(anchored) isochrons in IsoplotR

Pieter Vermeesch[1]

[1]Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, United Kingdom

Correspondence: Pieter Vermeesch (p.vermeesch@ucl.ac.uk)

Abstract. Isochrons are usually fitted by 'York regression', which uses a weighted least squares approach that accounts for correlated uncertainties in both variables. Despite its tremendous popularity in modern geochronology, the York algorithm has two important limitations that reduce its utility in several applications. First, it does not provide a satisfactory mechanism to deal with so-called 'errochrons', i.e. datasets that are overdispersed with respect to the analytical uncertainties. Second, York regression is not readily amenable to anchoring, in which either the slope or the intercept of the isochron is fixed based on some external information. Anchored isochrons can be very useful in cases where the data are insufficiently spread out to constrain both the radiogenic and non-radiogenic isotopic composition.

This paper addresses both of these issues by extending a maximum likelihood algorithm that was first proposed by Titterington and Halliday (1979, Chemical Geology 26.3-4: 183-195). The new algorithm offers the ability to attribute any excess dispersion to either the inherited component ('model 3a') or to diachronous closure of the isotopic system ('model 3b'). It provides an opportunity to anchor isochrons either to a fixed age, or to a fixed nonradiogenic composition. Last but not least, it allows the user to attach meaningful analytical uncertainty to the anchor. The new method has been implemented in IsoplotR for immediate use in Ar/Ar, Pb/Pb, U/Pb, Th/Pb, Rb/Sr, Sm/Nd, Lu/Hf, Re/Os, K/Ca and U-Th-He geochronology.

1   Introduction

Isochrons are mixing lines between radiogenic and inherited isotopic endmembers. They are an essential component of radiometric geochronology and exist in several forms. Sections 1-7 of this paper will deal with the simple case of 'P/D-isochrons', which apply to geochronometers that are based on the decay of a single radioactive parent nuclide (P; e.g., ^{87}Rb, ^{40}K, ^{147}Sm) to a particular daughter nuclide (D; e.g., ^{87}Sr, ^{40}Ar, ^{143}Nd). Sections 8 and 9 will discuss Pb/Pb- and U/Pb-isochrons, respectively. These are based on the paired decay of two parents (^{238}U and ^{235}U) to two daughters (^{206}Pb and ^{207}Pb, respectively) and require linear regression in two or three dimensions.

P/D-isochrons come in two kinds. 'conventional' isochrons are straight line relationships of the following kind:

$$\left[\frac{D}{d}\right]_i = \left[\frac{D}{d}\right]_0 + \left[\frac{P}{d}\right]_i \left(e^{\lambda t} - 1\right) \tag{1}$$

where i is the aliquot number ($1 \leq i \leq n$), d is a non-radiogenic sister isotope of D, $[D/d]_0$ is the inherited component, λ is the decay constant and t is the time elapsed since isotopic closure. A full list of definitions is provided in Appendix A. 'inverse'

isochrons are obtained by permuting P, D and d:

$$\left[\frac{d}{D}\right]_i = \left[\frac{d}{D}\right]_0 \left\{ 1 - \left[\frac{P}{D}\right]_i \left(e^{\lambda t} - 1\right) \right\} \tag{2}$$

The choice between conventional and inverse isochrons depends on the relative precision of the mass spectrometer measurements. Inverse isochrons are preferred if $d \ll D$. Otherwise, conventional isochrons are fine (Li and Vermeesch, 2021).

Equations 1 and 2 can be recast as a generic linear function:

$$y_i = a + bx_i \tag{3}$$

Table A1 maps the parameters of this generic equation to the parameters of Equations 1, 2 and subsequent isochron equations. Equation 3 is usually solved by York et al. (2004) regression (hereafter simply referred to as 'York regression'). York regression uses a least squares algorithm to estimate the intercept (a) and slope (b) of the isochron line from an $n \times 5$ table of paired isotopic ratio measurements, their standard errors and their error correlations. Appendix B provides an example of such an input table.

Although this paper will use the York parameters a and b to fit the isochron, all results will be presented in terms of the geologically more meaningful inherited endmember ratio $[D/d]_0$ and radiogenic endmember ratio $[D/P]^* \equiv e^{\lambda t} - 1$, where $[D/d]_0 = a$ and $[D/P]^* = b$ for conventional isochrons, and $[D/d]_0 = 1/a$ and $[D/P]^* = -b/a$ for inverse isochrons (Li and Vermeesch, 2021).

The most accurate and precise results are obtained from samples that are evenly spread along the isochron line, spanning the entire range of values from the inherited to the radiogenic endmember. Unfortunately, this condition is not always fulfilled. It is not uncommon for most or all aliquots in a sample to cluster together at one point along the isochron, making it difficult to accurately estimate the endmember compositions. For example, no precise isochron ages can be obtained from samples whose radiogenic daughter component is dwarfed by the inherited daughter component. Conversely, the composition of the inherited component cannot be precisely estimated in extremely radiogenic samples. Finally, when the data cluster together in the middle, then neither endmember component can be reliably determined (Figure 1).

Sometimes, such poorly constrained isochrons can be fixed using external information. For example, if the composition of the inherited component is known through some independent means (e.g., by analysing a cogenetic mineral that is naturally poor in P and rich in D and d), then this information can be used to anchor the isochron. Anchoring reduces the number of unknown parameters by one, benefitting numerical stability and precision. Conversely, if the age of the sample is known, then the inherited component can be estimated by anchoring to the radiogenic endmember.

There currently exists no formally documented way to anchor York regression. A commonly used 'hack' is to add an extra data point with infinite precision representing either the inherited or radiogenic endmember component. However, this hack does not provide a satisfactory mechanism to assign uncertainty to the anchor. This paper solves that problem. Section 2 introduces a maximum likelihood formulation of York regression, which is amenable to anchoring with uncertainty. Section 3 shows how this formulation can be used to model 'errorchrons' that are overdispersed with respect to the analytical uncertainties.

Section 4 provides further details about the implementation of maximum likelihood regression, which can attribute overdispersion to either the intercept ('model 3a') or slope ('model 3b') of the linear fit. It turns out that model 3b is computationally more challenging than model 3a. Section 5 shows how this issue can be avoided by inverting and/or flipping the isochron axes.

Section 6 shows how anchored isochron regression represents a trivial special case of the maximum likelihood algorithm.
60   Section 7 attaches two different geological interpretations to the uncertainty of isochron anchors. These two interpretations lead to two different fit models ('model 1' and 'model 3', which are so named for historical reasons). Sections 8 and 9 apply the same logic to Pb/Pb and U/Pb isochrons, respectively. Finally, Section 10 shows how the new algorithms can be used within the `IsoplotR` software package.
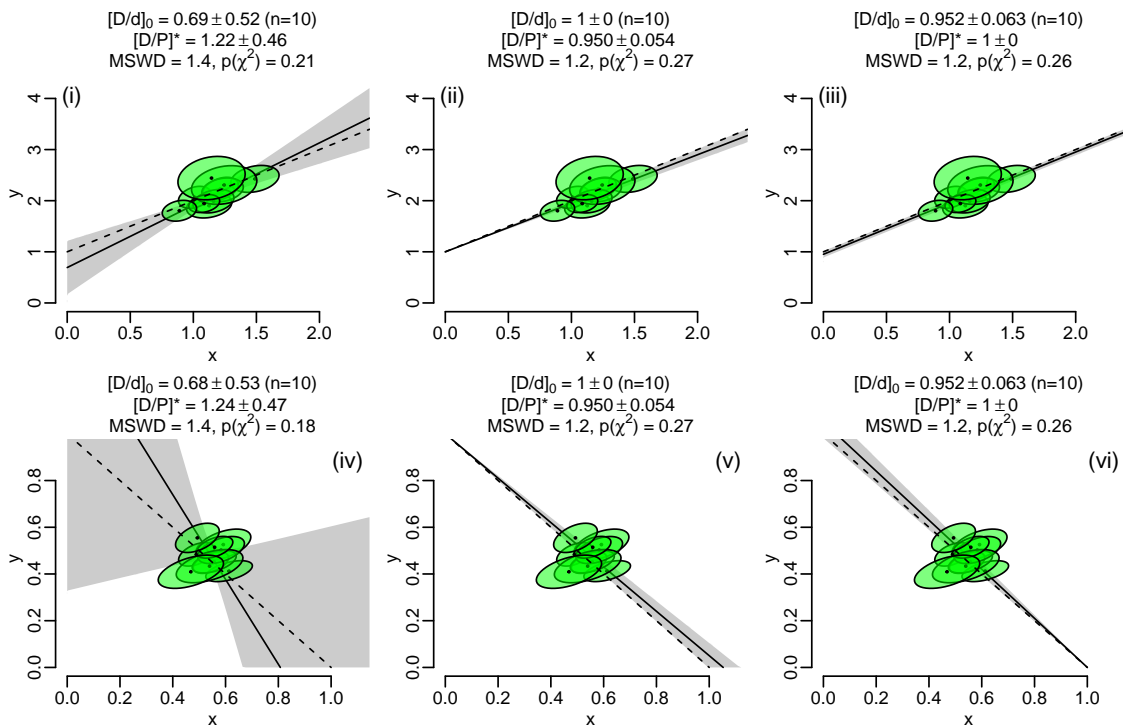
Figure 1. P/D-isochron diagrams for the synthetic dataset of Appendix B. The true slope and intercept are $a = 1$ and $b = 1$ for the conventional isochrons (i–iii), and $a = 1$ and $b = -1$ for the inverse isochrons (iv–vi). These parameter values correspond to $[D/d]_0 = [D/P]^* = 1$. The true trends are shown as dashed lines and the estimated trends as solid lines, with 95% confidence envelopes shown in grey. The estimated parameter values are shown above the panels. Unconstrained isochron regression (i and iv) yields imprecise and inaccurate results. Fixing the inherited component (ii and v) or the radiogenic component (iii and vi) greatly improves both aspects of the fit.

2 Maximum likelihood formulation of York regression

65 York et al. (2004) use the method of least squares to fit the general problem of weighted regression with correlated uncertainties in both variables. Titterington and Halliday (1979) obtained identical results using the method of maximum likelihood. The latter approach offers greater flexibility than least squares and will be used in the remainder of this paper.

The method of maximum likelihood is a standard statistical technique to estimate the parameters of a probability distribution from a set of measurements. In the case of two-dimensional isochron regression, the parameters are the intercept a, the slope

70 b and the true (but unknown) x_i-values. The data are the X_i and Y_i-measurements along with their assumed uncertainties and error correlations (see Appendix A for a full list of definitions). For a dataset of n aliquots, there are $2n$ measurements (n X_i-values and n Y_i-values) and $n + 2$ parameters (a, b and n x_i-values). That leaves $2n - n - 2 = n - 2$ 'degrees of freedom' to estimate the parameters.

This overconstrained problem can be solved by minimising the paired differences ('residuals') between the true x_i- and

75 y_i-values (where $y_i = a + bx_i$) and the measured X_i- and Y_i-values:

$$\begin{cases} X_i = x_i + \epsilon_{x,i} \\ Y_i = a + bx_i + \epsilon_{y,i} \end{cases} \qquad (4)$$

To make the problem tractable, the vector of residuals $\boldsymbol{\Delta}_i$ is assumed to be drawn from a bivariate normal distribution with zero mean and covariance matrix $\boldsymbol{\Sigma}_i$:

$$\boldsymbol{\Delta}_i \equiv \begin{bmatrix} X_i - x_i \\ Y_i - y_i \end{bmatrix} = \begin{bmatrix} \epsilon_{x,i} \\ \epsilon_{y,i} \end{bmatrix} \text{ and } \boldsymbol{\Sigma}_i \equiv \begin{bmatrix} s[X_i]^2 & s[X_i, Y_i] \\ s[X_i, Y_i] & s[Y_i]^2 \end{bmatrix}. \qquad (5)$$

80 The product of the normal probabilities for all the aliquots (from $1 \leq i \leq n$) is called the 'likelihood' of the parameters given the data. Maximising this number produces the most accurate possible parameter estimates $\hat{a}$, $\hat{b}$ and $\hat{x}_i$. Alternatively, and equivalently, the same estimates can be obtained by maximising the logarithm of the likelihood. This is the preferred approach as it improves numerical stability and facilitates the estimation of the parameter uncertainties. The log-likelihood function can be formally defined as:

85 $$\mathcal{LL}(\text{parameters}|\text{data}) = -\frac{1}{2} \sum_{i=1}^{n} \left(2\pi + \ln|\boldsymbol{\Sigma}_i| + \boldsymbol{\Delta}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\Delta}_i \right) \qquad (6)$$

This formulation allows for correlated uncertainties between the variables ($s[X_i, Y_i] \neq 0$ in Equation 5), but not between aliquots ($s[X_i, Y_j] = 0$ if $i \neq j$). Daëron and Vermeesch (2023) discuss the generalised case of 'omnivariant' regression, where this requirement has been relaxed and Equation 6 is replaced with a single matrix expression.

3 Dealing with overdispersion

90 The degree to which the residuals $\boldsymbol{\Delta}_i$ are consistent with the assumed analytical uncertainties $\boldsymbol{\Sigma}_i$ can be assessed using a Chi-square test and MSWD parameter. Readers who are not familiar with these concepts are referred to Appendix C for

details. Isochrons that exhibit significant overdispersion with respect to the analytical uncertainties are colloquially referred to as 'errorchrons'. Whether it is right to use such a pejorative term for this common phenomenon is debatable (Schaen et al., 2021). Four approaches may be used to deal with errorchrons:

95 **model 1**: Inflate the analytical uncertainties by a factor $\sqrt{\mathrm{MSWD}}$.

model 2: Ignore the analytical uncertainties and replace York regression with orthogonal regression or a similar technique. This approach will not be discussed further in this paper.

model 3: Quantify the dispersion as an additional free parameter. There are two options for doing so:

model 3a: Attribute the excess dispersion to variability of the inherited component and, hence, the isochron intercept.
100   For conventional isochrons, this means that the true isotopic ratios of the cogenetic aliquots do not belong to a single isochron line, but to a family of parallel isochron lines whose intercepts are normally distributed (Titterington and Halliday, 1979). The standard deviation of this normal distribution (σ_a) can be used to quantify the dispersion of the data around the 'central' isochron line.

model 3b: Attribute the excess dispersion to diachronous closure of the isotopic system. This mechanism produces
105   families of conventional isochrons that share a common intercept but differ in their slopes. Again, the distribution of the slopes can be assumed to follow a normal distribution, with mean b and standard deviation σ_b. The latter value can be used to quantify the dispersion of the data, which in this case captures the degree of diachroneity.

Model 3a isochron regression can be formalised by modifying Equation 4 as follows:

$$\begin{cases} X_i = x_i + \epsilon_{x,i} \\ Y_i = a + bx_i + \epsilon_a + \epsilon_{y,i} \end{cases} \tag{7}$$

110   The dispersion parameter σ_a can be estimated (as $\hat{\sigma}_a$) by incorporating it into the covariance matrices:

$$\boldsymbol{\Sigma}_{a,i} \equiv \begin{bmatrix} s[X_i]^2 & s[X_i, Y_i] \\ s[X_i, Y_i] & s[Y_i]^2 + \sigma_a^2 \end{bmatrix}. \tag{8}$$

The equivalent expressions for model 3b regression are

$$\begin{cases} X_i = x_i + \epsilon_{x,i} \\ Y_i = a + (b + \epsilon_b)x_i + \epsilon_{y,i} \end{cases} \tag{9}$$

and

115 $$\boldsymbol{\Sigma}_{b,i} \equiv \begin{bmatrix} s[X_i]^2 & s[X_i, Y_i] \\ s[X_i, Y_i] & s[Y_i]^2 + (x_i\sigma_b)^2 \end{bmatrix} \tag{10}$$

respectively.

4 Specific cases

Three specific cases of Equation 6 are of interest in the present discussion:

$$
\begin{aligned}
\textbf{model 1:} \quad & \mathcal{LL}_1(a,b,\mathbf{x}|\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma}) = \text{constant} - \tfrac{1}{2}\sum_{i=1}^{n}\boldsymbol{\Delta}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Delta}_i \\
\textbf{model 3a:} \quad & \mathcal{LL}_{3a}(a,b,\sigma_a,\mathbf{x}|\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma}) = \text{constant} - \tfrac{1}{2}\sum_{i=1}^{n}\left(\ln|\boldsymbol{\Sigma}_{a,i}| + \boldsymbol{\Delta}_i^T\boldsymbol{\Sigma}_{a,i}^{-1}\boldsymbol{\Delta}_i\right) \\
\textbf{model 3b:} \quad & \mathcal{LL}_{3b}(a,b,\sigma_b,\mathbf{x}|\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma}) = \text{constant} - \tfrac{1}{2}\sum_{i=1}^{n}\left(\ln|\boldsymbol{\Sigma}_{b,i}| + \boldsymbol{\Delta}_i^T\boldsymbol{\Sigma}_{b,i}^{-1}\boldsymbol{\Delta}_i\right)
\end{aligned}
\tag{11}
$$

120   Note that the (logged) determinant of the covariance matrices is not included in the constant for model 3 fits, because $\boldsymbol{\Sigma}_{a,i}$ and $\boldsymbol{\Sigma}_{b,i}$ are both functions of $\mathbf{x}$, whereas $\boldsymbol{\Sigma}_i$ is not. $\mathcal{LL}$ can be maximised numerically using an iterative two-step procedure. For model 1 regression, the first step maximises $\mathcal{LL}_1(\mathbf{x}|a,b,\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma})$ to find $\mathbf{x}$ for any pair of a- and b-values. The second step repeats the first step for different values of a and b until the overall optimum is reached.

    This procedure can easily be adapted to model 3 regression, by adding σ_a or σ_b to the unknown parameters. It works well
125 for model 3a, where the first step has a direct solution. However, it is slow for model 3b regression, in which finding $\mathbf{x}$ requires many iterations. See Appendix D for further details. Fortunately, in practice model model 3b is rarely or never needed for geochronology, as explained next.

5 Flipped isochron regression

For inverse isochrons, model 3b regression does not actually provide any useful information. Unlike conventional isochrons,
130 whose chronological information is contained in the slope, the chronological information of inverse isochrons is contained in their horizontal intercept. This information can be unlocked by flipping the axes of the isochron diagram around and treating the P/D-ratio as the dependent variable. Applying model 3a regression to the flipped data yields a dispersion estimate σ_a that can be used to quantify the age dispersion.

    The same trick can be used to estimate the slope uncertainty of a conventional isochron. A pragmatic way to avoid the slow
135 convergence rate of model 3b regression is to carry out the numerical optimisation in inverse isochron space, flip the dependent and independent variables around and carry out a model 3a regression on the transformed data. The resulting $\hat{\sigma}_a$-estimate can be converted to a $\hat{\sigma}_b$ estimate:

$$
\hat{\sigma}_b(\text{conventional isochron}) \approx \frac{\hat{\sigma}_a}{\hat{a}^2}(\text{flipped inverse isochron})
\tag{12}
$$

6 Anchored isochron regression

140 Anchored isochron regression requires just a trivial modification of the maximum likelihood algorithm. It suffices to treat the anchored parameter as data. For example, the slope and intercept of a model 1 isochron can be anchored by maximising $\mathcal{LL}_1(a,\mathbf{x}|b,\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma})$ and $\mathcal{LL}_1(b,\mathbf{x}|a,\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma})$, respectively. Note that these calculations assume that the anchor is known

exactly. This is reflected in the zero uncertainties of the nonradiogenic component Figures 1.ii and 1.v, and of the radiogenic components of Figures 1.iii and 1.vi. Such absolute certainty is unrealistic in the noisy world of geology. Section 7 introduces

145 two ways to formally account for uncertainty in anchored isochron regression.

7 Accounting for anchor uncertainty

When assigning uncertainty to statistical parameters, it is important to clearly define the meaning of this uncertainty. In the case of anchored isochron regression, the uncertainty of the intercept or slope can carry two meanings. For example, when the intercept of an isochron is anchored at a value $\bar{a} \pm \sigma_{\bar{a}}$, this can either mean that:

150 **model 1**: the data are underlain by a single isochron whose intercept is only approximately known, with a most likely value of $\bar{a}$ and a precision (standard error) of $\sigma_{\bar{a}}$.

model 3: the data were drawn from a family of isochron lines whose intercepts follow a normal distribution with mean $\bar{a}$ and standard deviation $\sigma_{\bar{a}}$.

These two different approaches can be implemented by replacing the log-likelihood functions of Equation 11 with

155
$$
\begin{aligned}
\textbf{model 1}: \quad & \mathcal{LL}_1(a,b,\mathbf{x}|\bar{a},\sigma_{\bar{a}},\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma}) = \text{constant} + \mathcal{LL}_1(a,b,\mathbf{x}|\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma}) - \frac{1}{2}\left(\frac{a-\bar{a}}{\sigma_{\bar{a}}}\right)^2 \\
\textbf{model 3}: \quad & \mathcal{LL}_{3a}(b,\mathbf{x}|\bar{a},\sigma_{\bar{a}},\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma})
\end{aligned}
\tag{13}
$$

The equivalent expressions for anchored slopes with uncertainty are

$$
\begin{aligned}
\textbf{model 1}: \quad & \mathcal{LL}_1(a,b,\mathbf{x}|\bar{b},\sigma_{\bar{b}},\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma}) = \text{constant} + \mathcal{LL}_1(a,b,\mathbf{x}|\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma}) - \frac{1}{2}\left(\frac{b-\bar{b}}{\sigma_{\bar{b}}}\right)^2 \\
\textbf{model 3}: \quad & \mathcal{LL}_{3b}(a,\mathbf{x}|\bar{b},\sigma_{\bar{b}},\mathbf{X},\mathbf{Y},\boldsymbol{\Sigma})
\end{aligned}
\tag{14}
$$

Note that model 1 treats the slope and intercept as unknowns, so their maximum likelihood estimate generally do *not* equal the anchored values $\bar{a}$ and $\bar{b}$. However, the smaller $\sigma_{\bar{a}}$ and $\sigma_{\bar{b}}$ are relative to $\bar{a}$ and $\bar{b}$, the closer that $\hat{a}$ and $\hat{b}$ are to $\bar{a}$ and $\bar{b}$. In

160 contrast, anchored model 3 regression fixes the dispersion of the anchored parameters, so that $\hat{\sigma}_a = \sigma_{\bar{a}}$ for anchored intercepts, whereas $\hat{\sigma}_b = \sigma_{\bar{b}}$ for anchored slopes.

Figure 2 applies the two approaches to the synthetic dataset of Appendix B and Figure 1, using anchors of $[D/d]_0 = 1 \pm 0.05$ and $[D/P]^* = 1 \pm 0.05$. As expected, models 1 and 3 produce slightly different outcomes, with model 1 resulting in $[D/d]_0$- and $[D/P]^*$-estimates that are slightly different than the anchored values, whereas model 3 reproduces the anchors exactly. The

165 model 3 fits produce more precise estimates for the unanchored quantities. This reflects the fact that some of the uncertainty in the isochron is captured by the dispersion parameter, which is absent from the model 1 fit.

Model 3 partitions the analytical and geological uncertainty between the standard errors of $\hat{a}$ and $\hat{b}$ on the one hand, and the dispersion parameter on the other hand. It is not so easy to visualise this partitioned uncertainty. Figure 2 uses a grey confidence envelope to show the analytical uncertainty, and black error bars to visualise the geological dispersion. Together

170 these two items represent the total uncertainty budget.

It is, of course, also possible to treat the dispersion parameter as an unknown by maximising $\mathcal{LL}_{3a}(b, \sigma_a, \mathbf{x} | \bar{a}, \mathbf{X}, \mathbf{Y}, \boldsymbol{\Sigma})$ or $\mathcal{LL}_{3b}(a, \sigma_b, \mathbf{x} | \bar{b}, \mathbf{X}, \mathbf{Y}, \boldsymbol{\Sigma})$.
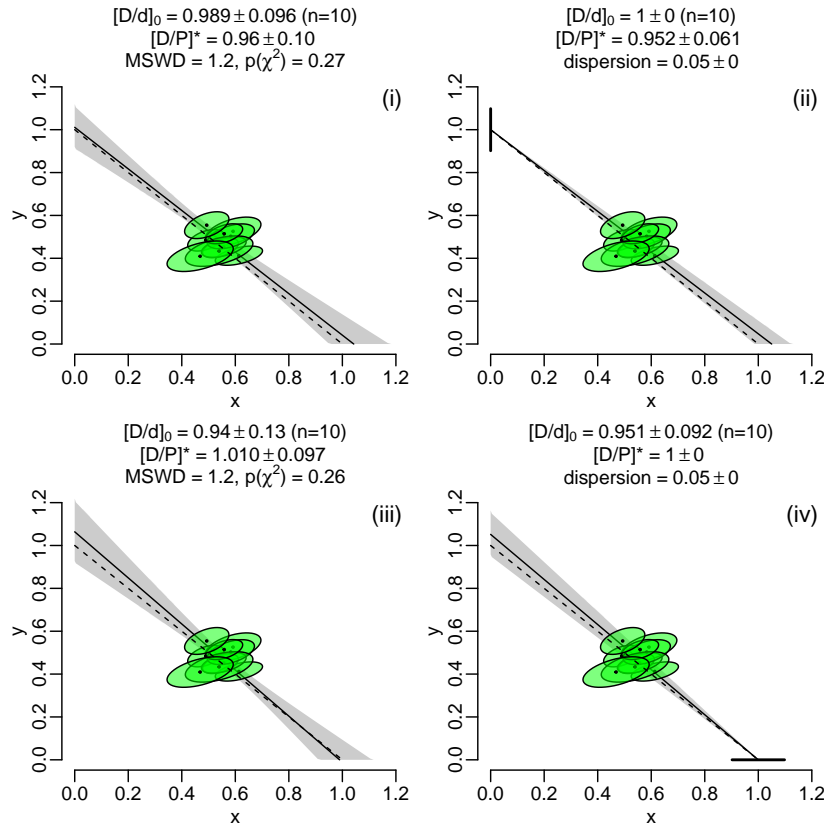
Figure 2. Anchored isochrons for the same data of Figure 1, shown on inverse isochrons. Panels i and ii anchor the non-radiogenic endmember composition to $[D/d]_0 = 1 \pm 0.05$. Panels iii and iv anchor the radiogenic endmember composition to $[D/P]^* = 1 \pm 0.05$. Panels i and iii use model 1 isochron regression, whereas panels ii and iv use model 3. Note how the estimated values for $[D/d]_0$ and $[D/P]^*$ are exactly the same as the anchored values for model 3, but not for model 1. The vertical error bar in panel ii and the horizontal error bar on panel iv represent the fixed dispersion of the vertical and horizontal intercept, respectively.

8 Pb/Pb isochrons

The ^{207}Pb/^{206}Pb-method is based on the paired decay of two radioactive parents P_1 and P_2 (^{238}U and ^{235}U) to two daughters D_1 and D_2 (^{206}Pb and ^{207}Pb, respectively) in the presence of a non-radiogenic sister isotope d (^{204}Pb). This gives rise to the

175

following implicit age equation:

$$\frac{[D_2/d]_i - [D_2/d]_0}{[D_1/d]_i - [D_1/d]_0} = \left[\frac{P_2}{P_1}\right]\frac{e^{\lambda_2 t} - 1}{e^{\lambda_1 t} - 1} \tag{15}$$

where λ_1 and λ_2 are the decay constants of P_1 and P_2, respectively, and $[P_2/P_1]$ is constant. Equation 15 can be recast into the generic form of Equation 3 to form a 'D/D-isochron' using the mapping of Table A1. This gives rise to conventional 180 (^{207}Pb/^{204}Pb vs. ^{206}Pb/^{204}Pb) and inverse (^{207}Pb/^{206}Pb vs. ^{204}Pb/^{206}Pb) isochrons.

Unlike inverse P/D-isochrons, which are characterised by negative slopes, inverse D/D-isochrons have positive slopes. There is a perfect symmetry between conventional and inverse D/D-isochrons in the sense that the intercept of a conventional isochron equals the slope of an inverse isochron and *vice versa*. This slightly changes and, in fact, simplifies the procedure for anchored isochron regression.

185   There is no need for flipped isochron regression of Pb/Pb data. Instead, all possible scenarios can be handled by straightforward model 1 and model 3a regression. Isochrons can be anchored to the nonradiogenic component in conventional isochron space, and to the radiogenic component in inverse isochron space.

9   U/Pb isochrons

The U/Pb method, like the Pb/Pb method, is based on the paired decay of ^{238}U to ^{206}Pb and of ^{235}U to ^{207}Pb. The two 190 chronometers can be treated separately and plotted as P/D-isochrons. These two-dimensional isochrons can be anchored to the radiogenic and nonradiogenic composition using the methods described in Sections 2–7.

$$\begin{cases} \left[\frac{^{206}\text{Pb}}{^{204}\text{Pb}}\right]_i = \left[\frac{^{206}\text{Pb}}{^{204}\text{Pb}}\right]_0 + \left[\frac{^{238}\text{U}}{^{204}\text{Pb}}\right]_i \left(e^{\lambda_{238} t} - 1\right) \\ \left[\frac{^{207}\text{Pb}}{^{204}\text{Pb}}\right]_i = \left[\frac{^{207}\text{Pb}}{^{204}\text{Pb}}\right]_0 + \left[\frac{^{235}\text{U}}{^{204}\text{Pb}}\right]_i \left(e^{\lambda_{235} t} - 1\right) \end{cases} \tag{16}$$

Alternatively, the two ingrowth equations can also be coupled, forming a three dimensional 'total-Pb/U isochron'. This problem can be solved using the method of maximum likelihood (Ludwig, 1998). There is just one important difference 195 between this solution and the York-like algorithm of Section 2. York regression is formulated in terms of a generic intercept (a) and slope (b). In contrast, the 'Ludwig' algorithm is formulated in terms of the actual non-radiogenic ^{206}Pb/^{204}Pb ($= \alpha$) and ^{207}Pb/^{204}Pb ($=\beta$) ratio, as well as the age (t) of the system.

It is relatively straightforward to generalise the concept of model-3 regression to total-Pb/U isochrons when the dispersion is attributed to the isochron age. However, it is no immediately clear how to do the same for the nonradiogenic composition, 200 because this would require partitioning the dispersion between the α- and β-parameters. This paper will not attempt to answer this questions. Instead, let us conclude the technical discussion by shifting to a far more common type of U/Pb data.

Most published U/Pb studies do not report ^{204}Pb because this nuclide is rare and difficult to measure. Instead, 'semitotal-Pb/U' regression is done in Wetherill or Tera-Wasserburg concordia space, which are akin to conventional and inverse isochron

space, respectively. In the absence of a nonradiogenic sister isotope of Pb, the isochron is then redefined as as a mixing line
205 between an inherited ^{207}Pb/^{206}Pb-ratio and the radiogenic ^{206}Pb/^{238}U- and ^{207}Pb/^{235}U-ratios.

The semitotal-Pb/U problem can be cast into a maximum likelihood format by replacing Equation 4 with

$$
\begin{cases}
X_i = x_i + e^{\lambda_{238}t} - 1 + \epsilon_{x,i} \\
Y_i = a + bx_i + \epsilon_{y,i}
\end{cases}
\tag{17}
$$

where X_i and Y_i are the measured ^{207}Pb/^{235}U- and ^{206}Pb/^{238}U-ratios of the ith aliquot, respectively; x_i is the true ^{207}Pb$_0$/^{235}U-
ratio of the ith aliquot; $a = e^{\lambda_{238}t} - 1$ and $b = [^{206}$Pb/^{207}Pb$]_0[^{235}$U/^{238}U]. Equation 17 can be solved using the same recipe
210 as the York reformulation of Sections 2–7, with the caveat that the isochron terminates on the concordia line at the U/Pb
composition corresponding to t.

Figure 3 shows a model 3a anchored semitotal-Pb/U isochron in Tera-Wasserburg space. The true age for this synthetic
dataset is 1500 Ma. The unconstrained model-1 isochron produces an imprecise isochron age of 1400 ± 350 Ma, which is
improved to 1522 ± 99 Ma by anchoring to $[^{207}$Pb/^{206}Pb$]_0 = 1.1$. Just like the P/D-isochrons of Figure 2, the uncertainty
215 budget for the U/Pb-isochrons of Figure 3 is partitioned between the analytical and geological dispersion, where the latter
is treated as a free parameter with a maximum likelihood estimate of $\hat{\sigma}_a = 0.091 \pm 0.783$. Note the large uncertainty of the
dispersion parameter, which reflects the relatively low sensitivity of the log-likelihood function to it.
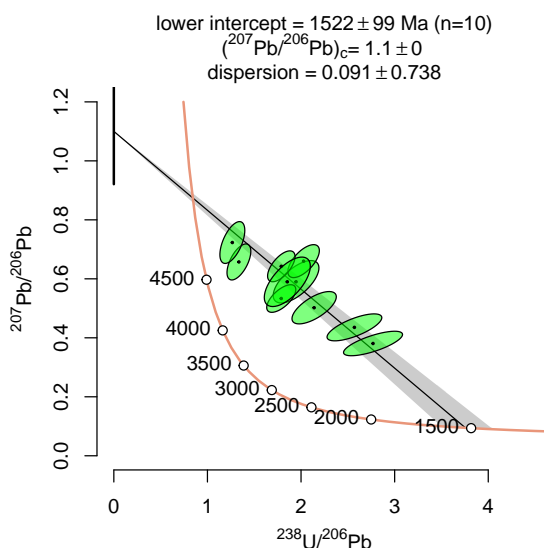
Figure 3. Model 3a semitotal-Pb/U isochron for a synthetic dataset with a true age $t = 1500$ Ma, anchored to ^{207}Pb/^{206}Pb$= 1.10$ with the
dispersion parameter treated as a free parameter. The data are provided in Table B2.

10   Implementation in `IsoplotR`

All the algorithms presented in this paper have been implemented in the free and open geochronological toolbox IsoplotR
220   (Vermeesch, 2018). The ability to fit isochrons with model 1, 2 and 3 dates back to IsoplotR's first public release. Versions 2.1
and 5.2 added model 1 anchors to U/Pb ischrons without and with uncertainty, respectively. Anchored York regression and
model 3 anchors with uncertainty were added to IsoplotR 6.0. These funtions can be accessed via the GUI or from the command
line.

In the GUI, anchored regression is available from the 'isochron' menu (so not from the 'concordia' menu for U/Pb data). The
225   relevant settings can be selected from the 'Options' menu and should be self explanatory. Using a built-in ^{40}Ar/^{39}Ar-dataset to
illustrate the command-line functionality, and starting with an unanchored inverse isochron fit using York regression:

```
library('IsoplotR')
attach(examples)
isochron(ArAr)
```

230   To carry out model 3a isochron regression anchored to a variable initial ^{40}Ar/^{36}Ar-ratio of 300 ± 5 (1se):

```
settings('iratio','Ar40Ar36',300,5)
isochron(ArAr,anchor=1,model=3,
         xlim=c(0,0.25),ylim=c(0,0.0035))
```

where `settings(...)` is used to change the atmospheric argon composition, `anchor=1` fixes the isochron to the inherited
235   component, and `model=3` tells IsoplotR to assign the dispersion to the reported uncertainty of the atmospheric ^{40}Ar/^{36}Ar-
ratio.

Finally, to carry out model 3b isochron regression anchored to a radiogenic composition that was set at 62 ± 1 Ma (1se):

```
isochron(ArAr,anchor=c(2,62,1),model=3,
         xlim=c(0,0.25),ylim=c(0,0.0035))
```

240   The dispersion can be changed to a free parameter by setting the third element of the `anchor` argument to a non-positive
value. Documentation for additional options and settings can be accessed from the console, using the `help(isochron)` or
`?isochron` commands.

11   Conclusions

This paper builds on previous work by McIntyre et al. (1966), Titterington and Halliday (1979) and Ludwig (2003). McIntyre
245   et al. (1966) introduced the concept of model 3 isochron regression, attributing the excess dispersion of errorchrons to vari-
ability of the inherited component. This approach is implemented in Isoplot (Ludwig, 2003) and is referred to as model 3a

regression in this paper. Note that McIntyre et al. (1966)'s definition of model 2 regression differs from that of Ludwig (2003). It is more akin (but not identical) to the $\sqrt{\text{MSWD}}$ approach of Section 3.

Titterington and Halliday (1979) recast the model 3 algorithm of McIntyre et al. (1966) in a maximum likelihood framework. This paper extends the maximum likelihood approach to a second type of geological scenario, in which excess dispersion of the data around the isochron is not attributed to the inherited component but to the radiogenic component. This approach is referred to as model 3b regression.

The maximum likelihood formulation of isochron regression can also be used to anchor isochrons to either the inherited or the radiogenic component, and to assign geologically meaningful uncertainty to the anchor. In reality it is possible that some samples are affected by both mechanisms, so that different aliquots of the same sample differ in their initial ratios as well as the timing of their isotopic closure. Unfortunately, it is not possible to simultaneously capture both types of dispersion using the algorithms of this paper.

The difference between model 1 and model 3 regression represents two contrasting views of geological reality. The model 1 approach assumes that the isotopic composition of minerals represents discrete components recorded at distinct events. In contrast, model 3 isochrons (anchored or not) represent a 'fuzzier' reality, in which the initial composition or timing of isotopic closure are allowed to vary within a rock. Under the latter model, the concept of an 'errorchron' no longer makes sense.

Code availability. IsoplotR is available from CRAN (https://CRAN.Rproject.org/package=IsoplotR) and GitHub (https://github.com/pvermees/IsoplotR)

Appendix A: Definitions

P: a radioactive parent (e.g., ^{87}Rb).

D: the radiogenic daughter of P (e.g., ^{87}Sr).

d: a non-radiogenic sister isotope of D (e.g., ^{86}Sr).

a, b, x_i, y_i: the true (but unknown) values of the intercept, slope and isotopic ratio of the ith aliquot (out of n aliquots) from a
    sample.

$\hat{a}$, $\hat{b}$, $\hat{x}_i$, $\hat{y}_i$: the estimated values of a, b, x_i and y_i, respectively.

X_i, Y_i: the measured values of x_i and y_i, respectively.

$s[X_i]$, $s[Y_i]$, $s[X_i, Y_i]$: the standard errors and covariance of X_i and Y_i.

$\epsilon_{x,i} = X_i - x_i$, $\epsilon_{y,i} = Y_i - y_i$: the residuals of the data around the best fit line.

$\boldsymbol{\Sigma}_i$: the 2×2 covariance matrix of X_i and Y_i

$\mathbf{x}, \mathbf{y}, \mathbf{X}, \mathbf{Y}, \mathbf{\Sigma}$: arrays and sets containing all n x_i, y_i, X_i, Y_i and $\mathbf{\Sigma}_i$-values.

Table A1. Assignment of the isotopic ratios for P/D- and D/D-isochrons to the generic parameters of Equation 3.

	P/D		D/D	
parameter	conventional	inverse	conventional	inverse
x_i	$[P/d]_i$	$[P/D]_i$	$[D_1/d]_i$	$[d/D_1]_i$
y_i	$[D/d]_i$	$[d/D]_i$	$[D_2/d]_i$	$[D_2/D_1]_i$
a	$[D/d]_0$	$[d/D]_0$	$[D_2/d]_0$	$[D_2/D_1]^*$
b	$[D/P]^*$	$-[d/D]_0/[P/D]^*$	$[D_2/D_1]^*$	$[D_2/d]_0$

Appendix B: Data

The synthetic data of Figures 1 and 2 are provided in Table B1 and were generated as follows:

1. Generate $\mathbf{x}$ by drawing ten random numbers from a logit-normal distribution where the mean and standard deviation of the logits are 0 and 0.2, respectively.

2. Generate $\mathbf{y}$ by plugging $\mathbf{x}$ into Equation 3 with $a = b = 1$.

3. Turn the resulting ten pairs of $\{x_i, y_i\}$-values into ten trios of $\{P_i, D_i, d_i\}$-probabilities that add up to unity.

4. Use the probabilities obtained in the previous step carry out between 250 and 2500 Bernoulli experiments, resulting in ten trios of Poisson variables.

5. Use the Poisson variables obtained in the previous step to form ten pairs of $\mathbf{X}$- and $\mathbf{Y}$-measurements and their (correlated) uncertainties.

6. Add some excess dispersion by shrinking the uncertainties by 20%.

Note that this procedure fits the definition of a model 1 isochron, even though this paper also uses the resulting data to illustrate model 3 regression.

Appendix C: Testing overdispersion

Whether a dataset is overdispersed with respect to the analytical uncertainties can be formally assessed using the Chi-statistic and test. To this end, calculate the χ^2-statistic using the sum of squares term of Equation 6.

$$\chi^2 = \sum_{i=1}^{n} \mathbf{\Delta}_i^T \mathbf{\Sigma}_i^{-1} \mathbf{\Delta}_i \tag{C1}$$

Table B1. Synthetic data for Figure 1, provided in an inverse isochron format. See Li and Vermeesch (2021) for details on how to convert these data to a conventional isochron format.

i	$X_i = [P/D]_i$	$s[X_i]$	$Y_i = [d/D]_i$	$s[Y_i]$	$r[X_i, Y_i]$
1	0.59	0.04	0.53	0.03	0.48
2	0.49	0.03	0.48	0.02	0.46
3	0.61	0.04	0.41	0.02	0.51
4	0.54	0.03	0.45	0.02	0.48
5	0.56	0.03	0.45	0.02	0.50
6	0.58	0.04	0.50	0.02	0.48
7	0.54	0.05	0.43	0.03	0.50
8	0.56	0.03	0.51	0.02	0.48
9	0.49	0.03	0.55	0.03	0.46
10	0.47	0.05	0.41	0.03	0.50

Table B2. Synthetic data for Figure 3.

i	$X_i = [^{238}\text{U}/^{206}\text{Pb}]_i$	$s[X_i]$	$Y_i = [^{207}\text{Pb}/^{206}\text{Pb}]_i$	$s[Y_i]$	$r[X_i, Y_i]$
1	1.34	0.05	0.66	0.02	0.59
2	1.95	0.10	0.59	0.03	0.65
3	2.57	0.12	0.44	0.02	0.72
4	1.27	0.05	0.72	0.03	0.58
5	2.03	0.07	0.66	0.02	0.64
6	1.79	0.06	0.53	0.02	0.65
7	2.14	0.10	0.50	0.02	0.68
8	1.79	0.06	0.64	0.02	0.63
9	2.77	0.13	0.38	0.02	0.73
10	1.85	0.10	0.59	0.03	0.63

If analytical uncertainty is the only source of dispersion in the dataset, then χ^2 is expected to follow a Chi-square distribution with $df = n - 2$ (or, for anchored regression, $df = n - 1$) degrees of freedom. This is called the 'null distribution'. The probability of observing a value greater than χ^2 under this distribution is called the p-value. A p-value cutoff of 0.05 is generally used to distinguish isochrons ($p \geq 0.05$) from errorchrons ($p < 0.05$). Dividing Equation C1 by the number of degrees of freedom (df) produces a 'reduced Chi-square statistic', which is known to geologists as an MSWD (McIntyre et al., 1966):

$$\text{MSWD} = \frac{\chi^2}{df} \tag{C2}$$

In the absence of overdispersion, and for sufficiently large samples ($n > 20$), the expected null distribution of the MSWD is approximately normal with a mean of 1 and a standard deviation $\sqrt{2/df}$ (Wendt and Carl, 1991). This information can be

300 used as an alternative way to test the null hypothesis.

Appendix D: Further details about the maximum likelihood solution

Section 4 described how the log-likelihood functions of Equation 11 can be maximised in a two step process: (1) find the x_i-values that maximise $\mathcal{LL}$ for any pair of a- and b-values, and (2) search the space of a- and b- values until the overall maximum is found. For model 1 regression, the first step has a direct solution. Introducing some shorthand notation for the inverse of the

305 covariance matrix Σ_i:

$$\mathbf{\Sigma}_i^{-1} \equiv \begin{bmatrix} \Omega_{1,1}^i & \Omega_{1,2}^i \\ \Omega_{2,1}^i & \Omega_{2,2}^i \end{bmatrix}, \tag{D1}$$

the fitted points x_i can be found by solving:

$$\left.\frac{\partial \mathcal{LL}_1}{\partial x_i}\right|_{\hat{x}_i} = 0 \tag{D2}$$

which yields

310 $$\hat{x}_i = \frac{X_i(\Omega_{1,2}^i + \Omega_{1,2}^i) + (Y_i - a)(\Omega_{2,1}^i + b\Omega_{2,2}^i)}{(\Omega_{1,2}^i + \Omega_{1,2}^i) + b(\Omega_{2,1}^i + b\Omega_{2,2}^i)}. \tag{D3}$$

The same approach can be used for model 3a regression (with $\partial\mathcal{LL}_{3a}/\partial x_i = 0$). However, it does not work for model 3b isochrons because $\partial^2\mathcal{LL}_{3b}/\partial x_i\partial x_j \neq 0$ when $i \neq j$. Therefore, the x_is are interdependent and cannot be solved separately. Although it is possible to jointly optimise the entire $\mathbf{x}$-vector, this takes considerably longer than model 1 and model 3a regression. Fortunately, as explained in Section 5, model 3b regression is rarely or never needed.

315 *Author contributions.* PV is sole author of this contribution

Competing interests. PV is an Associate Editor of this journal

Acknowledgements. This work was completed during a sabbatical stay at the Institute of Geology and Geophysics, Chinese Academy of Sciences, which was supported by CAS' President's International Fellowship Initiative (PIFI). The author would like to thank Prof. Xian-Hua Li of IGG-CAS and Dr. Yang Li of Peking University for hosting him at IGG-CAS. He would also like to thank IsoplotR users

320 Guilhem Hoareau (CNRS, France), Shitou Wu (IGG-CAS, China) and Stijn Glorie (Adelaide, Australia) for discussing the utility of anchored isochrons, and moving this feature up the priority list for IsoplotR.

References

Daëron, M. and Vermeesch, P.: Omnivariant generalized least squares regression: Theory, geochronological applications, and making the case for reconciled $\Delta 47$ calibratibons, Chemical Geology, p. 121881, 2023.

325   Li, Y. and Vermeesch, P.: Inverse isochron regression for Re–Os, K–Ca and other chronometers, Geochronology, 3, 415–420, 2021.

Ludwig, K. R.: On the treatment of concordant uranium-lead ages, Geochimica et Cosmochimica Acta, 62, 665–676, https://doi.org/10.1016/S0016-7037(98)00059-3, 1998.

Ludwig, K. R.: User's manual for Isoplot 3.00: a geochronological toolkit for Microsoft Excel, Berkeley Geochronology Center Special Publication, 4, 2003.

330   McIntyre, G. A., Brooks, C., Compston, W., and Turek, A.: The Statistical Assessment of Rb-Sr Isochrons, Journal of Geophysical Research, 71, 5459–5468, 1966.

Schaen, A. J., Jicha, B. R., Hodges, K. V., Vermeesch, P., Stelten, M. E., Mercer, C. M., Phillips, D., Rivera, T. A., Jourdan, F., Matchan, E. L., et al.: Interpreting and reporting ^{40}Ar/^{39}Ar geochronologic data, Bulletin, 133, 461–487, 2021.

Titterington, D. M. and Halliday, A. N.: On the fitting of parallel isochrons and the method of maximum likelihood, Chemical Geology, 26,

335   183–195, 1979.

Vermeesch, P.: `IsoplotR`: a free and open toolbox for geochronology, Geoscience Frontiers, 9, 1479–1493, 2018.

Wendt, I. and Carl, C.: The statistical distribution of the mean squared weighted deviation, Chemical Geology: Isotope Geoscience Section, 86, 275–285, 1991.

York, D., Evensen, N. M., Martínez, M. L., and De Basabe Delgado, J.: Unified equations for the slope, intercept, and standard errors of the

340   best straight line, American Journal of Physics, 72, 367–375, 2004.