Designation: D6122 – 15

Standard Practice for Validation of the Performance of Multivariate Online, At-Line, and Laboratory Infrared Spectrophotometer Based Analyzer Systems[1]

This standard is issued under the fixed designation D6122; 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.

INTRODUCTION

Operation of a laboratory or process stream analyzer system typically involves four sequential activities. *(1)* **Analyzer Calibration**—When an analyzer is initially installed, or after major maintenance has been performed, diagnostic testing is performed to demonstrate that the analyzer meets the manufacturer's specifications and historical performance standards. These diagnostic tests may require that the analyzer be adjusted so as to provide predetermined output levels for certain reference materials. *(2a)* **Correlation, where analyzer and Primary Test Method (PTM) measure the same material**—Once the diagnostic testing is completed, process stream samples are analyzed using both the analyzer system and the corresponding PTM. A mathematical function is derived that relates the analyzer output to the PTM. The application of this mathematical function to an analyzer output produces a Predicted Primary Test Method Result (PPTMR) for the same material. *(2b)* **Correlation, where analyzer measures a material which is subjected to treatment before being measured by the PTM**—Once the diagnostic testing is completed, the process stream samples are analyzed by the analyzer system. The same samples are subjected to a consistent treatment, and the treated samples are analyzed by the PTM. A mathematical function is derived that related the analyzer output for the untreated sample to the Primary Test Method Result (PTMR) for the treated material. The application of the mathematical function to the analyzer output for the untreated material produces a PPTMR for the treated material. *(3)* **Probationary Validation**—Once the relationship between the analyzer output and PTMRs has been established, a probationary validation is performed using an independent but limited set of materials that were not part of the correlation activity. This probationary validation is intended to demonstrate that the PPTMRs agree with the PTMRs to within user-specified requirements for the analyzer system application. *(4)* **General and Continual Validation**—After an adequate number of PPTMRs and PTMRs have been accrued on materials that were not part of the correlation activity, a comprehensive statistical assessment is performed to demonstrate that the PPTMRs agree with the PTMRs to within user-specified requirements. Subsequent to a successful general validation, quality assurance control chart monitoring of the differences between PPTMR and PTMR is conducted during normal operation of the process analyzer system to demonstrate that the agreement between the PPTMRs and the PTMRs established during the General Validation is maintained. This practice deals with the third and fourth of these activities.

"Correlation where analyzer measures a material which is subjected to treatment before being measured by the PTM" as outlined in this practice is intended primarily to be applied to biofuels where the biofuel material is added at a terminal or other facility and not included in the process stream material sampled by the analyzer at the basestock manufacturing facility. The "treatment" shall be a constant percentage addition of the biofuels material to the basestock material.

1. Scope*

1.1 This practice covers requirements for the validation of measurements made by laboratory or process (online or at-line) near- or mid-infrared analyzers, or both, used in the calculation of physical, chemical, or quality parameters (that is, properties) of liquid petroleum products and fuels. The properties are calculated from spectroscopic data using multivariate modeling methods. The requirements include verification of adequate instrument performance, verification of the applicability of the calibration model to the spectrum of the sample under test, and

*A Summary of Changes section appears at the end of this standard

Copyright © ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959. United States

verification that the degree of agreement between the results calculated from the infrared measurements and the results produced by the PTM used for the development of the calibration model meets user-specified requirements. When there is adequate variation in property level, the statistical methodology of Practice D6708 is used to provide general validation of this equivalence over the complete operating range of the analyzer. For cases where there is inadequate property variation, methodology for level specific validation is used.

1.1.1 For some applications, the analyzer and PTM are applied to the same material. The application of the multivariate model to the analyzer output (spectrum) directly produces a PPTMR for the same material for which the spectrum was measured. The PPTMRs are compared to the PTMRs measured on the same materials to determine the degree of agreement.

1.1.2 For other applications, the material measured by the analyzer system is subjected to a consistent treatment prior to being analyzed by the PTM. The application of the multivariate model to the analyzer output (spectrum) produces a PPTMR for the treated material. The PPTMRs based on the analyzer outputs are compared to the PTMRs measured on the treated materials to determine the degree of agreement.

1.2 Performance Validation is conducted by calculating the precision and bias of the differences between results from the analyzer system (or subsystem) produced by application of the multivariate model, (such results are herein referred to as PPTMRs), versus the PTMRs for the same sample set. Results used in the calculation are for samples that are not used in the development of the multivariate model. The calculated precision and bias are statistically compared to user-specified requirements for the analyzer system application.

1.2.1 For analyzers used in product release or product quality certification applications, the precision and bias requirement for the degree of agreement are typically based on the site or published precision of the PTM.

NOTE 1—In most applications of this type, the PTM is the specification-cited test method.

1.2.2 This practice does not describe procedures for establishing precision and bias requirements for analyzer system applications. Such requirements must be based on the criticality of the results to the intended business application and on contractual and regulatory requirements. The user must establish precision and bias requirements prior to initiating the validation procedures described herein.

1.3 This practice does not cover procedures for establishing the calibration model (correlation) used by the analyzer. Calibration procedures are covered in Practices E1655 and references therein.

1.4 This practice is intended as a review for experienced persons. For novices, this practice will serve as an overview of techniques used to verify instrument performance, to verify model applicability to the spectrum of the sample under test, and to verify equivalence between the parameters calculated from the infrared measurement and the results of the primary test method measurement.

1.5 This practice teaches and recommends appropriate statistical tools, outlier detection methods, for determining whether the spectrum of the sample under test is a member of the population of spectra used for the analyzer calibration. The statistical tools are used to determine if the infrared measurement results in a valid property or parameter estimate.

1.6 The outlier detection methods do not define criteria to determine whether the sample or the instrument is the cause of an outlier measurement. Thus, the operator who is measuring samples on a routine basis will find criteria to determine that a spectral measurement lies outside the calibration, but will not have specific information on the cause of the outlier. This practice does suggest methods by which instrument performance tests can be used to indicate if the outlier methods are responding to changes in the instrument response.

1.7 This practice is not intended as a quantitative performance standard for the comparison of analyzers of different design.

1.8 Although this practice deals primarily with validation of infrared analyzers, the procedures and statistical tests described herein are also applicable to other types of analyzers which employ multivariate models.

1.9 *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:*[2]

D1265 Practice for Sampling Liquefied Petroleum (LP) Gases, Manual Method

D3764 Practice for Validation of the Performance of Process Stream Analyzer Systems

D4057 Practice for Manual Sampling of Petroleum and Petroleum Products

D4177 Practice for Automatic Sampling of Petroleum and Petroleum Products

D6299 Practice for Applying Statistical Quality Assurance and Control Charting Techniques to Evaluate Analytical Measurement System Performance

D6708 Practice for Statistical Assessment and Improvement of Expected Agreement Between Two Test Methods that Purport to Measure the Same Property of a Material

[1] This practice is under the jurisdiction of ASTM Committee D02 on Petroleum Products, Liquid Fuels, and Lubricants and is the direct responsibility of Subcommittee D02.25 on Performance Assessment and Validation of Process Stream Analyzer Systems.

Current edition approved June 1, 2015. Published February 2016. Originally approved in 1997. Last previous edition approved in 2010 as D6122 – 13. DOI: 10.1520/D6122-15.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

D7278 Guide for Prediction of Analyzer Sample System Lag Times
D7453 Practice for Sampling of Petroleum Products for Analysis by Process Stream Analyzers and for Process Stream Analyzer System Validation
D7808 Practice for Determining the Site Precision of a Process Stream Analyzer on Process Stream Material
D7717 Practice for Preparing Volumetric Blends of Denatured Fuel Ethanol and Gasoline Blendstocks for Laboratory Analysis
E131 Terminology Relating to Molecular Spectroscopy
E275 Practice for Describing and Measuring Performance of Ultraviolet and Visible Spectrophotometers
E456 Terminology Relating to Quality and Statistics
E932 Practice for Describing and Measuring Performance of Dispersive Infrared Spectrometers
E1421 Practice for Describing and Measuring Performance of Fourier Transform Mid-Infrared (FT-MIR) Spectrometers: Level Zero and Level One Tests
E1655 Practices for Infrared Multivariate Quantitative Analysis
E1866 Guide for Establishing Spectrophotometer Performance Tests
E1944 Practice for Describing and Measuring Performance of Laboratory Fourier Transform Near-Infrared (FT-NIR) Spectrometers: Level Zero and Level One Tests

3. Terminology

3.1 *Definitions:*

3.1.1 For definitions of terms and symbols relating to IR spectroscopy, refer to Terminology E131.

3.1.2 For definitions of terms and symbols relating to multivariate calibration, refer to Practices E1655.

3.1.3 For definitions of terms relating to statistical quality control, refer to Practice D6299 and Terminology E456.

3.1.4 *between-method reproducibility (R_{XY}), n*—a quantitative expression of the random error associated with the difference between two results obtained by different operators using different apparatus and applying the two methods *X* and *Y*, respectively, each obtaining a single result on an identical test sample, when the methods have been assessed and an appropriate bias-correction has been applied in accordance with this practice; it is defined as the 95 % confidence limit for the difference between two such single and independent results. **D6708**

3.1.5 *control limits, n*—limits on a control chart which are used as criteria for signaling the need for action, or for judging whether a set of data does or does not indicate a state of statistical control. **E456**

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *action limit, n*—the limiting value from an instrument performance test, beyond which the analyzer is expected to produce potentially invalid results.

3.2.2 *analyzer, n*—all piping, hardware, computer, software, instrumentation and calibration model required to automatically perform analysis of a process or product stream.

3.2.3 *analyzer calibration, n*—see *multivariate calibration*.

3.2.4 *analyzer site precision, n*—a statistical measure of the expected long-term variability of analyzer results for samples whose spectra are neither outliers, nor nearest neighbor inliers.

3.2.5 *analyzer model, n*—see *multivariate model*.

3.2.6 *analyzer repeatability, n*—a statistical measure of the expected short-term variability of results produced by the analyzer for samples whose spectra are neither outliers nor nearest neighbor inliers.

3.2.7 *analyzer result, n*—the numerical estimate of a physical, chemical, or quality parameter produced by applying the calibration model to the spectral data collected by the analyzer.

3.2.8 *analyzer validation test, n*—see *validation test*.

3.2.9 *calibration transfer, n*—a method of applying a multivariate calibration developed on one analyzer to a different analyzer by mathematically modifying the calibration model or by instrument standardization.

3.2.10 *check sample, n*—a single, pure liquid hydrocarbon compound or a known, reproducible mixture of liquid hydrocarbon compounds whose spectrum is constant over time such that it can be used in a performance test.

3.2.11 *exponentially weighted moving average control chart, n*—a control chart based on the exponentially weighted average of individual observations from a system; the observations may be the differences between the analyzer result, and the result from the primary test method.

3.2.12 *individual observation control chart, n*—a control chart of individual observations from a system; the observations may be the differences between the analyzer result and the result from the primary test method.

3.2.13 *inlier, n*—see *nearest neighbor distance inlier*.

3.2.14 *inlier detection methods, n*—statistical tests which are conducted to determine if a spectrum resides within a region of the multivariate calibration space, which is sparsely populated.

3.2.15 *in-line probe, n*—a spectrophotometer cell installed in a process pipe or slip stream loop and connected to the analyzer by optical fibers.

3.2.16 *instrument, n*—spectrophotometer, associated electronics and computer, spectrophotometer cell and, if utilized, transfer optics.

3.2.17 *instrument standardization, n*—a procedure for standardizing the response of multiple instruments such that a common multivariate model is applicable for measurements conducted by these instruments, the standardization being accomplished by way of adjustment of the spectrophotometer hardware or by way of mathematical treatment of the collected spectra.

3.2.18 *line sample, n*—a process or product sample which is withdrawn from a sample port in accordance with Practices D1265, D4057, D4177, or D7453, whichever is applicable, during a period when the material flowing through the analyzer is of uniform quality and the analyzer result is essentially constant.

3.2.19 *moving range of two control chart, n*—a control chart that monitors the change in the absolute value of the difference between two successive differences of the analyzer result minus the result from the primary test method.

3.2.20 *multivariate calibration, n*—an analyzer calibration that relates the spectrum at multiple wavelengths or frequencies to the physical, chemical, or quality parameters.

3.2.21 *multivariate model, n*—a multivariate, mathematical rule or formula used to calculate physical, chemical, or quality parameters from the measured infrared spectrum.

3.2.22 *nearest neighbor distance inlier, n*—a spectrum residing within a gap in the multivariate calibration space, the result for which is subject to possible interpolation error.

3.2.23 *optical background, n*—the spectrum of radiation incident on a sample under test, typically obtained by measuring the radiation transmitted through the spectrophotometer cell when no sample is present, or when an optically thin or nonabsorbing liquid is present.

3.2.24 *optical reference filter, n*—an optical filter or other device which can be inserted into the optical path in the spectrophotometer or probe producing an absorption spectrum which is known to be constant over time, such that it can be used in place of a check or test sample in a performance test.

3.2.25 *outlier detection limits, n*—the limiting value for application of an outlier detection method to a spectrum, beyond which the spectrum represents an extrapolation of the calibration model.

3.2.26 *outlier detection methods, n*—statistical tests which are conducted to determine if the analysis of a spectrum using a multivariate model represents an interpolation of the model.

3.2.27 *outlier spectrum, n*—a spectrum whose analysis by a multivariate model represents an extrapolation of the model.

3.2.28 *performance test, n*—a test that verifies that the performance of the instrument is consistent with historical data and adequate to produce valid results.

3.2.29 *physical correction, n*—a type of post-processing where the correction made to the numerical value produced by the multivariate model is based on a separate physical measurement of, for example, sample density, sample path length, or particulate scattering.

3.2.30 *post-processing, v*—performing a mathematical operation on an intermediate analyzer result to produce the final result, including correcting for temperature effects, adding a mean property value of the analyzer calibration, and converting into appropriate units for reporting purposes.

3.2.31 *prediction deviations* (Δ), *n*—calculated differences (including algebraic sign) between predicted primary test method result and primary test result, defined as (PPTMR – PTMR).

3.2.31.1 *Discussion*—This is also referred to as prediction residuals in Practice D6708.

3.2.32 *pre-processing, v*—performing mathematical operations on raw spectral data prior to multivariate analysis or model development, such as selecting wave length regions, correcting for baseline, smoothing, mean centering, and assigning weights to certain spectral positions.

3.2.33 *Primary Test Method (PTM), n*—the analytical procedure used to generate the reference values against which the analyzer is both calibrated and validated; Practices E1655 uses the term reference method in place of the term primary test method.

3.2.34 *Primary Test Method Result (PTMR), n*—test result produced from an ASTM or other established standard test method that is accepted as the reference measure of a property.

3.2.35 *Predicted Primary Test Method Result (PPTMR), n*—result from the analyzer system, after application of any necessary correlation, that is interpreted as predictions of what the primary test method results would have been, if it was conducted on the same material.

3.2.36 *process analyzer system, n*—see *analyzer*.

3.2.37 *process analyzer validation samples, n*—see *validation samples*.

3.2.38 *spectrophotometer cell, n*—an apparatus which allows a liquid hydrocarbon to flow between two optical surfaces which are separated by a fixed distance, the sample path length, while simultaneously allowing light to pass through the liquid.

3.2.39 *test sample, n*—a process or product sample, or a mixture of process or product samples, which has a constant spectrum for a finite time period, and which can be used in a performance test; test samples and their spectra are generally not reproducible in the long term.

3.2.40 *transfer optics, n*—a device which allows movement of light from the spectrophotometer to a remote spectrophotometer cell and back to the spectrophotometer; transfer optics include optical fibers or other optical light pipes.

3.2.41 *validation samples, n*—samples that are used to compare the analyzer results to the primary test method results through the use of control charts and statistical tests; validation samples used in the initial validation may be line and test samples, whereas validation samples used in the periodic validation are line samples.

3.2.42 *validated result, n*—a result produced by the analyzer for a sample whose spectrum is neither an outlier nor a nearest neighbor inlier that is equivalent, within control limits to the result expected from the primary test method, so that the result can be used instead of the direct measurement of the sample by the primary test method.

3.2.43 *validation test, n*—a test performed on a validation sample that demonstrates that the result produced by the analyzer and the result produced by the primary test method are equivalent to within control limits.

4. Summary of Practice

4.1 This section describes, in summary form, the steps involved in the validation of an infrared analyzer over the long term. Before this practice may be undertaken, certain preconditions shall be satisfied. The preconditions are described in Section 7. This practice consists of four major procedures.

4.2 Each time a spectrum of a sample is collected using a laboratory or process analyzer, statistical tests are performed to verify that the multivariate model is applicable to the spectrum. Only spectra whose analysis represents interpolation of the multivariate model and which are sufficiently close to spectra in the calibration may be used in the analyzer validation.

4.3 When the analyzer is initially installed, or after major maintenance is concluded, performance tests are conducted to verify that the instrument is functioning properly. The intent of these tests is to provide a rapid indication of the state of the instrument. These tests are necessary but not sufficient to demonstrate valid analyzer results.

4.4 After the initial performance test is successfully completed, a probationary validation test is conducted on at least 15 samples that were not used in developing the multivariate model. The purpose of this probationary validation is to verify that the results produced by the analyzer (the PPTMRs) agree with the results from the primary test method (the PTMRs) to within user-defined limits for bias and precision. The PPTMRs and PTMRs are a compared using the statistical methodology of Practice D6708, recognizing that this is only a preliminary assessment. Precision and bias statistics on the prediction deviations (Δ) are generated for 15 samples whose spectra are not outliers nor nearest neighbor inliers, and the bias is assessed against pre-specified performance criteria. The system or subsystem performance is considered to be probationary validated for materials and property ranges representative of those used in the validation if the prediction deviations are in statistical control, and bias performance statistic meets pre-specified criterion providing that the spectra used in generating the results are neither outliers or nearest neighbor inliers.

4.5 After probationary validation is achieved, continued statistical quality control chart monitoring and analyses on Δ are carried out with new production samples to ensure ongoing prediction performance of the PPTMR meets the levels established from the probationary validation.

4.6 Once the total number of (PPTMR/PTMR/Δ) data sets for samples from probationary and continual validation reaches 30, a general validation is conducted using the statistical methodology of Practice D6708. The samples used in this general validation should only include those whose spectra are not outliers or nearest neighbor inliers relative to the multivariate model. The objective of the general validation is to demonstrate that the PPTMRs agree with the PTMRs to within user-defined limits for bias and precision on at least 30 samples covering a wider operating envelope, or, to confirm outcome from probationary validation with more accrued data.

4.7 During routine operation of the analyzer, validation tests are conducted on a regular, periodic basis to demonstrate that the analyzer results remain in statistical agreement with results for the primary test method. Prediction deviations (Δ) are monitored using statistical quality control charts at a frequency that is commensurate with the criticality of the application. Between validation tests, performance tests are conducted to verify that the instrument is performing in a consistent fashion.

5. Significance and Use

5.1 The primary purpose of this practice is to permit the user to validate numerical values produced by a multivariate, infrared or near-infrared laboratory or process (online or at-line) analyzer calibrated to measure a specific chemical concentration, chemical property, or physical property. *The validated analyzer results are expected to be statistically indistinguishable, over diverse samples whose spectra are neither outliers or nearest neighbor inliers, to those produced by the primary test method to within control limits established by control charts for the prespecified statistical confidence level.*

5.2 Procedures are described for verifying that the instrument, the model, and the analyzer system are stable and properly operating.

5.3 A multivariate analyzer system inherently utilizes a multivariate calibration model. In practice the model both implicitly and explicitly spans some subset of the population of all possible samples that could be in the complete multivariate sample space. The model is applicable only to samples that fall within the subset population used in the model construction. A sample measurement cannot be validated unless applicability is established. Applicability cannot be assumed.

5.3.1 Outlier detection methods are used to demonstrate applicability of the calibration model for the analysis of the process sample spectrum. The outlier detection limits are based on historical as well as theoretical criteria. The outlier detection methods are used to establish whether the results obtained by an analyzer are potentially valid. The validation procedures are based on mathematical test criteria that indicate whether the process sample spectrum is within the range spanned by the analyzer system calibration model. If the sample spectrum is an outlier, the analyzer result is invalid. If the sample spectrum is not an outlier, then the analyzer result is valid providing that all other requirements for validity are met. Additional, optional tests may be performed to determine if the process sample spectrum falls in a sparsely populated region of the multivariate space covered by the calibration set, too far from neighboring calibration spectra to ensure good interpolation. For example, such nearest neighbor tests are recommended if the calibration sample spectra are highly clustered.

5.3.2 This practice does not define mathematical criteria to determine from a spectroscopic measurement of a sample whether the sample, the model, or the instrument is the cause of an outlier measurement. Thus the operator who is measuring samples on a routine basis will find criteria in the outlier detection method to determine whether a sample measurement lies within the expected calibration space, but will not have specific information as to the cause of the outlier without additional testing.

6. Apparatus and Considerations for Quantitative IR Measurements

6.1 *Infrared or Near-Infrared Spectrophotometer:*

6.1.1 The analyzer covered by this practice is based on an infrared spectrophotometer, double-beam or single-beam, suitable for recording accurate measurements in the near-infrared

(780 nm to 2500 nm, 12820.5 cm^{-1} to 4000 cm^{-1}) or mid-infrared (4000 cm^{-1} to 400 cm^{-1}) regions, or both. The spectral range measured by the analyzer shall be the same or greater than that measured by the instrument used in collecting the spectral data upon which the multivariate calibration model is based. Complete descriptions of the instrumentation and procedures that are required for quantitative online process IR measurements are beyond the scope of this practice. Some general guidelines are given in Annex A1. (**Warning—**There are inherent dangers associated with the use of electrical instrumentation, online processes, and hydrocarbon materials. The users of this practice should have a practical knowledge of these hazards and employ appropriate safeguards.)

6.1.2 *In developing spectroscopic methods, it is the responsibility of the user to describe the instrumentation and the performance required to achieve the desired repeatability, reproducibility, and accuracy for the application.*

6.2 *Process Analyzer System*—The process analyzer system typically includes the spectrophotometer, transfer optics, the hardware for sample handling, the hardware for introduction of reference standards and solvents, the computer for controlling the spectrophotometer and calculating results, and the multivariate model. The system configuration should be compatible with the mid-infrared or near-infrared IR measurement and this practice.

6.3 *Collection of Line Samples:*

6.3.1 Withdraw line samples in accordance with accepted sampling methods as given by Practices D1265, D4057, D4177, or D7453, whichever is applicable. Flush the entire sample loop with the process stream sample prior to withdrawal of the line sample.

6.3.2 The intent of this practice is to collect samples that correspond directly to the spectra being collected by the analyzer. Collect the sample at a port close to the optical probe and at a time correlated with the collection of the sample spectrum. This practice requires that parameters that can impact the result also be recorded at the time of sample collection and the effect of these parameters is properly accounted for when comparing the results with the primary test method result. For a more detailed discussion of the various lag times that can influence the correspondence between the analyzer measurement and collection of line samples, see Practice D3764 and Guide D7278.

6.3.2.1 If line samples covering the composition and property range of interest cannot be acquired within a reasonable length of time once the validation process begins, consider using process-derived validation reference materials (VRMs) to extend the composition and property range of the validation sample set. A suitable process-derived VRM may simply be a batch of material obtained at a time prior to the start of the validation procedure, but one that was not used in calibrating either the analyzer or the primary test method. In general, the composition of a VRM used for validation should be similar to a composition that is anticipated for the process stream at some future time.

6.3.2.2 In cases where it is necessary to include the sample loop, or the sample conditioning unit, or both, in the validation procedure, VRMs should not be used to the exclusion of line sample unless it is practical to use the VRMs to validate both sample system and analyzer (this is generally not practical). The sample system can be excluded from the validation procedure if it is known that the sample system does not materially alter the composition or condition of the sample presented to the analyzer and if the sample system response time can be estimated with reasonable certainty. Guidance on how to meet these conditions is beyond the intended scope of this practice. If these conditions cannot be met and if VRMs are needed to extend the property and composition range of the validation set, it is recommended that the user conduct two probationary validations, one using line samples and the other using VRMs, to demonstrate that VRM procedure adequately reflects corresponding performance for actual process materials. Once demonstrated, the statistical quality control charting for continual validation can be done using VRM procedures, with a periodic line sample procedure mixed in over time to demonstrate that both procedures continue to provide similar and acceptable performance.

6.3.3 Sample storage for extended time periods is not recommended if there is likelihood that samples degrade with time. Chemical changes occurring during storage will cause changes in the spectrum, as well as changes in the property or quality parameter measured by the primary test method.

6.3.4 If possible, at the time of line sample withdrawal, collect sufficient quantity of sample material to allow for multiple measurements of the property or quality parameter by the primary test method, should such measurements be required.

7. Preconditions

7.1 Certain preconditions shall be met before this practice can be applied.

7.1.1 Install the analyzer in accordance with manufacturer's instructions.

7.1.2 Maintain analyzer and monitor per manufacturer's guidelines to assure proper peak shift and baseline management.

7.1.3 Develop and validate the multivariate calibration model used on the process analyzer using methods described in Practices E1655. If a calibration transfer method is used to transfer the model from one analyzer to another, verify the transferred model as described in Practices E1655.

NOTE 2—It is permissible to conduct the validation of the multivariate calibration model and the analyzer simultaneously using the same set of validation samples providing these samples meet the requirements of both Practices E1655 and this practice.

7.1.4 A quality assurance program for the primary test method is required in order to determine the usability of values generated by the primary test method in the validation of analyzer performance using this practice (see Section 8).

8. Reference Values and the Quality Assurance Program for the Primary Test Method

8.1 The property reference value against which analyzer results are compared during validation is established by applying the primary measurement method which was used in the

model development to line samples representing the process stream either directly, or after consistent treatment depending on the application.

8.1.1 If the line sample is treated prior to measurement by the primary method, such treatment should be done via procedures described in an appropriate ASTM standard such as Practice D7717. In the absence of an appropriate ASTM standard, the user shall document the treatment procedure in sufficient detail to ensure its consistent application.

8.2 A quality assurance program for the primary test method is required for values generated by this method to be used in analyzer validation. See Practice D6299 for reference.

8.2.1 Carefully check the laboratory apparatus used for primary test method measurement before these tests are performed to ensure compliance with the requirements of the primary test method.

8.2.2 Test control materials of known composition and quality on a regularly scheduled basis. Plot the primary test method results on control charts to ensure the long-term performance of the primary test. Individual values, exponentially weighted moving average, and moving range of two control charts are all recommended for charting the performance of the primary test method. Calculate the values for these control charts using equations given in Sections 12 and 13. Plot the differences between the primary test method result, and the expected value for the standard sample. Determine the historical precision of the primary test method from these regular tests, and compare it to published values for the method to determine if the test is within expected limits. Compare the historical precision to the analyzer precision using statistical tests.

9. Procedure

9.1 A flowchart for the steps involved in this practice, as it applies to process analyzers, is shown in Figs. 1-3.

9.2 *Initial Performance Tests:*

9.2.1 After the multivariate process analyzer has been installed (or reinstalled following major maintenance), check the performance of the instrument. Refer to manufacturer's instructions to ensure sufficient signal to noise ratio, peak positioning, and baseline management. The objective of the check is to determine that current performance of the instrument is consistent with performance which is known to produce valid analyses. Collect spectra of 20 check or test samples and analyze them using one or more of the Level 0, Level A, or Level B performance tests described in Annex A2 and Practice E1866.

9.2.2 Performance test results should be plotted on control charts and examined for trends. Such trend analysis may provide early warnings of possible analyzer problems. See Annex A2 and Practice D6299.

9.2.3 Compare the results for the initial performance tests to performance test action limits. These action limits may be based on historical data for the same tests, on simulations of the effects of performance changes on the analyzer results, or on a combination of historical and simulated data. Methods for establishing action limits are discussed in Annex A2 and Practice E1866.

9.2.3.1 If the performance test results are within action limits, then the procedure continues with the initial validation tests. If the performance test results are not within action limits, check installation, instrument standardization or calibration transfer, or combination thereof, and correct the cause of the inadequate performance. Repeat the initial performance tests.

9.2.3.2 If action limits for performance tests have not been established, use the results for the initial performance tests to generate an initial historical database against which future tests can be compared, and continue the validation procedure with the steps described in 9.3. In the absence of historical data or performance simulations, the performance of the instrument cannot be verified, but shall be assumed. Should the analyzer fail to validate, inadequate instrument performance could be responsible.

9.3 *Probationary Validation* (see Section 12 for details):

9.3.1 For an online or at-line process analyzer, once the initial performance tests are completed, collect spectra of 15 line and test samples and analyze them using the multivariate model. In order for the results to be used in the initial validation test, the spectra of the 15 line or test samples shall not be either outliers or nearest neighbor inliers (see Section 11 and Annex A3). Replace samples whose spectra are outliers or nearest neighbor inliers with other line or test samples.

9.3.2 Withdraw line samples from the process using methods described in Practices D1265, D4057, or D4177, whichever is applicable. The line sample shall correspond directly to the spectrum collected in 9.3.1.

9.3.3 For a laboratory analyzer, collect spectra of sampled line product and analyze them using the multivariate infrared method.

9.3.4 If appropriate, treat the line samples in a consistent fashion to produce treated line samples.

9.3.5 Analyze the line samples or treated line samples using the PTM.

9.3.6 Perform a preliminary Practice D6708 assessment of the agreement between the PPTMRs and the PTMRs. If there is insufficient variation among the 15 samples, conduct a level-specific probationary validation.

NOTE 3—If line samples covering the composition and property range of interest cannot be acquired within a reasonable length of time once the validation process begins, consider using either process-derived validation reference materials (VRMs) to extend the composition and property range of the validation sample set. A suitable process-derived VRM may simply be a batch of material obtained at a time prior to the start of the validation procedure which was not used in developing the multivariate calibration model nor for calibrating the primary test method. In general, the composition of a VRM used for validation should be similar to a composition that is anticipated for the process stream at some future time.

9.3.7 Compare values calculated by the analyzer to those obtained by the primary test method using statistical tests described in Section 12. If the values are in statistical control, and there is no significant bias, then the analyzer passes probationary validation and can be used to analyze line samples within the range over which the validation was conducted. If the values are not within statistical agreement, then the installation, instrument standardization or calibration

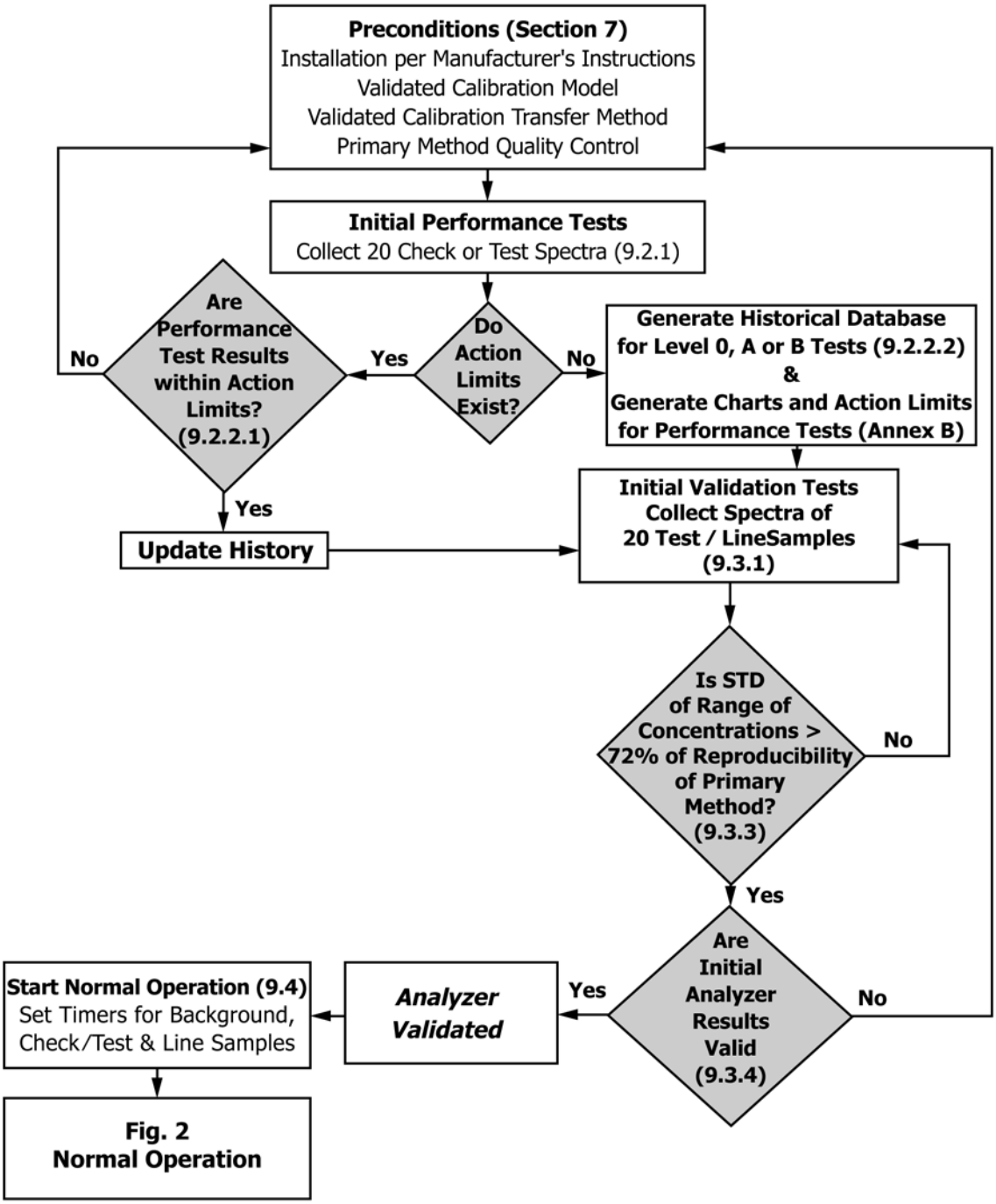

FIG. 1 Flowchart of Process Analyzer Validation Practice Initial Startup and Restart after Maintenance

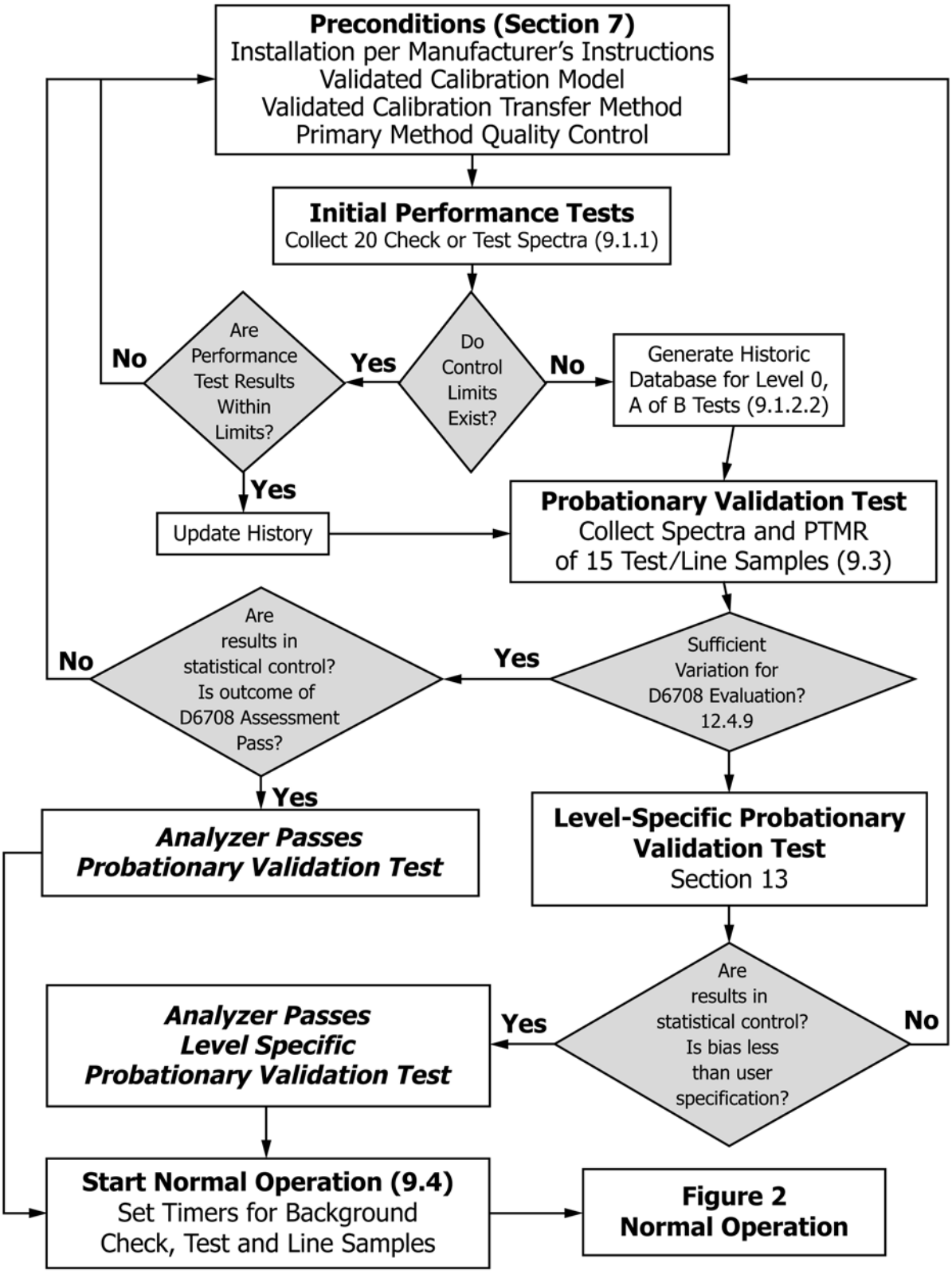

FIG. 1 Flowchart of Process Analyzer Validation Practice Initial Startup and Restart after Maintenance *(continued)*

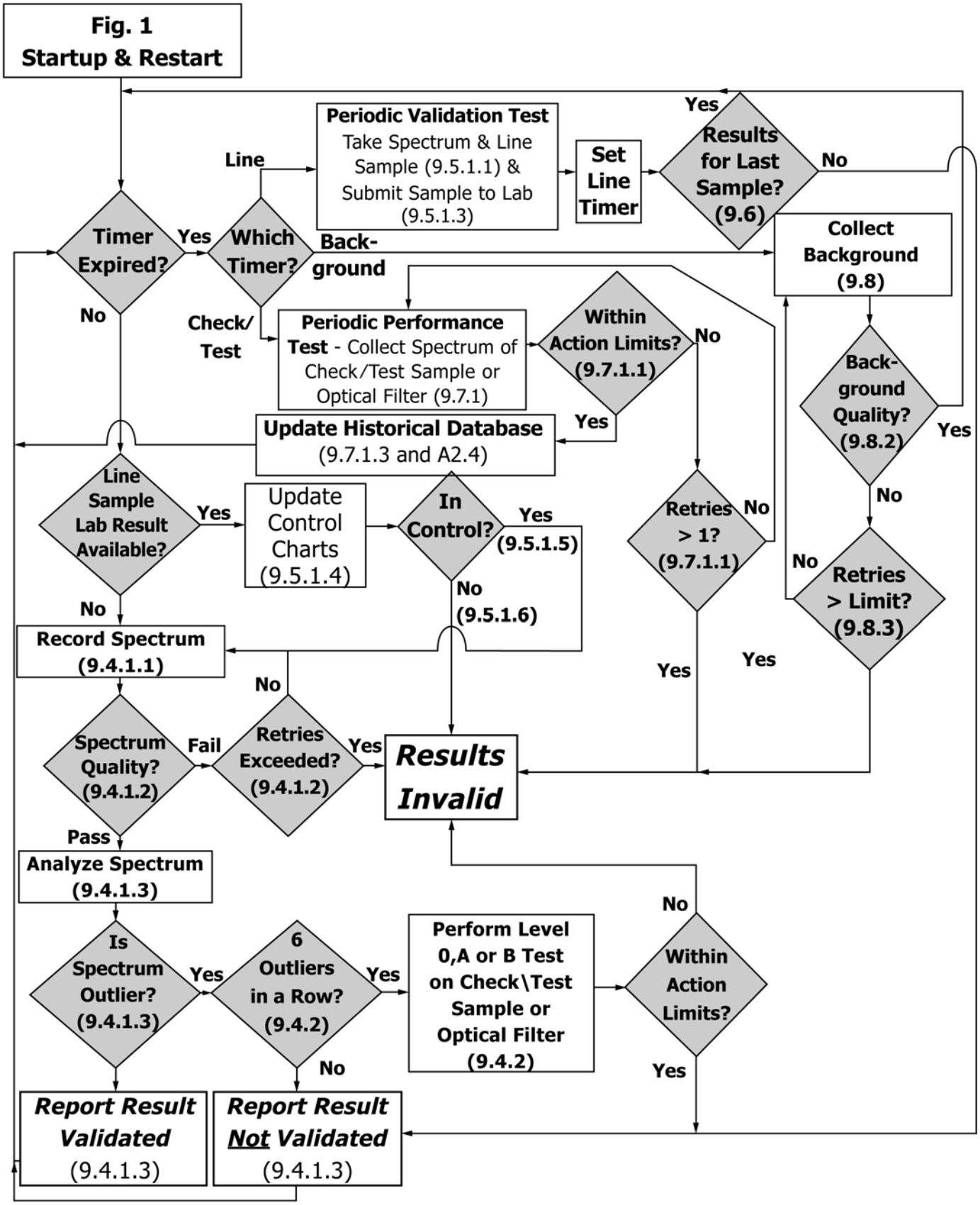

FIG. 2 Flowchart of Process Analyzer Validation Practice Normal Operation

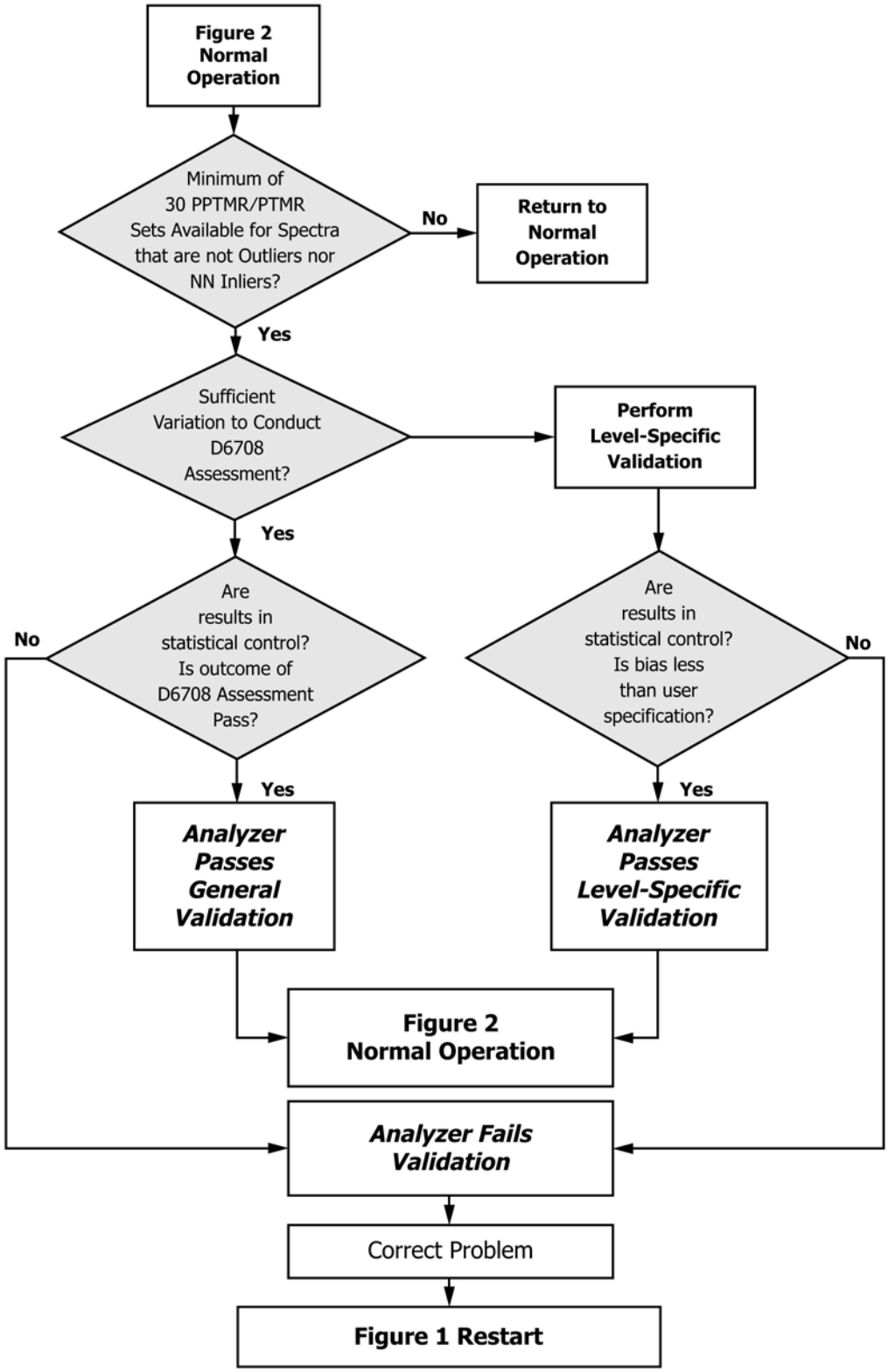

FIG. 3 Flowchart of Process Analyzer Validation Practice General Validation

transfer, or combination thereof, are checked and corrected, and the procedure starts over with initial performance tests as described in 9.2.

9.4 *Normal Operation:*

9.4.1 Once the probationary analyzer system validation is completed, normal operations for analysis of process samples may be conducted. Conduct tests of the performance of the analyzer and of the validity of the analyzer results on a periodic, regularly scheduled basis. When these tests are not scheduled, the normal application of the analyzer for online, at-line or laboratory analysis proceeds as follows:

9.4.1.1 Collect a spectrum of the sample.

9.4.1.2 Optionally, conduct tests on the spectrum in order to determine that the quality of the spectrum is adequate for use in estimating results by way of application of the multivariate model. Spectrum quality tests are generally defined by the instrument manufacturer or model developer, or both. If spectrum quality tests are used, allow a finite number of retries on the spectrum collection before the analyzer is considered inoperative, and the results produced invalid.

9.4.1.3 Analyze the spectrum using the calibration model, to produce one or more results, possibly uncertainties in these results, and statistics which are used to determine if the spectrum is an outlier or nearest neighbor inlier relative to the calibration (training) set used in the development of the multivariate model (see Section 11 and Annex A3). If the spectrum recorded during normal operation of the analyzer is not an outlier or nearest neighbor inlier, then the calculated property values produced are considered valid as long as the analyzer quality control charts are up to date and the differences between the analyzer results and the primary test method results are within control limits. If the spectrum recorded during the normal operation of the analyzer is an outlier or nearest neighbor inlier, then the specific results associated with that spectrum are considered to be invalid.

9.4.2 When spectra recorded during the normal operation of the analyzer are outliers, performance tests are conducted to determine if the instrument performance is within action limits (see 10.3.3).

9.4.2.1 For on-line process analyzers, where spectra are collected on a more or less continuous basis, conduct performance tests when six successive spectra are all outliers.

9.4.2.2 For applications which use at-line or laboratory analyzers, the tolerance of the application to outliers may vary. It is the user's responsibility to establish a trigger limit (number of successive outlier spectra) above which performance tests are required. For critical applications, performance tests can be performed after a single outlier spectrum.

9.5 *Periodic Continual Validation Tests:*

9.5.1 Conduct periodic analyzer validation tests at regularly scheduled intervals, preferably once a week (see Section 13).

9.5.1.1 Withdraw a line sample from the process. For a process analyzer, simultaneously collect a spectrum of the process stream with the process analyzer. For an at-line or laboratory analyzer, collect a spectrum of the line sample.

9.5.1.2 Analyze the spectrum using the multivariate model to produce a result, and to produce outlier and nearest neighbor inlier statistics. If the spectrum is an outlier or nearest neighbor inlier, it cannot be used for the validation test, and the procedure starts over with 9.5.1.

9.5.1.3 If appropriate, treat the line sample in a consistent fashion to produce a treated line sample.

9.5.1.4 Analyze the line sample or treated line sample by the primary test method used in the development of the calibration.

9.5.1.5 Compare the analyzer and primary test method results by plotting their difference on control charts as described in Section 13.

9.5.1.6 If the difference is within control limits, then the predicted result for the analyzer is considered to be valid.

9.5.1.7 If the difference is not within control limits, then the result for the analyzer is invalid. Check the control charts for the primary test method (see Section 8) to ensure that the primary test method is within control limits. If the primary test method is not within control limits, determine and correct the cause of the error, and repeat the primary test method test. If the primary test method is within control limits, conduct performance tests on the infrared analyzer to check if the instrument performance is within action limits. If the instrument performance is not within action limits, service to the analyzer may be necessary.

9.5.2 The exact period between validation samples will depend on the nature of the analyzer application. At minimum, collect and analyze a validation sample at least once within each seven-day period. More frequent validation testing may be appropriate for applications where analyzers are being used to certify products. The period between validation samples should not be less than the typical time required obtaining the reference data by way of the primary test method.

9.6 For a process analyzer, if the primary test method results for a line sample are not available by the time the next validation sample is scheduled to be collected, then the results produced by the analyzer are to be considered invalid until such time as the overdue results become available and the control charts are updated.

9.7 *Performance Tests:*

9.7.1 It is recommended that performance tests be conducted on a regularly scheduled basis, preferably daily, between the periodic analyzer validation tests. The objective of the test is to demonstrate that the analyzer performance is consistent between validation tests. In the absence of manufacturer's instructions, details on performance tests are given in Section 10, Annex A2, and Practice E1866.

9.7.1.1 If the results for the performance tests are within action limits, continue operation of the analyzer.

9.7.1.2 If the results of the performance tests are not within action limits, then repeat the test. If the results of the repeat test are not within action limits, then the analyzer results are considered invalid, and the analyzer should be serviced.

9.7.1.3 If action limits have not been established for the performance tests, it is recommended that validation tests be performed more frequently to establish the historical database against which the limits can be set (see Annex A2 and Practice E1866).

9.8 *Optical Backgrounds:*

9.8.1 Collect new optical backgrounds on a regularly scheduled interval, or when indicated by analyzer performance results.

9.8.2 Tests may be conducted on the collected optical background to determine its quality. Background quality tests are generally defined by the instrument manufacturer or model developer, or both.

9.8.3 If background quality tests are used, allow a finite number of retries on the spectrum collection before the analyzer is considered inoperative, and the results produced invalid.

9.9 *General Validation:*

9.9.1 Once a total of at least 30 probationary and continual validation data sets are available, a general validation may be attempted using the methodology of Practice D6708.

10. Performance Tests

10.1 Performance tests are conducted to determine whether the performance of the instrument (the spectrophotometer, the optical cell, and all transfer optics in between) is adequate to produce spectra of the quality sufficient for valid analyses. Typically, check or test samples are introduced into the analyzer, the spectra of these samples are analyzed using the appropriate Level 0, Level A, or Level B performance test, and the results are plotted on charts and compared to action limits. For analyzers equipped with in-line probes, it may be impractical to remove the probe to conduct performance tests. For such analyzers, alternative procedures described in Annex A2 and Practice E1866 may be used to conduct performance tests. Adequacy of the spectra is determined by comparison to a historical database of spectra of sufficient and insufficient quality. Alternatively, simulations of possible changes in instrument performance can be used to define the performance that is adequate for a given application. A description of Level 0, A, and B tests, and of methods for setting action limits for performance tests based on historical data and on simulations, are described in detail in Annex A2 and Practice E1866.

10.2 When conducting the performance tests, operate the instrument in the most stable and reproducible conditions attainable, as defined by the manufacturer. Allow sufficient warm-up time before the commencement of any measurements. If the calibration model was based on spectra of samples held within a specified temperature range, then allow all samples, including check and test samples, to equilibrate to this temperature prior to spectral measurement. If possible, the optical configuration used for measurements of test and check samples should be *identical* to that used for measurement of line samples. If identical optical configurations are not possible due to analyzer design, the user should recognize that the performance tests may not measure the performance of the entire instrument. Data collection and computation conditions should be equivalent to those used in the collection of the spectra used in the calibration model. Introduce fresh reference material into the spectrophotometer cell for each measurement. Flow through the cell during the measurement is not required. Date and time stamp the spectral data used in performance tests, and store the results of the tests in a historical database.

10.3 *Timing of Analyzer Performance Tests:*

10.3.1 Conduct performance tests on a regularly scheduled basis, preferably daily, to test instrument performance consistency between validation tests. Compare the results of the performance tests with action limits for the tests. If a significant change in the performance is observed, conduct a second analysis to verify the change. If the significant change in performance is verified, mark analyzer results not validated until the cause and effect of the change can be determined. If the change in performance is not verified, conduct analyses of five additional checks or test samples to demonstrate that the first occurrence was an anomaly, before continuing with normal operation.

10.3.1.1 The significance of a change in instrument performance may be unknown in the absence of historical data or simulations. In such case, more frequent validation testing may be required to demonstrate the relationship between analyzer performance and valid analyses. If, after a change in instrument performance is observed, the analyzer results remain in control, the change is not adversely affecting analyzer results. If, however, the analyzer results go out of control relative to the primary test method, the change is adversely affecting analyzer results.

10.3.1.2 If historical data or simulations exist to demonstrate that change in performance is sufficient to produce invalid analyses, then service the analyzer to correct the problem. Service of this type is considered major maintenance, and initial performance and validation tests are required before resuming analyzer operation.

10.3.2 When an analyzer is installed, or after major maintenance has been performed, conduct 20 instrument performance tests using the check or test sample over a 24 h period to capture any diurnal performance variations. Compare the performance test results for the 20 samples with performance test action limits to determine if the analyzer performance is adequate. Add the test results for the 20 analyses to the historical database against which future performance tests are compared. Once these performance tests have been successfully completed, initiate the initial validation of the analyzer.

10.3.3 If, during the course of normal operation, the spectra of six successive samples are determined to be spectral outliers (see Section 11 and Annex A3), it is recommended that performance tests be conducted to demonstrate that the outlier diagnostics are responding to chemical changes in the process stream and not to changes in the instrument performance. If the results for the performance tests are outside action limits, then the outlier diagnostics may be responding to instrument performance and the analyzer should be serviced. If the results for the performance tests are within action limits, then the outlier diagnostics are most likely responding to changes in the process which are producing materials outside the range of the current calibration. If the process remains outside the range of the calibration for extended periods, it is recommended that the instrument performance be verified periodically using performance tests, until such time as the process returns to a state where the model is again applicable. If the process has changed so as to be permanently outside the range of the calibration, then a new model should be developed following Practices

E1655. Revalidate the analyzer with the new model following the procedure described herein.

10.3.4 Conduct performance tests if a bias is observed between the analyzer and primary test method values to determine if the bias is the result of a change in instrument performance.

10.4 *Reference Materials for Instrument Performance Tests:*

10.4.1 Check samples are generally used for conducting performance tests. Check samples are single, pure, liquid hydrocarbon compounds or mixtures of liquid hydrocarbon compounds of definite composition. An alternate to using a check sample is to use an actual process sample called a test sample. For systems equipped with in-line probes, optical filters may be used as reference materials for instrument performance tests.

NOTE 4—Performance tests conducted on test samples are only intended to check the stability of analyzer performance over time. While the analyzer results for the test sample can be compared to the results for the primary test method, such comparisons are not a substitute for the validation tests described in Sections 12 and 13. Analyzer results for test samples can be used in the calculation of the analyzer intermediate precision (see Section 16).

10.4.2 Details on reference materials for instrument performance tests are given in Annex A2 and Practice E1866.

11. Verification that the Model is Applicable to the Spectrum of the Process Stream Sample—Spectral Outlier and Nearest Neighbor Inlier Detection

11.1 The spectra of the calibration samples define a set of variables that are used in the calibration model. If, when unknown samples are analyzed, the variables calculated from the spectrum of the unknown sample lie within the range of the variables for the calibration, the estimated value for the unknown sample is obtained by interpolation of the model. If the variables for the unknown sample are outside the range of the variables in the calibration model, the estimate represents an extrapolation of the model. Additionally, if the spectrum of the sample under test contains spectral features that were not present in the spectra of the calibration samples, then these features represent variables that were not included in the calibration, and the analysis of the sample spectrum represents an extrapolation of the model.

11.2 For the purpose of this practice, an analyzer result is considered valid only if the analysis involves an interpolation of the multivariate calibration model. Outlier detection methods are used to determine if an analysis represents an interpolation or an extrapolation of the multivariate model. The mathematics involved in outlier detection is described in Practices E1655 and in Annex A3. The calculation of outlier statistics is by necessity an integral part of the analyzer software since these calculations shall be conducted each time the multivariate model is applied to a spectrum to produce a result. Appropriate limits for outlier tests will generally be set by the calibration model developer based on statistics from the calibration set.

11.2.1 A Mahalanobis Distance (leverage statistic) or scores range test shall be employed to determine if the spectrum being analyzed represents an interpolation or extrapolation of the variable space defined by the calibration model.

11.2.2 A spectral residuals statistic shall be employed to detect extrapolation of the calibration model due to spectra features which were not present in the spectra of the calibration set.

11.2.3 Optionally, a Nearest Neighbor Distance statistic can be employed to determine when the spectrum being analyzed falls in a sparsely populated region of the multivariate calibration space. While analyses of such spectra represent interpolation of the model, there may be insufficient information in the model to produce valid analyses for these samples. The use of a Nearest Neighbor Distance statistic is recommended if the calibration samples are highly clustered in the multivariate space. It is the responsibility of the model developer to determine if use of a Nearest Neighbor Distance statistic is appropriate. If a Nearest Neighbor Distance statistic is employed, then the results for any sample whose Nearest Neighbor Distance exceeds the predetermined limit are considered invalid. Such samples are referred to as Nearest Neighbor Inliers.

11.3 Annex A3 discusses available outlier detection methods. Further details on outlier methods and on notations used in their calculations are in Practices E1655. Users may substitute other outlier detection methods providing they are at least as rigorous as those described in Annex A3 and Practices E1655. If alternative outlier detection methods are substituted, it is the user's responsibility to demonstrate that any analyzer results that are marked as invalid by the tests described herein are also marked as invalid by the substituted methods.

11.4 While it is generally preferable that the outlier statistics be generated using the same modeling method that was used to generate the calibration model, this is not required. For instance, MLR models do not provide spectral residual statistics. If an MLR model is used as the calibration model, an additional PCR or PLS model may be used to provide the necessary residuals statistics. If a supplementary model is used to generate outlier statistics, construct the supplementary model using the same set of calibration samples used for the predictive model, and apply the outlier statistics which will be used on the process analyzer system in the validation of the model in accordance with Practices E1655.

11.4.1 Outlier tests detect differences in the spectrum of the process sample relative to the spectra of the calibration samples. These spectral differences may be due to differences in the chemistries of the samples, or due to differences in the performance of the spectrometer used to collect the spectra. Table 1 discusses inferences that may be drawn from outlier test results. The outlier tests by themselves do not distinguish between the instrument and the sample being the cause of the outlier result. Instrument performance tests may be used to help determine if the outlier test is responding to changes in the process or in the instrument.

12. Analyzer System Probationary and General Validation

12.1 Probationary validation of the analyzer system is done using the statistical methodology of Practice D6708.

TABLE 1 Inferences Related to Outlier Detection or Instrument Failure

Mahalanobis Distance and Scores Range Tests	Spectral Residual Test	Inferences	Status of Analyzer Result
Less than limit(s)	Less than limit	Spectrum within range of calibration spectra	Result valid if control charts are current and within control limits
Greater than limit(s)	Less than limit	Possible instrument malfunction or model extrapolation due to sample component outside range for calibration	Invalid result
Less than limit(s)	Greater than limit	Possible instrument malfunction or model extrapolation due to sample absorption not present in calibration spectra	Invalid result
Greater than limit(s)	Greater than limit	Possible instrument malfunction or model extrapolation	Invalid result

12.2 Obtain the site precision for the primary test method for the materials and property levels of interest using procedures outlined in Practice D6299. Obtain the analyzer intermediate precision using procedures described in Section 17.

12.3 Collect PPTMRs and PTMRs for a minimum of 15 samples in accordance with line sample procedures (see 6.3) at a frequency of no more than once per day. Avoid taking this sample at the same time of day to ensure any time-of-day related effect is captured in the data set. If it is impractical to acquire line samples covering the composition and property range of interest in a reasonable length of time once validation begins, refer the discussion in Sections 6.3.2.1 and 6.3.2.2 on the use of validated reference materials to supplement line samples.

12.4 Assess the degree of agreement between the PTMRs and the PPTMRs using site and analyzer precision estimates and the statistical principles/calculations of Practice D6708 (see Annex for an example of assessment) to answer the following questions:

12.4.1 For each line sample collected whose spectrum is not an outlier or nearest neighbor inlier, calculate the prediction deviation Δ, where Δ = (PPTMR – PTMR).

12.4.2 After a minimum 15 line sample data sets are collected, conduct a preliminary Practice D6708 assessment, using the PPTMR, PTMR, and site precision standard deviations for the PPTMR and PTMR which have been established using control charts on suitable quality control materials as per Practice D6299.

NOTE 5—Site precision standard deviation is not to be confused with standard deviation of the PPTMR results for the line samples.

12.4.3 Follow the Practice D6708 outcome decision flowchart in Fig. 4. If the outcome is a "fail," then, the system that produced the PPTM results is deemed to have failed the validation requirements of this practice.

12.4.3.1 In the Practice D6708 assessment, an Anderson-Darling test is used to determine if the distribution of the prediction deviations (residuals), including sample specific biases and other errors, is nominally Gaussian (normal). Follow the instructions in Practice D6708 to determine assessment outcome.

12.4.3.2 If the PTMR or PPTMR are not reported to a sufficient number of significant digits, then the Anderson-Darling test may not be applicable. If the Practice D6708 assessment fails because the prediction deviations (Δ) fail the Anderson-Darling test, visually inspect the prediction deviations (Δ) to determine how many unique values are present. If there are fewer than 4 unique values, the Anderson-Darling test is not applicable. If there are 4 to 6 unique values, the Anderson-Darling test may or may not be applicable. For these situations, the range of the deviations (max-min) should be compared to user requirements to determine if the PPTMR can be used to meet the analyzer application requirement.

12.4.4 If the Practice D6708 outcome decision from above is a "pass," follow the instructions in Practice D6299 (section on Procedure for Pretreatment, Assessment, and Interpretation of Test Results) and assess all Δ following the quality control (QC) sample results protocol. Interpret the control chart generated and determine whether the Δ exhibit in statistical control behavior. Investigate the out-of-control points and take appropriate corrective actions to address the root cause(s). Replace the out-of-control points by repeating the line sampling procedure.

12.4.5 If the Δ exhibit in statistical control behavior, the system that produced the PPTMR is deemed to have passed probationary validation.

12.4.6 Continue to collect validation samples and populate the statistical control chart for Δ as described in Section 14 and Practice D6299.

12.4.7 A reassessment using Practice D6708 techniques as described above shall be conducted when data from a total of 30 line samples whose spectra are neither outliers nor nearest neighbor inliers have been accrued (including the probationary data).

12.4.8 If the Practice D6708 reassessment outcome (Fig. 4) is a "pass," and, if all Δ results exhibit in-statistical-control behavior, compare R_{XY} from the Practice D6708 outcome to the required precision performance. Ensure the comparison is carried out on the same unit basis (that is, compare reproducibility to reproducibility, not standard deviation). The analyzer system is deemed to have met the General Validation requirements of this practice if the precision performance criterion is met. A failure of the Practice D6708 outcome, or, out of control behavior of the Δ results, or failure to meet the precision performance criteria will be deemed as a failure to meet the General Validation requirements of this practice.

12.4.9 If the Practice D6708 assessment concludes that there is insufficient variation in the sample set, proceed with a level-specific validation later in Section 13. If line samples covering the composition and property range of interest cannot be acquired within a reasonable length of time once the validation process begins, consider using process-derived validation reference materials (VRMs) to extend the composition and property range of the validation sample set. A suitable process-derived VRM may simply be a batch of material obtained at a time prior to the start of the validation procedure, but was not used in developing the multivariate calibration model nor for calibrating the primary test method. In general, the composition of a VRM used for validation should be

TABLE 2 95th and 97.5th Percentiles of the Student's *t*-Distribution

Degrees of Freedom	t_{95}	$t_{97.5}$	Degrees of Freedom	t_{95}	$t_{97.5}$	Degrees of Freedom	t_{95}	$t_{97.5}$
1	6.3138	12.7062	28	1.7011	2.0484	75	1.6654	1.99210
2	2.9200	4.3027	29	1.6991	2.0452	80	1.6641	1.99006
3	2.3534	3.1824	30	1.6973	2.0423	85	1.6630	1.98827
4	2.1318	2.7764	31	1.6955	2.0395	90	1.6620	1.98667
5	2.0150	2.5706	32	1.6939	2.0369	95	1.6611	1.98525
6	1.9432	2.4469	33	1.6924	2.0345	100	1.6602	1.98397
7	1.8946	2.3646	34	1.6909	2.0322	105	1.6595	1.98282
8	1.8595	2.3060	35	1.6896	2.0301	110	1.6588	1.98177
9	1.8331	2.2622	36	1.6883	2.0281	115	1.6582	1.98081
10	1.8125	2.2281	37	1.6871	2.0262	120	1.6577	1.97993
11	1.7959	2.2010	38	1.6860	2.0244	125	1.6571	1.97912
12	1.7823	2.1788	39	1.6849	2.0227	130	1.6567	1.97838
13	1.7709	2.1604	40	1.6839	2.0211	135	1.6562	1.97769
14	1.7613	2.1448	41	1.6829	2.0195	140	1.6558	1.97705
15	1.7531	2.1314	42	1.6820	2.0181	145	1.6554	1.97646
16	1.7459	2.1199	43	1.6811	2.0167	150	1.6551	1.97591
17	1.7396	2.1098	44	1.6802	2.0154	155	1.6547	1.97539
18	1.7341	2.1009	45	1.6794	2.0141	160	1.6544	1.97490
19	1.7291	2.0930	46	1.6787	2.0129	165	1.6541	1.97445
20	1.7247	2.0860	47	1.6779	2.0117	170	1.6539	1.97402
21	1.7207	2.0796	48	1.6772	2.0106	175	1.6536	1.97361
22	1.7171	2.0739	49	1.6766	2.0096	180	1.6534	1.97323
23	1.7139	2.0687	50	1.6759	2.0086	185	1.6531	1.97287
24	1.7109	2.0639	55	1.6730	2.0040	190	1.6529	1.97253
25	1.7081	2.0595	60	1.6706	2.0003	195	1.6527	1.97220
26	1.7056	2.0555	65	1.6686	1.9971	200	1.6525	1.97190
27	1.7033	2.0518	70	1.6669	1.9944	∞	1.6449	1.96000

TABLE 3 Critical Values of λ for Generalized ESD Procedure

n	λ_1	λ_2	λ_3
20	2.71	2.68	2.65
25	2.82	2.80	2.78
30	2.91	2.89	2.88

similar to a composition that is anticipated for the process stream at some future time.

12.5 If there is no statistically significant bias for the general validation process, then the 95 % confidence limit on the absolute value of the difference between the measurements by the validated analyzer and by the primary test method is given approximately by R_{XY}. This limit applies only to primary test method results produced by the same laboratory which provided the data used in the validation. *Comparisons of the analyzer results to primary test method results for other laboratories may produce larger differences.*

12.5.1 Optionally, the analyzer validation results may be compared to those obtained during the validation of the multivariate model to determine if the analyzer performance is consistent with that expected based on the model validation.

12.5.1.1 Compare the standard deviation of the between-method reproducibility for the analyzer to that which was obtained for the validation of the model using an *F*-test. The *SEV* is the Standard Error of Validation for the model. The *SEV* was calculated as part of the validation of the model following procedures described in Practices E1655.

12.5.1.2 Calculate the value *F* as follows:

$$\sigma_{XY} = \frac{R_{XY}}{2.77} \tag{1}$$

$$F = \frac{\sigma_{XY}^2}{SEV^2} \text{ for } \sigma_{XY} > SEV \tag{2}$$

$$F = \frac{SEV^2}{\sigma_{XY}^2} \text{ for } SEV > \sigma_{XY} \tag{3}$$

12.5.1.3 Compare the value of *F* with the limiting *F* value given in Table 4. If Eq 2 is used, the number of degrees of freedom for the numerator and denominator are *n*–1 (where *n* is the number of analyzer validation samples) and *v*–1 (where *v* is the number of model validation samples), respectively. If Eq 3 is used, the number of degrees of freedom for the numerator and denominator are *v*–1 and *n*–1, respectively.

12.5.1.4 If the calculated value *F* is less than the limiting value *F* in Table 4, SE_a^2 is not significantly greater than SEV^2, and the performance of the analyzer is consistent with that expected for the multivariate model.

12.5.1.5 If the calculated value *F* is greater than the limiting value *F* from Table 4, then there is a statistically significant difference between SE_a^2 and *SEV*. If Eq 2 was used, the performance of the analyzer may be poorer than would be expected on the basis of the model validation results. Conduct further investigation of the analyzer function and operation to resolve the source of the poor performance. If Eq 3 was used, the performance of the analyzer may be better than would be expected on the basis of the model validation results.

13. Analyzer System Level Specific Validation

13.1 If there is inadequate property level variation to conduct a general initial validation (Section 12), then the level specific validation of the analyzer is performed by comparing the analyzer and primary test method results for a set of at least 15 validation samples whose spectra are neither outliers nor nearest neighbor inliers as defined in Section 11.

13.1.1 For each of the 15 line samples collected, calculate the prediction deviation (Δ).

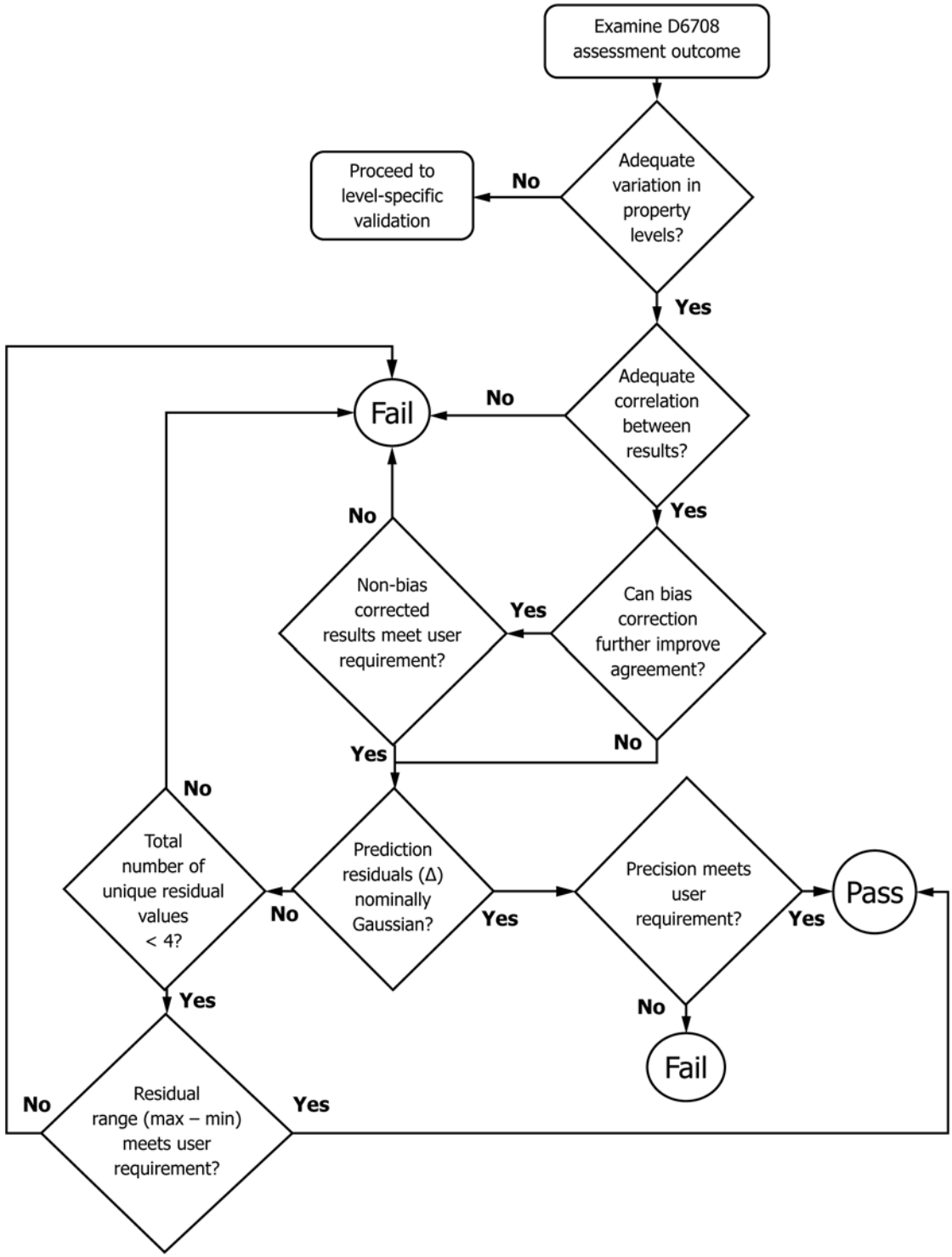

FIG. 4 Practice D6708 Outcome Assessment

13.1.2 Follow the instructions in Practice D6299 (section on Procedure for Pretreatment, Assessment, and Interpretation of Test Results) and assess all the Δ results following the quality control (QC) sample results protocol. Interpret the control chart generated and determine if the Δ results exhibit in statistical control behavior.

TABLE 4 95th Percentiles of the *F*-Distribution

		Degrees of Freedom Numerator (Number of Analyzer Validation Samples)						
		20	21	22	23	24	25	30
Degrees of Freedom Denominator	8	3.15	3.14	3.13	3.12	3.12	3.11	3.08
	12	2.54	2.53	2.52	2.51	2.51	2.50	2.47
	16	2.28	2.26	2.25	2.24	2.24	2.23	2.19
	20	2.12	2.11	2.10	2.09	2.08	2.07	2.04
	24	2.03	2.01	2.00	1.99	1.98	1.97	1.94
	28	1.96	1.95	1.93	1.92	1.91	1.91	1.87
	32	1.91	1.90	1.88	1.87	1.86	1.85	1.82
	36	1.87	1.86	1.85	1.83	1.82	1.81	1.78
	40	1.84	1.83	1.81	1.80	1.79	1.78	1.74
	44	1.81	1.80	1.79	1.78	1.77	1.76	1.72
	48	1.79	1.78	1.77	1.76	1.75	1.74	1.70
	52	1.78	1.76	1.75	1.74	1.73	1.72	1.68
	56	1.76	1.75	1.74	1.72	1.71	1.70	1.66
	60	1.75	1.73	1.72	1.71	1.70	1.69	1.65
	64	1.74	1.72	1.71	1.70	1.69	1.68	1.64
	68	1.73	1.71	1.70	1.69	1.68	1.67	1.63
	72	1.72	1.70	1.69	1.68	1.67	1.66	1.62
	76	1.71	1.70	1.68	1.67	1.66	1.65	1.61
	80	1.70	1.69	1.68	1.67	1.65	1.64	1.60
	84	1.70	1.68	1.67	1.66	1.65	1.64	1.59
	88	1.69	1.68	1.66	1.65	1.64	1.63	1.59
	92	1.69	1.67	1.66	1.65	1.64	1.63	1.58
	96	1.68	1.67	1.65	1.64	1.63	1.62	1.58
	100	1.68	1.66	1.65	1.64	1.63	1.62	1.57
	∞	1.57	1.56	1.54	1.53	1.52	1.51	1.46

13.1.3 If the Δ results are in statistical control, proceed with calculation of system precision and bias statistics. Otherwise, investigate the out-of-control points and take appropriate corrective actions to address the root cause(s). Replace the out-of-control points by repeating the line sampling procedure.

13.1.4 Assess the bias by performing a one-sample *t*-test using all the Δ results in accordance with Practice D6299. If the bias is not statistically significant, the system that produced the PPTMR is deemed to have passed probationary validation, limited to materials representative of the line samples used in the assessment.

13.1.5 If a statistically significant bias is observed, and is of a magnitude that exceeds user's requirement, for the purpose of this practice, the system that generated the PPTM results is considered to have failed to meet the probationary validation requirements. However, the average of the Δ results may be interpreted as the best estimate of the bias between the PTM and the analyzer system at the specific property level. Users may choose to reestablish the correlation, thus changing the PPTM process, and repeat the aforementioned probationary validation procedures.

13.2 Continue to collect validation samples and populate the control chart with new Δ results.

13.3 When the total number of validation sample data sets reaches 30, conduct a Practice D6708 assessment as per the protocol described under General Validation earlier in this practice.

13.3.1 If there is sufficient variation to conduct the Practice D6708 assessment, proceed with the general validation as described in Section 12.

13.3.2 If there is still insufficient variation for a successful Practice D6708 assessment, then, the data set is considered insufficient for a General Validation but can be used for assessing level specific performance.

13.3.3 Calculate the precision (standard deviation of the Δ) and the bias (mean of the Δ) and compare them to user specified requirements to form a conclusion for a level-specific validation outcome as follows:

13.3.3.1 Compare the precision (2× standard deviation) of the Δ results to the user-specified precision requirement (expressed as a reproducibility) to determine if the precision meets performance requirement.

13.3.3.2 Compare the bias (mean of the Δ) to the precision via a one-sample *t*-test to determine if the bias is statistically significant. If the bias is statistically significant, compare the bias value to user-specified bias requirement to determine if it is of practical significance.

13.3.3.3 If the precision meets user-specified precision requirements, and the bias is not of practical significance relative to user specified bias requirements, the analyzer passes level-specific validation and may be employed for analysis of materials within the narrow range covered by the level-specific validation materials.

13.3.3.4 If the analyzer precision does not meet user-specified precision requirements, or if the bias is practically significant, the analyzer fails level-specific validation. The cause of the failure should be investigated and corrected and validation can be restarted at the probationary level.

NOTE 6—For the purpose of this practice, if it is necessary to add a bias correction to a model to bring analyzer and primary test method results

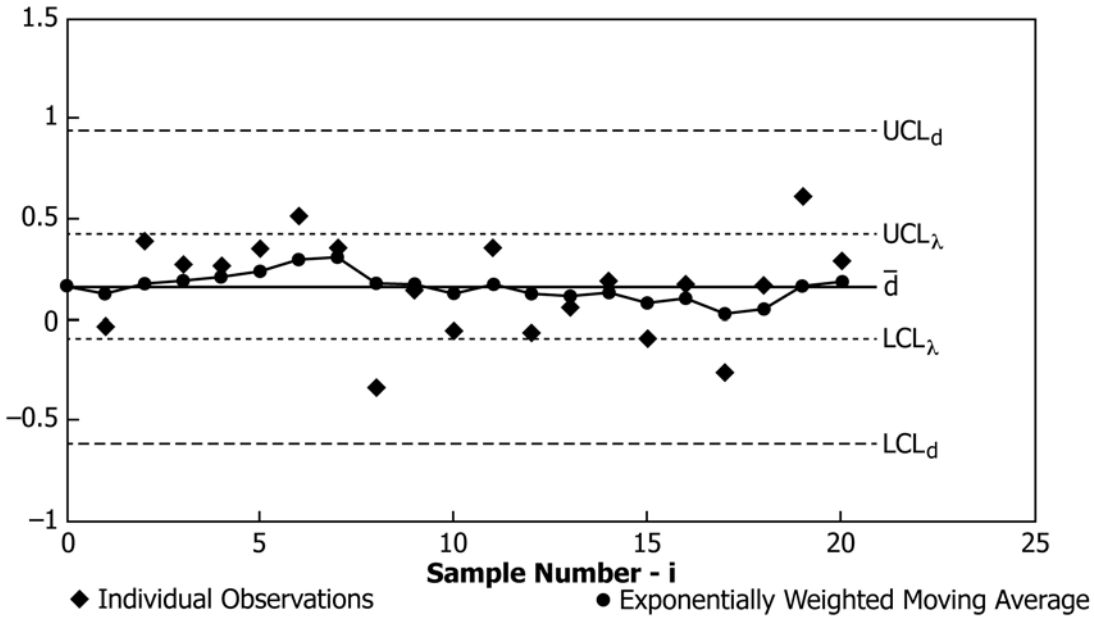

FIG. 5 Individual Observations and EWMA Charts

into agreement, the addition of the bias correction is considered to produce a new model. Validate this new model as described in Practices E1655. Once the new model has been validated, install it on the analyzer and validate the analyzer performance in accordance with the procedures described herein. If the bias is changed, it again produces a new model which again shall be revalidated in accordance with Practices E1655, and the analyzer performance shall again be validated.

14. Periodic Continual Validation by Plotting Control Charts of the Differences Between Methods

14.1 If the analyzer passes the initial validation test described in 12.1 or 13.1, periodically check the stability of the differences between the analyzer and primary test method using the control charts. Use the three types of control charts described in Practice D6299.

14.2 *Individual Values Control Chart for the Differences*—Begin by plotting the difference between the analyzer and primary method results, d_i, for the initial validation set of 20 samples as points, but do not connect the points. Calculate the mean difference and the standard deviation of the differences as instructed in Practice D6299, and use these to establish the upper and lower control limits.

14.3 *Exponentially Weighted Moving Average (EWMA) Control Chart:*

14.3.1 Overlay the Individual Values chart with an Exponentially Weighted Moving Average (EWMA) Control Chart for the differences as described in Practice D6299 (**1**).[3]

14.3.2 Calculate the control limits for the Exponentially Weighted Moving Average chart using a weight (lambda) of 0.2 to 0.4 as follows as described in Practice D6299. See Fig. 5.

14.4 *Moving Range of Two Control Chart:*

14.4.1 Construct a separate Moving Range of Two Control Chart for the initial 20 differences using the procedures described in Practice D6299.

14.4.2 See Fig. 6. Calculate control limits for the MR_2 chart as described in Practice D6299, and add them to the chart.

14.5 Collect a line sample at the appropriate validation interval. If the line sample spectrum is not an outlier or nearest neighbor inlier, determine the sample value by the primary test method. Compute the d_i, w_i, and MR_i values and plot them on the Individual Differences Control Chart, the Exponentially Weighted Moving Average Control Chart, and the Moving Range of Two Control Chart, respectively. Interpret the control charts in accordance with the instructions in Practice D6299.

15. Updating Control Limits

15.1 After a set of 20 additional periodic validation line samples have been collected, reevaluate the control limits for the three control charts to see if a statistically significant change in performance has occurred. Follow *Scenario 1 for Updating of Control Chart Parameters* in Practice D6299.

15.1.1 Calculate the bias and variance of the 20 new differences as described in Practice D6299. Perform a *t* test to see if any bias calculated is statistically significant as described in Practice D6299.

15.1.1.1 If the bias is not statistically significant, then the analyzer is expected to give essentially the same average result as the primary test method.

15.1.1.2 If the bias is statistically significant, then the user can be 95 % confident that the analyzer and the primary test method are not giving the same average results. The analyzer and primary test method validity are both suspect. Conduct further investigation of the analyzer function and operation and of the primary test method measurement to resolve the source of the bias. *Bias corrections of multivariate models are not permitted within the scope of this practice* (see Note 6).

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

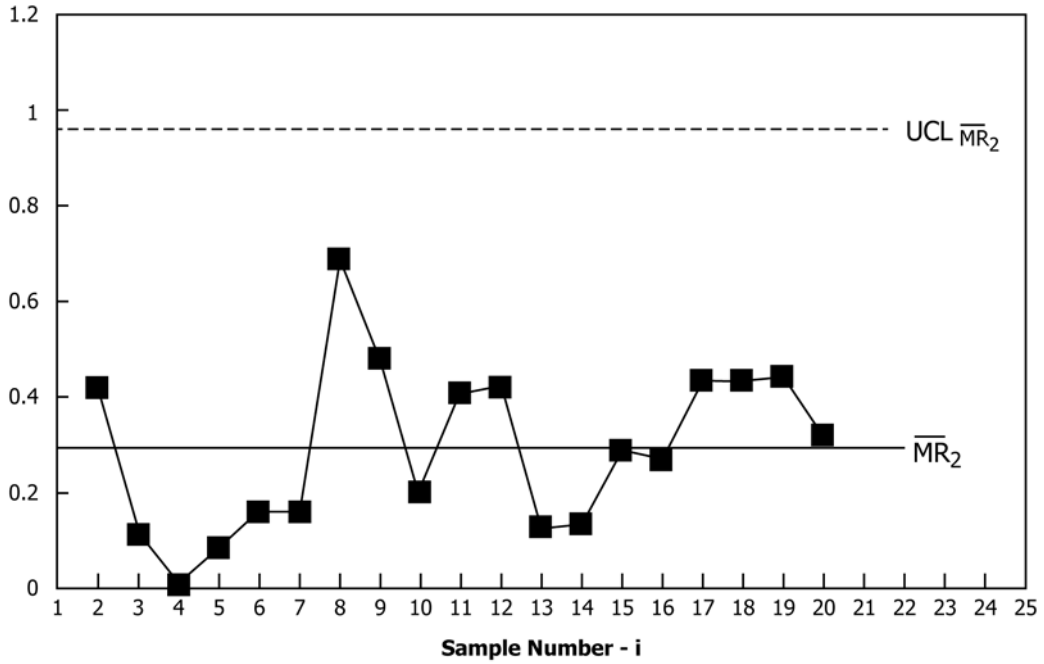

FIG. 6 Moving Range of Two (MR2) Chart

15.1.2 Compare the variance of the 20 new differences *y* to the variance previously calculated by an *F*-test as described in Practice D6299.

15.1.2.1 If the *F* value calculated is less than the critical *F* value as described in Practice D6299, and if the standard deviation of the new results is at least 72 % of the reproducibility of the primary test method, then the variance calculated for the 20 new results belongs to the same population as the previous variance. Pool the new results with the previous results to calculate a new analyzer variance as described in Practice D6299, and recalculate new control limits for all three control charts based on the pooled results.

15.1.2.2 If the *F* value calculated is greater than the critical value as described in Practice D6299, then there is a 95 % probability that the 20 new results come from a population that does NOT have the same variance as that estimated from the previous results, which suggests that a change has occurred in the entire validation process. Further investigation of key elements and procedures including, but not limited to the performance of the analyzer, the primary test method, and the sampling process, is warranted.

15.1.2.3 If the standard deviation for the new set of results is in the numerator in calculating the *F* value, and the *F* value is greater than the critical value, then the variance of the validation process has increased. Identify and correct the cause of the increase before continuing with the validation process. If no cause can be identified, it is recommended that the validation process be restarted with 20 new initial validation samples (see Section 13), and that analyzer results be marked invalid until the initial validation has been successfully completed.

15.1.2.4 If the standard deviation for the new set of results is in the demominator in calculating the *F* value, and the *F* value is greater than the critical value, then the variance of the validation process has decreased. Attempt to identify the cause for the improvement to determine if it can be maintained. If the improvement is not due to a special cause, and if the standard deviation of the 20 results is at least 72 % of the reproducibility of the primary method, then combine the results for the 20 new samples with the previous results to produce a new estimate of the validation process variance. Update the control limits appropriately. If the standard deviation of the results is not at least 72 % of the reproducibility of the primary test method, do NOT adjust the variance estimate or control limits.

NOTE 7—It is the user's responsibility to ensure that the procedures described in the latest revision of Practice D6299 are used in conjunction with this practice.

16. Analyzer Repeatability

16.1 Analyzer repeatability can be estimated directly from the analyzer results during periods when the process sample is relatively constant. Once a minimum of 25 analyses have been obtained, the results are plotted on control charts and statistically analyzed to estimate the analyzer repeatability.

16.1.1 Visually screen the results for unusual values. Use the Generalized Extreme Standardized Derivative method (see 13.3.2) to test for outliers among the results. Plot the results in chronological order and examine them for nonrandom patterns. Use a (normal) probability plot (see 14.3.2) to check that the results are normally distributed.

16.1.2 If the results are normally distributed, construct Individual Observation, Moving Range of Two, and Exponential Weighted Moving Average control charts for the results and establish limits (see Section 14).

16.2 Estimate the standard deviation for the analyses from the control charts as:

$$\hat{\sigma} = 0.89\,\overline{MR} \quad (4)$$

The analyzer repeatability is obtained by multiplying $\hat{\sigma}$ by 2.77.

NOTE 8—Practices E1655 defines a procedure for estimating the precision of the multivariate model. Since the Practices E1655 procedure generally involves spectral measurements of static samples under laboratory conditions, the Practices E1655 precision is expected to be somewhat better than what can be achieved in online application of the model. Similarly, statistical analysis of repetitive Level B performance tests may be used as an indication of analyzer repeatability and analyzer intermediate precision. However, since such performance tests do not necessarily include all potential sources of variation associated with the online measurement, the instrument precision may be somewhat better than what can be achieved during online measurement. If the analyzer repeatability measurements discussed previously cannot be performed, then the Practices E1655 model precision or instrument performance test precision may be used as an arbitrarily optimistic estimate of analyzer repeatability.

17. Analyzer Site Precision

17.1 The application of Practice D6708 procedures for the validation of analyzer performance requires that the analyzer site precision be known, or that the measurement of all validation samples be done in duplicate.

17.2 The analyzer site precision may be estimated using procedures described in Practice D7808.

18. Keywords

18.1 control chart; infrared analyzer; infrared spectrophotometers; IR spectroscopy; multivariate process; NIR spectroscopy; statistical quality assurance; validation

ANNEXES

(Mandatory Information)

A1. CONSIDERATIONS FOR QUANTITATIVE ONLINE PROCESS IR MEASUREMENTS

A1.1 Spectral data collection and computation parameters used for the collection of process sample spectra should generally be identical to those used in collecting the calibration spectra on which the multivariate model is based.

A1.1.1 The wavelength (frequency) range over which process sample spectra are collected shall be the same or greater than the range over which the calibration sample spectra were collected. If the range is greater, the additional data collected is discarded prior to application of the model.

A1.1.2 The optical and digital resolution at which process sample spectra are collected should be identical to that used in the collection of the calibration sample spectra.

A1.1.3 For instruments such as FT-IR where the spectra are obtained by mathematical processing of the raw spectral data, the processing conditions (for example, apodization, zero-filling, and so forth) employed in calculating the process sample spectra should be identical to those used in calculating the calibration sample spectra.

A1.1.4 Absorbances for the bands specified in this test method are expected to fall within the linear operating range of the spectrophotometer, as defined by the manufacturer, typically less than 1.0 absorbance units.

A1.1.5 If the measurement time (for example, number of averaged scans) is not the same for the process sample spectral measurement as for the measurement of the calibration sample spectra, then the user shall determine what effect this change has on the precision of the analyzer results.

A1.2 Spectrophotometer Cells and Other Infrared Sampling Methods

A1.2.1 One common process infrared measurement involves transmitting the infrared light through the sample while the sample is contained in a spectrophotometer cell. The spectrophotometer cell consists of two infrared transparent windows held apart at a fixed distance, the sample path length. The sample may flow through the cell during the spectral measurement, or the flow may be interrupted for the duration of the measurement.

TABLE A1.1 Common Path lengths for Liquid Hydrocarbon Analysis in the Infrared Region

NOTE 1—The path length used for the process measurements should be nominally the same as that used in collecting the data on which the calibration model is based.

Wavelength, nm	Frequency, cm^{-1}	Path length, mm
800-1100	12 500-9091	20-100
1100-1600	9091-6250	7-12
1600-2200	6250-4545	1-3
2000-6250	5000-1600	0.5
2500-25 000	4000-400	0.01-0.05

A1.2.2 Inspect spectrophotometer cells and verify that the cells contribute minimally to the measured absorbance of the sample. If contamination or deposition on the cell windows is suspected, clean windows with an appropriate solvent, or replace if necessary. Contamination can sometimes be detected by an increased baseline. Cell windows should also be examined for scratches and cracks.

A1.2.3 The optical path length is an important consideration in infrared spectroscopic measurements. Appropriate path lengths depend on the spectral range employed. Path lengths are chosen to keep the absorbance at analytical wavelengths within the linear operating range of the spectrophotometer. The most common path lengths for the infrared region are given in Table A1.1. The path length used for the process measurements should be nominally the same as that used in collecting the data on which the calibration model is based.

NOTE A1.1—Liquid viscosities may limit the use of flow cells in the

4000 to 400 cm^{-1} region. Internal reflection spectroscopy (see 6.2) may be more practical in this frequency range.

A1.2.4 Other sampling methods may be applicable to measurements conducted in some parts of the infrared region.

A1.2.5 The sample being analyzed may be contacted with an internal reflection element such that attenuated total reflectance occurs at the interface. Mid-infrared spectra are then measured by way of internal reflection spectroscopy.

A1.2.6 Transflection involves a measurement wherein the infrared radiation transmitted through the sample is returned through the sample by means of an external reflector. Some fiber optic probes employ transflection. Transflection doubles the effective path length of the cell since light passes through the sample twice.

A1.2.7 When check or test samples are being introduced, it is generally preferable to wash out the current sample with the next sample. The volume of sample used to flush the cell should be at least five times the volume between the sample inlet and cell exit point(s). When measurements are conducted on flowing samples, the flow through the optical cell should be high enough to ensure that a fresh sample is present for each spectral measurement.

A1.3 Fiber Optics

NOTE A1.2—Not all process IR analyzers are installed with fiber optics. This section applies only to analyzers that use fiber optics.

A1.3.1 Fiber optics, single-strand or multiple-strand fibers, can be employed to transmit light from the spectrophotometer to the sample and from the sample back to the spectrophotometer.

A1.3.2 Consult fiber and instrument manufacturers for proper selection, installation, and maintenance of fiber optic cables.

A1.4 In-Line Probes

A1.4.1 An in-line probe may be considered a spectrophotometer cell installed in a process pipe or side loop and connected to the spectrophotometer by optical fibers. In-line probes may be used in cases where the analysis is desired at process conditions (pressure and temperature), where it is difficult to install the required slip stream piping to permit safe withdrawal of a line sample for analysis, or where disposal of the sample after analysis may create an environmental hazard.

A1.4.1.1 Where possible, in-line probes should be installed to allow for their complete removal for the purpose of collecting backgrounds or instrument performance test data, and to allow for inspection of the probe for fouling or physical damage.

A1.4.1.2 For some installations, removal of the in-line probe involves excessive work, or exposes personnel to increased hazards. In this case, the probe cannot be inspected manually for fouling or physical damage. The total energy throughput of the system, and the baseline of the sample absorption spectra should be continuously monitored for evidence of fouling or damage.

A1.5 Sample Temperature

A1.5.1 Sample temperature greatly impacts the reproducibility of spectral measurements due to density changes and intermolecular interactions, and may consequently affect predicted values. The significance of temperature effects shall be separately established for every composition, component, or property measured.

A1.5.2 Temperature control of the reference material, samples, and process stream should be incorporated such that the temperature of all materials introduced into the spectrophotometer cell are constant and known. Alternatively, temperature variation over a specified range can be compensated for either in the multivariate calibration model or through pre-processing or post-processing, and the temperature for the process stream shall be controlled to within the range used for the calibration.

A2. INSTRUMENT PERFORMANCE TESTS

A2.1 Reference Materials for Instrument

A2.1.1 *Check Samples*—Check samples are generally used for conducting performance tests. Check samples are single, pure, liquid hydrocarbon compounds or mixtures of liquid hydrocarbon compounds of definite composition.

NOTE A2.1—If mixtures are utilized as check samples, they shall be prepared in a repeatable manner and, if stored, stored such that the mixture is stable over long periods of time. In preparing mixtures of liquid hydrocarbon materials, components should be accurately pipetted or weighed at ambient temperature. It is recommended that mixtures be independently verified for composition prior to use.

A2.1.1.1 The check sample is chosen such that its absorption spectrum is similar to the petroleum matrix of the application of interest.

A2.1.1.2 When possible, the check sample should contain the major functional groups associated with the process stream of interest.

NOTE A2.2—The near-infrared spectral region is a simplified spectrum for petroleum products in that the major bands are: aromatic, olefin, methyl, methylene, and oxygenates. For example, toluene is a frequently chosen reference material for gasoline range petroleum products or intermediates. Toluene contains two major functional groups associated with gasoline, aromatic, and methyl functional groups.

A2.1.1.3 The check sample should have significant absorbance at the wavelength(s) of interest. In order to adequately determine the photometric linearity of the instrument, the peak absorbance of a check sample should be similar to, and preferably slightly greater than, the largest absorbance expected from the process fluids.

A2.1.1.4 Mixtures can be used as check samples but their spectra may be adversely affected by temperature-sensitive interactions that may manifest themselves by wavelength and absorbance changes. Additionally, mixture composition may change with time due to differential evaporation if samples are not stored properly.

A2.1.2 *Test Samples*—A test sample is a process or product sample, or a mixture of process or product samples, whose spectrum is expected to be constant for the time period it is used in performance testing.

A2.1.2.1 Store the test sample in bulk quantities in controlled conditions such that the material is stable over time.

A2.1.2.2 Since test samples cannot be synthetically reproduced, they can only be used for performance testing for limited time periods. If test samples are used for this purpose, collection of historical data on a new test sample should be initiated before previous test samples are depleted. It is recommended that new test samples be analyzed sequentially with old test samples at least 15 times before they are used to replace the old test sample. The 15 analyses shall be performed over a time period that does not exceed one month in duration.

A2.1.3 *Optical Filters*—An optical reference filter is an optical filter or other optical device located in the spectrophotometer or the sample probe which produces an absorption spectrum which is known to be constant over time. This filter may be automatically inserted into the optical path to allow instrument performance tests to be performed.

A2.1.3.1 Optical filters are used principally with in-line probes when removal of the probe is inconvenient, precluding the use of check or test samples for routine instrument performance testing.

A2.1.3.2 If an optical filter is used routinely to check or correct the spectral data collection or computation, then the same filter cannot be used for instrument performance testing. For example, polystyrene filters are used to continuously check and correct the wavelength scale of some dispersive NIR spectrophotometers. For such systems, polystyrene filters should not be employed for instrument performance testing.

A2.2 Types of Performance Tests

A2.2.1 Three types of performance tests are described herein. ASTM Committee E13 has defined Level 0 tests to consist of a series of univariate instrument performance tests. The Level A and Level B tests defined herein are multivariate instrument performance tests intended to be a rapid pass/fail measure of the instrument performance.

A2.2.2 The Level 0, A, and B tests are intended to check the following spectrophotometer variables: baseline, path length, wavelength, spectroscopic resolution, and photometric precision and linearity.

A2.2.3 Level A tests involve the mathematical comparison of the spectrum of a check or test sample against a historical spectrum of the same material. Level B tests apply the actual process stream calibration model to analyze a check sample spectrum, a test sample spectrum, or the spectrum of an optical filter.

A2.2.4 Some Level 0 tests are specific for the type of spectrophotometer in use (Fourier transform, diode array, monochromator, acousto-optic tunable filter, and so forth), whereas, Level A and Level B tests are applicable to all spectrophotometers. Level 0 tests for some specific instruments have been suggested or approved by Committee E13. Tests that might be useful in Level 0 procedures include those discussed in Practices E275, E932, E1421 and E1944.

A2.2.5 If the Level A or B test fails, it may be useful to perform a Level 0 test to provide diagnostics which might pinpoint the cause of the failure.

A2.2.6 Level 0 and A tests can be developed prior to and utilized during and after the development of the process calibration model. Although, by its very nature, the Level B test can only be used after the process calibration model is developed, it has the added advantage of providing some information about the sensitivity of the calibration model to instrument performance parameters, especially when it is applied to test samples rather than check samples or optical filters.

A2.2.7 *Level 0 Tests:*

A2.2.7.1 Level 0 tests are not intended to provide absolute measures of instrument performance, but rather useful diagnostics that can be used to detect changes in instrument performance.

A2.2.7.2 Level 0 tests measure various significant instrument parameters by specific univariate type measurements performed on the spectrum of a check sample, a test sample, or an optical filter. Parameters most frequently checked are wavelength precision, spectral resolution or bandwidth, baseline levels, photometric noise, and photometric linearity. All of the parameters measured should be plotted on control charts and compared to historical values. The information derived is directly related to instrument performance and can be used for troubleshooting.

A2.2.7.3 If the manufacturer has not provided instructions or means, or both, for a level 0 test, and Committee E13 has not specified an appropriate test procedure for the specific type of instrument used in the analyzer, or if the sample specified by the Committee E13 procedure is incompatible with process operation, then the following guidelines can be used to develop a practical Level 0 test.

NOTE A2.3—A variety of algorithms can be used to calculate peak positions, photometric noise, baseline stability, and resolution from spectral data. Not all algorithms produce results of sufficient precision to be useful for instrument performance testing. The calculations in the following guidelines are intended as examples. The algorithms used in calculations of performance test results should be tested to demonstrate that they accurately track changes in instrument performance.

A2.2.7.4 A wavelength (or frequency) stability test is conducted by monitoring the positions of one or more absorbance peaks for a check sample, a test sample or an optical filter. It is recommended that the peak position be determined by the following steps:

(1) Compute the first derivative of the spectrum by applying the appropriate digital filter to the spectrum. A commonly used filter has been defined by Savitzki and Golay (**2**) with

corrections by Steiner, Termonia, and Deltour (**3**), with application criteria discussed by Willson and Polo (**4**). The latter reference discusses optimum filter parameters based upon the relationship between spectral bandwidth and digitization interval.

(2) Identify the zero crossing associated with the peak absorbance and compute its location by linear interpolation between the two adjacent points straddling the zero crossing. The zero crossing is taken as a measure of the peak position.

NOTE A2.4—The preceding test of wavelength stability can be affected by photometric noise. To minimize the effect of noise, the peaks used for wavelength stability testing should be less than 1.0 absorbance, and preferably below 0.7 absorbance.

A2.2.7.5 Photometric noise tests are conducted at two or more spectral regions, preferably areas of minimum absorbance. A spectral region used in the test covers at least eleven adjacent points. Two successive absorbance spectra of the check sample, the test sample, or the optical filter are collected. The second spectrum is subtracted from the first to generate a difference spectrum. The average value in each spectral region is calculated for the difference spectrum and the standard deviation about the average is calculated. The photometric noise is the standard deviation about the average.

$$\text{Photometric noise} = \sqrt{\frac{\sum_{i=1}^{n}\left(A_i - \bar{A}\right)^2}{n-1}} \qquad \text{(A2.1)}$$

A_i is the absorbance value of the difference spectrum at the ith data point, $\bar{A}$ is the average absorbance over the n data points.

NOTE A2.5—The preceding measurement of photometric noise can be affected by wavelength instability if there is a significant change in absorbance across the region used in the noise calculation.

A2.2.7.6 Baseline stability is calculated for the same regions used in the photometric noise test. For a single spectrum, the mean absorbance for each region is computed and compared to historical data. Variation from the historical value is taken as an indication of baseline instability.

A2.2.7.7 Spectral resolution at one or more peaks in the spectrum of a check sample, a test sample, or an optical filter should be monitored for stability. It is recommended that the spectral resolution of each peak be determined by the following steps.

(1) Compute the second derivative of the spectrum by applying an appropriate digital filter to the spectrum. A commonly used filter has been defined by Savitzki and Golay (**2**) with corrections by Steiner, Termonia, and Deltour (**3**), with application criteria discussed by Willson and Polo (**4**). The latter reference discusses optimum filter parameters based upon the relationship between spectral bandwidth and digitization interval.

(2) Identify the zero crossing on both sides of the minimum associated with the peak absorbance and computing their locations by linear interpolation from the two adjacent points straddling the zero crossing. The difference in the locations of the two zero crossings is taken as a measure of the spectral resolution.

NOTE A2.6—The preceding test of spectral resolution can be affected by photometric noise.

A2.2.7.8 Photometric linearity is tested using two peaks in the absorbance spectrum, one of which is the peak of maximum absorbance. The second peak is preferably less than half the absorbance of the maximum peak. Linear baselines for each peak are calculated from points of minimal absorbance on opposite sides of the peaks. The maximum absorbance for each peak is corrected for the baseline, and the ratio of the absorbances for the two peaks is calculated. The ratio is used to track changes in the photometric linearity.

NOTE A2.7—This test is sensitive to wavelength instabilities. A significant change in the ratio can be taken as evidence of a change in photometric linearity, only if wavelength stability has been demonstrated.

A2.2.8 *Level A Tests:*

A2.2.8.1 A Level A performance test is a pass/fail test that is sensitive to all of the Level 0 parameters. Level A tests do not identify specific failure modes, but merely indicate if the instrument performance is within historical bounds. In this test, the spectrum of a check sample, a test sample or an optical filter is compared to a historical spectra of the check sample, the test sample, or the optical filter by multivariate methods (least squares fitting or a PCR/PLS model). This procedure can provide some information about specific instrument parameters, but essentially looks for deviations in the residual spectrum as compared to the historical residual spectra. The spectral range used in Level A tests should be comparable to that used in the calibration model. If the spectrum of the check sample, the test sample, or the optical filter used in the Level A test contains absorptions that are significantly higher than those of the calibration samples, then these peaks can be excluded from the Level A fit.

A2.2.8.2 *Level A Tests Using a Least Squares Method*—In a Level A test, a least square fit of the current spectrum of the check sample, test sample, or optical filter is conducted against a historical spectrum of the same material. Baseline terms may be included in the fit to compensate for variations in baseline, and scaling may be applied to compensate for path length variations. The types of compensations (baseline or path length) used in the fit should be similar to those employed in the multivariate model used for the actual analyzer measurement. Methodology for calculating the least square fit is discussed by Blackburn (**5**) and by Antoon, Koenig, and Koenig (**6**). A typical least squares model could be:

$$g = ah + b\lambda + c1 \qquad \text{(A2.2)}$$

where g is the vector containing the current spectrum of the check sample, the test sample, or the optical filter, h is the vector containing the historical spectrum of the check sample, the test sample or the optical filter, λ (v for frequency based spectra) is the vector of the wavelength axis values for spectra g and h, and 1 is a vector of ones. a is a coefficient for scaling the historical spectrum to match the current spectrum. b is a coefficient which scales λ to provide a baseline correction which is linear in wavelength (or frequency). c is a coefficient for a baseline offset. The coefficients a, b, and c are first determined and then used to estimate the spectrum of the current sample $\hat{g}$. The residuals from the fit are the difference

between the measured and estimated values for the data points, $g - \hat{g}$. The residuals from the fit are squared, and summed. The resulting measure, herein referred to as the spectral residual, is used as a measure of changes in the instrument performance. This spectral residual should be plotted on control charts. Additionally, the scaling and baseline coefficients can be monitored as an additional measure of instrument performance.

NOTE A2.8—Any function of the sum of the squares of the residuals can be used, for example, the square root.

A2.2.8.3 *Level A Tests Using a PCR or PLS Method*—To perform a Level A test using PCR or PLS, one shall first develop an appropriate model. A series of historical spectra for the check sample, the test sample or the optical filter are analyzed without mean centering by a PCR or PLS regression algorithm using 100 % for the compositional value to generate the Level A model. Generally, only one variable should be retained in the model since all the spectra are of the same material. The type of pre-processing or post-processing done in the Level A test model should be comparable to that done in the multivariate calibration models being used on the analyzer. The principal component or latent variable resulting from this model is applied to a current spectrum of the check sample, the test sample, or the optical filter to generate a calculated spectrum of the test sample, the check sample, or the optical filter. From this calculated spectrum, the spectral residual can be computed as described previously. The spectral residual can be charted to determine if the instrument is operating within historical specifications.

NOTE A2.9—Chemometricians might refer to the analysis described in A2.2.8.3 as Principal Component Analysis rather than Principal Components Regression. However, the object here is to allow the Level A test to be developed and applied using the same chemometric software employed in the development and application of the multivariate calibration model.

A2.2.9 *Level B Tests:*

A2.2.9.1 A Level B performance test analyzes the spectrum of a check sample, a test sample, or an optical filter against the models in use on the analyzer system. As such, Level B tests can not be performed during calibration. Level B tests monitor the instrument performance for deviations to which the calibration model is sensitive. Tests on a limited number of samples are not rigorous, but failure in these tests are indicative that the analyzer operation has changed. The spectrum of the check sample, the test sample, or the optical filter is analyzed using the multivariate model normally applied to line samples. The predicted value (property or component concentration), the Mahalanobis distance, and the spectral residuals are again compared to historical values to detect any change in the analyzer performance.

A2.3 Performance Test Charts

A2.3.1 Performance test results should be plotted on charts and examined for trends. Such trend analysis may provide early warnings of possible analyzer problems.

A2.3.2 Individual Value Control Charts, Exponentially Weighted Moving Average Control Charts, and Moving Range of Two Control Charts (see Section 13) can be used to detect statistically significant changes in instrument performance. However, the statistical control limits associated with these charts will not necessarily be used to judge the performance test results. Instead, some performance test results are typically compared to action limits as described in A2.4.

A2.3.3 For some performance tests, the test results are expected to trend continuously in one direction until such time as the analyzer is serviced. For example, the energy output of an infrared source is expected to decrease continuously as the source ages, until such time, as the source is cleaned or replaced. The decreased energy may be observed as an increase in the Level 0 photometric noise, or as an increase in the Level A spectral residual. The daily change in energy, noise, or residual may be large relative to the precision with which these values can be measured, but have tolerable effect on the accuracy or precision of the analyzer results. For such tests, control charts and limits as discussed in Section 13 are inappropriate. An action limit for such tests needs to be determined from historical data or simulations as discussed in A2.4.

A2.3.4 For some performance tests, the test results are expected to vary randomly about a fixed point. For example, for a properly operating instrument, the Level 0 wavelength value might be expected to vary randomly about some average value. For such tests, the control charts and control limits described in Section 13 can be usefully employed to set initial action limits in the absence of historical data. Such initial action limits may be loosened if statistically significant performance changes detected by the control charts are not found to have significant effect on the validity of analyzer results.

A2.3.5 Since Level B composition or property results for check or test samples are most directly comparable to actual analyzer results, the Level B composition or property estimates are most amenable to statistical control charting. Action limits for Level B composition or property estimates can be set to the control limits described in Section 13.

A2.4 Performance Test Action Limits

A2.4.1 Calibration models differ greatly in their sensitivity to various aspects of instrument performance, and each application differs in what constitutes an acceptable tolerance to changes in the results caused by variations in instrument performance. Although instrument performance tests are useful in their own right, the process analyst should be concerned with how changes in the instrument performance propagate through the calibration model and affect the calculated results. Historical databases or simulations that define acceptable performance for one application may not be appropriate for another application. In addition, the level of performance required by an application may be changed by the updating of the calibration model.

A2.4.2 *Setting Action Limits Based on Historical Data for Performance Tests:*

A2.4.2.1 Performance tests provide measures of instrument performance. These measures can be compared to historical data for the same tests in order to judge the adequacy of analyzer performance. If historical data exist, limits for each test can be set and the performance can be judged against these

limits. If historical data do not exist, it will be necessary to collect it as a standard part of the analyzer operation, and such collection will eventually allow performance limits to be established. The collection of the historical database for performance tests is an integral part of the analyzer operation, and continues for the life of the analyzer.

A2.4.2.2 If the analyzer results for validation samples are in agreement with the results from the primary method, then the results for the performance tests conducted during the same time period should be considered an example of acceptable instrument performance and added to the historical database.

A2.4.2.3 If the analyzer results are not in agreement with the results from the primary method, and if the primary method is within statistical quality control, the results from the performance tests may be examples of unacceptable instrument performance, particularly if the results from the performance tests are inconsistent with the historical database. Examples of unacceptable instrument performance can be used to set action limits for future performance tests.

A2.4.2.4 It is strongly recommended that, at the time the multivariate model is developed, spectra of the check sample, the test sample, or the optical filter be collected along with spectra of the calibration and model validation samples. Performance tests can be applied to this data to determine the level of performance at the time of calibration. If a calibration model was developed and validated, then the level of performance measured during the calibration period is adequate to produce the precision demonstrated during calibration and validation of the model.

A2.4.2.5 Changes in analyzer performance that are detected by Level 0 tests may or may not produce a significant change in the results produced by the analyzer. Different types of multivariate models differ significantly in their sensitivity to various aspects of analyzer performance. By plotting the Level 0 test results against analyzer results on control charts, conditions that lead to invalid analyzer results can be identified, and action limits for each Level 0 test can eventually be established.

A2.4.2.6 Increases in the spectral residuals that are detected by Level A tests will generally reflect some change in the results produced by the analyzer. Even if the analyzer result does not change, the spectral residuals measured as part of the outlier testing will generally be expected to increase. The level of increase that can be tolerated can be determined by plotting the Level A test spectral residuals against analyzer results, and determining the maximum level at which valid analyzer results are produced.

A2.4.2.7 Changes in the values produced by a Level B test are the most straightforward to interpret since the values are directly comparable to the analyzer results. If the analysis of the spectrum of the check sample, the test sample, or the optical filter is an interpolation of the model, then limits can be set directly based on the desired performance of the analyzer. If the analysis of the spectrum of the check sample, the test sample, or the optical filter is an extrapolation of the model, exercise care in setting limits since the extrapolated result may be more sensitive to small changes in instrument performance than analyses that are interpolations of the model. This is known as leverage. In this case, initial limits should be confirmed by plotting the Level B results against analyzer results and determining the levels at which valid analyzer results are produced.

NOTE A2.10—Any one test sample, check sample, or optical filter only tests a small portion of the multivariate model space, and may not be sensitive to all aspects of analyzer performance. The Level 0, A, and B tests are intended to detect possible analyzer failure modes. Acceptable performance as measured by Level 0, A, and B tests is necessary but not sufficient by themselves for demonstrating valid analyzer performance. Comparison of analyzer results to in control, primary method laboratory values is also necessary to demonstrate the validity of analyzer results.

A2.4.3 *Determining Performance Action Limits by Simulating Instrument Response Changes:*

A2.4.3.1 An alternative procedure for determining action limits for instrument performance tests is to take actual, diverse, but representative spectra that are predicted well by the model, and to mathematically modify these spectra to simulate the expected variations in the instrument performance. The model sensitivity, for example, the change in the results per unit change in a performance parameter, can be estimated and used to establish action limits for each performance parameter based on the error tolerance for the application. Instrument performance parameters which can be modeled include wavelength (frequency) shifts, baseline shifts, changes in photometric noise, resolution changes, and detector linearity changes. The importance of different performance parameters is both application and instrument type dependent. Historical data for Level 0 performance tests are the best guide to the type of response changes that should be modeled for a given instrument type.

A2.4.3.2 For example, the sensitivity of an analyzer to baseline drift can be simulated by adding various baselines to a set of representative spectra, analyzing these spectra with the calibration model, and determining the change in the results as a function of the added baseline. The added baseline can, for example, be parameterized in terms of offset, slope, and curvature so that the effects of each can be determined.

A2.4.3.3 For example, the sensitivity of an analyzer to wavelength (frequency) shift can be simulated by shifting the wavelength (frequency) of a set of representative spectra, analyzing these spectra with the calibration model, and determining the change in the results as a function of the shift. If the shift is accomplished by way of interpolation of the spectra, exercise care that the interpolation function does not smooth or deresolve the spectra.

A2.4.3.4 Changes in instrument performance seldom affect only one aspect of that performance. If simulations are used to set action limits for performance tests, it is essential that multiple performance parameters be varied simultaneously. The magnitude of the changes to the performance parameters that should be simulated are best obtained from examination of historical data on Level 0 performance tests conducted on the type of instrument used in the analyzer.

A2.5 Tests for In-Line Probes

A2.5.1 *Option A*—Removal of the in-line probe from the process.

A2.5.1.1 Whenever possible, it is preferable to remove the in-line probe from the process for the purpose of conducting an instrument performance test.

A2.5.1.2 Removal of the in-line probe allows the entire optical path to be examined during the performance test. Fouling or physical damage to the probe is more readily detected.

A2.5.1.3 It will generally be necessary to clean the probe before conducting instrument performance tests to remove any residual process sample which could contaminate the check or test sample used in the tests. Similarly, it may be necessary to clean the probe after the tests if the check or test sample used in the tests is incompatible with the process being measured.

A2.5.2 *Option B*—Temporarily disconnecting the in-line probe.

A2.5.2.1 For probes connected to the analyzer by optical fibers, disconnect the fibers at the probe. Reconnect the fibers to an auxiliary probe, to a cuvette holder equipped with appropriate collimating optics or to a similar device. Collect the spectrum of the check or test sample and continue with the Level 0, A, or B test as described previously. Following the tests, reconnect the fibers to the in-line probe.

NOTE A2.11—If test samples are used for the instrument performance tests, it may be preferable to enclose the sample in a sealed cuvette to prevent differential evaporation of components and thus change in the chemical composition of the sample with time. Sealing these mixtures does not necessarily protect against thermal or photochemical degradation which can also alter chemical composition.

A2.5.2.2 When the in-line probe is not included in the optical path during the instrument performance tests, the integrity of the probe with respect to fouling and physical damage are not tested. Either of these two problems could contribute to invalid results.

A2.5.2.3 Fouling or contamination of the probe surface can sometimes be detected as changes in the baseline of the sample absorbance spectrum. The baseline should be monitored during normal operation for evidence of fouling.

A2.5.2.4 Physical damage to the probe could contribute to invalid results. The total energy throughput of the optical system should be monitored during normal operation for evidence of probe damage.

A2.5.3 *Option C—Using a Reference Channel:*

A2.5.3.1 If the analyzer is equipped with multiple optical channels, an alternative procedure is to dedicate one of the optical channels as a reference channel for use in instrument performance tests. Level 0, A, or B tests are performed over the reference channel at the required interval.

A2.5.3.2 Performance tests conducted over a reference channel do not test the entire optical path used for online analyses. While such tests may detect changes in source, spectrophotometer or detector performance, they are not affected by any changes in the fibers or probe in the online sample channel, nor are changes seen in the reference channel necessarily mirrored in the other channels. It is the analyzer vendor or user's responsibility to demonstrate that performance measured on the reference channel is representative of performance on other channels.

A2.5.3.3 Energy throughput and sample absorbance spectrum baseline should be monitored on each online channel for evidence of probe fouling or physical damage, or for changes in fiber transmittance.

A2.5.4 *Option D—Use of An Optical Reference Filter:*

A2.5.4.1 The spectrum of the optical reference filter is obtained by first acquiring a spectrum of the current online sample, inserting the filter into the optical path, and collecting a spectrum of the filter (plus sample). The absorbance spectrum of the filter is calculated as follows:

$$A_{filter} = -log\left(\frac{Spectrum_{filter+sample}}{Spectrum_{sample}}\right) \quad \text{(A2.3)}$$

Implicit to the successful use of this option is the assumption that the sample composition does not change significantly over the time required to collect the two spectra. The spectra should be collected as quickly as possible and in rapid succession. Testing should be performed during periods when the process is relatively stable to avoid compositional changes in the sample spectrum. Results of the tests using this option are comparable only when the tests are run with identical spectral acquisition times.

A2.5.4.2 Level 0, A, or B tests are conducted on the absorbance spectrum of the optical filter. Tests conducted in using this option should be designed to avoid spectral ranges where the sample absorptions will be strong (>1.0 absorbance), since the absorbance spectrum of the filter may be excessively noisy in such regions.

A2.5.4.3 Option D will not detect fouling of the probe since the optical effects of such fouling will be present in both of the spectra ratioed in Annex A3. Performance tests conducted using optical filters can be supplemented with baseline and photometric noise tests done on online spectra. Such tests should be performed in regions where the sample absorbance is known to be minimal.

A2.5.5 If options B-D are used for instrument performance testing, then it is recommended that the in-line probe be removed periodically for inspection and cleaning. The period between such removals will depend on the usage, and will, by necessity, be based on experience from the same or similar installations. For new applications, a period of one month is suggested until longer (or shorter) times are justified by process experience.

TABLE A3.1 Outlier Detection Methods

Type Test	Method	Computation	Outlier Detection Limit	Reference
Leverage or Scores Range Tests[A] (Mandatory)	h D^2	$x^t(XX^t)^{-1}x$	h_{max} or D^2_{max} model from calibration	E1655
	D	$\sqrt{D^2}$	D_{max} model	E1655
	M-Distance Ratio	$\frac{D^2}{\min\left(\frac{2k}{n}, 1\right)}$	$\frac{D^2_{max}}{\min\left(\frac{2k}{n}, 1\right)}$	Ref (3)
	Scores Range Test	For PCR, $S_{cal} = X^tL\Sigma^{-1}$, $s_u = x^tL\Sigma^{-1}$ For PLS, S_{cal} = *PLS scores for model* $s_u = W^tx_u$	< max (S_{cal}) and > min (S_{cal})	E1655
Spectral Residual Tests[B] (Mandatory)	RMSSR	$RMSSR = \sqrt{\frac{(\hat{x}-x)^t(\hat{x}-x)}{f}}$	$\left[\sum \frac{RMSSR_{anal}(i)}{RMSSR_{cal}(i)}\right] \times RMSSR_{max}$	E1655
	F-Ratio Test	$F_{ratio} = \frac{(\hat{x}-x)^t(\hat{x}-x)\,n}{\sum_{i=k+1}^{n}\lambda_i}$	F-Test	Ref (5)
Nearest Neighbor Test (Optional)	Nearest Neighbor	local D^2 = $\min[(x-x_i)^t(XX^t)^{-1}(x-x_i)]$	local D^2_{max} model	Ref (4)

[A] One leverage test is required for each sample during measurement.
[B] One Spectral Residual test is required for each sample during measurement.

A3. OUTLIER DETECTION METHODS

A3.1 Outlier detection methods are given in Table A3.1.

A3.2 Leverage (Mahalanobis Distance) and Scores Range Tests

A3.2.1 In this practice, a leverage test or a scores range test shall be used for detection of spectral outliers during analysis.

A3.2.1.1 The leverage statistic, h, is sometimes seen in the form of D^2 which is the Mahalanobis Distance squared. A discussion of the calculation of h is described by Eqs 61 to 66 in Practices E1655. If x is a vector containing the spectrum being analyzed, and X is a matrix whose columns are the calibration spectra, then a general expression for the calculation is given as follows:

$$h = x^t\left(XX^t\right)^{-1}x \qquad (A3.1)$$

NOTE A3.1—Commercial software packages use numerous variations on the leverage statistic. The leverage statistic is sometimes referred to as the hat matrix or as the Mahalanobis Distance, D^2 (although it is actually the square of the distance). Various commercial software packages may use D instead of D^2. Some software packages may scale h (or D^2) by n(or $n - 1$ if mean-centered) to obtain a statistic that is independent of the number of calibration samples. If this scaled statistic is further multiplied by $(n\text{-}k\text{-}1)/(nk)$, a statistic that has an F distribution is obtained (Eq 4). The leverage statistic, h, is preferred here since it is easily related to the number of samples and variables. Model developers should attempt to verify exactly what is being calculated. Both mean-centered and not-mean-centered definitions for h exist, with the mean-centered approach preferred. Regardless of whether mean centering of data is performed, the statistic designated h has valid utility for outlier detection.

Each row of X corresponds to a specific wavelength (or frequency) which was included in the calibration model. In many applications, the rows in X will be a subset of the spectral elements collected by the instrument. The matrix (XX^t) in [Eq A3.1] cannot be inverted unless the number of wavelengths (rows) in X is less than the number of calibration samples (columns) in X. Thus Eq A3.1 is strictly applicable only to *MLR*. If the number of wavelengths (or frequencies) exceeds the number of calibration samples, then the inverse of (XX^t) is approximated. The PCR and PLS involve two different methods for estimating this inverse. The corresponding equations for calculating h are obtained by substituting the PCR and PLS approximations into Eq A3.1. For more details, the user is referred to Practices E1655.

A3.2.1.2 During analysis, h is the leverage statistic of the sample spectrum.

A3.2.1.3 Other leverage functions can be used rather than h or D^2 as a valid outlier detection statistic. For example, the ratio of h to $2k/n$ is sometimes used. Samples for which the ratio exceeds $h_{max}/(2\ k/n)$ are then considered outliers.

A3.2.2 The scores range test is an alternative to the leverage test. Each of the scores for the sample being analyzed is compared to the range for that score for the calibration (training) set. If all of the scores for the sample being analyzed fall within the ranges for the calibration (training) set, then the spectrum is considered to be within the spectrum population. For the number of factors used, if any score for the sample being analyzed is greater than the maximum score or less than the minimum score from the calibration (training) set, then the spectrum is deemed to be an outlier.

A3.2.3 The analyzer results for a sample which lies outside the h_{max} for the calibration are considered to be invalid since they represent an extrapolation of the model. D^2 and D^2_{max} are the squares of the Mahalanobis Distance for the sample spectrum and the maximum Mahalanobis Distance for the calibration respectively. Either h, D or D^2 can be used as an outlier diagnostic.

NOTE A3.2—h will generally be less than $3k/n$ where k is the number of variables (MLR wavelengths, PCR Principal Components, PLS latent variables, and so forth) used in the model and n is the number of calibration samples. In most cases, calibration samples with h greater than $3\ k/n$ should have been eliminated as outliers during the development of the model if Practices E1655 was followed. Exceptions to this rule occur when repeated application of the $3\ k/n$ rule to successively smaller models continues to identify outliers past the point where 10 % of the calibration samples have been eliminated. In this case, the model built with 90 % of the original calibration samples may have a h greater than $3k/n$.

A3.3 Spectral Residuals

A3.3.1 Spectral residuals shall be used to detect when the spectrum being analyzed contains absorptions that were not present in the calibration samples. Such spectra are extrapolations of the calibration model.

A3.3.2 The spectral residual is given by $\hat{x} - x$ where $\hat{x}$ is the spectrum estimated from the model loadings and x is the measured spectrum. For example, for PCR, the spectral residual is given by:

$$s^t \sum L^t - x^t \tag{A3.2}$$

where $\hat{x}^t = s^t \sum L^t$ is the calculated spectrum for the sample under test based on the calibration model (see Eqs. 68 to 70, in Practices E1655). s is the vector of scores for the sample being tested and $\sum$ and L are the singular values and loading vectors for the calibration model. The Root Mean Square Spectral Residuals (RMSSR) is calculated as follows:

$$RMSSR = \sqrt{\frac{(\hat{x} - x)^t (\hat{x} - x)}{f}} \tag{A3.3}$$

where f is the number of data points (wavelengths or frequencies) per spectrum used in the model. The Upper Control Limit for an individual measurement can be calculated using:

$$\left[\sum \frac{RMSSR_{anal}(i)}{RMSSR_{cal}(i)} \right] \times RMSSR_{max} \tag{A3.4}$$

as shown in Table A3.1. $RMSSR_{anal}(i)$ are RMSSR values for replicate spectra of samples which were used in the calibration model, $RMSSR_{cal}(i)$ are the RMSSR values for the calibration spectra of the same samples, and $RMSSR_{max}$ is the maximum RMSSR value for the calibration (see Practices E1655, Section 16).

A3.3.3 *Residual F-Ratio Test*—The F_{ratio} test may be used to test spectral residuals. The F_{ratio} value calculated for based on the spectral residuals is compared to $F(\alpha,1,f)$. f is the number of degrees of freedom in the calibration model. $f=n-k$ if the model is not mean centered, and $f=n-k-1$ if the model is mean centered, where n is the number of calibration samples and k is the number of variables in the model. The value for the F_{ratio} is calculated as follows:

$$F_{ratio} = \frac{(\hat{x} - x)^t (\hat{x} - x) n}{trace\left[\left(\hat{X} - X \right)^t \left(\hat{X} - X \right) \right]} \tag{A3.5}$$

For a PCR model, the F_{ratio} for spectral residuals is calculated as:

$$F_{ratio} = (\hat{x} - x)^t (\hat{x} - x) n / \sum_{i=k+1}^{n} \lambda_i \tag{A3.6}$$

where the summation is over the λ_i eigenvalues for principal components that were left out of the model. The F_{ratio} value is calculated and compared to $F(\alpha,1,f)$. An $F_{ratio} \geq F(\alpha,1,f)$ is considered to be significant, indicating that analyzer result obtained for this sample are invalid.

A3.4 Nearest Neighbor Distance

A3.4.1 If the calibration sample spectra are distributed relatively uniformly over the variable space of the calibration model, then the leverage statistic discussed above is adequate to determine if a spectrum being analyzed is an interpolation of the model. If the spectrum produces an h less than h_{max} (and a RMSSR less than the limit), then it is reasonable to assume that the sample belongs to the same population as the calibration samples. However, if the calibration sample spectra are clustered within the variable space, the spectrum being analyzed can have an h less than h_{max} yet fall into a relatively unpopulated portion of the calibration space. In this case, the sample spectrum may not belong to the same population as the calibration sample spectra, and the results produced by application of the model may be invalid. Under these circumstances, it is desirable to employ a Nearest Neighbor Distance test to detect samples that fall within voids in the calibration space.

A3.4.2 Nearest Neighbor Distance, or the relative D or D^2, measures the distance between the spectrum being analyzed and individual spectra in the calibration set.

$$\text{Relative } D^2 = \min \left[(x - x_i)^t (X X^t)^{-1} (x - x_i) \right] \tag{A3.7}$$

A3.4.3 Relative D^2 values are calculated for all the calibration sample spectra. A maximum relative D^2 value is determined. This value represents the largest distance between calibration sample spectra.

A3.4.4 During analysis, the relative D^2 is calculated for the process sample spectrum. If the calculated value is greater than the maximum relative D^2 from A3.4.3, then the minimum distance between the process sample spectrum and the calibration spectra is greater than the largest distance between calibration sample spectra, the process sample spectrum falls within a sparsely populated region of the calibration space. Such samples are referred to as Nearest Neighbor Inliers.

REFERENCES

(1) Hunter, S. J., "The Exponentially Weighted Moving Average," *J. Quality Tech.*, Vol 18, No. 4, 1986, pp. 203–209.

(2) Savitzky, A., and Golay, M. J. E., "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Anal. Chem.*, Vol 36, 1964, pp. 1627–1639.

(3) Steiner, J., Termonia, Y., and Deltour, J., "Comments on Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Anal. Chem.*, Vol 44, 1972, p. 1906.

(4) Willson, P. D., and Polo, S. R., "Polynomial Filters of Any Degree," *J. Opt. Soc. Am.*, Vol 71, 1981, pp. 599–603.

(5) Blackburn, J. A., *Anal. Chem.*, Vol 37, 1965, p. 1000.

(6) Antoon, M. K., Koenig, J. H., and Koenig, J. L., *Appl. Spectrosc*, Vol 31, 1977, p. 578.

(7) Iglewicz, B., and Hoaglin, D. C., *Volume 16: How to Detect Outliers*, ASQC Quality Press, Milwaukee, WI, 1993.

(8) Shapiro, S. S., *Volume 3: How to Test Normality and Other Distribution Assumptions*, ASQC Quality Press, Milwaukee, WI, 1990.

(9) Quesenberry, C. P., "SPC Q-Charts for Start-up Processes and Short or Long Runs," *J. Quality Tech.*, Vol 23, No. 3, 1991, p. 215.

(10) Martens, H., and Naes, T., *Multivariate Calibration*, John Wiley and Sons, New York, NY, 1989, pp. 274–296.

(11) Malinowski, E. R., *Journal of Chemometrics*, Vol 3, 1988, p. 49.

(12) Shenk, J. S., and Westerhaus, M. O., *Crop Science*, Vol 31, 1991, p. 469.

SUMMARY OF CHANGES

Subcommittee D02.25 has identified the location of selected changes to this standard since the last issue (D6122 – 13) that may impact the use of this standard. (Approved June 1, 2015.)

(1) Revised the Introduction, Section 1, 5.1, 8.1, 9.3, and 9.5.

(2) Added Practice D7717 to Referenced Documents.

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 (e-mail); or through the ASTM website (www.astm.org). Permission rights to photocopy the standard may also be secured from the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Tel: (978) 646-2600; http://www.copyright.com/