ELSEVIER

Chemical Geology 185 (2002) 191–204

CHEMICAL GEOLOGY
INCLUDING
ISOTOPE GEOSCIENCE

www.elsevier.com/locate/chemgeo

Improving isochron calculations with robust statistics and the bootstrap

Roger Powell*, Janet Hergt, Jon Woodhead

School of Earth Sciences, The University of Melbourne, Melbourne, Victoria 3010, Australia

Received 3 November 2000; accepted 18 October 2001

Abstract

Typical isochron calculations involve using a least squares analysis of the data. If the scatter about a line through the data is perfectly Gaussian, then least squares provides an optimal handling of the data. However, if the data are Gaussian only in the centre of the distribution, but depart from it only slightly in the tails of the distribution, then least squares is not optimal and can easily degrade seriously. For the size of datasets that are used in isochron calculations, such non-Gaussian behaviour is impossible to test for. This is important because there are numerous sources of uncertainty which might result in subtly non-Gaussian behaviour. Therefore, to defend against degradation, it is proposed that isochron calculations be modified by the use of a robust statistical method so that non-Gaussian tail behaviour can be accounted for. For data that are actually Gaussian-distributed, the new isochron calculation method will generally give identical results to least squares. The improvement given by the new method is illustrated by the use of simulations.

Keywords: Isochron calculation; Bootstrap; Non-Gaussian behaviour; Robust statistics; Bootstrap

1. Introduction

Much of what is currently practised in the calculation of isochrons emanates from the classic papers of York (1966) and McIntyre et al. (1966). Their work evolved from what might be called the classical statistics tradition, in which a least squares analysis is undertaken, in the context of analytical uncertainties on both the $\boldsymbol{x}$ and $\boldsymbol{y}$ coordinates of the data in an isochron diagram (York, 1969). The least squares method is optimal if the data scatter about the linear trend through the data is Gaussian-distributed but for the usual size of isotopic datasets, this is impossible to test for. What is needed is an isochron calculation method that works well, in the sense of finding the linear trend through the majority of the data, regardless of the structure of the data scatter (i.e. perfectly Gaussian or not) about the trend. However, the least squares analysis normally used in isochron calculations cannot be relied upon to provide this trend (e.g., Rock et al., 1987). Non-Gaussian behaviour rapidly degrades the least squares analysis, as will be illustrated below.

In a least squares analysis, not only is an age and its uncertainty calculated, but also a statistical measure, MSWD,[1] which was originally used to say whether the

* Corresponding author. Tel.: +61-39-344-6520; fax: +61-39-344-7761.

E-mail address: powell@unimelb.edu.au (R. Powell).

[1] Mean squared weighted deviates, as defined in Eq. (5) of the text.

0009-2541/02/$ - see front matter
PII: S0009-2541(01)00403-X

data should indeed be combined (i.e. have age significance) (McIntyre et al., 1966). Out of this grew a practice that only if the MSWD passes an appropriate statistical test will the calculated age be termed an isochron. Otherwise, it is termed an errorchron (Brooks et al., 1972). The MSWD is used to determine if all the observed scatter in the data is consistent with analytical uncertainties. If MSWD "fails", then geological error is assumed to be involved, and the data as an errorchron are deemed to have less or no age significance. Although the strict application of this isochron–errorchron distinction based on the MSWD is not commonly applied, recognising that geological scatter need not destroy the age significance of data, there is nevertheless a reverence for the MSWD. For example, the MSWD is still used in error-expansion of age uncertainties, and in deciding whether or not error expansion should be applied. However, the MSWD cannot normally be relied upon, as is shown below. This is important because, with increasing MSWD through a certain value, the standard procedure of error expansion involves a step-wise change in age uncertainty. As well as finding reliably the linear trend through the majority of the data, an isochron calculation method is needed that provides age uncertainties that increase smoothly with increasing data scatter.

In this paper, the classical statistical isochron calculation methods are examined and shown to be sub-optimal, and alternative methods are proposed.

2. The least squares method

The isochron methods considered are those in which the data form a linear trend on an $\boldsymbol{x}$–$\boldsymbol{y}$ plot, the slope of the trend relating to the age when the isotope system closed, assuming that this took place at one time and subsequently remained undisturbed. These methods include such commonly employed systems, such as, Rb–Sr, Sm–Nd, Pb–Pb, U–Pb, Lu–Hf and Re–Os. For all real cases, the data defining the trend show some level of scatter, and estimation of the age involves regression of the data to determine the slope.

Considering these data as (column) vectors, $\boldsymbol{x}$ and $\boldsymbol{y}$, the vector of residuals, e, the vertical distance of the points to any line, is given by

$$e = \boldsymbol{y} - \boldsymbol{X\theta}$$

in which $\boldsymbol{\theta}$ is a column vector with two elements, the intercept and slope, and $\boldsymbol{X}$ is a matrix with two columns, the first being a column vector of ones (denoted 1), the second being $\boldsymbol{x}$, so $\boldsymbol{X} = [1\ \boldsymbol{x}]$. The classical "least squares" approach involves finding the $\boldsymbol{\theta}$ that minimises the sum of the squares of the residuals

$$\min_{\boldsymbol{\theta}}\ e^T e$$

in which T denotes the transpose of the vector (or matrix). A wide range of data uncertainty structures can be accommodated by considering the vector of residuals premultiplied by a weight matrix, $\boldsymbol{W}$, so that the least squares problem involves solving

$$\min_{\boldsymbol{\theta}}\ e^T \boldsymbol{W}^2 e. \tag{1}$$

In isochron calculations in which the elements of $\boldsymbol{x}$ and $\boldsymbol{y}$ have uncertainties associated with them, and the elements are correlated, $\boldsymbol{W}$ is diagonal, with diagonal elements, $\boldsymbol{W}_{ii}$

$$\boldsymbol{W}_{ii} = \frac{1}{\sqrt{\sigma_{\boldsymbol{y}_i}^2 + b^2\sigma_{\boldsymbol{x}_i}^2 - 2b\sigma_{\boldsymbol{x}_i}\sigma_{\boldsymbol{y}_i}\rho_{\boldsymbol{x}_i\boldsymbol{y}_i}}} \tag{2}$$

in which b is (an estimate of) the slope, and σ_{xi} is the standard deviation of $\boldsymbol{x}_i$, σ_{yi} is the standard deviation of $\boldsymbol{y}_i$, and $\rho_{x_i}\boldsymbol{y}_i$ is the correlation between $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ (e.g., York, 1969; Wendt and Carl, 1991). Various special cases can be found by making appropriate substitutions in Eq. (2).

The minimisation problem in Eq. (1) can be tackled directly, or in equation form. Thus, differentiating Eq. (1) with respect to $\boldsymbol{\theta}$, gives

$$\boldsymbol{X}^T \boldsymbol{W}^2 e = 0. \tag{3}$$

Substituting for e and rearranging gives the least squares estimator of $\boldsymbol{\theta}$

$$\boldsymbol{\theta} = (\boldsymbol{X}^T \boldsymbol{W}^2 \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}^2 y. \tag{4}$$

Use of Eq. (4) is iterative if the weights depend on $\boldsymbol{\theta}$, as in the case of the York regression using Eq. (4) with Eq. (2). The implementation of the tanh estimator advocated below can also be accommodated in the form of Eq. (4) — see below.

A measure of the scatter of the data about the regression line that features prominently in the geochronological literature is the "mean squared weighted deviates," or MSWD, a statistical measure of the scatter of the data about the line, normalised to the assigned uncertainties on the data. MSWD is simply

$$\mathrm{MSWD} = \frac{e^T \boldsymbol{W}^2 e}{n-2} \quad (5)$$

in which n is the number of data points. More commonly used in the statistical literature is σ_{fit} (or just σ) given by

$$\sigma_{\mathrm{fit}} = \sqrt{\mathrm{MSWD}} = \sqrt{\frac{e^T \boldsymbol{W}^2 e}{n-2}}. \quad (6)$$

In what follows, scatter is referred to in terms of MSWD, while in the equations σ_{fit} is used, it being clear that they are simply related.

In weighted least squares, the estimate of the uncertainties on $\boldsymbol{\theta}$, in the form of the covariance matrix of $\boldsymbol{\theta}$, $V_{\boldsymbol{\theta}}$, is provided by

$$V_{\boldsymbol{\theta}} = (\boldsymbol{X}^T \boldsymbol{W}^2 \boldsymbol{X})^{-1}. \quad (7)$$

The way in which the above least squares equations are normally used in geochronology can be summarised as follows

(1) collect data (make vectors $\boldsymbol{x}$ and $\boldsymbol{y}$)
(2) assign analytical uncertainties (make $\boldsymbol{W}$ with Eq. (2))
(3) calculate age (via the slope, $\boldsymbol{\theta}_2$, from Eq. (4))
(4) calculate MSWD (via Eq. (5))
(5) calculate uncertainty on age (via Eq. (7))

The primary assumptions made in following this approach are that, (a) all the uncertainties in the data are of analytical origin, and, (b), that these uncertainties are Gaussian in form. The first of these says that there is no scatter in the data of geological origin. It is also implicit that the analytical uncertainties are well known. In many studies, this is not true, as discussed in the next section. Yet this knowledge is critical to the use of MSWD in geochronology (e.g., Kullerud, 1991). Assuming that the uncertainties are Gaussian and assuming that the uncertainties are well known, the approach often followed is to compare the MSWD to its distribution under Gaussian uncertainties (Wendt and Carl, 1991), and if this test passes at the 95% level, then the data have age significance, and the uncertainty on the slope is given by Eq. (7). If it fails, then the data may not (or do not) have age significance. As noted earlier, the former case are called isochrons; the latter errorchrons.

Now consider that the weight matrix is multiplied by an unknown factor, f. The factor does not affect $\boldsymbol{\theta}$ (and therefore the calculated age), Eq. (4), but it does appear in the uncertainty on the age. The factor is given by

$$f = \sigma_{\mathrm{fit}} = \sqrt{\frac{e^T \boldsymbol{W}^2 e}{n-2}} \quad (8)$$

and

$$V_{\boldsymbol{\theta}} = f^2 (\boldsymbol{X}^T \boldsymbol{W}^2 \boldsymbol{X})^{-1}. \quad (9)$$

However, if f is to represent scatter over and above that represented by known analytical uncertainties (e.g., unassessed analytical uncertainties, or geological scatter), then f must be bounded below by, for example, $\alpha = 1$ (or, better, a value of α provided via a χ^2 test.[2]) Then, to find the value of f to use in Eq. (9), using Eq. (6), $f \equiv 1$ if $\sigma_{\mathrm{fit}} \le \alpha$, but $f = \sigma_{\mathrm{fit}}$ if $\sigma_{\mathrm{fit}} > \alpha$. For $\sigma_{\mathrm{fit}} \le \alpha$ the age uncertainty is constant, and this corresponds to York Model 1. For $\sigma_{\mathrm{fit}} > \alpha$, the age uncertainty increases with σ_{fit}. This is York Model 2, with the age uncertainty "magnified." So the above summary may be augmented

(4) calculate MSWD (via Eq. (5))
(5) use MSWD to decide if data scatter is appropriate:

- if OK, data define an isochron;
- if not OK, data define an errorchron

(6) calculate uncertainty on age, multiplying this by $\sqrt{\mathrm{MSWD}}$ if it is an errorchron

In the light of Eq. (8), if the MSWD is used to distinguish between isochrons and errorchrons, the importance of correctly assigning uncertainties to the data is obvious, as emphasised, for example, by Kullerud (1991). If the input uncertainties are doubled, say,

[2] $\alpha = \sqrt{2.5}$ is often used.

then $\sigma_{\rm fit}$ is halved, possibly converting an errorchron into an isochron. This is obviously unsatisfactory if input uncertainties are not sufficiently well known, and MSWD is given this special significance. Moreover, the statistical test is only saying that MSWD should be less than a certain value *with 95% confidence*. Using this to make a black–white distinction between errorchron and isochron is clearly inappropriate.

A considerably more profound difficulty with MSWD, however, is that arising from the assumption that the uncertainties on the data are Gaussian-distributed. The MSWD is particularly sensitive to non-Gaussian behaviour, as illustrated below. Such behaviour is commonplace in real data (e.g., Hampel et al., 1986, pp. 25–28). In isotopic data it will arise through the combination of the different sources of analytical and geological uncertainty, as well as from the nature of these uncertainties themselves. Critically, non-Gaussian behaviour is impossible to test for in the small to minute datasets that are typically employed in isochron calculations. Not only does non-Gaussian behaviour affect MSWD, but it compromises the accuracy of the least squares estimator. So alternative methods are needed, as outlined below. With such a method, the shortcomings of least squares in isochron calculations are emphasised, before returning to the behaviour of MSWD.

3. Towards a practical algorithm

The vast majority of the geochronological literature takes a "hard" classical-statistics stance regarding isochron calculations (see Rock et al., 1987, for an exception). However, there are alternatives, and ones that behave much more reliably than least squares with data that depart from Gaussian uncertainties. These robust statistical methods automatically down-weight or even reject outliers, rather than treating all data in the same way as in least squares, or relying on manual cleaning of the data. For an excellent introduction to robust statistics, the reader is referred to Hampel et al. (1986), pp. 1–71 and pp. 397–416.

An intuitive illustration of the problem, and the solution by using a robust method (for example, the tanh estimator outlined below) is provided in Fig. 1. A dataset of 10 points corresponding to 2500 Ma (Pb–Pb dating) was created with Gaussian uncertainties (σ_y= 0.01, representing all of the denominator in Eq. (2)). In Fig. 1, the high-end data point was progressively displaced from the trend, above and below, and various estimators applied to the resulting datasets. The least squares age [(Eq. (2))] simply follows the point, with the MSWD ($\sigma_{\rm fit}$) flagging the extent to which this datum departs from the trend. At some stage, the point will clearly (e.g., visibly) be displaced from the trend through the rest of the data and be considered an outlier. Such a datum may then be removed from the dataset, however this relies on a subjective decision made by the geochronologist performing the calculation. For hard rejection (e.g., manual cleaning) of the point once it is deemed an outlier, the result follows the

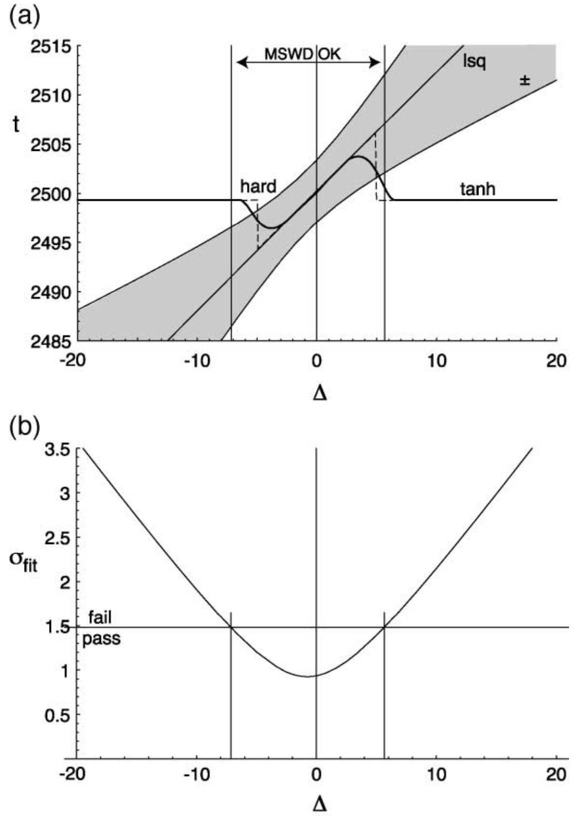

Fig. 1. (a) Age versus departure of a data point from the trend through the data, "Δ" giving the departure as a multiple of σ_y. The point is at one end of the trend, "lsq" is the least squares result, and its 95% confidence band; "tanh" is the tanh estimator result (heavy solid line); and (hard) is the hard rejection result (dashed line), where the point is thrown out once the point is "far" from the line. (b) Corresponding $\sigma_{\rm fit}$ versus departure relationships.

least squares age until the point is omitted, then moves discontinuously to the least squares age given by the dataset without that point. Such discontinuous behaviour, combined with the problem of deciding when to omit the data point (i.e. when is a point an outlier?), makes the approach less than desirable. In contrast, the robust age (tanh), follows the least squares line when the point is close to the trend, but varies smoothly away from the least squares age as the point leaves the trend and the influence of the point is downgraded, eventually being the fit of the data without that data point when the point is determined to be an outlier by the model. Such smooth handling of data off the trend is attractive in an isochron calculation, and has the added advantage that no subjective decisions regarding outliers are required. If the point displaced is in the middle of the trend, the effect on the age is similar to that in Fig. 1 but is less pronounced.

A way of introducing appropriate robust methods is to note that Eq. (3) is a member of a general family of regression estimators called M-estimators (e.g., Hampel et al., 1986, p. 315), which have the form

$$\boldsymbol{X}^T\psi(u) = 0 \tag{10}$$

in which $\boldsymbol{u}$ is the vector of residuals, e, normalised to a scale that reflects the scatter of the data, s. So $\boldsymbol{u} = e/s$. This family includes the estimator advocated below—the so-called tanh estimator, as well as least squares, with $\psi(u) = u$. The calculation of the data scale, s, is considered in the next paragraph. In the context of (implicit), $\psi(u)$ is chosen such that its value does not increase without bound as u increases, as it does with least squares. The motivation for an important group of M-estimators, the so-called redescending estimators (Hampel et al., 1986), is that (a) the optimal properties of least squares need to be utilised for smaller residuals, (b) larger residuals need to be excluded, and (c) intermediate residuals need to have a contribution that is transitional between (a) and (b). This effect can be seen in Fig. 2(a) in which, for $|u| < 1.634$, the contribution is the same as for least squares; for $|u| > 4$, the residual does not contribute, and in the intermediate region there is a smooth transition from a full contribution to no contribution. This transitional region represents residuals that can be considered to be progressively more doubtful and whose influence is therefore progressively downgraded (Hampel et al., 1986, Fig. 1, p. 61).

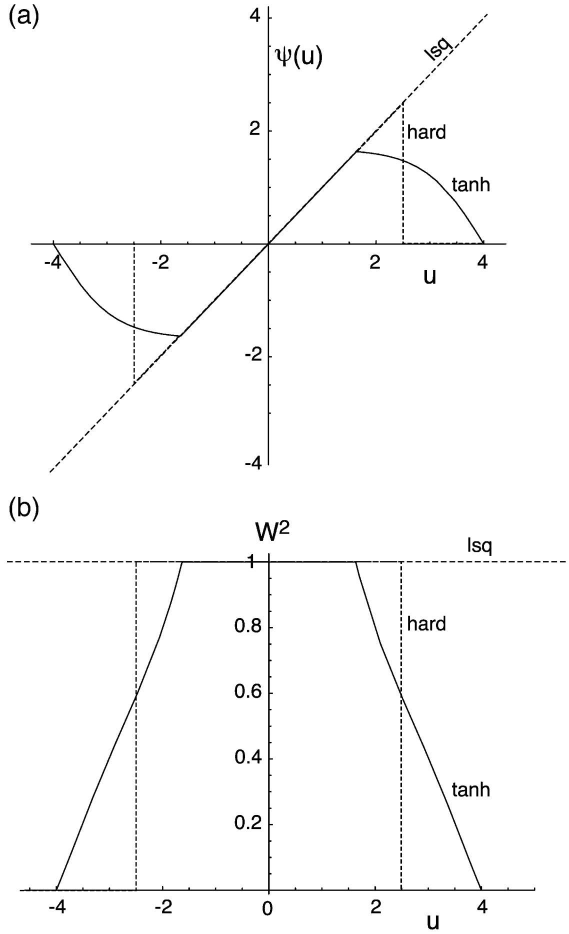

Fig. 2. (a) The function $\psi(u)$ plotted against $\boldsymbol{u}$, in which $\boldsymbol{u}$ is a residual normalised to scale, for least squares (lsq), hard rejection of outliers for $u > 2.5$ (hard; dashed line), and for a redescending estimator (tanh; solid line). The tuning constants used in the tanh estimator are $r = 4$ and $p = 1.634$ (Hampel et al., 1986, Table 2, p. 163). (b) The weighting $\mathbf{W}^2 = \psi(u)/u$ for the w-estimator form of the tanh estimator (solid line), as well as the weighting for least squares and hard rejection (dashed line).

In order to function effectively, methods such as the tanh estimator need to have both a scale to normalise the residuals, and an estimate of the trend of the data. Both of these requirements can be met with highly resistant methods that are little affected by (many) outliers in the data (Rock et al., 1987). The reason why these methods are not in themselves excellent ways of determining isochrons is that they tend to be highly inefficient, meaning that when applied to data that are actually Gaussian, for example in simulations, they have the undesirable effect of producing substantially wider confidence intervals on calculated ages (as in Fig. 5). Application of a redescending estimator from such an inefficient estimator substantially improves the efficiency, particularly for larger samples. The resistant method used here to determine the trend through the data is the least median of squares (LMS), although other such methods could be used. The LMS is found by determining the position of the minimum width band that includes half of the data, the width of the band being measured in the $\boldsymbol{y}$ direction (Rousseeuw and Leroy, 1987, pp. 197–201). The scale estimate, s, used is a standard resistant one based on the median of the absolute values of the residuals, known as nMAD

$$s = 1.4826\left(1 + \frac{5}{n-2}\right)\sqrt{\mathrm{med}\,\boldsymbol{e}^2} \tag{11}$$

in which $\boldsymbol{e}$ is the vector of residuals to the LMS line, $e = \boldsymbol{y} - \mathbf{X}\boldsymbol{\theta}_{\mathrm{LMS}}$. The factor (1.4826) and the term in the number of data points, n, make s approximate the standard deviation for Gaussian-distributed data (Rousseeuw and Leroy, 1987, p. 202). Thus

$$u = \frac{y - \boldsymbol{X}\theta_{\mathrm{LMS}}}{s}. \tag{12}$$

The tanh estimator takes various forms, based on Eq. (10) and the definition of $\psi(u)$ (Hampel et al., 1986, p. 160), Fig. 2(a). A fully iterated form has its problems because (implicit) may have several solutions, or it can blow up from an injudiciously chosen starting $\boldsymbol{\theta}$. Most of the efficiency gains can be made by taking one step from the LMS estimate, rather than carrying out a full iteration. The form of the 1-step tanh estimator used here is a w-estimator, e.g., Goodall (1983), involving one-step based on a rearrangement of (implicit) with u given by Eq. (12). Forming a diagonal weight matrix in terms of u, with diagonal elements, $\boldsymbol{W}_{ii}$

$$\boldsymbol{W}_{ii}^2 = \frac{\psi(u_i)}{u_i} = \begin{cases} 1 & \text{if } |u| < p \\ \frac{|u|-r}{p-r} & \text{if } |u| > p \text{ and } |u| < r \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

with $p = 1.634$ and $r = 4$, as shown in Fig. 2(b). These p and r values are suggested to be reasonable defaults by Hampel et al. (1986), p. 163. Then, in the form of Eq. (3),

$$\boldsymbol{X}^T\boldsymbol{W}^2u = 0$$

so Eq. (4) gives $\boldsymbol{\theta}_{\mathrm{tanh}}$

$$\boldsymbol{\theta}_{\mathrm{tanh}} = (\boldsymbol{X}^T\boldsymbol{W}^2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}^2y. \tag{14}$$

As a general scheme, this can be summarised as follows, with the actual methods proposed in brackets, starting with the collection of data, then

(2) calculate trend through data with a resistant estimator (LMS)
(3) calculate resistant scale of scatter about trend (nMAD)
(4) weigh data according to an efficient robust estimator, based on the scale of scatter about trend (tanh)
(5) adjust trend using weighted data; calculate age from slope of trend.

Note that these robust estimators are based on the actual scatter of the data about the linear trend, and not the analytical uncertainties on the data.

Such a scheme can include the use of other methods, including hard rejection of outliers, for example using

$$\boldsymbol{W}_{ii} = \begin{cases} 1 & \text{if } |u| < 2.5 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

as long as these weights are based on resistant estimates of the trend and the scatter about the trend. Whereas such an approach is already a considerable improvement on least squares, it is less efficient than the tanh estimator (and has the discontinuous behaviour noted above).

4. Simulations

The effectiveness of the tanh estimator, compared with that of least squares, can be best investigated by simulation. The data uncertainty distributions used are

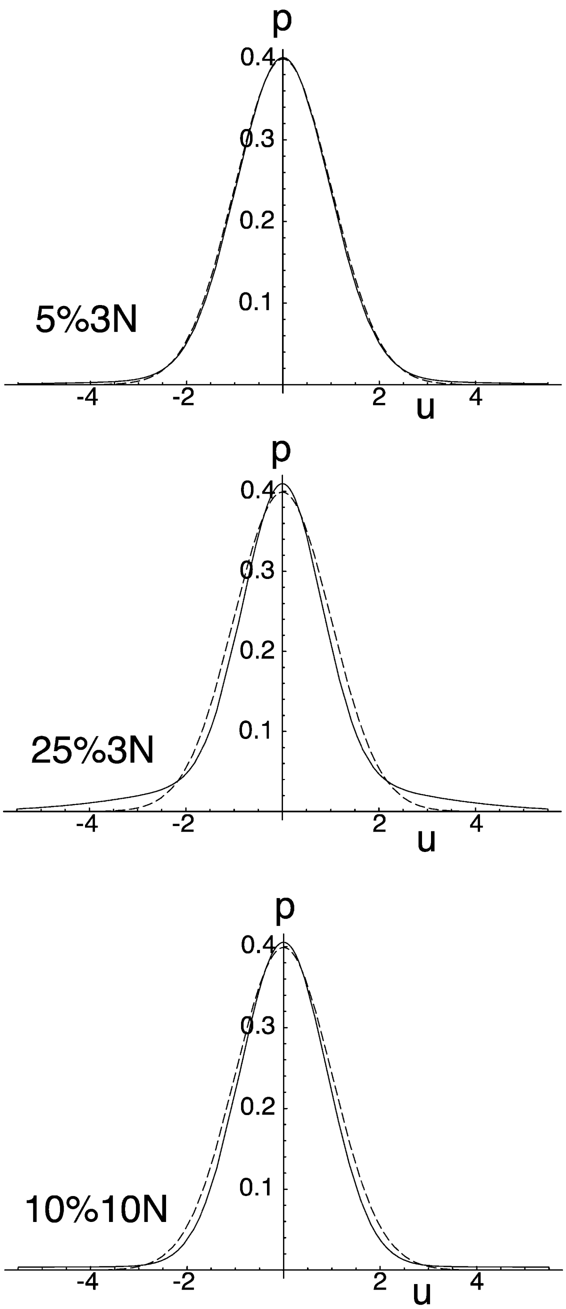

Fig. 3. The four probability distributions used in the simulations, Appendix A, (solid lines), compared with the Gaussian distribution (dashed lines). The distributions are interquartile matched.

described in Appendix A, and are shown in Fig. 3. These distributions were chosen to contribute a small amount of non-Gaussian behaviour to the uncertainties. They are all Gaussian in the centre, but are heavier-tailed than the Gaussian, with tail weights greater than 1 (Rosenberger and Gasko, 1983). Although it is not implied that isotope datasets have this precise uncertainty structure, our point is that such datasets are typically too small to know their true distribution with any certainty. Thus, if least squares cannot adequately handle data with such a structure, then an alternative needs to be found. An example of a comparison between least squares and the tanh estimator is shown in Fig. 4 for 10% 10N uncertainties, with this notation for uncertainties explained in Appendix A. The figure illustrates how much better the tanh estimator handles this moderate departure from the Gaussian compared with the least squares approach. A full comparison, in terms of the width of 95% confidence intervals on age, normalised to the interval for least squares and Gaussian-distributed data, is shown in Fig. 5. The open circles indicate the tail weight, for given sample size, at which the tanh estimator becomes more efficient than least squares. As can be seen, for larger sample sizes, this is at very small tail weight, meaning that even at small departures from Gaussian-distributed data, least squares has been sufficiently degraded to be less good than robust methods like the tanh estimator. As the sample size decreases, the transition tail weight becomes rather larger, and it is likely that for sample sizes, $n < 5$ or 6, an alternative approach to the tanh estimator will be required.

It could be argued that if MSWD properly flags datasets that are not Gaussian, and the MSWD is used to contribute to the width of the 95% confidence intervals on age when least squares is used, then least squares might still be viable. This can also be investigated with simulation, and the results are plotted in Fig. 6. This shows that MSWD is unreliable, even with only small departures from Gaussian behaviour. For example, for 10 data points with a structure 25% 3N, the MSWD has lost any predictive power with 35% of the simulations flagged as failing the statistical test, whereas the least squares results have only degraded marginally, with confidence intervals half as big again as for Gaussian-distributed data, Fig. 5.

In conclusion, MSWD is a flawed measure of data scatter, unless the data are *known* to be Gaussian, a fact

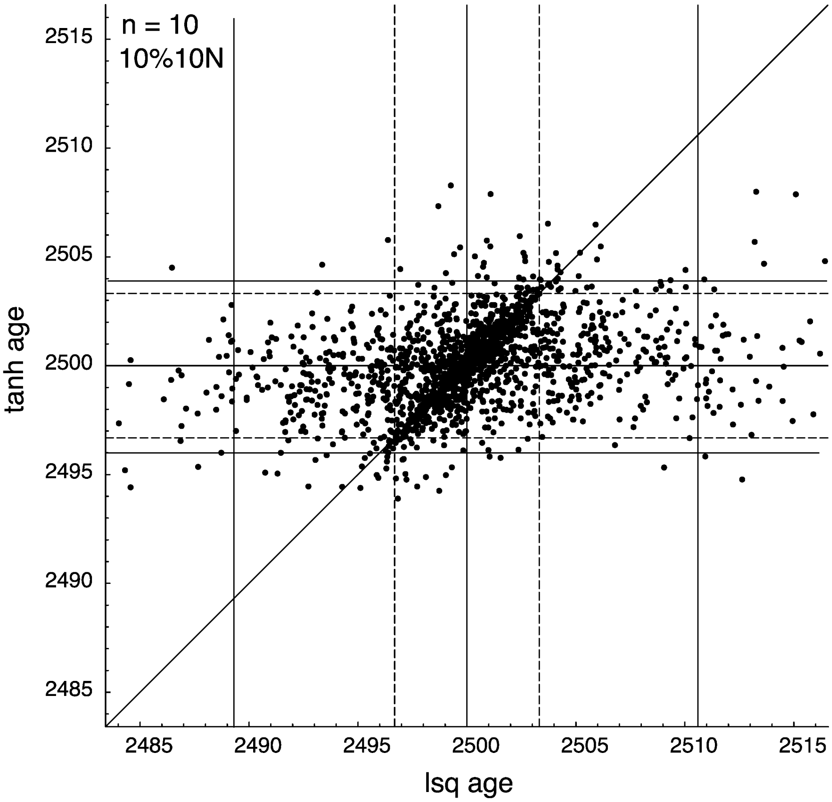

Fig. 4. Comparison of ages calculated by least squares (lsq) and the tanh estimator (tanh), for 2000 10-point isochrons, for uncertainties generated from 10% 10 N (see Appendix A). The dashed lines are the expected 95% confidence limits if the data were Gaussian distributed; the solid lines are the actual 95% confidence limits from simulations involving 9999 isochrons.

we argue cannot be determined. Contrary to common assumption, it does not provide the means of determining whether or not the scatter is contributed by analytical uncertainties alone, even if these are well known. Nor does it provide a reliable measure in the context of error expansion of age uncertainties. Given that the departures from Gaussian-distributed data as portrayed on the x-axis of Fig. 5 are not unlikely, it is preferable to abandon least squares and to adopt a robust estimator like the tanh estimator.

5. Uncertainties in isotopic data

While it is argued here that the importance of analytical uncertainties in isochron calculations cannot survive in the light of the discussion and results above, these uncertainties still play a role in isochron calculations. Although a critical ingredient in the type of calculations proposed here is the use of the actual scatter of the data to provide a scale for the scatter, it is important to use some estimate of the analytical uncertainties to provide a minimum scale. This is to guard against the data fortuitously having a small scatter, a problem that can easily arise for small samples.

Scatter of data is produced by a combination of analytical and geological factors, as discussed below. From the point of view of data having age significance, and in the light of the robust statistical approach followed above, a dataset is required to have a backbone of data that define a trend whose slope reflects the age. Robust statistical methods are good at finding such trends, regardless of the scatter about the trend, and particularly regardless of outliers from the trend. The backbone of data, which in the limit of larger datasets need only involve half of the data, not only

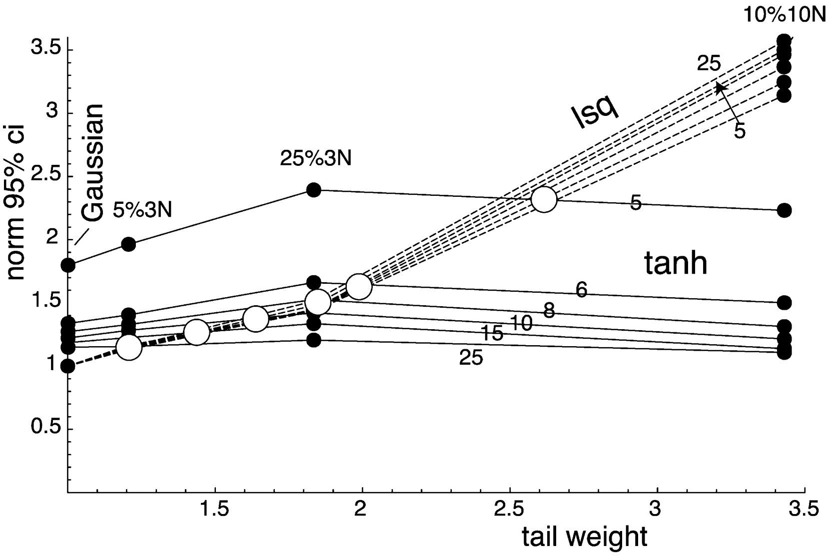

Fig. 5. The width of the 95% confidence interval on age, normalised to that of the least squares 95% confidence interval for Gaussian distributed uncertainties on the data, plotted against tail weight. Dashed lines are for least squares; solid lines are for tanh. The simulation results are plotted as bullets; the lines joining them are indicative; the line labels are the dataset sizes which vary from 5 to 25 data points.

provides the slope of the trend, but the scatter about the trend. Outliers, i.e. points off the trend, can be anything from adjacent to the trend, to gross outliers far from the trend. The former are of more concern as they cannot be reasonably or easily excluded from the calculations, whereas the latter can be discarded. The Gaussian mixture distributions such as 25% 3N (Appendix A) provide a way of simulating scatter with such moderate

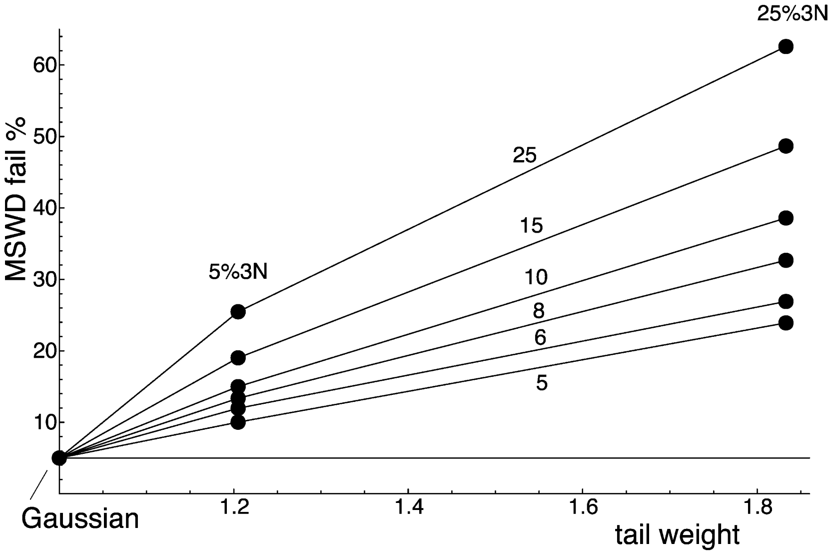

Fig. 6. The percentage of simulations in which MSWD fails plotted against tail weight for the two least non-Gaussian situations studied. The simulation results are plotted as bullets; the lines joining them are indicative; the line labels are the dataset sizes.

outliers. As shown in the last section, small departures from Gaussian behaviour have significant effects on simple statistics such as MSWD.

Analytical uncertainties arise from contributions from sample preparation (e.g., weighing errors, blank correction) through to the analysis itself (e.g., mass fractionation correction, errors relating to signal size). Of these, the instrumental factors are generally monitored routinely as the acquisition of an individual isotope ratio represents the mean of numerous separate measurements during a run. The analyst is thus presented with a value and associated uncertainty quoted as "in run" or "internal" precision (2σ). Additional factors (e.g., amplifier gain calibrations, collector efficiencies) then contribute further error, quantifiable as "external" precision or "reproducibility," normally measured on standard reference materials.

It could be argued, however, that there is no easy way of accurately determining true analytical uncertainties for real samples unless, for example, the analyst is prepared to undertake multiple analyses of each data point (thereby incorporating all sources of random errors). As a result, a commonly adopted strategy in isochron calculations is to apply an error which is proportional to the ratio, e.g. 0.1% on Pb isotope ratios, 0.005% on Sr isotope ratios, 0.5% on Rb/Sr, and so on.

Whereas it is reasonable to suppose that analytical uncertainties have a Gaussian form in the centre, it is impossible to determine whether or not they are Gaussian in the tail. In the light of the above, even in the absence of geological error, MSWD is still likely to fail if analytical uncertainties have tails that are heavier than Gaussian. For this reason alone, MSWD is unhelpful.

Geological error reflects a breakdown in the assumptions involved in an "ideal" isochron calculation: (1) that the individual samples analysed (rocks, minerals) started with precisely the same isotopic composition, (2) that the samples closed isotopically at the same time, and (3) that the samples all remained closed to isotope exchange since that time. Clearly, it is undesirable to construct isochrons from data that are grossly affected by such deviations from ideality. As noted above, the critical feature of the data for the bias on the age to be within a confidence interval on the intended age (i.e., has age significance) is that the effect of the geological factors does not significantly affect more than half of the data. Geological error, if present, may have any distribution, and the assumption that it is Gaussian is unwarranted. Certainly, the distribution can easily be such that the data no longer give an age that has any significance, through for example being skewed or biased.

While we would suggest that true uncertainties on individual data points can, at best, only be approximated, it is certainly possible to provide a measure of the minimum value. For the purpose of the isochron calculations proposed here, estimates of minimum values for the analytical uncertainties are all that must be assigned (for scaling purposes; see earlier in this section), as the robust estimators are based on the actual scatter of the data about the linear trend and not the analytical uncertainties. The assignment of precise values depends upon the application and analytical protocol in question (e.g., mass fractionation effects in the case of Pb data, blank correction in the case of low-level samples). Nevertheless, we would suggest that, for many purposes, the 2σ within-run precision reported for each analysis may form a useful starting point for the approach described herein, as the minimum analytical uncertainties need only be approximations for the purposes of scaling.

6. Age uncertainties: bootstrap confidence intervals and pictures

The isochron calculation problem is semiparametric in that the functional form of the data is specified to be linear (as the data are deemed to have age significance), but the distribution of the residuals is not specified. In this situation, the bootstrap (Hall, 1992; Davison and Hinkley, 1997) provides a way of obtaining a confidence interval on the age, given that the classical statistics uncertainties on parameters are dependent on the data being Gaussian-distributed. There are two ways of applying the bootstrap to fit a line to data: case resampling and error resampling (Davison and Hinkley, 1997, pp. 261–266).

In case of resampling, used in an isochron calculation by Kalsbeek and Hansen (1989), the $\{\boldsymbol{x},\boldsymbol{y}\}$ pairs that constitute the data are resampled. For n data pairs, consider an urn containing n balls each with a different $\{\boldsymbol{x},\boldsymbol{y}\}$ pair written on it. One resample involves repeating n times the taking of a ball from the urn, noting its $\{\boldsymbol{x},\boldsymbol{y}\}$ pair, then returning it to the urn. In other words, $n\{\boldsymbol{x},\boldsymbol{y}\}$ pairs are chosen at random with replacement. The isochron calculation is then applied to the resample. This process is then repeated, for say $B=999$ resamples. The collected B isochron results are then processed to give the slope (i.e., age) uncertainty. The data "design", the way the points are distributed along the linear trend, will vary between resamples, and will vary from the original data. Taking the view that the data design is intrinsic to the data, and needs to be preserved in the resample, the error resampling method is preferred here. In this, the distribution of the points along the trend is taken as a given, and it is to these points that the resampled residuals are applied. To start with, the predicted values of $\boldsymbol{y},\hat{y}=\mathbf{X}\boldsymbol{\theta}_{\mathrm{tanh}}$, are calculated. The residuals, $e=y-\hat{y}$, are the focus of the resampling. Each of B bootstrap datasets is formed from $\boldsymbol{x}$ and $y^*=\hat{y}+e^*$, in which e^* is a resample of the residuals. This resample is chosen at random with replacement from e. The bootstrap dataset is then fitted to give $\boldsymbol{\theta}^*$, and a scale s^*. Focussing on the slope, the original tanh slope is $b=\boldsymbol{\theta}_{\mathrm{tanh2}}$ and the bootstrap slope is $b^*=\boldsymbol{\theta}_2{}^*$. The pivotal statistic, ϕ, is defined as

$$\phi=\sqrt{n}\left(\frac{b^*-b}{s^*}\right). \tag{16}$$

This process is repeated $B=999$ times, say, giving 999 ϕ values. These ϕ values are then sorted, giving ϕ^s, and the 95% confidence interval of ϕ is simply the value of the 25th, $\phi_{25}{}^s$, and the 975th, $\phi_{975}{}^s$, elements. These values are converted back to slopes to give a confidence interval

$$\left\{b-\frac{s\phi_{975}^s}{\sqrt{n}},b+\frac{s\phi_{25}^s}{\sqrt{n}}\right\}$$

referred to as the percentile-t bootstrap 95% confidence interval of the slope, b, (Hall, 1992, pp. 16–17). This bootstrap confidence interval is effectively independent of the data uncertainties because it depends on the size of the residuals, which remain more or less constant as the data uncertainties are varied. The bootstrap estimate of slope uncertainty is undertaken using a pivotal statistic, rather than bootstrap slopes directly, because this reduces bias on the slope uncertainty.

Ideally, in the fitting of each resample, the fitting to get the slope, b^*, and the associated scale, s^*, would be undertaken using the same approach used to fit the original data (i.e., LMS, followed by tanh, for b^*, and using nMAD for s^*). However, a simpler approach that is much faster computationally is advocated, supported by a comparison with the ideal approach based on simulations. The simpler approach involves first cleaning the data of outliers, identified via the original robust fit of the data. The residuals associated with outliers are not included in the resampling. The fitting of the bootstrap datasets is then done using least squares, with s^* being the usual least squares estimate (i.e., σ_{fit}, Eq. (6)). The residuals used are chosen to have $|\boldsymbol{u}|\leq 2.5$, Eq. (12), in which u is formed from the residuals divided by the nMAD of the residuals. However, if this nMAD is smaller than the average of the analytical uncertainties

$$\frac{1}{n}\sum_i^n\sqrt{\sigma_{\boldsymbol{y}_i}^2+b^2\sigma_{\boldsymbol{x}_i}^2-2b\sigma_{\boldsymbol{x}_i}\sigma_{\boldsymbol{y}_i}\rho_{\boldsymbol{x}_i\boldsymbol{y}_i}}$$

then this average is used instead of the nMAD of the residuals. This is to guard against a situation in which the scatter on the data is smaller than that expected with the analytical uncertainties.

In addition to providing a confidence interval, bootstrap results of age may be viewed as a histogram (Fig. 7). This will show whether the distribution of the ages is symmetric or not, and whether it is "fat-tailed" compared to a Gaussian-distribution. An aid in the viewing of histograms such as Fig. 7 is provided by smoothing using a kernel estimator (e.g. Hall 1992; Wand and Jones, 1995). Working with the ϕ values generated above, the probability distribution of ϕ, $p(\phi)$, is built up by replacing each of the B ϕ values by a shape (kernel), centred on the value. The shape used here is the Epanechnikov kernel, $3/4(1-\boldsymbol{x}^2)$ with $-1<\boldsymbol{x}<1$, (Wand and Jones, 1995, p. 30). $\boldsymbol{x}$ is centred on ϕ and scaled so that the resulting probability distribution is smooth. Although approaches have been suggested to automate the determination of this scaling, trial and error is still advocated (Wand and Jones, 1995, p. 58). For Fig. 7, $p(\phi)$ is transformed back to the age using the same equation used in generating the confidence interval, for comparison with the histogram of ages. Such bootstrap probability distribution diagrams are called "confidence pictures" (Hall, 1992).

For larger datasets, the histograms/confidence pictures are more symmetric, and in fact, the reweighted least squares confidence interval often agrees relatively well with the bootstrap result. However, for

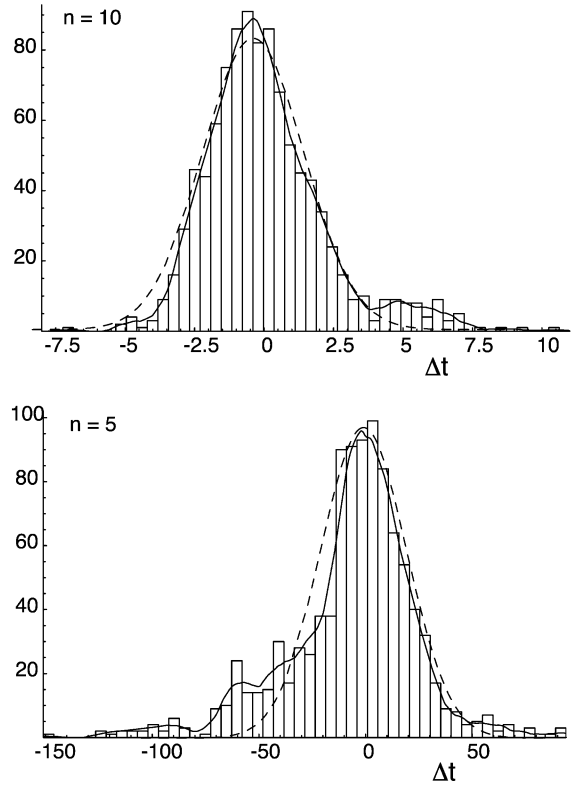

Fig. 7. Example histograms ($B = 999$) for simulated datasets (with $\sigma_y = 0.01$) and $n = 10$ and $n = 5$, showing the confidence picture (solid line) and the Gaussian approximation (dashed line). The y-axis is the probability, p.

small datasets, say for $n = 6$ or less, the reweighted least squares confidence interval provides a poor approximation of the bootstrap result, and the histograms can be quite asymmetric, or even bimodal. The small dataset situation is discussed further below. Two examples are shown in Fig. 7.

7. Discussion and conclusions

The approach advocated here can be summarised as follows

(1) collect data
(2) assign analytical uncertainties
(3) calculate age (using LMS, nMAD and 1-step w-estimator tanh)
(4) calculate age uncertainty using the bootstrap

These calculations are available from the authors as a stand-alone application for various computer platforms, as well as a Mathematica™ function.

In the calculation of the age, this approach represents a modification of the standard least squares procedure, in that data which scatter too much from the trend have their effect on the age automatically and smoothly downplayed. If there is no such scatter, the least squares and tanh estimators are expected to give the same ages, and many published isochron datasets do give identical results. With increasing scatter, the tanh estimator gives results that still follow the main linear trend in the data, whereas least squares may well not. Thus, we would suggest that the tanh estimator should be used routinely, as a precaution, instead of least squares.

In the calculation of the age uncertainty, the bootstrap gives the same result as the standard least squares procedure if the data scatter corresponds to the analytical uncertainties. With increasing scatter, the age uncertainty increases smoothly, in contrast to error expansion.

MSWD is not used in the approach advocated here, having been shown above that this measure of scatter is very sensitive to the nature of the scatter, particularly for scatter arising from probability distributions that are more long-tailed than Gaussian. Aspects of isochron calculations that rely on MSWD may well not work as expected, including error expansion and the isochron–errorchron distinction. Whereas datasets that give a low MSWD (e.g., less than 2.5) may reasonably be assumed to involve scatter dominated by essentially Gaussian analytical uncertainties, larger values may be due to non-Gaussian behaviour as well as geological error. Intermediate values of MSWD are likely to arise from uncertainties that are heavier-tailed than Gaussian analytical uncertainties and/or small geological contributions, whereas large values are likely to be dominated by geological contributions.

One example of the application of the new approach is presented here. The example chosen, using literature data from Russell (1995), is one in which there is a difference between the results of the least squares and tanh estimators (Fig. 8). Whereas the Nicolaysen plot in Fig. 8(a) shows that there is some data scatter about the linear trend, the Provost-like plot (Provost, 1990) in Fig. 8(b) brings out the nature of the scatter. It shows that the York model 2 age is pulled towards the two

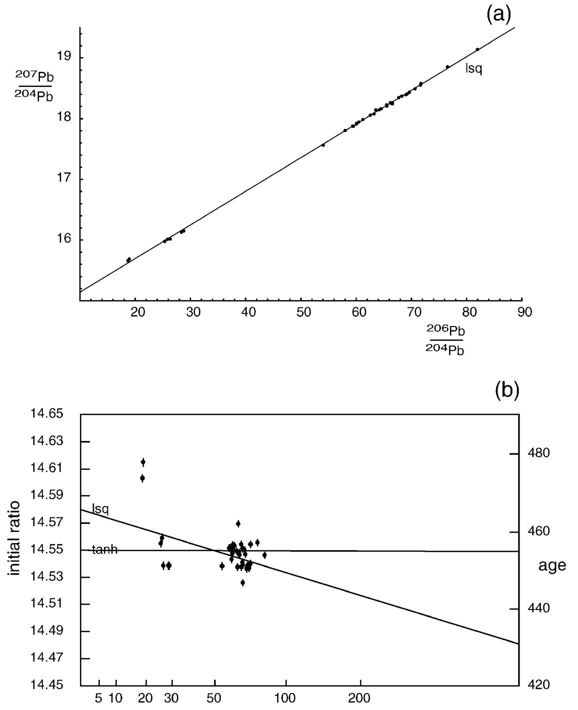

Fig. 8. Example ($^{207}Pb/^{204}Pb$)–($^{206}Pb/^{204}Pb$) isochron using data from Russell (1995), using all 36 data points, plotted (a) in the standard way, and (b) in the manner of Provost (1990), with ($^{207}Pb/^{204}Pb$) intercept on the left upright, and age on the right upright. "lsq" is a York model 2 age of 431 ± 15 Ma, while "tanh" is the tanh estimator age of 455 ± 11 Ma. See text for discussion.

lowest ($^{206}Pb/^{204}Pb$) values, whereas the tanh estimator treats these data as outliers, and downweights their influence on the age. In this case, it is a moot point whether these two points are outliers, or the next five lowest ($^{206}Pb/^{204}Pb$) values. The least squares method has put an average, and possibly meaningless, line through the data, whereas the tanh estimator has recognised that there is structure in the data, and acted accordingly. Whereas it is clear that manual removal of the two points will give the same result as the tanh estimator, the latter provides the result directly and "objectively" (in the context of the definition of this robust estimator), without the inevitable subjectivity of choosing which data to remove manually.

The ideal isochron calculation method would not only behave in a reliable and smooth way with increasing scatter, but also with sample size. The robust method advocated here will fail to behave appropriately with sample sizes less than 5 or 6. It is an open question what will provide the best method at small sample sizes, but it is reasonably clear that this will not be the least squares method. Whatever the method, it is likely to be much more dependent on the uncertainties assigned to the data than in the approach outlined in this paper.

Acknowledgements

We would like to thank Ken Ludwig for his detailed and thoughtful review, his enthusiasm and support in this work, and for forcing us to explain properly what we are trying to do.

Appendix A. Simulations

Simulations of isochron data were made varying

(1) size of dataset ($n = 5,6,8,10,15,25$)
(2) distribution of the uncertainties on the data (N, 5% 3 N, 25% 3 N, 10% 10 N—see below).

for a fixed data "design." The range in $\boldsymbol{x}$ used is 20–80, $\boldsymbol{y}$ prior to applying uncertainty is generated from the $\boldsymbol{x}$ and a $\boldsymbol{\theta}$ of $\{12.5,0.164261\}$ (corresponding to an age of 2500 Ma in PbPb dating). The data uncertainties, assigned only to $\boldsymbol{y}$, are based on $\sigma_y = 0.01$. An equal spacing of data points was used. In each of the 24 combinations, 9999 datasets were generated, and each was fitted with least squares (lsq) and with the tanh estimator (tanh). The focus was to look at the MSWD, in the lsq case, and to look at the 95% confidence interval on the calculated ages in both the lsq and tanh. The confidence interval in each of the 24 combinations was obtained by simply choosing the 250th and 9750th values of the 9999 ordered ages.

The uncertainty structures used were (a) the Gaussian distribution (N), (b) a Gaussian distribution contaminated with 5% of a Gaussian distribution with a standard deviation three times larger (5% 3N), (c) a Gaussian distribution contaminated with 25% of a Gaussian distribution with a standard deviation three times larger (25% 3 N), and (d) a Gaussian distribution contaminated with 10% of a Gaussian distribution with a standard deviation 10 times larger (10% 10N). The distributions are shown in Fig. 3. This series of distributions are in order of increasing tail weight, having thicker tails than the Gaussian distribution. The tail weights are, in order, 1, 1.205, 1.833, and 3.429 (Rosenberger and Gasko, 1983). These are used in the figures representing the simulations to reflect departure from the Gaussian. In the simulations, the distributions were normalised to match the interquartile range of the Gaussian distribution: the multipliers with respect to the Gaussian scale are 0.9624, 0.8118 and 0.8951, respectively.

References

Brooks, C., Hart, S.R., Wendt, I., 1972. Realistic use of two-error regression treatments as applied to Rubidium–Strontium data. Reviews of Geophysics and Space Physics 10, 551–577.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and their Application. Cambridge Univ. Press, Cambridge, 582 pp.

Goodall, C., 1983. M-estimators of location: an outline of the theory. In: Hoaglin, D.C., Mosteller, F., Tukey, J.W. (Eds.), Understanding Robust and Exploratory Data Analysis. Wiley, New York, pp. 339–403.

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer-Verlag, New York, 352 pp.

Hampel, F.R., Rousseeuw, P.J., Ronchetti, E.M., Stahel, W.A., 1986. Robust Statistics. Wiley, New York, 502 pp.

Kalsbeek, F., Hansen, M., 1989. Statistical analysis of Rb–Sr isotope data by the "bootstrap" method. Chemical Geology (Isotope Geoscience Section) 73, 289–297.

Kullerud, L., 1991. On the calculation of isochrons. Chemical Geology 87, 115–124.

McIntyre, G.A., Brooks, C., Compston, W., Turek, A., 1966. The statistical assessment of Rb–Sr isochrons. Journal of Geophysical Research 71, 5459–5465.

Provost, A., 1990. An improved diagram for isochron data. Chemical Geology (Isotope Geoscience Section) 80, 85–99.

Rock, N.M.S., Webb, J.A., McNaughton, N.J., Bell, G.D., 1987. Nonparametric estimation of averages and errors for small datasets in isotopic geoscience: a proposal. Chemical Geology (Isotope Geoscience Section) 66, 163–177.

Rosenberger, J.L., Gasko, M., 1983. Comparing location estimators: trimmed means, medians and trimean. In: Hoaglin, D.C., Mosteller, F., Tukey, J.W. (Eds.), Understanding Robust and Exploratory Data Analysis. Wiley, New York, pp. 297–338.

Rousseeuw, P.J., Leroy, A.M., 1987. Robust Regression and Outlier Detection. Wiley, New York, 329 pp.

Russell, J., 1995. Direct dating of Silurian macrofossils from Gotland, Sweden. In: Dunay, R.E., Hailwood, E.A. (Eds.), Non-Biostratigraphical Methods of Dating and Correlation. Geological Society Special Publication, vol. 89, pp. 175–200.

Wand, M.P., Jones, M.C., 1995. Kernel Smoothing. Chapman & Hall, London, 212 pp.

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

York, D., 1966. Least squares fitting of a straight line. Canadian Journal of Physics 44, 1079–1086.

York, D., 1969. Least squares fitting of a straight line with correlated errors. Earth and Planetary Science Letters 5, 320–324.