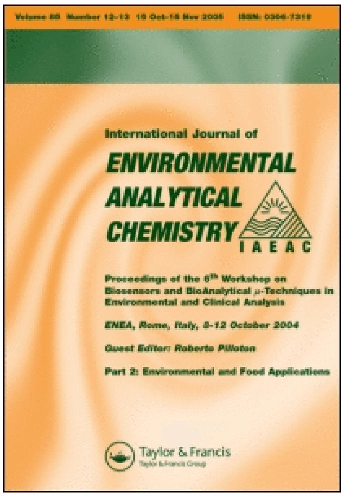

International Journal of Environmental Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713640455

Evaluating 'goodness-of-fit' for linear instrument calibrations through the origin

Thomas Georgian[a]
[a] US Army Corps of Engineers, Environmental and Munitions Center of Expertise, Omaha, Nebraska, United States

To cite this Article Georgian, Thomas(2009) 'Evaluating 'goodness-of-fit' for linear instrument calibrations through the origin', International Journal of Environmental Analytical Chemistry, 89: 5, 383 — 388
To link to this Article: DOI: 10.1080/03067310802627247
URL: http://dx.doi.org/10.1080/03067310802627247

Intern. J. Environ. Anal. Chem.
Vol. 89, No. 5, 20 April 2009, 383–388

Taylor & Francis
Taylor & Francis Group

Evaluating 'goodness-of-fit' for linear instrument calibrations through the origin

Thomas Georgian*

US Army Corps of Engineers, Environmental and Munitions Center of Expertise, 1616 Capitol Ave, Omaha, Nebraska, United States

(*Received 14 August 2008; final version received 13 November 2008*)

Instrument calibrations for environmental analyses frequently entail fitting straight lines forced through the origin, where either the correlation coefficient, Pearson's r, or the percent relative standard deviation, %RSD, for a set of response factors is used to measure the 'goodness-of-fit'. However, these two approaches do not produce comparable linear calibrations. To do this, a weighted regression line needs to be calculated. A weighted regression coefficient is subsequently defined to evaluate the 'goodness-of-fit' and is expressed as function of the %RSD.

Keywords: instrument calibration; response factors

1. Introduction

Response factors and the method of least squares (regression analysis) are commonly used to produce calibration lines for environmental chromatographic methods. For example, there are the two linear calibration options in SW-846 Method 8000C ('Determinative Chromatographic Separations'). To do 'linear calibration using the average calibration or response factor' a mean response factor, $\overline{RF}$, is calculated using at least five calibration standards:

$$\overline{RF} = \sum_{i=1}^{n} RF_i/n = \sum_{i=1}^{n} (y_i/x_i)/n \tag{1}$$

The quantities x_i and y_i denote the known amount (e.g. concentration or weight) and observed instrumental response (e.g. peak area or peak height) of the i-th calibration standard, respectively; n denotes the number of calibration standards (points). The 'goodness-of-fit' is evaluated using the percent relative standard deviation (%RSD) of the set of response factors:

$$\%\mathrm{RSD} = \left(s/\overline{RF}\right) \times 100 \tag{2}$$

The SD of the response factors, s, is calculated from the equation:

$$s = \sqrt{\sum_{i=1}^{n} \left(\overline{RF} - y_i/x_i\right)^2/(n-1)}. \tag{3}$$

*Email: thomas.georgian@usace.army.mil

ISSN 0306–7319 print/ISSN 1029–0397 online

DOI: 10.1080/03067310802627247
http://www.informaworld.com

If the %RSD $\leq$ 20%, linearity through the origin is assumed and the mean response factor is used to determine concentrations of samples using the calibration line:

$$y_i = \overline{RF}\,x_i \tag{4}$$

An alternative calibration option, 'linear calibration using a least squares regression', consists of using (in general) weighted regression to construct a calibration line of the form:

$$y = m'x + b \tag{5}$$

The quantities m' and b are the slope and intercept of the regression line, respectively. The slope and intercept are calculated by minimising the sum of the squared residuals $(\hat{y}_i - y_i)$ [1]:

$$Q = \sum_{i=1}^{n} w_i^2 (\hat{y}_i - y_i)^2 \tag{6}$$

where $\hat{y}_i$ is the predicted (calculated) instrumental response from Equation (5) $(\hat{y}_i = m'x_i + b)$, y_i is the observed response and w_i is a weighting factor. If all the weighting factors $w_i = 1$, then the correlation coefficient Pearson's r is used to evaluate 'goodness-of-fit'. The regression line is acceptable if $r^2 \geq 0.99$. However, if other weighting factors are used (e.g. $w_i = 1/x_i$), then the 'goodness-of-fit' is evaluated using the 'coefficient of determination' (COD) [2]

$$\mathrm{COD} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2 - ((n-1)/(n-p)) \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{7}$$

where $\bar{y}$ is the arithmetic mean of the set of the set of n observed measurements y_i. The calibration line is considered acceptable if COD $\geq$ 0.99. The quantity p is the degrees of freedom for the regression fit ($p = n - 2$). If the calibration line is forced through the origin ($y = m\ x$), $p = n - 1$ and COD $= r^2$.

2. Discussion

There is a problem with the approach for evaluating 'goodness-of-fit' in Method 8000C; namely, the %RSD and r^2 (COD) are not comparable measures of linearity even when the regression line is forced through the origin. In particular, when $r^2 \geq 0.99$, the %RSD is not necessarily $\leq$20%; when the %RSD $\leq$20%, r^2 is not necessarily $\geq$0.99. This is illustrated in Tables 1 and 2. Each table lists the %RSD and r^2 for the non-weighed calibration regression line $y = mx$ ($w_i = 1$) and the response factor line $y = \overline{RF}\,x$ ($w_i = 1/x_i$) for hypothetical calibration data.

Table 1 shows that r^2 can be less than 0.99 when %RSD < 20% and Table 2 shows that the %RSD can be greater than 20% when $r^2 \geq 0.99$. (Figures 1 and 2 are plots of the calibration data in Tables 1 and 2, respectively). However, a weighted correlation coefficient can be calculated to avoid comparability problems between the %RSD and Pearson's r.

A weighted correlation coefficient, r_w, can be defined and expressed as function of %RSD. To do this, it is first noted that the average response factor RF is identically equal

Table 1. Calibration data for which the %RSD < 20% and $r^2 < 0.99$.

x_i	y_i	$RF_i = y_i/x_i$
0.2	0.5	2.5
8	20	2.5
16	50	3.125
24	45	1.875
40	100	2.5
Calibration line	$y = mx$	$y = \overline{RF}\,x$
Slope	$m = 2.42$	$\overline{RF} = 2.5$
Goodness-of-fit	$r^2 = 0.979$	%RSD = 17.7%

Table 2. Calibration data for which the %RSD > 20% and $r^2 > 0.99$.

x_i	y_i	$RF_i = y_i/x_i$
0.2	0.9	4.5
8	20	2.5
16	45	2.8125
24	55	2.2917
40	100	2.42
Calibration line	$y = mx$	$y = \overline{RF}\,x$
Slope	$m = 2.48$	$\overline{RF} = 2.92$
Goodness-of-fit	$r^2 = 0.997$	%RSD = 30.9%

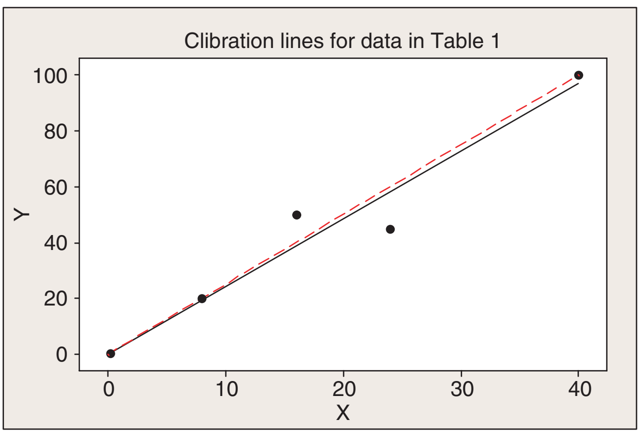

Figure 1. Regression line through the origin (solid line) and average response factor line (dashed line) for calibration data in Table 1.

to the slope m of a weighted regression line forced through the origin when the weighting factors $w_i = 1/x_i$:

$$m = \frac{\sum_{i=1}^{n} w_i^2 x_i y_i}{\sum_{i=1}^{n} w_i^2 x_i^2} = \frac{\sum_{i=1}^{n} (y_i/x_i)}{n} = \overline{RF} \tag{8}$$

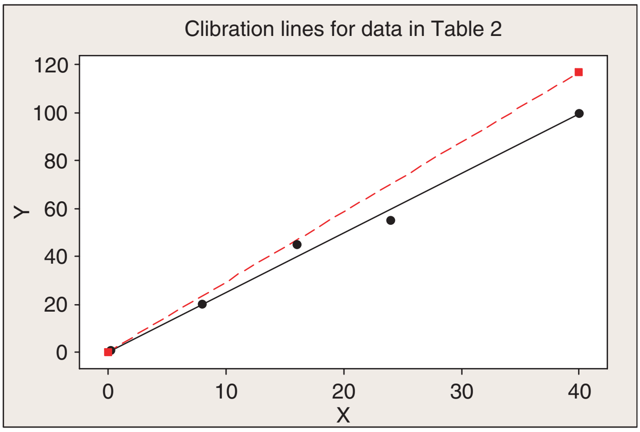

Figure 2. Regression line through the origin (solid line) and average response factor line (dashed line) for calibration data in Table 2.

(This is equivalent to assuming that the variance of the residuals is not homogeneous but is proportional to $1/x^2$). The SD of the response factors is equal to the SD of the residuals of the weighted regression line:

$$s = \sqrt{\frac{Q}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} w_i^2(\hat{y}_i - y_i)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} \left(\overline{RF} - y_i/x_i\right)^2}{n-1}} \tag{9}$$

Therefore, the %RSD of the response factors (Equation (3)) is equal to the SD of the residuals divided by the slope of the weighted regression line. The %RSD is a measure of the 'goodness-of-fit' of the regression line. When there is a 'perfect fit', the sum of the residuals $Q = 0$ and %RSD = 0. However, there is no upper bound for the %RSD when there is no relationship between x and y. Unlike a correlation coefficient, the %RSD is not a normalised quantity that directly measures the proportion of the explained variation of the regression line. Therefore, it seems desirable to calculate a correlation coefficient for the weighted regression line.

The square of the weighted correlation coefficient r_w^2 is defined as the ratio of the explained variation to the total variation about a weighted regression line forced through the origin:

$$\begin{aligned} r_w^2 &= \frac{\sum_{i=1}^{n} w_i^2 y_i^2 - \sum_{i=1}^{n} w_i^2 (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} w_i^2 y_i^2} \\ &= \frac{\sum_{i=1}^{n} w_i^2 y_i^2 - \left[\sum_{i=1}^{n} w_i^2 y_i^2 - m^2 \sum_{i=1}^{n} w_i^2 x_i^2\right]}{\sum_{i=1}^{n} w_i^2 y_i^2} \\ &= \frac{nm^2}{\sum_{i=1}^{n} (y_i/x_i)^2} \end{aligned} \tag{10}$$

It follows from Equations (3) and (8) that

$$s = \sqrt{\frac{\sum_{i=1}^{n} (m - y_i/x_i)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} (y_i/x_i)^2 - nm^2}{n-1}} \tag{11}$$

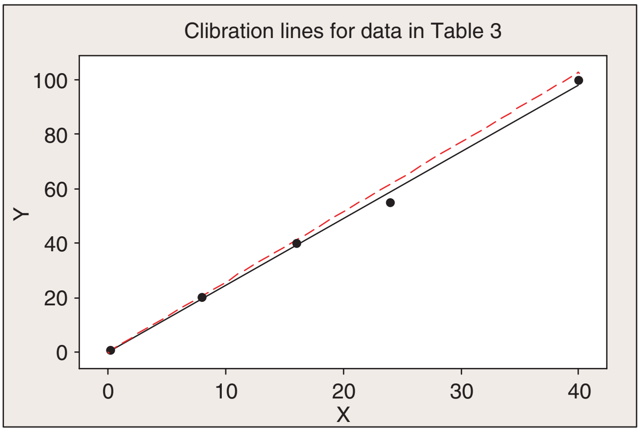

Figure 3. Regression line through the origin (solid line) and average response factor line (dashed line) for calibration data in Table 3.

From Equation (10), it follows that Equation (11) can be re-written as:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (y_i/x_i)^2 - nm^2}{n-1}} = \sqrt{\frac{(nm^2/r_w^2) - nm^2}{n-1}} = m\sqrt{\frac{n(1/r_w^2 - 1)}{n-1}} \tag{12}$$

Substituting the right-hand side of Equation (12) into Equation (2) gives:

$$\%\text{RSD} = \sqrt{\frac{n(1/r_w^2 - 1)}{n-1}} \times 100 \tag{13}$$

Assuming that only positive correlations are meaningful for calibration, solving for r_{w} gives:

$$r_w = \sqrt{\frac{1}{1 + [(n-1)/n](\%\text{RSD}/100)^2}} \tag{14}$$

Therefore, 'perfect fit', %RSD = 0%, corresponds to $r_{\text{w}}^2 = 1$ (and vice versa). Conversely, when there is no correlation, the %RSD becomes infinitely large and r_{w}^2 approaches zero. For $n = 5$ points (standards), an acceptance criterion of $r_{\text{w}}^2 \geq 0.99$ (i.e. no more than 1% of the variation is unexplained by the linear model) implies that the %RSD ≤ 11.2%. On the basis of this criterion, the calibrations summarised in Tables 1 and 2 would be unacceptable, as each %RSD exceeds 11.2%. Figure 3, which was drawn from the calibration data in Table 3, shows a response factor line for which the %RSD is about 11%. Note that both the response factor line (dashed line) and the ordinary least squares line forced through the origin (solid line) fit the calibration data reasonably well.

3. Conclusions

Environmental test methods typically require r^2 to be at least 0.99 or the %RSD of the response factors to be no greater than 15–20% for instrument calibrations. However, r^2 and the %RSD are not comparable measures of 'goodness-of-fit'. There is a lack of

Table 3. Calibration data for which %RSD ≤ 11.2 ($r_w^2 \geq 0.99$).

x_i	y_i	$RF_i = y_i/x_i$
0.2	0.61	3.05
8	20	2.5
16	40	2.5
24	55	2.29167
40	100	2.5
Calibration line	$y = mx$	$y = \overline{RF}\,x$
Slope	$m = 2.45$	$RF = 2.568$
Goodness-of-fit	$r^2 = 0.999$	%RSD = 11.1%

a well-defined relationship between r^2 and the %RSD. The acceptance criterion $r^2 \geq 0.99$ does not imply the %RSD ≤ 15 and vice versa. However, this issue can be resolved for linear calibrations through the origin by defining a weighted correlation coefficient r_w.

Using the average response factor option for calibration is equivalent to constructing a weighted regression line forced through the origin using the weighting factors $1/x_i$. These weighting factors can be used to calculate a weighted correlation coefficient r_w. The %RSD also may be used to evaluate 'goodness-of-fit', as it measures the relative error of the regression line through the origin. However, unlike the square of a correlation coefficient, it is not a normalised quantity that directly measures the proportion of the variation explained by the linear model. Like r^2, r_w^2 ranges from 0 to 1 and is equal to the proportion of variation that is explained by weighted regression line.

The correlation coefficient r_w is a more appropriate measure of 'goodness-of-fit' than Pearson's r for a calibration line calculated from an average response factor, as Pearson's r implies the weighting factors for the line $w_i = 1$ rather than $1/x_i$. Furthermore, unlike the relationship between r^2 and the %RSD, there is one-to-one correspondence (i.e. functional and inverse functional relationship) between r_w^2 and the %RSD. A typical acceptance criterion for calibration such as $r^2 \geq 0.99$ (i.e. 99% of the variation must be explained by the linear model), suggests that r_w^2 should also be at least 0.99. If five standards are used to construct a calibration line (e.g. according to Method 8000C, $n \geq 5$) then $r_w^2 \geq 0.99$ if and only if the %RSD $\leq 11\%$ (not 15–20%). For analytical methods that allow calibrations through the origin using response factors, it is recommended that the weighted correlation coefficient r_w be used to evaluate 'goodness-of-fit' (e.g. in the same manner that r^2 would be used for ordinary least squares fits).

References

[1] J. Neter, W. Wasserman, and M.H. Kutner, *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*, 2nd ed. (Richard D. Irwin, Homewood, IL, 1985).

[2] US Environmental Protection Agency, Office of Solid Waste. EPA SW-846: Test Methods for Evaluating Solid Waste, Physical/Chemical Methods. Third Edition, Final Update IV (2007). <http://www.epa.gov/epawaste/hazard/testmethods/sw846/new_meth.htm#8000C>.