Basic Statistical Methods for Analytical Chemistry
Part 2. Calibration and Regression Methods*
A Review

James N. Miller
Department of Chemistry, Loughborough University of Technology, Loughborough, Leicestershire LE11 3TU, UK

Summary of Contents
Linear calibration
    Correlation coefficient
    'Least squares' line
    Errors and confidence limits
Method of standard additions
Limit of detection and sensitivity
Intersection of two straight lines
Residuals in regression analysis
Regression techniques in the comparison of analytical methods
Robust and non-parametric regression methods
Analysis of variance in linear regression
Weighted linear regression methods
Partly straight, partly curved calibration plots
Treatment of non-linear data by transformations
Curvilinear regression
Spline functions and other robust non-linear regression methods

Keywords: *Analytical calibration method; statistics and rectilinear graph; curve fitting method; robust and non-parametric method; review*

Introduction

Most methods in modern analytical science involve the use of optical, electrical, thermal or other instruments in addition to the manipulative 'wet chemistry' skills which are an essential part of the analyst's training. Instrumental methods bring chemical benefits, such as the ability to study a wide range of concentrations, achieve very low limits of detection and perhaps study two or more analytes simultaneously. They also bring the practical benefits of lower unit costs and increased speed of analysis, perhaps through partial or complete automation. The results of instrumental analyses are evaluated by using calibration methods that bring about and reflect these advantages and are, to some extent, distinct from the statistical approaches discussed in Part 1 of this review.[1] Nonetheless many of the concepts summarized in Part 1 are also applied in the statistics of calibration methods, and familiarity with these concepts is assumed here.

A typical calibration experiment (a single analyte–multivariate calibration has recently been surveyed[2]) is performed by making up a series of standard solutions containing known amounts of the analyte and taking each solution separately through an instrumental analysis procedure with a well defined protocol. For each solution, the instrument generates a signal, and these signals are plotted on the *y*-axis of a calibration graph, with the standard concentrations on the *x*-axis. A straight line or curve is drawn through the calibration points and may then be used for the determination of a test ('unknown') sample. The unknown is taken through exactly the same analysis protocol as the standards, the instrument signal is recorded and the test concentration estimated from the calibration graph by interpolation—and not, with one

special exception described below, by extrapolation. It is apparent that one calibration graph can be used in the determination of many test samples, provided that instrument conditions and the experimental protocol do not change. This approach thus offers the desired feature of being able to analyse many samples rapidly over a range of concentrations; less obvious advantages include the ability to estimate limits of detection (see below) and eliminate the effects of some types of systematic error. For example, if the monochromator in a spectrophotometer has an error in its wavelength scale, errors in calculated concentrations using this instrument should cancel out between the standards and the samples.

This approach to the determination of concentrations poses several problems. What type of line—straight, curved, or part-straight, part-curved—should be drawn through the calibration points? Given that the instrument signals obtained from the standards will be subject to random errors, what is the best straight line or curve through those points? What are the errors in test concentrations determined by interpolation? What is the limit of detection of the analysis? These and other statistical questions posed by calibration experiments still generate new methods and excite considerable controversy. Not surprisingly, it is in the area of curve-fitting that most new procedures are being introduced, but linear regression methods also generate their own original literature, as will become apparent.

Linear Calibration

Correlation Coefficient

Many analytical procedures are carefully designed to give a linear calibration graph over the concentration range of interest, and analysts who use such methods routinely may assume linearity with only occasional checks. In the develop-

---

* For Part 1 of this series see reference 1.

ment of new methods, and in any other case where there is the least uncertainty, the assumption of linearity must be carefully investigated. It is always valuable to inspect the calibration graph visually on graph paper or on a computer monitor, as gentle curvature that might otherwise go unnoticed is often detected in this way (see below). Here, and in many other aspects of calibration statistics, the low-cost computer programs available for most personal computers are very valuable. As will be seen, it is important to plot the graph with the instrument response on the y-axis and the concentrations of the standards on the x-axis. One of the calibration points should normally be a 'blank', i.e., a sample containing all the reagents, solvents, etc., present in the other standards, but no analyte. It is poor practice to subtract the blank signal from those of the other standards before plotting the graph. The blank point is subject to errors as are all the other points and should be treated in the same way. As shown in Part 1 of this review,[1] if two results, x_1 and x_2, have random errors e_1 and e_2, then the random error in $x_1 - x_2$ is not $e_1 - e_2$. Thus, subtraction of the blank seriously complicates the proper estimation of the random errors of the calibration graph. Moreover, even if the blank signal is subtracted from the other measurements, the resulting graph may not pass exactly through the origin.

Linearity is often tested using the correlation coefficient, r. This quantity, whose full title is the 'product-moment correlation coefficient', is given by

$$r = \frac{\sum_i [(x_i - \bar{x})(y_i - \bar{y})]}{\{[\sum_i (x_i - \bar{x})^2][\sum_i (y_i - \bar{y})^2]\}^{\frac{1}{2}}} \quad (1)$$

where the points on the graph are $(x_1, y_1), (x_2, y_2), \ldots (x_i, y_i)$, $\ldots (x_n, y_n)$, and $\bar{x}$ and $\bar{y}$ are, as usual, the mean values of x_i and y_i respectively. It may be shown that $-1 \leq r \leq +1$. In the hypothetical situation when $r = -1$, all the points on the graph would lie on a perfect straight line of negative slope; if $r = +1$, all the points would lie exactly on a line of positive slope; and $r = 0$ indicates no linear correlation between x and y. Even rather 'poor' calibration graphs, i.e., with significant y-direction errors, will have r values close to 1 (or -1), values of $|r| <$ about 0.98 being unusual. Worse, points that clearly lie on a gentle curve can easily give high values of $|r|$. So the magnitude of r, considered alone, is a poor guide to linearity. A study of the 'y-residuals' (see below) is a simple and instructive test of whether a linear plot is appropriate. A recent report of the Analytical Methods Committee[3] provides a useful critique of the uses of r, and suggests an alternative method of testing linearity, based on the weighted least squares method (see below).

'Least Squares' Line

If a linear plot is valid, the analyst must plot the 'best' straight line through the points generated by the standard solutions. The common approach to this problem (not necessarily the best!) is to use the unweighted linear least squares method, which utilizes three assumptions. These are (i) that all the errors occur in the y-direction, i.e., that errors in making up the standards are negligible compared with the errors in measuring instrument signals, (ii) that the y-direction errors are normally distributed, and (iii) that the variation in the y-direction errors is the same at all values of x. Assumption (ii) is probably justified in most experiments (although robust and non-parametric calibration methods which minimize its significance are available, see below), but the other two assumptions merit closer examination.

The assumption that errors only occur in the y-direction is effectively valid in many experiments; errors in instrument signals are often at least 2–3% [relative standard deviation (RSD)], whereas the errors in making up the standards should be not more than one-tenth of this. However, modern automatic techniques are dramatically improving the precision

of many instrumental methods; flow injection analysis, for example, shows many examples of RSDs of 0.5% or less.[4] In such cases, it may be necessary either to abandon assumption (i) (again, suitable statistical methods are available—see below), or to maintain the validity of the assumption by making up the standards gravimetrically rather than volumetrically, i.e., with an even greater accuracy than usual. If the assumption is valid, the line calculated as shown below, is called the line of regression of y on x, and has the general formula $y = bx + a$, where b and a are, respectively, its slope and intercept. This line is calculated by minimizing the sums of the squares of the distances between the standard points and the line in the y-direction. (Hence the term 'least squares' for this method.) It is important to note that the line of regression of x on y would seek to minimize the squares of x-direction errors, and therefore would be entirely inappropriate when the signal is plotted on the y-axis. (The two lines are not the same except in the hypothetical situation when all the points lie exactly on a straight line.) The y-direction distances between each calibration point and the point on the calculated line at the same value of x are known as the y-residuals and are of great importance in several calculations, as will be shown later in this paper.

Assumption (iii), that the y-direction errors are equal, is also open to comment. In statistical terms it means that all the points on the graph are of equal weight, i.e., equal importance in the calculation of the best line—hence the term 'unweighted' least squares. In recent years this assumption has been tested for several different types of instrumental analysis, and in many cases it is found that the y-direction errors tend to increase as x increases, though not necessarily in linear proportion. Such findings should encourage the use of weighted least squares methods, in which greater weight is given to those points with the smallest experimental errors. These points are discussed further in a later section.

If assumptions (i)–(iii) are accepted then the slope, b, and intercept, a, of the unweighted least squares line are found from

$$b = \frac{\sum_i [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_i (x_i - \bar{x})^2} \quad (2)$$

$$a = \bar{y} - b\bar{x} \quad (3)$$

The equations show that, when b has been determined, a can be calculated by using the fact that the fitted line passes through the centroid, $(\bar{x}, \bar{y})$. These results are proved in reference 5, a classic text on the mathematics of regression methods. The values of a and b can be simply applied to the determination of the concentration of a test sample from the corresponding instrument output.

Errors and Confidence Limits

The concentration value for a test sample calculated by interpolation from the least squares line is of little value unless it is accompanied by an estimate of its random variation. To understand how such error estimates are made, it is first important to appreciate that analytical scientists use the line of regression of y on x in an unusual and complex way. This is best appreciated by considering a conventional application of the line in a non-chemical field. Suppose that the weights of a series of infants are plotted against their ages. In this case the weights would be subject to measurement errors and to inter-individual variations (e.g., all 3 month old infants would not weigh the same), so would be correctly plotted on the y-axis: the infants' ages, which would presumably be known exactly, would be plotted on the x-axis. The resulting plot would be used to predict the average weight (y) of a child of given age (x). That is, the graph would be used to estimate a y-value from an input x-value. The y-value obtained would of course be subject to error, because the least squares line itself

is subject to uncertainty. The graph would not normally be used to estimate the age of a child from its weight!

In analytical work, however, the calibration graph is used in the inverse way—an experimental value of y (y_0, the instrument signal for a test sample) is input, and the corresponding value of x (x_0, the concentration of the test sample) is determined by interpolation. The important difference is that x_0 is subject to error for two reasons, (1) the errors in the calibration line, as in the weight *versus* age example, and (2) the random error in the input y_0 value. Error calculations involving this 'inverse regression' method[5] are thus far from simple and indeed involve approximations (see below).

First, we must estimate the random errors of the slope and intercept of the regression line itself. These involve the preliminary calculation of the important statistic $s_{y/x}$, which is given by

$$s_{y/x} = \left[\frac{\sum\limits_i (y_i - \hat{y}_i)^2}{n-2} \right]^{\frac{1}{2}} \qquad (4)$$

In this equation, each y_i value is a measured signal value from the analytical instrument, while the corresponding $\hat{y}_i$ is the value of y on the fitted straight line at the same value of x. Each $(y_i - \hat{y}_i)$ value is thus a y-residual (see above). It is clear that equation (4) is similar to the equation for the standard deviation of a series of replicate results, except that the term $(n-2)$ appears in the denominator as the number of degrees of freedom of the data, rather than $n-1$. This difference is explained below in the discussion of analysis of variance applied to regression calculations. After $s_{y/x}$ has been determined, the standard deviation of the slope, s_b, and the standard deviation of the intercept, s_a can be determined from

$$s_b = \frac{s_{y/x}}{[\sum\limits_i (x_i - \bar{x})^2]^{\frac{1}{2}}} \qquad (5)$$

$$s_a = s_{y/x} \left[\frac{\sum\limits_i x_i^2}{n \sum\limits_i (x_i - \bar{x})^2} \right]^{\frac{1}{2}} \qquad (6)$$

These standard deviations can then be used to estimate the confidence limits for the true slope and intercept values. The confidence limits for the slope are given by $b + ts_b$, where the value of t is chosen at the desired confidence level (two-tailed values) and with $n-2$ degrees of freedom. Similarly, the confidence limits for the intercept are given by $a \pm ts_a$. These confidence limits are often of practical value in determining whether the slope or intercept of a line differs significantly from a particular or predicted value. For example, to test whether the intercept of a line differs significantly from 0 at the 95% confidence level, we need only see whether or not the 95% confidence interval for a includes zero.

The statistic $s_{y/x}$ is also used to provide equations for the confidence interval of the mean value of y_0 at a particular x_0 value, and for the (wider) confidence interval for a new and single value of y_0 measured at $x = x_0$. These equations are of limited value in analytical work, as already noted, and examples are given in standard texts.[5–7] Estimating the confidence limits for the entire line is more complex, as a combined confidence region for a and b is required. This problem was apparently first addressed by Working and Hotelling[8] in 1929, and there is a useful summary of their method and of related studies in the often-cited paper by Hunter.[9] The general form of the confidence limits is shown in Fig. 1(a), from which it is clear that the confidence limits are at their narrowest (best) in the region of $(\bar{x}, \bar{y})$, as the regression line must pass through this point.

We can now reconsider the principal analytical problem, that of estimating the standard deviation, s_{x_0}, and the confidence interval of a single concentration value x_0 derived from an instrument signal y_0. As shown diagrammatically in

Fig. 1, the confidence interval for this x_0 value results from the uncertainty in the measurement of y_0, combined with the confidence interval for the regression line at that y_0 value. The standard deviation s_{x_0} is given by

$$s_{x_0} = \frac{s_{y/x}}{b} \left[1 + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum\limits_i (x_i - \bar{x})^2} \right]^{\frac{1}{2}} \qquad (7)$$

It can be shown[5] that this equation is an approximation that is only valid when the function

$$g = \frac{t^2}{\left\{ b \middle/ \left[\frac{s_{y/x}^2}{\sum\limits_i (x_i - \bar{x})^2} \right]^{\frac{1}{2}} \right\}^2} \qquad (8)$$

has a value less than about 0.05. For g to have low values it is clearly necessary for b and $\sum\limits_i (x_i - \bar{x})^2$ to be relatively large and $s_{y/x}$ to be small. In an analytical experiment with reasonable precision and a good calibration plot these results are indeed obtained; for example the data given in reference 9 yield a g value of 0.002.

In a typical analysis, the value of y_0 might be obtained as the mean of m observations of a test sample, rather than as a single observation. In which case, the (approximate) equation for s_{x_0} becomes

$$s_{x_0} = \frac{s_{y/x}}{b} \left[\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum\limits_i (x_i - \bar{x})^2} \right]^{\frac{1}{2}} \qquad (9)$$

After s_{x_0} has been calculated, the confidence limits for x_0 can be determined as $x_0 \pm ts_{x_0}$, with t again chosen at a desired confidence level and $n-2$ degrees of freedom. Inspection of equations (7) and (9) provides important guidance on the performance of a calibration experiment, presuming that we wish to minimize s_{x_0}. In cases where $m = 1$, the first of the three terms within the bracket in these equations is generally the largest. Thus, making only a small number of replicate determinations of y_0 can dramatically improve the precision of x_0. Similarly, increasing the number of calibration points, n, is beneficial. If considerations of time, material availability, *etc.* limit the total number of experiments $(m + n)$ that can be
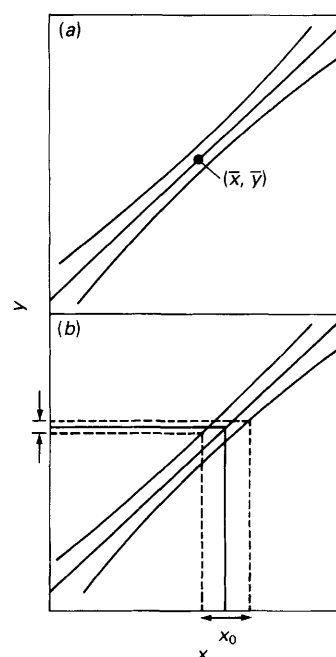

Fig. 1 Confidence limits in linear regression: (a) shows the hyperbolic form of the confidence limits for a predicted y-value; and (b) shows how these confidence limits combine with the uncertainty in y_0 to yield a confidence interval for a predicted x-value, x_0

performed, the sum of the first two components of the bracketed term in (7) and (9) is minimized by setting $m = n$. However, small values of n are to be avoided for a separate reason, *viz.*, that the use of $n - 2$ degrees of freedom then leads to very large values of t and correspondingly wide confidence intervals. Calculation shows that, in the simple case where $y_0 = \bar{y}$, then for any given values of $s_{y/x}$ and b, the priority (at the 95% confidence level) is to avoid values of $n < 5$ because of the high values of t associated with <3 degrees of freedom. When $n \geqslant 5$, maximum precision from a fixed number of measurements is obtained when $m = n$.

The last bracketed term in equations (7) and (9) shows that precision (for fixed m and n) is maximized when y_0 is as close as possible to $\bar{y}$ (this is expected in view of the confidence interval variation shown in Fig. 1), and when $\sum_i (x_i - \bar{x})^2$ is as large as possible. The latter finding suggests that calibration graphs might best be plotted with a cluster of points near the origin, and another cluster at the upper limit of the linear range of interest [Fig. 2(a)]. If n calibration points are determined in two clusters of $n/2$ points at the extremes of a straight line, the value of the term $\sum_i (x_i - \bar{x})^2$ is increased by a factor $[3(n - 1)/(n + 1)]$ compared with the case in which the n points are equally spaced along the same line [Fig. 2(b)]. In practice it is usual to use a calibration graph with points roughly equally distributed over the concentration range of interest. The use of two clusters of points gives no assurance of the linearity of the plot between the two extreme x values; moreover, the term $[(y_0 - \bar{y})^2/b^2 \sum_i (x_i - \bar{x})^2]$ is often the smallest of the three bracketed terms in equations (7) and (9), so reducing its value further may have only a marginal over-all effect on the precision of x_0.

Method of Standard Additions

In several analytical methods (*e.g.*, potentiometry, atomic and molecular spectroscopy) matrix effects on the measured signal demand the use of the method of standard additions. Known amounts of analyte are added (with allowance for any dilution effects) to aliquots of the test sample itself, and the calibration graph (Fig. 3) shows the variation of the measured signal with the amount of analyte added. In this way some matrix effects are equalized between the sample and the standards. The concentration of the test sample, x_e, is given by the intercept on the x-axis, which is clearly the ratio of the

y-axis intercept and the slope of the calibration line, calculated using equations (2) and (3), *i.e.*,

$$x_e = a/b \tag{10}$$

The standard deviation of x_e, s_{x_e}, is given by a modified form of equation (7):

$$s_{x_e} = \frac{S_{y/x}}{b}\left[\frac{1}{n} + \frac{\bar{y}^2}{b^2\sum_i (x_i - \bar{x})^2}\right]^{\frac{1}{2}} \tag{11}$$

This standard deviation can as always be converted into a confidence interval using the appropriate t value. It might be expected that such confidence intervals would be wider for this extrapolation method than for a conventional interpolation method. In reality, however, this is not so, as the uncertainty in the value of x_e derives only from the random errors of the regression line itself, the corresponding value of y being fixed at zero in this case. The real disadvantages of the method of standard additions are that each calibration line is valid for only a single test sample, larger amounts of the test sample may be needed and automation is difficult.

The slope of a standard additions plot is normally different from that of the conventional calibration plot for the same sample. The slope ratio is a measure of the proportional systematic error produced by the matrix effect, a principle used in many 'recovery' experiments.[10] The use of the conventional standard additions method has been discussed at length by Cardone.[11,12] The generalized standard additions method (GSAM)[13] is applicable to multicomponent analysis problems, but belongs to the realm of chemometrics.[14]

Limit of Detection and Sensitivity

The ability to detect minute amounts of analyte is a feature of many instrumental techniques and is often the major reason for their use. Moreover, the concept of a limit of detection (LOD) seems obvious: it is the least amount of material the analyst can detect because it yields an instrument response significantly greater than a blank. Nonetheless, the definition and measurement of LODs has caused great controversy in recent years, with additional and considerable confusion over nomenclature, and there have been many publications by statutory bodies and official committees in efforts to clarify the situation. Ironically, the significance of LODs, at least in the strict quantitative sense, is probably overestimated. There is clearly a need for a means of expressing that (for example) spectrofluorimetry at its best is capable of determining lower amounts of analytes than absorptiometry, and the principal use of LODs in the literature appears to be to show that a newly discovered method is indeed 'better' than its predecessors. But there are many reasons why the LOD of a particular method will be different in different laboratories, when
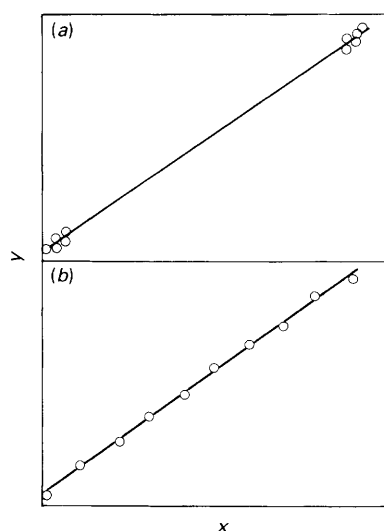

Fig. 2 Calibration graphs with: (a) clusters of standards at high and low concentrations; and (b) equally spaced standards
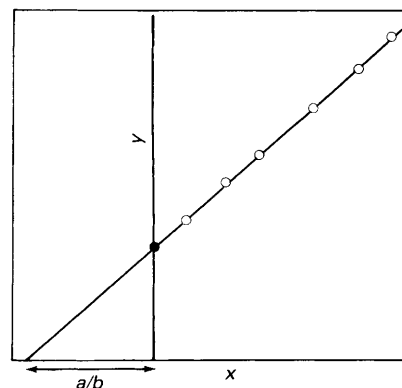

Fig. 3 Calibration graph for the method of standard additions. The point ● is due to the original sample; for details see text

applied to different samples or used by different workers. Not least among the problems is, as always, the occurrence of hidden and possibly large systematic errors, a point rightly emphasized by a recent report of the Analytical Methods Committee of the Analytical Division of the Royal Society of Chemistry.[15] It can thus be very misleading to read too much into the absolute value of a detection limit.

One principle on which all authorities are agreed is that the sensitivity of a method is not the same as its LOD. The sensitivity is simply the slope of the calibration plot. As calibration plots are often curved (see below) the concentration range over which the sensitivity applies should be quoted. In practice the concept of sensitivity is of limited value in comparing methods, as it depends so much on experimental conditions (e.g., the sensitivity of a spectrophotometric determination can simply be increased by increasing the optical path length). Comparisons of closely related methods —for example, of spectrophotometric methods for iron(III) using three organic chelating reagents with different molar absorptivities in 10 mm cuvettes[16]—may be of value.

The most common definitions of the LOD take the form in which the lowest detectable instrument signal, y_L, is given by

$$y_L = y_B + ks_B \qquad (12)$$

where y_B and s_B are, respectively, the blank signal and its standard deviation. Any sample yielding a signal greater than y_L is held to contain some analyte, while samples yielding signals $<y_L$ are reported to contain no detectable analyte. The constant, k, is at the discretion of the analyst, and it cannot be emphasized too strongly that there is no single, 'correct', definition of the LOD. It is thus essential that, whenever an LOD is quoted, its definition should also be given. After y_L has been established, it can readily be converted into a mass or concentration LOD, c_L, by using the equation

$$c_L = ks_B/b \qquad (13)$$

This equation shows the relationship between the LOD and sensitivity, the latter being given by the slope of the calibration graph, b, if the graph is linear throughout.

Kaiser[17] suggested that k should have a value of 3 (although other workers have, at least until recently, used $k = 2$, $k = 2^{3/2}$, etc.). This recommendation has been reinforced by the International Union of Pure and Applied Chemistry (IUPAC)[18] and others, and is now very common. It is important to clarify the significance of this definition. Fig. 4(a) illustrates the distribution of the random errors of y_B, the standard deviation σ_B being estimated by s_B, as always. The probability that a blank sample will yield a signal greater than $y_B + 3s_B$ is given by the shaded area, readily shown, using tables of the standard normal distribution, to be 0.00135, i.e., 0.135%. This is the probability that a false positive result will occur, i.e., that analyte will be said to be present when it is in fact absent. This is analogous to a type I error in conventional statistical tests.[1] However, the second type of error (type II error), that of obtaining a false negative result, i.e., deducing that analyte is absent when it is in fact present, can also occur. If numerous measurements are made on a solution that contains analyte at the LOD level, c_L, the instrument responses will be normally distributed about y_L, with the same standard deviation (estimated by s_B) as the blank. (This assumption of equal standard deviations was noted earlier in this review and has thus far been used throughout.) Half the measurements on a sample with concentration c_L will thus yield instrument responses below y_L. If any sample, yielding a signal less than y_L, is reported as containing no detectable analyte, the probability of a type II error is clearly 50%; this would always be true, irrespective of the separation of y_B and y_L.

Many workers, therefore, separately define a 'limit of decision', a point between y_B and y_L, to establish more sensible levels of type I and type II errors. This procedure is
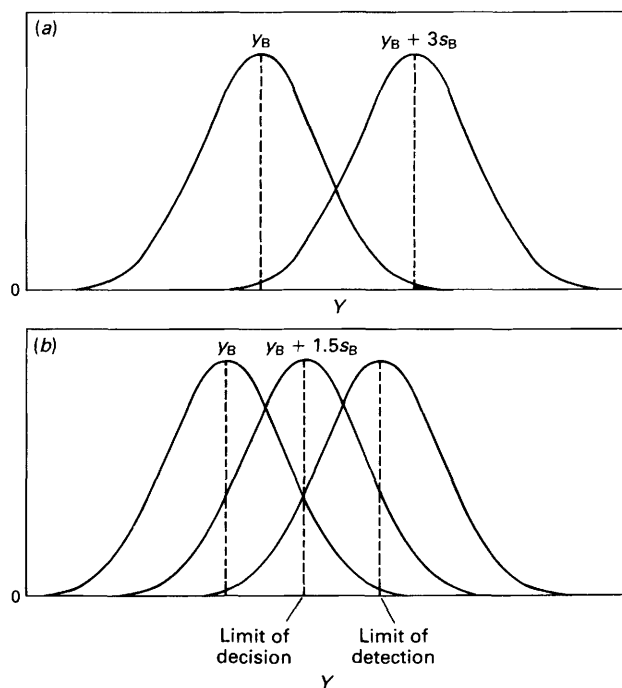

Fig. 4 Limits of detection; for details see text

again analogous to simple statistical tests; in this case the null hypothesis is that there is no analyte present, and the alternative hypothesis is that an analyte concentration c_L is present. The critical value for testing these hypotheses is set by the establishment of the limit of decision. In most cases the assumption is made that type I and type II errors are equally to be minimized (although it is easy to imagine practical instances where this is not appropriate), so the limit of decision would be at $y_B + 1.5s_B$ [Fig. 4(b)]. If analyte is reported present or absent at y-values, respectively, above and below this limit, there is a probability of 6.7% for each type of error. Many analysts feel this to be a reasonable criterion; it is not much different from the 95% confidence level routinely used in most statistical tests. Clearly, if the probability of both types of error is to be reduced to 0.135%, the limit of decision must be at $y_B + 3s_B$, and the LOD at $y_B + 6s_B$.[16]

Some workers further define a 'limit of determination', i.e., a concentration which can be determined with a particular RSD. Using the IUPAC definition of LOD, it is clear that the RSD at the limit of detection is 33.33%. A common definition of the limit of determination is $y_B + 10s_B$, indicating an RSD of 10%. It is to be noted that this result again assumes that the standard deviation s_B applies to measurements at all levels of y. The effects on LODs of departures from a uniform standard deviation have been considered by Liteanu and Rica[19] and by the Analytical Methods Committee.[15]

If the LOD definition of $y_B + 3s_B$ is accepted, it remains to discuss the estimation of y_B and s_B themselves. The blank signal, y_B might be obtained either as the average of several readings of a 'field blank' (i.e., a sample containing solvent, reagents and sample matrix but no analyte, and examined by the same protocol as all other samples), or by utilizing the intercept value, a, from the least squares calculation. If all the assumptions involved in the latter calculation are valid, these two methods should yield values for y_B that do not differ significantly. Repeated measurements on a field blank will also provide the value of s_B. Only if a field blank is unobtainable (a not uncommon situation) should the intercept, a, be used as the measure of y_B, in this situation $s_{y/x}$ will provide an estimate of s_B.

Intersection of Two Straight Lines

Analytical scientists frequently use methods requiring the determination of the point of intersection of two straight lines. This approach is used in Job's method[20] and in other studies of molecular interactions such as drug–protein binding. The usual requirement is to determine the x-value (often a concentration ratio rather than a single concentration in this case) of the point of intersection. If the two lines, each determined by the methods described above, are given by $y = b_1x + a_1$ and $y = b_2x + a_2$, the intersection point, x_x, is easily shown to be given by

$$x_x = (a_2 - a_1)/(b_1 - b_2) \qquad (14)$$

The confidence interval for x_x has been calculated in several ways (reviewed in reference 19), and continues to excite interest;[21] it is clearly related to the hyperbolic curves representing the confidence intervals for each line (Fig. 5). For the line $y = b_1x + a_1$ these curves are given by

$$y = (b_1x + a_1) \pm ts_{(y/x)1}\left[1 + \frac{1}{n_1} + \frac{(x - \bar{x}_1)^2}{\Sigma(x - \bar{x}_1)^2}\right]^{\frac{1}{2}} \qquad (15)$$

This equation yields the confidence limits for the true mean value of y at any given value of x. The t value is taken at the desired confidence level (usually 95%) and $n_1 - 2$ degrees of freedom. A similar equation applies to the line $y = b_2x + a_2$. One reasonable definition for the lower confidence limit for x_x, (x_L), is the abscissa value of the point of intersection of the upper confidence limit of line 1 and the lower confidence limit of line 2 (Fig. 5). At this point

$$b_1x_L + a_1 + ts_{(y/x)1}\left[1 + \frac{1}{n_1} + \frac{(x_L - \bar{x}_1)^2}{\Sigma(x - \bar{x}_1)^2}\right]^{\frac{1}{2}}$$
$$= b_2x_L + a_2 - ts_{(y/x)2}\left[1 + \frac{1}{n_2} + \frac{(x_L - \bar{x}_2)^2}{\Sigma(x - \bar{x}_2)^2}\right]^{\frac{1}{2}} \qquad (16)$$

which can be solved for x_L. An analogous equation can be written for x_u, which is similarly defined by the intersection of the lower confidence limit for line 1 and the upper confidence limit for line 2

$$b_1x_u + a_1 - ts_{(y/x)1}\left[1 + \frac{1}{n_1} + \frac{(x_u - \bar{x}_1)^2}{\Sigma(x - \bar{x}_1)^2}\right]^{\frac{1}{2}}$$
$$= b_2x_u + a_2 + ts_{(y/x)2}\left[1 + \frac{1}{n_2} + \frac{(x_u - \bar{x}_2)^2}{\Sigma(x - \bar{x}_2)^2}\right]^{\frac{1}{2}} \qquad (17)$$

As the confidence intervals for lines 1 and 2 may be of different width, and as the two lines may interesect at any angle, the confidence limits for x_x may not be symmetrical about x_x itself. It should also be noted that the confidence limits for x_x derived from (for example) the 95% confidence limits for the two separate lines are not necessarily the 95% confidence limits for x_x. As the estimation method used above assumes the worst case in combining the random errors of the two lines, the derived confidence limits are on the pessimistic (i.e., realistic!) side. Finally it is important to note that the practical applications of this method utilize extrapolations of the two straight lines to the intersection point. These extrapolations are generally short, and care is usually taken to perform the experiments in conditions where the extrapolations are believed to be valid. However, if this belief is erroneous (e.g., in studies of drug–protein binding where there is more than one class of binding site instead of the single class often assumed), even the best statistical methods cannot produce chemically valid results.

Residuals in Regression Statistics

Previous sections of this review have shown that the unweighted regression methods in common use in analytical chemistry are based on several assumptions which merit
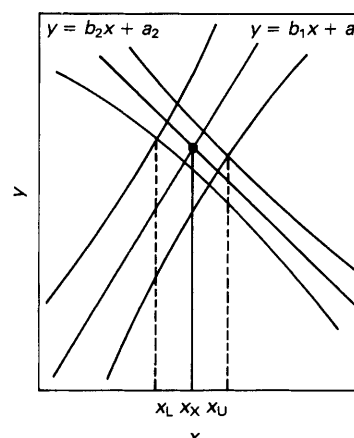

Fig. 5 Confidence limits for the point of intersection of two straight lines

critical examination and that it is not a straightforward matter to decide whether a straight line or a curve should be drawn through a set of calibration points. Important additional information on both these topics can be obtained from the y-residuals, the $(y - \hat{y})$ values which represent the differences between the experimental y-values and the fitted y-values. The residuals thus represent the random experimental errors in the measurements of y, if the statistical model used (the unweighted regression line of y on x) is correct. Many statistical tests can be applied to these residuals (a comprehensive survey is given in reference 5) but for routine work it is often sufficient to plot the individual residuals against $\hat{y}$ or against x. Many regression programs for personal computers offer this facility and some provide additional refinements, e.g., the inclusion of lines showing the standard deviations of the residuals.

It can be shown that, if the calibration line is calculated from the equation $y = bx + a$ (but not if it is forced through the origin by using the form $y = bx$), the residuals always total zero, allowing for rounding errors. As already noted, the residuals are assumed to be normally distributed. Fig. 6(a) shows the form that the residuals should thus take if the unweighted regression line is a good model for the experimental data. Fig. 6(b) and (c) indicates possible results if the unweighted regression line is inappropriate. If the residuals tend to become larger as y (or x) increases, the use of a weighted regression line (see below) is indicated, and if the residuals tend to fall on a curve, the use of a curved calibration graph rather than a linear one is desirable. In the latter case the signs (+ or −) of the residuals, which should be in random order if an appropriate statistical model has been used, will tend to occur in sequence ('runs'); in the example given, there is clearly a sequence of positive residuals, followed by a sequence of negative ones followed by a second positive sequence. The number of 'runs' (three in the example given) is thus significantly less than if the signs of the residuals had been + and − in random order. The Wald–Wolfowitz method tests for the significance of the number of runs in a set of data[5,22] by comparing the observed number of runs with tabulated data,[23] but it cannot be used if there are fewer than nine points in the calibration graph. Like the other residual diagnostic methods described here, the test is thus of restricted value in instrumental analysis, where the number of calibration points is frequently less than this. Tests on residuals are not, however, limited to linear regression plots: they can also be applied to non-linear plots, and indeed to any situation in which experimental data are fitted to a statistical model and some unexplained variations occur.

Examination of the residuals may shed light on a further problem, that of outliers among the data. The first part of this
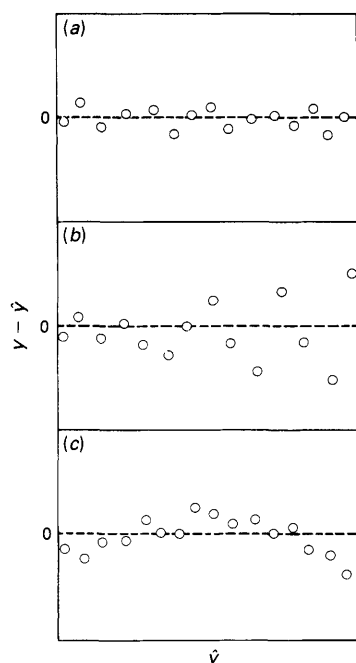
Fig. 6 Residuals in regression; for details see text

review[1] emphasized the importance of examining possible outliers carefully before rejecting them, if only because an observation that appears to be an outlier if one statistical model (*e.g.*, linear regression with normally distributed errors) is used, might not be an outlier if an alternative model (*e.g.*, a weighted or a polynomial regression equation) is fitted. After the residuals of a calibration graph have been calculated, it is usually easy to identify any that are exceptionally large. Again, many personal computer programs 'flag' such data points automatically. Unfortunately it is not legitimate simply to examine the residuals by the Q-test[1] or related outlier tests, as the residuals are not independent measurements (they must total zero). However, several methods have been developed for studying potential outliers in regression.[5,7] These methods transform the residuals before examination, and will not be treated in detail here. Perhaps the best-known approach involves the estimation of 'Cook's Distance'[24] for the suspect point. This distance is a measure of the influence of an observation, *i.e.*, of how much the regression line would be altered by omission of the observation from the usual calculations of *a* and *b*. A discussion of this method, along with a BASIC computer program which implements it, has recently been published.[25] The problem of outliers can alternatively be by-passed by the use of the robust and non-parametric methods described below.

Regression Techniques in the Comparison of Analytical Methods

When a novel analytical method is developed there is a natural desire to validate it by comparing it with well established methods. This is normally achieved by applying both new and established methods to the analysis of the same group of test samples. As calibration methods are designed for use over wide concentration ranges, these samples 'will properly contain widely differing amounts of the analyte under study. The question then arises, how are the paired results (*i.e.*, each sample examined by each of the two methods) evaluated for systematic errors? The paired *t*-test[1] cannot be used, as it ascribes the same weight to any given difference between a pair of results, irrespective of the absolute value measured.

The approach most commonly used is to plot the results of the two methods on the two axes of a regression graph; each point on the graph thus represents a single sample measured by the two techniques being compared. It is clear that, if both methods give identical results for all samples, the resulting graph will be a straight line of unit slope and zero intercept, with the correlation coefficient $r = +1$. Some departure from these idealized results is inevitable in practice, and the usual requirements if the new method is to be regarded as satisfactory are that r is close to $+1$, that the confidence interval for the intercept, a, includes zero, and that the confidence interval for the slope, b, includes 1. The new method is tested most rigorously if the comparison is made with a considerable number of samples covering in a roughly uniform way the concentration range of interest.

There are several ways in which the plotted line can deviate from the ideal characteristics summarized above. Sometimes one method will give results which are higher or lower than those of the other method by a constant amount (*i.e.*, $b = 1$, $a >$ or <0). In other cases there is a constant relative difference between the two methods ($a = 0$, $b >$ or < 1). These two types of error can occur simultaneously ($a >$ or < 0, $b >$ or < 1), and there are instances in which there is excellent agreement between the two methods over part of the range of interest, but disagreement at, *e.g.*, very high or very low concentrations. Finally, there are experiments where some of the points lie close to the ideal line ($b = 1$, $a = 0$), but another group of samples give widely divergent points; speciation problems are the most probable cause of this result.

These possibilities have been summarized by Thompson,[26] whose paper also studies an important problem in the use of conventional regression lines in method comparisons. The line of regression of y on x assumes that the random errors of x are zero. This is clearly not the case when two experimental methods are being compared, so despite its almost universal use in this context the conventional regression line is not a proper statistical tool for such comparisons. (The line of regression of x on y would be equally unsuitable.) It has, however, been shown that by using Monte Carlo simulation methods,[26] the consequences of this unsoundness are not serious provided that at least ten samples covering the concentration range of interest fairly uniformly are used in the comparison, and the results from the method with the smaller random errors are plotted on the x-axis. A rigorous solution of the method comparison problem would be a calculation of the best straight line through a series of points with both x and y values subject to random errors. Over a century ago, Adcock[27] offered a solution which assumed that the x- and y-direction errors were equal. A complete solution, based on maximum likelihood methods, has been proposed by Ripley and Thompson.[28] Their technique utilizes the statistical weight of each point (thus requiring more information—see below) and as it is a computer-based iterative approach, it may not command ready acceptance for routine use.

Robust and Non-parametric Regression Methods

Previous sections of this review have shown that the unweighted least-squares regression line of y on x may not be appropriate if the y-direction errors are not normally distributed, or if both x- and y-direction errors occur (as in method comparisons). Moreover, the calculation and interpretation of this line are complicated by the presence of possible outliers. Statisticians use the term robustness to describe the property of insensitivity to departures from an assumed statistical model; a number of robust regression methods have been developed in recent years. One obvious approach is to use non-parametric methods, which do not assume any particular distribution for the population of which the experimental data are a sample. (Note that robust methods are not necessarily non-parametric, but non-parametric methods are generally robust.)

Perhaps the best-known non-parametric regression method is that developed by Theil.[29] He suggested that the slope, b, of the regression line should be estimated by determining the median of the slopes of all the lines joining pairs of points. (The median of a set of measurements is the middle value when an odd number of measurements is arranged in numerical order, or the average of the two middle values when there is an even number of measurements.) A graph with n points will thus have $[n(n + 1)/2]$ independent estimates of the slope. After b has been determined, n estimates of the intercept, a, can be obtained from the equation $a_i = y_i - bx_i$. The median of these n values of a is taken as the value of the intercept estimate. Theil's procedure is open to two objections. Firstly, it seems that the slope estimates of points with well separated x_i values should carry more weight than those from neighbouring pairs of points; Jaeckel[30] has proposed a modified method that achieves this. Secondly, computation of the median slope value becomes tedious even for fairly moderate values of n; a graph with ten points will yield 55 separate estimates of the slope to be determined and sorted into numerical order. This problem is overcome by the use of a shorter technique (Theil's abbreviated or incomplete method) in which slope estimates are obtained from x_i and the first point above the median value of x, from x_2 and the second point above the median, and so on. (If n is odd, the middle x_i value is not used at all in the slope calculation.) After the slope of the line has been estimated in this way, the intercept is estimated as in the 'complete' method. An example of this approach, which also illustrates its robustness towards outliers, is given in reference 22. The Theil methods make no assumptions about the directions of the errors, so are suitable for method comparisons (see above). Hussain and Sprent[31] have shown that the Theil (complete) method is almost as efficient as the least-squares method when the errors are normally distributed, and much more efficient, especially when n is small, when the errors are not normally distributed. (Efficiency in statistics is a relative concept: it allows the comparison of two statistical tests in their ability to detect alternative hypotheses which are close to the null hypothesis, H_0.[32]) Maritz[33] has reviewed Theil's method and Sprent[32] and Maritz[34] have surveyed other robust and non-parametric regression methods, including those which handle curved graphs. It should be noted that the spline techniques of curve-fitting discussed below are non-parametric methods.

When a line is to be drawn through a large number of points (*e.g.*, in method comparisons), a relatively rapid and preliminary method of plotting it may be of value. The points are divided as nearly as possible into three equal groups of x-values, and the median x-value for each of the three groups is identified. These three data points are known as the summary points. The slope of the resistant line (*i.e.*, resistant to outliers) through all the points is then estimated by the slope of the line joining the two outermost summary points. The intercept of the line is calculated as the average of the three intercepts obtained from the determined slope and the three summary points (*cf.* Theil's method above). The values of the slope and intercept may be polished by iterative minimization of the y-residuals. This method (and many other 'quick and dirty' methods used in exploratory data analysis) is discussed in reference 35.

In some experiments the results cannot be expressed in quantitative terms, but only in terms of a rank order.[1] Examples—perhaps infrequent in analytical work—include the preferences of laboratory workers for different pieces of equipment, the state of health of laboratory animals or the taste quality of a food or drink sample. Relationships in such cases are studied using rank correlation methods. Spearman's rank correlation coefficient, ρ,[36] is famous as being the first statistical method to use ranks, and is readily shown[33] to be the product-moment correlation coefficient, r, converted for use when both x and y variables are expressed as ranks.

Spearman's ρ is given by

$$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2 - 1)} \tag{18}$$

and, like r, lies in the range $-1 \leqslant \rho \leqslant +1$. In equation (18) d_i is the difference between the x and y rankings for the ith measurement. If the calculated value of ρ exceeds the critical value (taken from tables) at the appropriate confidence level and value of n, a significant correlation between x and y is established (although not necessarily a causal relationship). As in other ranking methods, tied ranks (*i.e.*, observations of equal rank in x or y) are given mid-rank values. Thus, if several food samples were ranked, with the two best samples judged equal in quality, the ranking would be $1.5, 1.5, 3, 4, \ldots$ instead of $1, 2, 3, 4, \ldots$

The Kendall rank correlation coefficient, τ,[37] is based on a different idea. If, in a ranking experiment, high x-values are generally associated with high y-values, we expect that $y_j > y_i$ if $x_j > x_i$. Pairs of observations having this property are said to be concordant, while observations where the x and y values are in opposite order are said to be discordant (if $x_i = x_j$ or $y_i = y_j$ a tie has occurred). Kendall's method involves examining each of the $[n(n - 1)/2]$ pairs of data and evaluating n_c, the number of concordances, and n_d, the number of discordances. The rank correlation coefficient is then given by

$$\tau = \frac{n_c - n_d}{n(n - 1)/2} \tag{19}$$

Again, it is evident that $-1 \leqslant \tau \leqslant +1$, the value -1 corresponding to all the pairs of data points giving discordances, and $+1$ corresponding to complete concordance. Intermediate values are again compared with tabulated values. Kendall's method has the advantage that the data do not have to be converted into ranks for n_c and n_d to be calculated. Moreover, the computation can be further simplified[38] if the term $[n(n - 1)/2]$ is omitted, the test statistic being simply calculated as $T = n_c - n_d$. It is rare, however, for the results of the Kendall and Spearman methods to disagree, and both have been used successfully as tests for trend, *i.e.*, to examine whether there is a correlation when one of the variables is time. The concept of concordance introduced by Kendall can be extended to problems with more than two variables; examples are given in standard texts.[38]

Analysis of Variance in Linear Regression

Analysis of variance (ANOVA) is a powerful and very general method which separates the contributions to the over-all variation in a set of experimental data and tests their significance. The sources of variation (one of which is invariably the random measurement error) are each characterized by a sum of squares (SS), *i.e.*, the sum of a number of squared terms representing the variation in question, a number of degrees of freedom (DF) (as defined in reference 1), and a mean square, which is the former divided by the latter and can be used to test the significance of the variation contribution by means of the F-test.[1] The mean square and the number of degrees of freedom for the over-all variation are, respectively, the sums of the mean squares and degrees of freedom of the several contributing sources of variation: this additive property greatly simplifies the calculations, which are now widely available on personal computer software.

In analytical calibration experiments, variation in the y-direction only is considered. This variation is expressed as the sum of the squares of the distances of each calibration point from the mean y value, $\bar{y}$, *i.e.*, by $\Sigma(y_i - \bar{y})^2$. This is the total SS about $\bar{y}$. There are two contributions to this over-all variation. One is the SS due to regression, *i.e.*, that part of the variation due to the relationship between y and x; each term in this SS is clearly of the form $(\hat{y}_i - \bar{y})^2$. This SS has just one DF,

as just one function of the y_i values, *i.e.*, the slope, b, will calculate $\Sigma(\hat{y}_i - \bar{y})^2$ from $b^2 \Sigma(x_i - \bar{x})^2$. The second source of variation is the SS about regression, *i.e.*, the variation due to deviations from the calibration line, each term in the SS being of the form $(y_i - \hat{y}_i)^2$. This SS has $(n - 2)$ DF, reflecting the fact that the residuals come from a model requiring the estimation of two parameters, a and b. In accordance with the additivity principle described above, it is possible to show[5] that

Total SS about $\bar{y}$ = SS due to regression + SS about regression (20)

Moreover, the number of DF for the total SS is $(n - 2) + 1 = (n - 1)$. This result is expected, as only $(n - 1)$ y_i values are needed to determine the total SS about y, as $\Sigma(y_i - \bar{y}) = 0$ by definition. A typical ANOVA table for a linear regression plot is shown in Table 1. This is a one-way ANOVA calculation, there being only one source of variation in addition to the inevitable experimental error. The significance of the correlation can be tested by using the F-test, *i.e.*, by calculating

$$F_{1, (n-2)} = \text{MS}_{\text{reg}}/\text{MS}_{\text{res}} \quad (21)$$

In practice this is rarely necessary (though readily available in software packages), as the F-values are generally vastly greater than the critical values. A more common estimate of the goodness of fit is given by the statistic R^2, sometimes known as the (multiple) coefficient of determination or the (multiple) correlation coefficient. The prefix 'multiple' occurs because R^2 can also be used in curvilinear regression (see below). If the regression line (straight or curved) is to be a good fit to the experimental points, the SS due to regression should be a high proportion of the total SS about $\bar{y}$. This is expressed quantitatively using the equation

$$R^2 = \text{SS due to regression/total SS about } \bar{y} \quad (22)$$

R^2 clearly lies between 0 and 1 (although it can never reach 1 if there are multiple determinations of y_i at given x_i values[5]), however, it is often alternatively expressed as a percentage— the percentage of goodness of fit provided by a regression equation. It can be shown[5] that, for a straight line plot, $R^2 = r^2$, the square of the product-moment correlation coefficient. The application of R^2 to non-linear regression methods is considered further below.

Weighted Linear Regression Methods

The preceding discussion of regression methods has assumed that all the points on the regression line have equal weight (*i.e.*, equal importance) when the regression line is plotted. This is a reflection of the assumption that the y-direction errors are equal at all the values of x used in the calibration graph. In practice this is often an unrealistic assumption, as it is very common for the standard deviations of the measurements to alter with x. As already noted, RSDs will be high at analyte levels just above the LOD. However, there may be

other variations at much higher concentrations. In some cases the standard deviation is expected to rise in proportion to the concentration, *i.e.*, the RSD is approximately constant,[39] while in other cases the standard deviation rises, though less rapidly than the concentration.[40] Many attempts have been made to formulate rules and equations for this concentration related behaviour of the standard deviation for different methods.[41,42] In practice, however, it will frequently be better to rely on the analyst's experience of a particular method, instrument, *etc.* in this respect.

If experience suggests that the standard deviation of replicate measurements does indeed vary significantly with x (heteroscedastic data), a weighted regression line should be plotted. The equations for this line differ from equations (2)–(7) because a weighting factor, w_i, must be associated with each calibration point x_i, y_i. This factor is inversely proportional to the variance of y_i, s_i^2, and must either be estimated from a suitable model (see above), or determined directly from replicate measurements of y_i:

$$w_i = s_i^{-2}/(\sum_i s_i^{-2}/n) \quad (23)$$

Equation (23) conveniently scales the weighting factors so that their sum is equal to n, the number of x_i values. The slope and intercept of the weighted regression line are then given, respectively, by

$$b = \frac{\sum_i w_i x_i y_i - n\bar{x}_w\bar{y}_w}{\sum_i w_i x_i^2 - n(\bar{x}_w)^2} \quad (24)$$

$$a = \bar{y}_w - b\bar{x}_w \quad (25)$$

Both these equations use the coordinates of the weighted centroid, $(\bar{x}_w, \bar{y}_w)$, given by $\bar{x}_w = \sum_i w_i x_i/n$ and $\bar{y}_w = \sum_i w_i y_i/n$, respectively; the weighted regression line must pass through this point. The standard deviation, $s_{x_{0w}}$, and hence a confidence interval of a concentration estimated from a weighted regression line is given by

$$s_{x_{0w}} = \frac{s_{(y/x)_w}}{b}\left\{\frac{1}{w_0} + \frac{1}{n} + \frac{(y_0 - \bar{y}_w)^2}{b^2[\sum_i w_i y_i^2 - n(\bar{x}_w)^2]}\right\}^{\frac{1}{2}} \quad (26)$$

where w_0 is an interpolated weight appropriate to the experimental y_0 value, and $s_{(y/x)_w}$ is given by

$$s_{(y/x)_w}\left\{\frac{[\sum_i w_i y_i^2 - n(\bar{y}_w)^2] - b^2[\sum_i w_i x_i^2 - n(\bar{x}_w)^2]}{n-2}\right\}^{\frac{1}{2}} \quad (27)$$

The confidence limits for weighted regression lines have the general form shown in Fig. 7, with the weighted centroid closer to the origin than its unweighted counterpart.

Calculations of weighted regression lines are evidently more complex than unweighted regression computations, and only the more advanced computer software packages provide suitable programs. The slope and intercept of a weighted regression line are often very similar to those obtained when

Table 1 Anova table for linear regression

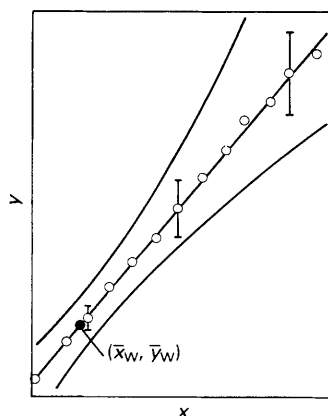| Source of variation	Degrees of freedom	Sum of squares	Mean square (MS)
Regression	1	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	
About regression (*i.e.*, residual)	$n - 2$	$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$s_{\hat{y}/x}^2 = \dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}$
Total	$n - 1$	$\sum_{i=1}^{n}(y_i - \bar{y})^2$	

Fig. 7 Confidence limits in weighted regression. The point ● is the weighted centroid $(\bar{x}_w, \bar{y}_w)$

unweighted calculations are applied to the same data, but only the weighted line provides proper estimates of standard deviations and confidence limits when the weights vary significantly with x_i. Weighted regression calculations must also be used when curvilinear data are converted into rectilinear data by a suitable algebraic transformation (see below).

The Analytical Methods Committee has recently suggested[3] the use of the weighted regression equations to test for the linearity of a calibration graph. The weighted residual sum of squares (i.e., the squares of the y-residuals, multiplied by their weights and then summed) should, if the plot is linear, have a chi-squared distribution with $(n - 2)$ DF. Significantly high values of chi squared thus suggest non-linearity. The same principle can be extended to test the fit of non-linear plots (see below).

Partly Straight, Partly Curved Calibration Plots

In many analytical methods a typical calibration plot is linear at low concentrations, but shows curvature towards the x-axis at higher analyte levels. This is often because an intrinsically non-linear relationship between instrument signal and concentration approximates to a linear function near the origin (e.g., fluorescence spectrometry). In such cases it would be logical to fit all the data, including the low-concentration points, to a curve. In practice, however, linear regression equations have been regarded as so much simpler to calculate than non-linear ones that the former are usually used over as wide a concentration range as possible. This gives rise to the question—to what upper concentration limit is the linear approximation satisfactory?

A simple exploratory approach to this problem is exemplified in reference 22. It involves calculating with equations (1)–(3) the correlation coefficient, slope, intercept and SS of the residuals, first for all the calibration points, and then for the data sets with the highest, next-highest, etc. points being omitted successively. If the highest point(s) lie on a significantly non-linear portion of the plot, omitting them from the calculations will produce large reductions in the SS of residuals, significant increases of r towards 1 (note again that absolute values are of little significance) and changes of a and b towards the values suggested by the calibration points near the origin. When the omission of further points produces only minor changes in r and the other parameters mentioned, then the linear portion of the graph has been successfully identified. During this stepwise removal of calibration points, a situation may arise where the omission of a point produces only slight improvements in the value of r etc., at the expense of a further restriction in the accepted linear range of the experiment.

Judgement must then be exercised to balance the advantages of an increased linear range against the possible loss of accuracy and precision.

Treatment of Non-linear Data by Transformations

A large number of analytical methods are known to produce non-linear calibration graphs. In some cases these curvilinear relationships are only important at higher analyte concentrations (see above), but in other cases (e.g., immunoassays) the entire plot is non-linear. Methods for fitting such data to curves are well established (see below) but the simplicity of linear methods has encouraged numerous workers to try to transform their data so that a rectilinear graph can be plotted. The most common transformations involve the use of logarithms (i.e., plotting y against ln x, or ln y against ln x) and exponentials. Less commonly used transformations include reciprocals, square roots and trigonometric functions. Two or more of these functions are sometimes used in combination, especially in calibration programs supplied with commercial analytical instruments. This topic has been surveyed by Bysouth and Tyson.[43] Draper and Smith[5] have reviewed such procedures at some length, and logical approaches to the best transformations have been surveyed by Box and Cox[44] and by Carroll and Ruppert.[45]

It is important to note that all such transformations may affect the relative magnitudes of the errors at different points on the plot. Thus a non-linear plot with approximately equal errors at all values of x (homoscedastic data) may be transformed into a linear plot with heteroscedastic errors. It can be shown[5,46] that if a function $y = f(x)$ with homoscedastic errors is transformed into the function $Y = BX + A$, then the weighting factors, w_i used in equations (24)–(27) above are calculated from

$$w_i = \left[1 / \left(\frac{dY_i}{dy_i} \right) \right]^2 \tag{28}$$

In some cases transformations make the use of a weighted regression plot less necessary. Thus a line of the form $y = bx$ with y-direction errors dependent on x may be subjected to a log–log transformation. The errors in log y then depend less markedly on log x and, therefore, a homoscedastic approach may be reasonable. Similarly, Kurtz et al.[47] have applied a series of power functions to transform chromatographic data to constant variance.

In some analytical methods data transformations are so common that specialist software may be available to perform the calculations and present the results graphically. This is particularly true in the field of competitive binding immunoassays, where several different (but related) transformations are in common use, including those involving the logit function $\{\text{logit } x = \ln[x/(1 - x)]\}$ and logistic functions such as $y = A/(B + Ce^{-Dx})$, with the A–D values to be found by iteration. (Immunoassay practitioners also use spline functions—see below.) These methods have been reviewed.[48]

Curvilinear Regression

In many analytical methods, non-linear regression plots arise from several experimental factors, making it impossible to predict a model equation for the curve. For example in molecular fluorimetry, theory[49] shows that a plot of fluorescence intensity against concentration should be linear—but only as the result of several assumptions about the optical system used and the sample under study, and with the aid of a mathematical approximation. In practice some or all of these assumptions will fail, but to an unpredictable extent, giving a calibration graph that may approximate to a straight line near the origin (see above) but is, in reality, a curve throughout. In such cases it is sensible to adopt an empirical fit of a curve to

the observed data. This is most commonly attempted by using a polynomial equation of the form

$$y = a + bx + cx^2 + dx^3 + \ldots \qquad (29)$$

The advantage of this type of equation is that, after the number of terms has been established, matrix manipulation allows an exact solution for a, b, c, etc. if the least-squares fitting criterion (see above) is used. Most computer packages offer such polynomial curve-fitting programs, so, in practice the major problem for the experimental scientist is to decide on the appropriate number of terms to be used in equation (29). The number of terms must clearly be $<n$ for the equation to have any physical meaning and common sense suggests that the least number of terms providing a satisfactory fit to the data should always be used (quadratic or cubic fits are frequently excellent).

Several approaches to this problem are available, the simplest (though probably not the best) being the use of the coefficient of determination, R^2. As described above, this coefficient expresses the extent to which the total SS about $\bar{y}$ can be explained by the regression equation under scrutiny. Values of R^2 close to 1 (or 100%—computer packages often present the result in this form) are thus apparently indicative of a good fit between the chosen equation and the experimental data. In practice we would thus examine in turn the values of R^2 obtained for the quadratic, cubic, quartic, etc. fits, and then make a judgement on the most appropriate polynomial.

This method is open to two objections. The first is that, like the related correlation coefficient, r (see above), R^2 can take very high values ($\gg 0.9$) even when visual inspection shows that the fit is obviously indifferent. More seriously still, it may be shown that R^2 always increases as successive terms are added to the polynomial equation, even if the latter are of no real value. (Draper and Smith[5] point out that this presents particular dangers if the data are grouped, i.e., if we have several y-values at each of only a few x-values. The number of terms in the polynomial must then be less than the number of x-values.) Thus, if this method is to be used, it is essential to attach little importance to absolute R^2 values and to continue adding terms to the polynomial only if this leads to substantial increases in R^2.

An alternative and probably more satisfactory curve-fitting criterion is the use of the 'adjusted R^2' statistic, given by[50]

$$R^2 \text{ (adjusted)} = 1 - \text{(residual MS/total MS)} \qquad (30)$$

The use of the mean square (MS) instead of SS terms, allows the number of degrees of freedom $(n - p)$, and hence the number of fitted parameters (p), to be taken into account. For any given data set and polynomial function, adjusted R^2 always has a lower value than R^2 itself; many computer packages provide both calculations.

Among several other methods for establishing the best polynomial fit[5,7] the simple use of the F-test[1] has much to commend it. In this application (often referred to as a partial F-test), F is used to test the null hypothesis that the addition of an extra polynomial term to equation (29) does not significantly improve the goodness of fit of the curve, when compared with the curve obtained without the extra term. Thus

$$F = \frac{(\text{extra SS due to adding } x^n \text{ term})/1}{\text{residual MS for } n\text{-order model}} \qquad (31)$$

The calculated F value is compared with the tabulated value of $F_{1, n-p}$ at the desired probability level; p is again the number of parameters to be determined, e.g., 3 (a, b, c), in a test to see whether a quadratic term is desirable. This test is simplified if the ANOVA calculation breaks down the SS due to regression into its component parts, i.e., the SS due to the term in x, the additional SS due to the term in x^2 etc.

It is worth noting that, if the form of a curvilinear plot is known, the calculation of the parameter values can be regarded as an optimization problem and can be tackled by methods such as simplex optimization.[51] This approach offers no particular benefit in cases where exact solutions are also available by matrix manipulation, such as the polynomial equation (29), but may be very advantageous in other cases where exact solutions are not accessible. Again commercially available software is plentiful. A recent paper[52] describes the calculation of confidence limits for calibration lines calculated using the simplex method.

Whichever model and calculation method is chosen to plot a calibration graph, it is desirable to examine the residuals generated and use them to study the validity of the chosen model. If the latter is suitable the residuals should show no marked trend in sign, spread or value when plotted against corresponding x or y values. As for linear regression, outlying points can also be studied.

Spline Functions and Other Robust Non-linear Regression Methods

As non-linear calibration plots often arise from a combination of physico-chemical effects (see above) and failure of mathematical approximations etc., it is perhaps unrealistic to expect that any single curve will adequately describe such data. It may thus be better to plot a curve which is a continuous series of shorter curved portions. The most popular approach of this type is the cubic spline method, which seeks to fit the experimental data with a series of curved portions each of cubic form, $y = a + bx + cx^2 + dx^3$. These portions are connected at points called 'knots', at each of which the two linked cubic functions and their first two derivatives must be continuous. In practice, the knots may coincide with experimental calibration points, but this is not essential and a variety of approaches to the selection of the number and positions of the knots is available. Spline function calculations are provided by several software packages, and their application to analytical problems has been reviewed by Wold[53] and by Wegscheider.[54]

A group of additional regression methods now attracting considerable attention relies on the use of fitting criteria other than or in addition to the least squares approach. In particular the 'reweighted least squares' method described in detail by Rousseeuw and Leroy[55] utilizes the least median of squares criterion (i.e., minimization of the median of the squared residuals) to identify large residuals. The least squares curve is then fitted to the remaining points. These modern robust methods can of course be applied to straight line graphs as well as curves, and despite their requirement for more advanced computer programs they are already attracting the attention of analytical scientists (e.g., reference 56). Such developments provide timely reminders that the apparently simple task of fitting a straight line or a curve to a set of analytical data is still provoking much original research.

References

1   Miller, J. C., and Miller, J. N., Analyst, 1988, **113**, 1351.
2   Martens, H., and Naes, T., Multivariate Calibration, Wiley, Chichester, 1989.
3   Analytical Methods Committee, Analyst, 1988, **113**, 1469.
4   Růžička, J., and Hansen, E. H., Flow Injection Analysis, Wiley, New York, 2nd edn., 1988.
5   Draper, N. R., and Smith, H., Applied Regression Analysis, Wiley, New York, 2nd edn., 1981.
6   Edwards, A. L., An Introduction to Linear Regression and Correlation, Freeman, New York, 2nd edn., 1984.
7   Kleinbaum, D. G., Kupper, L. L., and Muller, K. E., Applied Regression Analysis and Other Multivariable Methods, PWS-Kent, Boston, 2nd edn., 1988.

8 Working, H., and Hotelling, H., *J. Am. Stat. Assoc. (Suppl.)*, 1929, **24**, 73.
9 Hunter, J. S., *J. Assoc. Off. Anal. Chem.*, 1981, **64**, 574.
10 Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., and Kaufman, L., *Chemometrics: A Textbook*, Elsevier, Amsterdam, 1988.
11 Cardone, M. J., *Anal. Chem.*, 1986, **58**, 433.
12 Cardone, M. J., *Anal. Chem.*, 1986, **58**, 438.
13 Jochum, C., Jochum, P., and Kowalski, B. R., *Anal. Chem.*, 1981, **53**, 85.
14 Brereton, R. G., *Analyst*, 1987, **112**, 1635.
15 Analytical Methods Committee, *Analyst*, 1987, **112**, 199.
16 Specker, H., *Angew. Chem. Int. Ed. Engl.*, 1968, **7**, 252.
17 Kaiser, H., *The Limit of Detection of a Complete Analytical Procedure*, Adam Hilger, London, 1968.
18 IUPAC, Nomenclature, Symbols, Units and Their Usage in Spectrochemical Analysis-II, *Spectrochim. Acta, Part B*, 1978, **33**, 242.
19 Liteanu, C., and Rica, I., *Statistical Theory and Methodology of Trace Analysis*, Ellis Horwood, Chichester, 1980.
20 Hadjiioannou, T. P., Christian, G. D., Efstathiou, C. E., and Nikolelis, D. P., *Problem Solving in Analytical Chemistry*, Pergamon Press, Oxford, 1988.
21 Jandera, P., Kolda, S., and Kotrly, S., *Talanta*, 1970, **17**, 443.
22 Miller, J. C., and Miller, J. N., *Statistics for Analytical Chemistry*, Ellis Horwood, Chichester, 2nd edn., 1988.
23 Swed, F. S., and Eisenhart, C., *Ann. Math. Stat.*, 1943, **14**, 66.
24 Cook, R. D., *Technometrics*, 1977, **19**, 15.
25 Rius, F. X., Smeyers-Verbeke, J., and Massart, D. L., *Trends Anal. Chem.*, 1989, **8**, 8.
26 Thompson, M., *Analyst*, 1982, **107**, 1169.
27 Adcock, R. J., *Analyst (Des Moines, Iowa, USA)*, 1878, **5**, 53.
28 Ripley, B. D., and Thompson, M., *Analyst*, 1987, **112**, 377.
29 Theil, H., *Proc. K. Ned. Akad. Wet., Part A*, 1950, **53**, 386.
30 Jaeckel, L. A., *Ann. Math. Stat.*, 1972, **43**, 1449.
31 Hussain, S. S., and Sprent, P., *J. R. Stat. Soc., Part A*, 1983, **146**, 182.
32 Sprent, P., *Applied Nonparametric Statistical Methods*, Chapman and Hall, London, 1989.
33 Maritz, J. S., *Aust. J. Stat.*, 1979, **21**, 30.
34 Maritz, J. S., *Distribution-Free Statistical Methods*, Chapman and Hall, London, 1981.
35 Velleman, P. F., and Hoaglin, D. C., *Applications, Basics and Computing of Exploratory Data Analysis*, Duxbury Press, Boston, 1981.
36 Spearman, C., *Am. J. Psychol.*, 1904, **15**, 72.
37 Kendall, M. G., *Biometrika*, 1938, **30**, 81.
38 Conover, W. J., *Practical Nonparametric Statistics*, Wiley, New York, 2nd edn., 1980.
39 Francke, J.-P., de Zeeuw, R. A., and Hakkert, R., *Anal. Chem.*, 1978, **50**, 1374.
40 Garden, J. S., Mitchell, D. G., and Mills, W. N., *Anal. Chem.*, 1980, **52**, 2310.
41 Hughes, H., and Hurley, P. W., *Analyst*, 1987, **112**, 1445.
42 Thompson, M., *Analyst*, 1988, **113**, 1579.
43 Bysouth, S. R., and Tyson, J. F., *J. Anal. At. Spectrom.*, 1986, **1**, 85.
44 Box, G. E. P., and Cox, D. R., *J. Am. Stat. Assoc.*, 1984, **79**, 209.
45 Carroll, R. J., and Ruppert, D., *J. Am. Stat. Assoc.*, 1984, **79**, 321.
46 Thompson, M., and Howarth, R. J., *Analyst*, 1980, **105**, 1188.
47 Kurtz, D. A., Rosenberger, J. L., and Tamayo, G. J., in *Trace Residue Analysis*, ed. Kurtz, D. A., *American Chemical Society Symposium Series No.* 284, American Chemical Society, Washington, DC, 1985.
48 Wood, W. G., and Sokolowski, G., *Radioimmunoassay in Theory and Practice, A Handbook for Laboratory Personnel*, Schnetztor–Verlag, Konstanz, 1981.
49 Miller, J. N., in *Standards in Fluorescence Spectrometry*, ed. Miller, J. N., Chapman and Hall, London, 1981.
50 Chatfield, C., *Problem Solving: A Statistician's Guide*, Chapman and Hall, London, 1988.
51 Morgan, S. L., and Deming, S. N., *Anal. Chem.*, 1974, **46**, 1170.
52 Phillips, G. R., and Eyring, E. M., *Anal. Chem.*, 1988, **60**, 738.
53 Wold, S., *Technometrics*, 1974, **16**, 1.
54 Wegscheider, W., in *Trace Residue Analysis*, ed. Kurtz, D. A., *American Chemical Society Symposium Series No.* 284, American Chemical Society, Washington, DC, 1985.
55 Rousseeuw, P. J., and Leroy, A. M., *Robust Regression and Outlier Detection*, Wiley, New York, 1987.
56 Phillips, G. R., and Eyring, E. M., *Anal. Chem.*, 1983, **55**, 1134.