific sampling situation.

1.2 This practice does not purport to define an absolute bias. Bias defined by this practice is the difference between the population mean of the mechanical sampler test results and the accepted reference value.

1.3 *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[2]

D 2013 Method of Preparing Coal Samples for Analysis[2]

D 2234 Practice for Collection of a Gross Sample of Coal[2]

D 4621 Guide for Accountability and Quality Control in the Coal Analysis Laboratory[2]

D 4702 Guide for Inspecting Crosscut, Sweep-Arm, and Auger Mechanical Coal Sampling Systems for Conformance with Current ASTM Standards[2]

E 105 Practice for Probability Sampling of Materials[3]

E 122 Practice for Choice of Sample Size to Estimate a Measure of Quality for a Lot or Process[3]

E 456 Terminology Relating to Quality and Statistics[3]

E 691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method[3]

## 3. Terminology

3.1 *Definitions*—For additional definitions of terms used in this practice refer to Terminologies D 121 and E 456.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *bias*, *n*—the difference between the population mean of the mechanical sampler test results and the accepted reference value.

3.2.2 *confidence interval*, *n*—a numeric interval with a lower limit and a higher limit within which the true parameter value is estimated to fall. The confidence interval percentage indicates the percentage of time the true value will fall within the interval if the procedure is continuously repeated.

3.2.3 *correlation*, *n*—a measure of the linear dependence between paired system and reference measurements. Correlation frequently is expressed by the correlation coefficient, which can take a value from minus one (perfect negative linear relationship) to plus one (perfect positive linear relationship).

3.2.4 *delimitation error*, *n*—a material error that occurs when all the elements in a cross section of a coal stream do not have an equal probability of being intercepted (captured) by the sampler cutter during increment collection.

3.2.5 *ellipsoidal region*, *n*—an area that is formed by plane sections of ellipses that are defined by the values selected for the largest tolerable bias of each coal characteristic used in the bias test. The region will be used to determine if the system is biased.

3.2.6 *Hotelling's $T^2$ test*, *n*—a statistical test that is used to evaluate multivariate data. It is the multivariate equivalent of the Student's *t*-test.

3.2.7 *largest tolerable bias (*LTB), *n*—an interval whose upper and lower bounds represent the limits of an acceptable bias.

3.2.8 *mechanical sampling system*, *n*—a single machine or series of interconnected machines whose purpose is to extract mechanically, or process (divide and reduce), or a combination thereof, a sample of coal.

3.2.9 *paired data set*, *n*—system and reference values observed on samples collected and compared from the same batch of material.

3.2.10 *reference sample*, *n*—a sample used in testing of a mechanical sampling system which is comprised of one or more increments collected from the test batch or lot of coal by the stopped belt method as described in Practice D 2234.

3.2.11 *reject stream*, *n*—the coal flow within a mechanical sampling system, which occurs at each stage of division, before and after reduction, and is not included in the system sample.

3.2.12 *save stream*, *n*—the coal flow within a mechanical sampling system which occurs at each stage of division, before and after reduction, and after the final stage of division becomes the system sample.

3.2.13 *statistical independence*, *n*—two sample values are statistically independent if the occurrence of either one in no

[1] This practice 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 Jan. 10, 2000. Published April 2000.

[2] *Annual Book of ASTM Standards*, Vol 05.06.

[3] *Annual Book of ASTM Standards*, Vol 14.02.

way affects the probability assigned to the occurrence of the other.

3.2.14 *surrogate sample*, *n*—a sample, used in the evaluation of a mechanical sampling system, which is comprised of one or more increments collected from a coal stream within the mechanical sampling system in accordance with Practice D 2234, Conditions" A" or "B." Such a sample may be considered acceptable for evaluation of a mechanical sampling system's components, excluding the primary cutter, when demonstrated to be equivalent to the reference sample.

3.2.15 *system sample*, *n*—a sample collected from a test batch or lot of coal by the mechanical sampling system being tested for bias.

3.2.16 *Walsh averages*, *n*—given a series of observations (differences) $x1$, $x2$, ... $xn$, the $n$ $(n + 1)/2$ pair-wise averages given by:

$$(xi + xj)/2, 1 \leq i \leq j \leq n \quad (1)$$

3.2.16.1 *Discussion*—As an example of Walsh averages, assume one has three observations (differences) designated as $x1$, $x2$, and $x3$. There are then a total of $3(4)/2 = 6$ Walsh averages. They are as follows: $x1$, $x2$, $x3$, $(x1 + x2)/2$, $(x1 + x3)/2$, and $(x2 + x3)/2$.

## 4. Summary of Practice

4.1 This practice consists of procedures for comparing material collected by mechanical sampling systems to reference or surrogate samples collected by alternate procedures from individual batches or lots of coal, numbered 1 through *n*, in chronological order, providing *n* sets of samples. After collection, the test samples are prepared and analyzed using applicable ASTM test methods. For each measured characteristic, a numerical difference in the measurements between the observed system value and the observed reference value is calculated for each set of samples. Using the statistical procedures described in this practice, the set of differences from the *n* sets is then examined for evidence of bias between the mechanical system and reference measurements.

4.2 This practice is based on matched-pair experimental designs. The practice describes two procedures of sample collection, paired increment and paired test batch, and two statistical procedures for assessing bias: nonparametric and parametric. The Wilcoxon signed rank test procedure is a nonparametric test, assuming only symmetry of each of the univariate differences, the Hotelling's $T^2$ test is a parametric test assuming multivariate normality of the differences, and the Student's *t*-test is a parametric univariate test assuming normality of the differences.

## 5. Significance and Use

5.1 It is intended that these procedures be used to provide an estimate of the bias of a mechanical sampling system used to collect samples of coal. Mechanical coal-sampling systems are used extensively in industry for collecting samples while coal is being conveyed or transported in various stages of production, shipment, receipt, and use. The bias of the sampling system, in the measurement of coal quality, can have significant commercial and environmental consequences.

5.2 Bias as determined by these procedures need not be a constant or fixed value and can reflect the bias only under the conditions, which prevailed during the test period. Variables including, but not limited to, changes in the operation of the sampling system, the coal transfer operation, or the coal-sampling characteristics can cause changes in test results; therefore, if system bias is unacceptable, correct the cause rather than compensate for it.

5.3 A single bias test may not provide a meaningful generalized expectation of past or future system performance but an ongoing testing program can. Such a program may be established by mutual agreement of the interested parties.

5.4 Pairs of observed values used to draw conclusions regarding bias are subject to sources of error other than sampling. Valid conclusions are dependent upon the extent of bias, which may occur during sample handling, preparation, and analysis; thus, the importance of carefully following ASTM standards for sample preparation, laboratory analysis, and the importance of exercising careful quality control, must be emphasized.

5.5 In all cases, the test plan should approximate normal system operation and not be a source of bias itself. This is especially critical when the sampling system batch processes several consecutive increments at any stage. In this case, the system samples should consist of all the coal from an entire batch.

5.6 Since this practice includes several different methods of sample collection and statistical procedures, the procedures used for both sample collection and statistical processing must be chosen before the test is conducted. This does not preclude subjecting historical test data to alternate statistical procedures for alternative purposes.

## 6. Apparatus

6.1 *Sample Collection Devices*:

6.1.1 *Stopped-Belt Divider*—A device similar to that illustrated in Fig. 1. The width between the divider plates must be the same throughout the divider, and no less than three times the nominal top size of the coal. Assure the width is sufficient, and the design of the mechanism adequate, to enable quick and easy removal of all coal lying on the conveyor belt between the divider plates, including very fine material.

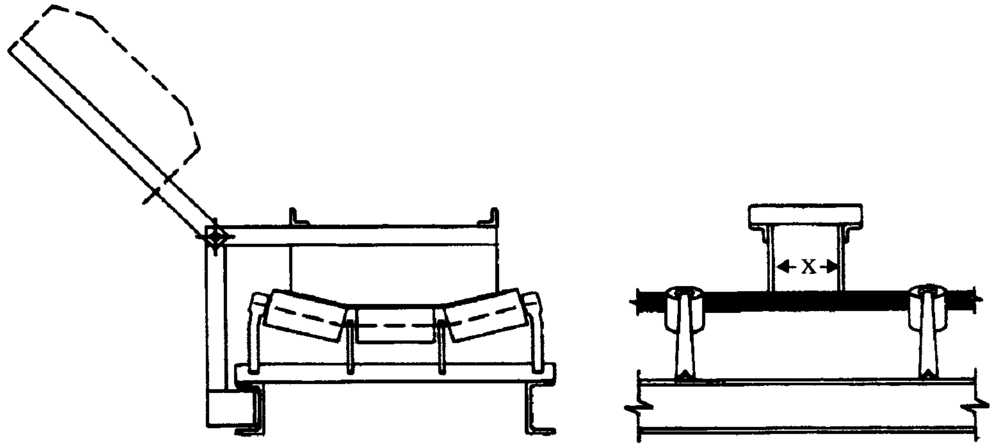
6.1.2 *Surrogate Reference Sample Collection Tools*—Devices used to subsample internal coal flows of a mechanical sampling system. These devices must be capable of extracting a full stream Type I-A-1 or I-B-1 increments (see Practice D 2234) from a mechanical sampling system stream of coal.

6.2 *Sample Preparation Equipment*—All bias test samples should be prepared using equipment as specified in Method D 2013.

## 7. Description of Test Procedures

7.1 *Sample Collection*:

7.1.1 This practice offers three basic test designs for bias testing of mechanical sampling systems. They are referred to as the paired increment, the paired test batch, and the intraphase test designs. The basic distinguishing features of the designs are given in 7.1.2.1-7.1.4.3. (**Warning**—Collecting test samples on multistage sampling systems, or testing individual system components or combinations of components on multistage systems, by either paired increment or paired batch




NOTE 1—The "*x*" dimension shall be no less than three times the nominal top size of the coal but of sufficient width to enable quick and easy removal of all coal lying on the conveyor belt between the divider plates, including very fine coal.

**FIG. 1 Bias Test Stopped-Belt Divider**

experimental designs can result in atypical moisture losses because of a disturbance or disruption of routine operating conditions. Disturbance or disruption of routine operating conditions is generally related to one or more of the following: the time interval involved in extraction of increments, interruption of internal flow within the sampling system, and induced ventilation within the sampling system. Every effort must be made to minimize adverse effects of such factors.)

7.1.2 *Paired Increment Design*:

7.1.2.1 Paired increment procedures involve the collection of system increments and reference samples, which are paired for comparison purposes. Collect reference samples from the same area of the conveyor or as near as possible to the location where the corresponding sampling system's primary increment(s) is extracted so a close physical association is created. Some variations of the test design can be collecting one reference sample for each sampling system primary increment and bracketing the location from where the system's primary increment is withdrawn with two reference samples or if the system's primary increments are normally batched through the remainder of the sampling system. Another option may be collecting multiple system primary increments within the bracket of reference samples.

7.1.2.2 The paired increment experimental design requires intermittent operation of the coal handling and sampling systems because of the need to stop the conveyor to remove reference samples.

7.1.2.3 Operating the sampling system under the control of system logic is the preferred practice. This procedure involves operating under system logic until it initiates collection of a primary increment, then manually tripping the conveyor system by pushing the stop button to shut it down. This technique requires that only the main conveyor shut down, while the sampling system purges under the routine operating settings of system logic, and may or may not, shut down. System logic timers should continue to operate without interruption.

7.1.2.4 Collect reference increments using a systematic collection scheme.

7.1.2.5 The paired increment design can be used to test individual system components.

7.1.3 *Paired Test Batch Design*:

7.1.3.1 Test batch procedures involve collection of paired reference and system samples during a specific time or tonnage the sampling system is operated at predesignated routine operating settings. The reference sample may be composed of one or more reference increments while the number of system primary increments depends on the preselected sampling system operating settings relative to the time or tonnage interval.

7.1.3.2 The reference sample and mechanical system sample originate from the same test batch of coal.

7.1.3.3 Operate the sampling system at the operating settings preselected for the test, the same for every batch.

7.1.3.4 Use a random sampling scheme developed according to the requirements of Practice E 105. A random start followed by systematic selection of increments thereafter is acceptable practice.

7.1.3.5 The paired test batch design often is used to test the overall mechanical system.

7.1.4 *Intraphase Test Design*:

7.1.4.1 This testing pertains to obtaining the overall sampling system bias estimate by combining data from two or more separate test phases, one phase of which includes a reference sample. Each test phase obtains data on one or more components or subsystems. The data from the separate test phases are statistically combined for an estimate of the overall system bias. This approach is useful when interruptions to the sampling system would impose an experimentally induced moisture loss. The sampling system uses batch processing instead of linear processing. This approach is also useful when it is necessary to diagnose the cause of a bias discovered by one of the other test procedures.

NOTE 1—A typical two-phase test would include a paired increment test of the primary sample cutter, under static conditions, followed by a paired test batch design (under normal operating conditions), using surrogate samples. Surrogate samples may be obtained during the per-increment phase by extracting surrogate samples in the same manner, as they will be collected during the batch phase.

7.1.4.2 Phased testing takes advantage of the fact that mechanical coal sampling and on-line preparation is a linear process and the overall results of this linear process can be determined by separately investigating the individual parts. The data obtained from individual process parts is combined statistically to obtain an estimate of the overall systems performance.

7.1.4.3 The test data, from the separate phases, are combined by algebraically adding the mean differences and by obtaining an estimate of the overall standard deviation by summing the variances associated with each phase and taking the square root.

7.2 *Statistical Procedures*—The matched pairs experimental design of the test for bias reflects the underlying requirements for meaningful assessment of bias test data. This practice supports both parametric and nonparametric procedures, either of which can encompass univariate or multivariate statistical analysis for assessment and interpretation of results, both of which assume independence of individual differences. The distinction between parametric and nonparametric statistical analysis lies in the assumptions regarding the distribution of the population of differences. Parametric statistical analysis is predicated on a normal distribution.

7.2.1 *Wilcoxon Signed Rank Nonparametric Test, Nonparametric Analysis*—This test is based on creating a superset of the population of differences by differencing every possible combination of the observed differences and sorting them in ascending order. The median of this distribution is taken as the point estimate of bias. Two-sided confidence limits for univariate and multivariate analysis for up to five variables are established based on the Bonferroni inequality, using Table A2.12. Interpretation of the results depends on whether or not the confidence interval encompasses zero for the univariate case, and on whether or not the confidence interval of any one of the variables encompasses zero in the multivariate case.

7.2.2 *Student's t and Hotelling's $T^2$ Parametric Analysis*—The parametric method requires computation of the mean and the standard deviation of the differences of the variable in question for the univariate case or of each of the variables for the multivariate case. The mean(s) are taken as the point estimate(s) of bias. The confidence interval for the univariate case and the confidence regions for the multivariate case are established using the corresponding standard deviation and Hotelling's $T^2$ values. Interpretation of the results depends on whether or not the confidence interval falls within the predetermined tolerable bias region (see 8.2.3).

7.2.3 *Variance Addition for Intraphase Test*—Intraphase statistical analysis is conducted using Student's *t*-test for the paired difference between two means. Mean differences for each test phase are added to arrive at an overall mean difference for the system. The estimated standard deviation of the combined phases is obtained by addition of the corresponding variances of the phase tests and taking the square root. Interpretation of the results depends on whether or not the confidence interval encompasses zero.

## 8. Organization and Planning

8.1 *Data Required to Plan Test*:

8.1.1 Obtain information pertinent to operation of the mechanical sampling system so that detailed test procedures can be prepared.

8.1.2 Obtain the layout of the associated coal handling system including description of coal conveyor widths, belt speeds, troughing idler angles, coal flow rates, availability, and permissible conveyor stops and restarts.

8.1.3 Obtain complete sampling system operating information, including sample cutter widths, sample cutter operating intervals and velocities, sample extraction rates for each stage of sampling, sample crusher product top sizes, accessibility for sample collection, and typical lot sizes. Identify adjustments typically made to accommodate different lot sizes or other operating conditions. Sources of information can include design parameters, or physical measurements, or both.

NOTE 2—The condition of and operation of the sampling system can be determined before doing a bias test. It is recommended that the inspection be done in accordance with Guide D 4702, by personnel familiar with the operation of the mechanical sampling system and knowledgeable in ASTM standards.

8.1.4 Obtain a description of coals typically sampled. Include the nominal coal top size, typical quality profiles of the test coal, and a description of the type of coal preparation, such as washed, crushed run-of-mine, or blended coal.

8.2 *Select Test Conditions*—Make the following decisions and selections before the test:

8.2.1 *Selection of Test Coal*—If coals of different quality are available for use in the bias test, a selection of the specific coal(s) to be used must be made before collection of test samples. Efforts should be made to keep the coal quality as consistent as is practical during the test. The user of this practice is cautioned that a change in coal quality could invalidate the statistical results and that bias can change with coal quality.

8.2.2 *Selection of Analytical Test Parameters for the Test*:

8.2.2.1 Make the same analytical determinations on both the reference and system samples. Use the observed values for each of these coal characteristics to make inferences concerning system bias of the sampling system against the chosen reference. A bias test using this practice can be based on one or more characteristics measured for the test comparison. As many as five coal characteristics can be used when testing for bias using the statistical practices in 7.2.1.

8.2.2.2 The greater the number of coal characteristics used in the statistical inference for a fixed number of paired data sets the larger the confidence interval widths will be; thus, the user should give consideration to limiting the number of coal characteristics to those which would yield a reasonable evaluation of the sampling system. Arguments can be made that only determinations of moisture and dry ash are necessary for evaluating bias of a sampling system, and that it is unlikely bias of other coal characteristics would exist independent of bias of either moisture or dry ash.

8.2.2.3 The specific coal quality characteristics to be used in the bias test should be selected before the test.

8.2.3 *Selection of the Largest Tolerable Bias*—Select the largest tolerable bias limits (LTB) for the coal characteristics that will be used in the statistical analysis, or review the width of the confidence interval, or size of the confidence ellipsoid, and conclude if the test is sufficiently precise. Refer to the Annexes on statistical procedures for guidance.

8.2.4 *Selection of Sampler Operating Mode*—Sampler operation and coal transfer rate should not change during the course of the test. If the sampler has the ability to operate in different modes (different lot sizes, tonnage rates, time or mass basis, and so forth), the user must select the mode or modes of operation in which the sampler is to be tested.

8.2.5 *Selection of Collection Method*—Under Section 7, the user will need to select an increment collection method. The methods listed and described in 7.1.2 and 7.1.3 are collection of paired data on an increment basis and collection of paired data on a test-batch basis.

8.2.6 *Selection of Reference Method of Sample Collection*—Practice D 2234 lists several different methods for increment collection. Condition "A" (Stopped-Belt Cut), in which a full cross-section of coal is removed from the stopped main conveyor belt, is considered the reference method and is the highest order of sampling methods available. For the purpose of this practice, surrogate samples can be obtained from increments collected by methods other than stopping the main coal flow belt. Such surrogate samples, collected in accordance with Practice D 2234, Conditions "A" or "B" and when proven free of significant bias relative to reference samples may be considered acceptable for evaluation of a mechanical sampling system's components, excluding the primary cutter.

8.2.7 *Selection of Statistical Procedures*—Select a statistical procedure by which to evaluate the data from the bias test. The statistical procedures listed and described in Annex A2 are as follows: Wilcoxon Signed Rank Nonparametric test (nonparametric), and Hotelling's $T^2$ (parametric), and combined variance for intraphase tests. When there is only one coal characteristic used for the test, then the Hotelling's $T^2$ is equivalent to the Student's *t*-test.

8.2.8 *Selection of Number of Paired Data Sets*—In the absence of information on the variance of differences of the paired data sets, it is not possible to estimate, before the test, how many data sets are needed to detect a bias at the largest tolerable bias (LTB) chosen for the test. Recognizing this lack of information, it has been a common practice in the industry to initially collect between 20 and 40 sets of data, with the actual number being determined by perception of the variability of the coal and the use to be made of the test results. At any time during the test, analysis of current data collected can enable the user to determine if additional data sets are needed to reach specified test precision. Alternatively, if information is available on the sampling variance, or on the variance of differences of similarly collected paired samples from the test coal or similar coals, the information can be used to optimize test design. Using such information, Practice E 122 can be helpful in planning the number of paired data sets.

8.2.9 *Selection of Test Batch Size*—Select the batch size of coal that will be used for the data sets. For the paired increment test design, this can be the region of coal on the conveyor from which the reference and system sample increments are to be collected. For a paired test-batch design, this can be based on time or tonnage. In either test design, the batch size should be approximately the same throughout the entire test period. Test batch size should take into consideration the mass of retained system sample and the necessity to ensure that small retained samples are not adversely affected by the sample collection process (change in moisture, etc.). See Annex A1-Annex A3 for additional information regarding the selection of test-batch size.

8.2.10 *Selection of Number of Reference Increments/Samples per Test Batch of Coal*—Select the number of reference increments/samples per test batch. For a paired increment test design this can be one or more increments such that the reference sample is collected nearby or brackets the region of coal from which the system's increment(s) are to be obtained. For a paired test-batch design, one or more reference increments can be collected during the chosen batch interval. In general, the fewer the number of increments per test batch, the higher the variance of paired sample differences and the lower the power of the test for a given number of paired sets. For coal relatively uniform within individual test batches, only one or two reference increments might be adequate. For a coal with characteristics highly variable within individual test batches, it may be necessary to take more reference increments from each test batch.

8.2.11 *Selection of Reference Sample Collection Times and Preparing a Collection Schedule*:

8.2.11.1 Prepare a schedule for collection of reference increments from test batches before beginning the collection of bias test samples.

8.2.11.2 The reference increments should be collected from the test batch interval such that all coal within that test batch interval has an opportunity to be collected over the course of the test. Selection of timing for collection of the reference samples must be by a random method.

8.2.11.3 Operate the mechanical sampling system continuously during the processing of each test batch. If the test batch size is smaller than a lot, consider operating the system continuously while processing several consecutive test batches.

8.2.11.4 The test batch interval should include only the cumulative time during which coal is flowing.

8.2.11.5 Precautions should be taken, in the choice of increment collection times, that test sample collection minimally affects the coal flow through the sampling system.

8.2.11.6 Samples collected for a bias test should be collected in accordance with Practice D 2234 (Conditions" A" or "B").

8.3 *General Sample Handling*—As rapidly as possible, all test samples should be sealed in moisture proof containers, identified, weighed, and stored in a protected area before beginning the next test batch interval. Some coals are more susceptible to oxidation, which may require additional precautions such as vapor and gas impervious storage containers.

NOTE 3—Any unaccounted for moisture change in the test samples, that results from collection and handling, will show up as either an under or over-estimates of any moisture difference attributed to the sampling system.

8.3.1 *Preparation of Test Samples*:

8.3.1.1 Minimum final masses (after preparation), which conform to the limits specified in Method D 2013 are recommended. It is recognized that this will not be possible in all cases with the system sample. Samples with masses less than those specified in Method D 2013 shall only be used by mutual agreement of the interested parties. It must be recognized that the use of system sample masses, which substantially are less than those recommended can decrease the ability of the test to detect a bias or cause false detection of bias. Small sample masses could be detrimental especially to the determination of moisture bias if the samples are not handled with special care to preserve moisture.

8.3.1.2 Reweigh all reference increments, all reference sieve increments, and all system samples before combining, crushing, or dividing. List each weight in the bias test report.

8.3.1.3 Multiple reference samples, collected during a single test batch, can be physically composited, prepared, and analyzed or individually prepared and analyzed, and the weighted average analysis result of the individual samples used as the reference value.

8.3.1.4 Sample preparation can be performed wholly or in part either at the test site or at the testing laboratory. In either case, the sample preparation procedures shall be consistent with all test samples subject to conditions imposed by Practice D 2234 and Method D 2013.

8.3.1.5 Measure and include in the total moisture result the moisture condensation adhering to the interior of the sample containers used for transporting and storing samples.

8.3.2 *Laboratory Analysis of the Test Samples*:

8.3.2.1 Use consistent procedures for laboratory analysis throughout the test for bias.

8.3.2.2 Every effort should be made to analyze the test samples quickly to avoid deterioration of the test samples as a result of lengthy storage time.

8.3.2.3 All test samples from a test batch shall be concurrently processed and analyzed. The purpose is to minimize introducing systematic error resulting from differences in treatment during preparation and analysis.

8.3.2.4 Laboratory record keeping and quality control practices shall be in accordance with Guide D 4621. Record and report the results of all analytical determinations on each test sample.

8.4 *Information to Be Obtained and Reported*:

8.4.1 A log of test sample collection activities during sample collection should be kept. Include the following information:

8.4.1.1 Weather conditions, including temperature and state of precipitation.

8.4.1.2 Date, starting and ending time of the collection of each test sample.

8.4.1.3 The weight and number of increments comprising each test sample shall be recorded. It is recommended that all test samples be weighed before and after preparation to monitor preparation losses.

8.4.1.4 Identification of responsible personnel involved in the test sample collection process.

8.4.1.5 A general description of the origin and identification of the coal used during the test for bias.

8.4.1.6 Date, time, and description of failures of mechanical sampling equipment or coal-handling equipment, and duration of downtime.

8.4.1.7 Description of the sampling system and its operation during the test.

8.4.1.8 Description of test design, sample collection, sample handling, and statistical methods used for the test.

8.4.1.9 Analytical test results on each sample.

8.4.1.10 Results of all statistical analysis.

## 9. Keywords

9.1 coal sampling; mechanical sampling; statistical analysis

# ANNEXES

**(Mandatory Information)**

## A1. COLLECTION OF REFERENCE SAMPLES

### A1.1 Reference Samples

A1.1.1 A stopped-belt sample provides the best possible reference sample and is the accepted method for bias test reference sample collection.

A1.1.2 One or several reference increments can be collected per test set (see Table A1.1). If more than one reference increment is collected, the system samples can be compared to each individual reference increment or to the weighted average composite value of the combined reference increments. It is acceptable to take two reference increments that bracket the area from which the sampling system's primary increment will be withdrawn. When two or more reference increments are collected per belt stoppage, they should be collected simultaneously to minimize moisture differences between the two samples. Reference samples can be collected from a physical location upstream or downstream of the primary sample cutter of the mechanical sampling system.

**TABLE A1.1 Schedule for Collection of Stopped-Belt Increments**

NOTE 1—Test batch interval = 60 min
Stopped-belt increments per test batch = 3
$k = 60/3 = 20$ min
$U = 2(20) = 40$

| Test Batch | Random No. | $T_1$ | $T_2$ | $T_3$ |
|---|---|---|---|---|
| 1 | 16 | 8.0 | 28.0 | 48.0 |
| 2 | 37 | 18.5 | 38.5 | 58.5 |
| 3 | 2 | 1.0 | 21.0 | 41.0 |
| 4 | 35 | 17.5 | 37.5 | 57.5 |
| 5 | 26 | 13.0 | 33.0 | 53.0 |
| – | – | – | – | – |
| – | – | – | – | – |

A1.1.3 It is important that reference increments be collected rapidly, without delay, and that coal-handling conveyors are shut down and coal flow stopped during as short a time interval as is practical when moisture is being tested as one of the coal characteristics.

A1.1.4 If the location of the collection of reference increments is exposed to the environment, do not collect reference samples during precipitation or strong winds which potentially could cause sample loss or sample contamination.

### A1.2 Surrogate Samples

A1.2.1 A surrogate sample is composed of one or more increments collected from a coal stream within the mechanical sampling system. In bias testing, the surrogate samples normally are extracted from a coal stream at the discharge of a feeder. Surrogate samples can be collected from primary increment streams, as well as from other system sample streams.

A1.2.2 The mass of each surrogate increment should be in accordance with Practice D 2234, Conditions“ A” or “B.” Each surrogate increment should consist of a complete cross section of the flowing coal stream. Care must be taken to minimize the effects of surrogate sampling on components downstream of the sampling point. See Practice D 2234 for guidance on the number and mass of increments, which are needed for secondary increment collection.

## A2. STATISTICAL PROCEDURES

### A2.1 Nonparametric Statistical Procedure

A2.1.1 As many as five coal characteristics can be used when testing for bias by this procedure.

A2.1.2 *Step 1*—As illustrated by the example in Tables A2.1-A2.3, tabulate the reference observations and system observations for all coal characteristics. Then, compute and tabulate the individual differences between reference and system values for each test batch as shown in the columns of the tables. In computing differences, subtract each reference value from each corresponding system value, retaining the sign of the result. Compute the sample average of the reference values, the sample average of the system values, and the sample average of the differences for each coal characteristic.

A2.1.3 *Step 2*—For each coal characteristic, arrange the $n$ differences in ascending order, as illustrated in Table A2.4. Determine the sample median value. When there are an odd number of differences, the median is the $(n + 1)/2^{th}$ ordered difference. When there are an even number of differences, the median is the average of the $n/2^{th}$ difference and the $(n + 2)/2^{th}$ difference. For the example illustrated in Table A2.4, the sample median for each characteristic is the average of the $8^{th}$ and $9^{th}$ differences.

A2.1.4 *Step 3*—Prepare a graph of the differences by consecutive test batch number, beginning with the first batch, and ending with the $n^{th}$ batch. Plot the sample median of the differences as a straight line across the graph.

**TABLE A2.1 Observed Moisture Values**

| | SB Ref | Mech System | Sys-Ref | Above(+) Below(–) Median | Run No. |
|---|---|---|---|---|---|
| 1 | 5.66 | 5.66 | 0.00 | + | 1 |
| 2 | 9.22 | 9.29 | 0.07 | + | 1 |
| 3 | 8.52 | 8.52 | 0.00 | + | 1 |
| 4 | 9.00 | 8.75 | –0.25 | – | 2 |
| 5 | 8.47 | 8.38 | –0.09 | – | 2 |
| 6 | 8.46 | 8.62 | 0.16 | + | 3 |
| 7 | 9.26 | 9.28 | 0.02 | + | 3 |
| 8 | 9.24 | 9.49 | 0.25 | + | 3 |
| 9 | 8.58 | 8.44 | –0.14 | – | 4 |
| 10 | 5.85 | 5.80 | –0.05 | + | 5 |
| 11 | 6.15 | 5.77 | –0.38 | – | 6 |
| 12 | 9.03 | 9.01 | –0.02 | + | 7 |
| 13 | 9.68 | 9.40 | –0.28 | – | 8 |
| 14 | 11.25 | 10.08 | –1.17 | – | 8 |
| 15 | 9.41 | 9.20 | –0.21 | – | 8 |
| 16 | 5.75 | 5.66 | –0.09 | – | 8 |
| Sample Average | 8.346 | 8.209 | –0.136 | | |

**TABLE A2.2 Observed Dry Ash Values**

| | SB Ref | Mech System | Sys-Ref | Above(+) Below(–) Median | Run No. |
|---|---|---|---|---|---|
| 1 | 8.92 | 8.89 | –0.03 | – | 1 |
| 2 | 8.22 | 8.28 | 0.06 | + | 2 |
| 3 | 8.90 | 9.09 | 0.19 | + | 2 |
| 4 | 9.16 | 9.05 | –0.11 | – | 3 |
| 5 | 9.00 | 9.08 | 0.08 | + | 4 |
| 6 | 9.03 | 9.03 | 0.00 | – | 5 |
| 7 | 8.20 | 8.21 | 0.01 | – | 5 |
| 8 | 8.10 | 8.26 | 0.16 | + | 6 |
| 9 | 8.74 | 8.89 | 0.15 | + | 6 |
| 10 | 8.53 | 8.58 | 0.05 | – | 7 |
| 11 | 8.80 | 8.73 | –0.07 | – | 7 |
| 12 | 9.04 | 9.00 | –0.04 | – | 7 |
| 13 | 8.16 | 8.38 | 0.22 | + | 8 |
| 14 | 8.49 | 8.47 | –0.02 | – | 9 |
| 15 | 8.11 | 8.23 | 0.12 | + | 10 |
| 16 | 8.67 | 8.75 | 0.08 | + | 10 |
| Sample Average | 8.629 | 8.683 | 0.053 | | |

**TABLE A2.3 Observed Dry Sulfur Values**

| | SB Ref | Mech System | Sys-Ref | Above(+) Below(–) Median | Run No. |
|---|---|---|---|---|---|
| 1 | 2.788 | 2.790 | 0.002 | | |
| 2 | 2.858 | 2.895 | 0.037 | + | 1 |
| 3 | 2.703 | 2.705 | 0.002 | | |
| 4 | 2.690 | 2.685 | –0.005 | – | 2 |
| 5 | 2.688 | 2.740 | 0.052 | + | 3 |
| 6 | 2.698 | 2.700 | 0.002 | | |
| 7 | 2.805 | 2.805 | 0.000 | – | 4 |
| 8 | 2.843 | 2.855 | 0.012 | + | 5 |
| 9 | 2.673 | 2.655 | –0.018 | – | 6 |
| 10 | 2.705 | 2.700 | –0.005 | – | 6 |
| 11 | 2.745 | 2.740 | –0.005 | – | 6 |
| 12 | 2.630 | 2.605 | –0.025 | – | 6 |
| 13 | 2.850 | 2.875 | 0.025 | + | 7 |
| 14 | 2.890 | 2.905 | 0.015 | + | 7 |
| 15 | 2.758 | 2.775 | 0.017 | + | 7 |
| 16 | 2.788 | 2.790 | 0.002 | | |
| Sample Average | 2.757 | 2.764 | 0.007 | | |

A2.1.5 *Step 4 Test for Independent Differences*—To draw

TABLE A2.4 Ordered Sample Differences

| | Moisture | Dry Ash | Dry Sulfur |
|---|---|---|---|
| 1 | –1.17 | –0.11 | –0.025 |
| 2 | –0.38 | –0.07 | –0.018 |
| 3 | –0.28 | –0.04 | –0.005 |
| 4 | –0.25 | –0.03 | –0.005 |
| 5 | –0.21 | –0.02 | –0.005 |
| 6 | –0.14 | 0.00 | 0.000 |
| 7 | –0.09 | 0.01 | 0.002 |
| 8 | –0.09 | 0.05 | 0.002 |
| 9 | –0.05 | 0.06 | 0.002 |
| 10 | –0.02 | 0.08 | 0.002 |
| 11 | 0.00 | 0.08 | 0.012 |
| 12 | 0.00 | 0.12 | 0.015 |
| 13 | 0.02 | 0.15 | 0.017 |
| 14 | 0.07 | 0.16 | 0.025 |
| 15 | 0.16 | 0.19 | 0.037 |
| 16 | 0.25 | 0.22 | 0.052 |
| Median | –0.070 | 0.055 | 0.002 |

TABLE A2.5 Significance Values for Number of Runs

| $p = 1$ | | | | | |
|---|---|---|---|---|---|
| $n_1,n_2$ | $l,u$ | $n_1,n_2$ | $l,u$ | $n_1,n_2$ | $l,u$ |
| 3,5 | 3,– | 9,10 | 7,14 | 13,18 | 12,20 |
| 3,6 | 3,– | 9,11 | 7,14 | 13,19 | 12,21 |
| 3,7 | 3,– | 9,12 | 8,15 | 14,14 | 11,19 |
| 4,4 | 3,7 | 9,13 | 8,15 | 14,15 | 11,20 |
| 4,5 | 3,8 | 9,14 | 8,16 | 14,16 | 12,20 |
| 4,6 | 4,8 | 10,10 | 7,15 | 14,17 | 12,21 |
| 4,7 | 4,8 | 10,11 | 8,15 | 14,18 | 12,21 |
| 4,8 | 4,– | 10,12 | 8,16 | 14,19 | 13,22 |
| 5,5 | 4,8 | 10,13 | 9,16 | 14,20 | 13,22 |
| 5,6 | 4,9 | 10,14 | 9,16 | 15,15 | 12,20 |
| 5,7 | 4,9 | 10,15 | 9,17 | 15,16 | 12,21 |
| 5,8 | 4,10 | 11,11 | 8,16 | 15,17 | 12,21 |
| 5,9 | 5,10 | 11,12 | 9,16 | 15,18 | 13,22 |
| 6,6 | 4,10 | 11,13 | 9,17 | 15,19 | 13,22 |
| 6,7 | 5,10 | 11,14 | 9,17 | 15,20 | 13,23 |
| 6,8 | 5,11 | 11,15 | 10,18 | 16,16 | 12,22 |
| 6,9 | 5,11 | 11,16 | 10,18 | 16,17 | 13,22 |
| 6,10 | 6,11 | 11,17 | 10,18 | 16,18 | 13,23 |
| 7,7 | 5,11 | 12,12 | 9,17 | 16,19 | 14,23 |
| 7,8 | 5,12 | 12,13 | 10,17 | 16,20 | 14,24 |
| 7,9 | 6,12 | 12,14 | 10,18 | 17,17 | 13,23 |
| 7,10 | 6,12 | 12,15 | 10,18 | 17,18 | 14,23 |
| 7,11 | 6,13 | 12,16 | 11,19 | 17,19 | 14,24 |
| 7,12 | 7,13 | 12,17 | 11,19 | 17,20 | 14,24 |
| 8,8 | 6,12 | 12,18 | 11,20 | 18,18 | 14,24 |
| 8,9 | 6,13 | 13,13 | 10,18 | 18,19 | 15,24 |
| 8,10 | 7,13 | 13,14 | 10,19 | 18,20 | 15,25 |
| 8,11 | 7,14 | 13,15 | 11,19 | 19,19 | 15,25 |
| 8,12 | 7,14 | 13,16 | 11,20 | 19,20 | 15,26 |
| 9,9 | 7,13 | 13,17 | 11,20 | 20,20 | 16,26 |

Legend:
$p$ = number of coal characteristics tested,
$n_1$ = number of fewest like signs,
$n_2$ = number of most like signs,
$l$ = lower significance value, and
$u$ = upper significance value.

inference about system bias correctly using the procedures of this practice, the sample differences must be independent. When the hypothesis of independence is rejectable, the process used to draw inference about bias can be suspect or viewed as inconclusive.

A2.1.5.1 Determine the number of runs ($r$) for each characteristic by first subtracting the sample median value found in Step 2 from each difference. If the result is positive, record a plus sign, and if negative record a minus sign, as illustrated by the columns of Tables A2.1-A2.3. Ignore differences equal to the median. Runs are sequences of values all above the median, as indicated by a series of positive signs, or all below the median, as indicated by a series of negative signs. After the number of runs has been determined, count the number of positive signs and the number of negative signs.

A2.1.5.2 When there are not an equal number of positive and negative signs, let $n_1$ denote the smallest number of like signs (all positive or all negative), and let $n_2$ denote the largest number of like signs. Observe that often the number of positive and negative signs will be equal, in which case set both $n_1$ and $n_2$ equal to the common number of like signs.

A2.1.5.3 Let $p$ denote the number of coal characteristics used in the bias test. For each coal characteristic, obtain the lower and upper significance values $l$ and $u$ from Tables A2.5-A2.9 using the appropriate values of $n_1$ and $n_2$. If, for any tested coal characteristic:

$$r < l \text{ or } r > u$$

the data fails the test for independent differences and one concludes there is evidence the individual differences are not independently distributed. In all such cases in which the data fails the test for independent differences, include the following statement in the bias test report:

*There is evidence the series of differences between reference and system measurements are not independent; therefore, it is possible the conclusions reached below about system bias are not correctly drawn because the assumptions made for the statistical test procedure are not fulfilled.*

A2.1.5.4 When it is believed the reason is known why the measurements are not independent, state what is known in the bias test report. If the cause of lack of independence is unknown, include the following statement in the bias test report:

*It might prove useful to undertake investigations to determine a cause (or causes) for the apparent lack of independence of the differences.*

A2.1.5.5 For the illustrative data of Tables A2.1-A2.4, using Tables A2.5-A2.9, use of the procedure gives:

| | $r$ | $n_1$ | $n_2$ | $l$ | $u$ |
|---|---|---|---|---|---|
| Moisture | 8 | 8 | 8 | 5 | 13 |
| Dry ash | 10 | 8 | 8 | 5 | 13 |
| Dry sulfur | 7 | 6 | 6 | 4 | 10 |

For each of the coal characteristics given in the illustration, the number of runs falls between the lower and upper points of significance; thus, there is insufficient evidence to conclude the observations within each series are not independent.

A2.1.5.6 For each coal characteristic, denote the individual differences for the $n$ test batches of coal by $x_1, x_2, ..., x_i, ..., x_n$. Then calculate the following $n(n-1)/2$ different averages of two observations:

$$(x_1 + x_2)/2, (x_1 + x_3)/2, ..., (x_{n-1} + x_n)/2 \qquad \text{(A2.1)}$$

Include these $n(n-1)/2$ averages with the original $n$ differences, yielding a total of $w = n(n + 1)/2$ values, which are the Walsh Averages. Next, sort the Walsh Averages low to high, and index them consecutively by order. Table A2.10 illustrates sorted Walsh Averages for the 16 moisture differences given in Table A2.1.

TABLE A2.6 Significance Values for Number of Runs

| $p = 2$ | | | | | |
|---|---|---|---|---|---|
| $n_1,n_2$ | $l,u$ | $n_1,n_2$ | $l,u$ | $n_1,n_2$ | $l,u$ |
| 3,5 | 2,– | 9,10 | 6,15 | 13,18 | 11,21 |
| 3,6 | 3,– | 9,11 | 7,15 | 13,19 | 11,22 |
| 3,7 | 3,– | 9,12 | 7,15 | 14,14 | 10,20 |
| 4,4 | 3,– | 9,13 | 7,16 | 14,15 | 10,21 |
| 4,5 | 3,– | 9,14 | 8,16 | 14,16 | 11,21 |
| 4,6 | 3,8 | 10,10 | 7,15 | 14,17 | 11,22 |
| 4,7 | 3,8 | 10,11 | 7,15 | 14,18 | 11,22 |
| 4,8 | 4,– | 10,12 | 8,16 | 14,19 | 12,22 |
| 5,5 | 3,9 | 10,13 | 8,17 | 14,20 | 12,23 |
| 5,6 | 4,9 | 10,14 | 8,17 | 15,15 | 11,21 |
| 5,7 | 4,9 | 10,15 | 8,17 | 15,16 | 11,22 |
| 5,8 | 4,10 | 11,11 | 8,16 | 15,17 | 12,22 |
| 5,9 | 4,11 | 11,12 | 8,17 | 15,18 | 12,23 |
| 6,6 | 4,10 | 11,13 | 8,18 | 15,19 | 12,23 |
| 6,7 | 4,11 | 11,14 | 9,18 | 15,20 | 13,24 |
| 6,8 | 4,11 | 11,15 | 9,18 | 16,16 | 12,22 |
| 6,9 | 5,12 | 11,16 | 9,19 | 16,17 | 12,23 |
| 6,10 | 5,12 | 11,17 | 10,19 | 16,18 | 12,24 |
| 7,7 | 4,12 | 12,12 | 8,18 | 16,19 | 13,24 |
| 7,8 | 5,12 | 12,13 | 9,18 | 16,20 | 13,24 |
| 7,9 | 5,13 | 12,14 | 9,19 | 17,17 | 12,24 |
| 7,10 | 6,13 | 12,15 | 9,19 | 17,18 | 13,24 |
| 7,11 | 6,13 | 12,16 | 10,20 | 17,19 | 13,25 |
| 7,12 | 6,13 | 12,17 | 10,20 | 17,20 | 14,25 |
| 8,8 | 5,13 | 12,18 | 10,20 | 18,18 | 13,25 |
| 8,9 | 6,13 | 13,13 | 9,19 | 18,19 | 14,25 |
| 8,10 | 6,14 | 13,14 | 10,19 | 18,20 | 14,26 |
| 8,11 | 6,14 | 13,15 | 10,19 | 19,19 | 14,26 |
| 8,12 | 7,15 | 13,16 | 10,20 | 19,20 | 14,26 |
| 9,9 | 6,14 | 13,17 | 11,21 | 20,20 | 15,27 |

Legend:
$p$ = number of coal characteristics tested,
$n_1$ = number of fewest like signs,
$n_2$ = number of most like signs,
$l$ = lower significance value, and
$u$ = upper significance value.

TABLE A2.7 Significance Values for Number of Runs

| $p = 3$ | | | | | |
|---|---|---|---|---|---|
| $n_1,n_2$ | $l,u$ | $n_1,n_2$ | $l,u$ | $n_1,n_2$ | $l,u$ |
| 3,5 | –,– | 9,10 | 6,15 | 13,18 | 10,22 |
| 3,6 | –,– | 9,11 | 6,15 | 13,19 | 11,22 |
| 3,7 | 3,– | 9,12 | 7,16 | 14,14 | 10,20 |
| 4,4 | –,– | 9,13 | 7,16 | 14,15 | 10,21 |
| 4,5 | 3,8 | 9,14 | 7,17 | 14,16 | 10,22 |
| 4,6 | 3,– | 10,10 | 6,16 | 14,17 | 11,22 |
| 4,7 | 3,– | 10,11 | 7,16 | 14,18 | 11,22 |
| 4,8 | 3,– | 10,12 | 7,17 | 14,19 | 11,23 |
| 5,5 | 3,9 | 10,13 | 7,17 | 14,20 | 12,23 |
| 5,6 | 3,10 | 10,14 | 8,17 | 15,15 | 10,22 |
| 5,7 | 4,10 | 10,15 | 8,18 | 15,16 | 11,22 |
| 5,8 | 4,10 | 11,11 | 7,17 | 15,17 | 11,23 |
| 5,9 | 4,– | 11,12 | 8,17 | 15,18 | 11,23 |
| 6,6 | 4,10 | 11,13 | 8,18 | 15,19 | 12,24 |
| 6,7 | 4,11 | 11,14 | 8,18 | 15,20 | 12,24 |
| 6,8 | 4,11 | 11,15 | 9,19 | 16,16 | 11,23 |
| 6,9 | 4,12 | 11,16 | 9,19 | 16,17 | 12,23 |
| 6,10 | 5,12 | 11,17 | 9,19 | 16,18 | 12,24 |
| 7,7 | 4,12 | 12,12 | 8,18 | 16,19 | 12,24 |
| 7,8 | 5,12 | 12,13 | 8,19 | 16,20 | 13,25 |
| 7,9 | 5,13 | 12,14 | 9,19 | 17,17 | 12,24 |
| 7,10 | 5,13 | 12,15 | 9,20 | 17,18 | 12,25 |
| 7,11 | 5,14 | 12,16 | 9,20 | 17,19 | 13,25 |
| 7,12 | 6,14 | 12,17 | 10,20 | 17,20 | 13,26 |
| 8,8 | 5,13 | 12,18 | 10,21 | 18,18 | 13,25 |
| 8,9 | 5,14 | 13,13 | 9,19 | 18,19 | 13,26 |
| 8,10 | 6,14 | 13,14 | 9,20 | 18,20 | 14,26 |
| 8,11 | 6,15 | 13,15 | 10,20 | 19,19 | 14,26 |
| 8,12 | 6,15 | 13,16 | 10,21 | 19,20 | 14,27 |
| 9,9 | 6,14 | 13,17 | 10,21 | 20,20 | 14,28 |

Legend:
$p$ = number of coal characteristics tested,
$n_1$ = number of fewest like signs,
$n_2$ = number of most like signs,
$l$ = lower significance value, and
$u$ = upper significance value.

A2.1.6 *Step 5*—Determine the point estimate of the bias and the confidence interval.

A2.1.6.1 The point estimate of the bias is the median of the $w$ (Walsh Averages). If $w$ is an odd integer, the median is the $(w + 1)/2^{th}$ ordered value. If $w$ is an even integer, the median is the average of the $w/2^{th}$ value and the $(w + 2)/2^{th}$ value. For the illustrative data of Table A2.10, $w$ is the even integer 136; thus, the median is the average of the $136/2^{th}$, or the $68^{th}$ value, which is –0.090, and the $(136 + 2)/2^{th}$, or $69^{th}$ value, which is also –0.090. Therefore, the median and point estimate of the bias is –0.090.

A2.1.6.2 The confidence interval is given as the closed interval $[L_d, U_d]$, where:
$L_d$ = the $d$th smallest value of the Walsh Averages and
$U_d$ = the $d$th largest value of the Walsh Averages.

A2.1.6.3 The value of $d$ is read from Table A2.11 using the appropriate values of $n$, the number of test batches, and $p$, the number of coal characteristics tested. Using the illustrative data of Table A2.10, moisture, dry ash, and dry sulfur were tested with 16 batches of coal; thus, $p = 3$, $n = 16$, and the table value of d is the integer 22. Therefore, $L_d$ is the 22nd value of Table A2.10 or –0.265, and $U_d$ is the $[n (n + 1)/2] + 1 - d =$ 115th value or 0.035. The closed confidence interval for moisture then is [–0.265, 0.035].

## A2.2 Interpretation of Nonparametric Results and Adequacy of Data

A2.2.1 Concluding statements for the test are made as follows:

A2.2.1.1 *Statement A*—If a chance error which, before the test had a maximum probability of occurring equal to no more than about 1 in 20, did not occur, biases of mechanically collected samples against reference samples lie within the closed intervals given below.

| | |
|---|---|
| Moisture | $L_d (m) \leq \beta(m) \leq U_d (m)$ |
| Dry ash | $L_d (da) \leq \beta(da) \leq U_d (da)$ |
| (continue with other characteristics tested) | |

where $\beta(m)$ and $\beta(da)$ denote moisture bias and dry ash bias, respectively.

Use Statement B or Statement C (below), as appropriate.

A2.2.1.2 *Statement B*—The confidence interval for each coal characteristic includes the value zero; thus, this test offers insufficient evidence to reject a hypothesis of no bias of system samples against reference samples.

A2.2.1.3 *Statement C*—The confidence interval(s) for (insert here the name of one or more characteristics) does not (do not) cover the value zero; thus, there is evidence of bias of mechanical system samples against reference samples. The sample estimate of the bias is (report the point estimate(s) as determined by A2.1.6.1).

A2.2.2 For the example test data given in Tables A2.1-A2.3,

TABLE A2.8 Significance Values for Number of Runs

| | | $p$ = 4 | | | |
|---|---|---|---|---|---|
| $n_1$,$n_2$ | $l$,$u$ | $n_1$,$n_2$ | $l$,$u$ | $n_1$,$n_2$ | $l$,$u$ |
| 3,5 | –,– | 9,10 | 6,15 | 13,18 | 10,22 |
| 3,6 | –,– | 9,11 | 6,16 | 13,19 | 10,22 |
| 3,7 | –,– | 9,12 | 6,16 | 14,14 | 9,21 |
| 4,4 | –,– | 9,13 | 7,16 | 14,15 | 10,21 |
| 4,5 | –,8 | 9,14 | 7,17 | 14,16 | 10,22 |
| 4,6 | 3,– | 10,10 | 6,16 | 14,17 | 10,22 |
| 4,7 | 3,– | 10,11 | 7,16 | 14,18 | 11,23 |
| 4,8 | 3,– | 10,12 | 7,17 | 14,19 | 11,23 |
| 5,5 | 3,9 | 10,13 | 7,17 | 14,20 | 11,24 |
| 5,6 | 3,10 | 10,14 | 8,18 | 15,15 | 10,22 |
| 5,7 | 3,10 | 10,15 | 8,18 | 15,16 | 10,23 |
| 5,8 | 4,– | 11,11 | 7,17 | 15,17 | 11,23 |
| 5,9 | 4,– | 11,12 | 7,18 | 15,18 | 11,24 |
| 6,6 | 3,11 | 11,13 | 8,18 | 15,19 | 11,24 |
| 6,7 | 4,11 | 11,14 | 8,19 | 15,20 | 12,24 |
| 6,8 | 4,12 | 11,15 | 8,19 | 16,16 | 11,23 |
| 6,9 | 4,12 | 11,16 | 9,19 | 16,17 | 11,24 |
| 6,10 | 4,12 | 11,17 | 9,20 | 16,18 | 12,24 |
| 7,7 | 4,12 | 12,12 | 8,18 | 16,19 | 12,25 |
| 7,8 | 4,12 | 12,13 | 8,19 | 16,20 | 12,25 |
| 7,9 | 5,13 | 12,14 | 8,19 | 17,17 | 12,24 |
| 7,10 | 5,13 | 12,15 | 9,20 | 17,18 | 12,25 |
| 7,11 | 5,14 | 12,16 | 9,20 | 17,19 | 12,25 |
| 7,12 | 5,14 | 12,17 | 9,21 | 17,20 | 13,26 |
| 8,8 | 5,13 | 12,18 | 10,21 | 18,18 | 12,26 |
| 8,9 | 5,14 | 13,13 | 8,20 | 18,19 | 13,26 |
| 8,10 | 5,14 | 13,14 | 9,20 | 18,20 | 13,27 |
| 8,11 | 6,15 | 13,15 | 9,21 | 19,19 | 13,27 |
| 8,12 | 6,15 | 13,16 | 10,21 | 19,20 | 14,27 |
| 9,9 | 6,14 | 13,17 | 10,22 | 20,20 | 14,28 |

Legend:
$p$ = number of coal characteristics tested,
$n_1$ = number of fewest like signs,
$n_2$ = number of most like signs,
$l$ = lower significance value, and
$u$ = upper significance value.

TABLE A2.9 Significance Values for Number of Runs

| | | $p$ = 5 | | | |
|---|---|---|---|---|---|
| $n_1$,$n_2$ | $l$,$u$ | $n_1$,$n_2$ | $l$,$u$ | $n_1$,$n_2$ | $l$,$u$ |
| 3,5 | –,– | 9,10 | 6,15 | 13,18 | 10,22 |
| 3,6 | –,– | 9,11 | 6,16 | 13,19 | 10,23 |
| 3,7 | –,– | 9,12 | 6,16 | 14,14 | 9,21 |
| 4,4 | –,– | 9,13 | 7,17 | 14,15 | 9,22 |
| 4,5 | –,9 | 9,14 | 7,17 | 14,16 | 10,22 |
| 4,6 | 3,– | 10,10 | 6,16 | 14,17 | 10,23 |
| 4,7 | 3,– | 10,11 | 6,17 | 14,18 | 10,23 |
| 4,8 | 3,– | 10,12 | 7,17 | 14,19 | 11,23 |
| 5,5 | 3,9 | 10,13 | 7,18 | 14,20 | 11,24 |
| 5,6 | 3,10 | 10,14 | 7,18 | 15,15 | 10,22 |
| 5,7 | 3,10 | 10,15 | 8,18 | 15,16 | 10,23 |
| 5,8 | 3,– | 11,11 | 7,17 | 15,17 | 11,23 |
| 5,9 | 4,– | 11,12 | 7,18 | 15,18 | 11,24 |
| 6,6 | 3,11 | 11,13 | 7,18 | 15,19 | 11,24 |
| 6,7 | 4,11 | 11,14 | 8,19 | 15,20 | 12,25 |
| 6,8 | 4,12 | 11,15 | 8,19 | 16,16 | 11,23 |
| 6,9 | 4,12 | 11,16 | 8,20 | 16,17 | 11,24 |
| 6,10 | 4,– | 11,17 | 9,20 | 16,18 | 11,24 |
| 7,7 | 4,12 | 12,12 | 8,18 | 16,19 | 12,25 |
| 7,8 | 4,13 | 12,13 | 8,19 | 16,20 | 12,25 |
| 7,9 | 5,13 | 12,14 | 8,20 | 17,17 | 11,25 |
| 7,10 | 5,14 | 12,15 | 9,20 | 17,18 | 12,25 |
| 7,11 | 5,14 | 12,16 | 9,21 | 17,19 | 12,26 |
| 7,12 | 5,14 | 12,17 | 9,21 | 17,20 | 13,26 |
| 8,8 | 5,13 | 12,18 | 9,21 | 18,18 | 12,26 |
| 8,9 | 5,14 | 13,13 | 8,20 | 18,19 | 13,26 |
| 8,10 | 5,14 | 13,14 | 9,20 | 18,20 | 13,27 |
| 8,11 | 6,15 | 13,15 | 9,21 | 19,19 | 13,27 |
| 8,12 | 6,15 | 13,16 | 9,21 | 19,20 | 13,28 |
| 9,9 | 5,15 | 13,17 | 10,22 | 20,20 | 14,28 |

Legend:
$p$ = number of coal characteristics tested,
$n_1$ = number of fewest like signs,
$n_2$ = number of most like signs,
$l$ = lower significance value, and
$u$ = upper significance value.

the concluding statements are as follows:

A2.2.2.1 If a chance error with a maximum probability before the test of no more than about 1 out of 20 of occurring, did not occur, biases of mechanically collected samples against reference samples lie within the closed intervals given below:

| | |
|---|---|
| Moisture | $-0.265 \leq \beta(m) \leq 0.035$ |
| Dry ash | $-0.020 \leq \beta(da) \leq 0.120$ |
| Dry sulfur | $-0.005 \leq \beta(ds) \leq 0.020$ |

where $\beta(m)$, $\beta(da)$, and $\beta(ds)$ represent moisture, dry ash, and dry sulfur biases, respectively.

A2.2.2.2 The confidence interval for each coal characteristic includes the value zero; thus, this test offers insufficient evidence to reject a hypothesis of no bias of system samples against reference samples.

A2.2.3 It can turn out that one or more confidence intervals given by Statement A of A2.2.1.1 will be too wide for the test to be useful. For example, if the closed interval for moisture turns out to be [–0.450, +0.115], whereas a moisture bias of –0.40, is of practical significance, one will conclude there is a need to reduce the width of the confidence interval. Note that the width of the interval is inversely proportional (approximately) to the square root of the number of paired differences. If the variance of paired differences is so large that it is not economically feasible to reduce the width of the confidence interval sufficiently by continuing the test, one might be able to repeat the test under more favorable conditions. Taking more reference increments per test batch, or employing some other means to reduce the variance of differences between the reference and system measurements should be evaluated if the aforementioned situation does occur and complicates the evaluation process.

## A2.3 Parametric Statistical Procedures

A2.3.1 The statistical procedures produce, in effect, a list of all bias values that are plausible given the experimental bias test data. This list is in the form of a one-dimensional confidence interval if only one characteristic is selected. If several characteristics are selected, the list is in the form of an $n$-dimensional confidence region. This confidence interval or region can be checked against a largest tolerable bias (LTB) interval or region, which represents what is acceptable to both producer and consumer, and which may be agreed upon before the bias test is performed.

A2.3.2 *Using Student's t-Statistic for Bias Test Data*—Student's $t$-statistic may be used to quantify the uncertainty in bias tests when only one quality parameter, for example, ash, sulfur, Btu, specific size fraction, and so forth is measured. A confidence interval is constructed for the unknown bias that summarizes all the information contained in the bias test data pairs. The following steps outline the procedure.

A2.3.2.1 Before performing the bias test, all interested parties should agree to an interval whose upper and lower

**TABLE A2.10 Sorted Walsh Averages for Moisture Illustrative Data**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | –1.170 | 35 | –0.190 | 69 | –0.090 | 103 | 0.000 |
| 2 | –0.775 | 36 | –0.185 | 70 | –0.080 | 104 | 0.000 |
| 3 | –0.725 | 37 | –0.185 | 71 | –0.070 | 105 | 0.000 |
| 4 | –0.710 | 38 | –0.180 | 72 | –0.070 | 106 | 0.010 |
| 5 | –0.690 | 39 | –0.175 | 73 | –0.070 | 107 | 0.010 |
| 6 | –0.655 | 40 | –0.170 | 74 | –0.070 | 108 | 0.010 |
| 7 | –0.630 | 41 | –0.170 | 75 | –0.070 | 109 | 0.010 |
| 8 | –0.630 | 42 | –0.165 | 76 | –0.065 | 110 | 0.020 |
| 9 | –0.610 | 43 | –0.155 | 77 | –0.060 | 111 | 0.020 |
| 10 | –0.595 | 44 | –0.150 | 78 | –0.060 | 112 | 0.025 |
| 11 | –0.585 | 45 | –0.150 | 79 | –0.055 | 113 | 0.035 |
| 12 | –0.585 | 46 | –0.150 | 80 | –0.055 | 114 | 0.035 |
| 13 | –0.575 | 47 | –0.150 | 81 | –0.050 | 115 | 0.035 |
| 14 | –0.550 | 48 | –0.140 | 82 | –0.045 | 116 | 0.035 |
| 15 | –0.505 | 49 | –0.140 | 83 | –0.045 | 117 | 0.045 |
| 16 | –0.460 | 50 | –0.140 | 84 | –0.045 | 118 | 0.055 |
| 17 | –0.380 | 51 | –0.135 | 85 | –0.045 | 119 | 0.055 |
| 18 | –0.330 | 52 | –0.130 | 86 | –0.045 | 120 | 0.070 |
| 19 | –0.315 | 53 | –0.130 | 87 | –0.035 | 121 | 0.070 |
| 20 | –0.295 | 54 | –0.125 | 88 | –0.035 | 122 | 0.080 |
| 21 | –0.280 | 55 | –0.125 | 89 | –0.035 | 123 | 0.080 |
| 22 | –0.265 | 56 | –0.115 | 90 | –0.035 | 124 | 0.080 |
| 23 | –0.260 | 57 | –0.115 | 91 | –0.025 | 125 | 0.080 |
| 24 | –0.250 | 58 | –0.115 | 92 | –0.025 | 126 | 0.090 |
| 25 | –0.245 | 59 | –0.115 | 93 | –0.025 | 127 | 0.100 |
| 26 | –0.235 | 60 | –0.110 | 94 | –0.020 | 128 | 0.115 |
| 27 | –0.235 | 61 | –0.105 | 95 | –0.015 | 129 | 0.115 |
| 28 | –0.230 | 62 | –0.105 | 96 | –0.015 | 130 | 0.125 |
| 29 | –0.215 | 63 | –0.105 | 97 | –0.010 | 131 | 0.125 |
| 30 | –0.210 | 64 | –0.095 | 98 | –0.010 | 132 | 0.135 |
| 31 | –0.210 | 65 | –0.095 | 99 | –0.010 | 133 | 0.160 |
| 32 | –0.200 | 66 | –0.090 | 100 | –0.010 | 134 | 0.160 |
| 33 | –0.195 | 67 | –0.090 | 101 | 0.000 | 135 | 0.205 |
| 34 | –0.190 | 68 | –0.090 | 102 | 0.000 | 136 | 0.250 |

**TABLE A2.11 Counting Value *d***

| *n* | *p* = 1 | *p* = 2 | *p* = 3 | *p* = 4 | *p* = 5 |
|---|---|---|---|---|---|
| 10 | 9 | 6 | 5 | 4 | 4 |
| 11 | 11 | 9 | 7 | 6 | 6 |
| 12 | 14 | 11 | 10 | 9 | 8 |
| 13 | 18 | 14 | 12 | 11 | 10 |
| 14 | 22 | 18 | 16 | 14 | 14 |
| 15 | 26 | 21 | 19 | 18 | 17 |
| 16 | 30 | 25 | 22 | 20 | 18 |
| 17 | 35 | 29 | 26 | 24 | 22 |
| 18 | 41 | 34 | 31 | 28 | 26 |
| 19 | 47 | 39 | 36 | 33 | 31 |
| 20 | 53 | 45 | 41 | 38 | 36 |
| 21 | 60 | 51 | 47 | 44 | 42 |
| 22 | 67 | 58 | 53 | 49 | 47 |
| 23 | 74 | 64 | 59 | 56 | 54 |
| 24 | 82 | 72 | 66 | 63 | 60 |
| 25 | 90 | 79 | 74 | 70 | 67 |
| 26 | 98 | 87 | 81 | 77 | 74 |
| 27 | 107 | 96 | 90 | 85 | 82 |
| 28 | 116 | 105 | 98 | 93 | 90 |
| 29 | 126 | 114 | 107 | 102 | 99 |
| 30 | 137 | 124 | 116 | 111 | 108 |
| 31 | 147 | 134 | 126 | 120 | 117 |
| 32 | 159 | 144 | 136 | 130 | 127 |
| 33 | 170 | 155 | 147 | 141 | 137 |
| 34 | 182 | 166 | 158 | 151 | 147 |
| 35 | 195 | 178 | 169 | 162 | 158 |
| 36 | 208 | 190 | 181 | 174 | 169 |
| 37 | 221 | 203 | 193 | 186 | 181 |
| 38 | 235 | 216 | 206 | 198 | 193 |
| 39 | 249 | 229 | 219 | 211 | 206 |
| 40 | 264 | 243 | 232 | 224 | 219 |

bounds represent the limits to a negligible bias. This interval will be referred to as the LTB and will be based on operational and economic considerations. If the bias can be reasonably shown to fall within the LTB interval, the sampling system will be considered to be practically unbiased. For the bias test to have a chance to lead to the correct action, it is important that the LTB interval accurately define the limits of an acceptable bias. An LTB that is too wide will increase the chances of allowing a seriously biased sampling system to go uncorrected, while an LTB that is too narrow will increase the chances of making unnecessary sampling system modifications. Once the LTB is chosen, it should not be revised after the data is collected just so the sampling system can be declared acceptable.

A2.3.2.2 Let $x_r(i)$ represent the quality of the $i^{th}$ reference (stopped-belt) increment and $x_a(i)$ represent the quality of the $i^{th}$ system sample and $d(i) = x_a(i) - x_r(i)$ the corresponding difference, for $i = 1, 2, ..., n$ where $n$ is the number of sample pairs. Calculate the following statistics:

$$\bar{d} = \frac{\sum_{i=1}^{n} d(i)}{n} \tag{A2.2}$$

$$s^2 = \frac{\sum_{i=1}^{n} d(i)^2 - n\bar{d}^2}{(n-1)} \tag{A2.3}$$

$$s_{\bar{d}} = \frac{s}{\sqrt{n}} \tag{A2.4}$$

A2.3.2.3 Here, $\bar{d}$ is the estimated bias and mean difference between the reference and system samples, $s^2$ is the estimated variance of these differences, $s$ (the square root of $s^2$) is the estimated standard deviation of the differences, and $s$ is the estimated standard deviation of the mean difference $\bar{d}$. ($s_{\bar{d}}$ also is referred to as the estimated standard error of the mean difference $\bar{d}$.)

A2.3.2.4 Calculate either the 95 or the 99 % confidence interval:

$$\bar{d} \pm t_{1-\alpha/2,n-1}\, s_{\bar{d}} \tag{A2.5}$$

where $\alpha$ is the univariate risk that the confidence interval does not cover the unknown level of bias, $(1 - \alpha)100$ % is the percent confidence interval, $n - 1$ is the value of the degrees of freedom of the estimate, and $t$ is read from a table of Student's $t$.

A2.3.2.5 Compare the calculated confidence interval with the LTB interval. There are three possible results:

(*a*) If the confidence interval falls entirely within the LTB interval declare the bias to be negligible and the sampling system acceptable.

(*b*) If the confidence interval falls entirely outside the LTB interval, declare the bias nonnegligible and the sampling system unacceptable.

(*c*) If the LTB interval and the confidence interval overlap, declare the bias test inconclusive. In this case, there is not enough evidence to conclude the sampling system to be acceptable and more bias test increments should be collected or a new bias test with more sets of data must be performed to resolve the problem.

A2.3.3 *Using Hotelling's $T^2$ Test for Multivariate Bias Test Data*—When more than one measurement is made on each increment and bias results are desired for all quality parameters, the Student's $t$-test should no longer be used. The multivariate analog of the Student's $t$-test is known as Hotelling's $T^2$ statistic. This multivariate method can be used to produce a confidence region that correctly defines all plausible parameter bias values, those that are supported by the actual bias test data.

A2.3.3.1 The following four steps correspond to those for Student's $t$-test:

(*a*) A multidimensional LTB region should be established for the $p$ quality parameters for which biases are to be estimated. If the number of quality parameters is less than or equal to three, the region can be graphed. For example, if $p = 2$, then the region could be the area inside a rectangle or an ellipse. If $p = 3$, then the region could be a prism or ellipsoid. Ellipsoidal regions exclude jointly large biases (regardless of direction, that is, negative or positive), while rectangular and prismatic regions do not. (When $p = 1$, the region is a line interval so that the choice does not exist when using Student's $t$ method.) Ellipsoidal regions of $p$ dimensions are better suited to being LTB regions. In general, the $p$-dimensional LTB is given by:

$$\sum x_i^2/m_i^2 \leq 1 \quad \text{(A2.6)}$$

where $x_i$ is the coordinate for the $i^{th}$ parameter and $m_i$ is the largest tolerable bias (irrespective of sign) for the $i^{th}$ parameter. As an example, suppose $p = 2$ and bias estimates are desired for ash and Btu. Suppose further that a negligible ash bias is within ±0.15 % ash and a negligible Btu bias is within ±10 Btu. If a rectangular LTB region were used, this would be specified as follows:

–0.15 % ash ≤ ash bias ≤ 0.15 % ash and –10 Btu ≤ Btu bias ≤ 10 Btu

NOTE A2.1—This would allow the ash and Btu bias simultaneously to be as bad as– 0.15 % ash and –10 Btu, respectively. Other extremes also are possible. If a more appropriate two-dimensional ellipsoidal LTB region (inequality A2.7) were used, the LTB region would be specified as follows:

$$x_1^2/(0.15\ \%\ \text{ash})^2 + x_2^2/(10\ \text{Btu})^2 \leq 1 \quad \text{(A2.7)}$$

where $x_1$ is the coordinate for the ash bias and $x_2$ is the coordinate for the Btu bias.

This elliptical region does not include simultaneously large biases in both parameters. For example, not only would a simultaneous 0.15 % ash bias and 10 Btu bias not be tolerated, even a simultaneous 0.11 % ash bias and 7.5 Btu bias would not be tolerated. As the ash bias approaches 0.15 % ash, the Btu bias must approach 0 Btu. Conversely, as the Btu bias approaches 10 Btu, the ash bias must approach 0 % ash.

(*b*) Let $x_r(i, j)$ represent the $j^{th}$ quality of the $i^{th}$ reference sample and $x_a(i, j)$ represent the $j^{th}$ quality of the $i^{th}$ actual system sample, and $d(i, j) = x_a(i, j) - x_r(i, j)$ the corresponding difference for $i = 1, 2, ..., n$ where $n$ is the number of increment pairs and $j = 1, 2, ..., p$ where $p$ is the number of quality parameters. Calculate the following statistics for each quality parameter $j$:

$$\bar{d}_j = \frac{\sum_{i=1}^{n} d(i, j)}{n} \quad \text{(A2.8)}$$

$$s_j^2 = \frac{\sum_{i=1}^{n} d(i, j)^2 - n\bar{d}_j^2}{(n-1)} \quad \text{(A2.9)}$$

Here $\bar{d}_j$ is the mean for the $j^{th}$ quality parameter and $s_j^2$ is the variance for the $j^{th}$ quality parameter. Except for the extra subscript $j$ used to denote a specific quality parameter, Eq A2.8 is equivalent to Eq A2.2, and Eq A2.9 is equivalent to Eq A2.3. Also, for every pair of quality parameters $j$ and $j'$, $j \neq j'$, compute:

$$s_{jj} = \frac{\sum_{i=1}^{n} d(i,j)d(i,j') - n\bar{d}_j\bar{d}_{j'}}{(n-1)} \quad \text{(A2.10)}$$

Here $s_{jj}$ is the covariance between quality $j$ and quality $j'$ where $j \neq j'$. Create the following $(p \times 1)$ mean vector:

$$\bar{D}^t = [\bar{d}_1 \bar{d}_2 \ldots \bar{d}_p] \quad \text{(A2.11)}$$

where $D^t$ denotes the transpose of a vector or matrix $D$. Create the following $(p \times p)$ variance-covariance matrix $S$:

$$S = \begin{bmatrix} s_1^2 & s_{12} & \text{—} & s_{1p} \\ s_{21} & s_2^2 & \text{—} & s_{2p} \\ | & | & \backslash & | \\ s_{p1} & s_{p2} & \text{—} & s_p^2 \end{bmatrix} \quad \text{(A2.12)}$$

NOTE A2.2—$s_{jj'} = s_{j'j}$. In all, $p$ variances and $p\ (p - 1)/2$, covariances must be calculated. The correlation $r_{jj'}$ between quality parameter $j$ and $j'$ can be estimated as follows:

$$r_{jj'} = \frac{r_{jj'}}{s_j s_{j'}} \quad \text{(A2.13)}$$

The percentage of variance accounted for (or explained) by parameter $j$ for parameter $j'$ is given by:

$$100\ \%\ x\ (r_{jj'})^2 \quad \text{(A2.14)}$$

(*c*) Let $X^t = [x_1\ x_2\ \ldots\ x_n]$ represent the coordinates for the parameter biases. Then, a $100(1 - \alpha)$ % confidence region is given by:

$$n(\bar{D} - X)'\ S^{-1}\ (\bar{D} - X) \leq T^2_{p,n-1}(\infty) = \frac{(n-1)p}{(n-p)} F_{p,n-p}(\infty) \quad \text{(A2.15)}$$

where $S^{-1}$ is the matrix inverse of $S$ and $T^2_{p,n-1}(\infty)$ is taken from the $T^2$ table for $p$ and $n - 1$ df, or alternatively, $F_{p,n-p}(\infty)$ is taken from the more easily available $F$ table with $p$ and $n - p$ df.

Alternatively, without using matrix computations, the leftmost expression of inequality Eq A2.15 can be calculated as follows:

$$n[\sum \delta_j (\bar{d}_j - x_j)^2 + 2\sum\sum \delta_{jj'}(\bar{d}_j - x_j)(\bar{d}_{j'} - x_{j'})] \quad \text{(A2.16)}$$

where $j \neq j'$ and

$$S^{-1} = \begin{bmatrix} \delta_1 & \delta_{12} & \text{—} & \delta_{1p} \\ \delta_{21} & \delta_2 & \text{—} & \delta_{2p} \\ | & | & \backslash & | \\ \delta_{p1} & \delta_{p2} & \text{—} & \delta_p \end{bmatrix} \quad \text{(A2.17)}$$

The remaining task is to determine the relationship between the confidence region and the LTB region.

(*d*) If $p$ is 2 or 3, the confidence region and the LTB region can be graphed and easily visualized. If $p$ is 4 or greater, it can mathematically be determined whether the regions overlap or not but is more difficult to visualize. As for the univariate case, there are three possible results:

(*1*) If the confidence region falls entirely within the LTB region, declare the bias to be negligible and the sampling system acceptable.

(*2*) If the confidence region falls entirely outside the LTB region, declare the bias non-negligible and the sampling system is unacceptable. The sampling system must be scrutinized to determine the cause of the bias and then corrected and retested.

(*3*) If the LTB region and the confidence region overlap, declare the bias test inconclusive. In this case, there is not enough evidence to conclude the sampling system is acceptable and more bias test increments must be collected or a new bias test with more increments must be performed to resolve the problem.

A2.3.3.2 *Example of Hotelling's $T^2$ Statistics*—A bias test was planned for a mechanical coal sampling system at the mine loadout. Before the bias test was performed, the producer and consumer agreed that the sampling system was acceptable (produced a negligible bias) if it could be shown that almost definitely:

$$x_1^2 / (0.15\ \%\ \text{ash})^2 + x_2^2 / (10\ \text{Btu})^2 \leq 1 \quad \text{(A2.18)}$$

where $x_1$ is the coordinate for ash bias, and $x_2$ is the coordinate for Btu bias and is based on inequality Eq A2.6. It was further agreed that this would be accomplished if the 95 % confidence region falls entirely within the above LTB region. The data shown in Table A2.12 was collected when the bias test was performed. Dry ash differences are denoted by subscript $j = 1$, and as-received Btu differences by subscript $j = 2$. The number of increment pairs $n$ is equal to 30. Then, using Eq A2.8, the mean differences $\bar{d}_1 = -0.46$ and $\bar{d}_2 = 46$, and therefore, by Eq A2.11.

$$\bar{D}' = [-0.46\ 46] \quad \text{(A2.19)}$$

**TABLE A2.12 Actual Minus Stopped Belt Differences Collected in a Bias Test**

| Pair | Actual–Stopped Belt Dry Ash (%) | As-Received Btu |
|---|---|---|
| 1 | –1.13 | 114 |
| 2 | –0.81 | 182 |
| 3 | –0.01 | 10 |
| 4 | 0.07 | 58 |
| 5 | –0.37 | 4 |
| 6 | –0.64 | 57 |
| 7 | 0.06 | 53 |
| 8 | –0.67 | 196 |
| 9 | –0.82 | 108 |
| 10 | –0.61 | –40 |
| 11 | –1.24 | 209 |
| 12 | 0.00 | 50 |
| 13 | –0.25 | 77 |
| 14 | –0.44 | 66 |
| 15 | –0.79 | 140 |
| 16 | –1.39 | 115 |
| 17 | –1.26 | 177 |
| 18 | –0.10 | –71 |
| 19 | –0.53 | 151 |
| 20 | 0.20 | –32 |
| 21 | –0.10 | –31 |
| 22 | –0.39 | 75 |
| 23 | –1.05 | 121 |
| 24 | –1.16 | 78 |
| 25 | 0.58 | –123 |
| 26 | 0.16 | –54 |
| 27 | –1.54 | 121 |
| 28 | 0.85 | –207 |
| 29 | 0.02 | –58 |
| 30 | –0.37 | –165 |

Using Eq A2.9, the variances are $s_1^2 = 0.35$ and $s_2^2 = 112\ 651$. Using Eq A2.10, the covariance $s_{12} = -47.5$. (The correlation between dry ash and as received Btu can be calculated as the covariance divided by the square root of the product of the variances. This yields $r_{12} = r_{21} = -0.76$.)

The variance-covariance matrix (Eq A2.12) then is:

$$s = \begin{bmatrix} 0.35 & -47.50 \\ -47.50 & 11\ 265.10 \end{bmatrix} \quad \text{(A2.20)}$$

and its inverse (Eq A2.17)

$$s^{-1} = \begin{bmatrix} 6.679\ 434 & 0.028\ 164 \\ 0.028\ 164 & 0.000\ 208 \end{bmatrix} \quad \text{(A2.21)}$$

Using inequality Eq A2.15, the 95 % confidence region is:

$$30[6.679\ 434(-0.46 - x_1)^2 + 2(0.028\ 164)(-0.46 - x_1)(46 - x_2) \quad \text{(A2.22)}$$

$$+ 0.000\ 208(46 - x_2)^2] \leq 6.92$$

given that $T^2_{2,29}$ (0.05) = 6.92.

Alternatively, $F_{2,28}$ (0.05) = 3.34 so that (from right-hand side of inequality Eq A2.15 is as follows:

$$\frac{(30-1)\ 2}{(30-2)} \times 3.34 = 6.92) \quad \text{(A2.23)}$$

The LTB region (inequality Eq A2.18) and the 95 % confidence region (inequality A2.23) are computed. The fact that these regions do not overlap and the confidence region falls entirely outside of the LTB region indicates almost without doubt the sampling system is unacceptable according to the test agreed to by the producer and the consumer.

For illustration purposes, suppose that instead of the previous bias test, a bias test was planned which would only measure as received Btu. Suppose further that the following LTB interval was agreed upon before the start of the bias test:

$$-10\ \text{Btu} \leq \text{Btu bias} \leq 10\ \text{Btu}$$

and that the producer and the consumer would consider the bias to be negligible if the 95 % confidence interval falls entirely within the LTB interval.

Assume that the same as received Btu difference data was collected. Using Eq A2.2, the mean Btu difference $\bar{d}$ is 46 Btu. Using Eq A2.3, the variance $s^2$ is 11 265.1 (Btu)$^2$. The standard error for the mean difference $s_d$ (Eq A2.4) is 19.38 (Btu)$^2$. The 95 % confidence interval (Eq A2.5) then is as follows:

$$46 \pm 2.045(19.38)$$

or

$$6.37\ \text{to}\ 85.63\ \text{Btu}$$

In this case, the bias test is declared inconclusive because the LTB interval and the 95 % confidence interval overlap from 6.37 to 10 Btu. More increment pairs must be collected so that a conclusion can be reached as to whether the sampling system is almost without doubt either acceptable or unacceptable when measuring only Btu.

A2.3.4 *Discussion of Results*:

A2.3.4.1 Part of the same data set was used to illustrate the computations for two different bias tests. The univariate bias test only used the as-received Btu data and reached a very different conclusion (bias test inconclusive) compared to the multivariate bias test (sampling system unacceptable). In part, the reason for the difference is the additional dry ash information available to the multivariate test. Although had just dry ash been collected and submitted to a univariate bias test procedure, both the univariate and multivariate procedures would have happened to declare the sampling system unacceptable. A properly applied multivariate bias test will always yield more information, and hence, be more likely to lead to the correct action than a univariate bias test procedure applied to a single parameter. (The only exception would be the unlikely case where the several parameters are perfectly correlated either –1 or +1 correlations. In this case, there really is only one parameter and the univariate and multivariate test procedures must reach the same conclusion.)

A2.3.4.2 Suppose both dry ash and as-received Btu were both collected and univariate tests applied separately to each parameter. This would be incorrect for several reasons. First, the correlation (in this case a moderate correlation of –0.76) between the parameters would be ignored which results in a loss of information. The correlation or lack of it between the measured parameters supplies information on which pairs of values are more or less likely. The separate repeated univariate tests implicitly generate a rectangular confidence region, which is too small, and the actual confidence percentage is much less than the percentage used to construct the individual intervals. That is, if separate repeated 95 % confidence intervals are constructed for $p$ parameters, the actual joint confidence region may only be a 50 % region, not 95 %. Also, because the correlation is ignored, the orientation of the confidence region would incorrectly include some less plausible bias values and exclude more plausible values. Finally, the appropriate multidimensional confidence region supplied by the multivariate $T^2$ method gives the most complete information available on the likely value of the joint bias. The use of repeated univariate confidence intervals in place of the appropriate multivariate confidence region will lead to a greater likelihood of incorrect conclusions and also to the greater likelihood of inconclusive bias tests, even while the multivariate confidence region would have lead to a definitive conclusion on sampler acceptability.

A2.3.4.3 It is important to notice what happens if the true (but unknown) bias is near either the upper or lower LTB limit (either inside or outside the interval or region). In this case, the number of increment pairs needed to resolve whether the sampling device is, or is not acceptable may become impractically large. The conclusion, simply, is that the sampling device is close to the limit and it will be virtually impossible to determine whether it is technically acceptable. The conservative approach, at some point, would be to investigate the sampling system to try to discover flaws in the design or operation of the device, or both, that have inadvertently been overlooked, and then correct them rather than continue bias testing the existing system. The system can be subjected to another bias test after corrections are made. Because the bias test is used to estimate an unknown potential bias, there is no way to insure avoiding this problem.

A2.3.4.4 The essential determination in a bias test is to see if the plausible values of the bias (as defined by a confidence interval or region) fall entirely within what is acceptable (as defined by the LTB interval or region). When measuring more than one quality parameter (excluding size fractions), it is not very meaningful to try to place individual simultaneous confidence intervals around each parameter bias estimate (although simultaneous individual confidence intervals can easily be constructed). The multivariate confidence region is a single confidence region that correctly handles inferences for more than one parameter. Essentially, the multivariate confidence region is a list of all plausible parameter biases given the bias test data actually collected. The shape and orientation of the ellipsoidal region and its location with respect to the LTB region provides all the information that is available from the bias test procedure.

A2.3.4.5 Both the univariate Student's $t$ method and the multivariate Hotelling's $T^2$ method make some statistical assumptions about the data collected. Both methods assume that the difference observations are statistically independent. In the multivariate method, the difference observations are vectors. The vector observations are assumed statistically independent. This means, along with the normality assumption to be discussed in the Appendix, that there is no correlation across observations. Both methods also assume that the data are normally distributed. In the multivariate case, a multivariate normal distribution is assumed. Both the statistical independence assumption and the normality assumption can and should be examined, although the details are beyond the scope of this practice.

### A2.4 Combined-Variance for Intraphase Test

A2.4.1 Intraphase statistical analysis is conducted using Student's $t$-test for the paired difference between two means. This analysis tests the hypothesis that the mean difference between pairs of related elements is statistically equivalent to zero. This statistic assumes normally distributed differences. If the calculated $t$ value is equal to or exceeds the value of $t$ for the appropriate confidence level and degrees of freedom taken from a $t$ table, then the difference is determined to be statistically significant.

A mechanical coal-sampling system is essentially a linear process. This means that the difference between a product's characteristic at the beginning and the end of the process must equal the sum of the differences in the characteristic between all intermediate steps. Algebraically, this is shown by the following equation:

$$(y_1 - y_2) + (y_2 - y_4) = y_1 - y_4 \quad \text{(A2.24)}$$

A2.4.1.1 Because differences in values are being dealt with and not the absolute values themselves, this analysis can be extended across phase boundaries, assuming that equivalent reference points are used on either side of the phase boundary. In these two-phased tests of mechanical coal-sampling systems, the assumption is made that the primary subsample and the surrogate reference subsample, even though taken at different times, are equivalent; therefore, the mean difference between the endpoints of the "A" phase can be added to the mean difference between the endpoints of the "B" phase to arrive at an overall mean difference for the system. It is recognized, however, that any statistical analysis is incomplete without some measure of confidence. The classic method of determining confidence levels involves calculating the range about the mean determined by multiplying a factor (the Student's $t$) by the standard error of the estimate. The standard error of each phase can be determined, but as they involve the square root function, they can not be added directly because:

Given that the Standard Error ≈ Standard Deviation =

$$\sqrt{\text{variance}}, \sqrt{\text{variance}_a} + \sqrt{\text{variance}_b} = \sqrt{\text{variance}_a + \text{variance}_b} \quad \text{(A2.25)}$$

However, variances, as the square of the standard deviation, are additive:

$$(\text{variance}_a) + (\text{variance}_b) = (\text{variance}_a + \text{variance}_b) \quad \text{(A2.26)}$$

Therefore, a technique known as pooling variances is used to develop a pooled variance and thereby, a pooled standard deviation. This pooled standard deviation, when divided by the square root of the number of sets, becomes the standard error, which may be multiplied by the appropriate *t* value and used to estimate the confidence interval. Deriving an appropriate denominator for weighting variances and for calculating the standard error becomes somewhat problematic. If sample sizes are equivalent, and if the population variances are equivalent, it is permissible to calculate the degrees of freedom by summing the sample sizes and subtracting the number of phases **(1)**.[4] If sample sizes are unequal, a modification to this method is required. The reader is referred to a statistics textbook for a discussion of the welch approximation or the satterthwaite approximation.

[4] The boldface numbers in parentheses refer to the list of references at the end of this standard.

## A3. SELECTION OF TEST BATCH SIZE

A3.1 The following criteria are recommended for selecting the mass of coal from which reference and system samples are to be drawn.

A3.2 The laboratory sample prepared from the mechanical system sample collected during processing of a test batch should be approximately equal in mass to the laboratory sample prepared from the reference sample; thus, the test batch size must be large enough to assure that the system collects a sample of sufficient size to meet this condition.

A3.3 When stopped-belt increments are used as the reference, it is recommended that the minimum time interval between collection of successive reference increments from a test batch of coal be at least 20 min to avoid interrupting the sampling system's moisture equilibrium. This 20-min time period may be decreased if it does not adversely affect the sampling system's moisture equilibrium. Before beginning the test, approval of the conveyor belt-stopping procedure should be obtained from management responsible for the material handling system.

A3.4 Where the time for processing a lot of coal is on the order of 1 h or less, consideration should be given to making the test batch size equal to the lot size.

# APPENDIX

**(Nonmandatory Information)**

## X1. THEORETICAL CONSIDERATIONS AND SOURCES OF FORMULAS AND TABLES

X1.1 The Bonferroni inequality was used in preparing Tables A2.5-A2.9 and Table A2.11. Let $\Theta = (\Theta_1, \Theta_2, ..., \Theta_p)$ be a $(p \times 1)$ vector of parameters. Using the Bonferroni inequality, one may construct separate two-sided confidence intervals for each of *p* parameters, each with confidence coefficient $100(1 - \alpha/p)$. Then, if $A_1$ denotes the event that the interval for $\Theta_1$ includes the actual value of $\Theta_1$, it follows that the probability that every interval covers the value of the parameter it estimates is at least $(1 - \alpha)$. Thus, the family confidence coefficient is at least $100(1 - \alpha)\%$ **(1)**. For Tables A2.5-A2.9 and Table A2.11, the value 0.95, or 95 %, was chosen as a uniform value for the maximum two-sided family confidence coefficient.

X1.2 The test for independent differences, A2.1.4, Step 4, is the standard test for randomness based on the number of runs above and below the sample median **(2,3)**. Values for the probability distribution of the total number of runs for samples of various sizes used to prepare Tables A2.5-A2.9 were taken from Ref **(4)**.

X1.3 A nonparametric test based on an assumption of symmetry is used to draw conclusions about bias **(5)**. The reasons for using this type of test as opposed to a test based on normal theory are as follows:

X1.3.1 It is observed that while the results of some test data indicate the assumption of normally distributed paired differences may be a reasonable assumption, data from other tests indicate a distribution heavier in the tails than normal **(6)**. In general, the data available from a specific test will be insufficient to exercise good judgment concerning the shape of the distribution at hand or for determining an appropriate normalizing transformation; thus, a test robust to departures from normality is preferred, and only symmetry is assumed. In the event the differences for a given test turn out to have a parent normal distribution, the loss through use of the more robust approach is small. In particular, the asymptotic relative efficiency of the nonparametric coverage for the center of symmetry compared to the coverage based on one- and two-sample *t* statistics is $3/\pi = 0.955$ for normally distributed populations.

The asymptotic relative efficiency is generally greater than one for distributions whose tails are longer than those of a normally distributed population **(7)**.

X1.4 The procedure described in A2.1.6, Step 5, used to determine the joint confidence intervals, is the standard Tukey procedure based on the Wilcoxon Signed Rank Test **(8)**. In preparation of Table A2.11, for *n* equal to 15 or less, values of the probability distribution are taken from Table A4 of Ref **(8)**. For *n* greater than 15, the following approximation is used:

$$d = n(n + 1)/4 - z_{\alpha/2}[n(n + 1)(2n + 1)/24]^{1/2} \quad \text{(X1.1)}$$

where $z_{\alpha/2}$ is the point on a standard normal with probability $\alpha/2$ above it.

## REFERENCES


(**1**) Alt, F. B., "Bonferroni Inequalities and Intervals," in *Encyclopedia of Statistical Sciences*, Vol 1, Wiley, New York, 1982, pp. 294–300.
(**2**) Walsh, J. E., *Handbook of Nonparametric Statistics*, D. Van Nostrand, Princeton, NJ, 1962, p. 64.
(**3**) Johnson, N. L., and Leone, F. C., *Statistics and Experimental Design*, Vol 1, 2nd Ed., Wiley, New York, 1977, Chapter 9.
(**4**) Beyer, W. H., Ed., *CRC Handbook of Tables for Probability and Statistics*, 2nd Ed., CRC Press, Boca Raton, FL, 1968, Table X6.
(**5**) Randles, R. H., "Tukey's Confidence Interval for Location," in *Encyclopedia of Statistical Sciences*, Vol 9, Wiley, New York, 1988, pp. 359–361.
(**6**) Rose, C. D., "A New Approach to Bias Testing of Mechanical Coal Sampling Systems," *The Journal of Coal Quality*, Vol 9, No. 4, Oct.–Dec. 1990, pp. 124–129.
(**7**) Noether, G. E., "Nonparametric Confidence Intervals," in *Encyclopedia of Statistical Sciences*, Vol 6, Wiley, New York, 1985, pp. 319–324.
(**8**) Hollander, M., and Wolfe, D. A., *Nonparametric Statistical Methods*, Wiley, New York, 1973, pp. 35–38, and Table A4.