Unified equations for the slope, intercept, and standard errors of the best straight line

Derek York, Norman M. Evensen, Margarita López Martínez, and Jonás De Basabe Delgado

Citation: Am. J. Phys. **72**, 367 (2004); doi: 10.1119/1.1632486

View online: http://dx.doi.org/10.1119/1.1632486

View Table of Contents: http://ajp.aapt.org/resource/1/AJPIAS/v72/i3

Published by the American Association of Physics Teachers

Unified equations for the slope, intercept, and standard errors of the best straight line

Derek York[a] and Norman M. Evensen
Department of Physics, University of Toronto, 60 Saint George St., Toronto, Ontario M5S 1A7, Canada

Margarita López Martínez and Jonás De Basabe Delgado[b]
Departamento de Geología, CICESE, km 107 Carr. Tijuana-Ensenada Ensenada, Baja California, México 22860

(Received 22 January 2003; accepted 17 October 2003)

It has long been recognized that the least-squares estimation method of fitting the best straight line to data points having normally distributed errors yields identical results for the slope and intercept of the line as does the method of maximum likelihood estimation. We show that, contrary to previous understanding, these two methods also give identical results for the standard errors in slope and intercept, provided that the least-squares estimation expressions are evaluated at the least-squares-adjusted points rather than at the observed points as has been done traditionally. This unification of standard errors holds when both x and y observations are subject to correlated errors that vary from point to point. All known correct regression solutions in the literature, including various special cases, can be derived from the original York equations. We present a compact set of equations for the slope, intercept, and newly unified standard errors.

[DOI: 10.1119/1.1632486]

I. INTRODUCTION

Despite its seeming simplicity, the problem of finding the best straight line through an experimentally determined set of points on the $x-y$ plane has triggered the publication of hundreds of papers in a host of fields since Gauss' original development of the method of least-squares estimation (LSE). This situation led Press *et al.*[1] to comment that ''Be aware that the literature on the seemingly straightforward subject of this section [Straight-Line Data with Errors in Both Coordinates] is generally confusing and sometimes plain wrong ... York[2] and Reed[3] usefully discuss the simple case of a straight line as treated here ...'' (p. 664).

Although the traditional regression of y on x is widely used (and included in most hand calculators and spreadsheets), only rarely are the x values actually error-free. In reality, both x and y errors may be significant, and may vary from point to point. In addition, the errors in the two coordinates may be highly correlated. For instance, in radiometric age determination, use is universally made of isotope ratio plots in which a common normalizing isotope forms the denominator of both the x and y coordinates. This practice imposes an often very significant correlation between the errors in x and y.

York[2] gave the general solutions for the LSE best straight line in terms of its slope and y intercept and errors in these two parameters, when both observables (X_i, Y_i) are subject to errors which vary from point to point. In Ref. 4, a more general solution also allowed for the possible correlation between the x and y errors at each point. In both papers, it was pointed out that many previously published solutions (for example, those of Refs. 5–7) corresponded to differing ways of assigning weights $[\omega(X_i),\ \omega(Y_i)]$ or error correlations r_i to the observed data points (X_i, Y_i). Rarely were the published special solutions the correct ones to use in typical experimental situations! The widespread adoption of York's LSE algorithm[2] led to a great improvement in this situation in some fields. However the more general treatment in York,[4] published in an earth science journal, has escaped the attention of many statisticians and others writing on least-squares fitting of straight lines, including Ref. 1 and Reed.[3,8]

Although the method of LSE has remained in widespread use among experimental scientists for solving statistical problems, members of the statistical community have shifted to the use of the maximum likelihood estimation (MLE) approach. Titterington and Halliday[9] recommended its use for straight line fitting. They emphasized that, if all the data errors are normally distributed and all the (X_i, Y_i) pairs are independent, maximizing the likelihood function was exactly the same as minimizing the weighted sum of squares in LSE. Therefore, MLE and LSE agree in their estimates of the best slope and intercept in this case. However, they pointed out that MLE is ''... a general theory that allows us to obtain approximate variances and covariances for the parameter estimates... It is not possible to say in general which is best, but the [MLE] method... seems to be simpler for this problem, and with small samples there is probably not much to choose between the methods numerically. The previous methods [such as those of York, 1969] were based on Taylor expansions (Deming, 1966) and it is noticeable that MLE theory has tended to supersede this approach in the statistical literature on this problem.''

Titterington and Halliday[9] carried out numerical comparisons of the MLE and the LSE results[4] for the standard errors of the slopes and intercepts of a number of data sets. They concluded that ''In most cases the standard errors from the two methods used here are very similar...'' (p. 189).

Our purpose is to show the exact relationship between these two ''alternative'' methods of error calculation in straight-line regression when normal errors are assumed, and when it is presumed that, were it not for random experimental errors, the observed points would have been perfectly collinear. We show that despite the apparently very different underlying analytical approaches (MLE based on second-order derivatives of the likelihood function, LSE based on the first-order derivatives of the slope and intercept), the

MLE expressions obtained in Ref. 9 for the standard errors of the slope and intercept are algebraically identical with the York[4] LSE standard error estimates when the latter are evaluated, *not* at the observed points (X_i, Y_i) as is usually done in least squares, but at the least-squares-adjusted points (x_i, y_i).

The substitution of the observed points in the LSE error formulas probably came about because of the additional computational work involved in calculating the adjusted points in the pre-computer era. As early as 1943, Deming[10] recognized that the errors should ideally be calculated at the adjusted points. The fitting process assumes that the true data points would lie on a single straight line, so that consistency demands the use of the adjusted points, which satisfy this criterion, rather than the observed points, which are scattered about the fitted line. It is now no problem to evaluate the errors at the adjusted points as they should be. However, even those authors who have occasionally evaluated their LSE solutions using the adjusted points[8] were apparently unaware that using the adjusted points made the least-squares solutions for the errors in the intercept and slope numerically identical to the standard MLE solutions. Conversely, the practitioners of MLE did not realize that they were identically reproducing not merely the LSE values of the slope and the intercept, but also the adjusted-point LSE standard errors of these parameters. We can thus say that the least-squares and maximum-likelihood methods for the fitting of a straight line have finally been unified for the calculation of all four parameters a, b, σ_a, and σ_b.

II. THE UNIFICATION

The notation used in this paper is summarized in Table I. In particular, we adopt the notation of $\tilde{\sigma}$ for the standard error as calculated using the MLE method and σ as calculated using the LSE method of York.[4] Titterington and Halliday[9] derived the following expressions for $\tilde{\sigma}_a^2$ and $\tilde{\sigma}_b^2$ in terms of the expectation values x_i, of the observables X_i:

$$\tilde{\sigma}_a^2 = \frac{\Sigma W_i x_i^2}{(\Sigma W_i x_i^2)(\Sigma W_i) - (\Sigma W_i x_i)^2}, \tag{1a}$$

$$\tilde{\sigma}_b^2 = \frac{\Sigma W_i}{(\Sigma W_i x_i^2)(\Sigma W_i) - (\Sigma W_i x_i)^2}. \tag{1b}$$

The expectation values (x_i and y_i) are identical to the LSE-adjusted values of X_i and Y_i, and indeed Titterington and Halliday noted that their formulas for the expectation values x_i and y_i were identical to York's formulas for the adjusted values.

The unification proceeds by simplifying Eq. (1) for $\tilde{\sigma}_a^2$ and $\tilde{\sigma}_b^2$. We then show that the more complex LSE expressions for these variances given in York[4] when evaluated at the least-squares-adjusted points reduce exactly to these simplified versions of the variances in Ref. 9.

We simplify the MLE variances by introducing the quantity u_i, defined in Table I. By transforming to the u_i from the x_i (see Appendix A), Eqs. (1a) and (1b) become

$$\tilde{\sigma}_a^2 = \frac{1}{\Sigma W_i} + \bar{x}^2 \tilde{\sigma}_b^2, \tag{2a}$$

$$\tilde{\sigma}_b^2 = \frac{1}{\Sigma W_i u_i^2}. \tag{2b}$$

Table I. Summary of our notation. Note that although the expression for S, the weighted sum of squared residuals, appears identical to that for a standard weighted regression of y on x, the weight W_i actually involves the weights in both x and y, as well as the correlations between the x and y errors.

Symbol	**Meaning**
a, b	y intercept and slope of best line, $y=a+bx$
σ_a, σ_b	Standard errors of a and b
X_i, Y_i	Observed data points
x_i, y_i	Least-squares-adjusted points, expectation values of X_i, Y_i
$\sigma_a(X_i, Y_i)$, $\sigma_b(X_i, Y_i)$	LSE standard errors evaluated at the observed points (X_i, Y_i)
$\sigma_a(x_i, y_i)$, $\sigma_b(x_i, y_i)$	LSE standard errors evaluated at the adjusted points (x_i, y_i)
$\tilde{\sigma}_a$, $\tilde{\sigma}_b$	MLE standard errors
$\omega(X_i)$, $\omega(Y_i)$	Weights of X_i, Y_i
α_i	$\sqrt{\omega(X_i)\omega(Y_i)}$
r_i	Correlation coefficient between errors in X_i and Y_i
W_i	$\dfrac{\omega(X_i)\omega(Y_i)}{\omega(X_i)+b^2\omega(Y_i)-2br_i\alpha_i}$
$\bar{X}$	$\dfrac{\Sigma W_i X_i}{\Sigma W_i}$
$\bar{Y}$	$\dfrac{\Sigma W_i Y_i}{\Sigma W_i}$
U_i	$X_i-\bar{X}$
V_i	$Y_i-\bar{Y}$
$\bar{x}$	$\dfrac{\Sigma W_i x_i}{\Sigma W_i}$
$\bar{y}$	$\dfrac{\Sigma W_i y_i}{\Sigma W_i}$
u_i	$x_i-\bar{x}$
v_i	$y_i-\bar{y}$
β_i	$W_i\left[\dfrac{U_i}{\omega(Y_i)}+\dfrac{bV_i}{\omega(X_i)}-(bU_i+V_i)\dfrac{r_i}{\alpha_i}\right]$
$\bar{\beta}$	$\dfrac{\Sigma W_i \beta_i}{\Sigma W_i}$
S	$\Sigma W_i(Y_i-bX_i-a)^2$

Turning now to the LSE errors of York,[4] σ_b^2 was expressed as

$$\sigma_b^2(X_i, Y_i) = \sum \left[\left(\frac{\partial\varphi}{\partial X_i}\right)^2 \frac{1}{\omega(X_i)} + \left(\frac{\partial\varphi}{\partial Y_i}\right)^2 \frac{1}{\omega(Y_i)} + \frac{2r_i}{\alpha_i}\frac{\partial\varphi}{\partial X_i}\frac{\partial\varphi}{\partial Y_i}\right] \Big/ \left(\frac{\partial\varphi}{\partial b}\right)^2, \tag{3}$$

where φ is the left-hand side of the least-squares cubic, quadratic, or linear equations (see Appendix B). From the expressions in York[4] for $\partial\varphi/\partial X_i$, $\partial\varphi/\partial Y_i$, and $\partial\varphi/\partial b$ it can be shown that

$$\frac{\partial\varphi}{\partial X_i} = W_i[2b(\beta_i - \bar{\beta}) - V_i], \tag{4a}$$

$$\frac{\partial\varphi}{\partial Y_i} = W_i[U_i - 2(\beta_i - \bar{\beta})], \tag{4b}$$

$$\frac{\partial\varphi}{\partial b}=\frac{1}{b}\sum W_iU_iV_i+4\sum W_i(\beta_i-U_i)(\beta_i-\overline{\beta})$$
$$-\frac{1}{b}\sum W_i^2\frac{r_i}{\alpha_i}(bU_i-V_i)^2. \tag{4c}$$

The substitution of these expressions for the three partial derivatives into Eq. (3) yields

$$\sigma_b^2(X_i,Y_i)=\frac{\Sigma W_i^2\left[\frac{U_i^2}{\omega(Y_i)}+\frac{V_i^2}{\omega(X_i)}-\frac{2r_i}{\alpha_i}U_iV_i\right]}{D^2}, \tag{5}$$

where

$$D=\frac{1}{b}\sum W_iU_iV_i+4\sum W_i(\beta_i-U_i)(\beta_i-\overline{\beta})$$
$$-\frac{1}{b}\sum W_i^2\frac{r_i}{\alpha_i}(bU_i-V_i)^2. \tag{6}$$

The expression for σ_b^2 in Eq. (5) was given in Ref. 11, except for the omission of the third term on the right-hand side of D, which we presume to be a typographical error. Traditionally, in least-squares fitting, Eq. (5) for σ_b^2 would be evaluated by inserting the corresponding values of the observables $(X_i,\ Y_i)$ into it. Instead, let us now evaluate it at the least-squares-adjusted points x_i and y_i. In Appendix C we show that when we substitute x_i for X_i and y_i for Y_i, and therefore also substitute $u_i=x_i-\overline{x}$ for $U_i=X_i-\overline{X}$ and $v_i=y_i-\overline{y}$ for $V_i=Y_i-\overline{Y}$, the numerator of σ_b^2 in Eq. (5) becomes $\Sigma W_iu_i^2$, and the denominator becomes $(\Sigma W_iu_i^2)^2$, so that the LSE expression for σ_b^2 evaluated at the LSE adjusted points is

$$\sigma_b^2(x_i,y_i)=\frac{\Sigma W_iu_i^2}{(\Sigma W_iu_i^2)^2}=\frac{1}{\Sigma W_iu_i^2}=\tilde{\sigma}_b^2. \tag{7}$$

York's[4] LSE expression for σ_a^2 is

$$\sigma_a^2(X_i,Y_i)=\sum\left[\left(\frac{\partial a}{\partial X_i}\right)^2\frac{1}{\omega(X_i)}+\left(\frac{\partial a}{\partial Y_i}\right)^2\frac{1}{\omega(Y_i)}\right.$$
$$\left.+\frac{2r_i}{\alpha_i}\frac{\partial a}{\partial X_i}\frac{\partial a}{\partial Y_i}\right], \tag{8}$$

where

$$\frac{\partial a}{\partial X_i}=-\frac{bW_i}{\Sigma W_i}+(\overline{X}+2\overline{\beta})\left[\left(\frac{\partial\varphi}{\partial X_i}\right)\Big/\left(\frac{\partial\varphi}{\partial b}\right)\right], \tag{9}$$

and

$$\frac{\partial a}{\partial Y_i}=\frac{W_i}{\Sigma W_i}+(\overline{X}+2\overline{\beta})\left[\left(\frac{\partial\varphi}{\partial Y_i}\right)\Big/\left(\frac{\partial\varphi}{\partial b}\right)\right]. \tag{10}$$

If we substitute the expressions for $\partial\varphi/\partial X_i$, $\partial\varphi/\partial Y_i$, and $\partial\varphi/\partial b$ in Eq. (4) into the expressions for $\partial a/\partial X_i$ and $\partial a/\partial Y_i$, Eqs. (9) and (10), we can then substitute the resulting $\partial a/\partial X_i$ and $\partial a/\partial Y_i$ into Eq. (8) to obtain the following result for σ_a^2:

$$\sigma_a^2(X_i,Y_i)=\frac{1}{\Sigma W_i}+(\overline{X}+2\overline{\beta})^2\sigma_b^2(X_i,Y_i)$$
$$+\frac{2(\overline{X}+2\overline{\beta})\overline{\beta}}{D}, \tag{11}$$

in agreement with Ref. 11 if their D is corrected for the missing term. To evaluate σ_a^2 at the adjusted points, we note that in this case $\overline{\beta}=0$ (see Appendix C), but $D\neq 0$, $\overline{X}$ obviously transforms to $\overline{x}$ and $\sigma_b^2(x_i,y_i)$ becomes $\tilde{\sigma}_b^2$. Thus we see immediately from Eq. (11) that

$$\sigma_a^2(x_i,y_i)=\frac{1}{\Sigma W_i}+\overline{x}^2\tilde{\sigma}_b^2=\tilde{\sigma}_a^2. \tag{12}$$

Thus Eqs. (7) and (12) yield the new unification theorem: *If the least-squares estimates of York (1969) of the errors in slope and intercept of the best straight line are evaluated at the least-squares-adjusted points instead of at the observed points, the least-squares errors become identical to the maximum-likelihood errors*.

It also follows simply that the covariance of the slope and intercept, $\mathrm{cov}(a,b)$, calculated by traditional LSE, becomes identical with the MLE estimate of this covariance, when evaluated at the adjusted points. Thus in both cases, $\mathrm{cov}(a,b)=-\overline{x}\tilde{\sigma}_b^2$, and the correlation coefficient of a with b is $r_{ab}=-\overline{x}\tilde{\sigma}_b/\tilde{\sigma}_a=-\overline{x}/\sqrt{\overline{x^2}}$.

Titterington and Halliday[9] found slight numerical differences between the York[4] solutions for $\sigma_a^2(X_i,Y_i)$ and $\sigma_b^2(X_i,Y_i)$ and their own detailed MLE results. We now see that this difference is due entirely to the York LSE algorithm following the traditional route of evaluating the expressions for these parameters at the observed points rather than the adjusted points. These slight differences are simply reflections of the slight differences between the observed points (X_i,Y_i) and the adjusted points (x_i,y_i). Clearly such minor differences between the LSE and MLE values of σ_a^2 and σ_b^2 would be expected to increase somewhat as data points which are more scattered about a straight line are fitted, because greater differences would then exist between the observed points (X_i,Y_i) and the adjusted points (x_i,y_i).

Although the LSE (evaluated at the observed points) and MLE error estimates will generally differ slightly, there are two very significant exceptions: the cases of the regression (weighted if desired) of y on x and the regression (weighted if desired) of x on y. In each of these regressions, we find that the LSE (evaluated at the observed points) and MLE methods automatically yield identical solutions for σ_a^2 and σ_b^2, regardless of the scatter of the observed points about the best line. This apparent paradox is beautifully resolved when we note that in, say, the case of the regression of y on x, the LSE expressions for $\sigma_a^2(X_i,Y_i)$ and $\sigma_b^2(X_i,Y_i)$ reduce exactly to functions only of the X_i (which are perfectly accurate observables in this particular regression and therefore equal to the x_i by definition), so that the σ_a^2 and σ_b^2 are simultaneously evaluated at the observable and the adjusted abscissae, and their identical LSE and MLE values are thus found in one calculation (see Appendix D). By symmetry, the equivalent explanation applies to (optionally weighted) regression of x on y.

Whether the above unification of the LSE errors and MLE errors [in Eqs. (7) and (12)] applies to more general cases than straight-line fitting in two dimensions, we can only conjecture. But in any case, this unification of the LSE and MLE errors now obviates any necessity for choosing between these two estimates of error. Henceforth we shall simply use σ_a and σ_b to denote these unified error estimates, where $\sigma_a = \sigma_a(x_i, y_i) = \tilde{\sigma}_a$ and $\sigma_b = \sigma_b(x_i, y_i) = \tilde{\sigma}_b$.

III. CONCISE EQUATIONS FOR THE BEST-FIT LINE

We have shown that the equations of York[4] contain all least-squares and maximum-likelihood solutions to the problem of fitting a straight line to data with (possibly correlated) normally distributed errors in x and y. All correct solutions that we are aware of in the literature can be derived (often as special cases) from those equations. If the newly unified standard errors of slope and intercept are used, then the error expressions reduce to particularly simple forms, yielding the following extremely compact set of four equations:

$$a = \overline{Y} - b\overline{X}, \tag{13a}$$

$$b = \frac{\Sigma W_i \beta_i V_i}{\Sigma W_i \beta_i U_i}, \tag{13b}$$

$$\sigma_a^2 = \frac{1}{\Sigma W_i} + \overline{x}^2 \sigma_b^2, \tag{13c}$$

$$\sigma_b^2 = \frac{1}{\Sigma W_i u_i^2}. \tag{13d}$$

Equation (13b) for b must, in the general case, be solved iteratively. A typical sequence of operations is

(1) Choose an approximate initial value of b (for instance, by simple regression of y on x).

(2) Determine the weights $\omega(X_i)$, $\omega(Y_i)$ for each point. If the errors in x and y are known, then normally $\omega(X_i) = 1/\sigma^2(X_i)$ and $\omega(Y_i) = 1/\sigma^2(Y_i)$, where $\sigma(X_i)$ and $\sigma(Y_i)$ are the errors in the x and y coordinates of the ith point.

(3) Use these weights, with the value of b and the correlations r_i (if any) between the x and y errors of the ith point, to evaluate W_i for each point.

(4) Use the observed points (X_i, Y_i) and W_i to calculate $\overline{X}$ and $\overline{Y}$, from which U_i and V_i, and hence β_i can be evaluated for each point.

(5) Use W_i, U_i, V_i, and β_i in the expression for b in Eq. (13b) to calculate an improved estimate of b.

(6) Use the new b and repeat steps (3), (4), and (5) until successive estimates of b agree within some desired tolerance (for example, one part in 10^{15}).

(7) From this final value of b, together with the final $\overline{X}$ and $\overline{Y}$, calculate a from Eq. (13a).

(8) For each point (X_i, Y_i), calculate the adjusted values x_i, where $x_i = \overline{X} + \beta_i$. (Similarly, $y_i = \overline{Y} + b\beta_i$, although these values are not needed in this calculation.)

(9) Use the adjusted x_i, together with W_i, to calculate $\overline{x}$, and thence u_i.

(10) From W_i, $\overline{x}$, and u_i, calculate σ_b, and then σ_a.

Although it is impossible to guarantee convergence for any arbitrary data set, years of experience have shown that the iteration procedure converges remarkably rapidly, with about ten iterations for most data sets, and fewer than 50 for pathological data sets such as Reed's data set II.[3]

The above algorithm is straightforward to program, and students would find it illuminating to compare the parameters resulting from the above algorithm (possibly using data with both x and y errors which they have acquired in a laboratory experiment) with the results of the simple regressions of y on x and x on y built into most hand calculators and spreadsheet programs.

Note that Eq. (13) is symmetrical in x and y (their superficial appearance to the contrary notwithstanding). They will therefore produce the identical straight line and corresponding errors if x and y are interchanged. In our work with ^{40}Ar–^{39}Ar geochronology, where the x intercept is significant,[12] we normally interchange x and y data to obtain the original x intercept and its standard error. Of course the slope obtained after the interchange is the reciprocal of the original slope.

If we use Eq. (13) and the definitions of W_i and β_i in Table I, it is easy to derive simplified solutions for special cases, many of which have been dealt with in the literature, sometimes with closed-form (noniterative) solutions. Most of these special cases use uncorrelated errors $(r_i = 0)$. For example the so-called major-axis solution[6] is given simply by setting $r_i = 0$ and $W_i = 1$. This solution corresponds to minimizing the sum of the squares of the perpendicular distances of the observed points from the fitted line. Although widely used, this solution is not invariant under a change of scale. To correct this deficiency, Kermack and Haldane[5] suggested the "reduced major-axis" solution (which *is* invariant under a change of scale). This solution corresponds to setting $r_i = 0$ and $W_i = 1/(\sigma_Y^2 + b\sigma_X^2)$, where $\sigma_X^2 = \Sigma(X_i - \overline{X})^2/(n-1)$ and $\sigma_Y^2 = \Sigma(Y_i - \overline{Y})^2/(n-1)$, that is, σ_X^2 is the variance of the X_i taken as a group, and similarly for σ_Y^2. The ubiquitous regression of y on x is given simply by setting $r_i = 0$ and $W_i = \omega(Y_i)$, where $\omega(Y_i) = 1$ if the regression is unweighted.

An example of regression with nonzero error correlations is given by Brooks *et al.*[7] As York pointed out,[4] their solution implicitly assumes perfect inverse correlation of x and y errors, and can be obtained from Eq. (13) by setting $r_i = -1$.

IV. MONTE CARLO TESTS OF ACCURACY

Now that we have derived unified LSE–MLE estimates of the standard errors of the slope and intercept, it is reasonable to ask how accurate these unified error estimates are. To find an absolute standard against which to test the above analytical approximations, we must examine the probabilistic model of linear fitting that forms the basis of the above mathematical analysis.

In this model, we have assumed that there exists a set of true points that lie exactly along a straight line, itself having a particular true intercept and slope. However, we can only measure the positions of these points imperfectly. Each point has associated measurement errors, expressed as a binormal distribution parametrized by x and y errors, and a correlation between those errors. The measurement process has randomly selected an observed point from the appropriate binor-

Table II. Results of Monte Carlo modeling. Each data set was run for 10^7 Monte Carlo trials. The quantities $\hat{\sigma}_a$, $\hat{\sigma}_b$ are the standard deviations of the Monte Carlo distributions of the parameters $\hat{a}$ and $\hat{b}$ (y intercept and slope); $\hat{a}$ and $\hat{b}$ are defined in Appendix E. The quantity Δ_a is defined as $\Delta_a = 100(\sigma_a - \hat{\sigma}_a)/\hat{\sigma}_a$; a similar definition holds for Δ_b. σ_a, σ_b are the analytical errors calculated from Eq. (13). Data set 3 is from Ref. 6 with weights of York—Ref. 2 (zero error correlations). Data set 4 is from Ref. 14.

Data set	Number of points	$S/(n-2)$	y intercept			Slope		
			$\hat{a}$	$\hat{\sigma}_a$	Δ_a (%)	$\hat{b}$	$\hat{\sigma}_b$	Δ_b (%)
1	7	5.382	30.155 10	0.316 481	0.015 707 7	−189.642 7	3.030 313	0.000 175 9
2	10	0.489	0.000 61	0.000 110	−0.127 829 6	−0.003 337	0.000 769	−0.126 700 8
3	10	1.483	5.479 91	0.295 713	−0.251 151 0	−0.480 533	0.058 256	−0.464 447 3
4	11	9.273	11.869 85	0.012 981	0.037 238 6	0.210 121	0.000 631	0.037 166 2
5	13	0.403	−13.278 74	1.586 903	−1.141 174 7	0.090 209	0.012 579	−1.392 359 2
6	15	2.260	130.400 32	2.407 611	−0.003 576 8	0.691 756	0.013 089	−0.005 689 7
7	19	0.023	45.044 43	2.928 639	0.013 963 5	−1.333 268	0.139 543	0.013 701 2
8	26	5.824	−2.389 36	0.198 183	0.009 069 1	1.834 382	0.031 145	0.040 527 0
9	34	1.488	3.372 95	0.009 926	−0.028 522 2	1.327 717	0.009 818	−0.044 751 3

mal distribution centered on each true point. Such measurements have then generated, from the set of true points, a set of observed points (which will not, in general, be collinear). Our task in fitting these observed points is to reverse this process; to try to undo the effects of the measurement errors, and thus to recover our best estimate of the slope and intercept of the original straight line (together with an estimate of the uncertainty in recovering those parameters).

If it were practical, the best way to estimate the uncertainty of our estimates would be to repeat the above measurement process a large number of times, each time generating a new set of observed points from the true points. Each set of observed points would then be used to derive a new best-fit line. By comparing the resulting large set of fitted slopes and intercepts obtained under identical experimental conditions, with the original true values of these parameters, we could then determine the average uncertainty of estimating the slope or intercept.

In reality, we do not have access to the true parameters of the line or to the true positions of the data points we are attempting to measure. The whole object of the fitting process is to estimate these quantities. Furthermore, practical considerations limit the number of possible repetitions of the experiment. So our best practical estimate of the true uncertainties in evaluating the slope and intercept from a given data set comes from repeated numerical experiments, that is, from a Monte Carlo model of repeated measurement.

For a given observed data set, such a Monte Carlo model begins by fitting the data set with the first two members of Eq. (13), and using the parameters of this best-fit line as the true parameters $(\hat{a},\hat{b})$, and the least-squares adjusted positions of the n data points as the true points $(\hat{x}_i,\ \hat{y}_i)$. Each true point is then assigned the actual errors of measurement $(\sigma(X_i),\sigma(Y_i),\ i=1,2,\ldots,n)$ and correlation coefficient r_i associated with the corresponding observed point. (As mentioned, both LSE and MLE methods agree on the slope and intercept of the best-fit line, and also agree on their estimates of the true positions of the fitted points. In LSE these are termed the ''adjusted'' points,[10] and in MLE they are the ''expectation values'' of the points.[9]) During this initial fitting of the observed data set, we also calculate the uncertainties in the intercept and slope (σ_a,σ_b) from Eq. (13). These are the two uncertainties whose accuracy we wish to assess.

We then conduct a simulated measurement process on this set of postulated true points, by generating, from the binormal distributions associated with each of the n true points, a set of n random ''observed'' points. These observed points can then be used to obtain a best-fit line, characterized by a slope and intercept that will be different from $\hat{a}$ and $\hat{b}$, the parameters of the original true line. If we repeat this simulated measurement process N times on our set of true points, we obtain N estimated slopes and intercepts. The distributions of the N slopes and intercepts about the known true values, as N becomes very large (10^7 in our Monte Carlo models), are measures of the expected errors of estimating the slope and intercept in a single measurement, such as the actual physical measurement that we originally performed. If the observed distribution of the slopes or intercepts is Gaussian, then the standard deviation of the distribution is the standard error of the parameter (slope or intercept) being estimated. (In fact, for all nine data sets in Table II, the 10^7 pairs of a and b values yielded histograms almost perfectly matching Gaussian distributions.[13]) Note that the standard deviations should be calculated with respect to the true y-intercept $\hat{a}$ and slope $\hat{b}$, rather than the means $\overline{a}$ and $\overline{b}$ of the N intercepts and slopes; that is, $\hat{\sigma}_a=\Sigma(a_i-\hat{a})^2/N$ rather than $\Sigma(a_i-\overline{a})^2/(N-1)$ and similarly for $\hat{\sigma}_b$. These quantities, $\hat{\sigma}_a$ and $\hat{\sigma}_b$, act as the true values against which we test the estimates (σ_a,σ_b) calculated from Eq. (13).

We have used this approach to test the calculated unified LSE–MLE errors in the slope and intercept against the results of Monte Carlo modeling, using a variety of real, experimentally derived data sets. These include data sets having 7–34 data points, showing a range of $\sim 10^5$ in $\hat{a}$ and $\hat{b}$, a range from -0.9998 to 0.8728 in the correlation coefficient r_{ab}, and with a range of more than a hundred in the goodness-of-fit parameter $S/(n-2)$ discussed below. The details of the Monte Carlo calculations are given in Appendix E.

In Table II we summarize the results of nine Monte Carlo models. We compare these true errors $\hat{\sigma}_a$ and $\hat{\sigma}_b$ for $N=10^7$ with the errors σ_a and σ_b calculated from the original experimental data set using Eq. (13). The Δ_a and Δ_b values are the percent differences between these calculated and true errors. The maximum value of Δ observed in the nine data sets was less than 1.4%, and two-thirds of the values were well under 0.1%. In other words, the error estimates calculated using Eq. (13) with these data sets are themselves typically in error by less than a percent. Clearly the approximations made in deriving Eq. (13) are exceptionally good in practice.

We also wish to emphasize that our numerical results reinforce the conclusions of Ref. 9 that the MLE (our adjusted-point LSE) and traditional LSE (our observed-point LSE) errors are very similar. Although we plan a more elaborate exploration of this and other aspects of the Monte Carlo modeling in a later paper, we note here that the deviations between the LSE observed-point errors and the adjusted-point errors for the nine data sets presented are all less than 10%, and two-thirds are well under 1%. This agreement is surprisingly good for error estimates derived from a relatively small number of experimental points. We conclude that published results based on least-squares treatments (such as those of York[2,4]) which use observed-point errors will, in general, remain valid for all practical purposes.

V. GOODNESS OF FIT

In general, the deviations of the observed points from the fitted points should be on the order of the assigned errors of the observed points. This concept can be quantified by considering the weighted sum of deviations from the best-fit line (with error correlations taken into account). This quantity, $S=\Sigma W_i(Y_i-bX_i-a)^2$, is the same one minimized in the least-squares formulation of the fitting problem.[4] If n points are being fitted, the expected value of S has a χ^2 distribution for $n-2$ degrees of freedom, so that the expected value of $S/(n-2)$ is unity.

Without discussing in detail what to do if $S/(n-2)$ is appreciably different from unity, we simply note that it can be interpreted either as a statistical fluke (with a probability obtainable from a table of χ^2), or as a failure of the assumptions (for example, the presumed linear relation is incorrect, the errors of the observed points are wrongly assigned, or an unaccounted for factor, such as systematic error, has affected the measurements). One technique that is sometimes applied if $S/(n-2)$ is significantly larger than unity is to multiply the calculated σ_a and σ_b values by $\sqrt{S/(n-2)}$, which is equivalent to multiplying all the x and y errors, $\sigma(X_i)$ and $\sigma(Y_i)$, by the same factor. This makes $S=n-2$, without affecting the computed slope and intercept. This procedure, of course, should not be applied mechanically, without giving some thought to its appropriateness. One can easily imagine situations where alternative actions would be more reasonable.

ACKNOWLEDGMENTS

D.Y. and N.M.E. were supported in this research by a grant from the Natural Sciences and Engineering Research Council of Canada. J.D.D. was supported by a M.Sc. grant from the Consejo Nacional de Ciencia y Tecnología (México). Useful discussions were enjoyed with Patrick E. Smith and Enrique Gómez Treviño. Two anonymous referees made helpful comments that have significantly improved the presentation.

APPENDIX A: MAXIMUM LIKELIHOOD ESTIMATES OF $\tilde{\sigma}_a$ AND $\tilde{\sigma}_b$

The expression of Ref. 9 for $\tilde{\sigma}_b^2$ [our Eq. (1b)] can obviously be written

$$\tilde{\sigma}_b^2=\frac{1}{\Sigma W_i x_i^2-\dfrac{(\Sigma W_i x_i)^2}{\Sigma W_i}}=\frac{1}{\Sigma W_i x_i^2-\overline{x}\Sigma W_i x_i}, \tag{A1a}$$

and therefore

$$\tilde{\sigma}_b^2=\frac{1}{\Sigma W_i x_i(x_i-\overline{x})}. \tag{A1b}$$

Recall that $x_i=u_i+\overline{x}$. If we substitute this expression into Eq. (A1b) for $\tilde{\sigma}_b^2$, we find

$$\tilde{\sigma}_b^2=\frac{1}{\Sigma W_i(u_i+\overline{x})u_i}=\frac{1}{\Sigma W_i u_i^2+\overline{x}\Sigma W_i u_i}. \tag{A2}$$

But

$$\begin{aligned}\sum W_i u_i&=\sum W_i(x_i-\overline{x})\\&=\sum W_i x_i-\overline{x}\sum W_i\\&=\left(\sum W_i\right)\left(\frac{\Sigma W_i x_i}{\Sigma W_i}-\overline{x}\right)\\&=\left(\sum W_i\right)(\overline{x}-\overline{x})=0.\end{aligned} \tag{A3}$$

Therefore

$$\tilde{\sigma}_b^2=\frac{1}{\Sigma W_i u_i^2}, \tag{A4}$$

as in Eq. (2b).

If we divide our Eq. (1a) by Eq. (1b), we find for the MLE calculations of Ref. 9:

$$\tilde{\sigma}_a^2=\tilde{\sigma}_b^2\frac{\Sigma W_i x_i^2}{\Sigma W_i}. \tag{A5}$$

But $u_i=x_i-\overline{x}$ and $\Sigma W_i u_i=0$, so

$$\begin{aligned}\sum W_i x_i^2&=\sum W_i(u_i+\overline{x})^2\\&=\sum W_i(u_i^2+2\overline{x}u_i+\overline{x}^2)\\&=\sum W_i u_i^2+2\overline{x}\sum W_i u_i+\overline{x}^2\sum W_i\\&=\frac{1}{\tilde{\sigma}_b^2}+\overline{x}^2\sum W_i.\end{aligned} \tag{A6}$$

So

$$\frac{\Sigma W_i x_i^2}{\Sigma W_i}=\frac{1}{\tilde{\sigma}_b^2\Sigma W_i}+\overline{x}^2 \tag{A7}$$

and from Eq. (A5),

$$\tilde{\sigma}_a^2=\tilde{\sigma}_b^2\frac{\Sigma W_i x_i^2}{\Sigma W_i}=\tilde{\sigma}_b^2\frac{1}{\tilde{\sigma}_b^2\Sigma W_i}+\overline{x}^2\tilde{\sigma}_b^2=\frac{1}{\Sigma W_i}+\overline{x}^2\tilde{\sigma}_b^2, \tag{A8}$$

as in Eq. (2a).

APPENDIX B: LEAST-SQUARES CUBIC, QUADRATIC, AND LINEAR EQUATIONS WHEN ERRORS IN x AND y ARE CORRELATED

In York[4] the following cubic (B1), quadratic (B2), and linear (B3) equations for the best slope b were given for the case when X_i and Y_i are correlated:

$$b^3\sum \frac{W_i^2U_i^2}{\omega(X_i)}-b^2\left[2\sum \frac{W_i^2U_iV_i}{\omega(X_i)}+\sum \frac{W_i^2r_iU_i^2}{\alpha_i}\right]$$
$$-b\left[\sum W_iU_i^2-2\sum \frac{W_i^2r_iU_iV_i}{\alpha_i}-\sum \frac{W_i^2V_i^2}{\omega(X_i)}\right]$$
$$+\sum W_iU_iV_i-\sum \frac{W_i^2r_iV_i^2}{\alpha_i}=0, \tag{B1}$$

$$b^2\sum W_i^2\left[\frac{U_iV_i}{\omega(X_i)}-\frac{r_iU_i^2}{\alpha_i}\right]+b\sum W_i^2\left[\frac{U_i^2}{\omega(Y_i)}-\frac{V_i^2}{\omega(X_i)}\right]$$
$$-\sum W_i^2\left[\frac{U_iV_i}{\omega(Y_i)}-\frac{r_iV_i^2}{\alpha_i}\right]=0, \tag{B2}$$

$$b\sum W_i^2U_i\left[\frac{U_i}{\omega(Y_i)}+\frac{bV_i}{\omega(X_i)}-\frac{br_iU_i}{\alpha_i}\right]-\sum W_i^2V_i\left[\frac{U_i}{\omega(Y_i)}\right.$$
$$\left.+\frac{bV_i}{\omega(X_i)}-\frac{r_iV_i}{\alpha_i}\right]=0, \tag{B3}$$

that is,

$$b=\frac{\Sigma W_i^2V_i\left[\frac{U_i}{\omega(Y_i)}+\frac{bV_i}{\omega(X_i)}-\frac{r_iV_i}{\alpha_i}\right]}{\Sigma W_i^2U_i\left[\frac{U_i}{\omega(Y_i)}+\frac{bV_i}{\omega(X_i)}-\frac{br_iU_i}{\alpha_i}\right]}. \tag{B4}$$

All three of these equations (linear, quadratic, and cubic) yield identical values for best b and the errors σ_a and σ_b.

Equation (B4), York's linear algorithm, was the first such pseudo-linear solution of the general least-squares problem with correlated errors. It may also be written as

$$b=\frac{\Sigma W_i\beta_iV_i+bA}{\Sigma W_i\beta_iU_i+A}, \tag{B5}$$

where $A=\Sigma W_i^2U_iV_i(r_i/\alpha_i)$. By cross multiplication and collection of the explicit terms in b, we have

$$b=\frac{\Sigma W_i\beta_iV_i}{\Sigma W_i\beta_iU_i}, \tag{B6}$$

a form given in Ref. 11, which confirmed the result of Ref. 4.

APPENDIX C: EVALUATION OF σ_a^2 AND σ_b^2 AT LSE-ADJUSTED POINTS

By definition, all of the LSE-adjusted points (x_i,y_i) fall on the LSE best straight line.[4] Thus,

$$y_i=a+bx_i, \tag{C1a}$$

$$\sum W_iy_i=a\sum W_i+b\sum W_ix_i, \tag{C1b}$$

$$\overline{y}=a+b\overline{x}, \tag{C1c}$$

$$y_i-\overline{y}=b(x_i-\overline{x}). \tag{C1d}$$

That is,

$$v_i=bu_i \quad \text{for all} \quad i, \tag{C2}$$

where

$$u_i=x_i-\overline{x}, \tag{C3a}$$

$$v_i=y_i-\overline{y}. \tag{C3b}$$

Thus we evaluate σ_b^2 at the adjusted points by substituting u_i for U_i, v_i for V_i and $bu_i=v_i$ in Eq. (5). The numerator of Eq. (5) becomes

$$\begin{aligned}\text{numerator}&=\sum W_i^2\left[\frac{U_i^2}{\omega(Y_i)}+\frac{V_i^2}{\omega(X_i)}-\frac{2r_iU_iV_i}{\alpha_i}\right]\\&=\sum W_i^2\left[\frac{u_i^2}{\omega(Y_i)}+\frac{v_i^2}{\omega(X_i)}-\frac{2r_iu_iv_i}{\alpha_i}\right]\\&=\sum W_i^2\left[\frac{u_i^2}{\omega(Y_i)}+\frac{b^2u_i^2}{\omega(X_i)}-\frac{2r_ibu_i^2}{\alpha_i}\right]\\&=\sum W_i^2u_i^2\left[\frac{1}{\omega(Y_i)}+\frac{b^2}{\omega(X_i)}-\frac{2br_i}{\alpha_i}\right]\\&=\sum W_i^2u_i^2\frac{1}{W_i}=\sum W_iu_i^2,\end{aligned} \tag{C4}$$

when evaluated at the LSE-adjusted points.

The denominator in Eq. (5) is D^2, where

$$D=\frac{1}{b}\sum W_iU_iV_i+4\sum W_i(\beta_i-U_i)(\beta_i-\overline{\beta})$$
$$-\frac{1}{b}\sum W_i^2\frac{r_i}{\alpha_i}(bU_i-V_i)^2. \tag{C5}$$

When we evaluate Eq. (C5) at the adjusted points, we immediately see that the third term vanishes, because bU_i-V_i transforms to $bu_i-v_i=0$ for all i.

In the second term we have β_i and $\overline{\beta}$ to transform. Now,

$$U_i-\beta_i=U_i-W_i\left[\frac{U_i}{\omega(Y_i)}+\frac{bV_i}{\omega(X_i)}-\frac{(bU_i+V_i)r_i}{\alpha_i}\right], \tag{C6}$$

by the definition of β_i. If we evaluate Eq. (C6) at the adjusted values (x_i,y_i), we have

$$\begin{aligned}\text{adjusted }(U_i-\beta_i)&=u_i-W_i\left[\frac{u_i}{\omega(Y_i)}+\frac{bv_i}{\omega(X_i)}-\frac{2bu_ir_i}{\alpha_i}\right]\\&=u_i-W_iu_i\left[\frac{1}{\omega(Y_i)}+\frac{b^2}{\omega(X_i)}-\frac{2br_i}{\alpha_i}\right]\\&=u_i-W_iu_i\frac{1}{W_i}=u_i-u_i=0,\end{aligned} \tag{C7}$$

for all i. Thus the second term in D also vanishes. Then D, evaluated at the adjusted values (x_i,y_i), becomes

$$\text{adjusted } D=\frac{1}{b}\sum W_iu_iv_i=\frac{1}{b}\sum W_iu_ibu_i=\sum W_iu_i^2. \tag{C8}$$

Then the value of the σ_b^2, evaluated at the adjusted (x_i,y_i), becomes

$$\sigma_b^2(x_i,y_i)=\frac{\Sigma W_i u_i^2}{(\Sigma W_i u_i^2)^2}=\frac{1}{\Sigma W_i u_i^2}=\tilde{\sigma}_b^2, \tag{C9}$$

by Eq. (A4), thus proving Eq. (7).

For the case of σ_a, from Eq. (11),

$$\sigma_a^2=\frac{1}{\Sigma W_i}+(\bar{X}+2\bar{\beta})^2\sigma_b^2+\frac{2(\bar{X}+2\bar{\beta})\bar{\beta}}{D}. \tag{C10}$$

To evaluate this expression at the adjusted points, we have to evaluate $\bar{\beta}$ there. We will show that in this case $\bar{\beta}=0$ by proving that in general $\bar{\beta}=\bar{x}-\bar{X}$. We have from York,[4]

$$x_i=X_i-W_i(bU_i-V_i)\left(\frac{b}{\omega(X_i)}-\frac{r_i}{\alpha_i}\right). \tag{C11}$$

But it is easy to see from the definition of β_i that

$$W_i(bU_i-V_i)\left(\frac{b}{\omega(X_i)}-\frac{r_i}{\alpha_i}\right)=U_i-\beta_i. \tag{C12}$$

Then

$$x_i=X_i-(U_i-\beta_i)=X_i-(X_i-\bar{X}-\beta_i)=\bar{X}+\beta_i. \tag{C13}$$

Thus, $\bar{x}=\bar{X}+\bar{\beta}$, or $\bar{\beta}=\bar{x}-\bar{X}$. That is, when evaluated at the adjusted points (x_i,y_i), $\bar{\beta}=\bar{x}-\bar{x}=0$. If we substitute $\bar{\beta}=0$ and $\bar{X}=\bar{x}$ in Eq. (C10) for σ_a^2, we obtain

$$\begin{aligned}\sigma_a^2(x_i,y_i)&=\frac{1}{\Sigma W_i}+\bar{x}^2\sigma_b^2(x_i,y_i)\\&=\frac{1}{\Sigma W_i}+\bar{x}^2\tilde{\sigma}_b^2 \quad [\text{by Eq. (C9)}]\\&=\tilde{\sigma}_a^2, \quad [\text{by Eq. (A8)}],\end{aligned} \tag{C14}$$

which proves Eq. (12).

APPENDIX D: PROOF THAT $\sigma_a^2(X_i,Y_i)=\tilde{\sigma}_a$ AND $\sigma_b^2(X_i,Y_i)=\tilde{\sigma}_b$ IN REGRESSION OF y ON x AND x ON y

In the classical regression of y on x (weighted if desired) it is assumed that the X_i are free of error and all the scatter is attributed to errors in Y_i. Thus $r_i=0$ automatically, and W_i collapses to $\omega(Y_i)$. Furthermore, in this case,

$$\beta_i=W_i\left[\frac{U_i}{\omega(Y_i)}+\frac{bV_i}{\omega(X_i)}\right]=\omega(Y_i)\frac{U_i}{\omega(Y_i)}=U_i, \tag{D1}$$

because $\omega(X_i)\gg\omega(Y_i)$. Hence, σ_b^2, evaluated at the observed points (X_i,Y_i) as traditionally done, becomes, from Eq. (5),

$$\sigma_b^2(X_i,Y_i)=\frac{\Sigma\omega(Y_i)^2\dfrac{U_i^2}{\omega(Y_i)}}{\left[\dfrac{1}{b}\Sigma\omega(Y_i)U_iV_i\right]^2}=b^2\frac{\Sigma\omega(Y_i)U_i^2}{[\Sigma\omega(Y_i)U_iV_i]^2}. \tag{D2}$$

Now Eq. (B4) becomes

$$b=\frac{\Sigma\omega(Y_i)^2\dfrac{V_iU_i}{\omega(Y_i)}}{\Sigma\omega(Y_i)^2\dfrac{U_i^2}{\omega(Y_i)}}=\frac{\Sigma\omega(Y_i)U_iV_i}{\Sigma\omega(Y_i)U_i^2}, \tag{D3}$$

so that $\Sigma\omega(Y_i)U_iV_i=b\Sigma\omega(Y_i)U_i^2$. We substitute this value for $\Sigma\omega(Y_i)U_iV_i$ in Eq. (D2),

$$\sigma_b^2(X_i,Y_i)=\frac{1}{\Sigma\omega(Y_i)U_i^2}. \tag{D4}$$

But for this regression the X_i have no errors, that is, $X_i=x_i$ and $\bar{X}=\bar{x}$, so that $U_i=u_i$. Then

$$\sigma_b^2(X_i,Y_i)=\frac{1}{\Sigma\omega(Y_i)u_i^2}=\tilde{\sigma}_b^2 \tag{D5}$$

by Eq. (2b). Similarly, because $\bar{X}=\bar{x}$, $\bar{\beta}=0$ for this regression. But $D\neq 0$, hence from Eqs. (11) and (2a)

$$\begin{aligned}\sigma_a^2(X_i,Y_i)&=\frac{1}{\Sigma\omega(Y_i)}+\bar{x}^2\sigma_b^2(X_i,Y_i)\\&=\frac{1}{\Sigma\omega(Y_i)}+\bar{x}^2\tilde{\sigma}_b^2=\tilde{\sigma}_a^2.\end{aligned} \tag{D6}$$

By symmetry, when x is regressed on y, $\sigma_a^2(X_i,Y_i)$ and $\sigma_b^2(X_i,Y_i)$ are (despite both being evaluated at their observed values) also automatically identical with $\tilde{\sigma}_a^2$ and $\tilde{\sigma}_b^2$, respectively.

APPENDIX E: MONTE CARLO MODELING OF ERROR ESTIMATES

The modeling proceeds as follows.

(1) Take an experimental data set, consisting of observed points (X_i,Y_i), which are scattered about a line and have Gaussian errors with standard deviations $\sigma(X_i)$ and $\sigma(Y_i)$, where the errors have a correlation r_i, $-1\leqslant r_i\leqslant 1$. Use Eq. (13) and fit this data set to a line to obtain the y-intercept $\hat{a}$, the slope $\hat{b}$, and the standard errors σ_a and σ_b. Use the adjusted points (x_i,y_i), $(i=1,\ldots,n)$, as a set of true initiating points distributed along the fitted straight line $(\hat{a},\hat{b})$ which is now taken to be the underlying true straight line for this data set.

(2) For each collinear point (x_i,y_i) generate a new (X_i',Y_i') at random from the binormal distribution function $\mathcal{N}[\sigma(X_i),\sigma(Y_i),r_i]$ centered on (x_i,y_i). The new set of (X_i',Y_i') will of course *not* be collinear, and represents an observed data set in the Monte Carlo model.

(3) Use the (X_i',Y_i') data set to calculate with Eq. (13) a new best-fit line with parameters (a_j,b_j).

(4) With the original collinear (x_i,y_i) of step (1), repeat steps (2) and (3) for $j=1,2,\ldots,N$, where N is some large number (say, 10^7) to generate a sequence of intercepts, $\mathbf{a}=(a_1,a_2,\ldots,a_N)$, and of slopes, $\mathbf{b}=(b_1,b_2,\ldots,b_N)$.

(5) Calculate the standard deviation $\hat{\sigma}_a$ of the sequence $\mathbf{a}$ from $\hat{\sigma}_a^2=\Sigma(a_i-\hat{a})^2/N$, and similarly for the standard deviation $\hat{\sigma}_b$ of the sequence $\mathbf{b}$.

[a]Electronic mail: york@physics.utoronto.ca
[b]Currently at Facultad de Ingeniería, Universidad Autónoma de Baja California, Km. 107 Carr. Tijuana-Ensenada, Ensenada, Baja California, México 22860.
[1]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in FORTRAN: The Art of Scientific Computing* (Cambridge U.P., New York, 1992), 2nd ed.
[2]Derek York, ''Least-squares fitting of a straight line,'' Can. J. Phys. **44**, 1079–1086 (1966).
[3]B. Cameron Reed, ''Linear least-squares fits with errors in both coordinates,'' Am. J. Phys. **57**, 642–646 (1989).
[4]D. York, ''Least squares fitting of a straight line with correlated errors,'' Earth Planet. Sci. Lett. **5**, 320–324 (1969).
[5]K. A. Kermack and J. B. S. Haldane, ''Organic correlation and allometry,'' Biometrika **37** (1/2), 30–41 (1950).
[6]Karl Pearson, ''On lines and planes of closest fit to systems of points in space,'' Philos. Mag. **2** (6), 559–572 (1901).
[7]C. Brooks, I. Wendt, and W. Harre, ''A two-error regression treatment and its application to Rb-Sr and initial ^{87}Sr/^{86}Sr ratios of younger Variscan granitic rocks from the Schwarzwald Massif, Southwest Germany,'' J. Geophys. Res. **73**, 6071–6084 (1968).
[8]B. Cameron Reed, ''Linear least-squares fits with errors in both coordinates. II. Comments on parameter variances,'' Am. J. Phys. **60** (1), 59–62 (1992).
[9]D. Michael Titterington and Alex N. Halliday, ''On the fitting of parallel isochrons and the method of maximum likelihood,'' Chem. Geol. **26**, 183–195 (1979).
[10]W. Edwards Deming, *Statistical Adjustment of Data* (Wiley, New York, 1943).
[11]J.-F. Minster, L.-P. Ricard, and C. J. Allègre, ''^{87}Rb-^{87}Sr chronology of enstatite meteorites,'' Earth Planet. Sci. Lett. **44**, 420–440 (1979).
[12]Grenville Turner, ''^{40}Ar-^{39}Ar ages from lunar maria,'' Earth Planet. Sci. Lett. **11**, 169–191 (1971).
[13]J. De Basabe Delgado, ''Regresión lineal con incertidumbres en todas las variables: Aplicaciones en geocronología al cálculo de isocronas,'' M.Sc. thesis, CICESE, 2002.
[14]F. Albarède, *Introduction to Geochemical Modeling* (Cambridge U.P., Cambridge, 1995).