CALIBRATION CURVES: PROGRAM USE/NEEDS

FINAL
Forum on Environmental Measurements
October 2010

The following is a summary of program uses and needs for calibration curves as integral background information to establish greater consistency across the Agency.

OAR – Stationary Source/Ambient Air – Almost all of our test procedures in Parts 60, 61, and 63 use calibration curves/lines to relate instrument response to analyte concentration or mass. Many of the methods contain quality assurance requirements that the analyst must meet for the acceptability of the calibration curve. Our QA requirements have generally not relied on correlation coefficients, but have followed an approach closer to that of Dr. Burrows where the response of each calibration standard must be within a specified range of the value predicted by the calibration line. Here is a typical example taken from Method 7 for the measurement of NO_x emissions from Stationary Sources. In this case the analytical procedure is spectrophotometry in the visible range.

10.1.3 Spectrophotometer Calibration Quality Control. Multiply the absorbance value obtained for each standard by the K_c factor (reciprocal of the least squares slope) to determine the distance each calibration point lies from the theoretical calibration line. The difference between the calculated concentration values and the actual concentrations (*i.e.*, 100, 200, 300, and 400 μg NO_2) should be less than 7 percent for all standards.

We need a quality assurance requirement for calibration curves/lines that would insure that testers are using calibration data to generate calibration curves that meet minimum standards for "goodness of fit."

Air Stationary Source Test Methods That Require Calibration Curves include: Method 2B, Method 2E, Method 2F, Method 2H, Method 3A, Method 3C, Method 5E, Method 6C, Method 7, Method 7A, Method 7B, Method 7C, Method 7D, Method 7E, Method 10A, Method 10B, Method 12, Method 13A, Method 13B, Method 15, Method 16, Method 16B, Method 18, Method 20, Method 23, Method 25, Method 25A, Method 25B, Method 26, Method 26A, Method 29, Method 30A, Method 30B, Method 101, Method 101A, Method 102, Method 103, Method 104, Method 105, Method 106, Method 107, Method 107A, Method 108, Method 108A, Method 108B, Method 114, Method 204B, Method 204C, Method 204D, Method 204F, Method 205, Method 306, Method 306A, Method 308, and Method 316 .

OECA – National Enforcement Investigation Center – The diverse nature of NEIC's forensic investigations limits the development of a standardized method for calibration. In some cases, the use of a standard linear regression model for the inorganic and organic analyses is appropriate, while other cases require a more

diverse approach. Any analysis at or near the method detection limit will present a much higher uncertainty; therefore calibration will need to be investigated as a part of the overall quality control. For NEIC, understanding the uncertainty associated with analysis at a regulatory or permit level is essential. The use of all quality control measures, such as blanks, spikes, surrogates, reference materials, and alternate analysis, makes the calibration only a single component in the evaluation of the overall uncertainty. Calibration design is optimized for each analysis to minimize the contribution of this source of systematic and random error to the total uncertainty in the final analytical results. A prescriptive and rigid approach to the calibration design will have a negative impact on NEIC's data quality.

OPPTS – Office of Pesticide Program – Information and data are received from many different sources using a wide array of different methods. A variety of approaches are used to satisfy different needs and purposes. Best for these programs to retain the flexibility of not dictating any one approach, but open to possibilities of new approaches.

ORD – Office of Research and Development Program – ORD analytical methods typically utilize linear calibrations with a quality control limit established using the coefficient of determination, r^2. ORD would benefit from discussion and guidance on the relationships between detection limit, quantitation limit, and the low standard of the calibration curve; use of weighted vs. non-weighted and linear vs. quadratic calibration curves; requirements for independent and continuing calibration checks; use of isotope dilution (especially when labeled standards are not available for all analytes of interest); the value of using matrix-matched calibration curves vs. curves established with standards in solvent; and appropriate quality control parameters for assessing curves.

OSWER – Solid Waste Program – ORCR promotes two primary approaches for delineating the relationship between the amount or concentration of target analyte introduced into an analytical instrument and the corresponding instrument response, the selection of which depends on the chemical nature of the target analyte and the availability of standards:

1. **External standard calibration** involves the comparison of instrument responses from the sample to the responses from target analytes of known concentration in the calibration standards.

2. **Internal standard calibration** involves, in addition to comparing the instrument responses of samples to those of calibration standards, the normalization of instrument responses from the target analytes in the sample based on the responses of specific standards added to the sample prior to instrument injection. Internal standard calibration provides a practical means of correcting for instrument variation and drift. The use of mass spectrometric detectors makes internal standard calibration practical because the masses of the internal standards can be resolved from those of the target analytes. Internal

standards generally consist of brominated, fluorinated, stable, isotopically-labeled analogs of specific target analytes, or otherwise closely-related compounds, whose presence in environmental samples is highly unlikely.

For either calibration approach, three different calibration techniques may be invoked:

A. **Linear calibration through origin** – In this method, the mean calibration factor (CF) (ratio of instrument detector response to target analyte amount or concentration) of an external calibration or mean response factor (RF) (ratio of detector response of analyte to its amount or concentration times the ratio of internal standard concentration to its detector response) of an internal calibration is determined through the analysis of one or more calibration standards and used to quantify the amount or concentration of target analyte in a sample based on the sample detector response. The method is used in cases where the relative standard deviation of the CFs or RFs is less than or equal to 20%, the detector response is directly proportional to the target analyte amount or concentration and the calibration passes through the origin. External linear calibration through the origin is typically used for ICP metals, in which case the calibration curve consists of a blank and a single standard prepared at an appropriate concentration so as to effectively outline the desired quantitation range. External and internal linear calibrations are also used for certain GC and HPLC methods.

B. **Linear least squares regression** – A mathematical model invoked for calibration data that describes the relationship between expected and observed values via minimization of the sum of the squared residuals (deviations between observed and expected values) - The final outcome of the least squares regression is a linear calibration model of the form: $y = m_1x + m_2x^2 + m_3x^3 + \ldots + m_nx^n + b$ (where y = detector response and x = target analyte amount or concentration). Least squares regression calibrations are typically derived from a minimum of three standards of varying concentration and are applicable to data sets in which the measurement uncertainty is relatively constant across the calibration range. Most SW-846 methods rely on first-order least squares regression models ($y = mx + b$) for calibration. However given the advent of new detection techniques, and the fact that many techniques cannot be optimized for all analytes to which they are applied, or over a sufficiently wide working range, second-order ($y = m_2x^2 + m_1x + b$) or third-order ($y = m_3x^3 + m_2x^2 + m_1x + b$) linear regression models are often invoked for calibration. In any of these cases, SW-846 methods allow forcing a linear least squares regression through the origin, provided that the resulting calibration meets the acceptance criteria and can be verified by acceptable quality control results. External least squares regressions are typically used for ICP/MS metals calibrations. Internal least squares regressions are generally used for calibration in GC/MS applications.

C. **Weighted least squares regression** – A form of linear least squares regression invoked for modeling a calibration curve in which the measurement uncertainty is determined to be not constant across the calibration range (as established through the analysis of three or more replicates at each calibration point). Unlike non-weighted least squares, each term in a weighted least squares regression includes an additional weighting factor (the reciprocal of the estimated measurement variance, $1/\sigma^2$; where σ = the standard deviation of detector responses for a given set of calibration point replicates) that serves to minimize the influence of calibration points exhibiting greater measurement uncertainty, while maximizing the contribution of those having smaller uncertainty towards the fit of the regression line. Internal weighted least squares regressions are often used for calibration in certain GC/MS applications.

The selection of the specific calibration technique is made in one of two ways. The first is to begin with the simplest approach, linear calibration through the origin, and then progress through the other options until the calibration acceptance criteria are met. The second way to select the calibration technique is to use *a priori* knowledge of the detector response to the target analyte(s). Such knowledge may come from previous experience, knowledge of the physics of the detector, or specific manufacturer's recommendations.

In Method 8276 (posted on the SW-846 Methods Website in March 2010), ORCR included two ways to determine if the selected calibration model is acceptable. Specifically, 1) refitting (% difference) the calibration data back to the model; and 2) relative standard error (RSE) – which compares actual response of a calibration level with predicted response.

Program Needs: The SW-846 methods are being used by various programs, including RCRA, Superfund, TSCA, and Homeland Security. ORCR deals with complex wastes and materials that are being managed or used in many different ways (e.g., land filling, land application, incineration, recycling. Given the difficulty often involved in analyzing matrices of such complexity, ORCR promotes flexibility in the application and use of test methods and considers all SW-846 methods as guidance, except in cases where parameters are method-defined or otherwise required by regulation.

ORCR strongly supports the performance-based approach and follows this approach in the RCRA testing program to the extent feasible. In this context, the selection of calibration method and technique is made based on the characteristics of the method, the data set and the specific program needs. Data generators should sufficiently demonstrate that the calibration meets the necessary acceptance criteria for the desired target analyte(s). Additionally, the calibration curve should bracket the concentration range of the samples for which it is being applied.

OW – Office of Ground Water/Drinking Water Program – Our program allows multiple calibration models based on the method. The most common are relative response factor, linear, quadratic and weighted quadratic. Our position is that the calibration model should be method specific and is determined during method development. No matter what we do with calibration we need to eliminate the use of R^2 criteria as a measure for calibration curve quality. One other issue that we should consider when higher order models are allowed is that there should be criteria for how many calibration points are required (e.g. 3 points per order of magnitude).

OW – Office of Science and Technology – Language taken from, "Solutions to Analytical Chemistry Problems with Clean Water Act Methods" located at http://www.epa.gov/waterscience/methods/

Number of Calibration Points:

The 600-series methods specify a minimum of three calibration points. The lowest of these points is required to be near the MDL. The highest is required to be near the upper linear range of the analytical system, and the third point is approximately midway between the two. Some methods, such as Methods 1624 and 1625, require calibration at five specific concentrations for nearly all analytes, and three or four specific concentrations for the remaining analytes for which the methods are not as sensitive.

The lowest calibration point should be below the action level and the high standard should still be within the calibration range of the instrument.

The flexibility in selecting the levels of the calibration points in the 600-series methods has led to a wide variety of calibration ranges as each laboratory may determine its own calibration range. Some laboratories may establish a relatively narrow calibration range, for instance a five-fold concentration range such as 10 to 50 µg/L (ppb), because it makes it simpler to meet the linearity specifications of the 600-series methods. Other laboratories may choose wider calibration ranges, e.g., 10 to 200 µg/L (ppb), in order to minimize the number of samples that should be diluted and reanalyzed because the concentration of one or more analytes exceeds the calibration range.

The data reviewer will need to make certain that all measurements are within the calibration range of the instrument. Samples with analyte concentrations above the calibration range should have been diluted and reanalyzed. The diluted sample results need only apply to those analytes that were out of the calibration range in the initial analysis. In other words, it is acceptable to use results for different analytes from different levels of dilution within the same sample. Some flexibility may be exercised in acceptance of data that are only slightly above (<10%) the calibration range. Such data are generally acceptable as calculated.

If data from an analysis of the diluted sample are not provided, limited use should be made of the data that are above the calibration range (>10%). The response of the analytical instrument to concentrations of analytes will eventually level off at concentrations above the calibration range.

While it is not possible to specify at what concentration this will occur from the calibration data provided, it is generally safe to assume that the reported concentration above the calibrated range is a lower limit of the actual concentration. Therefore, if concentration above the calibration range is also above a regulatory limit, it is highly likely that the actual concentration would also be above that limit.

Linearity of Calibration:

The relationship between the response of an analytical instrument to the concentration or amount of an analyte introduced into the instrument is referred to as the "calibration curve." An analytical instrument can be said to be calibrated in any instance in which an instrumental response can be related to a single concentration of an analyte. The response factor (GC/MS methods) or calibration factor (GC, HPLC methods) is the ratio of the response of the instrument to the concentration (or amount) of analyte introduced into the instrument. The response factor and calibration factor concepts are used in many methods for organic contaminants, while methods for metals and some other analytes may employ different concepts such as linear regressions.

While the shape of calibration curves can be modeled by quadratic equations or higher order mathematical functions, most analytical methods focus on a calibration range where the response is essentially a linear function of the concentration of the analyte. An advantage of linear calibration is that the response factor or calibration factor represents the slope of the calibration line and is relatively constant, simplifying the calculations and data interpretation. Whichever approach is used, all the 600- and 1600-Series methods specify some criterion for determining linearity of calibration. When this criterion is met, the calibration is sufficiently linear to permit the laboratory to use an average response factor or calibration factor, and it is assumed that the calibration is a straight line that passes through the zero/zero calibration point. Linearity is determined by calculating the relative standard deviation (RSD) of the response factor or calibration factor for each analyte and comparing this RSD to the limit specified in the method. If the RSD does not exceed the specification, linearity is assumed.

In the 600- and 1600-Series methods, the linearity specification varies from method to method, depending on the quantitation technique. The typical limits on the RSD are as follows:
- 15% for the gas chromatography (GC) and high-performance liquid chromatography (HPLC) methods;

- 20% for analytes determined by the internal standard technique in the gas chromatography/mass spectrometry (GC/MS) methods (624, 625, 1624, and 1625);
- 20% for analytes determined by isotope dilution in Methods 1613, 1624, and 1625; and
- 15% for mercury determined by atomic fluorescence in Method 1631.

Metals methods that employ a linear regression specify a criterion for the correlation coefficient, r, such as 0.995.

If the calibration is not linear, as determined by the RSD of the response factor or calibration factor, a calibration curve should be used. This means that a regression line or other mathematical function should be employed to relate the instrument response to the concentration. However, properly maintained and operated lab instrumentation should have no difficulty in meeting linearity specifications for 600- and 1600-Series methods. Linear regression emphasizes the importance of higher concentration standards and that the correlation coefficient is little impacted by poor performance of calibration standards with low concentrations.

For determination of nearly all of the organic analytes using the 600- and 1600-Series methods, calibration curves are linear over a concentration range of 20–100 times the nominal concentration, depending on the detector being employed. Whatever calibration range is used, the laboratory should provide the RSD results by which one can judge linearity, even in instances where the laboratory is using a calibration curve. In instances where the laboratory employs a curve rather than an average response or calibration factor, the data reviewer should review each calibration point to assure that the response increases as the concentration increases. If it does not, the instrument is not operating properly, or the calibration curve is out of the range of that instrument, and data are not considered usable.

Calibration Verification

Calibration verification involves the analysis of a single standard, typically in the middle of the calibration range, at the beginning of each analytical shift. The concentration of each analyte in this standard is determined using the initial calibration data and compared to specifications in the method. If the results are within the specifications, the laboratory is allowed to proceed with analyses without recalibrating and to use the multi-point calibration data to quantify sample results. It is also recommended that a calibration verification at the action level is periodically analyzed.

Specifications for calibration verification are generally given as a range of concentrations, as a recovery range, or as a percentage difference from the test concentration. For the 600-series semivolatile GC and HPLC methods, the difference must be within 15%. For Method 625, the difference must be within 20%. The GC and GC/MS methods for volatiles and the 1600-Series methods

specify a range of concentrations or recoveries for each analyte. These ranges are based on interlaboratory method validation studies.

If calibration cannot be verified, the laboratory may either recalibrate the instrument or prepare a fresh calibration standard and make a second attempt to verify calibration. If calibration cannot be verified with a fresh calibration standard, the instrument should be recalibrated. If calibration is not verified, subsequent data are considered to be invalid until the instrument is recalibrated.

Method Detection Limit or Minimum Level

Although this requirement is not explicitly stated in EPA wastewater methods (e.g., 600 and 1600- Series methods) we recommend use of the method detection limit (MDL) concept to establish detection capabilities. Detailed procedures for determining the MDL are provided at 40 CFR Part 136, Appendix B. Although exact frequencies vary by method, most methods require that, at a minimum, laboratories conduct an MDL study as part of their initial demonstration of capability and whenever a modification is made to the method that might affect the detection limit and amends thereafter. Data reviewers should consult the methods used for specific requirements, or the requirements of their customers, auditors, etc.

The Minimum Level (ML) is used as a quantitation level, and is defined in most of the 1600-Series methods as the lowest level at which the entire analytical system gives a recognizable signal and acceptable calibration point. Therefore, each 1600-Series method specifies that the calibration range for each analyte encompass the method specified ML.

Many of the EPA wastewater methods provide specific requirements regarding reporting results that are below the ML or the method-specified quantitation limit when these data will be used for compliance monitoring. Since these requirements vary slightly, data reviewers should consult the specific method for details.

If the sample results are above the ML, but are below the facility's regulatory compliance level, then the laboratory should report the results to indicate that the pollutant has been detected but is compliant with a facility's permit, assuming all QC criteria are met. If sample results are above the regulatory compliance level, the data reviewer may wish to evaluate the laboratory QC sample results to verify that the reported concentration is not attributable to analytical bias. In addition, the data reviewer should evaluate all blank results to determine if the level of pollutant detected may be attributable to contamination.