Univariate calibration by reversed regression of heteroscedastic data: a case study

Qiaoling Charlene Zeng,[a] Elizabeth Zhang[a] and Joel Tellinghuisen[b]

First published as an Advance Article on the web 28th October 2008
DOI: 10.1039/b808667b

In a study of calibration with HPLC data for acetaldehyde-DNPH, we have collected replicate data (5–11 points each) for 33 samples spanning the range 0.0004–3 μg of detected analyte. Over most of this range, the data uncertainty is proportional to the signal, implying that weighted least squares is required to obtain the calibration function, since minimum-variance estimation requires weights inversely proportional to the data variance. When a variance function derived from an analysis of the replicate statistics is used to assign weights, $w_i = 1/\sigma_i^2$, the resulting values of χ^2 for the calibration fit are too large by a factor of 400. This implies that the method error is dominated by sample preparation rather than measurement uncertainty, and it means that in the calibration fit, the peak area should be taken as the independent variable and the amount as the dependent. In this reversed regression, the generalized LS method (GLS) is used to estimate the total method variance function from the residuals. The resulting method variance function resembles the instrumental variance, in containing constant and proportional error terms. The calibration data demand at least a cubic polynomial for adequate representation, but other response functions are statistically equivalent, with the result that this model uncertainty is comparable to the directly computed statistical uncertainty of the calibration function. In these computations, emphasis is placed on the virtues of χ^2 as a statistical figure of merit over the widely used R.

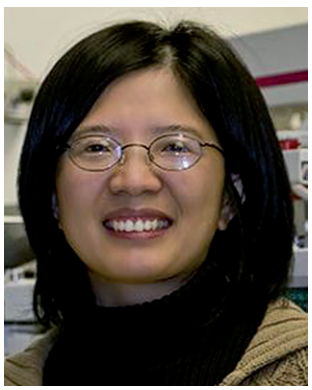
Introduction

In routine analytical work the workhorse calibration method is still classical univariate calibration, usually with assumed linear response functions, and often with neglect of possible data heteroscedasticity. The calibration data are fitted by linear least squares (LLS) to $y = a + bx$, and the unknown concentration or amount x_0 is obtained from $x_0 = (y_0 - a)/b$, where y_0 is the response measured for it. The uncertainty in x_0 is then estimated from[1,2]

$$\sigma_{x_0}^2 = \frac{\sigma^2}{b^2}\left[\frac{1}{r} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \Sigma (x_i - \bar{x})^2}\right], \quad (1)$$

where n is the number of calibration points, r is the number of independent measurements averaged to obtain y_0, and overbars denote averages. Here the data uncertainty σ is assumed to be known, *e.g.* from pooled data from prior similar measurements; then confidence limits are obtained using the normal distribution. More often, analysts take the 'ignorance' viewpoint about the data error and estimate it from the calibration fit itself. Then

[a]*Firmenich, Inc., 250 Plainsboro Road, Plainsboro, NJ 08536, USA*
[b]*Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235, USA*



Qiaoling Charlene Zeng

Qiaoling Charlene Zeng obtained her M.S. in Analytical Chemistry from Xiamen University, P.R. China and her Ph.D. from the Food Science Department of the University of Wisconsin at Madison in 1997. She has been a Senior Scientist in the R&D department at Firmenich Inc. since 1998. Her primary research areas include flavor and fragrance analysis, and analytical method developments and validations. She was a Visiting Scholar in Chemistry at Vanderbilt University from 2006 to 2008.

Joel Tellinghuisen

Joel Tellinghuisen has degrees from Cornell University (A.B.) and the University of California at Berkeley (Ph.D.) and has been Professor of Chemistry at Vanderbilt University since 1983. His primary research areas have been chemical physics and physical chemistry, and his major accomplishments include spectroscopic analysis and interpretation of excimer lasers. In his research he has emphasized computational methods, especially modern least-squares data analysis techniques.

σ is replaced by its estimate s, which is calculated from the sum of squared errors S [$= \Sigma(y_i - a - bx_i)^2$] using $s^2 = S/(n-2)$; and confidence limits are assessed using the t-distribution for $n-2$ ($\equiv \nu$) degrees of freedom. For relatively imprecise b, small correction terms are added to the right-hand side of eqn (1).[3–5]

It is worth recalling the assumptions behind these procedures. The calibration x values are error-free and the y measurements are unbiased estimates of their true values, with random normal (Gaussian) error of constant magnitude. Then the LLS estimates of a and b are normally distributed about their true values with standard errors that are known exactly at the outset if σ is known, or are estimated using s if the latter must be taken as an estimate of σ.[4,5] On the other hand, x_0 is not a linear estimator and in fact does not even have finite variance.[3,5] Thus σ_{x_0} in eqn (1) should be considered an asymptotic approximation; and the distribution of x_0 is inherently non-normal, necessitating the correction terms mentioned earlier when b is imprecise.

Real data often deviate from these assumptions, especially by (1) requiring response functions other than linear and (2) having non-constant uncertainty. One of us has recently described algorithmic procedures that straightforwardly extend the linear response, homoscedastic (constant σ) case to other response functions and heteroscedastic data.[6,7] In this approach, σ_{x_0} is a computational product of the analysis, valid (in the same asymptotic sense) for any response function and any data error structure. However, unless the data are assigned weights inversely proportional to their variances (*i.e.* $w_i \propto \sigma_i^{-2}$), eqn (1) and its equivalent for heteroscedastic data, as well as their computational counterparts in the algorithmic approach, are simply wrong.[8] Perhaps the most frequent exercise of this error is in the use of unweighted or ordinary LS (OLS) to analyze heteroscedastic data.

One assumption that is rarely challenged is the fundamental one: that x is error-free, dictating the choice of it as the independent variable and y as dependent. While some authors have dealt with the case of error in both variables (which makes the LS analysis non-linear),[9] some uncertainty in x is not a serious problem as long as x is significantly more precise than y in a relative sense – say a factor of 3. But many of today's instruments are so good that they can provide estimates of y that are much more precise than other aspects of the calibration procedure, especially the preparation of the calibration samples. This indeed turns out to be true in the present case study, where the method error dwarfs the instrumental error by a factor of 20.

This is actually a 'good news' result: When y can be taken as the independent variable, with regression of x upon y, the estimator of x_0 becomes linear, hence unbiased and normal provided that x has random normal error. Further, since y is now considered error-free, the only statistical error in x_0 is often the easily computed contribution from the calibration fit.

In the present work, we illustrate these considerations through a case study of calibration with precise but strongly heteroscedastic HPLC data for acetaldehyde-2,4-dinitrophenylhydrazone (acetaldehyde-DNPH) solutions spanning four orders of magnitude in the analyte. The data are weighted using an estimated variance function derived from an analysis of replicate data[10] and are fitted to polynomials of varying order in the search for an optimal response function. A cubic is judged best, but the resulting value of χ^2 is a factor of 400 too large, showing that other sources of error are dominant. Those sources can only involve the preparation of the calibration solutions (x), which means that the instrumental measurement of peak areas (y) is actually more precise. Accordingly, we reverse the regression. The replicate measurements cannot be used to assess the uncertainty in x, so we turn to generalized least squares (GLS), estimating the method variance function from the residuals themselves.[11–13] This function is found to have constant and proportional components, mirroring similar contributions to the instrumental variance function.

In the quest for a proper calibration response function, we emphasize χ^2 as a figure of merit for the LS fits and show how the popular R is less useful for this purpose.

Least-squares summary

The method of least squares is described in detail in many places, and the key computational relations were given in matrix notation in a recent study of heteroscedasticity and its neglect.[8] Accordingly these relations will not be reproduced here and indeed are not needed, since the computations described in this paper are all carried out with a data analysis program. We do need to emphasize that for a data set containing n points, the LS solution minimizes $S = \Sigma\, w_i\delta_i^2$ with respect to the p adjustable parameters, where the residuals $\delta_i = y_i - f(x_i)$ represent the differences between the measured values of the dependent variable and their calculated values for the response function $f(x)$. For minimum-variance estimation of the adjustable parameters, the weights w_i must be inversely proportional to the variances σ_i^2. If all σ_i are the same, we have OLS and can use $w_i = 1$; if not, we have weighted least squares (WLS). Clearly, knowledge of σ_i is key to the analysis. In particular, incorrect weighting leads to incorrect estimates of the parameter errors, as already noted. The extent of the damage from incorrect weighting in calibration depends upon factors like the x structure of the data, the range of their w_i, and the location of the unknown relative to the calibrants.[8]

The parameter estimates obtained from an LS fit are generally correlated, so to compute the statistical error in a function of those parameters, we need the full expression for error propagation,

$$\sigma_f^2 = \mathbf{g}^{\mathbf{T}}\mathbf{V}\mathbf{g}, \qquad (2)$$

where $\mathbf{V}$ is the variance-covariance matrix and the elements of the vector $\mathbf{g}$ are the derivatives of f with respect to the parameters. Here we are interested in the calibration function itself. Anticipating results below, we note that for a cubic polynomial, $\mathbf{g}^{\mathbf{T}} = (1\ x\ x^2\ x^3)$.

χ^2 and R^2

For LS fits in which the data are weighted $w_i = 1/\sigma_i^2$ and the fit model is correct, the sum of weighted, squared residuals, $S = \Sigma(\delta_i/\sigma_i)^2$, follows the chi-square (χ^2) distribution, which has an average value ν ($= n - p$) and variance 2ν. Accordingly, S/ν follows the reduced χ^2 distribution, which has an average 1 and variance $2/\nu$. It follows that in OLS of homoscedastic data, s^2 ($= S/\nu$) is distributed as a scaled χ^2 variate; thus it and the estimated

parameter variances have relative standard deviation (RSD) $(2/\nu)^{1/2}$, and the estimated parameter standard errors have RSD $(2\nu)^{-1/2}$. The latter, for example, shows that around 50 points are needed to yield estimated parameter uncertainties that are good to 10%; many routine calibrations employ as few as six points, giving uncertainties that are themselves uncertain by 35% in the customary approach of using the calibration fit itself to estimate σ^2.

Since estimated variances have uncertainty proportional to their magnitude, any fitting of variance-like quantities should employ weights inversely proportional to the squares of the fitted quantities.[14] This applies in particular to the estimation of variance functions from residuals, discussed below.

Analytical chemists commonly use R^2 (R) to judge the quality of calibration fits. R^2 and S are related by[15]

$$R^2 = 1 - \frac{S}{\Sigma w_i (y_i - \bar{y}_w)^2}, \quad (3)$$

where $\bar{y}_w$ is the weighted average of the dependent variable, $\bar{y}_w = \Sigma w_i y_i / \Sigma w_i$. Eqn (3) can be derived from the more specific definition of R for a straight-line fit and is commonly used by data analysis programs to produce R^2 values for fits to other functions. While it is customary to look for the value of R^2 closest to unity for the 'best' fit, the theoretical value of R^2 for a properly weighted fit should be less than 1, since the average of $S\,(=\chi^2) = \nu$ in a properly weighted fit.

Through eqn (3) R^2 and S contain equivalent information about the quality of the fit, and if nothing is known *a priori* about the data error, either can be used as a figure of merit: minimizing S (*i.e.* minimizing the estimated variance) is equivalent to making R^2 as close as possible to 1.

However, significant changes in S can be squeezed into surprisingly small changes in R^2, making the latter harder to 'read' in this context – a limitation among others that have been noted before.[16–20] When the data error is known *a priori* and the weights w_i are taken as ${\sigma_i}^{-2}$, χ^2 becomes a much more powerful tool for assessing the fit quality. The expected value of ν for S is easily remembered and deviations from the expected value can be subjected to a χ^2 test to assess their significance. In the simplest sense, when S exceeds ν by a given factor, it means that the apparent LS data variances exceed their prior estimates by this same factor. At the same time, values of R^2 are not only compressed into a small range near 1; they also have no simple predicted value, because of the denominator in the second term in eqn (3).

Generalized least squares

In the GLS method,[11–13] both the data variance function and the fit parameters are determined, through an iterative bootstrapping process in which the data are fitted, yielding residuals that are then analyzed to estimate the variance function. The latter is used to reassess the weights and the process is repeated, with adequate convergence typically coming in three to five cycles. Since the residuals are both positive and negative, their squares or absolute values are fitted in the variance function estimation (VFE) step. Here we fit the squared residuals, which itself requires weighting, since the squared residuals, like the variance, have the statistical properties of χ^2. We represent the method variance function, ${\sigma_{\text{meth}}}^2$, as

$${\sigma_{\text{meth}}}^2 = \sigma^2 h^2(\text{vars}, \boldsymbol{\theta}), \quad (4)$$

where h^2 is a function expressing the dependence on the variables and parameters ($\boldsymbol{\theta}$), and σ^2 is its scale.[11] From the properties of χ^2 discussed above, the standard deviation in ${\sigma_{\text{meth}}}^2$ is proportional to h^2, meaning that the weights in the VFE fit should be proportional to h^{-4}. At the same time, the weights in the data fit are proportional to h^{-2}. Both sets of weights are reassessed in each cycle, as h^2 evolves toward its final form.

Actually, the LS residuals, though they track the variance, do not all have the same scale. Rather, the ith residual has variance $(1 - H_{ii})\sigma_i^2$, where H_{ii} is from the 'hat' matrix.[8,11] It is customary to correct this scale difference by fitting the squares of $\delta_i/(1 - H_{ii})^{1/2}$, called 'Studentized residuals'. In the present case ($n = 33$), H_{ii} is nearly constant except for the few points near each end of the range. Since ΣH_{ii} = the number of adjustable parameters p, the main effect of this correction is to scale S up by the factor $n/(n - p)$.

We use only polynomials of varying order as response functions, in both the OLS and the GLS computations. Fits to polynomials are LLS, so all the guarantees of LLS apply. We do use non-linear fitting in the VFE stage of the GLS calculations. Although the statistical properties of estimated variance functions are unusual, Monte Carlo computations have confirmed that VFE yields near-optimal calibration functions from surprisingly few data points (<20).[8,14]

All LS fitting was done with the KaleidaGraph program (Synergy Software), using methods similar to those described in related earlier works and their supplements in this journal.[6–8,14]

Materials and methods

Chemicals and reagents

Acetaldehyde-2,4-dinitrophenylhydrazone (acetaldehyde-DNPH) solution (1025.94 μg acetaldehyde/mL in acetonitrile) was purchased from Supelco (Bellefonte, PA). The calibration samples were prepared from the standard by serial dilution with a 50% aqueous solution of HPLC-grade acetonitrile in 18 MΩ cm deionized water (EASYpure™ UV, Barnstead Thermolyne Corp., Dubuque, IA). The dilutions were carried out with small volumetric flasks (10–25 mL, Kimble) and pipettes, including Gilson Microman Pipettes M1000 and M100. The resulting calibration samples contained acetaldehyde in the concentration range 0.01–100 μg mL^{-1}.

HPLC experiments

The HPLC measurements were done with an Agilent 1100 series HPLC system equipped with a diode array detector. Separation was performed on a Phenomenex Luna C18 (2) column (250 × 4.6 mm, 5 μm) at 30 °C. The mobile phase was composed of 55% acetonitrile and 45% deionized water, and the flow rate was 1 mL min^{-1}. Acetaldehyde-DNPH was monitored at 360 nm.

Multiple measurements (5–11) were carried out on each sample using a fixed injection volume. The volume was 30 μL for most samples, but a few experiments employed volumes as low as 1.0 μL and as high as 90 μL. Peak areas were obtained from the autointegration routine in the instrument software. Independent checks of some of these peak areas by numerical integration of the raw chromatograms showed good consistency.

Results and discussion

Instrumental variance

The current calibration data set consists of 33 averaged peak areas spanning injected acetaldehyde amounts from 0.004 to 3 μg (Table 1). The estimated data variance function is derived from the replicate measurements of these points, augmented by additional measurements made later and analyzed in a comparison study of four HPLC analytes.[10] Fig. 1 illustrates the % standard deviations from Table 1, showing that proportional data error ($\sigma_i \propto y_i$, leading to constant % error) dominates at large signal. This dependence is frequently observed in chromatographic data,[13,21–25] but the more complex behavior over large dynamic range is less widely appreciated.

Our full analysis of the s^2 estimates yielded the estimated variance function,[10]

$$\sigma^2 = a^2 + (by)^2 + cy + (s_v/v)^2 y^2, \quad (5)$$

as a function of the peak area y (as output from the peak integration software, units as in Table 1), with $a = 0.20$, $b = 0.0018$, $c = 0.0043$, and s_v (the uncertainty in the injected volume) $= 0.0079$ μL. The structure of this variance function is like that expected and observed for spectrophotometric detection,[26,27] with the addition of the last term for the injection from the syringe. It is noteworthy that the last term exceeds the second term when the injected volume is less than 4 μL. It is also important to recognize that eqn (5) represents just the instrumental data variance function, as it was obtained by injecting a given volume of each sample solution into the instrument multiple times. This must represent the minimum variance for an analytical procedure using this instrument on this analyte.

Table 1 HPLC calibration data for acetaldehyde-DNPH, showing for each point: amount of acetaldehyde, number of replicates, injection volume, average HPLC peak area, and relative standard deviation in peak area

m/μg[a]	No.	v/μL	Area (arb.)[a]	RSD (%)
3.700 (−4)	6	30.0	1.420 (1)	2.18
3.730 (−3)	11	30.0	8.700 (1)	1.11
9.230 (−3)	6	30.0	2.377 (2)	0.43
1.524 (−2)	10	30.0	3.751 (2)	0.20
1.537 (−2)	10	90.0	3.713 (2)	0.12
1.539 (−2)	11	30.0	3.736 (2)	0.31
1.539 (−2)	11	3.0	3.755 (2)	0.40
1.539 (−2)	11	1.0	3.875 (2)	0.67
1.539 (−2)	10	10.0	3.793 (2)	0.23
1.539 (−2)	11	90.0	3.718 (2)	0.29
1.539 (−2)	10	1.0	3.903 (2)	1.22
1.693 (−2)	5	1.1	4.320 (2)	0.43
3.078 (−2)	10	30.0	7.744 (2)	0.25
3.078 (−2)	10	30.0	7.697 (2)	0.32
3.950 (−2)	6	35.0	9.676 (2)	0.43
7.633 (−2)	5	30.0	2.007 (3)	0.24
7.694 (−2)	5	30.0	1.955 (3)	0.23
1.539 (−1)	5	30.0	3.893 (3)	0.27
1.539 (−1)	5	30.0	3.974 (3)	0.14
3.080 (−1)	5	30.0	7.808 (3)	0.19
3.080 (−1)	5	30.0	7.800 (3)	0.29
6.156 (−1)	10	30.0	1.601 (4)	0.14
6.156 (−1)	10	30.0	1.557 (4)	0.17
1.231 (0)	5	30.0	3.138 (4)	0.27
1.231 (0)	5	30.0	3.094 (4)	0.20
1.538 (0)	5	30.0	4.130 (4)	0.06
1.539 (0)	5	30.0	4.062 (4)	0.10
1.847 (0)	5	30.0	4.717 (4)	0.40
1.847 (0)	5	30.0	4.830 (4)	0.17
2.462 (0)	5	30.0	6.130 (4)	0.12
2.462 (0)	5	30.0	6.223 (4)	0.20
3.078 (0)	10	30.0	7.436 (4)	0.21
3.078 (0)	10	30.0	7.401 (4)	0.24

[a] Figures in parentheses are multiplier powers of ten.

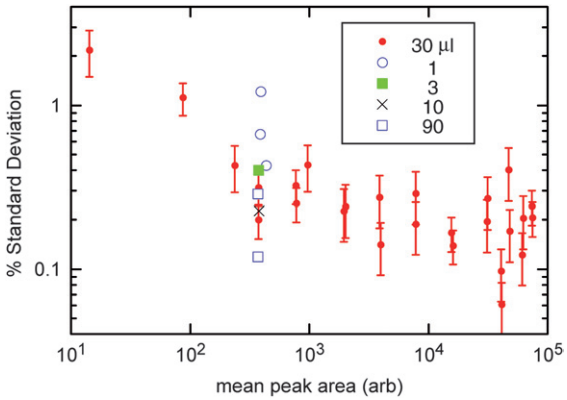


Fig. 1 Estimated % standard deviation (RSD × 100%), displayed as a function of signal and injection volume (legend). Error bars show uncertainty based on the experimental estimates, taken as $s_i/(2v)^{1/2}$ and displayed as symmetric.

Determining the calibration function

We next fit the calibration data from Table 1 to the response function using eqn (5) to compute the weights. Since each value is a mean of $r = 5$–11 measurements, these means are fitted, with weights taken as r/σ^2, *i.e.* as the inverse variances in the mean. A linear response function is inadequate, so we expand the fitting to polynomials of increasing order. Fig. 2 shows results for two goodness-of-fit indicators, χ^2/ν and R, as functions of the polynomial order.

Naive use of the R values in Fig. 2 as the figure of merit for this calibration might induce an analyst to stop with the linear response function, since even this is better than 'three nines'.

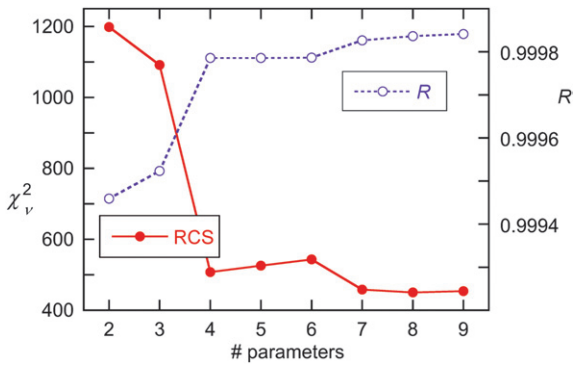


Fig. 2 Reduced χ^2 and R (scale right) for weighted fits of the calibration mean peak areas to polynomials of varying order, from linear ($p = 2$) to eighth order.

However, the χ_ν^2 values tell a different story. First, these drop sharply until $p = 4$, then rise slightly before dropping again at $p = 7$. From the standpoint of *ad hoc* trial-and-error fitting, we might conclude that $p = 7$ is 'best.' However, the three extra parameters give only a 5.5% reduction in χ_ν^2, at the cost of reduced extrapolating ability and the possible introduction of unphysical 'wiggles.' We pick $p = 4$ as a safer and simpler choice, reasonably close to minimum variance.

The scale of the χ_ν^2 values reveals a problem: if the data were truly limited by the instrumental error, χ_ν^2 should be *ca.* 1. Thus the true or method error here must be larger than the instrumental error by a factor of *ca.* 20. We see this in a different way in the residuals plot of Fig. 3, which shows a spread of normalized residuals about 20 times expected. The method error must involve the preparation of the calibration solutions from the stock standard, since there is no other operation involved in the experiments.

We emphasize that this information is not obvious from the R values. As noted earlier, the expected or ideal result for R is <1.00, but the precise value is not as simply predicted as that for χ^2, making R and R^2 of little value as quantitative figures of merit for the calibration fit.

Reversing the regression

The results in Fig. 2 and Fig. 3 imply that we should be fitting x as a function of y, since the latter is much more precisely determined. (Because the relative uncertainty in x is so much greater than that in y, the alternative approach of treating both as uncertain will yield virtually identical results.) The lack of any strong systematic trends in the residuals plot (Fig. 3) provides several pieces of information that are useful for this reversed regression. First, although solution preparation errors are fundamentally systematic in nature, they appear to have been largely randomized over the 33-point data set under study here; thus they can be treated like random statistical errors in the analysis. Second, the data variance function must be at least qualitatively correct; otherwise we might expect to see clear differences in the spread of residuals at the two ends of the range. Third, the cubic response function will probably suffice for the reversed regression. Regarding the second of these, we expect terms like the first two in eqn (5) and perhaps the third, but not the fourth, because the choice of injection volume affected the dilution scheme in only a complex indirect way. Also, our ability to define a variance function from just the 33 residuals in the GLS treatment is limited relative to that from the >300 replicate measurements that led to eqn (5). The efficacy of the cubic response function can be tested in the same trial-and-error procedure used to generate Fig. 2, after we have a handle on the effective data error for the reversed regression.

This last point is important, because in the GLS method the VFE step is perforce tied to the calibration step and hence dependent upon the assumed response function. We start with the cubic function and then check it in the end. In support of this procedure, the residuals shown in Fig. 3 do not change much as the fit order is increased beyond third order.

We first conducted the VFE fitting with neglect of the studentization correction of the residuals. The computations indicated that only two VF parameters could be statistically justified, so we chose the first two terms in eqn (5), rewritten in the form of eqn (4), as

$$\sigma_{\text{meth}}^2 = \sigma^2[1 + (dy)^2], \tag{6}$$

and obtained adequate convergence in three cycles. The resulting value of d was 0.0030(12), with $\sigma = 2.2(7) \times 10^{-4}$ μg. The second term in brackets exceeds the first beyond $y = 300$, confirming that proportional error dominates the method variance over most of the data range.

This analysis expresses the error in the amount m as a function of the mean peak area. Since it is m that is uncertain, we may expect its variance function to depend upon m rather than upon its related peak area, so we repeated the analysis taking m as the independent variable in the VFE computations. Results are shown in Fig. 4. Using these results to compute the needed elements H_{ii}, we then checked the importance of studentizing the residuals. This resulted in only small changes in σ and d, well within their stated uncertainties, and no significant change in the calibration function itself, shown in Fig. 5.

Fig. 6 shows the statistical error in the calibration function. Although the use of eqn (2) for this computation is straightforward, it is not easily implemented in many spreadsheet and data analysis programs. For polynomials there is a simple trick that

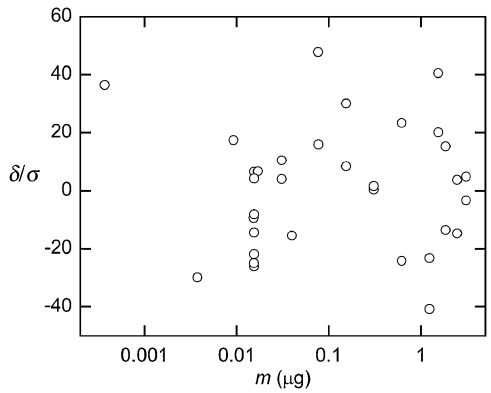


Fig. 3 Normalized fit residuals from fit to cubic response function ($p = 4$) of Fig. 2.

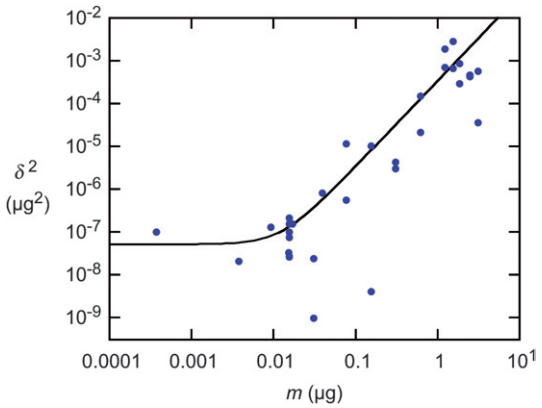


Fig. 4 Squared residuals from cubic calibration fit in converged GLS analysis. The fitted curve is $\sigma^2[1 + (dm)^2]$, with $\sigma = 2.3(8) \times 10^{-4}$ μg and $d = 81(34)$ μg^{-1}.

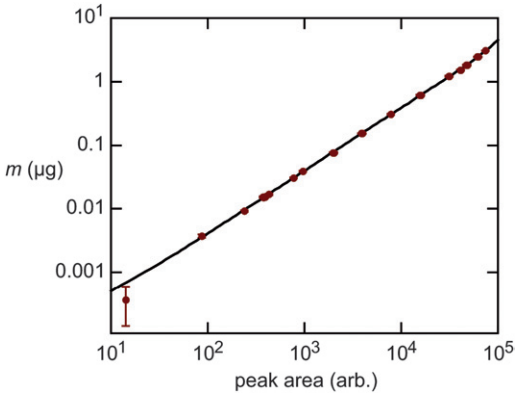


Fig. 5 Reversed calibration function, from weighted fit to a cubic function of the peak area using weights computed from the variance function given in Fig. 4. The fitted values of the parameters are (units μg) $c_0 = 0.00012(13)$, $c_1 = 3.987(30) \times 10^{-5}$, $c_2 = -9.6(2.3) \times 10^{-11}$, and $c_3 = 1.56(33) \times 10^{-15}$, with $\chi^2 = 33.0$. On the scale of this plot, the error bars show for only the first point.

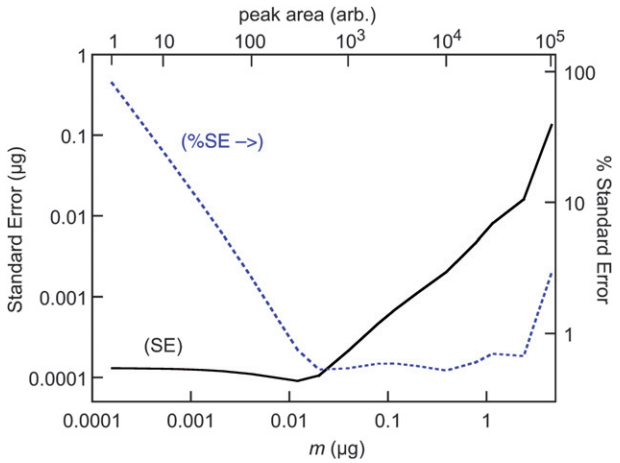


Fig. 6 Standard error (absolute and %) in the calibration function, as a function of the analyte amount and the corresponding peak area (top). The small-scale variability reflects the varying density of calibration points across the range of m.

permits one to compute the error bands without resorting to these matrix expressions: refit to the same polynomial in $(x - x_c)$, where x_c is any chosen value of x.[8] Fits to a polynomial in $(x - x_c)$ are statistically equivalent for all x_c. Since this function is centered at x_c, the constant c_0 is its value at x_c, and the standard error in c_0 is the desired σ_f at x_c. By varying x_c, one can thus compute the error bands over any desired range of x.†

Next we check the effect of altering the response function. Adding a term in x^4 increases χ_ν^2 and renders both c_2 and c_3 statistically insignificant (*i.e.* zero within their standard errors). Dropping the less significant of these terms (c_2x^2) does produce

† Dropping internal terms from a polynomial in $(x - x_c)$ makes the fits for different x_c inequivalent, requiring use of eqn (2) for rigorous error band computations in place of the simpler recentering approach. However, in practice this inequivalence is weak across the range of the calibration data, so the recentering-based estimates remain good approximations of the correct quantities.

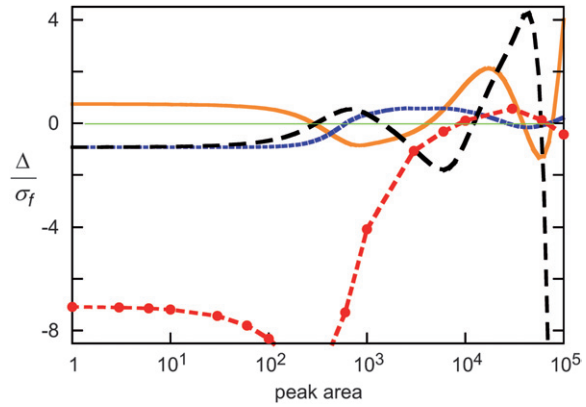


Fig. 7 Calibration function model error. Δ is the difference between the alternative fit model and the cubic polynomial of Fig. 5, and σ_f is the statistical error in the latter, from Fig. 6. The solid (orange) line is for a quartic polynomial missing the quadratic term; the fine-dashed (blue) line is for a cubic with $c_0 = 0$; the broad-dashed (black) line is for a sixth-order polynomial with $c_0 = 0$. The broken curve with points (red) is for an unweighted fit to a linear response function, scaled by a factor of $^1/_{10}$ for display purposes.

a four-parameter fit having marginally lower χ^2 (31.7) than that in Fig. 5. Alternatively, we can reduce χ_ν^2 by dropping the constant term in the calibration function of Fig. 5 (since it is within 1σ of zero). As a third possibility, we find a significant drop in χ_ν^2 for a sixth-order polynomial, with only c_0 insignificant; dropping it gives a six-parameter function with lowest χ_ν^2. To judge whether these changes in the response function are significant, we benchmark them against the statistical error in Fig. 7, which also includes results for OLS. We see that the four weighted calibration models, which are arguably equivalent, do differ in their predictions by amounts comparable to the statistical error; in contrast, the unweighted fit to a linear response function is systematically low by *ca.* 70σ over the first three decades of peak area – a dramatic consequence of the neglected weights. Lacking solid reasons for choosing one weighted model over another, we might reasonably elect to scale up the statistical error in Fig. 6 to roughly reflect this model uncertainty for the region of interest.

If the unknown samples are used directly, without any preprocessing, the errors in Fig. 6, perhaps expanded to include model error, would represent the standard deviation for a single determination using these calibration data. If the unknowns must be treated beforehand, using procedures like those used to prepare the calibration samples, each unknown would have a statistical variance comparable to the derived variance function in Fig. 4. The uncertainty of an unknown would then be estimated as the square root of the sum of this 'preparation' variance and the calibration variance.

Our initial target in this study was the instrumental variance function, so we paid little attention to the solution preparation procedures that are now found to dominate the method variance. In retrospect, the errors can be attributed largely to the volumetric instruments used in the dilution sequences. For example, even the volume uncertainty for the 10 mL volumetric flasks (many different ones of which were used) is 0.2%, which equals the dominant instrumental error for most of our samples. The

Gilson micropipettes are a greater source of uncertainty: at delivery volumes of 100 μL the M1000 can have systematic error as large as 3% and random error up to 1.6%. With the knowledge we now have of the instrumental error, we could devise solution preparation procedures that would significantly reduce the method error, if that were desirable.

Conclusion

Precise HPLC data for acetaldehyde are used to obtain a calibration curve spanning four orders of magnitude in the analyte. Over this broad range, the data exhibit strong heteroscedasticity, dominated by proportional error at large signal and constant at small error. When the calibration data are fitted using weights based on the instrumental data variance function, the χ^2 values are too large by a factor of 400, leading to the conclusion that uncertainty in sample preparation greatly exceeds that of measurement. The implication of this result is that the regression should be reversed. By contrast, the widely used R statistic provides little guidance here.

For the reversed regression, the GLS method is used to estimate the method variance function, which is represented as a sum of a constant term and a proportional error (2%) that dominates over most of the calibration range. The calibration data require a polynomial of order three for adequate representation. Comparison of predictions from this model and other, higher-order models of similar statistical quality show that the dependence on model choice can exceed the statistical uncertainty of calibration.

The method variance derived here is dominated by volumetric procedural uncertainties. Given the high precision with which many modern instruments operate, it is probably not rare to find the measurement error much smaller than other sources of uncertainty, which means that the reverse regression approach we have illustrated here should see wider use. Of course, knowledge of the instrumental error is required to make this decision. Then, excessively large values of χ^2 for weighted calibration fits signal the presence of other, larger sources of uncertainty.

When sample preparation is the main source of uncertainty in calibration, it will typically require more operator effort to characterize the method uncertainty. Here 33 samples sufficed to determine two variance function parameters to only about 40% RSD, but 33 is many more than obtained in typical calibration efforts. The use of sample replicates might help in estimating the method variance function, but then one must take care that the replications are truly random in their sampling of the procedures (*e.g.* that they effectively randomize the systematic errors in volumetric equipment). One possibility in this regard is the use of accumulated data from procedures used repeatedly in day-to-day routine work.[7,8]

Routine univariate calibration often requires much less than the 0.2% limiting instrumental uncertainty of the present study, and many of the present results might seem academic for such work. Indeed, weights can be neglected with little precision loss if the calibration data are structured to approximately center the unknown.[8] However, there are many situations in which the inherent precision is much less than 0.2%; and when a calibration function is desired over a wide operating range, approaches like those discussed here can be important for realizing the full potential of the analysis method.

References

1 J. N. Miller, *Analyst*, 1991, **116**, 3.
2 K. Danzer and L. A. Currie, *Pure Appl. Chem.*, 1998, **70**, 993.
3 G. K. Shukla, *Technometrics*, 1972, **14**, 547.
4 R. N. Draper and H. Smith, *Applied Regression Analysis*, 3rd edn, 1998, Wiley, New York.
5 J. Tellinghuisen, *Fresenius' J. Anal. Chem.*, 2000, **368**, 585.
6 J. Tellinghuisen, *Analyst*, 2000, **125**, 1045.
7 J. Tellinghuisen, *Analyst*, 2005, **130**, 370.
8 J. Tellinghuisen, *Analyst*, 2007, **132**, 536.
9 P. J. Ogren and J. R. Norton, *J. Chem. Educ.*, 1992, **69**, A130.
10 Q. C. Zeng, E. Zhang, H. Dong and J. Tellinghuisen, *J. Chromatogr., A*, 2008, **1206**, 147.
11 R. J. Carroll and D. Ruppert, *Transformation and Weighting in Regression*, 1988, Chapman and Hall, New York.
12 M. Davidian and P. D. Haaland, *Chemom. Intell. Lab. Syst.*, 1990, **9**, 231.
13 K. Baumann and H. Wätzig, *J. Chromatogr., A*, 1995, **700**, 9.
14 J. Tellinghuisen, *Analyst*, 2008, **133**, 161.
15 W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes*, 1986, Cambridge University Press, Cambridge, UK.
16 M. Mulholland and D. B. Hibbert, *J. Chromatogr., A*, 1997, **762**, 73.
17 R. de Levie, *J. Chem. Educ.*, 2003, **80**, 1030.
18 W. Huber, *Accredit. Qual. Assur.*, 2004, **9**, 726.
19 D. B. Hibbert, *Accredit. Qual. Assur.*, 2005, **10**, 300.
20 S. L. R. Ellison, *Accredit. Qual. Assur.*, 2006, **11**, 146.
21 E. L. Johnson, D. L. Reynolds, D. S. Wright and L. A. Pachla, *J. Chromatogr. Sci.*, 1988, **26**, 372.
22 N. V. Nagaraja, J. K. Paliwal and R. C. Gupta, *J. Pharm. Biomed. Anal.*, 1999, **20**, 433.
23 A. M. Almeida, M. M. Castel-Branco and A. C. Falcão, *J. Chromatogr., B*, 2002, **774**, 215.
24 M. K. Kiser and J. W. Dolan, *LC-GC North America*, 2004, **22**, 112.
25 Publication now found in the literature for the 1200 Series system, 'Agilent 1200 Series Injection Systems', locatable by entering this title in the Keywords box at the Literature Library page, http://www.chem.agilent.com/scripts/library.asp (accessed 21 October 2008).
26 J. D. Ingle, Jr and S. R. Crouch, *Spectrochemical Analysis*, 1988, Prentice-Hall, Englewood Cliffs, New Jersey.
27 J. Tellinghuisen, *Appl. Spectrosc.*, 2000, **54**, 431.