Available online at www.sciencedirect.com

ScienceDirect

Geochimica et Cosmochimica Acta 124 (2014) 237–249

Geochimica et Cosmochimica Acta

www.elsevier.com/locate/gca

Straight line regression through data with correlated uncertainties in two or more dimensions, with an application to kinetic isotope fractionation

Noah M. McLean

MIT, 77 Massachusetts Ave., Cambridge, MA 02139, United States

Received 15 August 2012; accepted in revised form 28 May 2013; available online 5 September 2013

Abstract

Straight line regression algorithms are used frequently for measured data that contain non-negligible uncertainties in each variable. For the general case of correlated measurement uncertainties between two variables that differ from one analysis to the next, the popular algorithm of York, 1968 calculates the maximum likelihood estimate for the line parameters and their uncertainties. However, it considers only two-dimensional data and omits the uncertainty correlation between the slope and *y*-intercept, an important term for evaluating confidence intervals away from the origin. This contribution applies the maximum likelihood method to straight line regression through data in any number of dimensions to calculate a vector-valued slope and intercept as well as the covariance matrix that describes their uncertainties and uncertainty correlations. The algorithm is applied to Pb data measured by TIMS with a silica gel activator that define a fractionation line in a three dimensional log-ratio space. While the log-ratios of even mass number Pb isotopes follow the slope predicted by mass-dependent fractionation with a Rayleigh or exponential law within calculated uncertainties, the log-ratio containing the odd mass number isotope ^{207}Pb diverges significantly, exhibiting mass-independent fractionation. The straight line regression algorithm is appropriate for fractionation lines that form linear trends in log-ratio space, but not for isochrons or mixing lines, which are predicted to be linear only when plotted as isotope or compositional ratios.

1. INTRODUCTION

It is not possible to measure a physical quantity with perfect accuracy and infinite precision. To estimate the degree to which the analysis corresponds to its true value, analysts assign uncertainties to the data, usually based on the variability of repeated measurements. Each repeated measurement often consists of multiple variables, like elemental or isotope ratios. The number of variables defines the dimension of the dataset, so that three variables can be plotted in a three-dimensional space, and so on. A series of repeated measurements of the same physical quantity, like the isotopic composition of a sample, can then be summarized by the mean of the repeated measurements and by their covariance matrix, which describes the uncertainty in each of the measured variables and the uncertainty correlations between them.

When multiple analyses fall on a line or plane in any number of dimensions, they can be described by a linear model $AX = B$, where A and B are predictor and response variables respectively, and elements of X are the model parameters, such as the slope and y-intercept. To calculate the best fit parameters, the definition of 'best' determines the regression technique used. The most familiar, ordinary least squares (OLS) in two dimensions, assumes that a noisy response variable, usually plotted on the y-axis, is a linear function of a predictor variable, usually plotted on the x-axis, which is not assigned uncertainty. The OLS algorithm can be extended to any number of dimensions, for instance determining the best fit plane to two predictor and one response variable. OLS minimizes the squares of the (vertical)

E-mail address: nmclean@mit.edu

0016-7037/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.gca.2013.08.035

distances from the measured data points to the regression line or plane, where this distance is measured on the uncertain response variable only. Since it ignores any uncertainty in the predictor variable, OLS is not an appropriate regression algorithm for data with non-negligible measurement uncertainties in all measured variables.

Another least squares approach, termed total least squares (TLS, Golub, 1973; Huffel and Markovsky, 1991), and also known as orthogonal regression, Deming regression, or reduced major axis regression (Sokal and Rohlf, 1995), minimizes the perpendicular distance between each point and the regression line or plane. While this model incorporates uncertainty in all variables, its simplest formulation assumes the uncertainties, for instance in both the x- and y-directions, are the same magnitude. Additionally, conventional TLS assumes the data are homoscedastic, or that each analysis has the same covariance matrix. For most applications in geo- and cosmochemistry, these assumptions do not hold: the variables plotted on each axis have different uncertainties that are often correlated, such as when the same isotope appears in the denominator of each isotope ratio, or all variables plotted are subject to the same correction. Separate analyses are frequently independent of one another but have different uncertainties, depending for instance on ion beam intensity, and thus are heteroscedastic.

In order to fit heteroscedastic data with correlated uncertainties, more sophisticated algorithms weight the influence of each measurement according to its uncertainty. For instance, a weighted OLS algorithm divides each squared (vertical) distance by its variance, then minimizes the sum of the resulting ratios. Weighted or generalized TLS approaches (e.g., Van Huffel and Vandewalle, 1989; Markovsky and Van Huffel, 2007), treat correlated uncertainties in both predictor variables and response variables, but not uncertainty correlations between them.

However, a variant termed elementwise-weighted total least squares (EW-TLS) solves the general regression case for multiple measurements of any linear system with heteroscedastic independent data points that have unique covariance matrices (Kukush and Van Huffel, 2004; Markovsky et al., 2006). When the uncertainties are assumed to be normally distributed, EW-TLS yields the maximum likelihood estimate (MLE) of the model parameters (Markovsky et al., 2006), meaning that the best fit parameters maximize a likelihood function that considers all the measured data, their uncertainties, and the normally-distributed measurement model, explained in detail in Section 2. Likewise, the general MLE of Sohn and Menke, 2002 to linear and non-linear regression models for correlated data additionally allow for uncertainty correlations between analyses. The highly cited York, 1968 algorithm is a special case of EW-TLS, calculating the MLE of the slope and y-intercept for a line in two dimensions, as well as their uncertainties.

This contribution focuses on a more general scenario than that of York, 1968: a straight line (one-dimensional) fit to heteroscedastic data with correlated, approximately normally distributed uncertainties for two or more variables. Recently published applications include, but are not limited to, kinetic and equilibrium isotope fractionation lines (Bizzarro et al., 2011; Zhang et al., 2011; Weyer and Seitz, 2012); trace element partitioning data (Balta et al., 2011); multiple isotope system plots, e.g., ϵ_{Hf} vs. ϵ_{Nd} in Vervoort et al., 2011; evaluation of correlations between elemental and isotopic data and instrumental mass fractionation during secondary ion or inductively-coupled plasma mass spectrometry (Hartley et al., 2012; Peel et al., 2008); empirical descriptions of temperature-dependent cation exchange reactions (Toner and Sletten, 2013); and use of concentrations of multiple elements in stream water and soil to estimate chemical weathering rates (Gaillardet et al., 1999; Ferrier et al., 2011). Although many two-component mixing lines, such as common Pb in $^{206}Pb/^{204}Pb$–$^{207}Pb/^{204}Pb$–$^{208}Pb/^{204}Pb$ space and isochrons, for instance in $^{207}Pb/^{206}Pb$–$^{204}Pb/^{206}Pb$ coordinates, create linear arrays of data, they are not suitable for linear regression because they contain ratios of major compositional components. This caveat is explored in Section 5.

Because the EW-TLS algorithm is a general case of the straight line model presented here, its previously derived statistical properties, such as convergence of the estimated fit parameters in probability to the true fit parameters as the number of measurements approaches infinity (Kukush and Van Huffel, 2004), apply here as well. However, by restricting the focus from multi-dimensional fit surfaces to a straight line, the resulting algorithm may be simplified and its solution optimized. For instance, terms involving the singular value or Cholesky decomposition of the inverse covariance matrix in (Markovsky et al., 2006) divide out in Eq. (12), avoiding costly linear algebra calculations.

The solution presented here, set in terms of the parametric equation for a line $\mathbf{p} = \mathbf{a} + \mathbf{v}t$ instead of its linear algebra $AX \approx B$ equivalent, is also easier to pose in terms of the parameters of interest. This is demonstrated in the example application to kinetic isotope fractionation lines in Section 4. The parametric equation for a line is also conveniently analogous to established methods of compositional data analysis (Egozcue et al., 2003).

2. ESTIMATION OF REGRESSION PARAMETERS

The general case of straight line regression involves determining the best fit line through n measured data points $\mathbf{p}_i$, $(i = 1 \ldots n)$, in k dimensions, where each measured datum has a unique covariance matrix $\mathbf{\Sigma}_i$. In this model, each $\mathbf{p}_i$ is a $k \times 1$ column vector for a single measurement that contains the value of each variable measured. $\mathbf{\Sigma}_i$ is its $k \times k$ covariance matrix, with variance terms for each variable on the diagonal and covariance terms that describe the uncertainty correlations between the variables on the off diagonals. The best fit line through the data can be expressed in terms of a direction vector, $\mathbf{v}$, and an arbitrary point on the line, $\mathbf{a}$, which form the parametric expression for a line, $\mathbf{a} + \mathbf{v}t$. This algorithm solves for the maximum likelihood estimates of $\mathbf{a}$ and $\mathbf{v}$, which are both $k \times 1$ column vectors, as well as their uncertainties and uncertainty correlations. Fig. 1 illustrates the straight line regression problem for $k = 3$ dimensions.

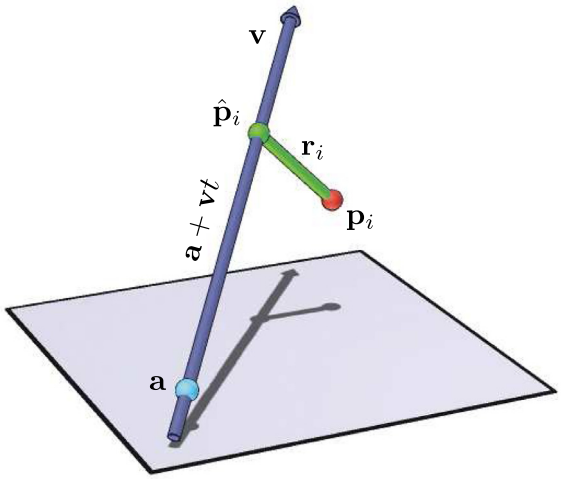

Fig. 1. Projection from a measured data point to the best fit line in three dimensions. The best fit line contains the point **a** (light blue), and follows the direction vector **v** (dark blue arrow). Any point on the line can be given by the expression $\mathbf{a} + \mathbf{v}t$. The closest point on the line to the measured data point $\mathbf{p}_i$ (red) is its perpendicular projection to the line $\hat{\mathbf{p}}_i$, in green. The vector difference between $\mathbf{p}_i$ and $\hat{\mathbf{p}}_i$ is known as the residual vector $\mathbf{r}_i$, shown as a green line segment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.1. Vector from a point to a line

Every regression algorithm depends on determining the difference, known as a residual, between a measured data point and the best fit line. To calculate the residual, first translating both the line and all measured points by $-\mathbf{a}$ simplifies the expression for the regression line. The new location of measured data point i is $(\mathbf{p}_i - \mathbf{a})$, and points on the line can now be expressed simply as $\mathbf{v}t$. Next, the perpendicular projection $\hat{\mathbf{p}}_i$ of each translated point onto the line is

$$\hat{\mathbf{p}}_i = \frac{\mathbf{v}\mathbf{v}^{\mathsf{T}}}{\mathbf{v}^{\mathsf{T}}\mathbf{v}}(\mathbf{p}_i - \mathbf{a}) \tag{1}$$

where $^{\mathsf{T}}$ denotes the transpose. Note that $\hat{\mathbf{p}}_i$ is in the translated coordinate system, so that $\hat{\mathbf{p}}_i + \mathbf{a}$ is the point's projection in the original coordinate system. The residual vector $\mathbf{r}_i$, or the misfit between the measured data point and its projection $\hat{\mathbf{p}}_i$, is

$$\mathbf{r}_i = (\mathbf{p}_i - \mathbf{a}) - \hat{\mathbf{p}}_i = (\mathbf{p}_i - \mathbf{a}) - \frac{\mathbf{v}\mathbf{v}^{\mathsf{T}}}{\mathbf{v}^{\mathsf{T}}\mathbf{v}}(\mathbf{p}_i - \mathbf{a}) \tag{2}$$

Because it is a vector difference, Eq. (2) does not depend on the reference frame. The magnitude of $\mathbf{r}_i$ represents the shortest Euclidean distance from the measured point to the best fit line. However, this distance does not consider the measured data uncertainties, embodied in the covariance matrices $\mathbf{\Sigma}_i$, and so does not necessarily represent the best reconciliation of the model line and the data. Likewise, when the data in any of the k axes are expressed in different units (e.g., meters and seconds), then this distance contains terms from each unit, mixing apples and oranges. The solution presented below, known as statistical normalization, can be used to construct probability distribution and likelihood functions for the measured data and line fit parameters.

2.2. Maximum likelihood and the minimum distance

The uncertainty in the measurement of the point $\mathbf{p}_i$ is quantified with its covariance matrix $\mathbf{\Sigma}_i$, which contains variances along its diagonal and covariances as off-diagonal elements. The shape of this uncertainty, the probability density function describes the relative probability that the true value of a measurement is in the vicinity of a point **x**. Here, **x** represents an arbitrary position, expressed as a $k \times 1$ column vector. If the data are normally distributed, the multivariate probability density function at any point **x** in the vicinity of the measurement $\mathbf{p}_i$ with covariance matrix $\mathbf{\Sigma}_i$ can be expressed as

$$f(\mathbf{x}\,|\mathbf{p}_i, \mathbf{\Sigma}_i) = \frac{1}{(2\pi)^{k/2}|\mathbf{\Sigma}_i|^{1/2}} \times \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{p}_i)^{\mathsf{T}}\mathbf{\Sigma}_i^{-1}(\mathbf{x} - \mathbf{p}_i)\right) \tag{3}$$

where $||$ is the matrix determinant and $\mathbf{\Sigma}^{-1}$ is the inverse of the covariance matrix. For a normally distributed univariate measurement ($k = 1$), the probability distribution function takes the shape of the well known 'bell curve'.

In any number of dimensions, a surface of constant probability density around the point $\mathbf{p}_i$ is the locus of all points **x** that satisfy

$$(\mathbf{x} - \mathbf{p}_i)^{\mathsf{T}}\mathbf{\Sigma}_i^{-1}(\mathbf{x} - \mathbf{p}_i) = c^2, \tag{4}$$

the exponential term in Eq. (3). The variable c in Eq. (4) is known as the Mahalanobis distance, and is analogous to the distance, in standard deviations, from **p** to **x**. In one dimension, the endpoints of a 1σ confidence interval have $c = 1$; in two dimensions, the ellipse with $c = 1$ is the 1σ uncertainty ellipse.

In Fig. 2a, the perpendicular projection from the measured point to the line $\mathbf{a} + \mathbf{v}t$ is not the best estimate of where the point actually falls on the line: in this case, the perpendicular vector intersects the line outside the 1σ uncertainty ellipse, but the line extends inside the ellipse elsewhere. The concept of orthogonal projection outlined above can be used only in the space where the Mahalanobis distance in Eq. (4) is the same as the Euclidean distance d, where

$$d^2 = (\mathbf{x} - \mathbf{p}_i)^{\mathsf{T}}(\mathbf{x} - \mathbf{p}_i). \tag{5}$$

In the space where the Euclidean and Mahalanobis distances agree, the covariance matrix and its inverse in Eq. (4) are both the identity matrix **I**, whose uncertainty ellipse is a circle (Fig. 2b), or a sphere or its higher-dimensional analog. The linear transform that maps the inverse covariance matrix to the identity matrix is given by **U**, the upper triangular matrix produced by Cholesky decomposition of the inverse covariance matrix:

$$\mathbf{\Sigma}^{-1} = \mathbf{U}^{\mathsf{T}}\mathbf{U} \tag{6}$$

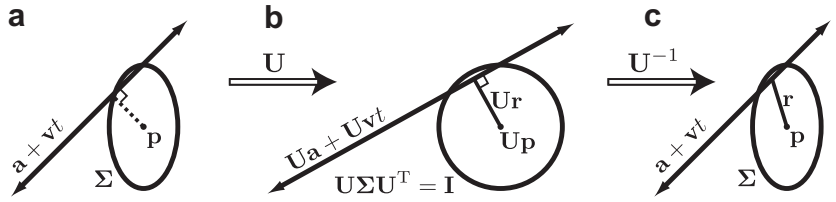

Fig. 2. (a) The point $\mathbf{p}$ with covariance matrix $\boldsymbol{\Sigma}$ depicted as a 1σ uncertainty ellipse (Mahalanobis distance = 1) is predicted to lie on the line $\mathbf{a}+\mathbf{v}t$. The perpendicular projection from the point to the line is not the best estimate of the point's location on the line. (b) The system in (a) with an applied linear transform defined by the matrix $\mathbf{U}$, here a stretching along the horizontal axis. In the transformed system, Euclidean distance is the same as Mahalanobis distance, and the best residual (solid line segment) is determined by the perpendicular projection algorithm of Section 2.1. (c) Although it is not part of the fitting algorithm, the inverse linear transform $\mathbf{U}^{-1}$ may be applied to restore the original system. The restored maximum likelihood residual is not the perpendicular vector between the point and the line.

where U has dimensions $k \times k$. Manipulating Eq. (6), it can be shown that the inverse covariance matrix becomes the identity matrix after a linear transformation $\mathbf{U}$,

$$\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{\mathrm{T}} = \mathbf{I}. \tag{7}$$

For a given measurement vector $\mathbf{p}$ and line parameters $\mathbf{a}$ and $\mathbf{v}$, the linear transformation is applied by premultiplying each column vector by $\mathbf{U}$. The residual vector in Eq. (2) becomes

$$\mathbf{U}\mathbf{r} = \mathbf{U}(\mathbf{p}-\mathbf{a}) - \mathbf{U}\hat{\mathbf{p}} = \mathbf{U}(\mathbf{p}-\mathbf{a}) - \frac{\mathbf{U}\mathbf{v}(\mathbf{U}\mathbf{v})^{\mathrm{T}}}{(\mathbf{U}\mathbf{v})^{\mathrm{T}}\mathbf{U}\mathbf{v}}\mathbf{U}(\mathbf{p}-\mathbf{a}) \tag{8}$$

and the squared Mahalanobis distance is the inner product of the residual with itself:

$$\|\mathbf{U}\mathbf{r}\|^2 = \left[\mathbf{U}(\mathbf{p}-\mathbf{a}) - \frac{\mathbf{U}\mathbf{v}(\mathbf{U}\mathbf{v})^{\mathrm{T}}}{(\mathbf{U}\mathbf{v})^{\mathrm{T}}\mathbf{U}\mathbf{v}}\mathbf{U}(\mathbf{p}-\mathbf{a})\right]^{\mathrm{T}} \left[\mathbf{U}(\mathbf{p}-\mathbf{a}) - \frac{\mathbf{U}\mathbf{v}(\mathbf{U}\mathbf{v})^{\mathrm{T}}}{(\mathbf{U}\mathbf{v})^{\mathrm{T}}\mathbf{U}\mathbf{v}}\mathbf{U}(\mathbf{p}-\mathbf{a})\right] \tag{9}$$

Eqs. (10)–(13) simplify this cumbersome expression. Factoring out a $\mathbf{U}$ from each term above and simplifying $(\mathbf{U}\mathbf{v})^{\mathrm{T}}$ to $\mathbf{v}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}$ yields

$$\|\mathbf{U}\mathbf{r}\|^2 = \left[\mathbf{U}\left((\mathbf{p}-\mathbf{a}) - \frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{v}}\mathbf{U}(\mathbf{p}-\mathbf{a})\right)\right]^{\mathrm{T}} \left[\mathbf{U}\left((\mathbf{p}-\mathbf{a}) - \frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{v}}\mathbf{U}(\mathbf{p}-\mathbf{a})\right)\right] \tag{10}$$

Transposing the two terms in brackets in the first term of Eq. (10) yields

$$\|\mathbf{U}\mathbf{r}\|^2 = \left((\mathbf{p}-\mathbf{a}) - \frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{v}}\mathbf{U}(\mathbf{p}-\mathbf{a})\right)^{\mathrm{T}}\mathbf{U}^{\mathrm{T}} U\left((\mathbf{p}-\mathbf{a}) - \frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{v}}\mathbf{U}(\mathbf{p}-\mathbf{a})\right) \tag{11}$$

Substituting $\boldsymbol{\Sigma}^{-1} = \mathbf{U}^{\mathrm{T}}\mathbf{U}$ from the definition of the Cholesky decomposition in Eq. (6) yields

$$\|\mathbf{U}\mathbf{r}\|^2 = \left((\mathbf{p}-\mathbf{a}) - \frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}}{\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{v}}(\mathbf{p}-\mathbf{a})\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}\left((\mathbf{p}-\mathbf{a}) - \frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}}{\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{v}}(\mathbf{p}-\mathbf{a})\right) \tag{12}$$

In this way, the upper Cholesky decomposition matrices factor out, precluding the need to calculate them. Finally, multiplying out the terms in parentheses and factoring yields a simplified expression for the squared Mahalanobis distance from the measured point to the best fit line

$$\|\mathbf{U}\mathbf{r}\|^2 = (\mathbf{p}-\mathbf{a})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{p}-\mathbf{a}) - \frac{\left(\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{p}-\mathbf{a})\right)^2}{\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{v}} \tag{13}$$

The sum of the expression for the squared Mahalanobis distance in Eq. (13) over all measured data points $\mathbf{p}_i$, $(i = 1 \ldots n)$ represents the 'sum of squares' to be minimized.

2.3. Probability density and likelihood functions

To demonstrate that minimizing the sum of the squares generates the maximum likelihood estimate for the line parameters, the expression in Eq. (13) can be used to build a probability density function for the each measurement, analogous to Eq. (3). The function

$$f(\mathbf{p}_i|\mathbf{a},\mathbf{v}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}_i|^{1/2}} \times \exp\left(-\frac{1}{2}\left[(\mathbf{p}_i-\mathbf{a})^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\mathbf{a}) - \frac{\left(\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\mathbf{a})\right)^2}{\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}}\right]\right) \tag{14}$$

describes the probability that a point in the vicinity of $\mathbf{p}_i$ belongs to the line $\mathbf{a}+\mathbf{v}t$.

Given a dataset of n independent measurements $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots \mathbf{p}_n)$ of the linear array $\mathbf{a}+\mathbf{v}t$, the joint density function $f(\mathbf{p}_1, \mathbf{p}_2, \ldots \mathbf{p}_n \,|\, \mathbf{a}, \mathbf{v}) = f(\mathbf{p}_1 \,|\, \mathbf{a}, \mathbf{v}) \cdot f(\mathbf{p}_2 \,|\, \mathbf{a}, \mathbf{v}) \cdot \ldots f(\mathbf{p}_n \,|\, \mathbf{a}, \mathbf{v})$ defines the probability function of the entire measured dataset $\mathbf{P}$ given any line parameters $\mathbf{a}$ and $\mathbf{v}$. The measured data are fixed, and we are instead interested in the likelihood function of the regression parameters given the measured data $\mathcal{L}(\mathbf{a}, \mathbf{v} \,|\, \mathbf{P})$, which is equal to the joint probability density function above:

$$\mathcal{L}(\mathbf{a}, \mathbf{v} \,|\, \mathbf{P}) = f(\mathbf{P} \,|\, \mathbf{a}, \mathbf{v}) = \prod_{i=1}^{n} f(\mathbf{p}_i|\mathbf{a}, \mathbf{v}) \tag{15}$$

Although the large product is not easily manipulated or differentiated, taking the natural logarithm converts the product of exponentials to a sum:

$$\ln \mathcal{L}(\mathbf{a}, \mathbf{v} \mid \mathbf{P}) = \sum_{i=1}^{n} \ln f(\mathbf{p}_i \mid \mathbf{a}, \mathbf{v}) \tag{16}$$

To avoid confusion with the traditional covariance matrix notation $\boldsymbol{\Sigma}$, all summation symbols will include their upper and lower limits explicitly, as above.

3. SOLVING FOR THE BEST FIT LINE

The best fit line maximizes the likelihood function $\mathcal{L}(\mathbf{a}, \mathbf{v} \mid \mathbf{P})$ of the regression parameters $\mathbf{a}$ and $\mathbf{v}$ given the measured data points $\mathbf{P}$. Taking the natural logarithm of the likelihood function is a monotone transformation, so that maximizing the log-likelihood $\ln \mathcal{L}(\mathbf{a}, \mathbf{v} \mid \mathbf{P})$ simultaneously maximizes the likelihood function itself. Substituting the probability density function for a point $\mathbf{p}_i$ on the line $\mathbf{a} + \mathbf{v}t$ in Eq. (14) into the likelihood function in Eq. (16) yields

$$\ln \mathcal{L}(\mathbf{a}, \mathbf{v} | \mathbf{P}) = -\frac{1}{2} \sum_{i=1}^{n} \left[(\mathbf{p}_i - \mathbf{a})^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} (\mathbf{p}_i - \mathbf{a}) - \frac{\left(\mathbf{v}^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} (\mathbf{p}_i - \mathbf{a})\right)^2}{\mathbf{v}^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} \mathbf{v}} + \ln |\boldsymbol{\Sigma}_i| + k \ln(2\pi) \right] \tag{17}$$

where $|\boldsymbol{\Sigma}_i|$ is the determinant of the measured covariance matrix.

The maximum likelihood estimate of the line parameters $\mathbf{a}$ and $\mathbf{v}$ for the line $\mathbf{a} + \mathbf{v}t$ maximizes the log-likelihood equation. This maximum occurs where the derivatives of Eq. (17) with respect to $\mathbf{a}$ and $\mathbf{p}$ are equal to zero. Differentiating yields a system of $2k$ equations, with the elements of $\mathbf{a}$ and $\mathbf{v}$ as $2k$ unknowns:

$$\frac{\partial \ln \mathcal{L}}{\partial \mathbf{a}} = \sum_{i=1}^{n} \frac{(\mathbf{p}_i - \mathbf{a})^{\mathsf{T}} \left(\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1} \mathbf{v} \mathbf{v}^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} \right)}{\mathbf{v}^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} \mathbf{v}} = \mathbf{0} \tag{18}$$

$$\frac{\partial \ln \mathcal{L}}{\partial \mathbf{v}} = \sum_{i=1}^{n} \frac{\mathbf{v}^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} \mathbf{v} \mathbf{v}^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} (\mathbf{p}_i - \mathbf{a}) \boldsymbol{\Sigma}_i^{-1} (\mathbf{p}_i - \mathbf{a}) - \left(\mathbf{v}^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} (\mathbf{p}_i - \mathbf{a})\right)^2 \boldsymbol{\Sigma}_i^{-1} \mathbf{v}}{\left(\mathbf{v}^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} \mathbf{v}\right)^2} = \mathbf{0} \tag{19}$$

where $\mathbf{0}$ is a vector of zeros with length k.

Although the system of Eqs. (18) and (19) contains $2k$ equations with $2k$ unknowns, the same line may be defined by any point $\mathbf{a}$ on the line and a direction vector $\mathbf{v}$ of any length. Thus the above system has an infinite number of solutions. A unique solution may be reached, for instance, by setting the first component of $\mathbf{a}$ to zero, the multi-dimensional analogue of a y-intercept, and setting the first component of $\mathbf{v}$ to one, the multi-dimensional analogue of the slope. With these substitutions, the above $2k$ equations in $2(k-1)$ variables will have a unique solution at a best fit $\mathbf{a}$ and $\mathbf{v}$.

However, for datasets that are far from the origin, setting a component of $\mathbf{a}$ to zero will result in a poorly conditioned covariance matrices for the line fit parameters, with highly correlated uncertainties between the components of $\mathbf{a}$ and $\mathbf{v}$. Numerical stability is improved by instead setting any one $\mathbf{a}$ component to its value in the multivariate weighed mean of the measured data $\mathbf{P}$. The weighted mean, derived in AppendixB, is also an excellent initial estimate for $\mathbf{a}$ in the numerical solution of Eqs. (18) and (19). Once a component of $\mathbf{v}$ has been specified, an initial estimate for each of the other direction vector components can be determined by ordinary least-squares regression on the dataset considered two dimensions at a time, always plotting on the x-axis the variable corresponding to the fixed component of $\mathbf{v}$.

The solution to Eqs. (18) and (19) may be reached quickly with a non-linear equation solver included in common mathematical software packages such as MATLAB. Alternately, the likelihood Eq. (17), and its first and second derivatives (Eqs. 18–19 and A.1–A.4) can be explicitly incorporated into a non-linear optimization scheme, such as the Levenberg–Marquardt algorithm (Levenberg, 1944; Marquardt, 1963). A MATLAB code and a Microsoft Excel Add-In that use this approach are included in the supplementary materials.

3.1. Uncertainty in fit parameters

The uncertainty in the best fit line parameters $\bar{\mathbf{a}}$ and $\bar{\mathbf{v}}$ are estimated by calculating the Fisher Information Matrix, $\mathcal{I}$, the negative second derivative of the likelihood function with respect to the best fit parameters. $\mathcal{I}$ takes the form

$$\mathcal{I} = -\begin{bmatrix} \frac{\partial^2 \ln \mathcal{L}}{\partial \bar{\mathbf{a}} \bar{\mathbf{a}}^{\mathsf{T}}} & \frac{\partial^2 \ln \mathcal{L}}{\partial \bar{\mathbf{a}} \bar{\mathbf{v}}^{\mathsf{T}}} \\ \frac{\partial^2 \ln \mathcal{L}}{\partial \bar{\mathbf{v}} \bar{\mathbf{a}}^{\mathsf{T}}} & \frac{\partial^2 \ln \mathcal{L}}{\partial \bar{\mathbf{v}} \bar{\mathbf{v}}^{\mathsf{T}}} \end{bmatrix} \tag{20}$$

Because the first components of vectors $\bar{\mathbf{a}}$ and $\bar{\mathbf{v}}$ have been specified explicitly, these components have zero uncertainty. For the other $k-1$ components of each vector, as well as the covariances between them, the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{av}}$ is calculated by removing the rows and columns 1 and $k+1$ corresponding to the specified components, generating a $2(k-1) \times 2(k-1)$ matrix, then calculating the matrix inverse of the shortened Fisher information matrix,

$$\boldsymbol{\Sigma}_{\mathbf{av}} = \mathcal{I}_r^{-1} \tag{21}$$

where $\mathcal{I}_r$ is $\mathcal{I}$ with the appropriate rows and columns removed. The same result is reached by calculating the pseudoinverse of $\mathcal{I}$ with the rows and columns corresponding to the specified components set to zero. The second derivatives in Eq. (20) are included in Appendix A.

To determine whether the uncertainties in the fit parameters are valid, or if the straight line model is reasonable for the observed data, a statistical goodness of fit test is required. The reduced chi-square (χ^2_{red}), known in the geochronology community as the MSWD or mean squared weighted deviation (Wendt and Carl, 1991), relates the average ratio of the squared offset between the measurement and the model to its variance, or expected value:

$$\chi^2_{red} = \frac{1}{df} \sum_{i=1}^{n} \frac{(\text{meas.} - \text{model})^2}{\sigma^2_{\text{meas.}}} \tag{22}$$

The expression inside the summation of Eq. (22) is analogous to Eq. (13), the squared distance from the measured point to the line weighted by its inverse covariance matrix. A line fit in any number of dimensions has $df = (k-1)(n-2)$ degrees of freedom, where $k-1$ represents the number of independent directions the measurement can deviate from the straight line and $n-2$ is the number of vector-valued measurements minus the two line fit parameters $\mathbf{a}$ and $\mathbf{v}$. Thus,

$$\chi^2_{red} = \frac{1}{(k-1)(n-2)} \sum_{i=1}^{n} \left[(\mathbf{p}_i - \mathbf{a})^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} (\mathbf{p}_i - \mathbf{a}) - \frac{\left(\mathbf{v}^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} (\mathbf{p}_i - \mathbf{a})\right)^2}{\mathbf{v}^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} \mathbf{v}} \right] \tag{23}$$

The expected value of χ^2_{red} is 1, with a probability density function that approaches a normal distribution with increasing degrees of freedom that has a standard deviation of $(2/df)^{1/2}$ (Wendt and Carl, 1991).

3.2. Other linear regression algorithms

The solutions to Eqs. (18) and (19) and the covariance matrix calculated with Eq. (21) can be shown to reduce to the two special cases of uncertainty-weighted line fitting currently used by the geochemistry and geochronology communities. A detailed history of parametric regression algorithms in earth sciences can be found in Howarth, 2001.

In a series of related papers, York, 1966,York, 1968 and York et al., 2004 offer incrementally improved algorithms for straight-line fitting in two dimensions by iteratively solving simultaneously for the MLE line parameters and uncertainties. Crucially, the objective equation presented here in Eq. (13) with $k = 2$ and $\mathbf{a}_1 = 0$, $\mathbf{v}_1 = 1$ simplifies to Eq. 2 of York, 1968. Note that the York, 1968 paper omits the covariance term between the uncertainties in the slope and y-intercept, although this is corrected in (York et al., 2004). This term becomes important when constructing confidence intervals for the regression line when the measured data is far from the origin.

When applied to linear models, the more general nonlinear regression algorithm presented in Sohn and Menke, 2002 is generalizable to higher dimensions, and is also equivalent to the MLE presented here. However, the narrow focus on straight line regression here results in an iterative algorithm that requires solving linear systems of rank $2k - 2$ instead of rank n at each step. This proffers faster convergence when using straight line regression as part of a larger, more computationally intensive algorithm, or with very large datasets such as the one presented in Section 4.

The maximum likelihood method used here is also employed by Titterington and Halliday, 1979 for two-dimensional linear regression through data with correlated uncertainties. The equations for the first and second derivatives of the likelihood equations in the Titterington and Halliday, 1979 appendix are special cases of Eqs. (18), (19) and Appendix A. York et al., 2004 show that the expressions for the uncertainties in the fit parameters for both the least-squares and MLE approach are equivalent.

The other multi-dimensional line fit algorithms commonly used in geochronology are for three-dimensional linear arrays, first described for ^{230}Th/U isochrons in Ludwig and Titterington, 1994 and utilized again for 'Total Pb/U Isochrons' in Ludwig, 1998. In the explicitly MLE-motivated former publication, the first (unnumbered) equation in the Appendix is equivalent to Eq. (17) for $k = 3$ and $\mathbf{a}_1 = 0$, $\mathbf{v}_1 = 1$, and the second partial derivatives given later are equivalent to the general case given here in Appendix A. In the latter publication, Eq. 33 is a special case of Eq. (13) above with the same constraints. Note the matrix form of the second partial derivatives given here in Appendix A avoid the rapidly rising number of partial derivative terms required when computing the partial derivatives separately for each line parameter. For instance, 18 separate partial second derivatives are not needed for $k = 4$, 32 for $k = 5$, and so on.

4. APPLICATION TO KINETIC ISOTOPE FRACTIONATION

Kinetic isotope fractionation between isotopologues, or isotopic molecules, occurs when a chemical reaction occurs irreversibly with different rate constants for each isotopologue (Young et al., 2002). The rate constants depend in large part on the effective masses of the isotopologues, and are often assumed proportional to the inverse square root of their effective mass (Bigeleisen, 1949; Young et al., 2002). In the case of evaporation of Pb from a hot filament in a thermal ionization mass spectrometer (TIMS), the reaction product Pb is instantaneously removed by ionization and acceleration by the mass spectrometer or by rapid pumping at high vacuum.

Thus, in addition to an instantaneous isotopic fractionation due to differing rate constants, there is a time-progressive depletion of the species with the faster rate constant (usually the lighter isotopologue), known as a reservoir effect. The two assumptions, that the ratio of the reaction constants depend on the inverse square root of the mass but not on time or temperature, and that Pb evaporated from the filament is instantaneously removed, generate a differential equation that relates the quantity of sample consumed to the isotopic composition of the instantaneous reaction products, which are analyzed by TIMS when ionized. The solution to this differential equation is known as the 'Rayleigh law' (Russell et al., 1978; Habfast, 1998).

4.1. Rayleigh and exponential fractionation equations

The Rayleigh law for two isotopes a and b with masses M_a and M_b is

$$\left(\frac{a}{b}\right)_{meas} = \left(\frac{a}{b}\right)_{corr} \cdot \sqrt{\frac{M_b}{M_a}} \cdot \left(\frac{b}{b_0}\right)^{\sqrt{M_b/M_a}-1} \tag{24}$$

where *corr* and *meas* denote true (fractionation-corrected) and measured isotope ratios, and (b/b_0) is the fraction of isotope b remaining on the filament.

Not all observations by TIMS of fractionating samples follow the predictions made by the Rayleigh law (e.g., Russell et al., 1978), perhaps because ions are being generated from multiple reservoirs on the filament that are depleted to differing extents. Additionally, the progressive depletion model does not apply to a constantly refreshed reservoir, as for inductively-coupled plasma mass spectrometry (ICP-MS). A popular (e.g., Boyet and Carlson, 2006; Amelin et al., 2009) empirical law, which is also an approximate solution to the Rayleigh equation, is the exponential fractionation law (Russell et al., 1978; Habfast, 1998; Albarède et al., 2004). The exponential fractionation equation for the same scenario above is

$$\left(\frac{a}{b}\right)_{corr} = \left(\frac{a}{b}\right)_{meas} \cdot \left(\frac{M_b}{M_a}\right)^{\beta} \quad (25)$$

where β is the mass fractionation factor, which may change during the course of an analysis and vary between analyses.

For both fractionation laws, it is possible to derive a linear equation relating the fractionation-corrected log-ratio of isotopes to the measured log-ratio. For the Rayleigh law, taking the natural logarithm of both sides and rearranging terms yields

$$\ln\left(\frac{a}{b}\right)_{corr} = \ln\left(\frac{a}{b}\right)_{meas} - \frac{1}{2}\ln\frac{M_b}{M_a} - \left(\sqrt{\frac{M_b}{M_a}} - 1\right) \cdot \ln\left(\frac{b}{b_0}\right) \quad (26)$$

and for the exponential law,

$$\ln\left(\frac{a}{b}\right)_{corr} = \ln\left(\frac{a}{b}\right)_{meas} + \beta \ln\left(\frac{M_b}{M_a}\right) \quad (27)$$

Three-isotope systems whose fractionation is governed by either the Rayleigh or exponential law (i.e., changing (b/b_0) or β) will generate linear arrays when plotted as $\ln(a/b)_{meas}$ vs. $\ln(c/b)_{meas}$. This relationship is often exploited to assess or correct the effects of isotopic fractionation (Maréchal et al., 1999; Bizzarro et al., 2011). For the simpler exponential fractionation equation, the line can be expressed parametrically as passing through the log-ratio of the corrected isotopic composition (IC) $\langle \ln(c/b)_{corr}, \ln(a/b)_{corr}\rangle$ with a two-component direction vector $\langle \ln(M_b/M_c), \ln(M_b/M_a)\rangle$ and fractionation factor β, where $\langle\rangle$ denotes a transposed column vector. It follows that the log-ratio of the corrected IC is analogous to **a** above, the direction vector, which depends on the ratios of the isotopic masses has its analog in **v**, and β becomes the location parameter t.

For the Rayleigh law, a plot of $\ln(a/b)_{meas}$ vs. $\ln(c/b)_{meas}$ passes through the point $\langle \ln(c/b) + \frac{1}{2}\ln(M_b/M_c), \ln(a/b) + \frac{1}{2}\ln(M_b/M_a)\rangle$ with a direction vector $\langle \sqrt{M_b/M_c} - 1, \sqrt{M_b/M_a} - 1\rangle$. Unlike the exponential law, the Rayleigh law does not predict that the observed fractionation line will go through the true IC of the sample, and the predicted slopes for the two are different. If the true IC is known to sufficient absolute precision, this property may be used to distinguish between the two laws. In the absence of a well-known standard IC, the difference in the predicted slopes provides a sensitive test for the applicability of each law.

The linear equations above may be scaled up to systems with any number of isotopes (e.g., Sr, Ti, Nd, Pb), producing straight lines in log-ratio space with dimension k equal to the number of isotopes present minus one. Isotopic fractionation behavior can be evaluated by calculating the means $\mathbf{p}_i$ of log-ratios to a common isotope along with their covariance matrices $\mathbf{\Sigma}_i$, then fitting a straight line to the data. It is important to note that if $\mathbf{p}_i$ and $\mathbf{\Sigma}_i$ are estimated from multiple measurements, then mean and covariance calculations must be performed *after* the log-ratios are evaluated. In other words, one must calculate the mean and covariance matrix of the log-ratios rather than the logarithm of the mean and covariance matrix of the measured ratios. Using a common compositional variable (e.g., an isotope abundance) as the denominator of log-ratios is the additive log-ratio transform (alr) of Aitchison, 1986, a well-established analytical technique for compositional data discussed in Section 5.2.

4.2. Application to Pb measurements by TIMS

Precise isotope ratio measurements of Pb by TIMS are used for high precision geo- and cosmochronology (Amelin et al., 2009; Davydov et al., 2010), geochemistry (Fekiacova et al., 2007; Abouchami et al., 2005), and isotope standard inter-calibration (Thirlwall, 2000; Doucelance and Manhès, 2001). To empirically evaluate isotopic fractionation of Pb during TIMS analysis, a large ($\sim$ 50 ng) sample of the standard NBS 982 (Catanzaro et al., 1968) was loaded with a silica gel activator (Gerstenberger and Haase, 1997) on a Re filament and run to exhaustion. The ion beams of ^{204}Pb, ^{206}Pb, ^{207}Pb, and ^{208}Pb were measured on a static Faraday array in 100-millisecond integrations, alternating between 40 seconds baselines measured at half-mass and 75 seconds on-peak. The entire run consists of 51000 on-peak and 27200 baseline integrations, provided in the Electronic annex EA-1.

After correcting for baselines and propagating baseline uncertainties, the natural logarithms of the ^{204}Pb, ^{207}Pb, and ^{208}Pb intensity divided by the ^{206}Pb intensity were evaluated for each integration. The mean and covariance matrix of every 25 log-ratios were calculated for a total of $n = 2040$ $\mathbf{p}_i$ and $\mathbf{\Sigma}_i$, where each $\mathbf{p}$ has the form $\langle \log(^{204}Pb/^{206}Pb), \log(^{207}Pb/^{206}Pb), \log(^{208}Pb/^{206}Pb)\rangle$. The short integration times and limited number of integrations per mean (2.5 s total per data point) attempts to separate the covariance due to random signal noise from the linear fractionation trend. The three-dimensional measured data, plotted as a pair of two-dimensional projections, is presented in Fig. 3.

The solution to the system of Eqs. (18) and (19) was calculated with the commercial mathematical software MATLAB using the built-in 'fsolve' function, which employs the Levenberg–Marquardt algorithm (Levenberg, 1944; Marquardt, 1963). The MATLAB code is provided as Electronic annex EA-2. For this calculation with $n = 2040$ means in $k = 3$ dimensions, the code takes $\sim$ 20 s to compile on a 2.4 GHz laptop computer. The maximum likelihood estimate of the regression parameters **a** and **v** and their uncertainties are listed in Table 1 and the best fit line is plotted in green in Fig. 3a, along with its 2σ uncertainty envelope in Fig. 3b.

For a straight line fit, one component of **a** and **v** must be explicitly specified, and the other components are calculated using these values. Here, the values of $\mathbf{a}_3$ and $\mathbf{v}_3$ were chosen to be 0.00016 and 1, respectively, so that **a** corresponds to the point on the observed line with a ^{208}Pb/^{206}Pb of 1.00016 (Catanzaro et al., 1968) and the first two direction vector components are the slopes of the first two variables when plotted against the third. Thus the first and second components of **a** and **v** are conditional upon the explicitly specified value of the third component.

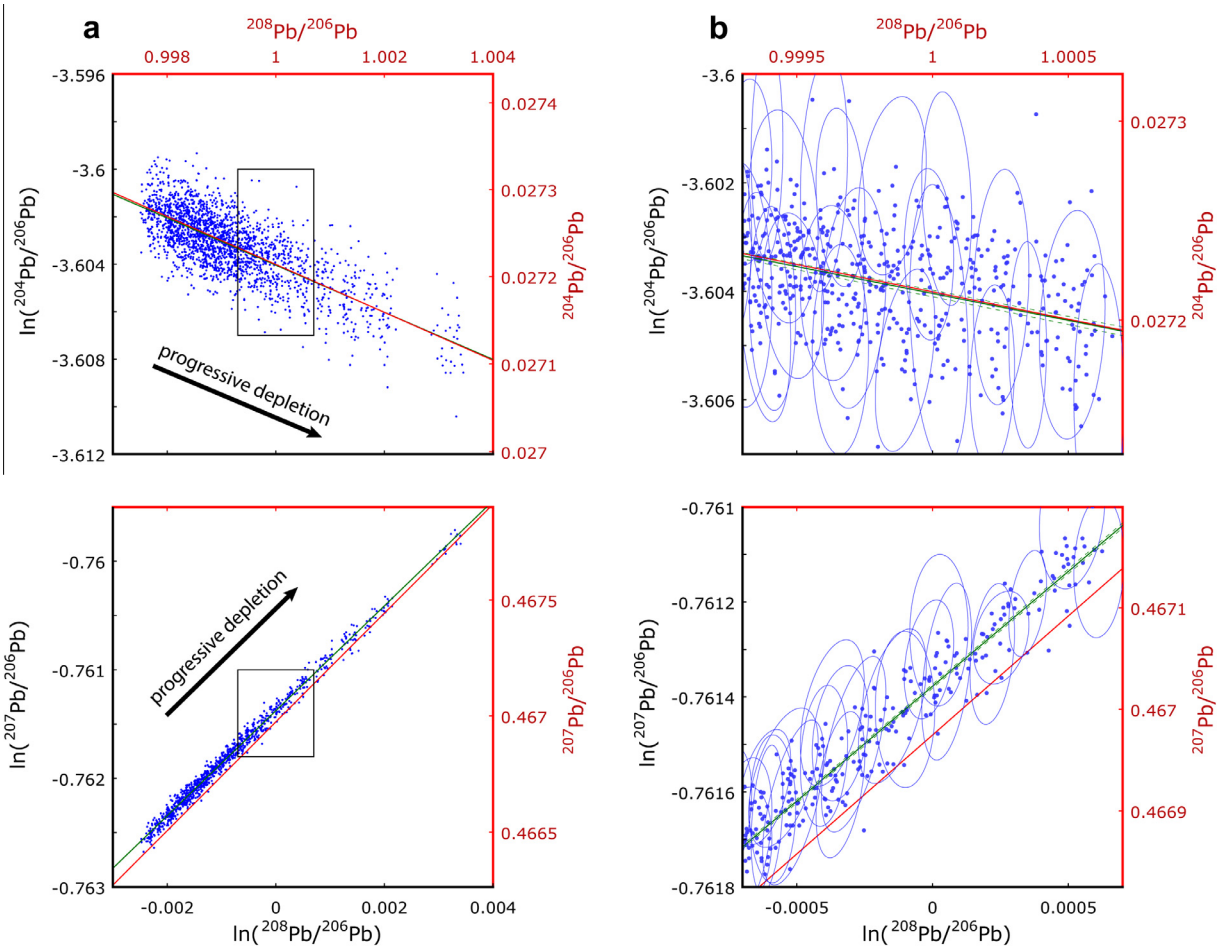

Fig. 3. Observed isotopic fractionation for a single analysis of Pb standard NBS 982, measured by TIMS using a silica gel activator. Data is plotted in log-ratio space (black axes), with isotope ratio values given on the opposite axis in red. (a) Plots of all measured data, as log-ratios $\log(^{204}Pb/^{206}Pb)$ and $\log(^{207}Pb/^{206}Pb)$ vs. $\log(^{208}Pb/^{206}Pb)$. Each blue measured value represents the mean of 25 consecutive 100-millisecond integrations. The dark green line is the best fit line to the three-dimensional data. For reference, the mass-dependent fractionation line predicted by the exponential equation is plotted in red. While the observed components of the direction vector in ^{204}Pb–^{206}Pb–^{208}Pb space, involving all even isotopes, is the same as the predicted MDF line, the $\log(^{207}Pb/^{206}Pb)$ slope component diverges significantly. Arrows indicate the trend in isotopic composition through time due to progressive depletion of the sample, and the black box marks the plot limits in (b). **(b)** A closeup of data in (a), plotted with the same conventions. A 2σ uncertainty ellipse is displayed for every twenty-fifth measured mean, with the others omitted for clarity. The 2σ uncertainty envelope for the best fit line is plotted as green dashed curves. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Likewise, the covariance matrix $\mathbf{\Sigma_{av}}$ describes the conditional probability distribution of the unknown parameters, here $\mathbf{a}_1, \mathbf{a}_2, \mathbf{v}_1$, and $\mathbf{v}_2$, given $\mathbf{a}_3 = 0.00016$ and $\mathbf{v}_3 = 1$. Changing the specified values of $\mathbf{a}_3$ and $\mathbf{v}_3$ would change both the values and uncertainties of the new best fit line parameters, but not the location of the best fit line or its uncertainty envelope.

The reduced chi-square statistic (χ^2_{red}, Eq. 23) for this linear fit is 1.197 for $(3-1)(2040-2) = 4076$ degrees of freedom. The predicted standard deviation of χ^2_{red} is $(2/4076)^{1/2} \approx 0.022$, so that the observed χ^2_{red} is significantly greater than its expected value of 1. Possible reasons for the elevated χ^2_{red} include heavy-tailed non-Gaussian probability distributions for the measured data or non-linear evolution of the log-ratios, such as intermittent presence of isobaric interferences or subtly changing fractionation behavior with time. Although the latter possibilities are difficult to rule out, the former is altogether likely considering the dearth of detailed noise characterization of mass spectrometer measurements, especially at the 100 ms conversion rate of the analog-digital converter. In this case, the robust linear regression algorithm for $k = 2$ dimensions developed by Powell et al., 2002 could be adapted to a higher-dimensional straight line model, such as that given by Sohn and Menke, 2002. Because the χ^2_{red} is so close to unity, inflating the uncertainties by $\sqrt{1.197} = 1.094$ or using a robust regression technique will not change the conclusions that follow.

In the upper two plots of Fig. 3, which are projections of the data onto the $\log(^{204}Pb/^{206}Pb)$ vs. $\log(^{208}Pb/^{206}Pb)$ plane, the best fit line agrees with the predicted mass-dependent fractionation (MDF) line using both the exponential or Rayleigh laws (see also Table 1). However, the observed

Table 1
Maximum likelihood estimates, uncertainties, and correlation coefficients for the best fit line parameters to the NBS 982 Pb fractionation data.

	Value	$\pm 2\sigma$ (*abs*)	Exp.[a]	Ray.[b]
$\mathbf{a}_1$	−3.604198	0.000061		
$\mathbf{a}_2$	−0.7613001	0.0000041		
$\mathbf{a}_3$[c]	*0.00016*	–		
$\mathbf{v}_1$	−0.993	0.040	−1.009	−1.014
$\mathbf{v}_2$	0.4836	0.0027	0.5014	0.5020
$\mathbf{v}_3$[c]	*1*	–		

Matrix of correlation coefficients for uncertainties in fit parameters

	$\mathbf{a}_1$	$\mathbf{a}_2$	$\mathbf{v}_1$	$\mathbf{v}_2$
$\mathbf{a}_1$	1			
$\mathbf{a}_2$	0.032	1		
$\mathbf{v}_1$	0.726	0.021	1	
$\mathbf{v}_2$	0.021	0.721	0.028	1

[a] Value predicted by exponential law.
[b] Value predicted by Rayleigh law.
[c] Value set explicitly.

and mass-dependent fractionation lines diverge significantly in the $\log(^{207}Pb/^{206}Pb)$ vs. $\log(^{208}Pb/^{206}Pb)$ projection.

The horizontal or vertical position of the mass-dependent fractionation line is uncertain for a NBS 982 Pb analysis: the absolute uncertainty of the IC is known to only to ca. 180 ppm (2σ) (Catanzaro et al., 1968), and the analysis itself necessarily includes a loading blank of uncertain mass and isotopic composition. Additionally, the sample was run as a static analysis, and no effort was made to cancel small differences in Faraday collector or amplifier efficiencies. None of these variables affect the predicted slope of the fractionation line, which depends on the masses of the isotopes alone (Albarède and Beard, 2004). Thus, the systematic difference in odd mass number ^{207}Pb behavior relative to the even masses can be reliably attributed to mass independent fraction (MIF) during the analysis.

MIF has been described in several heavy element isotopic systems, such as Cd, Hf, Hg, Tl, and U in laboratory and natural settings (see references in Epov et al., 2011; Thiemens et al., 2012), and may be the result of nuclear field shift (Bigeleisen, 1996; Fujii et al., 2009) or nuclear spin (Buchachenko, 1995) effects. Both mechanisms predict systematic differences in the fractionation of odd isotopes over even ones, as observed here. The magnitude of the nuclear field shift effect is inversely proportional to the absolute temperature of the system, which is typically > 1400 K during TIMS analysis, implicating the nuclear spin effect. However, the reactions responsible for producing Pb^+ ions from a silica gel activator are complex (Kessinger et al., 2001), and no theoretical predictions yet exist for the magnitude of the nuclear field shift effect (e.g., Schauble, 2007).

For large (> 1 ng) Pb loads analyzed with a silica gel activator by TIMS, MIF or an 'odd-even effect' has been postulated by several authors, including Thirlwall, 2000, Doucelance and Manhès, 2001, and most recently by Amelin et al., 2005, who also incorporate a ^{202}Pb–^{205}Pb tracer. In contrast to the data reported in these references, the magnitude of mass-independent fractionation observed here is constant during the course of the analysis, reflected by the linear fractionation trend. Thus the fractionation correction equations for log-ratios need not be abandoned, and may be amended with an additional term of the form

$$\ln\left(\frac{a}{b}\right)_{corr} = \ln\left(\frac{a}{b}\right)_{meas} + \beta\left[\ln\left(\frac{M_b}{M_a}\right) + \ln\left(\frac{\gamma_b}{\gamma_a}\right)\right] \quad (28)$$

where γ describes the ratio between an isotope's apparent atomic mass on the fractionation line and its true mass. The observed NBS 982 Pb fractionation trend is fit by this modified exponential law with $\gamma_{204Pb} = \gamma_{206Pb} = \gamma_{208Pb} = 1$ and $\gamma_{207Pb} = 0.999828$.

Accurate fractionation correction for precise Pb measurements by TIMS requires both a mass-dependent and a mass-independent correction. This may be true especially for Pb measurements made with a ^{205}Pb tracer, an additional odd-mass isotope. If the values of γ_{205Pb} and γ_{207Pb} are observed to vary between analyses, this additional uncertainty should be considered when interpreting high-precision datasets.

5. ISOCHRONS, MIXING LINES, AND THE STATISTICS OF COMPOSITIONAL DATA

The exponential and Rayleigh fractionation laws presented in Section 4 predict linear trends in log-ratio coordinates, but contain exponential terms in isotope ratio coordinates. In contrast, isotopic mixing lines and isochron relationships have linear trends only in the (non-logged) isotope ratio coordinates. Fitting a straight line to these models thus requires evaluating means and covariance matrices and minimizing weighted distances in isotope ratio space. There are a number of fundamental problems with all three calculations, which are outlined in full in Aitchison, 1986, as well as (e.g., Aitchison, 1984,Aitchison, 1992; Pawlowsky-Glahn and Egozcue, 2006), but have been largely ignored by the isotope geochemistry community.

The basic problem is that within a set of isotope ratios or compositional components is an implicit constraint: the sum of the (non-negative) fractional components, is equal to 1, or 100%. Recognizing the unit sum constraint has led to a rich literature on compositional data analysis, and ignoring this constraint manifests itself in several intuitively unpredictable ways, detailed in Aitchison, 1986, with two symptoms outlined briefly below. The non-linear regression algorithm required by isochron and convex mixing models is beyond the scope of this work.

5.1. Some pitfalls of conventional multivariate statistics for isotope ratios

Consider evaluating the arithmetic means $\bar{r}_1$ and $\bar{r}_2$ of n measurements of the isotope ratios $r_1 = (a/b)$ and $r_2 = (c/b)$. For the same n measurements, the arithmetic mean $\bar{r}_3$ of $r_3 = (a/c)$ is not $\bar{r}_1/\bar{r}_2$. Therefore, any statistical inference made, for instance an age derived from an isochron, will depend on the arbitrary choice of numerator and denominator isotopes when the ratios are measured: expressing the system in terms of r_2 and r_3 instead of r_1 and r_2 or their reciprocals would yield different results. Note that the same problem applies to ratios expressed in ϵ or δ notation.

This problem does not extend to the log-ratios utilized in Section 4. For instance, the arithmetic means $\bar{R}_1$ and $\bar{R}_2$ of n measured log-ratios $R_1 = \log(a/b)$ and $R_2 = \log(c/b)$ can be subtracted to calculate the mean $\bar{R}_3$ of $R_3 = \log(a/c)$. Transforming back to (non-logged) isotope ratio space by evaluating $\hat{r}_1 = \exp(\bar{R}_1), \hat{r}_2 = \exp(\bar{R}_2)$, and $\hat{r}_3 = \exp(\bar{R}_3)$ generates a self-consistent set of means, so that $\hat{r}_1/\hat{r}_2 = \hat{r}_3$. Because $\bar{R}$ is the arithmetic mean of the measured log-ratios, $\hat{r}$ is the geometric mean of the isotope ratios.

Using the covariance matrix of the measured isotope ratios to define a multivariate normal probability distribution function (e.g., Ludwig, 2003; McLean et al., 2011) has several related shortcomings. First, the multivariate normal distribution assigns a finite probability over all positive and negative real numbers, which is inappropriate for isotope ratios that are assumed to always be greater than or equal to zero. Although this effect is negligible for analyses with small uncertainties far from zero, it is more pronounced for measurements with larger uncertainties and low absolute values. For instance, the inverse Pb isochron in Fig. 2 of Amelin et al., 2009 features several highly radiogenic analyses with large leverage on the Pb–Pb date whose 2σ uncertainty ellipses overlap negative $^{204}Pb/^{206}Pb$ values.

In contrast, log-ratios are permitted to be negative, since the exponential function that transforms back to (non-logged) isotope ratios generates only positive values. The assumption of a normal distribution for the log-ratios is equivalent to assuming a log-normal distribution for the isotope ratios themselves. This difference becomes smaller for increasing precision and distance from a zero value, and the distinction for data near detection limits merits further investigation with empirical data. However, the fact that the log-normal distribution is defined only over positive values represents a step in the right direction.

5.2. Application of compositional data methods to exponential mass fractionation

The exponential fractionation correction Eqs. (25) and (27) in Section 4.1 are cumbersome, especially when the presence of three or more isotopes necessitate a system of pairwise equations. However, the same relationship for all isotopes can also be written succinctly as a linear function of compositional vectors, using the operators appropriate for compositional data:

$$\mathbf{x}_{corr} = \mathbf{x}_{meas} \oplus (\beta \otimes \mathbf{mass}) \tag{29}$$

In this equation, both $\mathbf{x}_{meas}$ and **mass** are compositional data vectors, which express the relative proportion of each component and sum to a constant, for instance 1 or 100%. Thus a typical measured Pb IC $\langle ^{204}Pb/^{206}Pb, ^{207}Pb/^{206}Pb, ^{208}Pb/^{206}Pb\rangle$ of $\langle 0.0273, 0.4668, 0.9989\rangle$ from Section 4 becomes $\mathbf{x}_{meas} \approx \langle 0.011, 0.4011, 0.1872, 0.4007\rangle$, where the four components correspond to the fractional abundances of ^{204}Pb, ^{206}Pb, ^{207}Pb, and ^{208}Pb.

Normalization to a constant sum, in this case unity, performs the same function as evaluating ratios instead of utilizing absolute intensities or abundances. Only the ratios of the isotopic masses are used in Eqs. (25) and (27) as well, and so they too form a compositional data vector and can be normalized to the same constant. Note that the arbitrary fractionation factor β used in Eq. (29) has the opposite sign as that used in Eqs. (25) and (27), so that the same corrected Pb IC would be calculated using for instance $\beta = 1.4$ in Eq. (25) and $\beta = -1.4$ in Eq. (29).

The two operations $\oplus$ and $\otimes$ in Eq. (29), known as perturbation and power transformation respectively, are the compositional analogues of addition and multiplication. As defined by Aitchison, 1986 for D total compositional components, they are

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}\langle x_1 y_1, x_2 y_2, \ldots x_D y_D\rangle \tag{30}$$

$$\beta \otimes \mathbf{y} = \mathcal{C}\langle y_1^{\beta}, y_2^{\beta}, \ldots y_D^{\beta}\rangle \tag{31}$$

where the closure operator $\mathcal{C}$ simply normalizes the resulting vector components to a constant sum.

Transforming the compositional data vectors into log-ratios with a common denominator (e.g., ^{206}Pb in the example) is the additive log-ratio transform (alr) of Aitchison, 1986, which is familiar to isotope geochemists as the natural sample space for plotting exponential fractionation trends as straight lines.

After the alr transformation, the analogy of the perturbation and power transform to addition and multiplication become clear. For instance, the $[\log(M_b/M_a) + \log(\gamma_b/\gamma_a)]$ term in Eq. (28) can be rewritten as an expression for a modified 'effective' mass vector for isotopes undergoing mass independent fractionation,

$$\mathbf{mass}' = \mathbf{mass} \oplus \mathbf{gamma} \tag{32}$$

Multiplying **mass**′ by β in Eq. (28) becomes a power transform, so that a modified form of the exponential equation, including mass independent effects, is

$$\mathbf{x}_{corr} = \mathbf{x}_{meas} \oplus (\beta \otimes \mathbf{mass}') \tag{33}$$

Thus, the complicated system of equations needed to describe the behavior of multiple isotopes exhibiting mass-independent fractionation takes an elegant linear form, analogous to the line $\mathbf{a} + \mathbf{v}t$. Here, the **a** is the measured isotopic composition represented by the $\mathbf{x}_{meas}$, the slope vector **v** is the effective masses **mass**′, and the position parameter t is the fractionation parameter β.

6. SUMMARY AND CONCLUSIONS

Although a number of straight line regression algorithms are used by the geochemical community, none treat the general case of data with correlated uncertainties in two or more dimensions. The best fit line parameters are those that maximize their likelihood function, and are found by setting the derivative of the log-likelihood function to zero and solving this system of nonlinear equations. Uncertainties in the regression parameters are then determined using the second derivatives of the log-likelihood equation.

Isotopic fractionation is often modeled by linear functions in log-ratio space, such as the Rayleigh and exponential laws. An experiment testing their applicability to Pb analysis by TIMS shows that progressive fractionation during an analysis is well fit by a linear trend. Although the mass-dependent fractionation prediction made by the linear and Rayleigh laws hold for even mass number Pb isotopes,

^{207}Pb exhibits constant-magnitude mass-independent fractionation throughout the analysis, described by a modified exponential fractionation law.

For compositional data, the linear regression algorithm is valid only when the data are treated as log-ratios. Models that are not linear in this space, such as isochrons and mixing lines, cannot be fit by a line because self-consistent means, covariance matrices, and distances cannot be calculated in isotope ratio space. The established field of compositional data analysis can be used to extend the findings herein.

ACKNOWLEDGMENTS

This manuscript benefited greatly from discussion and comments by S. Bowring and B. Schoene, helpful reviews by J. Rudge, P. Vermeesch, and D. Davis, and from expert editorial input from Y. Amelin. This research was supported by NSF Grants EAR 0451802 (The EARTHTIME Network) and EAR 0930166.

APPENDIX A. SECOND DERIVATIVES OF LOG-LIKELIHOOD EQUATION FOR BEST FIT LINE

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \mathbf{a}\mathbf{a}^{\mathrm{T}}} = \sum_{i=1}^{n} \frac{\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}}{\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}} - \boldsymbol{\Sigma}_i^{-1} \tag{A.1}$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \mathbf{a}\mathbf{v}^{\mathrm{T}}} = \sum_{i=1}^{n} \frac{2(\mathbf{p}_i-\mathbf{a})^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1} - \mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\left(\boldsymbol{\Sigma}_i^{-1}\mathbf{v}(\mathbf{p}_i-\mathbf{a})^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1} + (\mathbf{p}_i-\mathbf{a})^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\boldsymbol{\Sigma}_i^{-1}\right)}{\left(\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\right)^2} \tag{A.2}$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \mathbf{v}\mathbf{a}^{\mathrm{T}}} = \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \mathbf{a}\mathbf{v}^{\mathrm{T}}}\right)^{\mathrm{T}} \tag{A.3}$$

$$\begin{aligned}\frac{\partial^2 \ln \mathcal{L}}{\partial \mathbf{v}\mathbf{v}^{\mathrm{T}}} = \sum_{i=1}^{n} \Big[&\left(\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\right)^2 \left(2\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\mathbf{v})\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\mathbf{a})\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\right.\\ &+ \mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\mathbf{a})(\mathbf{p}_i-\mathbf{a})^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1} - \left(\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\mathbf{a})\right)^2\boldsymbol{\Sigma}_i^{-1}\\ &- 2\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\mathbf{a})\boldsymbol{\Sigma}_i^{-1}\mathbf{v}(\mathbf{p}_i-\mathbf{a})^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\Big)\\ &- 4\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\left(\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\mathbf{a})\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\mathbf{a})\right.\\ &\left.\left.- \left(\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\mathbf{a})\right)^2\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\right)\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\Big]\right.\\ &/\left(\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{v}\right)^4\end{aligned} \tag{A.4}$$

APPENDIX B. MULTIVARIATE WEIGHTED MEAN

Eq. (3) is the multivariate probability density function at any point $\mathbf{x}$ in the vicinity of the measurement $\mathbf{p}_i$ with covariance matrix $\boldsymbol{\Sigma}_i$. For a dataset of n measurements $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots \mathbf{p}_n)$, the multivariate weighted mean $\hat{\mathbf{p}}$ can be derived using a maximum likelihood calculation similar to that of the best fit line. Using the probability distribution function in (3) to form a likelihood function for $\hat{\mathbf{p}}$ in analogy to Eq. (3)

$$\begin{aligned}\ln \mathcal{L}(\hat{\mathbf{p}} \,|\, \mathbf{P}, \boldsymbol{\Sigma}) &\propto \sum_{i=1}^{n} \ln f(\mathbf{p}_i, \boldsymbol{\Sigma}_i \mid \hat{\mathbf{p}})\\ &= -\frac{1}{2}\sum_{i=1}^{n}\left[(\mathbf{p}_i-\hat{\mathbf{p}})^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\hat{\mathbf{p}})\right]\end{aligned} \tag{B.1}$$

Setting the derivative of the log-likelihood function with respect to $\hat{\mathbf{p}}$ to zero and solving for the maximum likelihood estimate of $\hat{\mathbf{p}}$ yields

$$\frac{\partial \ln \mathcal{L}}{\partial \hat{\mathbf{p}}} = \sum_{i=1}^{n} \boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\hat{\mathbf{p}}) = 0 \tag{B.2}$$

$$\hat{\mathbf{p}} = \left[\sum_{i=1}^{n}\left(\boldsymbol{\Sigma}_i^{-1}\right)\right]^{-1}\left[\sum_{i=1}^{n}\left(\boldsymbol{\Sigma}_i^{-1}\,\mathbf{p}_i\right)\right] \tag{B.3}$$

This explicit estimate for the multivariate weighted mean is a special case of the implicitly defined weighted mean in the presence of overdispersion in Vermeesch, 2010.

The covariance matrix for $\hat{\mathbf{p}}$ is the inverse of its Fisher Information Matrix $\mathcal{I}$,

$$\mathcal{I} = -\frac{\partial^2 \ln \mathcal{L}}{\partial \hat{\mathbf{p}}\hat{\mathbf{p}}^{\mathrm{T}}} \tag{B.4}$$

$$\boldsymbol{\Sigma}_{\hat{\mathbf{p}}} = \mathcal{I}^{-1} = \left[\sum_{i=1}^{n}\left(\boldsymbol{\Sigma}_i^{-1}\right)\right]^{-1} \tag{B.5}$$

This choice of $\hat{\mathbf{p}}$ minimizes

$$\chi^2_{red} = \frac{1}{k(n-1)}\sum_{i=1}^{n}\left[(\mathbf{p}_i-\hat{\mathbf{p}})^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_i-\hat{\mathbf{p}})\right] \tag{B.6}$$

the expression for the reduced chi-square sum, also known as the MSWD.

REFERENCES

Abouchami W., Hofmann A. W., Galer S. J. G., Frey F. A., Eisele J. and Feigenson M. (2005) Lead isotopes reveal bilateral asymmetry and vertical continuity in the Hawaiian mantle plume. *Nature* **434**(7035), 851–856.

Aitchison J. (1984) The statistical analysis of geochemical compositions. *Math. Geol.* **16**, 531–564.

Aitchison J. (1986) *The Statistical Analysis of Compositional Data*. Chapman & Hall Ltd., London.

Aitchison J. (1992) On criteria for measures of compositional difference. *Math. Geol.* **24**, 365–379.

Albarède F. and Beard B. (2004) Analytical methods for non-traditional isotopes. *Rev. Mineral. Geochem.* **55**(1), 113–152.

Albarède F., Telouk P., Blichert-Toft J., Boyet M., Agranier A. and Nelson B. (2004) Precise and accurate isotopic measurements using multiple-collector ICPMS. *Geochim. Cosmochim. Acta* **68**(12), 2725–2744.

Amelin Y., Connelly J., Zartman R. E., Chen J. H., Göpel C. and Neymark L. A. (2009) Modern U–Pb chronometry of meteorites: advancing to higher time resolution reveals new problems. *Geochim. Cosmochim. Acta* **73**(17), 5212–5223.

Amelin Y., Davis D. W. and Davis W. J. (2005) Decoupled fractionation of even- and odd-mass isotopes of Pb in TIMS. *Geochim. Cosmochim. Acta* **69**(10), A215.

Balta J. B., Asimow P. D. and Mosenfelder J. L. (2011) Manganese partitioning during hydrous melting of peridotite. *Geochim. Cosmochim. Acta* **75**(20), 5819–5833.

Bigeleisen J. (1949) The relative reaction velocities of isotopic molecules. *J. Chem. Phys.* **17**(8), 675–678.

Bigeleisen J. (1996) Nuclear size and shape effects in chemical reactions. isotope chemistry of the heavy elements. *J. Am. Chem. Soc.* **118**(15), 3676–3680.

Bizzarro M., Paton C., Larsen K., Schiller M., Trinquier A. and Ulfbeck D. (2011) High-precision Mg-isotope measurements of terrestrial and extraterrestrial material by HR-MC-ICPMS-implications for the relative and absolute Mg-isotope composition of the bulk silicate Earth. *J. Anal. At. Spectrom.* **26**, 565–577.

Boyet M. and Carlson R. W. (2006) A new geochemical model for the Earth's mantle inferred from ^{146}Sm–^{142}Nd systematics. *Earth Planet. Sci. Lett.* **250**(1–2), 254–268.

Buchachenko A. L. (1995) MIE versus CIE: comparative analysis of magnetic and classical isotope effects. *Chem. Rev.* **95**(7), 2507–2528.

Catanzaro E. J., Murphy T. J., Shields W. R. and Garner E. L. (1968) Absolute isotopic abundane ratios of common, equal-atom, and radiogenic lead standards. *J. Res. Nat. Burea Stand. – A. Phys. Chem.* **72A**(3), 261–267.

Davydov V. I., Crowley J. L., Schmitz M. D. and Poletaev V. I. (2010) High-precision U–Pb zircon age calibration of the global Carboniferous time scale and Milankovitch band cyclicity in the Donets Basin, eastern Ukraine. *Geochem. Geophys. Geosyst.* **11**(2).

Doucelance R. and Manhès G. (2001) Reevaluation of precise lead isotope measurements by thermal ionization mass spectrometry: comparison with determinations by plasma source mass spectrometry. *Chem. Geol.* **176**(1–4), 361–377.

Egozcue J., Pawlowsky-Glahn V., Mateu-Figueras G. and Barceló-Vidal C. (2003) Isometric logratio transformations for compositional data analysis. *Math. Geol.* **35**(3), 279–300.

Epov V. N., Malinovskiy D., Vanhaecke F., Bégué D. and Donard O. F. (2011) Modern mass spectrometry for studying mass-independent fractionation of heavy stable isotopes in environmental and biological sciences. *J. Anal. At. Spectrom.* **26**, 1142–1156.

Fekiacova Z., Abouchami W., Galer S. J. G., Garcia M. O. and Hofmann A. W. (2007) Origin and temporal evolution of Koolau Volcano, Hawaii: inferences from isotope data on the Koolau Scientific Drilling Project (KSDP), the Honolulu volcanics and ODP site 843. *Earth Planet. Sci. Lett.* **261**(1–2), 65–83.

Ferrier K. L., Kirchner J. W. and Finkel R. C. (2011) Estimating millennial-scale rates of dust incorporation into eroding hillslope regolith using cosmogenic nuclides and immobile weathering tracers. *J. Geophys. Res.* **116**(F3), F03022.

Fujii T., Moynier F. and Albarède F. (2009) The nuclear field shift effect in chemical exchange reactions. *Chem. Geol.* **267**, 139–156.

Gaillardet J., Dupré B. and Allègre C. (1999) Geochemistry of large river suspended sediments: silicate weathering or recycling tracer?. *Geochim. Cosmochim. Acta* **63**(23–24) 4037–4051.

Gerstenberger H. and Haase G. (1997) A highly effective emitter substance for mass spectrometric Pb isotope ratio determinations. *Chem. Geol.* **136**(3–4), 309–312.

Golub G. H. (1973) Some modified matrix eigenvalue problems. *SIAM Rev.* **15**(2), 318–334.

Habfast K. (1998) Fractionation correction and multiple collectors in thermal ionization isotope ratio mass spectrometry. *Int. J. Mass Spectrom.* **176**(1–2), 133–148.

Hartley M., Thordarson T., Taylor C. and Fitton J.EIMF (2012) Evaluation of the effects of composition on instrumental mass fractionation during SIMS oxygen isotope analyses of glasses. *Chem. Geol.* **334**(0), 312–323.

Howarth R. J. (2001) A history of regression and related model-fitting in the earth sciences (1636?–2000). *Nat. Resour. Res.* **10**, 241–286.

Kessinger G. F., Huett T. and Delmore J. E. (2001) Ag ion formation mechanisms in molten glass ion emitters. *Int. J. Mass Spectrom.* **208**(1–3), 37–57.

Kukush A. and Van Huffel S. (2004) Consistency of elementwise-weighted total least squares estimator in a multivariate errors-in-variables model $AX = B$. *Metrika* **59**(1), 75–97.

Levenberg K. (1944) A method for the solution of certain non-linear problems in least squares. *Q. J. Appl. Math.* **II**(2), 164–168.

Ludwig K. R. (1998) On the treatment of concordant uranium–lead ages. *Geochim. Cosmochim. Acta* **62**(4), 665–676.

Ludwig K. R. (2003) *Isoplot/Ex Version 3.00: A Geochronological Toolkit for Microsoft Excel*. Berkeley Geochronology Center, Special Publication 4, 73 p.

Ludwig K. R. and Titterington D. M. (1994) Calculation of isochrons, ages, and errors. *Geochim. Cosmochim. Acta* **58**(22), 5031–5042.

Maréchal C. N., Télouk P. and Albarède F. (1999) Precise analysis of Cu and Zn isotopic compositions by plasma-source mass spectrometry. *Chem. Geol.* **156**(1–4), 251–273.

Markovsky I., Luisa Rastello M., Premoli A., Kukush A. and Van Huffel S. (2006) The element-wise weighted total least-squares problem. *Comput. Stat. Data Anal.* **50**, 181–209.

Markovsky I. and Van Huffel S. (2007) Overview of total least-squares methods. *Signal Process.* **87**(10), 2283–2302, Special Section: Total Least Squares and Errors-in-Variables Modeling.

Marquardt D. W. (1963) An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* **11**(2), 431–441.

McLean N. M., Bowring J. F. and Bowring S. A. (2011) An algorithm for U–Pb isotope dilution data reduction and uncertainty propagation. *Geochem. Geophys. Geosyst.* **12**.

Pawlowsky-Glahn V. and Egozcue J. J. (2006) Compositional data and their analysis: an introduction. *Geol. Soc. Lond. Spec. Publ.* **264**(1), 1–10.

Peel K., Weiss D., Chapman J., Arnold T. and Coles B. (2008) A simple combined sample-standard bracketing and inter-element correction procedure for accurate mass bias correction and precise zn and cu isotope ratio measurements. *J. Anal. At. Spectrom.* **23**(1), 103–110.

Powell R., Hergt J. and Woodhead J. (2002) Improving isochron calculations with robust statistics and the bootstrap. *Chem. Geol.* **185**(3–4), 191–204.

Russell W. A., Papanastassiou D. A. and Tombrello T. A. (1978) Ca isotope fractionation on the Earth and other solar system materials. *Geochim. Cosmochim. Acta* **42**(8), 1075–1090.

Schauble E. A. (2007) Role of nuclear volume in driving equilibrium stable isotope fractionation of mercury, thallium, and other very heavy elements. *Geochim. Cosmochim. Acta* **71**(9), 2170–2189.

Sohn R. A. and Menke W. (2002) Application of maximum likelihood and bootstrap methods to nonlinear curve-fit problems in geochemistry. *Geochem. Geophys. Geosyst.* **3**(7).

Sokal R. R. and Rohlf F. J. (1995) *Biometry: the principles and practice of statistics in biological research*, third ed. W.H. Freeman and Co., New York.

Thiemens M. H., Chakraborty S. and Dominguez G. (2012) The physical chemistry of mass-independent isotope effects and their observation in nature. *Annu. Rev. Phys. Chem.* **63**, 155–177.

Thirlwall M. F. (2000) Inter-laboratory and other errors in Pb isotope analyses investigated using a ^{207}Pb–^{204}Pb double spike. *Chem. Geol.* **163**(1–4), 299–322.

Titterington D. M. and Halliday A. N. (1979) On the fitting of parallel isochrons and the method of maximum likelihood. *Chem. Geol.* **26**(3–4), 183–195.

Toner J. D. and Sletten R. S. (2013) The formation of Ca-Cl-rich groundwaters in the Dry Valleys of Antarctica: field measurements and modeling of reactive transport. *Geochim. Cosmochim. Acta* **110**, 84–105.

Van Huffel S. and Markovsky I. (1991) *The Total Least Squares Problem: Computational Aspects and Analysis, vol. 9 of Frontiers in Applied Mathematics*. SIAM, Philadelphia, Pennsylvania.

Van Huffel S. and Vandewalle J. (1989) Analysis and properties of the generalized total least squares problem $AX \approx B$ when some or all columns in A are subject to error. *SIAM J. Matrix Anal. Appl.* **10**(3), 294–315.

Vermeesch P. (2010) Helioplot, and the treatment of overdispersed (U–Th–Sm)/He data. *Chem. Geol.* **271**(3–4), 108–111.

Vervoort J. D., Plank T. and Prytulak J. (2011) The Hf–Nd isotopic composition of marine sediments. *Geochim. Cosmochim. Acta* **75**(20), 5903–5926.

Wendt I. and Carl C. (1991) The statistical distribution of the mean squared weighted deviation. *Chem. Geol.: Isot. Geosci. Sect.* **86**(4), 275–285.

Weyer S. and Seitz H. M. (2012) Coupled lithium- and iron isotope fractionation during magmatic differentiation. *Chem. Geol.* **294–295**, 42–50.

York D. (1966) Least-squares fitting of a straight line. *Canadian Journal of Physics* **44**, 1079.

York D. (1968) Least squares fitting of a straight line with correlated errors. *Earth Planet. Sci. Lett.* **5**, 320–324.

York D., Evensen N. M., Martinez M. L. and Delgado J. D. (2004) Unified equations for the slope, intercept, and standard errors of the best straight line. *Am. J. Phys.* **72**(3), 367–375.

Young E. D., Galy A. and Nagahara H. (2002) Kinetic and equilibrium mass-dependent isotope fractionation laws in nature and their geochemical and cosmochemical significance. *Geochim. Cosmochim. Acta* **66**(6), 1095–1104.

Zhang J., Dauphas N., Davis A. M. and Pourmand A. (2011) A new method for MC-ICPMS measurement of titanium isotopic composition: Identification of correlated isotope anomalies in meteorites. *J. Anal. At. Spectrom.*

Associate editor: Yuri Amelin