Provenance of sedimentary rocks

Problem Moderators:
Geoff Pritchard, University of Auckland.
Ken Russell, University of Wollongong
Bill Whiten, University of Queensland.
Paul Moloney, RMIT University

Abstract

Understanding the origins, or provenance, of a sedimentary deposit is an important aspect of geology. Sedimentary rocks are derived from the erosion of other rocks and thus provide important records of the geological environment at the time they were deposited. Some minerals found in sedimentary rocks, such as zircon particles, can be dated using uranium-lead techniques to trace the age of their parent rock thus providing useful information about the geological environment.

Statistical and mathematical analyses that can assist in the analysis of the distribution of ages of the zircon crystals are examined. Methods of defining a difference between the distributions of ages found in rock samples are proposed, and demonstrated in the division of multiple rock samples into clusters of similar types.

A test for the existence of a cluster is developed, and statistics for comparing different rock samples examined. Estimating an accurate age for the sedimentary deposit itself, proves to be difficult unless prior distributions providing significant extra information are available.

Keywords: age of sedimentary rocks, zircon ages, clustering, testing for clusters.

1 Introduction

Geoscience Australia is an agency of the Australian Government. It provides "geoscience information and services such as maps, earth monitoring and

strategic data about mineral and energy resources" [**3**]. It is very interested in the science of geochronology (analysing the age of geological events), which leads to a need to date rocks. This project is specifically interested in the dating of sedimentary rocks, such as sandstone, because these have been formed from the eroded detritus of other rocks and can thus provide a snapshot of the geological environment at the time when the detritus was deposited. Zircon crystals form in molten magna which cools to form igneous rocks. When the igneous rocks are eroded they form part of sedimentary rocks. The zircon crystals can have their age determined and thus provide valuable information on the components that have combined to form a sedimentary deposit.

As zircon crystals form in molten magma, they absorb uranium but not lead. Over time the uranium slowly decays to lead, providing a means to determine the ages of the zircon crystals. Dating a sample of the zircon crystals within a sedimentary rock produces an estimate of the distribution of ages of the igneous source rocks contributing to the sediment. This estimation is, of course, subject to the usual limitations of statistical sampling as the zircon grains whose ages are determined are only a small sample from all the grains in the deposit, and although great care is taken, the sample may have been contaminated by grains of other ages.

Two of the key questions are:

1. How many grains are needed to estimate the age distribution adequately?

2. How to compare or classify samples from different locations, on the basis of their age distributions?

Both questions imply the use of a metric to quantify the "distance" between two distributions. For the first, the term "adequately" implies that the estimated distribution is acceptably close to the true one. However this depends on the needs of the application, which will vary from case to case. For the second question, we need some objective measure of how similar one distribution is to another. Hence we focus on probabilities, distance measures and clustering of the age estimates.

2 Distance measures for probability distributions

A probability distribution μ can be thought of as a way of spreading a unit mass over the real number line $(-\infty, \infty)$. The mass assigned to (an interval or other set) A is $\mu(A) \in [0, 1]$. Often a distribution is identified with its cumulative distribution function (cdf) F, where $F(x) = \mu((-\infty, x])$. Some distributions can also be described by a density function f, where

$$\int_a^b f(x)\, dx \;=\; F(b) - F(a) \;=\; \mu((a, b])\;;$$

thus $f = F'$.

We will use several different metrics to quantify the dissimilarity between one distribution and another. In the following, suppose that μ_1, μ_2 are distributions; F_1, F_2 their cdfs; and (where appropriate) f_1, f_2 their densities.

Kolmogorov-Smirnov metric. Defined by

$$d_{KS}(\mu_1, \mu_2) \;=\; \max_x |F_1(x) - F_2(x)|\,.$$

Wasserstein metric. Defined by

$$d_W(\mu_1, \mu_2) \;=\; \int_{-\infty}^{\infty} |F_1(x) - F_2(x)|\; dx$$

or

$$d_W(\mu_1, \mu_2) \;=\; \int_0^1 |Q_1(u) - Q_2(u)|\; du,$$

where Q_1, Q_2 are (more or less) the inverses of F_1, F_2:

$$Q_i(u) \;=\; \min\{x : F_i(x) \geq u\}\,.$$

This is sometimes called the "earth-mover's distance": it is the average distance that mass must be moved to transform one distribution into the other.

The Kolmogorov-Smirnov and Wasserstein distances are illustrated in Figure 1. The Kolmogorov-Smirnov distance is represented by the length of the vertical arrow, and the Wasserstein distance by the area enclosed between the two curves.

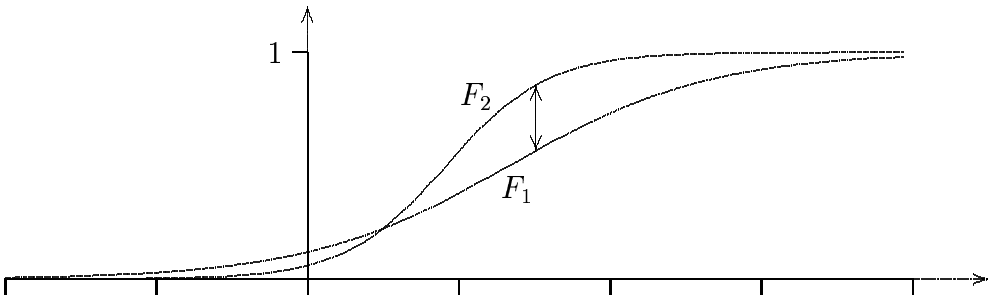

Figure 1: The Kolmogorov-Smirnov (maximum vertical difference) and Wasserstein (sum of areas between curves) metrics.

L_2 metric on densities. When dealing with distributions that have densities, we may use a metric on the density functions. For example, the L_2 metric, as used in [**5**]:

$$d_{L2}(\mu_1, \mu_2) = \left(\int_{-\infty}^{\infty} (f_1(x) - f_2(x))^2 \, dx \right)^{1/2}.$$

Normal mixtures. It can be assumed that the ages of crystals from one source follow a normal distribution, as they are based on counts of atoms, and the clustering procedures can overlap normal distributions to approximate the actual distribution of ages. When crystals from several sources are combined, this produces a mixture of Normal distributions of the form

$$\mu = \sum_{i=1}^{m} x_i \nu_i,$$

where $\nu_1, \ldots, \nu_m$ are normal distributions and $x_1, \ldots, x_m$ (the mixing proportions) are non-negative numbers with $\sum_{i=1}^{m} x_i = 1$. Within such a class of distributions, a metric may be based in a straight-forward fashion on the mixing proportions: if $\mu_1 = \sum_{i=1}^{m} x_i \nu_i$ and $\mu_2 = \sum_{i=1}^{m} y_i \nu_i$, then

$$d_{NM}(\mu_1, \mu_2) = \left(\sum_{i=1}^{m} (x_i - y_i)^2 \right)^{1/2}.$$

Example. To illustrate the differences between the various metrics, we consider the following example. Let μ_1 be a normal distribution with mean 0 and standard deviation w, and μ_2 a normal distribution with mean and standard

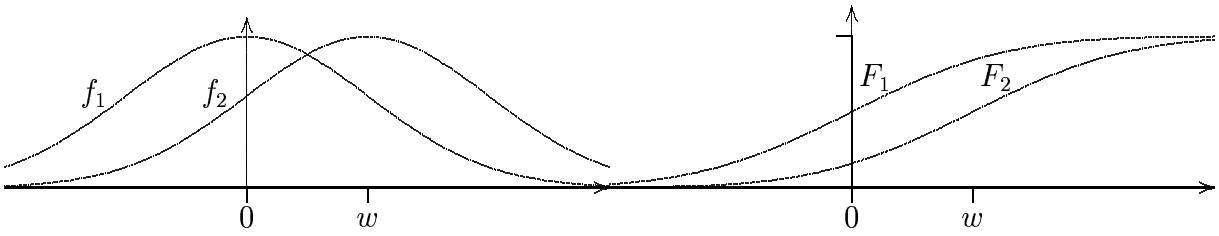

Figure 2: The distributions μ_1 and μ_2 showing densities f_1 & f_1 and cumulative distribution functions (F_1 & F_2).

deviation both equal to w (Figure 2). Then

$$\begin{aligned} d_W(\mu_1, \mu_2) &= w \\ d_{KS}(\mu_1, \mu_2) &= \Phi(1/2) - \Phi(-1/2) \approx 0.3829 \\ d_{L2}(\mu_1, \mu_2) &= Cw^{-1/2} \qquad \text{where } C \approx 0.3533 \\ d_{NM}(\mu_1, \mu_2) &= \sqrt{2}. \end{aligned}$$

Here Φ is the standard normal distribution function.

Suppose that w decreases to 0. Then the horizontal separation between μ_1 and μ_2 likewise decreases to 0, but only the Wasserstein metric recognizes this. On the other hand, as w is simply a scaling of the horizontal axis, the two distributions always intersect half way between the maxima, and thus the relative degree of overlap between the two distributions remains constant; this is recognized by the Kolmogorov-Smirnov metric. The L_2 metric has a tendency to blow up when distributions become concentrated (i.e. have high density values); according to this metric, μ_1 and μ_2 (rather unintuitively) move farther apart as w decreases. Finally, d_{NM} simply observes that our distributions are different components; neither the overlap nor the separation of the normal distributions is relevant.

3 Data

The primary data provided by Geoscience Australia consists of a collection red of 38 samples from across eastern and central Australia. The origins of

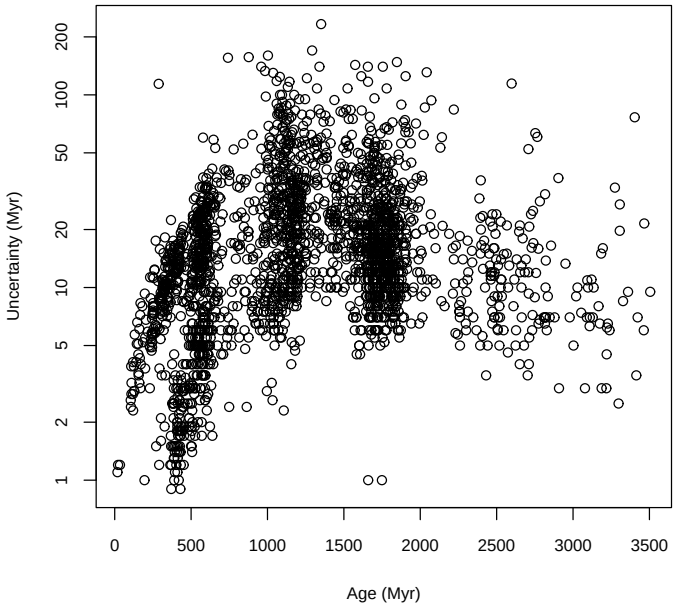

Figure 3: All the zircon ages and their uncertainties.

most of these samples are discussed in [**5**]. Each sample comprises age determinations for between 22 and 131 zircon grains taken from a single location, with each age determination having an associated 1-sigma uncertainty estimate. All 2374 age determinations in the dataset are illustrated in Figure 3.

A few additional samples were provided mainly for calibration purposes.

4 Clustering

Given a measure of the dissimilarity between two samples, based on the distribution of their zircon ages, we have a basis on which to compare and classify the samples within a larger collection. Most often, we would like to group the samples into a small number of clusters, with similar samples being grouped together. This is the problem of *clustering*, for which several standard algorithms are available; a good general introduction to the topic is [**4**].

Some clustering algorithms, including the popular k-means method, assume that the objects to be classified can be represented as points in some Euclidean space of fixed dimension. But in this application, we are clustering probability distributions (or estimates thereof), which cannot be easily represented in this way. We will therefore limit ourselves to clustering algorithms which can accept as input only a *dissimilarity matrix* giving the pairwise dissimilarities between the samples in our collection.

We investigate the use of four such algorithms commonly used in the statistical literature. Two, the "partitioning around medioids" (PAM) method and "fuzzy analysis clustering" ("fanny"), partition data into a pre-selected number of clusters. Two more, "agglomerative nesting" ("agnes") and "divisive analysis clustering" ("diana") construct a hierarchy of nested clusters. All four methods are described in [**4**] and implemented in R.

The output of the hierarchical methods can be represented as dendrograms, as shown in Figure 4 for our samples. These suggest that the distance measure used will have more effect on the result than the clustering method.

We attempt to group our set of 38 samples into five clusters, using each of our four methods and the Kolmogorov-Smirnov and Wasserstein distances. (This number of clusters was chosen because 25 of the samples are the same ones appearing in Section 6.4 of [**5**], where they are classified into four clusters. Most of the remaining samples contain younger grains, so tend to cluster

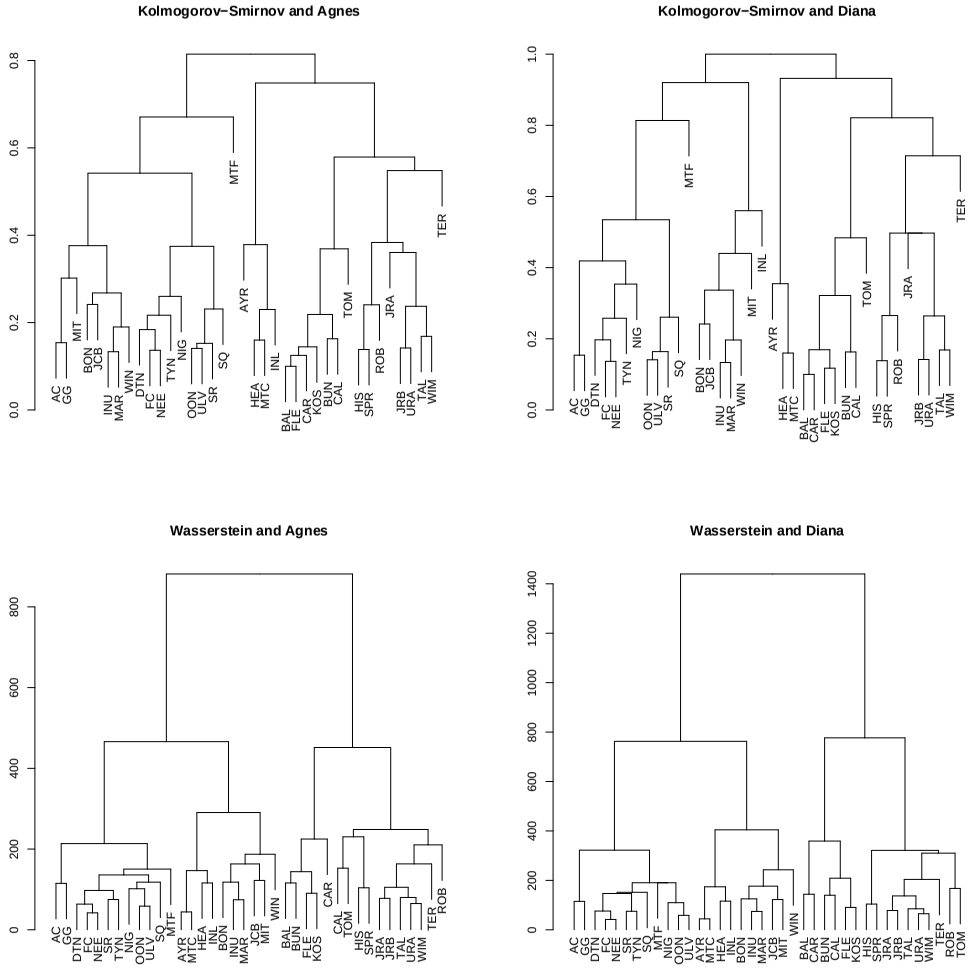

Figure 4: Hierarchical clustering.

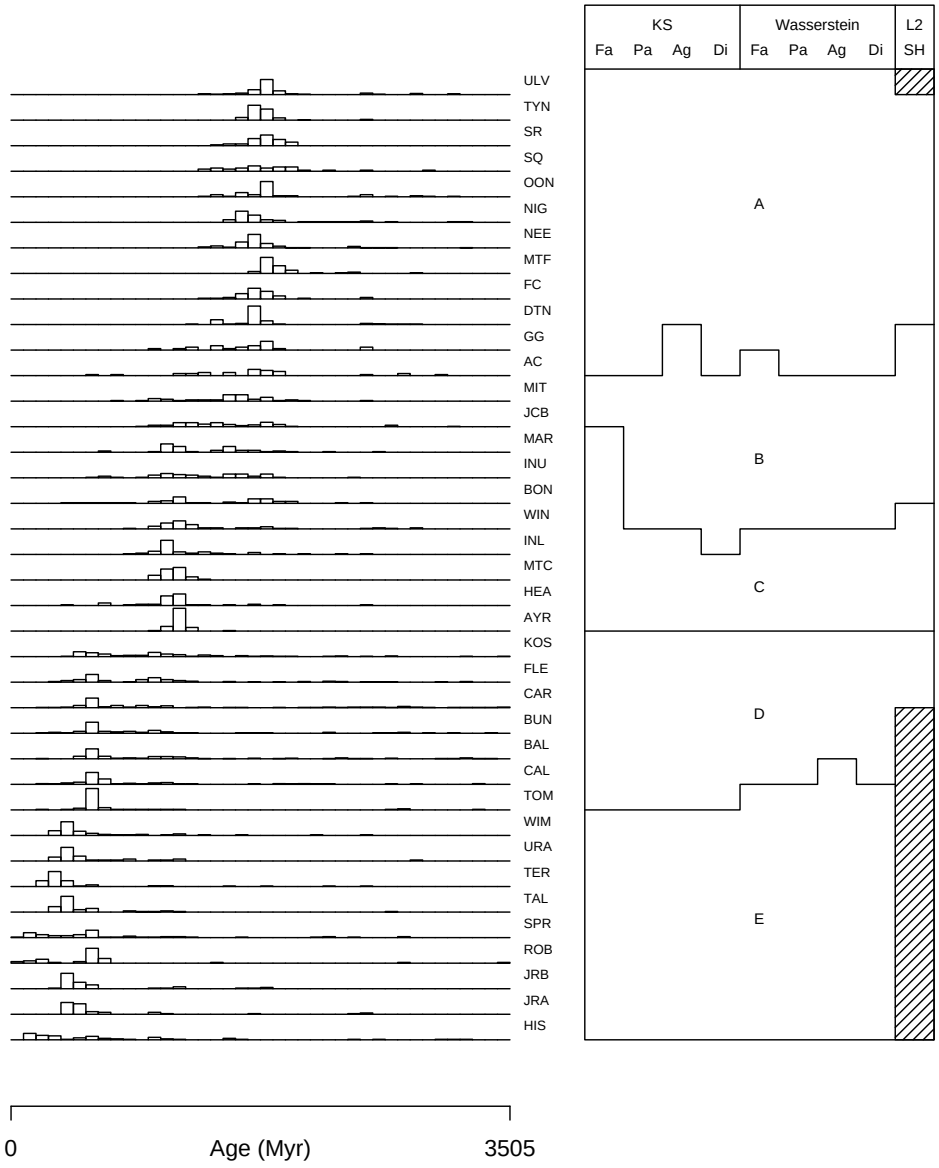

Figure 5: Age distributions of zircons sampled from 38 sites (identified by the codes in the central column, that indicate the location in central and eastern Australia from where the samples came), and classifications into five clusters. The rightmost column shows the results of [**5**] for 25 of the samples.

with each other.) Five recognisable clusters emerge (see Figure 5):

- *Cluster A* samples are dominated by relatively old grains (around 1750 Myr);
- *Cluster B* samples have the greatest variety of grain ages, often resembling a mixture of the 'A' and 'C' age distributions;
- *Cluster C* samples contain mostly grains around 1150 Myr old;
- *Cluster D* samples are dominated by younger grains (around 600 Myr);
- *Cluster E* samples contain the youngest grains of all (400 Myr or less).

The various methods and distance measures produce fairly similar results, with disagreement for only a few of the samples. The Kolmogorov-Smirnov metric has a tendency to isolate some samples. (For the Agnes method, it places MTF in a cluster all by itself, but we have chosen to ignore this rather than allow a sixth cluster.) The Wasserstein metric (and the L_2 metric of [**5**]) produce a more definite classification. However experience in the use of these methods will determine which is the most useful in providing geological information.

5 Determining if a cluster of ages exists

One question of interest is whether several ages that occur close together can be considered to form a cluster of ages that can be given a geological interpretation. Guidelines for determining when ages can be considered a cluster were developed, using the assumption that, if the ages are distributed with a uniform distribution, no cluster exists.

The width of a group of k ages relative to the total age range, can be used to test whether those ages form a cluster. Calculate the probability that this relative width will be achieved by k adjacent observations selected from n observations generated from the uniform distribution. If this probability is "too low" the alternative hypothesis that these k ages form a cluster becomes reasonable. Note that this is the probability of a Type I error, that is, inferring that a cluster is present when it is not present, as the probability is generated from a uniform distribution, which does not generate clusters.

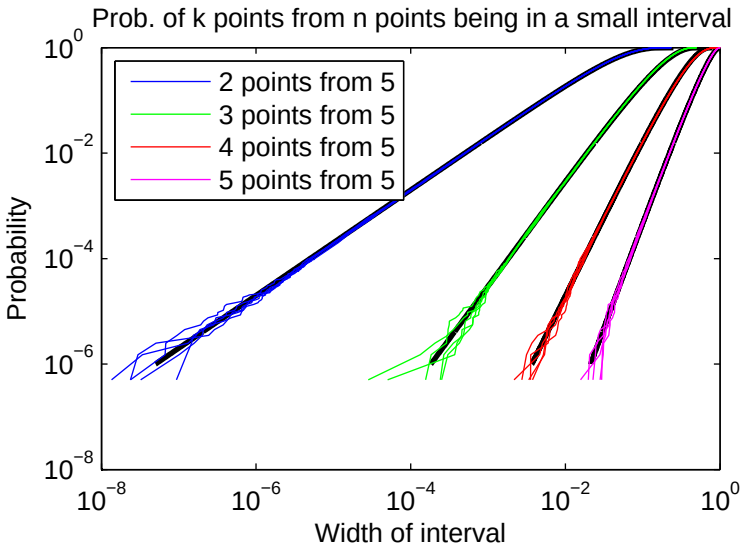

Figure 6: Simulated probabilities of a given width, on top of approximate formula in black.

The probability that $k = n$ values chosen at random from a uniform distribution exhibit a cluster can be calculated, either from theory (for instance using information from chapter 2 of [**1**]) or simulation.

The range (from minimum to maximum) spanned by the entire sample of n uniformly distributed points has a Beta distribution with parameters $n-1$ and 2. The probability that this range is less than x is thus

$$\frac{\int_0^x t^{n-2}(1-t)dt}{\int_0^1 t^{n-2}(1-t)dt} \tag{1}$$

which simplifies to:

$$nx^{n-1} - (n-1)x^n \quad , \tag{2}$$

and as we are only interested in small values of x this is approximately:

$$nx^{n-1} \quad . \tag{3}$$

The range spanned by k of the n uniformly distributed points, from the rth to the $(r+k-1)$th smallest for a fixed value of r, has a Beta distribution

with parameters $k-1$ and $n-k+2$ (note that this distribution does not depend on r). The probability that this range is less than x is thus

$$\frac{\int_0^x t^{k-2}(1-t)^{n-k+1}dt}{\int_0^1 t^{k-2}(1-t)^{n-k+1}dt} \tag{4}$$

Again simulations show an excellent agreement. However the actual case of interest is the minimum value of the width of any group of k points out of n. This appears to be more difficult to determine by theory but an examination of simulations suggests the behavior at small widths is very close to the above Beta$(k-1, n-k+2)$ distribution with the width multiplied by the factor $(n-k+1)^{1/(k-1)}$:

$$\frac{\int_0^{x(n-k+1)^{1/(k-1)}} t^{k-2}(1-t)^{n-k+1}dt}{\int_0^1 t^{k-2}(1-t)^{n-k+1}dt} \tag{5}$$

Figure 6 shows some typical cases of the simulations and this formula. Equation 5 gives, for a fixed width x, a higher probability than that given by equation 4, as the minimum width of any k points out of the n available is selected. When $k=n$ this formula is the same as given earlier. Similar to the previous case, this formula can be approximated for small values of x as:

$$\frac{n!\ x^{k-1}}{(n-k)!\ (k-1)!} \tag{6}$$

For example, given the ages 1000, 1200, 1210, 1230, 2000, 3000 million years over the range of 4000 million years, the pair 1200, 1210 has a probability of 0.05 of a chance occurrence, which gives a marginal suggestion of a cluster. The triple 1200, 1210, & 1230 has a probability of 0.0017 of occurring by chance, which gives a strong suggestion that these are not random but form a cluster. In both cases there is essentially no difference between the simulated result and the formulas 5 and 6.

6 Comparing groups of clusters

Having decided on the groups of ages that exist in samples, the number of ages found in each group can be determined. Comparing samples is then a

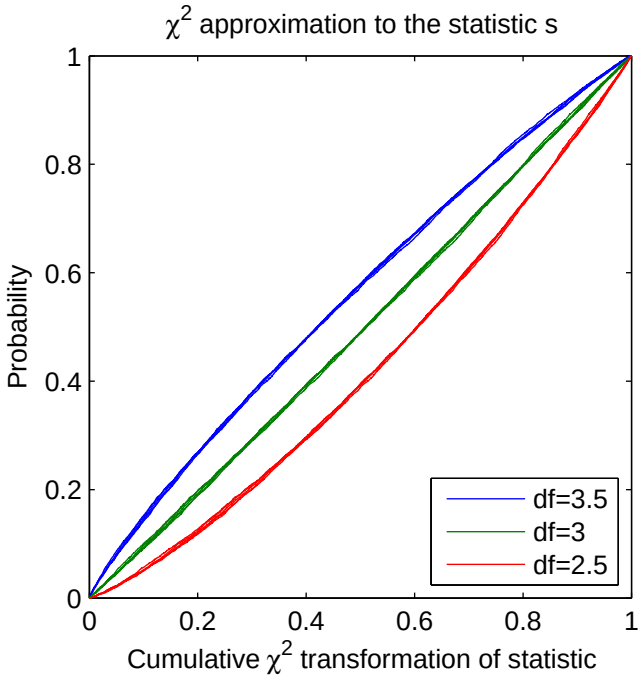

Figure 7: A comparison of the distribution of the statistic s of equation 9 after being transformed using cumulative χ^2 distributions.

matter of comparing the proportions in each group, as suggested earlier with the statistic d_{NM}.

Let there be k clusters. Suppose that sample j has $m^{(j)}$ distinct ages, and that $n_i^{(j)}$ of these ages are found in the ith cluster ($j = 1, 2$; $i = 1, \ldots, k$), with $\sum_{i=1}^{k} n_i^{(j)} = m^{(j)}$. The following statistic based on contingency table comparisons [2] can be used to compare the two samples:

$$s = \sum_{i=1}^{k} \frac{(n_i^{(1)} - r_i m^{(1)})^2}{r_i m^{(1)}} + \sum_{i=1}^{k} \frac{(n_i^{(2)} - r_i m^{(2)})^2}{r_i m^{(2)}} \tag{7}$$

where r_i is the combined estimate of the proportion in the ith cluster:

$$r_i = (n_i^{(1)} + n_i^{(2)})/(m^{(1)} + m^{(2)}) \tag{8}$$

This statistic can be simplified by eliminating r_i, giving:

$$s = m^{(1)} m^{(2)} \sum_{i=1}^{k} \frac{(n_i^{(1)}/m^{(1)} - n_i^{(2)}/m^{(2)})^2}{n_i^{(1)} + n_i^{(2)}} \tag{9}$$

which for larger values of $n_i^{(1)}$ and $n_i^{(2)}$ can be approximated by a χ^2 distribution with $k - 1$ degrees of freedom. Figure 7 shows a particular case compared to χ^2 distributions. Four clusters with proportions 0.5 0.25 0.15 & 0.1 and numbers of ages 100 & 50 in the samples were simulated 10000 times. Each curve on the graph consists of five overlapping repeats of this simulation. The approximation is good in this case, and seems to be only slightly less accurate for cases with lower numbers of ages. Thus this χ^2 distribution can be used to give a probability that provides a useful measure of the difference between distributions. This probability can be used in the clustering algorithms to divide the age distributions into groups with similar properties.

7 Estimating the age of sedimentary deposits

One piece of information that is of particular interest to geologists is the age of sedimentary deposits. Assuming no outlier has intruded, the age of the sedimentary deposit is clearly later than the most recent zircon age. Additional prior information can be added (Figure 8), on the likely time between the formation of the zircon grains $f(t)$ and the formation of the

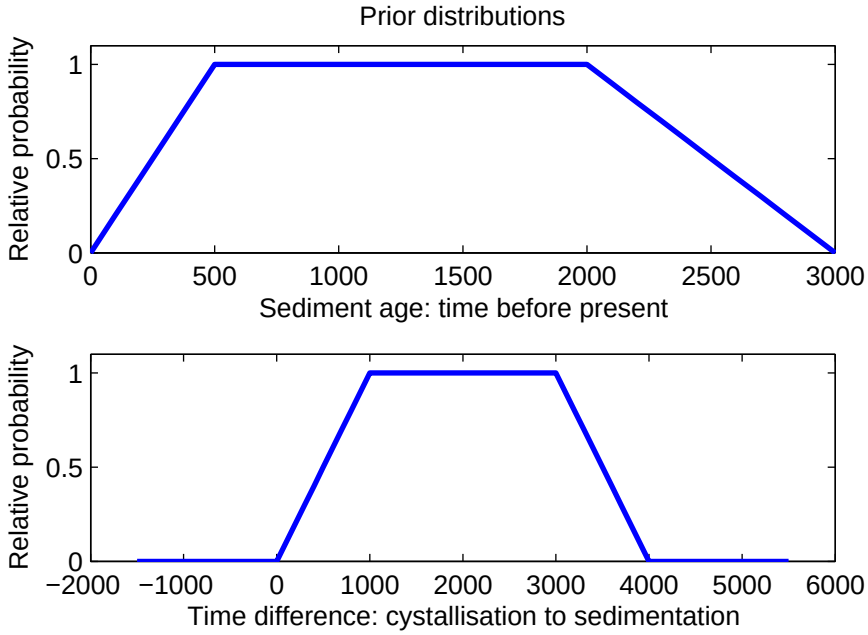

Figure 8: Prior distributions for sedimentary deposit age, and the time between formation of zircon crystals and formation of sedimentary deposit.

sedimentary deposit, and on the range of possible ages for the sedimentary deposit $a(t)$. As the probabilities will be normalised, both $f(t)$ and $a(t)$ can be expressed as relative values.

The observed ages of the zircon crystals z_i (Figure 9) can be combined with the prior for the time difference as the geometric mean of the individual probabilities for each crystal sample. This geometric mean provides an estimate that is not biased by the number of age samples, and is multiplied by the prior for the likely age of the deposit, to get an estimate proportional to the probability density function of possible ages of the deposit:

$$\prod_i a(x) \left(\prod_i^n f(z_i - x) \right)^{1/n} \tag{10}$$

for age x. Integrating and normalising this gives the cumulative distribution

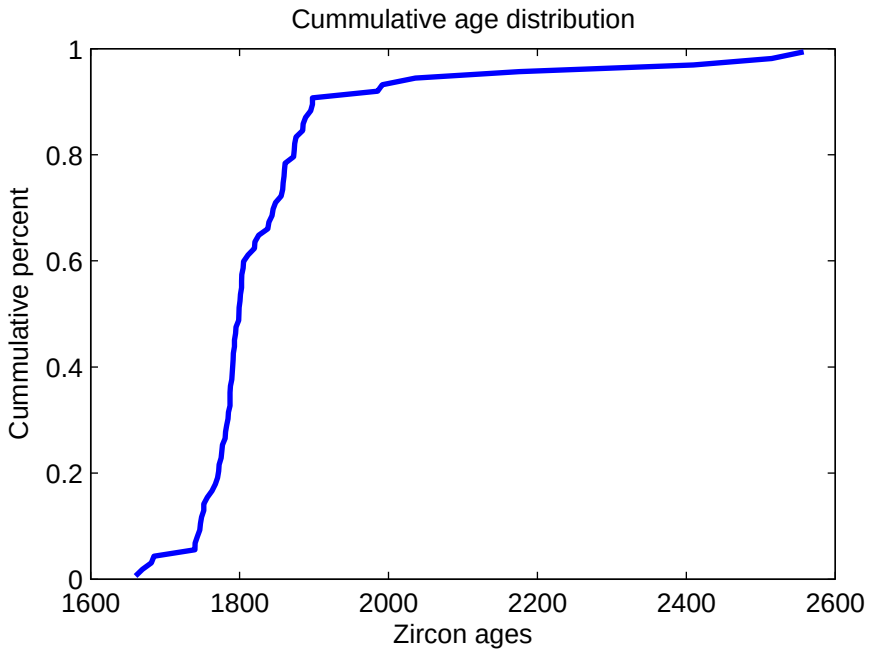

Figure 9: Cumulative distribution of zircon ages.

function:

$$\frac{\int_0^x a(x)\left(\prod_i^n f(z_i - x)\right)^{1/n} dx}{\int_0^{x_{max}} a(x)\left(\prod_i^n f(z_i - x)\right)^{1/n} dx} \tag{11}$$

For the priors given in Figure 8 and the zircon age distribution in Figure 9, Figure 10 gives the cumulative distribution of the estimated age of the sedimentary deposit. Figure 10 shows a wide spead for the possible ages of the sedimentary deposit. The zircon ages only provide an upper limit to the age of the sedimentary deposit. To obtain a more refined estimate, additional information needs to be provided in the assumed prior distributions. Then the accuracy of the prior distributions determine the accuracy of the estimate of the sedimentary deposit age.

To allow for the possibility of an occasional more recent outlier in the zircon ages, a small probability of negative age differences can be included in the prior.

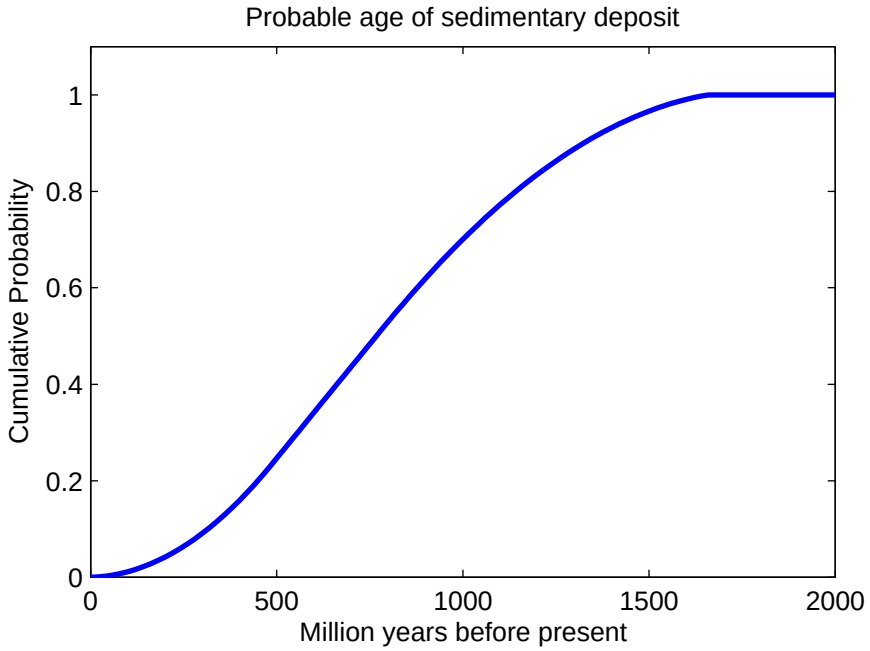

Figure 10: Estimated cumulative distribution for sedimentary deposit age. Calculated from information in figures 8 & 9, and equation 11.

8 Conclusions

Interpreting zircon age estimates is an interesting problem that introduces multiple mathematical aspects. The distribution of the ages can be defined directly as an empirical distribution using the data. Alternatively, standard smoothing or clustering techniques can be used to obtain an estimate of the underlying distribution of the age data. Expressing the distribution as the sum of Normal clusters has the advantage of allocating the ages into groups that may have geological significance.

Several different metrics can be used to quantify the difference between the distributions of ages found for different rock samples. These measures can then be used in a clustering program to estimate relations between the different rock samples. This provides a quantitative method of relating the rock samples.

A particular interest is when the measured ages form a cluster that can be examined for geological significance. Determining the probability of an apparent grouping of ages forming when the age are generated from a uniform distribution provides a method of testing for the existence of a cluster. The distribution of this statistic can easily be generated from a Monte Carlo simulation, and a useful approximation to the low probability part of the distribution is given.

Given the number of ages in clusters, a measure of the difference between rock samples can be specified in a manner similar to that used for contingency tables. A close approximation using the χ^2 distribution is available and more accurate probabilities, if needed, can be obtained by simulation.

Statistics with the help of prior distributions can be used to calculate a distribution of the probable age of the sedimentary deposit. Unfortunately, unless there is useful additional information input via the prior distributions, the deposit age estimate is not well determined. However, making the assumed information explicit as prior distributions gives a clear description of the assumptions made.

The worth of the different methods of analysing the zircon age data ultimately depends on the value they provide in the subsequent geological interpretation, and the extent to which they make the procedure used quantitative. It remains for geologists to evaluate which of the alternatives proposed add the greatest value to their work.

9 Acknowledgements

The project moderators are grateful to the industry representative Dr Keith Sircombe from Geoscience Australia for bringing this problem to MISG and for his patience and enthusiasm in presenting this problem to us. There was also an enthusiastic group that worked on this problem including Prof. David Griffiths, Dr John Ormerod (U. Wollongong), Kaye Marion, Dr Lynne McArthur, Petra Siskos (RMIT U.), Dr John Coghill (retired), Judith Shand (NSW Dept of Education and Training), and the two undergraduate trainees Wei Xian Lim (U. Wollongong), Fatima Hassib (La Trobe U.). The moderators are please to acknowledge the inputs from this group of people. Prof Geoff McLachlan suggested a possible relation to contingency tables.

References

[1] David, H. A. and Nagaraja, H. N., *Order Statistics,* Wiley (2003)

[2] Everitt B. S., *The analysis of contingency tables,* Chapman & Hall (1992)

[3] Geoscience Australia, *Briefing note for 2009 MISG project,* MISG 2009

[4] Kaufman, L. and Rousseeuw, P.J., *Finding Groups in Data: An Introduction to Cluster Analysis,* Wiley (1990)

[5] Sircombe K. N. and Hazelton M. L., *Comparison of detrital age distributions by kernel functional estimation,* Sedimentary Geology 171 (2004), 91-111.