

# Extraction of Physical Quantities from Isotope Mixing Diagrams.

Jamie Gilmour.

Last updated  
22 /6 / 2015



Imaging Science and Biomedical Engineering Division,  
Medical School, University of Manchester,  
Stopford Building, Oxford Road,  
Manchester, M13 9PT.

# Extraction of Physical Quantities from Isotope Mixing Diagrams.

Jamie Gilmour

School of Earth, Atmospheric and Environmental Science, University of Manchester.

## Abstract

I derive a formula for the square of the Mahalanobis distance that allows a straight line to be fitted to data where uncertainties in  $x$  and  $y$  are correlated. I show how errors in the resulting gradient and intercept can be calculated from the Cramér-Rao bound. I provide VBA Excel® macros that perform these calculations. Using them I demonstrate that this approach reproduces the results of the standard York fit as implemented in the Excel® add in “Isoplot”. I show that the choice of normalising isotope can have a major influence on the quantities derived from fitting and the likelihood calculated for the model. This occurs because normalisation to an isotope determined with low precision produces error distributions that are poorly approximated by normal distributions, and so a fundamental condition required for the validity of the fitting process is violated. For this reason, calculations of relevant quantities should be carried out using normalisation to the most precisely measured isotope, usually one of high relative abundance. Graphical representations can be constructed in other, more familiar, normalisations based on these calculations if desired, but it should be noted that the error bars may not accurately reflect underlying distributions.

## 1. Introduction.

Isotope geochemists work with ratios between isotope abundances (atoms/atoms, [ions counted]/[ions counted]). A common tool is the mixing diagram or isochron diagram, where two ratios that have a common denominator are plotted in an (x,y) coordinate system. The power of this diagram is that measured compositions that are mixtures between two underlying components plot on a straight line. This allows for scientifically valuable conclusions to be extracted by fitting straight lines to data. For instance, in idealised radioisotope dating, one component is an underlying composition present when the rock formed. Over time radioactive decay produced a second component in which the decay product is present in fixed proportion to the parent element and the third isotope is absent. Analyses yield varying mixtures of the two components which allows their compositions to be constrained. Formally, it is hypothesised that two components can account for a data set. If this is correct, the data will plot on a straight line in a mixing diagram. A goodness-of-fit statistic is used to assess whether this hypothesis is supported by the data. If it is, parameters to constrain the components can be extracted from the best fit (e.g. the relative abundance of daughter isotope to parent, from which an age can be calculated).

It is widely recognised that the use of a common denominator in the isotope ratios introduces a complication to the fitting process; the errors in the two isotope ratios are correlated; this complicates the construction of a goodness-of-fit statistic to be minimized by a fitting routine. The problem was famously addressed by York<sup>1</sup> and “York fits” are now the community standard.

Because three isotopic abundances are involved there is a choice of three common denominators in constructing a mixing diagram. Since the underlying data are identical and are always being used to evaluate the same model, rigorously consistent treatments ought to yield agreement in the goodness-of-fit statistic to within the rounding error of the calculation. This not always the case – the goodness-of-fit statistics, the physical parameters of the best-fit models and their uncertainties vary significantly with choice of normalising isotope. This implies that different assumptions are implicitly being made when the denominator is chosen. When this occurs, which (if any) treatment should be adopted? Agreement within one standard deviation is not an appropriate test for comparing two treatments of the same dataset and cannot justify adopting the one that yields the lowest calculated errors or the best fit.

In what follows I revisit the steps in deriving quantities from mixing diagrams, derive the fitting criteria and implement them in Excel®, show that implementing these procedures reproduces the values extracted

---

<sup>1</sup> York D. (1969) Least Squares Fitting of a Straight Line with Correlated Errors. *Earth and Planetary Science Letters* 5 320-324

from a widely used package distributed as an Excel® addin (“Isoplot”, K. R. Ludwig, Berkeley Geochronology Centre), and examine where the variation dependent on choice of denominator might originate. Rather than using the “error propagation” in a form familiar to most isotope geochemists, I frame the discussion in terms of covariance matrices, first illustrating their use in a familiar context. Finally, I show how choice of denominator can determine the validity of the fitting process and make recommendations about which to adopt for calculation of goodness-of-fit, associated parameters, and their errors.

## 2. Deriving Uncertainties on Isotope Ratios.

Suppose a mass spectrometer measures abundances of three isotopes –  $a$ ,  $b$  and  $c$ . The amounts correspond, ultimately, to counts in a detector. It is assumed that count levels are high enough that the measurements can be approximated by normal distributions with standard deviations  $\Delta a$  etc. It is also assumed that there is no covariance between the uncertainties in  $a$ ,  $b$  and  $c$ . Writing the variance in  $a$  as  $v_a (= [\Delta a]^2)$  (etc.), the covariance matrix is...

$$\mathbf{Cov}(a, b, c) = \begin{bmatrix} v_a & 0 & 0 \\ 0 & v_b & 0 \\ 0 & 0 & v_c \end{bmatrix} \quad (1)$$

In calculating isotope ratios  $R_1 = a/c$  and  $R_2 = b/c$  we have a Jacobian...

$$\mathbf{J} = \begin{bmatrix} \partial R_1 / \partial a & \partial R_1 / \partial b & \partial R_1 / \partial c \\ \partial R_2 / \partial a & \partial R_2 / \partial b & \partial R_2 / \partial c \end{bmatrix} = \begin{bmatrix} 1/c & 0 & -a/c^2 \\ 0 & 1/c & -b/c^2 \end{bmatrix} \quad (2)$$

The covariance matrix for  $R_1$  and  $R_2$  is then...

$$\begin{aligned} \mathbf{Cov}(R_1, R_2) &= \mathbf{J} \mathbf{Cov}(a, b, c) \mathbf{J}^T \\ &= \begin{bmatrix} v_a/c^2 + a^2 v_c/c^4 & ab v_c/c^4 \\ ab v_c/c^4 & v_b/c^2 + a^2 v_c/c^4 \end{bmatrix} \\ &= \begin{bmatrix} \text{var}(R_1) & \text{covar}(R_1, R_2) \\ \text{covar}(R_1, R_2) & \text{var}(R_2) \end{bmatrix} \end{aligned} \quad (3)$$

where “var” and “covar” denote variance and covariance respectively. (The variance is the square of the standard deviation, the correlation coefficient is the covariance divided by the product of the standard deviations of the two correlated quantities.) This allows us to calculate the variances on  $R_1$  and  $R_2$  and the covariance between them. The variances on  $R_1$  and  $R_2$  are identical to those calculated by the conventional error propagation method.

$$\frac{\Delta R_1^2}{R_1^2} = \frac{\Delta a^2}{a^2} + \frac{\Delta c^2}{c^2} \left( = \frac{v_a}{a^2} + \frac{v_c}{c^2} \right) \quad (4)$$

## 3. Fitting a straight line to data consisting of ratios with variances and covariances.

We have a data set of isotope ratios with a common denominator and wish to fit a straight line of the form  $y=mx+c$  where  $m$  is the gradient and  $c$  is the intercept with the  $y$  axis. We do this by choosing  $m$  and  $c$  to maximize the probability that the dataset was produced according to the underlying physical principle that predicts the line. This is the same as maximizing the likelihood of the model given the data. We need a measure of this likelihood/probability. This measure is the square of the Mahalanobis Distance. To adapt a quote from Wikipedia<sup>2</sup>:

The Mahalanobis distance of an observation  $x = (x_1, x_2, x_3, \dots, x_n)^T$  from a set of observations with mean  $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_n)^T$  and covariance matrix  $\mathbf{S}$  is defined as:

<sup>2</sup> [https://en.wikipedia.org/wiki/Mahalanobis\\_distance](https://en.wikipedia.org/wiki/Mahalanobis_distance)

$$D(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)} \quad (5)$$

The square of the Mahalanobis distance is a  $\chi^2$  statistic, in other words it is twice the natural logarithm of the likelihood of the model given the data. It quantifies the likelihood of the model given the data.

How do we use this? There are a series of steps...

**Step 1.** To translate this into our terms, our observations are  $R_1 (\equiv x_m)$  and  $R_2 (\equiv y_m)$ . (This subscript “m” is for “measured”.) We model these with the straight line  $y_m = mx_m + c$ .

**Step 2.** Given a particular straight line, we want to find the point on the line that minimizes  $D_x$ , the square of the Mahalanobis distance, for the point  $(x_m, y_m)$ . It is simpler to do this calculation after a coordinate transformation that moves the point to the origin, making the  $\mu$  in equation (5) all zero. The gradient of the line we are considering remains the same. The intercept of the line becomes  $c_F = mx_m - y_m + c$  and the new coordinates are  $(x, y)$ .

The covariance matrix of  $(x, y)$  is identical to that of  $(x_m, y_m)$ ; to see this...

$$y = y - y_m \text{ and } x = x - x_m \text{ so } \frac{\partial y}{\partial y_m} = \frac{\partial x}{\partial x_m} = -1.$$

The Jacobian for the transformation is thus

$$J = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

So

$$\mathbf{Cov}(x, y) = J \mathbf{Cov}(x_m, y_m) J^T \text{ and } J J^T = \mathbf{I} \text{ (the 2x2 identity matrix)}$$

The square of the Mahalanobis distance from a point on the line where  $x=x_L$  to the origin in the new coordinate system is then...

$$D_{x_L} = \begin{bmatrix} x_L \\ mx_L + c_F \end{bmatrix}^T \mathbf{Cov}(x, y) \begin{bmatrix} x_L \\ mx_L + c_F \end{bmatrix} \quad (6)$$

Finding  $x_L$  such that this is minimized, then substituting it back in to find the corresponding  $D_x$  is a standard bit of calculus and leads to...

$$D_x = D c_F^2 Q \quad (7)$$

Where

$$Q = v_x - \frac{A^2}{B}$$

$$A = (m v_x - c_v)$$

$$B = (v_y - 2m c_v + v_x m^2)$$

and D is the determinant of the covariance matrix *i.e.*

$$D = \frac{1}{(v_x v_y - c_v^2)}$$

$v_x$  and  $v_y$  are the variances in  $x$  and  $y$  respectively, and  $c_v$  is the covariance between errors in  $x$  and  $y$ .

**Step 3.** This minimum squared Mahalanobis distance can then be summed over all points in the dataset, and the resulting quantity minimized with respect to  $m$  and  $c$  to find the best fit line. (Note that  $c$ , the intercept of the line, enters the calculation through  $c_F$ .) There is no explicit solution to the minimization, but standard software such as Excel® (via the Solver® add in) has minimization routines or a custom minimization routine can be implemented (see Appendix).

Fig. 1 provides an Excel® macro to calculate the square of the Mahalanobis distance of a dataset from the line  $y = mx + c$ .

```

Public Function Mahalanobis(inrangel As Variant, m As Double, c As Double) As Variant
'This calculates the sum of the squares of the Mahalanobis distances for the data in the
'range "inrangel" given a line of gradient "m" and intercept "c". "inrangel" is five
columns
'wide with data in order x, sd(x), y, sd(y), covar(x,y). The covariance is used, not the
'correlation coefficient - the function could be edited to change this by the user. Output
'of this function can be minized using solver to fit a straight line to data with
correlated
'errors (c) Jamie Gilmour. No restrictions are imposed on the use of this function, but no
'guarantee is given and you use it at your own risk.
'July 2015
Dim cr, D, bit1, bit2, bit3, sum As Double
numrows = inrangel.Rows.Count
sum = 0
For i = 1 To numrows
x = inrangel(i, 1)
vx = inrangel(i, 2) ^ 2
y = inrangel(i, 3)
vy = inrangel(i, 4) ^ 2
cv = inrangel(i, 5)
cf = m * x - y + c
D = 1 / (vx * vy - cv * cv)
bit1 = D * cf * cf
bit2 = (vx * m - cv) ^ 2
bit3 = (vy - 2 * m * cv + vx * m * m)
sum = sum + bit1 * (vx - bit2 / bit3)
Next i
mahalanobis = sum
End Function

```

Fig 1: VBA function to calculate the square of the Mahalanobis distance of a line from a dataset. If minimizing this it is important to provide sensible starting  $m$  and  $c$ .

**Step 4.** We now need to estimate the variances and covariances on and between  $m$  and  $c$ . For this we use the Cramér–Rao lower bound (CRLB)<sup>3</sup>. Loosely, this tells us the smallest the variance and covariance on  $m$  and  $c$  can possibly be. These minima correspond to the case where the errors really are normally distributed, something we have tacitly assumed in using the theory above (and assume whenever we calculate uncertainties by error propagation). The lower bounds are therefore our error estimates. The CRLB tells us that the covariance matrix of  $m$  and  $c$  will be the inverse of the Fisher Information Matrix, or...

$$\begin{bmatrix} v_m & \text{Cov}(m, c) \\ \text{Cov}(m, c) & v_c \end{bmatrix} = \left[ \frac{1}{2} \begin{bmatrix} \sum \partial^2 D_x / \partial m^2 & \sum \partial^2 D_x / \partial m \partial c \\ \sum \partial^2 D_x / \partial m \partial c & \sum \partial^2 D_x / \partial c^2 \end{bmatrix} \right]^{-1} \quad (8)$$

Where the summations are over all data points and the factor of  $\frac{1}{2}$  arises because  $\chi^2$  is twice the natural logarithm of the likelihood. Again, with some standard but slightly long winded calculus it can be shown that...

$$\begin{aligned} \partial^2 D_x / \partial m^2 &= 2Dx^2Q - 8 \frac{DxQc_F A}{B} + 8 \frac{Dc_F^2 Q A^2}{B^2} - 2 \frac{Dc_F^2 Q v_x}{B} \\ \partial^2 D_x / \partial c^2 &= 2DQ \end{aligned} \quad (9)$$

$$\partial^2 D_x / \partial m \partial c = 2DQ \left( x - 2 \frac{Ac_F}{B} \right)$$

...where quantities are defined in (7) above. (If these formulae look daunting, compare the appendix of York, 1969.)

Once values of  $m$  and  $c$  that minimize the square of the Mahalanobis distance for a dataset have been found, the corresponding covariance matrix can be calculated using the Excel function in Fig. 2. In the Appendix I provide details of VBA software for Excel that can be used to apply this approach to data.

<sup>3</sup>[https://en.wikipedia.org/wiki/Cram%C3%A9r%E2%80%93Rao\\_bound](https://en.wikipedia.org/wiki/Cram%C3%A9r%E2%80%93Rao_bound)

```

Public Function CramerRao(inrangel As Variant, m As Double, c As Double) As Variant
'This calculates the variances and covariances of m and c once for the data in the
'range "inrangel" given a best fit line of gradient "m" and intercept "c". "inrangel" is
five columns
'wide with data in order x, sd(x), y, sd(y), covar(x,y). The covariance is used, not the
'correlation coefficient - the function could be edited to change this by the user.
'(c) Jamie Gilmour. No restrictions are imposed on the use of this function, but no
'guarantee is given and you use it at your own risk.
'July 2015
Dim cr, D, A, B, Q, t1, t2, t3, t4, sumd2dc2, sumd2dms, sumd2dm2 As Double
Dim CRdet As Double
numrows = inrangel.Rows.Count
sumd2dc2 = 0
sumd2dm2 = 0
sumd2dm2c = 0
For i = 1 To numrows
x = inrangel(i, 1)
vx = inrangel(i, 2) ^ 2
y = inrangel(i, 3)
vy = inrangel(i, 4) ^ 2
cv = inrangel(i, 5)
cf = m * x - y + c
D = 1 / (vx * vy - cv * cv)
A = m * vx - cv
B = vy - 2 * m * cv + vx * m * m
Q = (vx - A * A / B)
sumd2dc2 = sumd2dc2 + 2 * D * Q
t1 = 2 * D * x * x * Q
t2 = 8 * D * x * Q * cf * A / B
t3 = 8 * A * A * D * cf * cf * Q / (B * B)
t4 = 2 * D * cf * cf * Q * vx / B
sumd2dm2 = sumd2dm2 + t1 - t2 + t3 - t4
sumd2dm2c = sumd2dm2c + 2 * D * Q * (x - 2 * cf * A / B)
Next i
sumd2dc2 = 0.5 * sumd2dc2
sumd2dm2 = 0.5 * sumd2dm2
sumd2dm2c = 0.5 * sumd2dm2c
'Fisher information matrix calculated, now invert.
CRdet = 1 / (sumd2dm2 * sumd2dc2 - sumd2dm2c * sumd2dm2c)
varm = sumd2dc2 * CRdet
varc = sumd2dm2 * CRdet
Covar = -sumd2dm2c * CRdet
CramerRao = Array(varm, varc, Covar)
End Function

```

Fig 2: VBA function to calculate the variances and covariance of the gradient and intercept from a fitted line using the Cramér-Rao bound. (This function returns an array over one row and three columns.)

#### 4. Benchmarking against Isoplot.

Isoplot, written by K. R. Ludwig of the Berkeley Geochronology Centre, is a visual basic add in for Microsoft Excel®. Tests here were performed using MS Excel 2003 and version 3.75 of Isoplot.

A simulated dataset (Table 1) of 10 points defined to scatter about a mixing line with a MSWD (Mean Standard Weighted Distribution<sup>4</sup>) close to 1 was used to compare the output of the theory outlined in section 3 with the results of Isoplot. The Isoplot “other X-Y [-Z] plot” option was chosen and data, errors and correlation coefficients supplied. The square of the Mahalanobis distance was calculated using the function in Fig. 1, and minimized using the Solver® routine before the function in Fig. 2 was used to calculate variances and covariances.

Output from Isoplot is compared with output from the procedures and functions of section 2 in Table 2. Within reasonable calculation errors agreement is exact.

#### 5. Monte Carlo Test.

To further evaluate this approach it was tested against a simulated dataset.

<sup>4</sup> MSWD is  $\chi^2$  divided by the number of degrees of freedom. In this case, with 10 datapoints fitted by a model that has two parameters (intercept and gradient)  $MSWD = \chi^2 / (10 - 2) = \chi^2 / 8$  where  $\chi^2$  is the Mahalanobis distance.

$R_1$	$\Delta R_1$	$R_2$	$\Delta R_2$	Covariance	Correlation Coefficient
0.000692	4.07974E-05	0.384799	0.001132	6.40E-10	0.013861
0.001313	5.06921E-05	0.312079	0.000894	8.01E-10	0.017663
0.001603	5.14626E-05	0.263917	0.000742	6.98E-10	0.018284
0.001813	5.07916E-05	0.227369	0.00063	5.86E-10	0.018312
0.002028	5.03397E-05	0.199737	0.000547	5.05E-10	0.018358
0.002028	4.74391E-05	0.180834	0.000486	4.06E-10	0.017605
0.002214	4.69449E-05	0.165033	0.000437	3.63E-10	0.017689
0.002355	4.61747E-05	0.152667	0.000399	3.25E-10	0.017642
0.002367	4.42935E-05	0.142528	0.000367	2.79E-10	0.017165
0.002459	4.33319E-05	0.134064	0.00034	2.51E-10	0.017027

Table 1: Data used to illustrate comparison of Isoplot and the theory presented here. The correlation coefficient is the covariance divided by the product of the standard deviations or  $R_1$  and  $R_2$ ).

	Isoplot	This work
Gradient, $1\sigma$ error	-146.9	3.9
Intercept, $1\sigma$ error	0.4931	0.0076
Correlation coefficient between errors	-0.958	-0.958
MSWD	1.3	1.3

Table 2: Illustrative comparison between output from Isoplot and this work, precision adopted from Isoplot report.  $1\sigma$  errors have been calculated from Isoplot output by dividing quoted  $2\sigma$  errors by 1.96 (Isoplot Users Manual, p22).

	Isotope a, error	Isotope b, error	Isotope c, error
483.4951	27.48565	2.50E+05	6.25E+02
1009.258	39.71103	2.50E+05	6.25E+02
1490.265	48.25494	2.50E+05	6.25E+02
2046.34	56.54561	2.51E+05	6.26E+02
2480.018	62.24972	2.51E+05	6.26E+02
2987.806	68.32604	2.55E+05	6.31E+02
3571.19	74.69929	2.60E+05	6.38E+02
3987.503	78.93334	2.65E+05	6.43E+02
4438.672	83.2792	2.69E+05	6.49E+02
4990.344	88.30296	2.75E+05	6.55E+02

Table 3: Simulated isotope abundances for Monte Carlo

Quantity	Value
Standard deviation of calculated gradient	4.135
Average calculated uncertainty on gradient:	4.102
Standard deviation of calculated intercept	0.00820
Average calculated uncertainty on intercept	0.00809
Covariance between errors on gradient and intercept.	-0.0326
Average calculated covariance	-0.0320

Table 4: Monte Carlo test results.

A set of 10 abundances of three isotopes with associated uncertainties was constructed (Table 3). This dataset served as the basis for 1000 simulations. These were generated by adding to each abundance in Table 3 an offset calculated as the corresponding uncertainty in Table 3 multiplied by a factor sampled from a normal distribution. These data were therefore normally distributed about the mean abundances with standard deviations corresponding to the assigned errors. These new quantities and the uncertainties in Table 3 then served as the “Isotope Abundances” from which ratios  $R_1$  and  $R_2$  (see section 2) and their associated errors and covariance were calculated. A best fit line was then made to each of the 1000 simulated datasets using the Mahalanobis subroutine (Fig. 1), and the errors on the gradient and intercept

and their covariance calculated using the CramerRao subroutine (Fig. 2). Finally, the averages, standard deviations and covariance of the distributions of calculated errors and gradients were calculated, and the distributions compared to the averages of the predicted errors and their covariance (Table 4). Agreement is excellent.

While by no means demonstrating conclusively that the approach is flawless, the tests described here indicate that it is valid. However, like all such treatments, it relies on the assumption that the isotope ratio data are normally distributed about a mean.

## 6. Are isotope ratios normally distributed?

It is common to all the approaches isotope geochemists use to fit a straight line to data that the underlying distributions characterised by measurement errors are assumed to be normal (Gaussian distributions). In the above treatment, this is assumed in constructing the Mahalanobis distance from the covariance matrix and when the Cramér-Rao bound is adopted as a means of determining the error on slope and gradient. The derived uncertainties are lower bounds in the sense that if the input quantities are not normally distributed the uncertainties will be higher than have been estimated. It is also inherent in the use of conventional error propagation techniques, such as those used by York in his derivation. How appropriate is it to assume that isotope ratios and their errors represent the means and standard deviations of normal distributions?

Ultimately, isotope abundances are measured by mass spectrometers and correspond to a number of ions counted. The underlying distributions are thus Poisson distributions which, though not identical to normal distributions, can reasonably be approximated by normal distributions when significant numbers of counts have been detected.

However, the ratio of two normal distributions is not itself a normal distribution. In a special case, the ratio of two normal distributions centred on the origin is a Cauchy distribution, or Lorentzian. When the two normal distributions have non-zero means the resulting distribution is more complex. There are circumstances where the ratios of normal distributions measured by isotope geochemists cannot be approximated by another normal distribution, and knowing this should influence the way in which the calculation of correlation lines is performed.

This can be investigated analytically. Suppose an isotope ratio  $x = a/c$  has uncertainties that have close to a normal distribution, under what circumstances is the distribution of the reciprocal  $y = c/a$  closely approximated by a normal distribution? Given the normal distribution...

$$P(x)\Delta x = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x-x_0)^2}{2\sigma_x^2}\right) \Delta x$$

Where  $x_0$  is the mean and  $\sigma_x$  the standard deviation of the distribution,  $y$  has the distribution...

$$P(y)\Delta y = \frac{y_0^2}{y^2} \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\left(\frac{y}{y_0}\right)^2 \frac{(y-y_0)^2}{2\sigma_y^2}\right) \Delta y \quad (10)$$

This approximates a normal distribution for values of  $y$  close to  $y_0$ , and so can be reliably treated as normal when  $\sigma_y$  is small compared to  $y_0$ . Fig. 3 compares the form of this distribution to that of the normal distribution assumed when error propagation is used to calculate the uncertainty in  $y$ .

In Fig. 4 the results of an “experiment” are presented that test how well and under what conditions the distribution of an isotope ratio approximates a normal distribution. One thousand pairs of normally distributed isotope abundances ( $a,c$ ) were generated with the standard deviations and averages stated in Fig 4. The corresponding ratios were then calculated with each of  $a$  and  $c$  as the denominator. A histogram was generated from the ratios, and compared to a normal distribution centred on the calculated ratio ( $a/c$  or  $c/a$ ) with standard deviation calculated according to equation (4) (i.e. using error propagation). At each bin the difference between observed and predicted numbers of ratios was divided by a Poisson error based

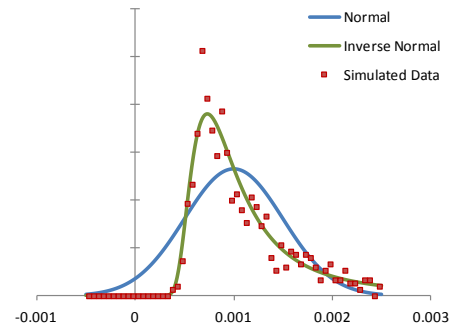


Fig. 3. Distribution of the reciprocal of a variable normally distributed with average 1000 and standard deviation 500.



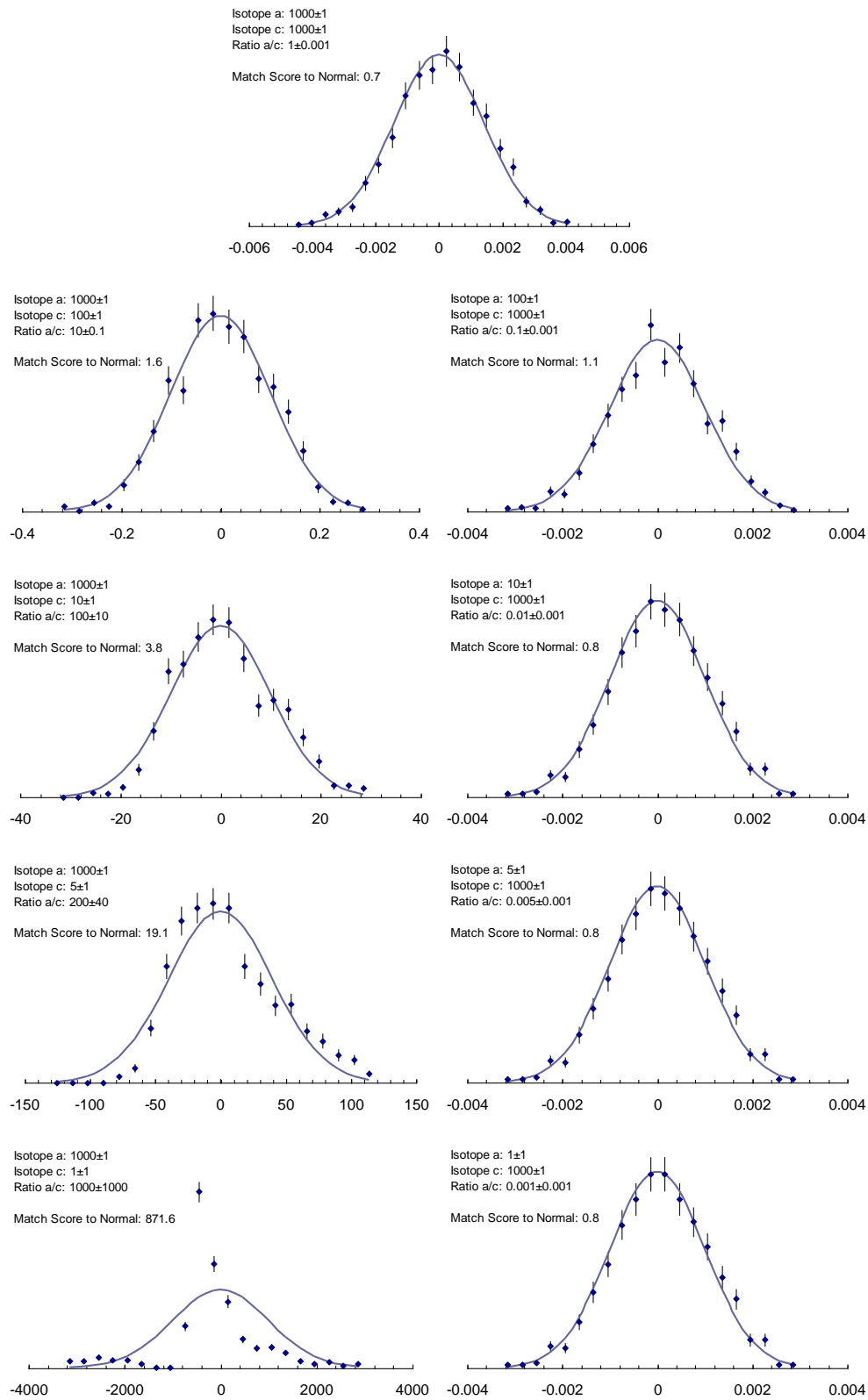


Fig. 3: Test of whether the ratio of two normally distributed isotopic abundances is itself normally distributed. Ratios on the left are the reciprocal of ratios on the right. In all cases, the assumption that the ratio is normally distributed (match score = 1) with the conventionally calculated error is a closer approximation when the less precisely measured isotope is the numerator than when it is the denominator.

on the number counted. These values were summed and divided by the number of isotope pairs input to produce the match score to normal listed on each figure alongside other input parameters. Values of this  $\chi^2$  close to 1 indicate that the ratios are normally distributed. When the precision of the denominator is low the match is not good. In all cases the ratio approximates more closely to a normal distribution when the less precisely measured isotope is the numerator than when it is the denominator. In other words, the assumption used in fitting that all errors are normally distributed is most valid when the isotope measured most precisely (usually the most abundant) is used as the denominator in calculating ratios.

Does this departure from normal distributions in isotope ratios have the potential to affect derived quantities? It can certainly account for one inconsistency in treating isotopic data – derived quantities and MSWD values that are not identical within calculation error between data treatments that adopt different normalising isotopes. In particular, MSWD and the square of the Mahalanobis distance are measures of the extent to which the data support the hypothesis (for instance, that they indicate two component mixing or isochronous behaviour). The answer ought not to depend on the choice of normalising isotope. That it does is a consequence of the failure of the assumption that the isotope ratios have normally distributed errors. The results displayed in Figure 4 suggest that the problem is most likely to arise where one isotope is present in only small quantities in comparison to others measured and so is poorly determined, and this is chosen as the denominator in the mixing diagram. Conversely, fits on mixing diagrams are most valid when the most precisely defined isotope is used for normalisation, because this makes the isotope ratio distributions most closely approximate normal distributions.

These circumstances can especially arise in noble gas analyses in the I-Xe and Ar-Ar systems. Samples are usually analysed by step heating, where samples are held at each of a series of increasing temperature steps and the gas released analysed. In each case, gas from some temperature steps can be dominated by radiogenic ( $^{40}\text{Ar}$  and  $^{129}\text{Xe}$  from the decay of  $^{40}\text{K}$  and  $^{129}\text{I}$  respectively) and nucleogenic ( $^{39}\text{Ar}$  and  $^{128}\text{Xe}$  produced after neutron capture on  $^{39}\text{K}$  and  $^{127}\text{I}$  during artificial irradiation respectively) components. The third isotope in each system (usually  $^{36}\text{Ar}$  or  $^{132}\text{Xe}$ ) can be present only in a trapped component, and it is not uncommon in some releases for these isotopes to be dominated by blank. The standard normalisation for isochron diagrams is to the minor isotope,  $^{36}\text{Ar}$  or  $^{132}\text{Xe}$  because this makes the age correspond to the gradient of the line in the mixing diagram.

The results shown in Fig. 4 suggest that the requirement that errors be normally distributed will be most closely met when one of the more abundant isotopes is used as the denominator in mixing diagrams, and that these should be the preferred choice. This is especially the case when the abundance of the trapped component in some data points is, within error, zero; in this case the use of  $^{129}\text{Xe}/^{132}\text{Xe}$  or  $^{40}\text{Ar}/^{36}\text{Ar}$  in calculating a best fit line invalidates the fit. A pathological example of this is illustrated in Table 5, where results from a simulated argon-argon dataset with two choices of normalising isotope are displayed. One simulated release (italicised) has no significant  $^{36}\text{Ar}$  above blank. In this case, the conventional normalisation to  $^{36}\text{Ar}$  indicates a well supported hypothesis of two component mixing, while the alternative normalization to  $^{40}\text{Ar}$  reveals that this hypothesis is not supported by the data at all. This can be understood with reference to Fig. 3. Calculating the Mahalanobis distance of a line from this point – the probability that the point lay on the line, on the assumption of a normal distribution would not give the correct answer.

Calculations should thus always be made in ratios normalised to an abundant isotope where all ratios approximate normal distribution. If required to produce conventional plots normalised to the least abundant isotope, *e.g.* “normal” isochron diagrams where the age corresponds to the gradient of the line, the calculation of the parameters of the line should be carried out using the alternative normalisation. It should, however, be recognised that the error bars may not adequately reflect the distributions of the data. The parameters of the line plotted in the conventional diagram can then be calculated from the fit to the data using the correct normalisation choice. For instance, in the I-Xe system, the conventional isochron diagram is a plot of  $^{129}\text{Xe}/^{132}\text{Xe}$  vs  $^{128}\text{Xe}/^{132}\text{Xe}$ . An I-Xe age is derived from the gradient of a line fitted to a linear array of points. However, model evaluation and parameter estimation (including age determination) should be carried out in a plot of  $^{129}\text{Xe}/^{129}\text{Xe}$  vs  $^{132}\text{Xe}/^{129}\text{Xe}$ , where the age is derived from the intercept with

the vertical axis. If the fitted line in this representation is  $y=mx+c$ , the equation of the line in the conventional representation is:

$$y_n = \frac{1}{c}x_n - \frac{m}{c} \quad (8)$$

*i.e.* the gradient is  $1/c$  and the intercept with the  $y$  axis is at  $-m/c$  (note that the gradient of the line in the alternative representation is expected to be negative, so this intercept is expected to be a positive number).

The  $R_{129}$  ( $= {}^{129}\text{Xe}/{}^{132}\text{Xe}$ ) composition of a trapped component with known  $R_{128}$  ( $= {}^{128}\text{Xe}/{}^{132}\text{Xe}$ ) known to within some uncertainty (*i.e.*  $R_{128} \pm \Delta_{128}$ ) can be calculated. For instance, in the alternative representation where  $x = {}^{132}\text{Xe}/{}^{129}\text{Xe}$  and  $y = {}^{128}\text{Xe}/{}^{129}\text{Xe}$  it is the intersection of the fitted line with the line  $y=R_{128}x$  leading to:

$$R_{129} = \frac{mR_{128}}{R_{128}} + c \quad \text{so} \quad R_{129} = \frac{c}{\left(1 - \frac{m}{R_{128}}\right)} \quad (9)$$

and the corresponding covariance “matrix” (it only has one element) is calculated from...

Isotopic Abundances						
<sup>36</sup> Ar	Err	<sup>39</sup> Ar	Err	<sup>40</sup> Ar	Err	Err
2.19E+00		3.83E+00		5.0182E+05		5.06E+02
1.1107E+03		2.38E+01		8.0046E+05		6.39E+02
1.5500E+03		2.81E+01		9.4773E+05		6.95E+02
2.0075E+03		3.20E+01		1.0973E+06		7.48E+02
2.3791E+03		3.48E+01		1.2505E+06		7.99E+02

Derived Ratios Relative to <sup>36</sup> Ar					
<sup>39</sup> Ar/ <sup>36</sup> Ar	Err	<sup>40</sup> Ar/ <sup>36</sup> Ar	Err	Covariance	Corr. Coeff.
1.1E+05		2.3E+05		8.01E+10	1.00E+00
2.2525E+02		4.84E+00		7.4563E+01	9.97E-01
1.6182E+02		2.94E+00		3.2567E+01	9.96E-01
1.2460E+02		1.99E+00		1.7309E+01	9.95E-01
1.0522E+02		1.55E+00		1.1860E+01	9.94E-01

Derived Ratios Relative to <sup>40</sup> Ar					
<sup>36</sup> Ar/ <sup>40</sup> Ar	Err	<sup>39</sup> Ar/ <sup>40</sup> Ar	Err	Covariance	Corr. Coeff.
4.4E-06		7.6E-06		2.22E-12	3.33E-04
1.3876E-03		2.98E-05		2.76E-10	1.82E-02
1.6355E-03		2.97E-05		2.33E-10	1.85E-02
1.8295E-03		2.92E-05		1.94E-10	1.84E-02
1.9025E-03		2.79E-05		1.55E-10	1.78E-02

Derived Quantities			
	Standard Isochron ( <sup>40</sup> Ar/ <sup>36</sup> Ar vs <sup>39</sup> Ar/ <sup>36</sup> Ar)		Inverse Isochron ( <sup>39</sup> Ar/ <sup>40</sup> Ar vs <sup>36</sup> Ar/ <sup>40</sup> Ar)
Gradient	1.636±0.074	Gradient	-149.8±1.5
Intercept	349±11	Intercept	0.5015±0.0014
Covariance	-0.77	Covariance	-0.0012
MSWD	0.88	MSWD	12
Probability <sup>1</sup>	0.45	Probability <sup>1</sup>	Zero
<sup>40</sup> Ar*/ <sup>39</sup> Ar*	<b>1.64±0.07</b>	<sup>40</sup> Ar*/ <sup>39</sup> Ar*	<b>1.994±0.006</b>

Table 5: Two treatments of the same simulated dataset that differ only in choice of normalising isotope – the derived <sup>40</sup>Ar\*/<sup>39</sup>Ar\* ratios are dramatically different even though the input data and, apart from normalisation, numerical methods are identical. The effect can be attributed to the presence of one point (italicised) where the <sup>36</sup>Ar release was indistinguishable from blank and the <sup>40</sup>Ar\*/<sup>39</sup>Ar\* ratio was an outlier. As shown in Fig. 3, propagation of errors to a ratio where this is the denominator dramatically overestimates the width of the distribution, allowing a lower calculated MSWD at the expense of an inaccurate fit (the data were simulated with <sup>40</sup>Ar\*/<sup>39</sup>Ar\* = 2). In contrast, ratios where this isotope is the numerator closely approximate a normal distribution resulting in a more realistic MSWD. Derived quantities were calculated with the routines presented here. Probability values have been extracted from Isoplot fits that yielded identical results.

$$(\Delta_{128})^2 = \mathbf{J} \mathbf{cov}(m, c, R_{128}) \mathbf{J}^T \quad (10)$$

$$\mathbf{J} = \begin{bmatrix} \frac{\partial R_{129}}{\partial m} & \frac{\partial R_{129}}{\partial c} & \frac{\partial R_{129}}{\partial R_{128}} \end{bmatrix}$$

$$\mathbf{cov}(m, c, R_{128}) = \begin{bmatrix} \text{var}(m) & \text{covar}(m, c) & 0 \\ \text{covar}(m, c) & \text{var}(c) & 0 \\ 0 & 0 & \text{var}(R_{128}) \end{bmatrix}$$

where J is the Jacobian, and it is assumed that uncertainty in the input ratio is not correlated with uncertainty in the line parameters.

## 7. Summary and Recommendations

An approach to fitting a straight line to data with correlated errors has been developed based on the square of the Mahalanobis distance and the Cramér-Rao bound. This approach makes the same assumptions as, and yields effectively identical answers to, the widely used York fit. Both approaches assume normally distributed errors around the input data, a condition that some choices of normalising isotope can lead to the data not adequately meeting. Whichever approach is adopted, to minimise the effects of isotope ratios not having normally distributed errors, fits should always be carried out in data normalised to an isotope of high relative abundance.

**Acknowledgements.** This work grew out of many conversations with Neil Thacker, whose explanations of the underlying statistical principles were invaluable. It was supported by the STFC through grant ST/M001253/1 and by the Leverhulme Trust through RPG-2014-019.

Manchester and Bolton, July 2015

### Appendix: Fitting Routines in Visual Basic for Applications®

The Microsoft Excel® macro-enabled workbook Mahalanobis\_Fit.xls is available from the author. It contains implementations of the approach to fitting in sections 3 and 4 of this document. This is in the form of the functions shown in Figs 1 and 2, and the Macro “Mahalanobis Fit” which includes a minimization routine and generates a report in a text box including the corresponding quantities derived from the fit. The functions are in Module 1. The fitting macro is in Module 4.

You use any of this software at your own risk and no guarantee is made. You are welcome to redistribute it provided any modifications are commented, attributed to the person who made them, and themselves documented with comments.

#### Module 1

Public Function Mahalanobis(inrange1 As Variant, m As Double, c As Double) As Variant

This can be used as a function in Excel just as any other function can. *inrange1* is a contiguous range of the worksheet with the data to be fitted. Each datapoint should be on a row. From left to right the 5 cells should be “x”, “standard deviation of x”, “y”, “standard deviation of y”, “covariance”. *m* and *c* should be cells with guesses of the gradient and intercept respectively. Note that Isoplot uses the correlation coefficient, not the covariance as used here. It returns a single value corresponding to the sum of the squares of the Mahalanobis distance from the line of gradient *m* and intercept *c*. If it is called with a single row it therefore returns the residual for that point. Solver® can be used to minimize the output of this function by varying the values in the cells corresponding to *m* and *c* (remember to select “minimum” as the target!). To work, this requires that the initial guesses are not too far away from the final solution.

Public Function CramerRao(inrange1 As Variant, m As Double, c As Double) As Variant

This is an array function (highlight three adjacent cells in one row, type the function, press CTRL+ENTER). Input parameters are the same as for the Mahalanobis function. It returns, from left to right, the variance of *m*, the variance of *c*, and the covariance. **These only have meaning if *m* and *c* have previously been optimised w.r.t. the input data range using the Mahalanobis function.** To get the 1  $\sigma$  errors take the

square roots of the variances. To get the correlation coefficient divide the covariance by the product of the two standard deviations.

#### Module 4

This contains a macro “Mahalanobis\_Fit” and some private routines that it uses. Highlight the data range(s) and run the macro – output will be made to a new textbox (Fig A1) located over the input data range (it can be repositioned). Data can be selected in multiple areas; from first to last, and then left to right within each area, they will be x, sdx, y, sdy, covar. It was written and tested in Microsoft Excel 2007®.

The macro performs the following operations:

- Find values to initialise  $m$  and  $c$ . It looks for the minimum and maximum  $x$  values in the input data. It uses these and the corresponding  $y$  values to work out a gradient and intercept. If the data aren't linear or one of these is an outlier there is a danger that the starting points will not lead to the optimization converging on the best fit but get stuck in a local minimum.
- Find  $m$  and  $c$  values that minimize the square of the Mahalanobis distance using Private Sub Optim(inrange,  $m$ ,  $c$ ,  $dm$ ,  $dc$ ). This is an implementation of the Nelder-Mead downhill simplex method as described by Wikipedia<sup>5</sup>, it uses the terminology and variables listed there. In essence, it draws a triangle in  $m$ - $c$  space around the initial values provided. It then moves and shrinks the triangle to minimize the sum of the squares of the Mahalanobis distances of the data at the vertices. It returns the dimension of the triangle as  $dm$  and  $dc$ , and these are reported at the end of the macro. It works out the Mahalanobis distances using a separate function contained in the same module – Private Function mahalanobis2(inrange1,  $m$ ,  $c$ ) As Double – not the function in Module 1.
- Once the optimum  $m$  and  $c$  have been returned, subroutine Private Sub CramerRao2(inrange1,  $m$ ,  $c$ ,  $vm$ ,  $vc$ ,  $cv$ ) works out the variances and covariance. Once again, this is separate from the function in Module 1.
- Subroutine Mahalanobis\_Fit then works out the parameters to be reported. It uses the worksheet function Chidist to calculate the probability of the data given the line based on the value of  $\chi^2$  and the number of degrees of freedom (for different versions of Excel you may need to change this). If you want to change reported errors to  $2\sigma$ , scale them with MSWD, or alter any other aspect of what is reported, this is where to do it.
- Finally, subroutine Mahalanobis\_Fit creates a textbox that has its upper left corner in the upper left of the first selected data area and reports the output of the fit and some other bookkeeping data.

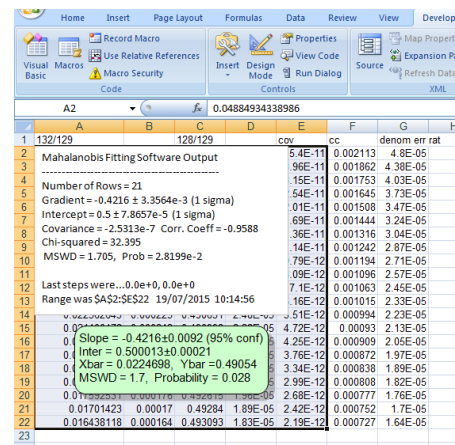


Fig A1. Example output from Mahalanobis\_Fit. Also shown – highlighted input range, output from Isoplot

<sup>5</sup>[https://en.wikipedia.org/wiki/Nelder-Mead\\_method](https://en.wikipedia.org/wiki/Nelder-Mead_method)