

TECHNIQUES FOR IMPROVING THE ACCURACY OF CALIBRATION IN THE ENVIRONMENTAL LABORATORY

Dennis A Edgerley

Quanterra Environmental Services, 880 Riverside Parkway, West Sacramento, California 95605

Abstract

Consistent and reliable procedures for generating calibrations are essential to accurate laboratory results. Unfortunately the interpretation of acceptable practice is often based on misunderstanding or derived from practices commonly utilized in non-environmental methods, and therefore does not provide a reliable means for maintaining data quality. This paper presents a demonstration that some common practices used in the calculation and evaluation of calibration factors, including the use of unweighted regression and the associated correlation coefficient, are inappropriate for environmental analysis due to high relative errors which result at the low end of the curve. Alternate criteria for evaluation of calibration curves are proposed based on the Relative Standard Error (%RSE). Statistical derivations and examples are presented to demonstrate how this approach provides an improved measure for the evaluation of calibration data based on weighted regression. Other related considerations for assessing acceptability of calibration data are also presented.

Introduction

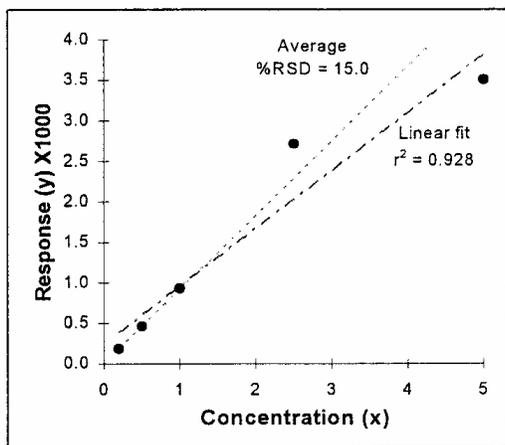
Any analytical measurement must employ reference elements to ensure traceability to relevant basic quantities. The quality of a calibration depends on the uncertainty of the reference, the appropriateness of the reference and how well the calculation procedures match the requirements of the analysis. A majority of the methods employed in the environmental laboratory are based on a relative calibration where standards of known content and concentration are analyzed by a suitable detector. The responses of samples analyzed under the same conditions are then used to calculate concentrations by numerical interpolation to a response curve from the calibration standards.

The only criterion for initial calibration in SW-846 method 8000A¹ reads, "If %RSD is less than 20 an average calibration factor can be used otherwise data should be fitted to a curve." There are many examples of well defined and reproducible calibrations which either due to nonlinearity, or a non-zero intersection of the axis will not meet this 20% criterion for acceptability. Recognizing this, update III to SW-846, Method 8000B² provides additional direction on use mid evaluation of least squares regression, adding criteria for higher order curves. However, several critical issues are not sufficiently considered and overall the current guidance remains incomplete with regard to error weighting and evaluation of acceptability.

Linear Calibration

The most commonly adopted option for handling calibration data, which is linear but does not meet the criteria for averaging, involves the calculation of coefficients for a linear equation of the form:

$$y = Ax + B \quad \text{Equation 1}$$



Concentration is defined here as the independent variable (x) and response as the dependent variable (y) in compliance with method 8000B. A least squares regression is employed and the value of the correlation coefficient (r) or the coefficient of determination (r^2) is evaluated as a measure of acceptability. A value of 1.00 represents a perfect correlation. Generally in practice, a value of r^2 greater than 0.990 is considered satisfactory.

The practical difficulty encountered in this approach is displayed in Figure 1. For the example data set, the line based on an average calculation (shown as a dotted line) easily meets the 20% acceptance criteria for %RSD, yet r^2 does not meet the criteria for acceptability with a value of 0.928.

Figure 1

To understand this inconsistency we must examine more closely the statistical difference between an average calculation and the regression line. The average calibration factor is determined according to equation 2.

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i \quad \text{Equation 2}$$

Where:

\bar{C} = Average Calibration factor
 C_i = Calibration factor for calibration level i (y_i/x_i)
 n = number of calibration levels

This and all following equations may also be adapted to internal standard methods by substituting the relative response calculated as in Equation 3 for the measured response (y_i).

$$y_i^{\text{Relative}} = y_i \times \frac{x_i^{\text{IS}}}{y_i^{\text{IS}}} \quad \text{Equation 3}$$

y_i = response of target analyte
 x_i^{IS} = concentration of internal standard
 y_i^{IS} = response of internal standard

Graphically the average and associated error limits based on ± 1 standard deviation are shown in Figure 2. This type of normalized plot makes visual examination of the data more straight-forward³ as values at the low end of the calibration are shown at the same relative scaling as those toward the high end.

The average represents the value which will minimize the variance in the calibration factors (s^2_c) for all calibration points⁴. For reference the calculation of variance is shown in Equation 4.

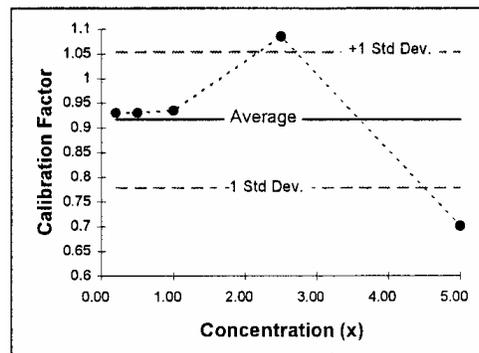


Figure 2

$$s^2_c = \frac{\sum_{i=1}^n (C_i - \bar{C})^2}{n-1} \quad \text{Equation 4}$$

If \hat{y}_i is defined as the expected response for calibration level i from the relationship $\hat{y}_i = x_i \bar{C}$ by substitution the variance can also be expressed as:

$$s^2_c = \frac{\sum_{i=1}^n \left[\frac{y_i - \hat{y}_i}{x_i} \right]^2}{n-1} \quad \text{Equation 5}$$

Using the method of least squares we can determine a mathematical relationship between the dependent and independent variables which minimizes the residual variance. By definition the residual variance represents the variability due to experimental error⁵ and does not include that contribution to variance which is attributable to differences in the independent variable. The residual variance of y on x is defined as:

$$s^2_{yx} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1} \quad \text{Equation 6}$$

A comparison of equations 5 and 6 demonstrates that the average value is the same as a coefficient derived from least squares regression if a weighting of $1/x^2$ is applied. The average calibration factor gives a result which is identical to that produced by the method of least squares using $(1/\text{concentration}^2)$ weighting and intercept forced through zero.

Alternatively a generalized equation for calculating the coefficient which fits the simple relationship $y = Ax$ and minimizes the residual variance without weighting is determined by substituting Ax for y in Equation 6 and setting the derivative with respect to A equal to zero. This gives for the calibration factor:

$$A = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \tag{Equation 7}$$

For the example data set, calibration coefficients and residual variances are compared in table 1. The coefficients determined from these two approaches are quite different and it is obvious that an average calibration coefficient does not minimize the residual variance of y on x .

Table 1

Calculation type	CF	s^2
Average	0.916	1.335
Linear fit	0.784	0.768

Transforming the vertical axis of Figure 2 to a non-normalized form gives the plot shown in Figure 3. A comparable plot for the unweighted regression line using the same set of example data is presented in Figure 4. Both figures represent graphically, with the dashed lines, effective error weighting based on ± 1 standard deviation.

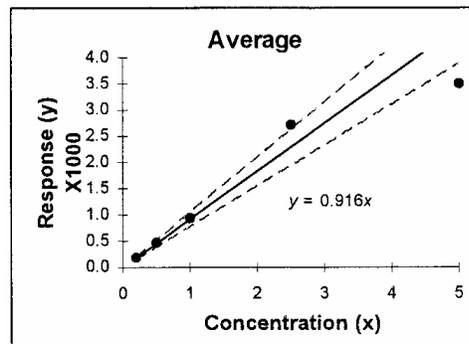


Figure 3

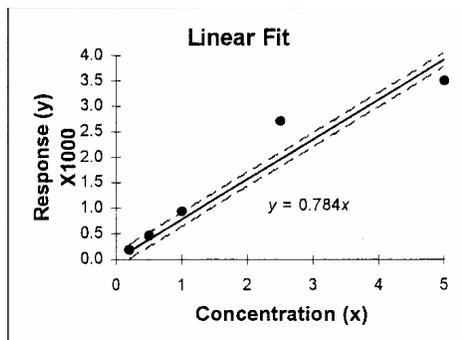


Figure 3 demonstrates that $1/x^2$ weighting gives a relationship that emphasizes precision at the low end of the calibration range. In environmental analyses frequently the objective is to ensure that target analytes do not exceed defined regulatory or action limits, hence reducing quantitation error at low concentrations is especially important.

Figure 4

Deriving the best form for a calibration curve must include consideration of the weighting factors which are appropriate to the requirements of the analysis. As was shown empirically for the single factor calibrations and can be proven for the general case⁶, the value of the coefficients is sensitive to the weighting. Method 8000B requires that at least three replicates at a minimum of 5 concentration levels are used to derive weighting factors which are defined as the inverse of the standard deviation squared for each concentration. In practice this is an especially burdensome requirement, both in the amount of data required and the computational difficulty which results when using individually derived weighting values for each calibration level. Faced with the options in method 8000B, most laboratories will probably choose curve fitting without any consideration of weighting regardless of the potential negative impact on data quality.

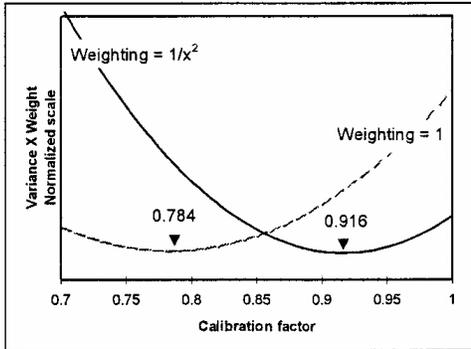
For many methods in the environmental laboratory errors in measurement are proportional to the magnitude of the parameter measured. Likewise it is common to consider percentage errors relative to the concentration as we have demonstrated in the case of the average calibration calculations. Rather than perform multiple replicates each time a calibration is run, the laboratory should make an initial determination of the relationship between concentration and the standard deviation throughout the calibration range. Where a direct proportionality is found, all subsequent calibrations should be based on the $1/\text{concentration}^2$ weighting which is consistent with average calibration curves. Conversely, if standard deviations are of equal absolute magnitude throughout the concentration range, calibrations should be unweighted and average calibration factors should not be used. When the weighting is expressed as a function of the concentration most data systems are capable of performing the calculation.

Percent Relative Standard Error

The magnitude of the residual variance with error weighting applied will provide a measure of the experimental error for the derived curve. Figure 5 shows the relationship between variance and calibration coefficient for a fitting equation of the form: $y = Ax$.

The value of relative standard deviation defined according to Equation 8 has widespread acceptance as a measure of the error associated with an average calibration.

$$\%RSD = 100x \frac{1}{C} x \sqrt{\frac{\sum_{i=1}^n (\bar{C} - C_i)^2}{n-1}} \quad \text{Equation 8}$$



Because standard deviation is equal to the square root of variance, it can be shown that %RSD is also equal to the square root of the weighted residual variance of y on x , calculated as a percentage, using a derivation similar to that for Equation 5.

Likewise the coefficient of determination (r^2) is normally, recognized as an indication of error associated with regression curves.

Figure 5

$$r^2 = \frac{\sum(y - \bar{y})^2 - \sum(y - \hat{y})^2}{\sum(y - \bar{y})^2} \quad \text{Equation 9}$$

Since for any set of data the $\sum(y - \bar{y})^2$ term is not dependent on the form of the calibration curve function or the coefficients, the value of $(1-r^2)$ will be directly proportional to the unweighted residual variance as defined in Equation 6.

It becomes apparent that for an average calibration, the coefficient of determination calculated will not provide an optimum measure of error as r^2 applies only to an unweighted least squares determination. The %RSD on the other hand is only meaningful as a measure of error when applied to an average calibration. A generalized indicator of error, *Percent Relative Standard Error (%RSE)*, which can be applied to any form of weighted regression function is derived similarly to %RSD making adjustment for the degrees of freedom in the relationship by replacing the $n - 1$ factor with $n - p$ where p is an integer equal to the number of coefficients as defined in Equation 10.

%RSE is equivalent to %RSD when calculated for the average. In Figure 6 the least squares line derived using $(1/\text{concentration}^2)$ weighting is compared to the average line and the unweighted least squares line for the same data set used in the previous figures. In this example the elimination of the zero point is responsible for an increase in the %RSE for the weighted line ($p = 2$).

$$\%RSE = 100x \sqrt{\frac{\sum_{i=1}^n \left[\frac{\bar{y}_i - y_i}{\bar{y}_i} \right]^2}{n-p}} \quad \text{Equation 10}$$

Where:

- y_i = Actual response of calibration level
- \bar{y}_i = Calculated response from curve
- p = number of terms in the fitting equation
(average = 1, linear = 2, quadratic 3)
- n = number of calibration points

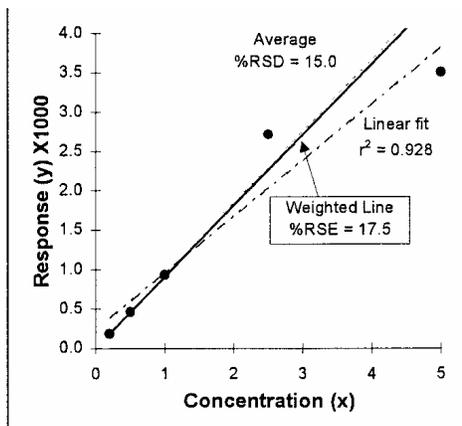


Figure 6

The improved low-end accuracy for this weighted line compared to the unweighted regression line is shown in Table 2. Concentrations measured close to the lowest calibration point would be reported with values approaching 100% lower by using an unweighted linear calibration as compared to the weighted curve.

Table 2

Concentration	% Error (calculated - True)	
	Weighted	Unweighted
0.2	1.7	109
0.5	1.2	29.9
1.0	2.7	2.8
2.5	16.7	25.1
5.0	28.7	8.9

Only recently, in the update to SW-846 method 8000B and also in a draft of guidance for development or modification of water methods⁷ has the EPA recognized the importance of considering weighting in calibration calculations. These documents do not, however provide options for evaluating weighted regression fits. The coefficient of determination (COD) used in method 8000B for evaluating polynomial curves is "weighted" only to adjust for the degrees of freedom in the fitting equation and does not provide a measure of error which is suitable for regression curves derived with error weighting.

Non-Linear Calibration

While it is reasonable to use the simplest mathematical relationship, which gives acceptable accuracy, it should not always be assumed that an average or linear curve is preferred. Some detector systems commonly used in the environmental laboratory are inherently non-linear. As an example the electron capture detector (ECD) commonly utilized for its sensitivity to chlorinated pesticides and herbicides can for most analytes provide calibrations which are linear over a 20X concentration range. It is often difficult to optimize detector conditions for multiple analytes with widely varying electron affinities and even under ideal conditions many ECD detectors show linearity within 20% RSD only over a narrow concentration range. The calibration data plotted in Figure 7 was curved for enhanced accuracy over a 32X range by applying a quadratic fit. While the %RSD for the average was not acceptable at 44.9% the quadrati curve gives a %RSE of only 3.7%.

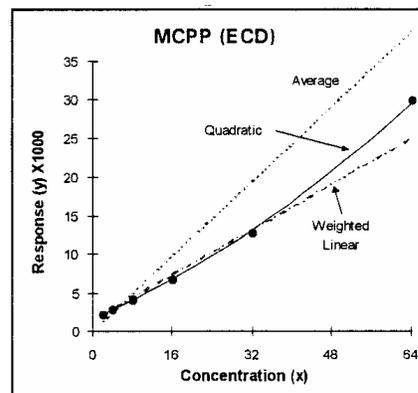


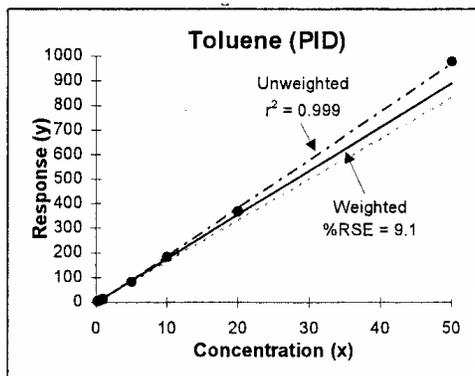
Figure 7

In lieu of higher order curve fitting the approach often defaults to reducing the calibration range, which normally requires re-analysis of some or all calibration points as well as increasing the number of sample dilutions, both of which will add to the cost of the analysis without necessarily providing significant improvement in quality. A simple recalculation of the curve parameters based on a quadratic fit allows the full concentration range of the calibration data to be utilized with improved accuracy. The use of second or higher order curves should not be applied simply to achieve minor improvements in the %RSE that will allow an otherwise unacceptable curve to be used. Justification for higher order calculations should be based primarily on an understanding of the performance characteristics of the detector or the method. The use of higher order curve fitting should also not be substituted for proper instrument maintenance, nor should an effort be made to extend calibrations beyond the detector saturation level.

Evaluation of Calibrations

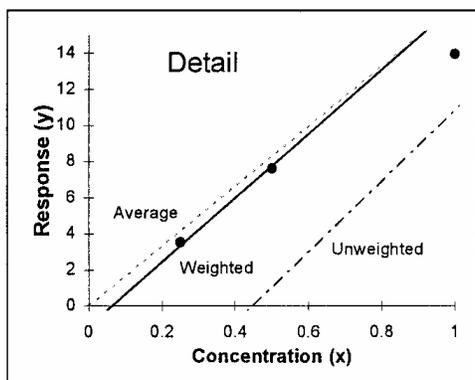
Whenever possible, data evaluation should be performed against rigid criteria that will prevent any tendency for analyst bias to affect reported results and allow automated data validation processes to be implemented. In a production-oriented laboratory setting, visual examination of every calibration curve is not routinely performed. Rather curves are evaluated against calculated criteria only. Modern data systems allow regression parameters to

be calculated with very little analyst effort. Although the %RSE calculation is not commonly provided it can usually be implemented either in a user function or by exporting the results to a spreadsheet.



The behavior of the calibration curve near the reporting limit should also be a primary consideration for environmental methods. According to the proposed update to SW-846 method 8000B, data should be considered unreliable for instrument response less than 3 times the y-intercept from the curve if the value is positive, or less than the concentration calculated from zero response if the y-intercept is negative. The use of weighted curve fitting will greatly reduce the possibility that the intercept exceeds these limits. The unweighted line in Figure 8a appears to provide a very good fit to the data with $r^2 = 0.999$. The expanded view near the low end of the calibration shown in Figure 8b clearly demonstrates the improved fit for the weighted line.

Figure 8a



Recommendations for the minimum number of data points needed to prepare a calibration vary from one to fifteen depending on the method and the order of the fitting equation. Normally five points are considered adequate for linear curves. Some problems may be corrected by elimination of data points from the calibration calculation. This should only be done if they are either the highest or lowest concentration and the number of points remaining meets the minimum requirements of the method. The analyst should also be aware that removing the low point might adversely affect the reporting limit, as quantitation must not be performed outside the concentration range of the calibration.

Figure 8b

When second order curves are evaluated, the acceptability should include an evaluation of the curve inflection points. In Figure 9, the quadratic curve (solid line) provides a significant improvement in %RSE over the weighted least squares line (dashed). However, the y-value of the curve inflection for this line is below the response of the highest data point, thus quantitation near the upper end of the calibration range could give erroneously high results.

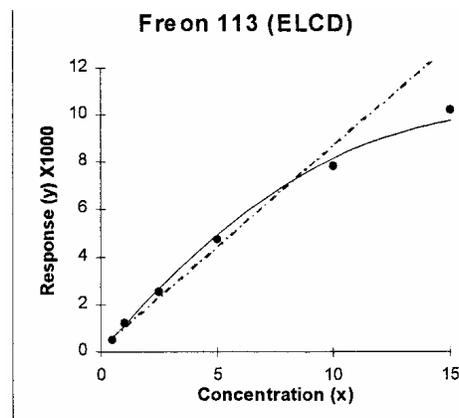


Figure 9

As a rule, all points in the calibration curve should demonstrate a consistent relationship between concentration and response (response increases with increasing concentration). According to SW-846 method 8000B "... the curve must be continuous, continuously differentiable and monotonic over the calibration range."

Summary

The correlation coefficient as a criterion for evaluating regression curves does not apply to weighted regressions that are necessary for accurate low concentration reporting of environmental data. The correlation coefficient is also not consistent with the %RSD criterion used to evaluate average curves, often leading to calibration data that is not acceptable for a linear regression based on the correlation coefficient but does meet %RSD criterion of an average calibration.

The Percent Relative Standard Error (%RSE) provides an improved criteria for evaluation of calibration curves in environmental laboratory methods. The advantages of this approach are as follows:

- All curve types can be calculated with the same ($1/\text{concentration}^2$) weighting applied, This places emphasis on relative error with improved accuracy at concentrations near the reporting limit.
- The %RSE criterion is consistent for evaluation of all curve fitting types. Interpretation of acceptability for calibration curves is simplified with the same criterion applied to all curve types and no conflicting criteria.
- The most appropriate curve fitting model can be applied to each set of calibration data with evaluation criteria between different curve types directly comparable.

The evaluation of all calibration data should include, as a minimum, the following checks:

- %RSE < maximum limit
- concentration level of low standard \leq reporting limit
- low point intersection values < reporting limit
- Number of calibration levels meets method requirements
- All points must be monotonically increasing
- Also for second or higher order curves inflection points should not be within the calibration range

With the widespread availability of powerful computer hardware and software in the laboratory, it is unnecessary to sacrifice data quality for the sake of simplification. Analysts should be familiar with the productivity and quality benefits of least squares curve-fitting algorithms. Clients and regulators must understand the importance of weighted curve fitting and the need for complete and consistent evaluation criteria, which will provide high quality results for the lowest cost.

References

1. Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, SW846, 3rd Edition, Final Update 1, July 1992, Section 8000A
2. Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, SW846, 3rd Edition, Revision 2, December 1996, Section 8000B
3. R. Cassidy and M. Janoski, LC-GC 10(9), 692 (1992)
4. Allen L. Edwards, An Introduction to linear Regression and Correlation, W.H. Freeman and Company, 1976
5. Maurice G. Kendall and William R. Buckland, A Dictionary of Statistical Terms, Hafner Publishing Co., 1972
6. J.C. Miller and J. N. Miller, Statistics for Analytical Chemistry, John Wiley, 1984
7. Guide to Method Flexibility and Approval of EPA Water Methods, U.S. EPA, Office of Water, Draft: December 1996.
8. Grant T. Wernimont, Use of Statistics to Develop and Evaluate Analytical Methods, Association of Official Analytical Chemists, 1985
9. G. Bulmer, Principles of Statistics, Dover Publications, 1979