ELSEVIER

Earth and Planetary Science Letters 128 (1994) 373–390

EPSL

Mixture modeling of multi-component data sets with application to ion-probe zircon ages

M.S. Sambridge, W. Compston

Research School of Earth Sciences, Institute of Advanced Studies, Australian National University, Canberra, ACT 0200, Australia

Received 15 December 1993; accepted 19 August 1994

Abstract

A method is presented for detecting multiple components in a population of analytical observations for zircon and other ages. The procedure uses an approach known as mixture modeling, in order to estimate the most likely ages, proportions and number of distinct components in a given data set. Particular attention is paid to estimating errors in the estimated ages and proportions. At each stage of the procedure several alternative numerical approaches are suggested, each having their own advantages in terms of efficiency and accuracy.

The methodology is tested on synthetic data sets simulating two or more mixed populations of zircon ages. In this case true ages and proportions of each population are known and compare well with the results of the new procedure. Two examples are presented of its use with sets of SHRIMP $^{238}U-^{206}Pb$ zircon ages from Palaeozoic rocks. A published data set for altered zircons from bentonite at Meishucun, South China, previously treated as a single-component population after screening for gross alteration effects, can be resolved into two components by the new procedure and their ages, proportions and standard errors estimated. The older component, at 530 ± 5 Ma (2σ), is our best current estimate for the age of the bentonite.

Mixture modeling of a data set for unaltered zircons from a tonalite elsewhere defines the magmatic $^{238}U-^{206}Pb$ age at high precision ($2\sigma \pm 1.5$ Ma), but one-quarter of the 41 analyses detect hidden and significantly older cores.

1. Introduction

The advent of rapid isotopic age determinations, for example through the use of the SHRIMP ion microprobe and of laser fusion gas extraction for $^{39}Ar-^{40}Ar$ dating, presents the problem of assessment of multiple age estimates for a single geological sample. For conventional U–Pb age determinations on zircon, assessment is made using comparatively few analyses per rock sample, with the emphasis placed on whether or not the coupled $^{238}U-^{206}Pb$ and $^{235}U-^{207}Pb$ ages are concordant. Even if just one of the few analyses is concordant, this is usually taken as proof that the age is correct. In contrast to this, SHRIMP geochronology for zircons that are younger than, say, 1000 Ma, does not often yield $^{235}U-^{207}Pb$ ages that are sufficiently precise for comparison with $^{238}U-^{206}Pb$. Consequently, the question of the validity of the age determination must be based on finding (ideally) that all replicate analyses are in agreement to within experimental error.

Inevitably, there will be outliers; both zircons that are too old, due to inheritance from the rock's magmatic source, and those that have not remained chemically closed, owing to later geochemical events. The identification of 'outliers'

SSDI 0012-821X(94)00179-0

has always been a troublesome area in statistics but it can be managed by amateurs if the data set is dominated by a single population with only a small minority of analyses that are much older or much younger. However, the real problem area is resolution of two or more populations of zircons which each comprise a sizeable proportion of the total that are *nearly equal* in their ages. Geologically, this is (or should be) quite a commonplace expectation; for example, for zircons in tuffs from a region with a long, continuous period of volcanism. It is quite likely that a given tuff sample will be a sedimentary mixture of detrital zircons from slightly older volcanics and those from the period of volcanism whose age is to be found. In our experience, detrital grains from older volcanics need show no surface abrasion features whatever, so that the older age of such zircons cannot be determined reliably by visual inspection prior to analysis.

These issues have been faced and solved by practitioners of fission-track dating (e.g., [1]), which has much in common with SHRIMP ^{238}U–^{206}Pb geochronology. In both methods many individual grains from a single rock are analysed, there can be unequal variance per analysis (due to highly variable U concentrations) and the age differences to be resolved are comparable with experimental precision. In addition, the precision of the latter is low in comparison with the extreme precision obtained for the much larger samples used for isotope dilution analysis and thermal ionization. Galbraith and Green [1] have shown how mixed ages can be resolved into two principal components and the proportions of each estimated, by a mathematical analysis based on probabilities and goodness-of-fit.

The need for an objective and quantitative approach to this problem has been most felt in SHRIMP ^{238}U–^{206}Pb dating applied to Phanerozoic time-scale calibration. Here the best samples geologically are (marine) tuffs interbedded with sediments that carry the diagnostic fossil fauna. However, such samples are especially susceptible to mixing with foreign zircons from detrital grains from 'normal' sedimentation and with those from wall-rock inclusions in the explosively produced volcanic ash. They are also susceptible to the effects of the low temperature geochemical alteration that accompanies the later alteration of volcanic glass to clay minerals, and to the later movement of pore fluids through the clays, which can interact with the original zircons. As a result outliers may be present that are younger than the event of interest. These may be in the physical form of either a thin rim of new zircon that overlies the magmatic zircon, or as rims and internal zones of altered zircon that have lost radiogenic Pb and/or gained U.

This paper addresses the problem of estimating the number of distinct components and their ages in a collection of zircons, although the procedure is quite general and can be used for any problem where a histogram of data is available and a multi-peaked distribution is sought. We have paid particular attention to estimating the likely errors in these parameters and several approaches are discussed. Since most numerical methods have a limited range of applicability, we feel it is more useful to draw conclusions based on the results of several competing methods than to rely on a single approach. For this reason a set of alternative methods are presented at each stage. One of these is closely related to that proposed by Galbraith and Green [1] for fission-track dating, while others represent an extension of their approach to incorporate different types of error statistics.

An attempt has been made to keep the paper largely self contained (with much of the mathematical detail placed in the appendices). Our main objective is to convey the methodology and an appreciation of the underlying theory, without forcing the reader to study extensive algebra. An introduction to mixture modeling appears in the next section, followed by full details of the estimation procedure in section 2.2. An efficient method is presented for the case of Gaussian error statistics and also non-Gaussian (more robust) statistics, together with examples. Section 2.4 deals with the estimation of errors in the maximum likelihood set of ages and proportions via the covariance matrix and confidence ellipsoid about the solution. Section 2.5 is concerned with estimating the most likely number of age components in the data; again several approaches are

described and all are illustrated with numerical examples. In section 3 the methodology is applied to the re-interpretation of the complex data set published for zircons from the Meishucun Bed 5 horizon in South China which has been previously interpreted using traditional single-component methods [2]. A second application to a real (unpublished) zircon data set appears in section 4. The comparison of real and synthetic data examples allows insight into how well the methods presented are likely to work in practice.

2. Mixture modeling for zircon ages and proportions

2.1. Theory: formulation of the maximum likelihood approach

The estimation of distinct age components from a sample of zircon ages is an example of what statisticians call a mixture modeling problem. (See [3] and the references therein for a review of the statistical literature on mixture modeling.) This problem can be formulated in the following way. If one has a set of n 'measurements' (in this case estimated ages), a_i, $(i = 1, \cdots, n)$ and their associated standard errors, σ_i, $(i = 1, \cdots, n)$, then we wish to determine the number of true age components, n_c, their values, t_j, $(j = 1, \cdots, n_c)$, and the proportion of the n measurements that belong to each class, π_j, $(j = 1, \cdots, n_c)$.

By definition, the sum of the proportions is unity:

$$\sum_{j=1}^{n_c} \pi_j = 1 \qquad (1)$$

and so we have a total of n data and $(2n_c - 1)$ unknowns. The value of n_c is unknown but we shall treat it as a known constant for now. In practice, we may have to repeat the estimation of the ages and proportions for different values of n_c. (Although we use the term 'measurement' for the age estimates, the values are usually the result of a set of measurements and calculations, which may be quite involved.)

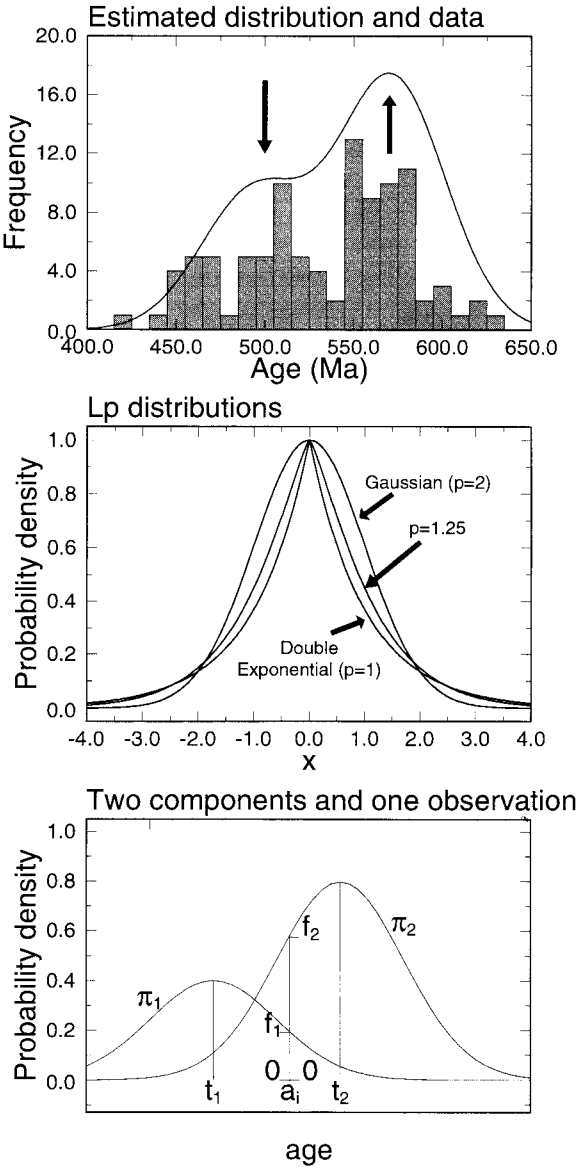

Fig. 1. Top panel: synthetic data set and estimated two-component distribution. The arrows indicate the two true ages. Middle panel: Three distributions for different values of p in Eq. (7). Lowest panel: a single measured age a_i and two distinct age components. The relative probabilities, f_1 and f_2 of it belonging to each age component is given by the height of each curve.

In practice, we will be presented with a histogram of measured ages, such as that in Fig. 1 (top panel). The question is then, does this repre-

sent a single population with an acceptable spread, given some knowledge of the standard error in the measured age, σ_i, or is it the result of several overlapping populations? If the number of data is large enough (and the component separated enough) to resolve two or more obvious peaks in the histogram then there is clear evidence for multiple populations and no need for mixture modeling. However, if this is not the case, then mixture modeling provides a way of finding the best fit (maximum likelihood) set of ages and proportions for any number of assumed components. To proceed it is necessary to know (or make a good guess at) the error distribution of the observations, because it is this distribution that will be used for the fitting procedure. Usually this distribution is assumed to be Gaussian but the theory presented here will hold for any other type of distribution, for example, the more robust distributions shown in Fig. 1 (middle panel)—robust means less affected by outliers.

Once the error distribution associated with each measured age is known, a way of choosing the 'best fitting' set of (true) ages and proportions is required. Since the distributions from each real age component overlap, every measured age could actually have come from any of the true age components. For example, consider the case when there are two components, as in Fig. 1 (lowest panel) ($n_c = 2$). The two true ages and proportions are (t_1, π_1), (t_2, π_2) and the measured age is a_i. The height of each curve at the measured age gives the relative likelihood of it coming from either t_1 or t_2. The proportion parameters represent the ratio of the areas under each curve in Fig. 1. By combining these quantities we obtain the likelihood function for the unknown true ages and proportions given by the single measured age, a_i:

$$f(a_i,\sigma_i) = \sum_{j=1}^{n_c} \pi_j f_i(t_j) \tag{2}$$

where the dependence of f_i on each unknown age, t_j, is shown explicitly. If all of the measured ages are independent then we can multiply these terms together to obtain the complete likelihood function of the unknown parameters, which we denote as L:

$$L = \prod_{i=1}^{n} f(a_i,\sigma_i) \tag{3}$$

The term 'likelihood function' is used because it gives the likelihood that the true ages and proportions have values (t_j, π_j), ($j = 1, \cdots, n_c$), when the measured ages have values a_i, ($i = 1, \cdots, n$). An important point to note is that the exact values of the true ages and proportions are not recoverable. The quantity L gives only the probability density for any particular set of parameter values. However, using the procedure described below it is possible to find the set of ages and proportions which maximize the likelihood, L, and then estimate their standard errors. In this sense we treat these as the 'best' estimation of the age and proportion parameters based on the available data.

2.2. Solving for the maximum likelihood solution

We wish to find a single set of 'best fit' ages and proportions which maximizes the likelihood function given by Eq. (3). Maximizing L is equivalent to maximizing the logarithm of L, where:

$$\ln L = \sum_{i=1}^{n} \ln f(a_i,\sigma_i) \tag{4}$$

substituting in Eq. (2) this becomes:

$$\ln L = \sum_{i=1}^{n} \ln\left(\sum_{j=1}^{n_c} \pi_j f_{ij} \right) \tag{5}$$

where we have written f_{ij} to represent $f_i(t_j)$. We have to maximize ln L under the constraint given by Eq. (1); that is, the proportions add to unity. This is a standard problem of constrained optimization, which can be solved with the method of Lagrange Multipliers (see, e.g. [4]). Using this method we arrive at a pair of non-linear equations to solve:

$$\pi_j = \frac{1}{n} \sum_{i=1}^{n} \frac{\pi_j f_{ij}}{S_i} \tag{6a}$$

and:

$$\sum_{i=1}^{n} \frac{\pi_j f'_{ij}}{S_i} = 0 \tag{6b}$$

where $S_i = \sum_{k=1}^{n_c} \pi_k f_{ik}$, and $f'_{ij} = (\partial f_{ij})/(\partial t_j)$. For a derivation of these equations see appendix A. So far, we have not specified f_{ij}, which describes the distribution of errors in the age measurements a_i. The simplest case is when the age measurements have Gaussian errors, and we have:

$$f_{ij} = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{ -(a_i - t_j)^2 / 2\sigma_i^2 \right\} \tag{7}$$

In appendix A it is shown that in this case Eq. (6b) becomes:

$$t_j = \frac{\sum_{i=1}^{n} \pi_j a_i f_{ij} / \sigma_i^2 S_i}{\sum_{i=1}^{n} \pi_j f_{ij} / \sigma_i^2 S_i} \tag{8}$$

Eqs. (6a) and (8) can now be used to solve for π_j and t_j for each j, $(j = 1, \cdots, n_c)$ iteratively. An initial trial solution (π_j^0, t_j^0) is found and used to evaluate the right-hand side of (6a) and (8) to give an improved pair of values. This process is then repeated until the changes in the parameters π_j and t_j are smaller than some prescribed value, for example, 0.1% of the previous value. (Often it is simply stopped after a fixed number of iterations.) It is important to note that, in general, there is no guarantee that this iterative scheme will converge. However, in our trials performed on real and synthetic problems these equations have never failed to converge.

In non-linear maximization problems the possibility always exists of multiple maxima in the likelihood function, which means that Eqs. (6) and (8) will be satisfied by more than one set of age and proportion parameters. The iterative process will usually converge to the nearest set of parameters to the starting set (π_j^0, t_j^0), and so the only effective way of locating the true global maximum is to repeat the iterative process from different starting points. In the examples presented below we always repeat the numerical scheme starting from randomly generated ages and proportions (π_j^0, t_j^0) and examine the complete distribution of solutions.

The general equations (6) and those for the special case of Gaussian errors, (8), were derived by Galbraith [5] and used by Galbraith and Green [1] to estimate the component ages in a finite mixture of grains analysed by fission-track dating. If the error distributions closely follow Gaussian statistics, then this is the most efficient and numerically stable method of finding the maximum likelihood set of age and proportions for each age component. If the errors in the measured ages are far from Gaussian then, ideally, one should find the error distribution, f_{ij}, which most closely describes their statistical behaviour. The new expression for f_{ij} should then be substituted in Eqs. (6) and a new iterative scheme found. Often, however, the true error statistics are poorly known, in which case it can be prudent to find the maximum likelihood solution using a range of robust error distributions. By using robust error statistics one ensures that the resulting set of ages and proportions will be more representative of the data as a whole, and less influenced by the occasional age measurement with abnormally large error. The well known Gaussian statistic, Eq. (7), is notoriously non-robust, in that it can be heavily influenced by an outlier age measurement, with the result that the maximum likelihood solution is biased towards the age measurements with largest error.

To extend the maximum likelihood procedure to robust error distributions we consider the generalized Gaussian distribution:

$$f_{ij} = c\,|x| \exp\left\{ -\frac{1}{p} \left| \frac{(a_i - t_j)}{\sigma_i} \right|^p \right\} \tag{9}$$

where $|x|$ represents the absolute value of x; c is a normalizing constant; and p is a number between 1 and 2 which determines the robustness of the distribution. For the case $p = 2$, f_{ij} becomes Gaussian. For $p < 2$ the distribution becomes more robust because its tails become higher (see Fig. 1). If $p = 1$ it is equivalent to the well known double exponential distribution. A set of equations, similar to those for the Gaussian case can be derived for the generalized Gaussian (see appendix A) and used for all values of p between 1 and 2. Alternatively, a maximization procedure, known as Powell's direction set method, may be used to improve numerical stability and efficiency. See [6] for a detailed description of this method. In the next section we present examples using both the direct solution approach and Pow-

ell's method with non-Gaussian error statistics on real data.

2.3. Synthetic examples of the basic algorithm

An example of maximum likelihood estimation

As an example of the estimation procedure consider Fig. 2, which summarizes the results of a test to estimate the maximum likelihood set of ages and proportions using 100 synthetically generated age measurements. The true solution has two component ages and proportions with values $t_1 = 500$ Ma, $\pi_1 = 0.4$, $t_2 = 570$ Ma and $\pi_2 = 0.6$. The synthetic age data was generated by randomly selecting one of two ages, with probabilities 0.4 and 0.6, and adding Gaussian noise with a standard deviation of 30 Ma. The maximum likelihood procedure was then applied to the 100 error-contaminated ages to recover the true ages and proportions. To avoid any dependence on the starting parameters the process was repeated 50 times, in each case starting from a randomly generated set of ages between 450 and 650 Ma and proportions between 0.0 and 1.0.

Fig. 2a shows the starting points, solutions and 95% confidence ellipse for the two ages after 50 iterations of Eqs. (6a) and (8). (The calculation of the confidence ellipse is discussed in section 4.) Fig. 2b is a similar plot for the final pairs of age and proportions values. The solid triangle represents the 'best' single solution of the 50; that is, with the highest value of log-likelihood, ln L, and the outlined triangle is the true solution. In both cases all 50 solutions are virtually identical (as the crosses are all plotted on top of one another), which indicates that the numerical procedure is independent of the starting point and has located the unique maximum likelihood solution. The estimation method has therefore performed well in this simple test. As one would expect, the maximum likelihood solution is not the same as the true solution but lies within the 95% confidence ellipse about the solution. This highlights the importance of placing a confidence region about the estimated solution.

To gain an understanding of the sensitivity of the maximum likelihood procedure to factors such as the number of data and the separation of true

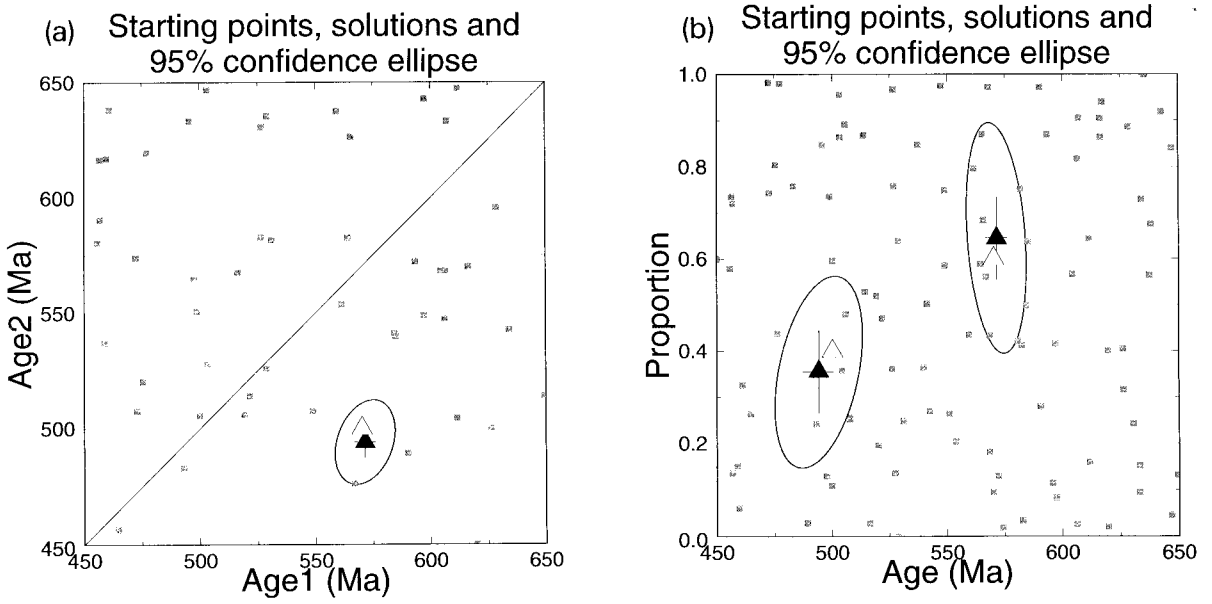

Fig. 2. (a) The starting points (gray squares), final solutions (+) and 95% confidence ellipse about the best of the 50 age/age solutions (▲) for the synthetic data problem. △ = true solution; + = the standard error on the parameter, all crosses are on top of one another. (b) As for (a) but showing both age/proportion parameter pairs. Only one of the proportion parameters is independent and the other is given the same standard error for convenience of display.

ages and proportions, we performed a large number of synthetic tests. Each test consisted of generating simulated age data (from two age components) with Gaussian noise and using the estimation procedure to recover the original ages and proportions. In each test one parameter was varied and the effect on the solution was examined. Overall the number of synthetic ages was varied from 10 to 100, the true age separation varied from 15 to 90 Ma (the Gaussian noise had a standard deviation of 30 Ma) and the ratio of the number of data generated from each component (the proportion parameters) varied from 0.3 to 0.7. For each combination of parameters the procedure was repeated 100 times and the performance to the method analysed. The results clearly showed that the proportion parameters were the most difficult to resolve. A significant trade-off was found between age and proportion parameters, which appears to be non-symmetric in that accurate proportions are dependent on having a good age estimate, while the reverse is not true. It is expected that the proportions parameters will be the most difficult to estimate in practice, and therefore it will be important to estimate the relative errors present in a single application of the maximum likelihood procedure.

2.4. Calculation of covariance matrix and confidence ellipses

Once a set of maximum likelihood ages and proportions are found, the next question one usually asks is: what are the standard errors on these values? This can be answered by finding the covariance matrix for the parameters. The covariance matrix is a symmetric square matrix of size equal to the number of independent parameters; that is, $2n_c - 1$. It is well known that in the case where the error statistics are Gaussian and the unknowns are linearly related to the data then the log-likelihood function, ln L, becomes a quadratic function of the unknowns. The errors in the solution parameters can then be adequately described by the a posteriori covariance matrix of the solution [7,8]. In this case, the inverse of the covariance matrix is directly related to the second derivatives of ln L with respect to the unknowns t_j and π_j, that is:

$$\{C^{-1}\}_{lm} = \frac{\partial^2(\ln L)}{\partial p_l \partial p_m} \tag{10}$$

where p_l and p_m represent the parameters t_j and π_j using indices in the range $l = 1, \cdots, 2n_c - 1$. Therefore, one need only differentiate the log-likelihood expression in (5) to obtain the appropriate formulae. The resulting expressions are given by Galbraith and Green [1]. Commonly, however, analytical expressions for second derivatives lead to numerical instabilities in taking the matrix inverse to obtain the covariance matrix, C. An alternative to these formulae is to calculate the derivatives in (10) using finite-difference methods. When the error distribution is not Gaussian, or when the problem is non-linear (as it is here) then the expression (10) is, itself, only an approximation of the covariance matrix and, in practice, finite-difference derivatives can be much simpler, and often more stable, than using explicit formulae. In this case we simply estimate the second derivatives numerically by evaluating ln L for small perturbations, $\pm\Delta\pi$ and $\pm\Delta t$, in each parameter about the maximum likelihood solution. For convenience, we shall write (π_j^{ML}, t_j^{ML}) for the set of maximum likelihood parameters; LL for the value of ln L evaluated at (π_j^{ML}, t_j^{ML}), and use superscripts $+$, $-$ and 0 to indicate the type of perturbation; that is, LL^{++} represents the value of ln L determined at $(\pi_j^{ML} + \Delta\pi, t_j^{ML} + \Delta t)$ and, similarly, LL^{0-} is evaluated at $(\pi_j^{ML}, t_j^{ML} - \Delta t)$. With this notation the elements of the covariance matrix can be approximated for any type of error statistics using the formulae:

$$\frac{\partial^2 \ln L}{\partial \pi_j \partial \pi_k} = \frac{1}{(\Delta\pi)^2}(LL^{+0} - 2LL + LL^{-0}) \tag{11}$$

$$\frac{\partial^2 \ln L}{\partial \pi_j \partial \pi_k} = \frac{1}{4\Delta\pi\Delta t}(LL^{++} - LL^{+-} - LL^{\mp} + LL^{--}) \tag{12}$$

$$\frac{\partial^2 \ln L}{\partial t_j \partial t_k} = \frac{1}{(\Delta t)^2}(LL^{0+} - 2LL + LL^{0-}) \qquad (13)$$

To use these formulae one need only calculate the value of ln L at the eight points about the maximum likelihood solution for each age group, j.

With the derivatives evaluated using either the analytical or the finite-difference method one can form a square matrix of size ($2n_c$-1) and the covariance matrix, C, is obtained by taking its inverse. The standard errors are the square roots of the corresponding diagonal terms in the covariance matrix:

$$\sigma_{\pi_k} = \sqrt{C_{kk}} \qquad (14)$$

and:

$$\sigma_{t_l} = \sqrt{C_{ll}} \qquad (15)$$

The off-diagonal elements in the covariance matrix give information on how well each parameter is independently resolved. A simple thing to do is to normalize the rows of the covariance matrix by the value of each diagonal element; that is, replace C_{km} with C_{km}/C_{kk}. If the magnitude of the off-diagonal elements are close to 1 (or -1) then the corresponding parameters are highly correlated (anti-correlated) and cannot be independently resolved by the data. Ideally, one would have a covariance matrix with its largest values down the diagonal. Examining off-diagonal elements can be useful in detecting whether two age particular age components are distinctly resolved in the data.

Another use for the covariance matrix is in determining confidence ellipses for any pair of parameters about the maximum likelihood solution. The confidence ellipsoid has as many dimensions as there are ages and proportions. It and represents all possible combinations of ages and proportions surrounding the maximum likelihood point within which one can have 95% confidence that the true solution exists. This ellipsoid can be found by determining the region about the maximum likelihood solution over which the integral of L is equal to 0.95. This type of calculation is also dependent on the assumptions regarding the nature of the error statistics present. For the ideal case of Gaussian errors and a linear problem the equation of the confidence ellipsoid is given by:

$$\Delta = \delta \mathbf{p}^T C^{-1} \delta \mathbf{p} \qquad (16)$$

where the vector $\delta \mathbf{p}$ is used to represent the change in the age and proportion parameters from the maximum likelihood solution (π_j^{ML}, t_j^{ML}); and Δ is the change in value of the L from the maximum likelihood point to the edge of the ellipsoid. The shape and orientation of the confidence ellipsoid is determined by the covariance matrix, C, while Δ controls its size, that is, which percentage it represents. Once the covariance matrix is determined, the only thing left to do is find the value of Δ corresponding to 95%. Again, for the Gaussian case standard statistical tables may be used. Once Δ is determined the confidence ellipsoid can be projected onto any chosen pair of parameters to produce a confidence ellipse (details of this procedure appear in appendix B). An example of the confidence ellipses which result can be seen in Fig. 2 for the synthetic data problem discussed in section 3. When Gaussian statistics are not appropriate the same procedure is often used to obtain an approximation of the confidence ellipse. In general, the only exact procedure is to perform the multi-dimensional integration numerically, which can be computationally very expensive (see [6]).

The covariance matrix and confidence ellipse are useful because they provide information on the likely errors in the final age and proportion parameters. In addition, they can be used to help estimate the number of distinct age components resolved by the data, which is discussed further in the next section. One must be careful, however, in placing too strong an emphasis on these error estimates, because they are often sensitive to the underlying statistical assumptions on the measurement errors. The effect of changing the likelihood function to something more robust (e.g., the generalized Gaussian, Eq. (9), with $p = 1.25$) is often to move the maximum likelihood solution by more than the original standard errors would allow. Therefore, in cases where the error statistics are not well known it is prudent to find the maximum likelihood solution and corresponding

standard errors for different types of error statistic, and examine the spread in solutions as well.

2.5. Estimating the number of age components

Once a mechanism is available to estimate 'best fit' ages and proportions, the next objective is to determine the number of distinct components resolvable by the data, n_c. There is always the possibility that real components will remain undetectable with a limited amount of data, and therefore we should always attempt to estimate the minimum number of components necessary to explain the distribution of age measurements. In this way the temptation to over-interpret the data is limited. Here we describe a series of simple tests that can be applied, and illustrate these with both synthetic and real data examples. Again, the general philosophy is not to rely heavily on any single method but rather to build up information on the likely number of components using a series of complimentary approaches.

The first procedure is to determine the confidence level at which the ellipse about two ages does not intersect the equal age line. When this condition is satisfied the two ages are distinct at the corresponding confidence level. For example, in Fig. 2a the 95% confidence ellipse does not intersect the diagonal equal age line and so the two estimated ages are distinct at the 95% confidence level. If the equal age line is intersected by the ellipse then the confidence level should be reduced until the ellipse shrinks to a point where it just touches the equal age line. The reduced confidence value gives the new level of confidence in the ages being distinct. Again, this process is dependent on the assumptions made about the type of the error statistics (i.e., whether they are Gaussian or not) and the accuracy of the confidence ellipse but it can be a useful indicator of when two or more ages are present in the data.

Another approach is to repeat the maximum likelihood estimation with different numbers of assumed age components. Fig. 3 shows the results

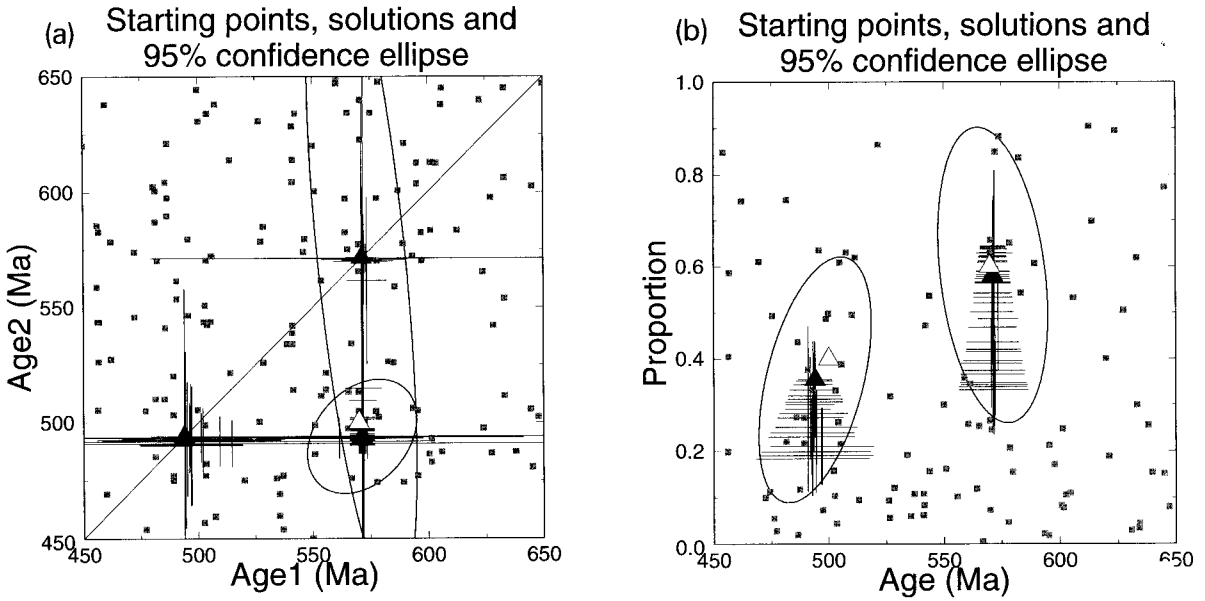

Fig. 3. (a) Details as for Fig. 2a but with three age components assumed (only two exist in the synthetic data). Two 95% confidence ellipses are plotted. For one pair the ages are distinct and the ellipse does not intersect the equal age line, for the other the ellipse is centred on the line and the ages are not distinct. Squares = randomly generated starting points; + = age/age and age/proportion solutions; △ = true solution; ▲ = the two calculated ages. (b) Same as in (a) but showing the two (of the three) most significant age/proportion solutions. Note the error in proportions can be large.

of assuming three age components in the synthetic data set, when only two actually exist. The maximum likelihood estimation was performed from 50 randomly generated starting points and all pairs of the age/age and age/proportion solutions are plotted. One can see that the standard errors are larger and the scatter in the solutions is much greater than when assuming the correct number of components (compare Fig. 2). In Fig. 3a, two 95% confidence ellipses are shown and one of these is centred on the equal age line. This strongly indicates that the two ages are not distinguishable. In addition, the final solutions clearly fall into three groups, two of which lie on the equal age line, which means that in nearly all of the 50 repeats two of the three age parameters have converged to the same value. Interestingly, the corresponding proportion parameters are not equal; usually one is much greater than the other. The best of the solutions, labelled S3 in Table 1, is a typical example. Again, the true (two-component) solution is given by the outlined triangle and this matches very well with the two distinct ages.

Another simple test is to examine the value of the negative log-likelihood of the best solution as a function of the number of assumed age components, n_c. This has been done in Fig. 4, the misfit of the best solution (i.e., value of $-\ln L$) as a percentage of the one component case, is plotted against n_c. The 'best misfit' curve, represented by the black squares, shows how the relative improvement in data fit trades off with the number of components. As the number of assumed components increases there are more parameters with which to fit the data and so we would expect the misfit of the best solution to decrease. The 'elbow' of the curve is usually a good indicator of the number of distinct age components. In Fig. 4 there is virtually no further reduction in misfit when n_c is greater than 2, which again indicates that only two distinct age components exist in the synthetic data set.

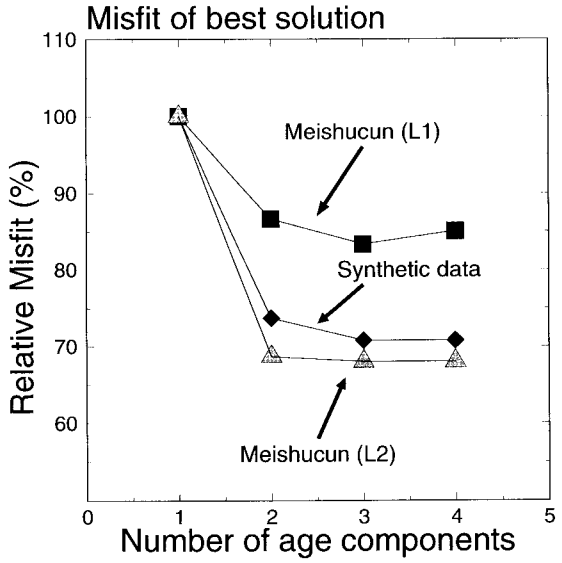

Fig. 4. Misfit of the best of 50 solutions against number of assumed age components for synthetic and Meishucun Bed 5 data sets. L1 = double-exponential statistics; L2 = Gaussian. The 'elbow' in the curve indicates that no significant improvement in fit is obtained by increasing the number of distinct components.

Table 1
Summary of results for synthetic data set

	Best solution from 50 trials (100 synthetic data)						
Experiment	*p*	Component	Age	Stn Err	Proportion	Stn Err	Misfit
S1	2.0	1	543.8	3.0	1.0		123.23
S2	2.0	1	494.0	6.5	0.36	/	90.08
		2	571.3	4.6	0.64	0.1	
S3	2.0	1	571.3	44.7	0.06		90.08
		2	494.0	7.4	0.36	0.1	
		3	571.3	7.1	0.58	0.1	
S4	2.0	1	571.3	30.1	0.12		90.08
		2	571.4	23.7	0.18	-	
		3	571.3	15.0	0.34	-	
		4	494.0	7.4	0.36	0.1	
True Sol	2.0	1	500.0		0.4		
		2	570.0		0.6		

Another test is to examine how well the final set of ages and proportions fit the histogram of the original data. In Fig. 5 we have assigned each measured age to one of the two age components found for the $n_c = 2$ case. This is done by calculating the probability density $\pi_j f_i$ for each measured age, a_i and age component, t_j, and then assigning age a_i to the component with the higher value. The two panels in Fig. 5 show the histogram of the classified ages, together with the distributions corresponding to the solution values. The total area under the curves are scaled by the number of data but the relative areas and positions of the distributions are equal to the solution proportions and ages respectively (see S2, Table 1). This classification procedure at-

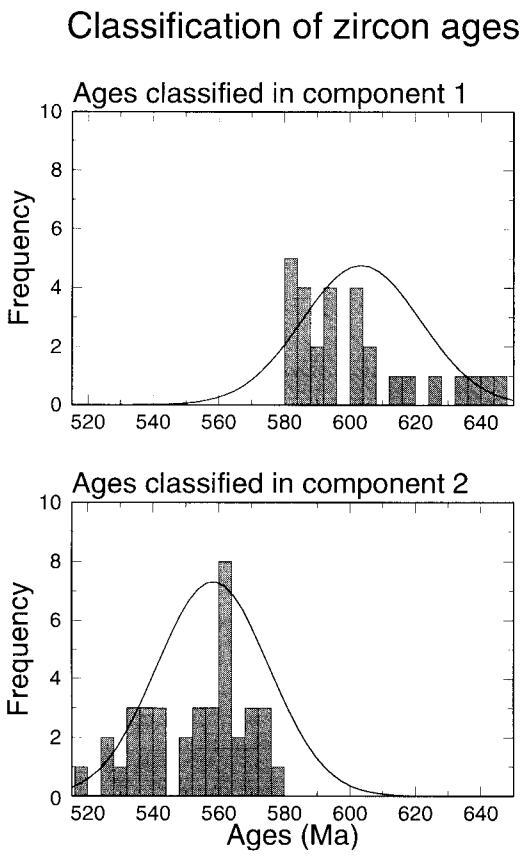

Fig. 5. Histograms of the raw (upper panel) and synthetic (lower panel) data classified into two groups according to the final solutions (shown as Gaussian distributions). The combined histogram and two-component distribution is shown in Fig. 1. The distribution of ages assigned to each component can be a good indicator of the relative significance of the individual components.

tempts to identify to which of the two age components each original datum belongs. The simple method employed here is subjective and can be incorrect for some of the ages in the parts where the two curves overlap. Nevertheless, the combined distribution (shown in Fig. 1) seems to provide a reasonable fit to the combined data histogram. The true ages are indicated by arrows and these are very close to the peaks in the combined distribution.

A more formal procedure that could be used here is the Kolmogorov–Smirnov (K–S) test (see chapter 13 of [6]), which directly measures the discrepancy between the original data histogram and the estimated distribution defined by the final ages and proportions (both shown in Fig. 1). The K–S test gives the probability that this discrepancy could have occurred by random chance and, therefore, once the value increases to an acceptable level (say 0.5) there is no support for adding further components. Our tests have shown that plotting the K–S probability as a function of the number of distinct age components is virtually identical to the misfit against n_c plot described above and so no extra information is gained. A similar situation arises with the 'binned Chi-square' test [6] which performs a similar task to the K–S test. Overall, the combination of confidence ellipse, multiple component tests and histogram fit seems to provide a useful series of tests to identify the number of distinct components in the synthetic data set.

3. Application to a complex zircon data set

The first example we have chosen to illustrate the mixture modeling methods described in this paper are the SHRIMP zircon ages published by Compston [2] for the Meishucun Bed 5 horizon in South China. These data are important geologically for their bearing on the numerical calibration of the Early Cambrian. They are important also as a challenge, as the data were recognized from the outset as a complex age population that was subject both to open-system behaviour, which produced variably young ages, and to detrital zircon inheritance. It is essential to be able to extract all the numerical information present in the data set from this particular sample, as it is an example of a volcanic layer whose stratigraphic relationship to the faunal assemblages within the Yangtze Platform is unambiguous. It should not be abandoned simply because the zircon ages show complications.

The mixture modeling methods described here were applied to the SHRIMP zircon ^{238}U–^{206}Pb ages shown in table 3 of Compston [2]. In the original analysis of this data set the lowest ages were successively deleted because they were judged as outliers due to variable Pb loss. This

produced a total of 41 analyses in the range 450–560 Ma and, assuming a single population, a mean age of 525 ± 7 Ma (2σ) was found. An unsatisfactory aspect of this procedure was that the standard deviation observed for the remaining age measurements was 2.7%, still suspiciously greater than the expected error, although not provably so. In the present study two subsets of the data were analysed. The first consisted of all 46 ages in the range 450–560 Ma and the second of all 58 ages between 350 and 600 Ma. The first set of 46 analyses had a lower mean of 521 ± 5 Ma (2σ) due to some lower ages being included. The larger set contains many analysis with lower ages which have variable U gain or Pb loss but are included in order to test the influence of outliers on the results.

The maximum likelihood procedure was performed using Gaussian error statistics, $p = 2$, and also with a more robust double-exponential $p = 1$ distribution (see Eq. 9). It is thought that the measurement errors are closer to Gaussian, but performing the procedure with the two most extreme distributions allows us to assess the influence of the statistical assumptions on the results. For each statistical distribution the estimation procedure was repeated assuming up to four distinct age components. In each case the procedure was repeated from 50 random starting values and from each starting value a solution was obtained after 50 iterations of Eqs. (6a) and (8) for the Gaussian case, and with Powell's technique for the $p = 1.0$ distribution. The solution with the lowest misfit (of the 50) for each value of n_c and p is summarized in Table 2 for the 46 sample data set.

A plot of the relative misfit reduction from Table 2 against the number of age components appears in Fig. 4. For the two-component Gaussian case, the starting points, solutions and confidence ellipses about each age/age and age/proportion pair are shown in Fig. 6a and b, respectively. The results for the two-component case are similar to the synthetic data in that all 50 solutions have converged to the same values and are indistinguishable from one another in Fig. 6. The confidence ellipse about the two ages is small and is far from the equal age line, indicating a high probability that the two age components are distinct. Even though the two statistical

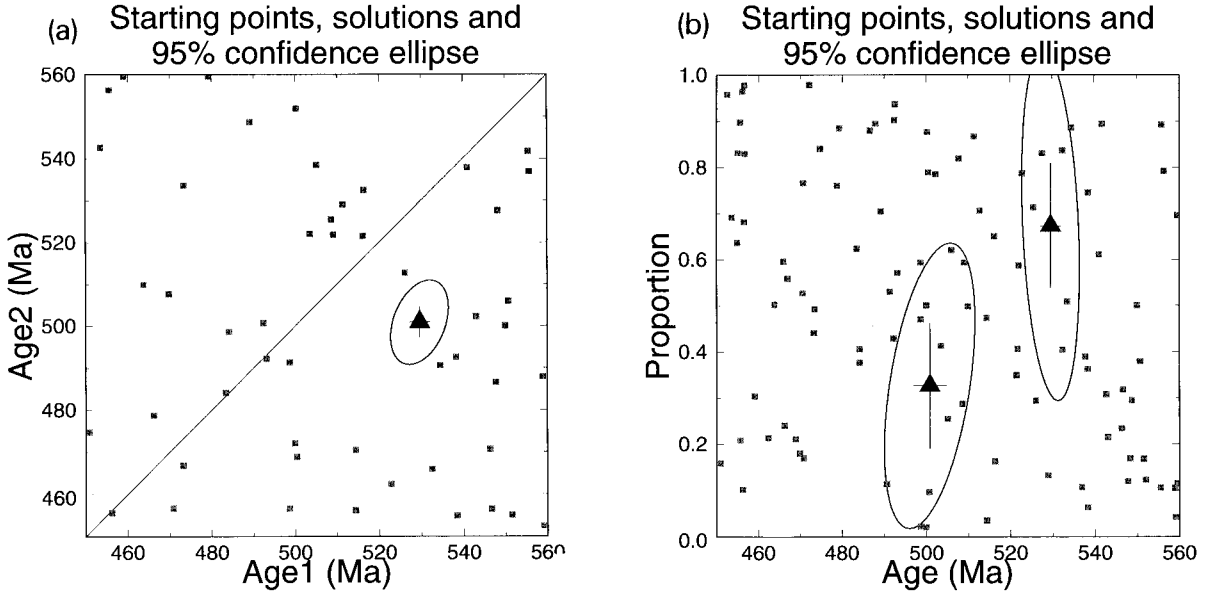

Fig. 6. (a) Results for age–age pairs with the Meishucun Bed 5 data set. Two age components have been assumed and symbols are the same as in Fig. 2. (b) Same as for (a) but showing both age proportion pairs.

Table 2
Summary of results for the Meishucun zircons

Best solution from 50 trials with 46 ages)							
Experiment	*p*	Component	Age	Stn Err	Proportion	Stn Err	Misfit
MB1	2.0	1	519.5	1.6	1.0		63.81
MB2	2.0	1	500.8	3.6	0.33		43.86
		2	529.6	2.5	0.67	0.1	
MB3	2.0	1	497.3	4.1	0.23		43.41
		2	516.4	10.2	0.26	0.3	
		3	532.4	4.9	0.51	0.3	
MB4	2.0	1	532.4	28.9	0.05		43.41
		2	497.3	5.5	0.23	0.14	
		3	516.3	12.3	0.26	0.27	
		4	532.3	5.9	0.46	0.26	
MB5	1.0	1	521.0	2.6	1.0		62.86
MB6	1.0	1	499.0	2.0	0.40		54.30
		2	529.0	1.7	0.60	0.1	
MB7	1.0	1	497.0	1.9	0.20		52.40
		2	517.0	2.2	0.30	0.2	
		3	535.0	3.4	0.50	0.2	
MB8	1.0	1	529.0	2.9	0.18		53.46
		2	538.0	2.9	0.25	0.1	
		3	497.0	2.0	0.26	0.1	
		4	518.0	2.3	0.35	0.2	
MB58	1.0	1	373.0	1.9	0.03		
		2	417.0	1.8	0.11	0.1	
		3	497.0	1.8	0.27	0.1	
		4	529.0	1.7	0.59	0.1	

MB58 represents results with 58 ages.

distributions differ in character they yield virtually identical two-component solutions: ages of 500 ± 6 Ma (2σ) and 530 ± 5 Ma (2σ) for the Gaussian case, and 499 ± 4 Ma (2σ) and 529 ± 3 Ma (2σ) in the exponential case.

When three components are assumed the ages from the two-component case are again recovered, together with a third component which appears to lie at the mean of these ages. The same feature occurs with both statistics and, since the third component produces only an extra 0.7% (Gaussian) to 3% (exponential) reduction in the misfit, it does not seem justified by the data. When four components are assumed there is no improvement in misfit reduction. (The misfit actually increases in the exponential case.) Again, the solutions show that the four-component solution is virtually identical to the three-component solution. In the Gaussian case one component (at 532 Ma) has been picked out twice. There is therefore no evidence for a fourth component in this data set.

All of these calculations were repeated for the larger data set of 58 analyses in the range 350–600 Ma. The primary purpose of this is to test whether the two components resolved by the smaller data set were recoverable when the variable Pb loss zircons were included in the procedure. Interestingly, the results showed that, as the number of assumed components was increased to four the two components around 500 and 530 Ma were indeed recovered in both the Gaussian and exponential cases. Table 2 contains a summary of these results. Reassuringly, these two components are the only two which appear with both statistics. It appears that the effect of adding the extra (lower) zircon ages to the procedure is to introduce a large number of outliers which, as one would expect, are poorly modeled by either statistical distribution. The mixture modeling process therefore has a lot more difficulty in picking out distinct components and can only recover the two most prominent components when enough degrees of freedom are introduced in the four-component case.

The most important geological conclusion from these results is the presence of a group of zircons within the Meishucun Bed 5 bentonite at 530 ± 5 Ma (2σ), slightly older and better defined than our previous result of 525 ± 7 Ma [2] based on a less objective statistical procedure. Interpreting the group to be precipitates from the magma that produced the bentonite, we infer that 530 ± 5 Ma is also the time of deposition for Bed 5 in the Meishucun section, which therefore provides a time point within the Tommotian. The revised age estimate is now numerically greater than the Cooper [9] result of 526 ± 4 Ma for the Atdabanian–Botomian boundary and it is consistent with the recent result by mass-spectrometric isotope dilution of 534.6 ± 0.7 Ma for the earliest Tommotian of the Kharaulakh section, Siberia [10].

The geological significance of the younger group of zircons at 501 ± 7 Ma is ambiguous. The group might designate a genuine, early geochemical event, during which parts of zircons lost most of their radiogenic Pb, or it might signify that many areas within the grains have gained U and/or lost Pb by a few percent in the recent past. It is quite clear that other areas within grains having apparent ages less than 450 Ma have lost very large fractions of their radiogenic Pb at times later than a possible 500 Ma event. The question of a 500 Ma group will be addressed fully during later assessment of unpub-

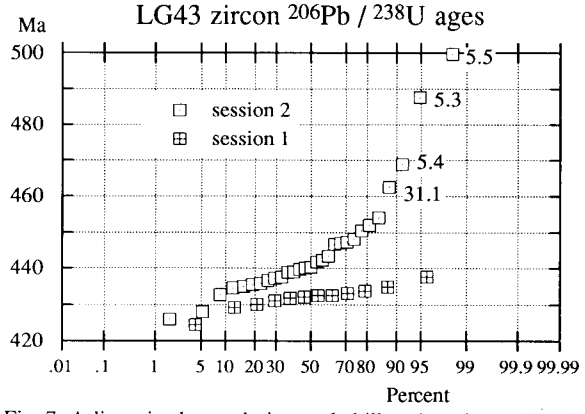

Fig. 7. A linearized cumulative probability plot of ages found in two separate sessions. The significance of this type of plot for detecting multiple components in geochemical data and its qualitative use for identifying components has been described by Sinclair [11].

lished data on zircons extracted from more altered exposures of the Bed 5 bentonite.

4. An application to a 'simple' zircon data set

Unpublished SHRIMP ^{238}U–^{206}Pb ages for zircons from an undeformed tonalite in northeast China are presented in Fig. 7. The tonalite has no useful prior age constraints. The zircons are long (> 5:1 length/breadth), euhedral, unaltered, unzoned, free of obvious zircon cores and collectively give the appearance of a single magmatic population. Single analyses on twelve grains using the SHRIMP II ion probe (analyst Li Xian-hua) produced an extremely well-grouped set of ages (session 1, Fig. 7). The standard deviation per age expected from counting errors was 4 ± 1 Ma (σ), as compared with 3.1 Ma as observed from the twelve. The mixture modeling program finds only a single age for this data set, as would be expected.

Could such a uniform sequence of SHRIMP ages be produced by a chance sampling effect? Suppose that 5% of the area within the LG43 zircons exposed to SHRIMP analysis is actually composed of older zircon cores, older whole grains and areas within grains that have lost Pb or gained U after crystallization. The chance that a randomly selected sequence of twelve spot analyses will find only the undisturbed magmatic grains is $(0.95)^{12}$, which equals 0.54 and is quite a high probability. Thus, the low dispersion found in the first analytical session does not signify that the LG43 sample is necessarily composed solely of undisturbed magmatic zircons.

Some of the original twelve grains and a number of new grains were re-analysed in a second session to assess the LG43 sample more thoroughly (Fig. 7, session 2). It is obvious that there is no single Gaussian population, and that analyses 5.5, 5.3, 5.4 and 31.1 signify the presence of older zircon. The results for grain 5 can be attributed to a 'hidden' zircon core, for which there are many precedents elsewhere. The high ages within grains 5 and 31 show that the differences are target-related, rather than due to uncontrolled instrumental effects. It is also plain that the first twelve analyses of LG43 give a misleading impression of the uniformity in zircon age. With one exception, the between-session replicate ages were slightly older than before.

If the ages from the two sessions are combined, but with the four prominently older ages deleted, the remainder define a straight line (Fig. 8) indicative of a single Gaussian population having a mean age of 437.7 ± 1.2 Ma (σ). However, this inference has two problems: (1) the age is distinctly older than the 432.0 ± 0.9 Ma found

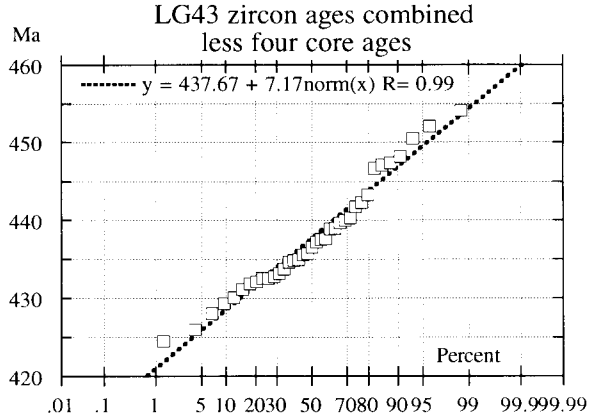

Fig. 8. A linearized cumulative probability plot for the combined ages shown in Fig. 7 with the four prominently older ages removed. The best-fit straight line is also shown.

originally, and (2) the standard deviation indicated per analysis (7.2 Ma) is well in excess of the expected standard deviation based on counting statistics plus other error sources.

Mixture modeling of the combined ages without any deletions finds a principal component at 435.2 ± 0.7 Ma (69%), another at 448.4 ± 1.2 Ma (21%), and two minor components at 489 ± 3 Ma and 466 ± 2 Ma. The two older components arise from the high 'outliers' in Fig. 7. It is likely that the multiple ages within grain 5 are not single components but due to an array of discordant ages within a core, or to variable overlaps of the areas analysed between the core and magmatic zircon. The separation into two younger components removes the excessive standard deviation of 7.2 Ma required by the single age assumption. In addition, 435.2 ± 0.7 Ma is tolerably close to the original 432.0 ± 0.9 Ma, although it remains detectably older. (It is possible that this is due to a small bias between the two analytical sessions caused by error in the calibration lines for the reference zircon). The program fails to find a stable solution for five components.

The above results lead to the following interpretation and conclusions:

(1) The best estimate of the true age for magmatic crystallization is the principal component at 435.2 ± 0.7 Ma (σ). The initial impression of extreme uniformity in age was a sampling effect.

(2) There is an apparent age-component older than this at 448 Ma. However, inspection of the within-grain replicate ages does not conclusively indicate the presence of a single population of earlier-crystallized zircon. Like the variable dispersion in age within grain 5, it could be due instead to overlap with small patches of inherited zircon older than 448 Ma within the magmatic grains.

(3) The precise definition of the principal age group (± 1.5 Ma for 2σ) shows that SHRIMP is fully capable of providing useful time-scale results, given a uniform target. The dispersion in age within altered zircons, such as those from Meishucun, can create the erroneous impression that the technique is faulty rather than the zircons.

(4) The indication from the zircons from tonalite LG43, as well as those from many other rocks studied using SHRIMP, is that more grains contain 'hidden' cores or slightly older zircon populations than geochronologists have hoped and expected. At least four of the 23 'magmatic' grains examined here contain hidden cores and 7 of the 41 analysed areas detect inheritance. For such zircons, analysis of a whole single grain has a significant probability of giving an age which is too old.

5. Concluding remarks

The combination of methods presented has produced a flexible approach for zircon (or other) ages. Several methods have been suggested for detecting the most likely number of distinct age components. None of these is infallible but it is expected that their combination will provide a useful indicator of the minimum number of components needed to satisfy the data.

We feel the main practical advantage of our methodology is that it allows a more complete analysis of the data set. Previously, in cases where two or more components were suspected of existing in the data, little could be done to estimate the ages of either, unless they were well separated and a large number of data were available. Our methodology is likely to prove most useful in the more difficult case where two populations have nearly equal ages and are of comparable size. If only one age population is assumed, when more exist, then simply taking an average will inevitably give a poor estimate of all true ages. The new approach allows the most likely set of ages to be determined for any given number of components.

Several extensions of the current approach are possible. Probably the most important would be to replace the calculation of the approximate covariance matrix discussed in section 2.4 with an exact procedure based on multi-dimensional integration using Monte Carlo methods [5]. This would be very computationally expensive at present; however, it would improve estimates of errors, which may be crucial in cases where age

components become difficult to resolve or when the error statistics are far from Gaussian.

Our methodology has performed well in synthetic problems, where we can compare performance with known true ages, and with real data that has previously been interpreted with traditional methods. The application to both real data sets has yielded more complex situations than in the synthetic problem and this is expected to be typical of most real data applications where error statistics are not well understood. However, the combination of methods has yielded a clear picture of the most likely number of distinct components, their ages and proportions. We expect the approach will prove a potentially powerful tool for analysing multi-component data sets of this kind.

Acknowledgments

The authors thank K. Gallagher and I. Wendt for their suggestions and constructive criticisms on an earlier version of this manuscript.[CL]

Appendix A: derivation of the maximum likelihood equations

The method of Lagrange multipliers may be used to find the set of ages and proportions which maximizes the log of the likelihood function, ln L, given by Eq. (5), under the constraint that the sum of the proportions is unity (Eq. 1). This constrained maximization problem can be solved by the unconstrained maximization of a new function, M:

$$M = \ln L + \lambda \left(\sum_{j=1}^{n_c} \pi_j - 1 \right) \tag{A.1}$$

where λ is a Lagrange multiplier that should be treated as an extra unknown. The stationary points of M are found by differentiating with respect to each unknown and setting the result to zero, which gives:

$$\frac{\partial M}{\partial \pi_j} = \frac{\partial \ln L}{\partial \pi j} + \lambda = 0 \tag{A.2}$$

$$\frac{\partial M}{\partial t_j} = \frac{\partial \ln L}{\partial t_j} = 0 \tag{A.3}$$

$$\frac{\partial M}{\partial \lambda} = \sum_{j=1}^{n_c} \pi_j - 1 = 0 \tag{A.4}$$

Differentiating ln L from Eq. (5) we get:

$$\frac{\partial \ln L}{\partial \pi_j} = \sum_{i=1}^{n} \frac{f_{ij}}{S_i} \tag{A.5}$$

and:

$$\frac{\partial \ln L}{\partial t_j} = \sum_{i=1}^{n} \frac{\pi_j f'_{ij}}{S_i} \tag{A.6}$$

where $S_i = \Sigma_{k=1}^{n_c} f_{ik}$, and $f'_{ij} = \dfrac{\partial f_{ij}}{\partial t_j}$. Substituting Eq. (A.5) into Eq. (A.2) and multiplying by π_j we get:

$$\sum_{i=1}^{n} \frac{\pi_j f_{ij}}{S_i} + \lambda \pi_j = 0 \tag{A.7}$$

By summing this equation over the j (i.e., the n_c age groups) and using the constraint Eq. (A.4) we get:

$$\lambda = - \sum_{j=1}^{n_c} \sum_{i=1}^{n} \frac{\pi_j f_{ij}}{S_i} \tag{A.8}$$

We can swap the order of the double summation to give:

$$\lambda = - \sum_{i=1}^{n} \sum_{j=1}^{n_c} \frac{\pi_j f_{ij}}{S_i} = - \sum_{i=1}^{n} \frac{S_i}{S_i} = - \sum_{i=1}^{n} 1 = -n \tag{A.9}$$

Having found the value of λ we can substitute it in Eq. (A.7) and solve for πj, which gives:

$$\pi j = \frac{1}{n} \sum_{i=1}^{n} \frac{\pi_j f_{ij}}{S_i} \tag{A.10}$$

The second equation is found by substituting Eq. (A.6) in Eq. (A.3), which gives:

$$\sum_{i=1}^{n} \frac{\pi_j f'_{ij}}{S_i} = 0 \tag{A.11}$$

Eqs. (A.10) and (A.11) form a pair of non-linear equations for each j that can be used to solve for the parameters (π_j, t_j) $(j = 1, \cdots, n)$.

For the special case of Gaussian error statistics the p.d.f. of the jth pair of parameters (π_j, t_j) due to ith measured age a_i is given by (7). Eq. (A.11) involves the first derivative of the Gaussian distribution with respect to t_j:

$$f'_{ij} = \frac{1}{\sigma_i^2}(a_i - t_j) f_{ij} \tag{A.12}$$

The second derivative is required by analytical formulae of the covariance matrix:

$$f''_{ij} = \frac{1}{\sigma_i^2}(a_i - t_j) f'_{ij} - \frac{f_{ij}}{\sigma_i^2} \tag{A.13}$$

In the case where the measurements are not independent these expressions become slightly more complicated (see [8]). Eq. (A.10) is already in a form suitable for numerical solution. To simplify the equations in the Gaussian case we can substitute these expressions in Eq. (A.11) and after simplification we obtain:

$$t_j = \frac{\Sigma_{i=1}^{n} \pi_j a_i f_{ij} / \sigma_i^2 S_i}{\Sigma_{i=1}^{n} \pi_j f_{ij} / \sigma_i^2 S_i} \tag{A.14}$$

Eqs. (A.10) and (A.14) form the basis of an iterative solution for the parameters π_j and t_j for the Gaussian distribution (Eq. 7)

For more robust non-Gaussian error distributions (A.10) is unchanged but the second equation (A.11) is more complicated and, for some forms of error distribution, it may not lead to an equation that can be solved numerically. Fortunately for a generalized Gaussian distribution (Eq. 9) the problem is tractable and the new expression to replace (A.14) is:

$$t_j = \frac{\Sigma_{i=1}^{n} \dfrac{\pi_j a_i f_{ij}}{\sigma_i^2 S_i} \left| \dfrac{(a_i - t_j)}{\sigma_i} \right|^{p-2}}{\Sigma_{i=1}^{n} \dfrac{\pi_j f_{ij}}{\sigma_i^2 S_i} \left| \dfrac{(a_i - t_j)}{\sigma_i} \right|^{p-2}} \tag{A.15}$$

Eqs. (A.10) and (A.15) are the new pair of equations to be solved for the generalized Gaussian distribution (for any value of p). Again, if $p = 2$ then this equation reduces to the Gaussian case and is exactly Eq. (A.14).

Appendix B: plotting the confidence ellipse

The procedure used to calculate the confidence ellipse about any pair of parameters is described in detail in [6]. To calculate the 95% confidence ellipse about two age parameters (as in Fig. 2) an appropriate value of Δ in Eq. (16) must be found. A 95% confidence region is, by definition, the region about the maximum likelihood solution for which there is a 95% probability that it contains the true solution. Therefore the integral of the probability density function (L in Eq. 3) within this region should be equal to 0.95. If the error statistics have a Gaussian distribution then this integral can be evaluated analytically, and a chi-squared table can be used to find the value of Δ which corresponds to any number of parameters, ν, and confidence probability, Pr. Chi-squared tables appear in books of standard statistical tables. However, in practice it is convenient to use a numerical root-finding routine, instead of a table, to evaluate Δ for given ν and Pr. An example can be found in Press ([6] chap. 14, p. 537). If the error statistics are not Gaussian then the value of Δ cannot be found from tables and the expensive numerical integration referred to in section 2 must be performed. Usually, however, the above procedure may still be used to get an approximate value for Δ but the confidence ellipse must then be treated with caution.

When Δ is known the $(2n_c - 1)$-dimensional ellipsoid must be projected onto the two-parameter axes of interest. This is done by taking the full covariance matrix, C, and copying the intersection of ν rows and columns corresponding to the parameters of interest into a $\nu \times \nu$ matrix denoted C_{proj}. The smaller matrix is then inverted to give the required equation of the projected confidence ellipse:

$$\Delta = \delta \mathbf{t}^T C_{\text{proj}}^{-1} \delta \mathbf{t} \tag{B.1}$$

where the vector $\delta \mathbf{t} = (t_1 - t_1^{\text{ML}}, t_2 - t_2^{\text{ML}})^{\text{T}}$ represents the change in the two ages from the maximum likelihood point $(t_1^{\text{ML}}, t_2^{\text{ML}})$. The same

procedure can be applied for any number of parameters, or any other pair of parameters.

Usually $\nu = 2$ and we wish to plot the ellipse given by (B.1) about the maximum likelihood point $(t_1^{\mathrm{ML}}, t_2^{\mathrm{ML}})$. This is done by converting to polar co-ordinates (r, θ), and finding the radius, r, for given values of θ about the maximum likelihood solution. We can write the relationship as:

$$t_1 = t_1^{\mathrm{ML}} + r\cos\theta; t_2 = t_2^{ML} + r\sin\theta \qquad \text{(B.2)}$$

If we write the elements of the 2×2 matrix C_{proj}^{-1} as

$$C_{\mathrm{proj}}^{-1} = \begin{pmatrix} C'_{11} & C'_{12} \\ C'_{21} & C'_{22} \end{pmatrix} \qquad \text{(B.3)}$$

then by substituting (B.3) and (B.2) into (B.1) and re-arranging we obtain:

$$r = \left([\Delta] / \left[\left(c'_{11}\cos^2\theta + (C'_{12} + C'_{21})\cos\theta\sin\theta + C'_{22}\sin^2\theta \right) \right] \right)^{1/2} \qquad \text{(B.4)}$$

This expression is used to find r for regular values of θ between 0 and 2π. The corresponding values of t_1 and t_2 are found from (B.1) and (B.2) and hence the ellipse can be plotted. In the case where the two parameters have very different sizes (e.g. if one is an age parameter and the other a proportion parameter, as in Fig. 2b) regular steps in θ around the ellipse will not produce an evenly distributed set of points on the plot. This is because θ depends on the relative scales of the two axes. To ensure a regular distribution of points (and therefore a well-defined ellipse) θ should be transformed to a second variable, θ', using:

$$\tan\theta' = \frac{\tan\theta}{s} \qquad \text{(B.5)}$$

where s represents the relative size of the two parameter axes. The procedure would then be to choose regular values of θ between 0 and 2π, find the corresponding value of θ' from (B.5) and use this instead of θ in (B.4). For Fig. 2b the scale is 200:1 and so we should set $s = 200$. An alternative value is to set s approximately equal to the ratio of the standard errors of the two parameters.

References

[1] R.F. Galbraith and P.F. Green, Estimating the component ages in a finite mixture, Nucl. Tracks Radiat. Meas. 17, 197–206, 1990.

[2] W. Compston, I.S. Williams, J.L. Kirschvink, Zhang Zichao and Ma Guogan, Zircon U–Pb ages for the Early Cambrian time-scale, J. Geol. Soc. London 149, 171–184, 1992.

[3] G.J. McLachlan and K.E. Basford, Mixture Models: Inference and Applications to Clustering, Marcel Dekker, New York, 1987.

[4] G. Arfken, Mathematical Methods for Physicists, Academic Press, San Diego, 3rd ed., 1985.

[5] R.F. Galbraith, Graphical display of estimates having differing standard errors, Technometrics 30, 271–281, 1988.

[6] W.H. Press, B.P. Flannery, A.T. Saul and W.T. Vetterling, Numerical Recipes, 702 pp., Camb. Univ. Press., Cambridge, 1987.

[7] W. Menke, Geophysical Data Analysis: Discrete Inverse Theory, Academic Press, San Diego, revised ed., 1989.

[8] A. Tarantola, Inverse Problem Theory, 613 pp., Elsevier, Amsterdam, 1987.

[9] J.A. Cooper, R.J.F. Jenkins, W. Compston and I.S. Williams, Ion-probe zircon dating of a mid-Early Cambrian tuff in South Australia, J. Geol. Soc. London 149, 185–192, 1992.

[10] S.A. Bowring, J.P. Grotzinger, C.E. Isachsen, A.H. Knoll, S.M. Pelechaty and P. Kolosov, Calibrating rates of Early Cambrian evolution, Science 261 (5126), 1293, 1993.

[11] A.J. Sinclair, Selection of threshold values in geochemical data using probability graphs, J. Geochem. Explor. 3, 129–149, 1974.

Classification of zircon ages

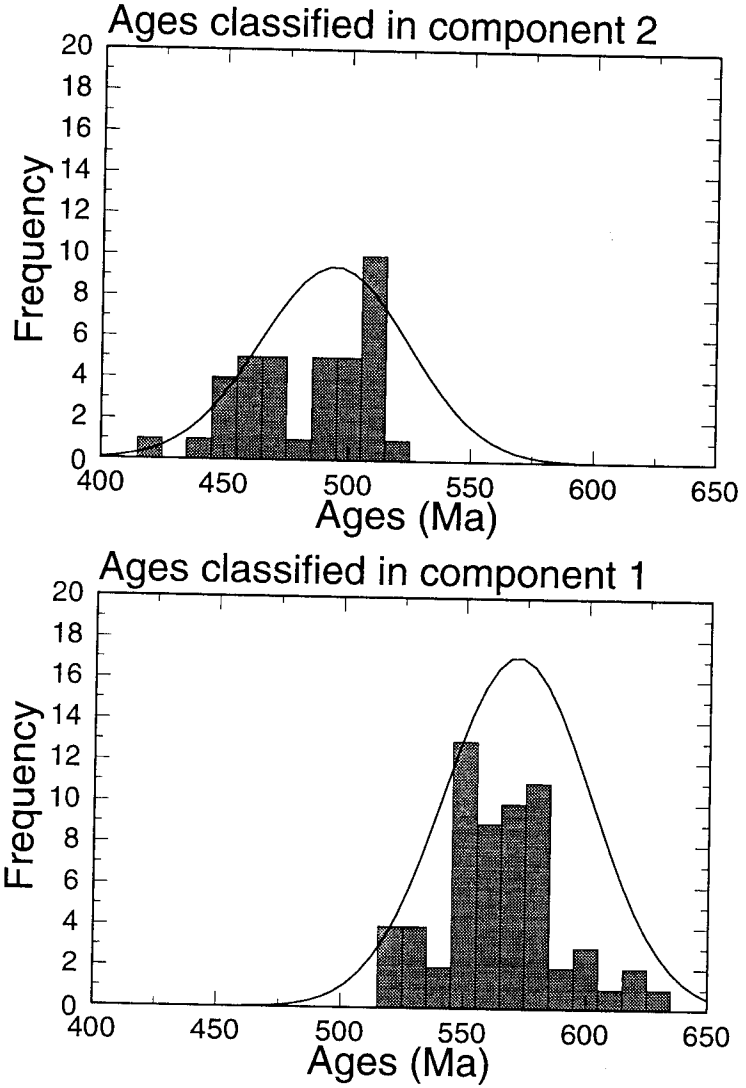

Fig. 5 Histograms of synthetic data classified into two components according to the final solutions (shown as Gaussian distributions). The combined histogram and two-component distribution is shown in Fig. 1. The number of ages assigned to each component can be a good indicator of the relative significance of the individual components.

(Erratum: This figure should replace figure 5.)