Designation: ~~D6122 – 20a~~ D6122 – 21

Standard Practice for Validation of the Performance of Multivariate Online, At-Line, Field and Laboratory Infrared Spectrophotometer, and Raman Spectrometer 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 five sequential activities. *(1)* **Correlation**—Prior to the initiation of the procedures described in this practice, a multivariate model is derived which relates the spectrum produced by the analyzer to the Primary Test Method Result (PTMR). *(1a)* **If the analyzer and Primary Test Method (PTM) measure the same material**, then the multivariate model directly relates the spectra to PTMR collected on the same samples. Alternatively *(1b)* **if the analyzer measures the spectra of a material that is subjected to treatment prior to being measured by the PTM**, then the multivariate model relates the spectra of the untreated sample to the PTMR for the same sample after treatment. *(2)* **Analyzer Qualification**—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 *(3)* **Local Validation**—A local validation is performed using an independent but limited set of materials that were not part of the correlation activity. This local validation is intended to demonstrate that the agreement between the Predicted Primary Method Test Results (PPTMRs) and the PTMRs are consistent with expectations based on the multivariate model. *(4)* **General Validation**—After an adequate number of PPTMRs and PTMRs have been accrued on materials that were not part of the correlation activity and which adequately span the multivariate model compositional space, a comprehensive statistical assessment can be performed to demonstrate that the PPTMRs agree with the PTMRs to within user-specified requirements. *(5)* **Continual Validation**—Subsequent to a successful local or 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, fourth, and fifth 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 can 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. The correlation is deemed valid only for the specific percentage addition and type of biofuel material used in its development.

[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 ~~Oct. 1, 2020~~May 1, 2021. Published ~~November 2020~~July 2021. Originally approved in 1997. Last previous edition approved in 2020 as ~~D6122 – 20.~~D6122 – 20a. DOI: ~~10.1520/D6122-20A.~~10.1520/D6122-21.

1. Scope*

1.1 This practice covers requirements for the validation of measurements made by laboratory, field, or process (online or at-line) infrared (near- or mid-infrared analyzers, or both), and Raman analyzers, 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 verification that the uncertainties associated with the degree of agreement between the results calculated from the infrared or Raman measurements and the results produced by the PTM used for the development of the calibration model meets user-specified requirements. Initially, a limited number of validation samples representative of current production are used to do a local validation. When there is an adequate number of validation samples with sufficient variation in both property level and sample composition to span the model calibration space, the statistical methodology of Practice D6708 can be used to provide general validation of this equivalence over the complete operating range of the analyzer. For cases where adequate property and composition variation is not achieved, local validation shall continue to be 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 additive 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.

1.1.3 In some cases, a two-step procedure is employed. In the first step, the analyzer and PTM are applied to the measurement of a blendstock material. In a second step, the PPTMRs produced in Step 1 are used as inputs to a second model that predicts the results obtained when the PTM is applied to the analysis of the finished blended product produced by additivation to the blendstock. If the analyzer used in the first step is a multivariate spectroscopic based analyzer, then this practice is used to access the degree of agreement between PPTMRs and PTMRs. Otherwise, Practice D3764 is used to compare the PPTMRs to the PTMRs for this blendstock to determine the degree of agreement. Since this second step does not use spectroscopic data, the validation of the second step is done using Practice D3764. If the first step uses a multivariate spectrophotometric analyzer, then only samples for which the spectra are not outliers relative to the multivariate model are used in the second step. Note that the second model might accommodate variable levels of additive material addition to the blend stock.

1.2 Multiple physical, chemical, or quality properties of the sample under test are typically predicted from a single spectral measurement. In applying this practice, each property prediction is validated separately. The separate validation procedures for each property may share common features, and be affected by common effects, but the performance of each property prediction is evaluated independently. The user will typically have multiple validation procedures running simultaneously in parallel.

1.3 Results used in analyzer validation are for samples that were not used in the development of the multivariate model, and for spectra which are not outliers or nearest neighbor inliers relative to the multivariate model.

1.4 When the number, composition range or property range of available validation samples do not span the model calibration range, a local validation is done using available samples representative of current production. When the number, composition range and property range of available validation samples becomes comparable to those of the model calibration set, a general validation can be done.

1.4.1 *Local Validation:*

1.4.1.1 The calibration samples used in developing the multivariate model must show adequate compositional and property variation to enable the development of a meaningful correlation, and must span the compositional range of samples to be analyzed using the model to ensure that such analyses are done via interpolation rather than extrapolation. The Standard Error of Calibration (SEC) is a measure of how well the PTMRs and PPTMRs agree for this set of calibration samples. SEC includes contributions from spectrum measurement error, PTM measurement error, and model error. Sample (type) specific biases are a part of the model error. Typically, spectroscopic analyzers are very precise, so that spectral measurement error is small relative to the other types of error.

1.4.1.2 During initial analyzer validation, the compositional range of available samples may be small relative to the range of the calibration set. Because of the high precision of the spectroscopic measurement, the average difference between the PTMRs and PPTMRs may reflect a sample (type) specific bias which is statistically observable, but which are less than the 95 % uncertainty of PPTMR, *U(PPTMR)*. Therefore, the bias and precision of the PTMR/PPTMR differences are not used as the basis for local validation.

1.4.1.3 Based on SEC, and the leverage statistic, a 95 % uncertainty for each PPTMR, *U(PPTMR)* is calculated. During validation, for each non-outlier sample, a determination is made as to whether the absolute difference between PPTMR and PTMR, |δ|, is less than or equal to *U(PPTMR)*. Counts are maintained as to the total number of non-outlier validation samples, and the number of samples for which |δ| is less than or equal to *U(PPTMR)*. Given the total number of non-outlier validation samples, an inverse binomial distribution is used to calculate the minimum number of results for which |δ| must be less than *U(PPTMR)*. If the number of results for which |Δ| is less than *U(PPTMR)* is greater than or equal to this minimum, then the results are consistent with the expectations of the multivariate model, and the analyzer passes local validation. The calculations involved are described in detail in Section 11 and Annex A4.

1.4.1.4 The user must establish that results that are consistent with the expectations based on the multivariate model will be adequate for the intended application. A 95 % probability is recommended for the inverse binomial distribution calculation. The user may adjust this based on the criticality of the application. See Annex A4 for details.

1.4.2 *General Validation:*

1.4.2.1 When the validation samples are of sufficient number, and their compositional and property ranges are comparable to that of the model calibration set, then a General Validation can be done.

1.4.2.2 General Validation is conducted by doing a D6708 based assessment 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. The system (or subsystem) is considered to be validated if the D6708 meets the following condition:

(1) No bias correction can statistically improve the agreement between the PPTMRs versus the PTMRs, and

(2) R_{xy} computed as per D6708 meets user-specified requirements.

1.4.2.3 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.

1.4.2.4 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.

1.5 This practice does not cover procedures for establishing the calibration model (correlation) used by the analyzer. Calibration procedures are covered in ~~Practices~~Practice ~~E1655~~D8321 and references therein.

1.6 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 that the degree of agreement between PPTMRs and PTMRs meet user requirements.

1.7 This practice specifies 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.

1.8 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.

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

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

1.11 *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, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.12 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D86 Test Method for Distillation of Petroleum Products and Liquid Fuels at Atmospheric Pressure
D1265 Practice for Sampling Liquefied Petroleum (LP) Gases, Manual Method
D1319 Test Method for Hydrocarbon Types in Liquid Petroleum Products by Fluorescent Indicator Adsorption
D2699 Test Method for Research Octane Number of Spark-Ignition Engine Fuel
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
D5599 Test Method for Determination of Oxygenates in Gasoline by Gas Chromatography and Oxygen Selective Flame Ionization Detection
D5769 Test Method for Determination of Benzene, Toluene, and Total Aromatics in Finished Gasolines by Gas Chromatography/Mass Spectrometry
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
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
D7915 Practice for Application of Generalized Extreme Studentized Deviate (GESD) Technique to Simultaneously Identify Multiple Outliers in a Data Set
D8321 Practice for Development and Validation of Multivariate Analyses for Use in Predicting Properties of Petroleum Products, Liquid Fuels, and Lubricants based on Spectroscopic Measurements
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:*

[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.

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

3.1.2 For definitions of terms and symbols relating to multivariate calibration, refer to Practices ~~E1655~~D8321.

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 that 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. **D6299**

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 model, n*—see *multivariate model*.

3.2.5 *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.

3.2.6 *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.

3.2.7 *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.

3.2.8 *analyzer validation status, n*—an indicator as to the validity of analyzer results produced by application of the multivariate model to spectra of the process sample.

3.2.8.1 *Discussion*—

Prior to the analyzer passing probationary local validation, the analyzer validation status and the validity of the results is unknown; once the analyzer passes probationary local validation, the analyzer validation status is pass, and results are validated as long as the spectrum is not an outlier or nearest neighbor inlier; if the analyzer fails probationary or continual validation, the analyzer status is fail, and analyzer results are not validated.

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

3.2.10 *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.

3.2.11 *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.

3.2.12 *expectations based on the multivariate model, n*—the absolute difference between the PPTMR and the PTMR for a set of validation samples will not exceed the uncertainty on the PPTMR more than one time in 20 in the long term.

3.2.13 *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.

3.2.14 *general validation, n*—a comprehensive evaluation of the agreement between the PPTMR and the PTMR done on a set of samples that adequately span the multivariate model composition space using the statistical methodology of Practice D6708 to demonstrate all required criteria in D6708 are met, and R_{xy} meets user requirements.

3.2.15 *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.

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

3.2.17 *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.

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

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

3.2.20 *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.

3.2.21 *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.

3.2.22 *local validation, n*—an evaluation of the agreement between the PPTMR and PTMR done on a set of samples that do not necessarily span the compositional space of the multivariate model so as to demonstrate that the agreement is consistent with expectations based on the multivariate model.

3.2.23 *model degrees of freedom (dof), n*—the dimension of the multivariate space defined by the number of calibration sample spectra, the number of model variables, and the number of variables used in defining the property level dependence of the Standard Error of Calibration (*SEC*).

3.2.23.1 *Discussion*—

For a multivariate model that is not mean-centered, $dof = n\text{-}k\text{-}c$, where n is the number of calibration samples, k is the number of model variables, and c is 0, 1 or 2 depending on whether *SEC* is level independent, has a linear dependence on property level, or has a power dependence. For a mean-centered model, $dof = n\text{-}k\text{-}c\text{-}1$.

3.2.24 *model variables, n*—the independent variables derived from the calibration spectra which are regressed against the calibration sample properties to produce the multivariate model.

3.2.24.1 *Discussion*—

For MLR, the model variables would be the absorbance at the selected wavelengths or frequencies; for PCR or PLS, the model variables are the Principal Components or latent variables.

3.2.25 *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.

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

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

3.2.28 *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.

3.2.29 *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.

3.2.30 *optical reference filter, n*—an optical filter or other device which can be inserted into the optical path in the spectrophotometer or probe producing a 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.

3.2.31 *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.

3.2.32 *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.

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

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

3.2.35 *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.

3.2.36 *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.

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

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

3.2.37.2 *Discussion*—
Local validation uses the absolute value of the prediction deviations, |δ|.

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

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

3.2.40 *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.

3.2.41 *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.

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

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

3.2.44 *spectrometer cell, Raman, n*—an apparatus which allows a liquid hydrocarbon to flow past an optical surface or surfaces that allow(s) transmission of the laser light into the sample and the generated Raman scattering light out of the sample.

3.2.45 *spectrophotometer cell infrared, 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.

3.2.46 *Standard Error of Calibration (SEC), n*—a measure of the agreement between PPTMR and PTMR for the samples used in developing a multivariate model, $SEC = \sqrt{\sum_{i=1}^{n}(PPTMR_i - PTMR_i)^2/dof}$, where *dof* is the model degrees of freedom and *n* is the number of calibration samples.

3.2.47 *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.

3.2.48 *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.

3.2.49 *uncertainty of Predicted Primary Test Method Result (U(PPTMR)), n*—the interval about PPTMR in which PTMR is expected to occur 95 % of the time in the long run.

3.2.49.1 *Discussion—*

PTMR is expected to fall in the range between PPTMR – *U(PPTMR)* and PPTMR + *U(PPTMR)* 95 % of the time over the long run.

3.2.49.2 *Discussion—*

$U(PPTMR) = t(p,dof) \cdot SEC(m) \cdot \sqrt{1+h}$ where *t* is the Student's T value for probability level *p* and model degrees of freedom *dof*,*SEC(m)* is the model Standard Error of Calibration at property level *m*, where *m* is the average of PPTMR and PTMR, and *h* is the leverage calculated for the spectrum being analyzed to produce PPTMR. For more details, see Annex A4.

3.2.50 *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.

3.2.51 *validation reference material (VRM), n*—a line sample retain, composite sample, or tank sample which is representative of current production, has a measured PTMR, and is used in place of a line sample during the validation process.

3.2.52 *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.

3.2.53 *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.

3.3 *Symbols:*

3.3.1 *c*—the number of coefficients needed to describe the property level dependence of *SEC* minus 1.

3.3.2 δ—difference between PPTMR and PTMR; |δ| indicates the absolute value of the difference; δ_i indicates the difference for the i^{th} sample.

3.3.3 $\hat{\delta}_i$—the estimate of δ_i produced by the SEC level dependence model.

3.3.4 ∇_j—the Root Mean Square Difference between PPTMR and PTMR for the samples in the j^{th} subset.

3.3.5 $\hat{\nabla}_j$—the estimate of ∇_j produced by a weighted regression model.

3.3.6 $\bar{\nabla}$—the weighted average of the ∇ over the s subsets.

3.3.7 *dof*—model degrees of freedom

3.3.8 *F*—an F-ratio; subscript indicates whether the ratio is between *WSSE* values for (*L-C*) linear and constant, (*P-C*) power and constant, or (*P-L*) power and linear *SEC* level dependence models.

3.3.9 *h*—leverage statistic

3.3.10 *k*—number of variables in multivariate model

3.3.11 m_i—the average of the PPTMR and PTMR for the i^{th} sample.

3.3.12 $\bar{m}_i$—the average of the m_i values for the samples in the j^{th} subset.

3.3.13 *n*—number of model calibration samples

3.3.14 *PTM*—Primary Test Method

3.3.15 *PTMR*—Primary Test Method Result

3.3.16 *PPTMR*—Predicted Primary Test Method Result

3.3.17 *RMSEC*—Root Mean Square Error of Calibration

3.3.18 R_{XY}—between method reproducibility

3.3.19 *s*—the number of subsets used in the weighted regression.

3.3.20 *SEC*—Standard Error of Calibration; *SEC(m)* indicates *SEC* at property level *m*; a subscript *C*,*L*, or *P* after the *SEC* indicates whether *SEC* has a constant, linear, or power dependence on property level.

3.3.21 *SSE*—the (unweighted) Sum of Squared Errors; a subscript *C*, *L*, or *P* indicates constant, linear, or power regression.

3.3.22 *t(p, dof)*—Student's T-value at probability *p* for *dof* degrees of freedom

3.3.23 *U(PPTMR)*—Uncertainty on PPTMR

3.3.24 *WSSE*—the Weighted Sum of Squared Errors; a subscript *C*,*L*, or *P* indicates constant, linear, or power regression.

3.3.25 *WSSM*—the Weighted Model Sum of Squares; a subscript *C*,*L*, or *P* indicates constant, linear, or power regression.

3.3.26 *WSST*—the Total Sum of Squares; a subscript *C*,*L*, or *P* indicates constant, linear, or power regression.

3.3.27 z_i—the transform of m_i; $z_i = \max(m_i) - m_i$.

3.3.28 $\bar{z}_j$—the transform of $\bar{m}_j$; $\bar{z}_j = \max(\bar{m}_j) - m$.

4. Summary of Practice

4.1 This section describes, in summary form, the steps involved in the validation of an infrared analyzer for prediction of a single physical, chemical, or quality property over the long term. If multiple properties are predicted from a single spectral measurement, the validation of each property prediction is considered a separate application of this practice. These separate applications of the practice may share certain features, but the analyzer performance for the prediction of each property is evaluated separately.

4.2 Before this practice may be undertaken, certain preconditions shall be satisfied. The preconditions are described in Section 7.

4.3 This practice consists of five major procedures.

4.3.1 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.

4.3.2 When the analyzer is initially installed, or after major maintenance is concluded, an analyzer qualification is performed. 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.

NOTE 2—Major maintenance is any change to the analyzer system hardware or software that is shown by historical data or simulations to cause a statistically observable change in the analyzer performance relative to before the maintenance. What constitutes major maintenance is specific to the analyzer hardware and software employed. Users should consult the analyzer manufacturer as to what types of maintenance should be considered major. Any maintenance which requires calibration transfer to be performed should be considered major maintenance. Any maintenance for which performance changes are routinely compensated for in analyzer software or in the multivariate model are not considered major maintenance.

4.3.3 After the analyzer qualification is successfully completed, a probationary local 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 expectations based on the multivariate model. As the spectra of these initial validation samples are collected, they are analyzed with the multivariate model to produce the PPTMRs. The absolute differences between PPTMR and PTMRs, $|\Delta|$, are compared to the 95 % uncertainties of the PPTMRs, *U(PPTMR)*. If $|\delta|$ does not exceed *U(PPTMR)* for more than two samples, probationary local validation continues until 15 validation samples have been processed. If $|\delta|$ is less than *U(PPTMR)* for at least 13 of the 15 initial validation samples, then the predictions are consistent with the expectations based on the multivariate model, and the system or subsystem performance is considered to be probationary validated for materials and property ranges representative of those used in the validation. providing that the spectra used in generating the results are neither outliers or nearest neighbor inliers. If $|\delta|$ is greater than *U(PPTMR)* for more than 2 of the initial validation samples, the analyzer system fails probationary validation. An investigation of the cause of the failure should be conducted, and corrective action taken. The validation process then restarts with initial performance testing.

NOTE 3—Probationary validation is done using the local validation procedure since 15 samples is too few to conduct a general validation. 15 samples would seldom if ever span the property and composition range of a multivariate model based on more than 3 variables.

4.3.4 After probationary local validation is achieved, continued statistical quality control chart monitoring and analyses of $|\delta|$ are carried out with new production samples to ensure ongoing prediction performance of the PPTMR meets the levels established from the probationary validation. The $|\delta|$ are compared to *U(PPTMR)*, and a count is maintained of the total number of non-outlier validation samples, and the number for which $|\delta|$ is less than or equal to *U(PPTMR)*. An inverse binomial distribution is used to calculate the minimum number of samples for which $|\delta|$ must be less than *U(PPTMR)*. As long as this minimum is met, the analyzer system passes continual local validation. If the minimum is not met, the analyzer fails local validation. An investigation of the cause of the failure should be conducted, and corrective action taken. The validation process then restarts with analyzer qualification.

4.3.5 Once the total number of (PPTMR/PTMR/δ) data sets for samples from probationary and continual validation reaches 4 times the number of variables in the multivariate model ($4k$), a general validation can be conducted using the statistical methodology of Practice D6708 providing that the available validation samples adequately span the full composition and property range of the multivariate model. 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 4*k* samples covering a wider operating envelope.

4.4 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. If the analyzer results agree with the primary test method to within limits based on the multivariate model for the user-prespecified statistical confidence level, these results can be considered 'validated' to the user pre-specified confidence limit for a specific application, and hence can be considered useful for that specific application.

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 t he 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 and Raman Measurements

6.1 *Infrared or Near-Infrared Spectrophotometer and Raman Spectrometer*

6.1.1 The analyzer covered by this practice includes those 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 Analyzers covered by this practice also include those based on a Raman spectrometer. A typical process Raman analyzer will include a laser excitation source, filters to block Raleigh scattered light, and a spectrograph consisting of a grating for dispersing the Raman scattered light and a multielement detector. The Raman analyzer will also contain a Raman spectrometer cell, and possibly optical fibers for transmitting the laser light to the cell and returning the Raman scattered light to the spectrograph.

6.1.3 *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 IR or Raman 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 For online systems, 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. If the analyzer system collects multiple spectra over the time it takes to collect a sample, then the average of the PPTMR values for these multiple spectra may be used as the line sample PPTMR.

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 (retained line sample, composite sample, or tank sample) 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~~Practice ~~E1655~~D8321. If a calibration transfer method is used to transfer the model from one analyzer to another, verify the transferred model as described in ~~Practices~~Practice ~~E1655~~D8321.

NOTE 4—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~~Practice ~~E1655~~D8321 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. 1-6.

9.2 *Analyzer Qualification (*Fig. 1*):*

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 qualify the analyzer by demonstrating that current performance of the instrument is consistent with performance which is known to produce valid analyses. Performance is checked by taking a spectrum of a check or test sample, and analyzing the spectrum using one or more of the Level 0, Level A, or Level B performance tests described in Annex A2 and Practice E1866. Performance tests are discussed in more detail in Section 10.

9.2.1.1 If test action limits for the performance test(s) have been established by the manufacturer or user, then the analyzer is qualified if the performance is within limits. If the initial performance is not within limits, then one retest is allowed. If the performance for the retest is within limits, the analyzer is qualified. Otherwise, corrective action must be taken to fix the cause of the performance issue.

9.2.1.2 If historical limits for the performance test have not been established by the manufacturer or user, then a minimum of 20 such tests must be conducted to generate an initial database from which limits can be established. In order that these initial performance tests be representative of real operation, the analyzer should be returned to normal use between performance tests, and the 20 results should be collected over a period of at least 24 hours to capture any diurnal performance variations.

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.

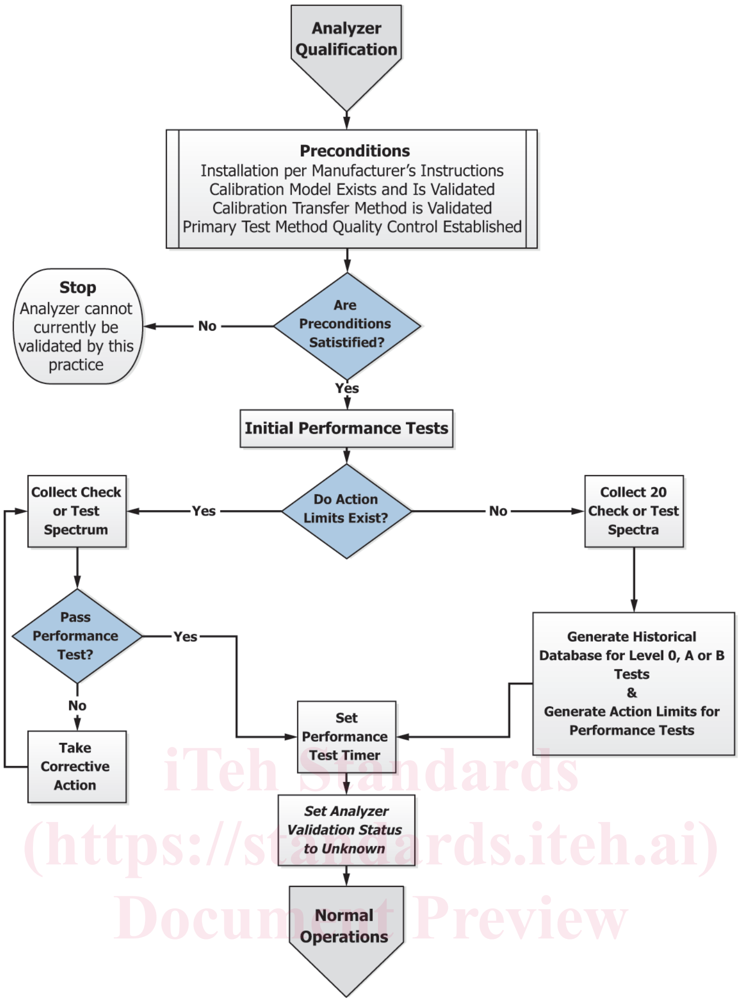

FIG. 1 Flowchart of Process Analyzer Validation Practice
Analyzer Qualification at Initial Startup. at Restart after Maintenance, and after Model Update

9.2.3 Performance test 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 analyzer validation status is set to unknown, and the procedure continues with the normal operations and probationary local 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.

9.2.4 Installation of a new multivariate model does not change the analyzer qualification, so no additional performance testing is required. However, the installation does restart the validation process with probationary local validation.

9.3 *Normal Operation* (Fig. 2)*:*

9.3.1 When validation tests are not being performed, normal operations for analysis of process samples may be conducted. Timers are used to determine when new backgrounds are collected, when performance tests are conducted, and when line samples are

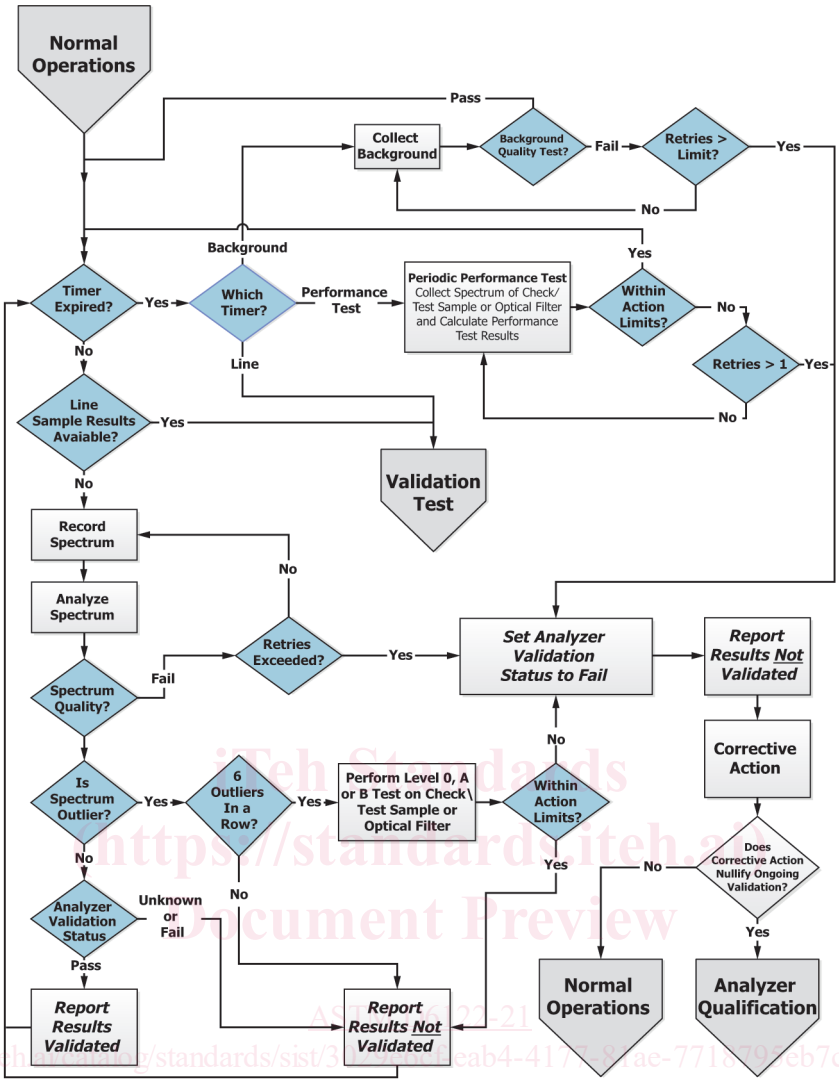

FIG. 2 Flowchart of Process Analyzer Validation Practice Normal Operations

taken for validation testing. For analyzers in continuous operation, the timers are set for regular intervals. For analyzers used in batch operation, the backgrounds and performance tests may be conducted at the beginning or end of a batch, whereas line samples for validation testing may be collected at intervals throughout the batch. The user must set appropriate intervals for all timers. When these tests are not scheduled, the normal application of the analyzer for online, at-line or laboratory analysis proceeds as follows:

9.3.2 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. If, after a process sample spectrum has been collected and processed, the performance test timer has expired, conduct a performance test.

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

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

9.3.2.3 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).