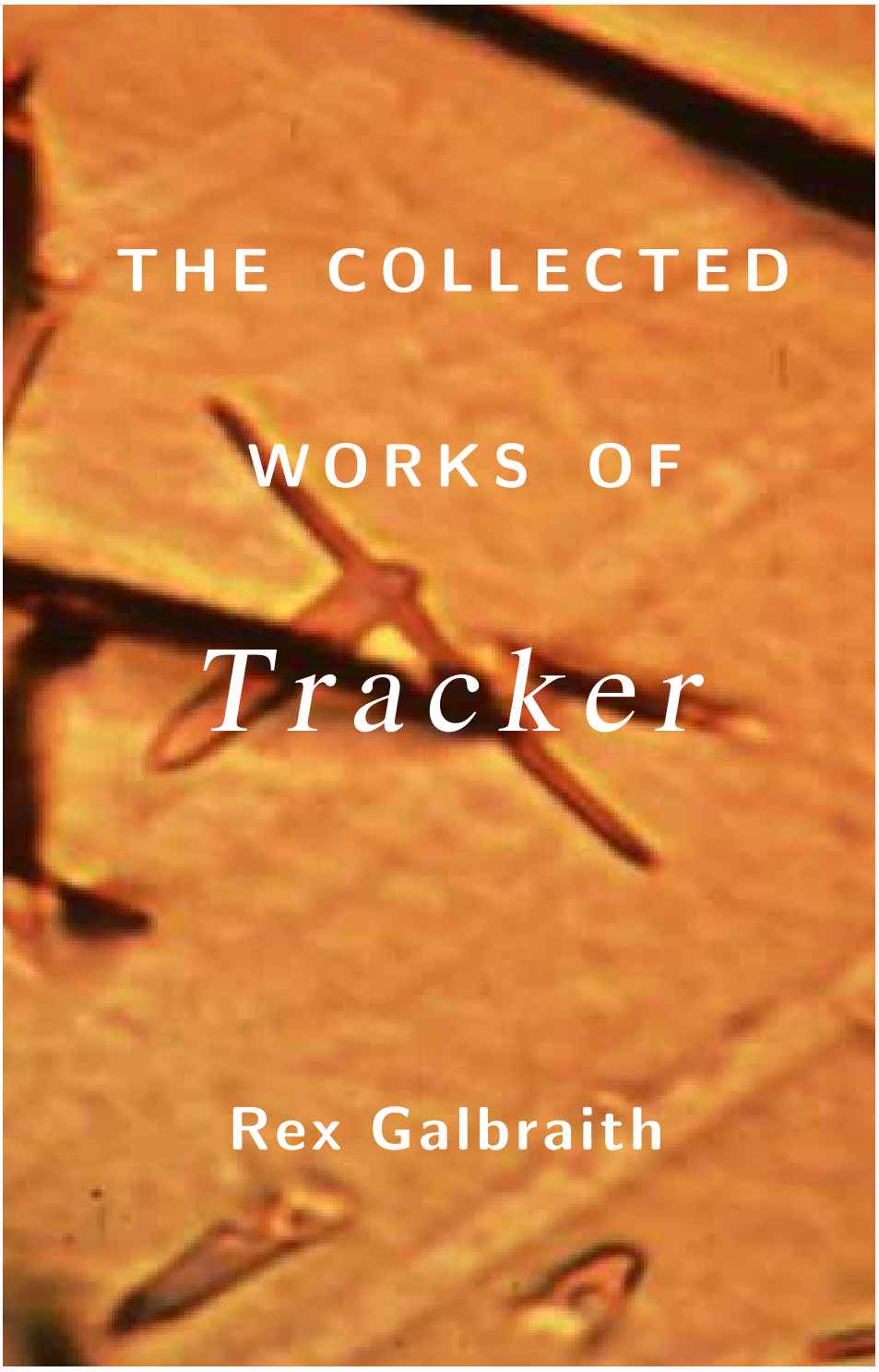

THE COLLECTED
WORKS OF
Tracker
Rex Galbraith

THE COLLECTED WORKS OF *Tracker*

Your man at the microscope

This edition published in 2017 by Rex Galbraith,
Department of Statistical Science, University College
London, Gower Street, London, WC1E 6BT, UK

ISBN 978-1-78808-177-1 (e-book)
ISBN 978-1-78808-816-9 (printed)

Typeset in LaTeX by the author

Cover design of fission tracks in apatite
adapted from an original photograph kindly
provided by Paul Green, Geotrack International,
37 Melville Rd, Brunswick West, Vic 3055, Australia

Printed in the UK by Dolman Scott Ltd
1 High Street, Thatcham, Berks, RG19 3JG

An electronic version of this book is available
in Open Access format at UCL Discovery
http://discovery.ucl.ac.uk/1576530/

Preface

THIS BOOK contains six articles that I wrote in the 1990s under the pen-name *Tracker*. They appeared in *On Track*, the Newsletter of the International Fission-Track Community. I have re-produced them here with minor corrections and edits, and I have re-drawn most of the figures. I hope that having them available in one volume will be appreciated by old and new trackers and their colleagues.

TRACKER'S ARTICLES were intended to be instructive — explanations of statistical concepts relevant to researchers in the fission track community (and to scientists in other fields) — and hopefully sufficiently light hearted to entice such readers. They arose mainly from topical questions put to me at scientific meetings and in correspondence. Fission track analysis was developing as a new science, not only for dating events on geological time scales but also for learning about geological thermal histories, provenance and

landscape evolution. Fission track analysis is inherently statistical in nature: tracks form continuously over time and are subject to heating and other processes that characterise the statistical distributions of their numbers, lengths and orientations — so ideas of probability and statistics are at the heart of the subject. I found it amazing how the microscopic fission track measurements can give such detailed quantitative information on geological time scales.

THERE ARE at least two further pieces that *Tracker* failed to write: *A Fit of Peak*, about finite mixture models, and *Projected Rights and Wrongs*, about inferences from projected fission track lengths and angles, conventionally denoted by r and ω. The latter was motivated (or would have been) by mathematical fallacies that can arise when line segments randomly generated in three dimensional space are observed in two dimensions; and also by a Gary Larson cartoon showing two scientists in white coats studying a complicated equation on a blackboard — having r on the right hand side and various terms such as $4(Wr)^2$ on the left — with the caption "Yes, yes, I *know* that Sydney ... *every*body knows that! ... But look: Four wrongs *squared*, minus two wrongs to the fourth power, divided by this formula *do* make a right." These

topics and many others are covered at greater depth in my book *Statistics for Fission Track Analysis*, Chapman & Hall/CRC, 2005.

THE FISSION TRACK community in the 1990s consisted almost entirely of geologists, physicists, geochronologists and geophysicists. The imaginary *Tracker* tried to pose as one of these but sometimes betrayed himself by an ignorance of geology and a surprising familiarity with the history of statistics at University College London. As most readers will know, I was a teacher and researcher in statistical science at UCL. I was introduced to fission track dating by Paul Green, when he worked there before moving to Melbourne, and I was lucky to have a brilliant statistical ally, Geoff Laslett, who lived in Melbourne. We worked closely with the leading scientists in this field, who accepted us as part of their community. We were both passionate about science and about the use of statistics in science, and we shared a love of literature, cricket and cryptic crosswords. Sadly Geoff died of cancer in 2010. I thank both Paul and Geoff for their inspiration and encouragement.

I HAVE ADDED a list of names of fission-trackers (*Tracker's* contemporaries) at the end of the book, in alphabetical order of given name, as that is how I

remember them. Like most such lists, it is somewhat arbitrary, and I apologise to anyone I may have omitted. I have personally met or corresponded with the vast majority of these people and I have learnt much from them — not only about this field but also about how science works. I thank them all for that and for the pleasure of their company. I also thank Jane Galbraith, my wonderful wife and colleague and much more.

Rex Galbraith, February 2017.

Contents

It's All a Plot

My famous namesake wrote *The Age of Uncertainty*, a lofty enterprise well suited, no doubt, to the whimsical world of economics, where time is money and the only certainties are death and taxes. But in the hard world of rock and acid, of hot times and revealing cleavages — down, so to speak, at the apatite face — it is the *uncertainty of ages* that is of greater concern. And not only the uncertainty but also the *variation*.

Fission track age

What is a fission track age? You need to distinguish between a *true* age and an *estimate*. An estimate is what you get from observed data. A true age is what your estimate is an estimate of: the value, if you like, that the age equation would deliver if zeta, rho-d, rho-s and rho-i were measured perfectly, with zero error (and let's not worry here about what we mean by error). The true age may or may not have a true meaning: it may correspond to the time since your sample cooled below some temperature, or it may be something more ethereal — a reflection of a

meandering thermal history, reminiscent, perhaps, of Mae West's invitation — full of promise but almost certainly later.

For a sample of grains, the age estimates will vary even when the true ages are the same. And when the true ages differ, the spread of estimates will be greater than the spread of true ages.

Chi-square test

SOME YEARS AGO, I was disconcerted to see a headline in the London Times:

Chi-Chi dying

The thought that this noble statistic might be falling into disuse, implausible though it seemed, was alarming indeed. One could well imagine how the report would go:

> "The world of fission track dating was rocked to its foundations today when it was revealed that . . . "

which would be echoed in other newspapers under headlines like "Test frequency fails to reach expectation" (Guardian), "Only 5% reported use of statistic" (Financial Times), "Dating agency drops age test" (Daily Mirror) and "Chi-square out for

the count" (Sun Sport). But it turned out to be a story about London Zoo's giant panda, which in its own way was no less tragic, I suppose. Today it is London Zoo itself that is dying.

WHAT DOES the chi-square test do? It tells you whether your sample age estimates are consistent with a single true age, or whether there is evidence of a spread in true ages. If your data are consistent with a single true age, this does not mean that there really is a single true age, but merely that there is little evidence to the contrary. And if there is evidence of a spread, it doesn't directly tell you how much spread. To put it another way, the upper chi-square tail probability (the P-value) measures the amount of *evidence*, not the amount of spread.

Of course, if there is a large spread (of true ages) and you have moderately large counts, you will get a large chi-square statistic. But you may also get a large chi-square statistic, or small P-value, even when there isn't much spread because you have large counts — you have a lot of evidence that there is *some* spread.

On the other hand, it often happens that your data "pass" the chi-square test even when there really is a spread, just because there isn't much evidence — usually because the spontaneous track

counts are small. Your data can be consistent with a single true age and, at the same time, be consistent with a spread of true ages. And if the counts are very small there could be quite a large spread.

So WHAT do you conclude? If you are dating an age standard, or determining your zeta, then you know that your sample age estimates should agree and you would expect your data to "pass". (I use the common jargon, though I prefer to quote the P-value and avoid the pass/fail language — after all, the messages from P = 0.04 and P = 0.06 are essentially the same.) If it does, and if you have sufficient data that you would have found a spread in ages if there was one, then you have increased confidence in your result. If it "fails" you would look to your experimental procedure for an explanation.

But if you are analysing a field apatite sample with an unknown thermal history, you may suspect that some track annealing has taken place and that there is a spread in true ages. So even if your data did "pass" the chi-square test, you would not *automatically* assume that there is no spread. Especially if you have small counts. You would normally consider track length measurements and related samples and view all the evidence in a

wider context. And if your data "fail" the chi-square test it could be due to genuine variation.

AT THIS POINT, I wish to *scotch*, once and for all, a pernicious rumour that heretofore has been largely contained beneath the earth's crust, even if, to pursue the metaphor, it has not been deeply buried, but of late seems to have been gaining credibility in certain quarters. I refer, of course, to the scurrilous notion that there is nothing new about the chi-square test — some have even said it is as old as the hills. I wish to state categorically that there is absolutely no truth in this totally unfounded allegation. It was as recently as 1900 that Karl Pearson, founder of the UCL Statistics Department, and almost certainly a genius, for R.A. Fisher disliked him intensely, first derived the chi-square test for assessing the agreement between observed and expected frequencies. Fisher's animosity — indeed it was mutual — was greatly reinforced in 1922 when he pointed out that Pearson had got his degrees of freedom wrong — and although Fisher provided the essential mathematics it was as late as 1957 that David Finney, renowned in statistical circles for his work on selecting insecticide doses that kill bugs half dead, explicitly applied it to pairs of counts from independent Poisson distributions when

comparing two paper cleaners. And even then it wasn't used for ages.

Radial Plots

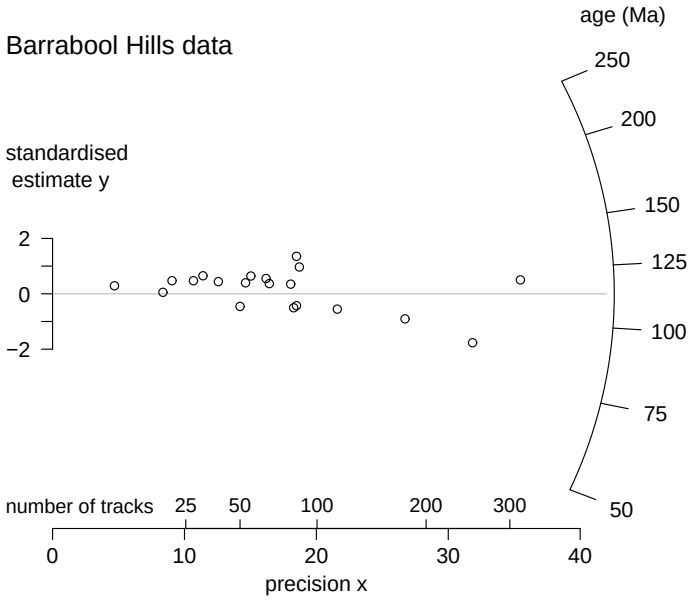
"A MAN and his *hobby-horse* . . . " quoth Tristram Shandy, though it would be difficult to find a man with such a hobby-horse as to compare with that of his Uncle Toby: yet will I follow his example who "would use no other argument to prove his *hobby-horse* was a *hobby-horse* indeed, but by getting upon his back and riding him about — leaving the world, after that, to determine the point as it thought fit".

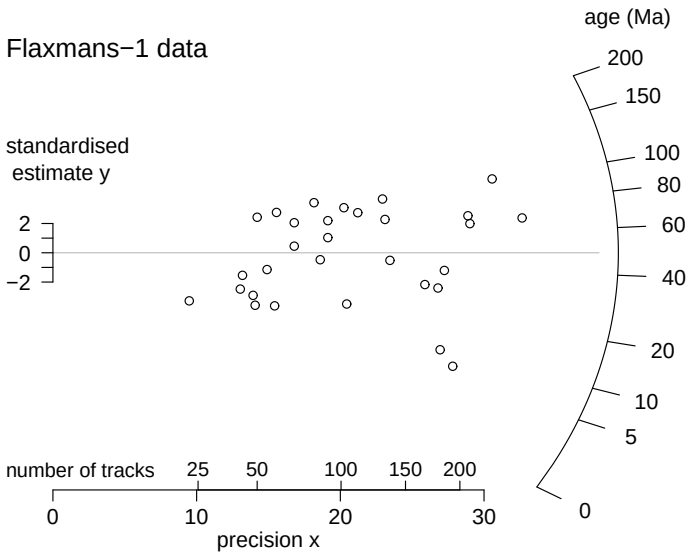
So "What the devil does the plot signify, except to bring in fine things?" The question could hardly be better put. If you have two estimates, one very precise and the other not, you place more credence in the precise one — you do not weight them equally. A radial plot displays your single grain ages so that they can be seen and compared *at face value*. It automatically gives them the right weights in accordance with their precisions. If you plot just the ages, ignoring their differing precisions, it is practically impossible to compare them sensibly.

In a general sense, there are two reasons why age estimates vary: because they differ from the true ages (there is *uncertainty*) and because the true

ages themselves vary (there is *variation*). In a radial plot, you can display the uncertainty so it is easier to assess the variation. You can judge whether single grain ages look homogeneous, and check your view, if necessary, by the chi-square test. And you can go further and assess visually the spread in ages, comparing it with possible explanations or possible models. It is a tool for assessing your data, and it is a tool for showing it to others.

BUT LET US climb upon our hobby-horse and ride him about.


Barrabool Hills data
age (Ma)
250
200
150
125
100
75
50
standardised
estimate y
2
0
−2
number of tracks
25
50
100
200
300
0
10
20
30
40
precision x


Flaxmans−1 data
age (Ma)
200
150
100
80
60
40
20
10
5
0
standardised
estimate y
2
0
−2
number of tracks
25
50
100
150
200
0
10
20
30
precision x

THE PLOT of age estimates from the Barrabool Hills, Otway Ranges — only a stone's throw from that delightful restaurant in Apollo Bay (opening hours are extended to 9:30 pm on Fridays) — illustrates apatite data with a clean thermal history. These grains were cooled at about 120 Ma and have not experienced significant heat since. The estimates look homogeneous on the graph and the chi-square test confirms that there is no evidence to the contrary. (The chi-square statistic is 10.0 with 19 degrees of freedom, giving a P-value of 0.95.) We have plenty of data and the picture seems clear: if there really was a spread in true ages we feel we would see more sign of it.

THE FLAXMANS-1 apatites, also from the Otway Basin, are from a depth of 2.6 km and a current temperature of 92°C. Fission tracks have been annealed by varying amounts and so the *true* ages vary between grains. There is clear heterogeneity in the graph; the estimates vary from 0 to 120 Ma and the pattern invites further study. The heterogeneity is so obvious that a formal test is not necessary. In fact the chi-square statistic is 241.4 with 29 degrees of freedom. It is large partly because there is a wide spread and partly because there is a lot of information. These two examples reinforce each other.

Radex and bar exercises

BEFORE the next plot, try the following exercises. Draw a straight line on a strip of transparency: this is your *radial extrapolator* or radex, for short. Place it on the graph so that it passes through the origin and through a point. Read off the age on the circular scale.

Make a copy of the ± 2 y-scale, also on a piece of transparency. This is your two-sigma error bar; the same bar applies to any point. Centre the bar, parallel to the y-scale, on a point; hold one end of your radex at the origin and move the other end up and down as much as possible so that it always intersects the bar. The resulting interval on the age scale is the two-sigma age confidence interval for that grain. Repeat this for other points. See how you get a shorter confidence interval from a more precise estimate. Optimists can do one-sigma confidence intervals in similar style.

WHAT ABOUT the pooled age and its standard error? For this you might need more space (your local bar will do nicely) and an extended radex. Simply plot the (x, y) point corresponding to the pooled age on your graph (probably off the page), centre your error bar *there* and repeat the previous exercise. (In the above graphs, the pooled age is at $y = 0$ and $x =$ twice the square root of the total

number of tracks — spontaneous plus induced — for all grains. This ignores error in zeta and rho-d but you can add that roughly in your head.)

To COMPARE several estimates together, place your radex line so that it goes from the origin through the middle of the points. Slide the centre of the error bar along the line, keeping the bar parallel to the y-axis; see if you can catch all (or nearly all) of the points inside the bar. Adjust the line and repeat if necessary. If you can, the estimates are roughly consistent with each other. There will be times when is isn't obvious whether or not they are; you can check formally using a chi-square test. You could do this with all of the points, or with a prescribed subset. With two lines you could see whether the estimates are consistent with a two-component mixture.

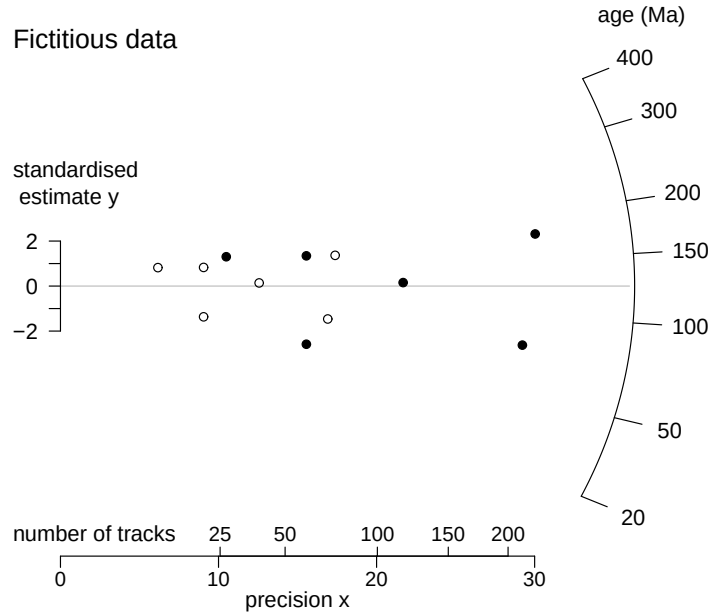
TURN the graph so that the y-scale is horizontal. Imagine that you are standing at the origin looking out to the age scale. Slowly tip the paper away from you so you are eventually looking along the line from the origin through the points. Tip the paper back and forth a few times. See how many points you can cover visually with the y error bar. Also, if the points with higher precision tend to fan out more than those with lower precision, that is

evidence of variation in true ages. This is dynamic graphics. Now lean on the bar ...

A convenient fiction

THE LAST GRAPH shows two sets of data I collected myself. In the first set, plotted as open circles, the pairs of counts representing numbers of spontaneous and induced tracks are $(2, 18)$, $(3, 6)$, $(6, 14)$, $(9, 30)$, $(11, 60)$ and $(22, 53)$. I confess I made them up. I used only six grains so the graph would not get too cluttered. They look reasonably homogeneous. The chi-square statistic is 7.0 with 5 degrees of freedom, giving a P-value of 0.22. There is not much evidence here of any heterogeneity.

TO GET the second set of data, plotted as closed circles, I multiplied every count in the first set by 3. The ratios N_s/N_i are exactly the same as before and so are the single grain ages. Each point has moved away from the origin along its radius. The mean age is the same for the two data sets, as is the standard deviation, or indeed any calculation done on just the ages. But the second set doesn't look homogeneous — the closed circles scatter too much in the y direction. The chi-square statistic is 21.1 with 5 degrees of freedom, giving a P-value of less than 0.001 and strong evidence of heterogeneity.



THERE ARE two messages here. One is that the precisions carry information and methods that ignore this are wrong in principle. Any calculation using just the ages would give the same answer for both sets of data. But in the second set, the precisions are higher; there is more information. Indeed we have counted more tracks!

The other message is the fallacy of assuming that a hypothesis is true just because data are consistent with it. We may have data something like the first set, but if we had counted bigger areas per grain we might have got something like the

second set. Of course this fallacy is inherent in all scientific inference and is well known, but it is still a common source of misjudgement.

SOMETIMES there are many grains with very few or even no spontaneous tracks. What about these? Can we put them on a radial plot? What happens to the chi-square test? Who wants to know? For answers to these and other mysteries we must visit Alaska, in the next chapter.

Sources

Fisher, R.A. (1922) On the interpretation of chi-square from contingency tables, and the calculation of P. *J. Roy. Stat. Soc.* **85,** 78–94.

Franklin, Benjamin (1748) "Time is money" *Advice to a Young Tradesman.*

Franklin, Benjamin (1789) "There are only two certainties: death and taxes." *Letter to Jean Baptiste Le Roy.*

Galbraith, J.K. (1977) *The Age of Uncertainty,* BBC television series and book.

Galbraith, R.F. (1988) Graphical display of estimates having differing standard errors. *Technometrics,* **30,** 271–281.

Galbraith, R.F. (1990) The radial plot: graphical assessment of spread in ages. *Nucl. Tracks Radiat. Meas.,* **17,** 207–214.

Green, Paul (1986) Fission track data from the Otway Ranges.

Laslett, Geoff (1991) Personal Communication.

Leander, E.K. and Finney, D.J. (1957) An extension of the use of the χ^2-test. *J. Roy. Statist. Soc.,* C, **5,** 132–136.

Pearson, K. (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag. Ser. 5* **50,** 157–172.

[Are you sure about this title? Ed.]

Sterne, Laurence (1759) *The Life and Opinions of Tristram Shandy, Gentleman.*

The Times (ca 1980) News item about Chi-Chi.

Villiers, George (1671) "What the devil does the plot signify except to bring in fine things?" *The Rehearsal.*

West, Mae (1928) "Come up and see me sometime." *Diamond Lil.*

First published in *On Track* in November 1991, by *Tracker,* Your man at the microscope with the tinted glasses and the geological watch (it shows the time in Ma, digital version only $2).

A Message from Alaska

"ONCE UPON A TIME there were no bears at all."

Our old school teacher would tell us that story when he was in a hurry. Not even Chi-Chi I would sometimes think, which at least made the story a bit longer. But why did they think there were no bears? Because there were no tracks? That might just be because there was no snow.

No fossil tracks

LIKEWISE if there are no fossil tracks, it might be because there are no fossils — the sample is young — or because there isn't much *uranium*. A common problem these days is how to estimate a single grain age when $N_s = 0$. For EDM data, the conventional formula estimates ρ_s/ρ_i as N_s/N_i with relative standard error

$$\sqrt{\frac{1}{N_s} + \frac{1}{N_i}}\,.$$

If we blindly put $N_s = 0$ in these formulae we get an age estimate of zero with infinite relative standard error, a result that is neither useful nor

sensible. These formulae are fine when applied to reasonably large counts but they break down when $N_s = 0$, and are not too good for very small non-zero counts either.

So several people have asked: "How do we assign an error to a zero age?" and "How can we put zero ages on a radial plot?" With inimitable charm (and pedantry) I usually start by pointing out that the phrase "zero age" is misconceived; for when $N_s = 0$, it does not mean that the true age is zero, or even that it is necessarily sensible to estimate it as zero — indeed it is easy to find a better estimate than zero. So let's talk about zero counts, not zero ages.

THE ESSENCE of the problem is easily seen. Consider two grains, both having $N_s = 0$ (no spontaneous tracks) but one has $N_i = 5$ and the other $N_i = 50$. With the $(0, 5)$ grain we might think that either it is young or else it might be quite old and just have a low uranium content. With the $(0, 50)$ grain though, we would be more inclined to think it was young. The two grains do not carry the same information. It follows that any method that treats them as if they do is wrong in principle. This applies also to non-zero counts, especially small ones; it matters less with large counts.

Interval estimates

TO ANSWER the first question, we must change our concept: when $N_s = 0$ it is not sensible to quote a point estimate with an associated standard error. Instead we can calculate a *confidence interval* — a range of values (ages) that the observed data are consistent with. ("He who has never ended a sentence with a preposition has never tried to explain what a confidence interval is" — New Zealand proverb). When $N_s = 0$, a 95% confidence interval for ρ_s/ρ_i goes from 0 to

$$20^{1/(N_i+1)} - 1$$

which can be substituted into the age equation to get a 95% confidence interval for the age.

So the idea is to calculate an upper 95% confidence limit for ρ_s/ρ_i and hence for the single grain age. When $N_s = 0$ the upper x% confidence limit is

$$\left(\frac{100}{100-x}\right)^{1/(N_i+1)} - 1\,.$$

To get a 95% upper age, put $x = 95$ in this formula and substitute the result in place of ρ_s/ρ_i in the age equation. Of course the age interval will depend on ζ and ρ_d also. Note that this formula does the right thing: it depends on N_i and it gets closer to zero as

Upper 95% confidence limits when $N_s = 0$
Ages are in Ma using $\zeta = 355$ and $\rho_d = 1.5 \times 10^6$ cm^{-2}

N_s	N_i	ρ_s/ρ_i	Age	N_s	N_i	ρ_s/ρ_i	Age
0	1	3.4721	863.9	0	40	0.0758	20.2
0	2	1.7144	441.0	0	50	0.0605	16.1
0	3	1.1147	290.2	0	60	0.0503	13.4
0	4	0.8206	214.9	0	80	0.0377	10.0
0	5	0.6475	170.1	0	100	0.0301	8.0
0	6	0.5341	140.7	0	120	0.0251	6.7
0	7	0.4542	119.8	0	150	0.0200	5.3
0	8	0.3950	104.3	0	200	0.0150	4.0
0	9	0.3493	92.3	0	300	0.0100	2.7
0	10	0.3130	82.8	0	500	0.0060	1.6
0	15	0.2059	54.6	0	1000	0.0030	0.8
0	20	0.1533	40.7				
0	25	0.1221	32.4				
0	30	0.1015	27.0				
0	35	0.0868	23.1				

N_i increases. If N_i is very large and $N_s = 0$ we really would think that the true age was close to zero.

THE TABLE opposite shows values of this formula for $x = 95$ and varying N_i. Such limits could be quoted in routine practice.

For illustration I have also shown the age limits you would get using $\zeta = 355$ and $\rho_d = 1.5 \times 10^6$. For example, the 95% interval for a $(0, 5)$ grain is 0–170 Ma while for a $(0, 50)$ grain it is 0–16 Ma. Quite a difference! (Go on, get out your calculator and check it: just put 0.6745 in place of ρ_s/ρ_i and similarly 0.0606 ... don't tell me, you've forgotten the age equation.) You can interpolate in this table if necessary. Put it on your wall next to Madonna or in the space where Fergie used to be.

BY THE WAY, these confidence intervals do not allow for error in estimating ζ and ρ_d. They are appropriate for comparing EDM age estimates for grains that were irradiated together. When calculating the standard error for the *pooled* age, the error in estimating ζ and ρ_d would normally be included. I have not yet seen data where the total N_s for all grains is zero, but no doubt it will happen one day.

Binomial and Poisson probabilities

WHERE does the above formula come from?

Consider the following statistical problem: suppose that in a large population a proportion θ of individuals have some attribute. A random sample of n individuals is chosen and r of them have the attribute. We wish to estimate θ, in particular when $r = 0$.

The observation $r = 0$ is consistent with the hypothesis that $\theta = 0$. It is also consistent with θ being small and non-zero. The probability of observing $r = 0$ is $(1 - \theta)^n$, which decreases as θ increases — and the value of θ for which this probability equals 0.05, say, gives an upper 95% confidence limit for θ. Solving the equation

$$(1 - \theta)^n = 0.05$$

gives

$$\theta = 1 - 0.05^{1/n}$$

or equivalently

$$\frac{\theta}{1 - \theta} = 20^{1/n} - 1 \,.$$

Actually this value of θ is known to be conservative in the sense that the interval has a slightly greater chance than 0.95 of including the true θ, and a more accurate result is obtained by using $n + 1$ in

place of n. In statistical jargon, the formula with $n+1$ is the Bayes estimate for a uniform prior.

WHAT has all of this got to do with N_s and N_i?

Imagine the $n = N_s + N_i$ tracks to be a sample from a population in which a proportion θ are spontaneous and the others are induced, where $\theta = \rho_s/(\rho_s + \rho_i)$. The observed number of spontaneous tracks is $r = N_s$. We want to estimate ρ_s/ρ_i in the age equation, which is the same as $\theta/(1-\theta)$. This is just what we have done above. Although this analogy seems contrived, it is in fact legitimate; it follows from the fact that N_s and N_i have independent Poisson distributions. As Goethe remarked, mathematicians are like Frenchmen: whatever you say to them they translate into their own language and forthwith it is something entirely different.

YOU CAN calculate "exact" confidence intervals from small non-zero counts also, though the formula, which uses the Binomial probability of observing N_s spontaneous tracks gets more complicated as N_s increases. I did this in Table 2 of my radial plot article but did not explain the method. Having read the above you may be pleased that I didn't.

Arcsin transformation

WHAT about putting zero counts on a radial plot?

I said earlier that when $N_s = 0$ it is not sensible to calculate a point estimate and standard error for the age. But we can just about calculate an estimate and standard error for something else, namely ν, where

$$\nu = \arctan \sqrt{\rho_s / \rho_i}\,.$$

If we know ν we can find $\rho_s/\rho_i = \tan^2 \nu$ and put this in the age equation. So we estimate the arctan square root of ρ_s/ρ_i, though this is commonly known as the arcsin transformation and written in the equivalent form

$$\nu = \arcsin \sqrt{\rho_s / (\rho_s + \rho_i)}\,.$$

Sir David Cox more stylishly calls it the angular transformation. Perfectionists may point out that it is not necessarily the best transformation, but it is quite good nevertheless.

Thus, for the radial plot we can use z, an estimate of ν, given by

$$z = \arctan \sqrt{\frac{N_s + 0.375}{N_i + 0.375}}$$

and its approximate standard error given by

$$\sigma = \frac{1}{2\sqrt{N_s + N_i + 0.5}}\,.$$

Of course z and σ must be calculated for each grain, not just those for which $N_s = 0$. These behave sensibly even when $N_s = 0$; both z and σ are positive and decrease as N_i increases. Moreover, even when $N_s = 0$, $\tan^2(z + 2\sigma)$ compares fairly well with the upper 95% limits in the above table. The comparison is a bit rough for a $(0, 1)$ grain, but there is very little information in a $(0, 1)$ grain. What about a $(0, 0)$ grain? Now there's a challenge — but Ian says that's not fission track dating!

ADDING 0.375 and 0.5 to the counts in the above formulae are refinements suggested by Frank Anscombe, who did for the arcsin transformation what J.B.S. Haldane did for the log odds transformation, and several years earlier apparently. Haldane was Professor of Biometry at UCL. I inherited his arm chair until it was taken over by mice. According to his colleague, Nobel Laureate Peter Medawar, Haldane was "in some ways the cleverest and in other ways the silliest man I have ever known". Nevertheless Haldane has a room named after him whereas Medawar has only a building. Anscombe worked at Rothamstead Experimental Station, so he probably visited UCL at least once.

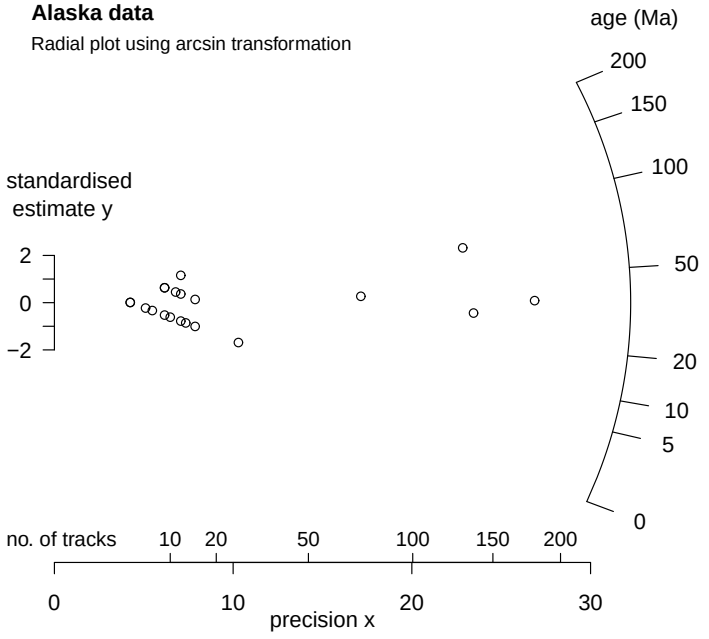
Chi-square test

THERE IS another point about small counts, which is that the chi-square test for homogeneity may need modification. When the counts become sufficiently small, the P-value obtained from the chi-square distribution ceases to be a good approximation. It is possible (but not easy) to calculate the exact P-value: you have to add up the multiple hypergeometric probabilities for all possible sets of data that have the same row and column totals as the actual data and which have a chi-square statistic larger than that for the actual data. There is statistical software available now (e.g., the StatExact package) that will do this.

A SIMPLER alternative is to pool grains to get larger counts. This has the benefit of increasing power also, but one must be objective about which grains to pool. It should be done on the basis of numbers of *induced* tracks, i.e., low uranium, not on the basis of low spontaneous track counts. Even a homogeneous sample may appear heterogeneous when counts from grains with like ages are pooled together.

Alaska at last

THE GRAPH shows a radial plot of some data from Alaska that are very instructive. Out of twenty grains there are ten with 0 spontaneous tracks, five with 1 spontaneous track and one each with 2, 6, 9, 14 and 18. The numbers of induced tracks vary from 4 to 166. The ages look homogeneous. The chi-square statistic is 17.6 with 19 degrees of freedom, giving a nominal P-value of 0.55, which seems to suggest no evidence of heterogeneity. The



zero count grains are easily recognised: they lie on a downward sloping line. Both the estimate and the upper two-sigma confidence limit decrease as N_i increases in accordance with good sense. This is what the "exact" formula does too, and what the usual "large count" formula does not do.

THE FIRST thing to say is that because many of these counts are so small, the P-value of 0.55 is unreliable. But even if we calculated an "exact" P-value, I don't think it would indicate any heterogeneity. Secondly, and more importantly perhaps, because there are so many small counts the chi-square test will not be very powerful. Suppose we ignore the grains with fewer than 20 induced tracks and look at just the remaining five. These *by themselves* don't look quite so homogeneous. The chi-square statistic for just them is 7.9 with 4 degrees of freedom and a P-value of 0.098. Not strong evidence of heterogeneity, but more suggestive than before.

A more powerful test is to pool the 16 grains with small numbers of induced tracks and use the total N_s and N_i from these. This ensures that all expected frequencies are at least 5, the well known rule of thumb learnt by first year statistics students. There is a *prima facie* case for doing this as these 16 grains clearly agree with a common age. This now

gives a chi-square statistic of 10.39 with 4 degrees of freedom and a P-value of 0.044, indicating there is some evidence of heterogeneity after all.

BUT the main message is that there is not much information about whether or by how much the true ages vary. Such small counts are simply not very informative. So it would be foolish to assume that there was no variation just because these data look homogeneous. In fact more evidence was obtained subsequently from this sample: more grains were dated and they really did show a spread of ages.

IT IS of interest too, not only to look for evidence of heterogeneity, but also to *estimate* how much spread there is. For these Alaska data, the true ages could easily vary by 30% or even 40% of their central value, even though they look homogeneous. How can we estimate the central age and age dispersion? As Uncle Toby said, that's another story.

Sources

Anscombe, F.J. (1948) Transformation of Poisson, binomial and negative binomial data *Biometrika,* **35,** 246–254.

Duddy, I.R. (Ian) (1989) Fission track data from Alaska.

Galbraith, R.F. (1990) The radial plot: graphical assessment of spread in ages. *Nucl. Tracks Radiat. Meas.,* **17,** 207–214.

Goethe, J.W. (ca 1800) "Mathematicians are like Frenchmen" quoted in *The World of Mathematics,* v. 3, p.1832 by James R. Newman, 1956.

Haldane, J.B.S. (1956) The estimation and significance of the logarithm of a ratio of frequencies. *Ann. Eugenics,* **20,** 309–311.

Laslett, Geoff (1991) Personal Communication.

Medawar, Peter (1987) *Memoirs of a Thinking Radish.*

Sterne, Laurence (1759) *The Life and Opinions of Tristram Shandy, Gentleman.*

First published in *On Track* in May 1992, by *Tracker,* Your man at the microscope with the healthy apatite and the prismatic face.

Middle Age Spread

When Noah's Ark finally reached land, back in the mid-Holocene, there was a lengthy disembarkation procedure. The animals lined up, two by two, for Noah to complete their customs, immigration and quarantine documentation, record personal statistics and tell them to "Go forth and multiply". Presently there came two lowly creatures, appearing from under the coiled rope (they have one on every ship's deck, for practising the art of rope coiling) who said "We can't multiply, we are only adders."

God's instructions on this delicate matter, as on many others, had been somewhat lacking in detail and there was little that Noah could do other than to keep them waiting under his table until he had finished dealing with the other animals. Naturally this took some time, though the formalities were fairly straightforward in most cases. There was a problem with the giant pandas, who didn't seem to know which was which (their names didn't help) and who's future seemed uncertain. Eventually Noah sent them away, trusting to faith, and, wondering how far faith would take him, looked

under his table for the two reptiles. They were both still there, along with numerous little ones, all looking very pleased with themselves. "I thought you said you were only adders and could not multiply" said Noah. "Ah yes," they replied "but this is a log table."

Napier's invention

THAT AWFUL story was told to us long ago by our school teacher in a desperate attempt to interest us in logarithms. I swear I have not repeated it until now and, as the average reader of *On Track* has probably never used a log table — at least for multiplication — there is little chance that it will go further. We had to use common logs, the only ones readily available, though we knew that John Napier in 1614 had suggested replacing 10, 100, 1000, . . . by 2.71828, 7.38906, 20.08554, . . . , which he said was natural.

Notwithstanding this obvious insight, Henry Briggs ten years later published the first great log table: logarithms to base 10, to ten decimal places, of all the numbers from 1 to 100,000. Exactly three hundred years later, A.J. Thompson calculated them to twenty decimal places with the aid of a computer — who, as was customary, was female and nameless. This was essentially an act of

chauvinism: the French had threatened to publish fourteen decimal places and the Germans fifteen. They are now published in two volumes by the Department of Statistical Science, UCL, under the title *Logarithmetica Britannica*, price £20, o.n.o. Their weight is ideal for flattening posters, door stops, etc., and together with the three-and-a-half foot high statue of Napier himself behind the common room door, they are an inspiring reminder of the single most useful advance in mathematics since "before Noah was a sailor".

Manuel's law

THEY ALSO inspired other tables: an industry that produced tables for the distribution of the correlation coefficient, for testing the significance in a 2×2 table, symmetric functions and allied tables (a popular title), beta functions and gamma functions which are incomplete. But it was Manuel I of Byzantium (1143–1180) who inspired the most wonderful result. Nine coins from his first coinage had the following silver contents, in parts per hundred:

5.9, 6.8, 6.4, 7.0, 6.6, 7.7, 7.2, 6.9 and 6.2.

You can see the coins in the Fitzwilliam Museum, Cambridge, UK, provided you promise not to

subject them to fission track analysis. The mean and standard deviation of these measurements are 6.74 and 0.543 pph, so the *relative* standard deviation is $0.543/6.74 = 0.0806$. Now if we find the *logarithm* (the natural log) of each measurement and calculate the ordinary standard deviation of these, it comes to 0.0806. The same!

AMAZING though this is, it is apparently no accident: here are the silver contents of seven coins from Manuel's fourth coinage:

5.3, 5.6, 5.1, 6.2, 5.5, 5.8 and 5.8.

These have a relative standard deviation of 0.0646 and the standard deviation of their logarithms is 0.0644. Practically the same! Moreover the mint precision has improved from 8.1% to below 6.5%, an achievement so commendable that one scarcely notices that the mean silver content has dropped from 6.74 to 5.61 pph. The Byzantine government, as well as originating and establishing a standard monetary unit, was apparently well versed in other modern fiscal practices, including printing money when needed and surreptitious devaluation — here by 17%, presumably without a formal announcement from the Bank of Constantinople.

BUT the really interesting thing about Manuel's law is that it has nothing whatever to do with Byzantine coins. It simply states that the relative standard deviation of any set of measurements approximately equals the standard deviation of their logarithms. Natural logs of course.

To prove it, just take any set of numbers you like (any ten numbers between 80 and 120, say) and calculate their relative standard deviation and the standard deviation of their logs. Hey presto! Practically the same result. Not *exactly* the same but close enough provided that the numbers are positive (of course) and their relative standard deviation is "small". Actually the word "small' is misleading here: the agreement is excellent even for a relative standard deviation as high as 0.50 (i.e., 50%). In practice, one's relative standard deviations are usually well below 50%, so the approximation works where we want it to work. (Adherents of Murphy's law may notice an apparent contradiction, but the general form of Murphy's law implies that Murphy's law itself can go wrong.)

It won't have escaped you that the great value of Manuel's law is not so much in the calculation (few would regard calculating the standard deviation of the logs the easier option) but in the *concept*. That is the real power of mathematics.

Central ages

WHAT IS a central age and how is it calculated?

Imagine a population of a large number of crystal grains. Imagine that each grain has a unique true value of ρ_s/ρ_i, the ratio of its spontaneous to induced track density. For example, imagine that ρ_s could be measured exactly, without error, for each and every grain, that they could then all be irradiated together and ρ_i measured exactly in an external detector for each grain — and hence each ratio ρ_s/ρ_i could be determined exactly. We don't have to know what all these values are, but just that they exist in principle.

Now let μ and σ be the mean and standard deviation of their logarithms. (Why? Because that's natural.) Then the central age is the age corresponding to μ. That is, the age obtained by putting ρ_s/ρ_i equal to $\exp(\mu)$ in the fission track age equation. We call it the central age because it is in the centre of a distribution of ages. Pedants might prefer to say that its logarithm is in the centre of a distribution of log-ages, but *anti-log central log-age* would never catch on, and in any case it is really *the age corresponding to the anti-log central log spontaneous to induced track density ratio*. Moreover, by Manuel's law, σ, being the standard deviation of $\log(\rho_s/\rho_i)$, is the *relative* standard deviation of

ρ_s/ρ_i. And because ρ_s/ρ_i is effectively proportional to the age, σ is the relative standard deviation of the population ages. We call σ the age dispersion. The central age and age dispersion *together* summarise the distribution of population ages.

IN PRACTICE, we just have a sample of grains from the population. If we could determine ρ_s/ρ_i exactly for each sample grain, we could then *estimate* μ and σ (e.g., from the sample mean and standard deviation of their logarithms). But we cannot determine ρ_s/ρ_i exactly, even for a sample grain; we can only *estimate* it from track counts. Nevertheless, we can still estimate μ and σ, though to do so is more comp- licated because (a) there is an extra component of variation from the track counts (Poisson variation) and (b) this component differs between grains.

An efficient method (i.e., one that uses the data well) is to assume that the population values of $\log(\rho_s/\rho_i)$ follow a Normal distribution, and to use the method of maximum likelihood. It is not possible to write down the estimates explicitly. A simple algorithm is available on request and is included in a written (and submitted) version of Galbraith and Laslett (1992). The estimate of σ comes from the "extra-Poisson" variation — i.e., over and above the Poisson variation.

To understand the estimate of μ, note that if σ is large and dominates the Poisson variation, then each grain is weighted equally, while if σ is small, the Poisson variation dominates and μ is estimated from the pooled track counts. In general, the estimate of μ lies between these two values. Also, the precisions of the estimates of μ and σ will depend not only on the number of sample grains, but also on how well each ρ_s/ρ_i is estimated — i.e., on the track counts.

Rufford well

When might the above model (for it is a model) be relevant?

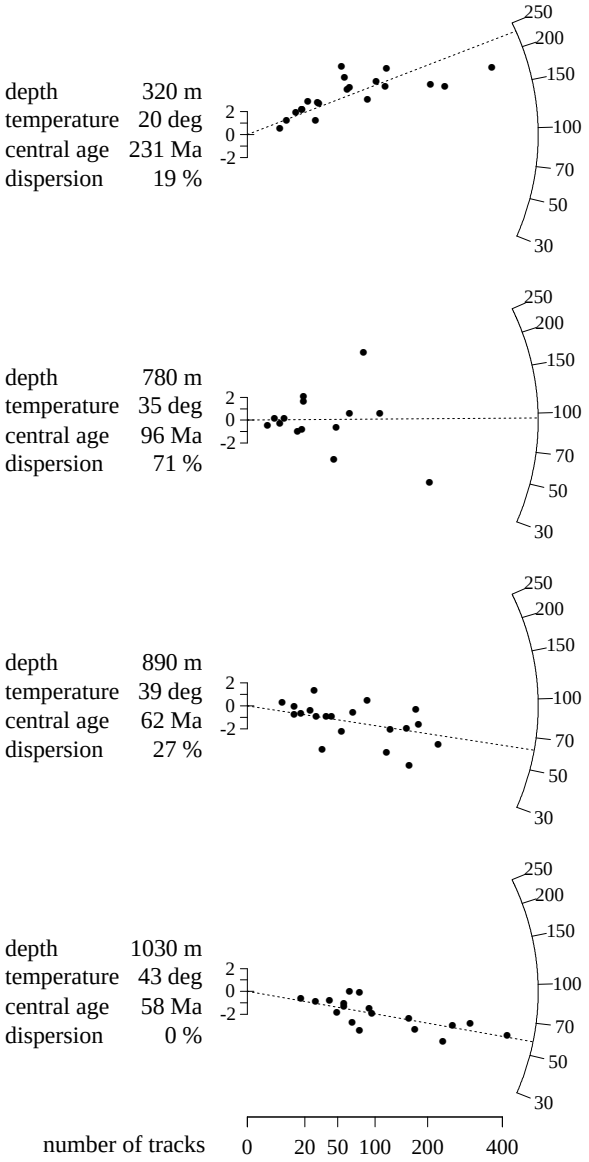
Suppose a number of contemporaneous apatite grains experienced a common thermal history severe enough to shorten tracks. And suppose that each grain had a different chlorine content, so that tracks would shorten by different amounts in different grains. Then the true values of ρ_s/ρ_i would vary between grains. Consequently, the true "ages" (as given by the age equation) would vary between grains. The age dispersion tells us how much they would vary, and the central age tells us the central age. For example, if the age dispersion was 0.20 (20%) and the central age was 100 Ma, then about 95% of the population ages would be in

the range

$$\exp(\log 100 \pm 2 \times 0.20) = 100\, e^{\pm 40}$$

i.e., between 67 and 149 Ma.

THIS IS just one scenario and there may be others for which the model is applicable. Usually we don't know much about how chlorine contents vary — their distribution may be very irregular — or about the precise effect of amount of chlorine on track length reduction at various temperatures. So to assume that $\log(\rho_s/\rho_i)$ is Normal is something of an act of faith. Nevertheless the parameters μ and σ can provide a useful description of a sample that can serve as a summary and as a basis for comparison with other samples.


depth 320 m
temperature 20 deg
central age 231 Ma
dispersion 19 %
depth 780 m
temperature 35 deg
central age 96 Ma
dispersion 71 %
depth 890 m
temperature 39 deg
central age 62 Ma
dispersion 27 %
depth 1030 m
temperature 43 deg
central age 58 Ma
dispersion 0 %
2
0
-2
250
200
150
100
70
50
30
number of tracks
0 20 50 100 200 400

THE GRAPH opposite shows radial plots of fission track ages for four samples from different depths in a well. The top sample, at depth 320 m and a current temperature of 20°C, has an estimated central age of 231 Ma and an age dispersion of 19%. There is evidence of a spread, suggesting that tracks have been shortened more in some grains than in others.

In the next sample, at depth 780 m and 35°C, some grains have ages comparable with the top sample, while others are much younger. Their distribution has a lower central age of 96 Ma and a higher dispersion of 71%. In the third sample, at depth 890 m and 39°C, many more of the ages are younger; the central age is now down to 62 Ma and the dispersion has gone down again to 27%.

In the bottom sample, at depth 1030 m and temperature 43°C, the single grain ages are consistent with each other and with the central age of 58 Ma. The age dispersion is now down to zero. (When the dispersion is zero, the central age estimate coincides with the usual pooled age.) The inference here is that there has been sufficient heat to anneal completely all tracks formed before 58 Ma, since there is no evidence of tracks being shortened by differing amounts, and that little or no shortening has taken place since that time.

Of course, such data should be viewed along with track length measurements and other relevant context information. For the record, the numbers of confined tracks measured, along with their means and standard deviations in cm were, from top to bottom (95, 11.63, 2.03), (19, 11.02, 2.93), (94, 12.62, 2.03) and (81, 12.17, 1.81).

This pattern of the age dispersion increasing and decreasing again as the central age decreases is quite common and has, I am sorry to say, been referred to as *Middle Age Spread*. It generally occurs, to differing degrees, in apatite samples that are in a sequence of increasing temperature — down a borehole of over an area — and may be a useful indicator of such a sequence.

To see the pattern on a graph, an important feature is that the different samples are plotted with respect to data frames having the same age scale, centred at the same place, and with the same x and y scales. Edward Tufte calls this type of graphical design "small multiples" and says "constancy of design puts emphasis on changes of data, not changes in data frames."

Such designs are useful for comparing data between samples, as opposed to within a sample, and are often a better alternative to superimposing

several samples on the same data frame. So we are back to multiplication.

Sources

Briggs, H. (1624) *Arithmetica Logarithmica.*

Galbraith, R.F. and Laslett, G.M. (1992) Statistical models for mixed ages. *Presented at the 7th International Workshop in Fission Track Thermochronology, Philadelphia, 13–17 July, 1992.*

Green, Paul (1989) Fission track data from the Rufford Well.

Hurford, Tony (1992) Middle Age Spread.

Laslett, Geoff (1992) Note from a school of adders: there is no collective noun for adders in general although, curiously, there are collective nouns for various types of adder:

Type of adder	Collective noun
Puff adder	Smoking room
Death adder	Valley
Common adder	House
Black adder	Series
Horned adder	Motorway
Chancellor of the Exchequer	Lamontation

That's enough adders Geoff.

Napier, J. (1614) *Mirifici logarithmorum canonis descripto.*

Pearson, K. (1922) Tables of the Incomplete Gamma Function. *Biometrika Trust, University College London.*

Pearson E.S. and Johnson, N. (1934) Tables of the Incomplete Beta Function. *Biometrika Trust, University College London.*

Shakespeare, William (1602) *Twelfth Night, III, ii.*

> FABIAN: I will prove it legitimate, sir, upon the oaths of judgement and reason.
> SIR TOBY: They have been grand jury men since before Noah was a sailor.

Thompson, A.J. (1924) *Logarithmetica Britannica.*

Tufte, E.R. (1990) *Envisioning Information.* Graphics press, Cheshire, Ct, USA.

First published in *On Track* in November 1992, by *Tracker,* Your man at the microscope with the geological time and inclination.

Trials and Errors

RECENT DISCLOSURES concerning bears and Byzantine coins have prompted several letters from readers:

> Grand Canyon Hotel, Denver, Colorado.
>
> Dear Tracker,
>
> My wife and I have been trying to reduce our errors for ages, but we find that the older we get, the bigger they become. Are we Normal?
>
> Yours, CW.
>
> P.S. Nancy says that our *relative* errors have gone down a bit, though.

> Oasis Bar, Surfers' Beach, California.
>
> Dear Tracker,
>
> I have discovered that I can reduce my standard errors by dating younger things. Unfortunately, this seems to lead to increased *relative* errors. Yesterday I dated a groovy piece that had an age of 0.0000030 Ma with relative standard error 4000%. Is this a record?
>
> Yours, Trevor.

Windsor Castle, Windsor, Victoria.

Dear Tracker,

In your otherwise admirable explanation of Manuel's law (*On Track, 2, ii*) it can't have escaped your attention that there is one feature that does not ring true. Not only was the silver content reduced in the later minting, but also the relative standard deviation went down. As every analyst knows, when the true value goes down, the variation also goes down, but the relative variation tends to go up. What is going on? I think we should be told.

Carry on, G.

The Hat Shop, Grizzly Mountain, Alaska.

Sir,

In a recent issue of your tracking magazine, you allude to bears that leave no tracks in the absence of snow, a conjecture whose authenticity is questionable to say the least. As one with no little experience of the heavier thick-furred plantigrade quadrupeds, I submit that evidence of their proximity can usually be inferred from a rudimentary knowledge of the geological environment in a variety of climatic

conditions. To quote a bard from an earlier age:

> Farewell!
> The day frowns more and more:
> thou art like to have
> A lullaby too rough. I never saw
> The heavens so dim by day. A
> savage clamour!
> Well may I get aboard! This is
> the chase:
> I am gone forever.
> *[Exit, pursued by a bear.]*

Furthermore, the giant panda is never seen in these parts: what sort of tracker are you?

I remain, yours, etc., D. Crockett.

THE LAST CORRESPONDENT is clearly receiving *On Track* in error, probably as a free offer, and is best ignored. The first three, though, have put their collective finger on a useful point. A relative standard error, being the ratio of a standard error to the value of the quantity being estimated, will tend, other things being equal, to be small when the true quantity is large and large when the true quantity is small. Of course, this is not the full story. In some experimental procedures, the absolute standard error is naturally dependent on

the value of the quantity being estimated, while the relative standard error is not, or not strongly so at least. This is the case for fission track ages, as observed by both CW and Trevor.

A fiscal mystery

OUR correspondent G is referring to the following data from the reign of Manual I:

Silver content x (pph)			
	mean	s.d.	s.d./mean
First coinage	6.74	0.543	0.081
Fourth coinage	5.61	0.363	0.065

He points out that while the mean and standard deviation both went down in the later coinage, so did the relative standard deviation, from 8.1% to 6.5%.

How did this happen? My theory is this: Manuel needed to produce more coins but did not wish to use more silver than before, so what did he do? He added more alloy. Let us look not at the silver content of each coin, but at its *alloy* content — the number of parts of alloy per part of silver:

Alloy content $100/x$			
	mean	s.d.	s.d./mean
First coinage	14.91	1.20	0.081
Fourth coinage	17.88	1.15	0.064

Now in the later coinage, the mean alloy content has gone up, the s.d. is much the same (surely evidence that we are on track) and the relative standard deviation has gone down, in accordance with G's desire.

INDEED, the two relative standard deviations are practically the same as before — and a moment's thought tells you that this must be so: by Manuel's law, the relative s.d. equals (to a close approximation) the s.d. of the logs; and furthermore, because $\log(100/x) = \log(100) - \log(x)$, the s.d. of $\log(100/x)$ must equal the s.d. of $\log(x)$. Manuel's law is powerful indeed!

A Poisson situation

BUT let us take a simpler situation: estimating a track density. Consider a count N that is a random quantity with expected value $A\rho$, where A is known (for example, a known area of crystal surface) and ρ is an unknown density to be estimated. Suppose N has a Poisson distribution. (Today we will leave aside the question of why we suppose this — contrary to popular opinion, it has little to do with deaths from horse kicks.) Then the variance of N is $A\rho$, equal to the mean. The standard deviation is therefore $\sqrt{A\rho}$ and the relative standard deviation (or coefficient of

variation) is

$$\sqrt{A\rho}/A\rho = 1/\sqrt{A\rho}\,.$$

Suppose we observe $N = r$ and we wish to estimate ρ. The natural estimate is

$$\hat{\rho} = r/A \tag{1}$$

which has standard error

$$\mathrm{se}(\hat{\rho}) = \sqrt{A\rho}/A = \sqrt{\rho/A} \tag{2}$$

and relative standard error

$$\mathrm{se}(\hat{\rho})/\rho = 1/\sqrt{A\rho}\,. \tag{3}$$

Notice that if the density ρ is fixed and the area A is large then both the standard error (2) and the relative standard error (3) will be small. If the expected count $A\rho$ is large then the relative standard error (3) will be small, though the absolute standard error (2) may not be. However, if ρ is small the standard error (2) will be small but the relative standard error (3) will be large — simply because it is relative to something small. To see how good your estimate is, you may need to express its precision more directly, possibly as a confidence interval.

An approximate standard error

NOW equations (2) and (3) depend on ρ which is unknown, so we can't calculate them. (And if we knew ρ we wouldn't want to calculate them as we wouldn't need to estimate ρ, an irony familiar to statistics students long before Yossarian's famous Catch 22.) In order to get an idea of the precision it is standard practice to substitute the estimate r/A to get approximate values

$$\mathrm{se}(\hat{\rho}) \approx \sqrt{r}/A \tag{4}$$

and relative standard error

$$\mathrm{se}(\hat{\rho})/\rho \approx 1/\sqrt{r}\,. \tag{5}$$

REMEMBER that (4) and (5) are approximate. They are likely to be poor approximations when r is small, and effectively nonsense when $r = 0$. It can easily happen (and does) that $r = 0$ even though ρ is not zero.

But there is another aspect to consider when r is small. It is not only that (4) and (5) are poor approximations, true though this is, but also (2) and (3) are themselves poor measures of precision in this circumstance. To understand why, we need to think about standard errors.

A probability distribution of possible estimates

THE STANDARD ERROR $\text{se}(\hat{\rho})$ is a measure of how imprecisely the parameter ρ has been estimated. It is the standard deviation of all of the possible estimates that the experiment might have produced. These estimates have a probability distribution which is derived from the Poisson distribution — i.e., the distribution of N/A, where A is fixed and N has a Poisson distribution. (In statistics jargon, this probability distribution is called the *sampling distribution* of $\hat{\rho}$ because it is the distribution from which $\hat{\rho}$ has been sampled.)

If the standard error is small, the probability distribution of estimates is closely concentrated around the true value; but the usefulness of the standard error as a measure of precision also depends on the shape of this distribution.

If $A\rho$ is large enough, the distribution of N/A is approximately Normal with mean ρ and standard deviation $\text{se}(\hat{\rho}) = \sqrt{\rho/A}$, as given by (2). Then, for example, $\hat{\rho} \pm 2\text{se}(\hat{\rho})$ is an approximate 95% confidence interval for ρ.

But if $A\rho$ is not large, the distribution of possible estimates is *not* approximately Normal. Then $\text{se}(\hat{\rho})$ is not such a useful measure of precision, and $\hat{\rho} \pm 2\text{se}(\hat{\rho})$ does not provide an approximate 95% confidence interval. This applies to (2), and therefore to (3), when r is small. There

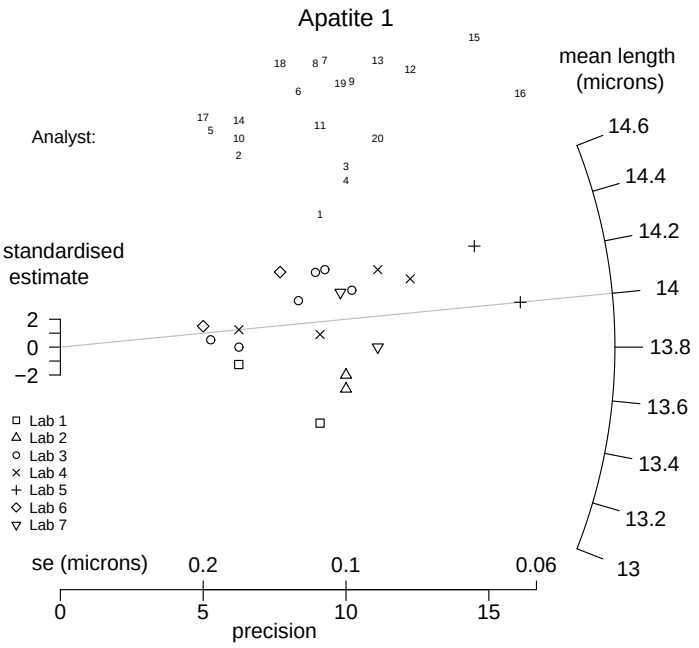
are better things to do in these circumstances, such as transforming to a different scale or using confidence intervals directly.

A mean length

So much for errors; what about trials?

Before the 1992 Philadelphia Workshop, aliquots of two apatite samples were sent to several labs for analysis, both for dating and for making track length measurements; and in several of the labs, the apatites were analysed by more than one analyst. For the present purpose, let us regard this simply as an exercise in estimating the *mean length* of tracks in the original apatite samples. And let us suppose that everyone is trying to estimate the same thing. In practice, people measured lengths of horizontal confined tracks according to their normal procedure, uncorrected for length-biased sampling.

The graph overleaf is a radial plot of the mean length estimates for Apatite 1, made by 20 analysts from 7 labs. For each analyst, the estimate is the mean of the length measurements made, and its standard error is their standard deviation divided by the square root of the number of tracks measured. Thus if an analyst measured the lengths of 100 tracks with a mean of 14 microns and a



standard deviation of 1 micron, the standard error used on the radial plot would be $1/\sqrt{100} = 0.1$ microns, and the point would plot close to (slightly above and to the right of) that by Analyst 11. It is clear from the graph that the estimates are not all in agreement with respect to this measure of precision; there are other sources of variation.

Where might variation come from? You will easily tell me. For example, differences in etchant used, etch time and microscope conditions will

lead to differences between labs. Differences in the way tracks are selected as well as actual differences between aliquots, may result in length differences between analysts. Even two different analysts from the same lab measuring exactly the same tracks may get systematically different mean lengths. (One person I know claims to have better eyesight than his colleagues and can see the track ends better.) It would be possible to estimate the sizes of systematic and random differences from suitably designed studies — and this could be a worthwhile exercise.

Within and between

LOOKING AT the Apatite 1 graph more closely, you can see that differences between analysts from the same lab are generally not so great as those from different labs. The 'within-lab' variation is less than the 'between-lab' variation. In fact, within some labs (2, 6 and possibly 1 and 3) there is little evidence of any extra variation between analysts. Treating all variation as being random, I have estimated the between-lab standard deviation to be 0.35 microns, and the between analyst (within lab) standard deviation to be 0.19 microns. These represent extra components of variation: for example, mean lengths measured by different analysts in the same lab, each having a sub-sample

standard error of 0.1 microns, will vary with a standard deviation of

$$\sqrt{0.19^2 + 0.10^2} = 0.21 \text{ microns}$$

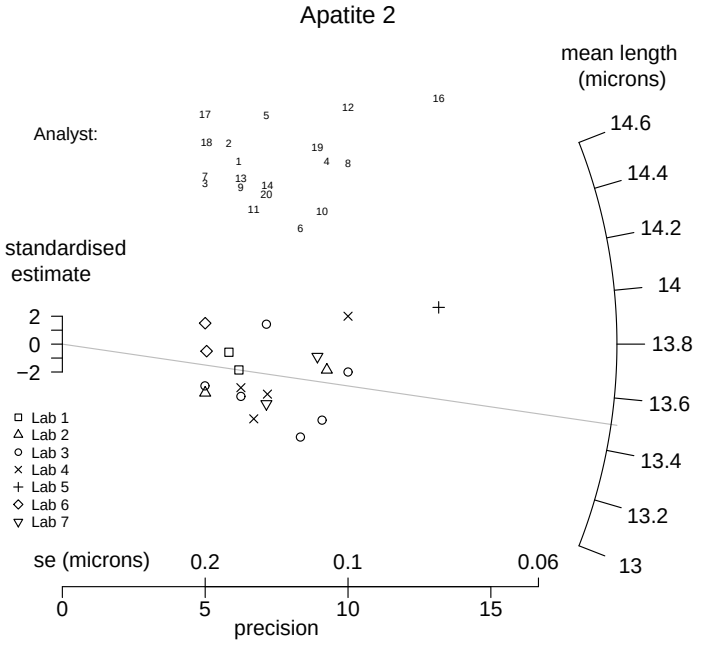
and for such analysts in different labs they would vary with a standard deviation of

$$\sqrt{0.35^2 + 0.10^2} = 0.41 \text{ microns}\,.$$

The practical importance of these extra components of variance will depend on the context. For example, if reduction in mean length is of interest, rather than the absolute mean length (as in some annealing studies) it will be important for the same analyst to measure all the samples to be compared, or to be able to combine measurements from different analysts without introducing systematic biases. Furthermore, differences between analysts in *reduction* in mean length might be much smaller than for absolute length. These things are unknown, but are relevant to studies that use data from different labs.

A second apatite

THE GRAPH below shows the estimates obtained by the same analysts for the second apatite sample (except for analyst 15 who didn't measure this sample).



Again there is extra variation between labs and some extra variation between analysts within labs. It is interesting to compare relative positions of the corresponding points for the two graphs. For both apatites, Analysts 17 and 18 (the two from Lab 6) are slightly higher than the overall average but not systematically different from each other. Analysts 19 and 20 (the two from Lab 7) are in similar relative positions, with 19 above the average and 20 below the average in both cases. Analysts 1 and 2

(from Lab 1) are in good agreement for both apatites, but somewhat lower than average for Apatite 2. Analysts 15 and 16 (both from Lab 5) quote the highest precisions.

There are other patterns, and exceptions also. It seems likely that there are both systematic and random effects between analysts and between labs.

An anova

THERE ARE two sorts of people. Ordinary folk like you and me, who derive simple pleasure from hydrothermal intrusions in a sedimentary basin, or an alpine fault in a parauthochthonous terrane. But to those other people, beauty is *analysis of variance*.

Here is what they would do with the data from the two graphs (leaving out Laboratory 5 for simplicity):

Source of variation	D.f.	Mean square
Apatites	1	2.8274
Labs	5	0.3765
Apatites × Labs	5	0.1794
Analysts (within labs)	12	0.0994
Residual	12	0.0706

Naturally d.f. stands for degrees of freedom and mean square stands for variance.

To read this table, start at the bottom. The residual mean square of 0.0706 represents the

variance between estimates having accounted for all systematic differences between apatites, labs and analysts. If there were no such differences, all the other mean squares would roughly equal this one. The residual standard deviation (the square root of 0.0706) is 0.27 microns. Actually, part of this is known (the part corresponding to the standard errors in the graphs) and if we take away this part (a less than trivial exercise) the remaining part comes to 0.21 microns. This represents the standard deviation of *random* differences between analysts (within labs) and agrees well with the figure of 0.19 microns found earlier.

The mean square for Analysts (within labs) is 0.0994, a bit larger than 0.0706. The extra bit is due to *systematic* differences between analysts in the same lab. The three other mean squares are even larger. The Apatites mean square of 2.8274 reflects the fact that the mean track lengths for the two apatites are different (14 microns for Apatite 1 and 13.5 microns for Apatite 2). If they were more different, it would be even larger. The Labs mean square of 0.3765 is also quite large, reflecting systematic differences between labs. And the Apatite $\times$ Labs mean square of 0.1794 reflects the fact that the differences in mean lengths between labs differ for the two apatites.

Beauty indeed!

A laboratory difference

IN THE ABSENCE of more information, the figure of about 0.2 microns for the extra within-lab standard deviation may have implications for current laboratory practice. Differences between labs, on the other hand, are usually relatively large and not so easily interpreted as being random. Naturally a laboratory will tend to trust its own results and discount others if they do not agree. As a laboratory track counter once remarked (in a slightly different context): "There are three sorts of laboratories: those that can count, and those that can't."

Sources

Bortkewitsch, L. von (1898) *Das Gesetz der kleinen Zahlen*, Teubner, Leipzig. (Data on numbers of deaths from horse kicks in the Prussian army.)

Heller, Joseph (1961) *Catch-22.*

Miller, D.S. et al. (1992) Results of interlaboratory comparison of fission track ages for the 1992 Fission Track Workshop. *Presented at the 7th International Workshop in Fission Track Thermochronology, Philadelphia, 13–17 July, 1992.*

Moore, D.S. (1989) "There are three sorts of mathematicians ..." *After dinner speech to the International Conference on Teaching Statistics, Dunedin, NZ.*

Shakespeare, William (c1608) *The Winter's Tale, III, iii.*

Tracker (1992) A Message from Alaska, *On Track, 2, i.*

Tracker (1992) Middle Age Spread, *On Track, 2, ii.*

First published in *On Track* in May 1993, by *Tracker,* Your man at the microscope with the Fish Canyon T-shirt and the Durango jeans.

A Negative Association

THE FOLLOWING letter arrived recently by express post:

> Dave's Hat Shop,
> Grizzly Mountain, Alaska,
> June, 1993.
>
> Dear Mr. Tracker,
>
> I admit that I have never fought the giant panda, but I do know something about hats. Your recent article Trials and Errors (*OT*, May 1993) apparently contains more Errors than intended. Practically every rho has a missing hat. Not only is this extremely untidy, as Ophelia would often complain:
>
> > "Lord Hamlet, with his doublet all unbraced;
> > No hat upon his head; his stockings fouled,
> > Ungart'red, and down-gyvéd to his ankle ..."
>
> but also it makes a complete nonsense of the whole hat industry, with which many of us have been associated from an early age. Furthermore, some rho's have gone Latin and your square roots are ambiguously truncated to say the least.
>
> I remain, yours etc.,
>
> D. Crockett.

A MAD HATTER it seems. But the honest fellow is perfectly correct. We use a Greek rho (ρ) to denote the real track density, and rho-hat (ρ with a $\hat{}$ on top) to denote the measurement, or estimate, of ρ. The Greeks naturally use the opposite convention, which accounts for the rarity of Greek statisticians. In the aforementioned article,[1] therefore, the *first* ρ in each equation (1) to (5), but not the others, should have a hat on. Also, a standard error refers to the estimate, not the true value, so when you see the expression se(ρ) or $\rho \pm 2\text{se}(\rho)$, the ρ should have a hat on. I leave you the exercise of retrieving your ancient issue of *On Track*, adding the appropriate hats, changing r to ρ in the first two of the three sentences before equation (4), and clarifying all square root signs. I regret the confusion and can only assume that our editor has Greek relations.
[τηατ ωασ α πρεϖιονσ εδιτορ, Ed.]

BUT WHY are we so fussy about hats? Why do we distinguish between the "true" value of a quantity and the value observed when we measure it? Because not to do so can cause misunderstanding and mistakes, as I was unsuccessfully trying to

[1] This refers to the original article in *On Track*, May 1993, which contained a number of typesetting errors. These have been corrected in the present volume.

explain. Plotting age against uranium is a good example.

Age and uranium

ANN BLYTHE, in the November issue of *On Track*, plotted fission track age against uranium concentration for three zircon samples, and found that older grains tended to have lower uranium concentration than younger grains. She conjectured a partial resetting of ages through metamictization of high-U zircons, but also noted a possible "counting bias" problem where old high-U grains (and young low-U grains) were more difficult to count. A similar negative association between age and U was noted by Andy Carter at Besançon, who also cited earlier references to it. He considered several possible causes, including variable annealing rates, poor thermalisation of the reactor, imperfect matching of detector, variable etching and counting bias.

Now, as AB and AC recognised, there will indeed tend to be a negative association between age and uranium within a sample due to selection of grains. Paul Green describes common practice for selecting zircons thus:

> "Start at one corner of the mount and traverse across looking at every grain and

> counting all (or enough) of those grains that fulfil the usual criteria (prismatic surface, well etched, etc.). The main criterion is that the track density should be countable — too high and you can't count it, too low and the tracks don't etch properly because the background alpha damage isn't sufficiently intense and the grain etches anisotropically. In practice, the band of acceptable track densities is probably around 5E5 to 5E6. I don't think anyone selects grains for uranium content, or takes account of grains with rho's that are too high to count. Chuck and Nancy published a paper some time ago (in *Nuclear Tracks* I think) suggesting different etch times to select grains of different track densities within one sample, but in practice I've never been convinced on this because when all is said and done the range of countable track densities is pretty narrow."

So a major limitation is that the spontaneous track density ρ_s should be in the acceptable range. Now ρ_s is roughly proportional to the product of age and U concentration, so if a grain is old, it stands to reason that it will need to have a lower U concentration in order to keep ρ_s to within the limits. And *vice versa*. Although I have expressed

this in terms of the underlying *true* track density, in practice grains will be selected on the basis of the *observed* counts per unit area. Thus for example, if uranium concentrations are high enough, the observed numbers of spontaneous tracks may vary sufficiently that some will be uncountable, even if the true ages are the same for each grain. And if the true ages also varied, we would expect the selection effect to be stronger.

IF YOU were trying to discover a real relationship between age and uranium, how could this bias be avoided? One way might be to select grains by looking at the *induced* tracks. Then you would sometimes select grains where spontaneous tracks are too dense to count (or inadequately etched). The spontaneous track counts for these grains would need to be treated as censored data — a statistical procedure much less exciting than it sounds. Nevertheless it could be done in principle.

Related measurements

BUT there is another, more subtle, statistical association that arises when you plot "age" against "uranium", even, indeed especially, when age and uranium are practically constant. What is really plotted is *measured* age against *measured* uranium

(i.e., age-hat against U-hat). The former is roughly proportional to the ratio n_s/n_i while the latter is proportional to n_i/A, where A is the area of crystal surface in which tracks are counted.

Even if the true ages and uranium contents were the same for each grain, the counts would differ because of their natural Poisson variation. Furthermore a larger than expected n_i would produce, other things being equal, a smaller than expected n_s/n_i and *vice versa*, resulting in a negative correlation. In other words, the measured ages and U-concs are not independent because they use the same n_i, and n_i is imprecise.

LET US say this with numbers. Let the true $\rho_s = 50$, and $\rho_i = 100$, so $\rho_s/\rho_i = 0.5$. And let us count tracks in equal areas, which we take to be unit area.

Suppose we observe $n_s = 50$ (fortuitously equal to the true value of ρ_s) and $n_i = 108$, which is a random value from a Poisson distribution with mean 100. Then our estimate of uranium concentration (in suitable units) will be 108, which is a bit larger than the true value 100, while our estimate of ρ_s/ρ_i will be $50/108 = 0.46$, a bit lower than the true value of 0.5. If instead we had observed $n_i = 94$, which we might well have, we would estimate the uranium concentration to be 94

(an under-estimate) and simultaneously estimate ρ_s/ρ_i as $50/94 = 0.53$ (an over-estimate).

For several grains with the same $\rho_s = 50$ and $\rho_i = 100$, induced track counts greater than 100 will tend to produce lower age estimates while induced track counts less than 100 will tend to produce higher age estimates. Thus when we plot the age estimates against the corresponding U estimates there will be a negative association.

In practice, other things also vary — spontaneous track counts, areas, uranium concentrations and possibly true ages — which may partially hide this effect, but it will still exist. It is essentially a correlation between the estimation errors of age and uranium concentration. It exists precisely because n_i is imprecise.

Two associations

So there are *two* negative associations. When assessing any relation between age and uranium, we need to understand not only how grains are selected but also how the measurements are related.

Which effect is greater? As far as selection is concerned, the counter must get some feel about whether a high proportion of grains are rejected

because ρ_s is outside the acceptable range. This effect seems likely to be relatively strong when both the true ages and uranium concentrations vary substantially. The correlation due to the related measurements, on the other hand, is easy to estimate and take account of if necessary. Its effect will be greatest when the true ages and Us are constant. And it will arise just as easily in apatite samples. For example, I happen to have results from Ian Duddy for the two apatite samples used for the Besançon interlaboratory trial. The sample correlations between the ages and uranium concentrations (20 grains each) are -0.41 and -0.42.

IN FACT it is possible to measure age and uranium "independently" by the external detector method. After counting n_s, irradiating and counting n_i, send your grains back to the reactor to irradiate them *again* (using a new detector) and obtain a second induced track count n_{i2} for the same matched areas. Use n_s/n_i to measure age and n_{i2} to measure uranium, and you will not see the latter negative relation. Furthermore, any relation you do see will then have some other cause.

SOME YEARS AGO Chuck Naeser sent me some fission track age determinations for zircons from the Fish Canyon Tuff, where two irradiations were

done to obtain duplicate induced track counts for the same 48 grains. The counts are beautiful, mostly between 200 and 400 for both spontaneous and induced tracks. The sample correlation between single grain ages (Ma) and uranium concentrations (ppm), using results from the first irradiation only, is -0.27. But the correlation between age from the first irradiation and uranium from the second irradiation is only -0.09, much nearer zero. I would expect the selection effect to be small here and the correlation of -0.27 to be essentially due to non-independence of the measurement errors.

Random numbers

YOU CAN demonstrate these correlations with random numbers, shown in the graphs overleaf.

Figure 1: constant age and uranium

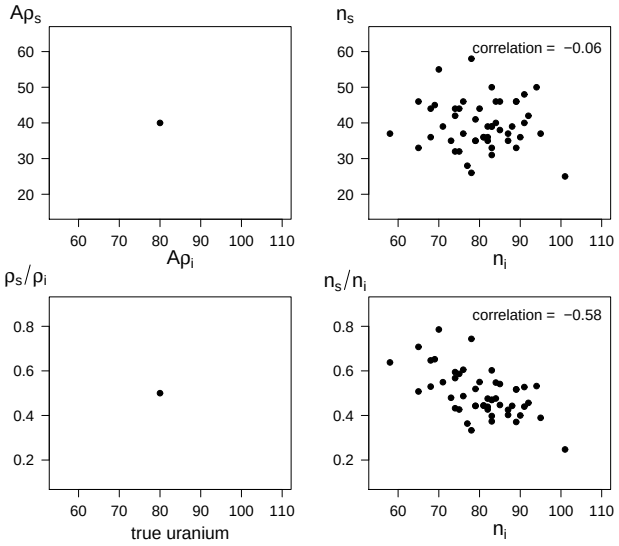
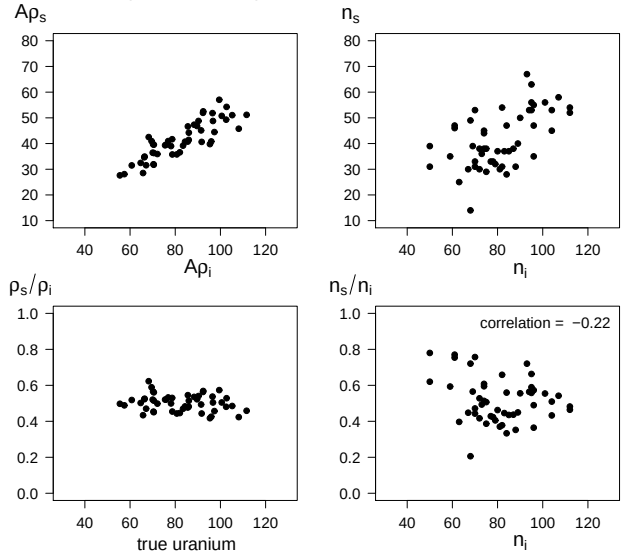


Figure 2: 10% age and 20% uranium variation



IN FIGURE 1 there are 50 notional grains each with the same true $A\rho_s = 40$ and $A\rho_i = 80$. The counts have random Poisson values with these means. For simplicity suppose that the same area A is always used.

The left panels plot the 'true' values: $A\rho_s$ against $A\rho_i$ (top left) and ρ_s/ρ_i against $A\rho_i$ (bottom left), all of which are the same for each grain. The right panels plot the corresponding 'observed' values: n_s against n_i, and n_s/n_i against n_i, where the Poisson variation is added.

The bottom left panel is effectively true age against true uranium while the bottom right panel is effectively measured age against measured uranium — and clearly shows the negative correlation that we now expect.

IN FIGURE 2, I have generated random values of $A\rho_i$ with a 20% coefficient of variation and independent random values of ρ_s/ρ_i with a 10% c.v.

The true ages and U-concentrations vary between grains, but are independent (bottom left), while the measured values are negatively correlated (bottom right). Now the association is not so strong. If the true ages and amounts of uranium vary enough, this variation will mask the Poisson variation and hide the correlation.

Figure 3: 50% age and 50% uranium variation

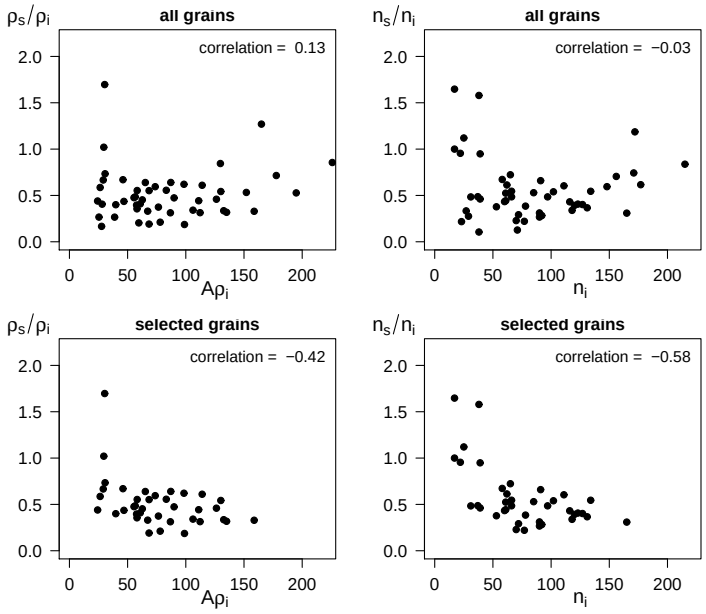


FIGURE 3 illustrates effects of both selection and related measurements when the true ages and uranium contents are independent and both have a 50% c.v. The top left panel plots independent random values of ρ_s/ρ_i against $A\rho_i$ for 50 grains, and shows a practically zero correlation. The top right panel plots the measured values n_s/n_i against n_i. The correlation is now slightly negative, but still quite near zero — the larger variation in true age and U has largely masked the Poisson variation.

In the bottom panels I have selected only those grains whose spontaneous track count n_s is between 10 and 80. The real selection criterion is of course much more subtle than this. For these grains, there is a clear negative correlation between the true ages and Us due to selection (bottom left panel) and an even stronger one between the measured values, due to both selection and non-independence (bottom right panel). All panels use the same scales, so those with keen eyes can discern which points have been selected and which have been most affected by Poisson's law.

A formula

I HAVE worked out an approximate formula for the correlation between n_s/n_i and n_i/A based on the following assumptions. Areas A are randomly chosen with mean ν and coefficient of variation δ; values of ρ_i are randomly chosen with mean μ and coefficient of variation α; and values of ρ_s/ρ_i are randomly chosen with mean λ and coefficient of variation β. Then the correlation between n_s/n_i and n_i/A is approximately

$$\begin{aligned} & 1/\sqrt{1+\nu\mu\alpha^2(1+\delta^2)} \\ \times \quad & 1/\sqrt{1+\lambda^{-1}+\nu\mu\beta^2(1+\alpha^2+\delta^2+\alpha^2\delta^2)}\,. \end{aligned}$$

A few simulations suggest that this formula slightly under-estimates the strength of the correlation, but by substituting reasonable values for the various means and coefficients of variation it will probably give a good idea of the size of this effect. If all areas are equal, put $\delta = 0$. Substituting $\delta = 0$, $\nu\mu = 80$, $\lambda = 0.5$, $\alpha = 0.2$ and $\beta = 0.1$, which are the parameters used for Figure 2, gives a correlation of -0.25, close to the empirical value of -0.23 that we found.

WHEN all true ages, uranium concentrations and areas are the same, then $\alpha = \beta = \delta = 0$ and the above formula reduces to

$$-1/\sqrt{1+\lambda^{-1}}$$

which can also be expressed as

$$-\sqrt{\rho_s/(\rho_s+\rho_i)}\,.$$

This is the formula cited by Andy Carter. Substituting $\lambda = 0.5$ gives a theoretical value of -0.58, in miraculous agreement with the empirical value in Figure 1. Of course it is not possible to give a general formula for the correlation due to selection.

Change against initial value

SPURIOUS negative associations have often arisen in the scientific literature, particularly in medical and social sciences, where measurements contain a substantial "random" component. Measurements on a number of individuals are taken before and after some treatment or intervention, and in order to see if the treatment effect might depend on the underlying level of the variable in question, the change (or treatment effect) is plotted against the initial measurement. A "significant" negative correlation is reported, often with much excitement, and it is concluded that the treatment reduces the response for those with an initial high value, but has an opposite effect for those with an initial low value. The fact is that there would be a negative correlation even if the treatment had no effect at all. What would be really interesting is if there were no correlation!

Eliminate the negative

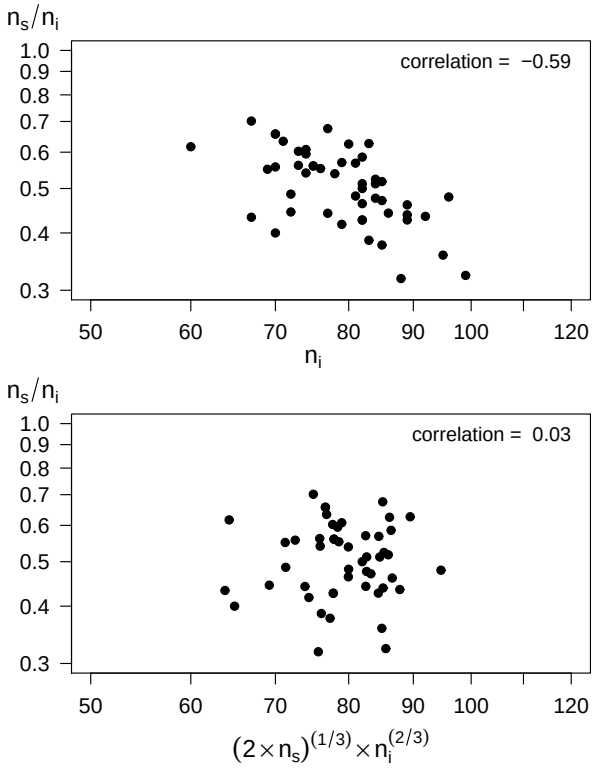
ACTUALLY, there is a clever trick for trying to see if any change depends on the size of the measurement, which is: plot the change (after minus before) against the mean of the before and after values. This seems intuitively wrong because the second measurement includes the treatment

effect. But when the treatment effect is additive and the random component of variation is about the same for both the before and after measurements, this has the magical effect of eliminating the negative correlation. A trend on this plot might thus be of real interest — the average treatment effect might depend on the initial level, or the treatment might change the variance of the responses. I attribute this idea to the late Patrick Oldham, though our American friends might claim it to be one of John Tukey's many inventions.

Now I can hear you thinking "If you plot the *logarithm* of n_s/n_i against the *logarithm* of n_i, is that not the same as plotting a difference against the initial value?"

So it is — and therefore can we not "Eliminate the Negative" by plotting $\log n_s/n_i$ against the mean of $\log n_s$ and $\log n_i$? Not quite, because n_s and n_i have different variances, but we could fix that up by using a weighted mean. For fun, I have worked out how to do it for the constant age and U case, as in Figure 1: any linear function of $\rho_s \log n_s + \rho_i \log n_i$ will do.

In the next figure, the upper panel shows n_s/n_i against n_i for 50 grains with the same parameters as for Figure 1, but now using log scales. The lower



panel uses for the x-axis effectively $\log n_s + 2 \log n_i$ instead of $\log n_i$, though I've first added $\log 2$ and then divided by 3 to get the same range of x as in the upper panel. Thus on the x-axis we have the cube root of $2n_s$ times the square of the cube root of n_i, plotted on a log scale. Hey presto! No correlation! Fun lovers are invited to find the right x-axis function for Figure 2.

Good news from Ghent

THE INTERNATIONAL WORKSHOP in Ghent contained much good news. No more hydrothermal intrusion, crustal cooling, tectonic exhumation, extensional unroofing or (regrettably) uplift and erosion. Just denudation!

But, as every English schoolboy remembers, Ghent is the famous origin of some earlier good news, though what this news actually was, he has long forgotten, as have the inhabitants of Aix. I was reminded of this some years ago while visiting my wife's aunt. The first course of dinner was a rather thin consommé with croutons — or so I thought, until I noticed some paper floating in my bowl. To my astonishment it had writing on it:

> I sprang to the stirrup, and Joris, and he;
> I galloped, Dirck galloped, we galloped all three

The text became somewhat faded, but I could just make out how it ended:

> No voice but was praising this Roland of mine,
> As I poured down his throat our last measure of wine,

Which (the burgesses voted by common
consent)
Was no more than his due who brought
good news from Ghent.

An English aunt is by nature eccentric, but this needed more explanation. “A wonderful idea”, I said, “to serve poetry with the soup, but isn’t it rather difficult to read?”

“Yes, I was curious too”, she replied, "but Mrs Beeton clearly said: If the Mixture fails to Thicken, put a little Browning in."

Sources

Beeton, I.M. (1859-61) *Mrs. Beeton's English Cookery and Household Management.*

Blythe, Ann (1996) Preliminary note on Hf and U concentrations and fission-track annealing in zircon. *On Track,* November, 1996.

Browning, Robert (1840-50) How they brought the Good News from Ghent to Aix. The Poetical Works of Robert Browning. Oxford University Press.

Carter, A. (1990) The thermal history and annealing effects in zircons from the Ordovician of North Wales. *Nucl. Tracks Radiat. Meas.* **17,** 309-313.

Duddy, Ian (1988) Fission track dating results for Besançon apatite age standards. Personal communication.

Green, Paul (1997) E-mail message on selecting zircons. Personal communication.

Mercer, Johnny (1944) "You've got to

Accentuate the Positive,
Eliminate the Negative,
Latch on to the Affirmative,
Don't mess with Mister In-Between."

Popular song.

Muir, Frank (ca.1968) "...put a little Browning in." *My Word,* BBC radio programme.

Naeser, Chuck (1983) Fission track age determinations for Fish Canyon Tuff zircon. Personal communication.

Oldham, P.D. (1962) A note on the analysis of repeated measurements of the same subjects. *Journal of Chronic Diseases*, **15,** 969–977.

Shakespeare, W.S. (1605) Hamlet, II, i, 78. The New Shakespeare. Cambridge University Press.

Tracker (1993) Trials and Errors, *On Track,* May 1993.

First published in *On Track* in May 1997, by *Tracker,* Your man at the microscope with the uncertain age and the U concentration.

How Many Greens in a Tony?

"WHAT are the units of zeta?"

Not a question that one expects to be asked in polite company, or indeed anywhere. We were revising our book chapter in response to an impressively meticulous analysis from our editor. A peaceful activity in which I would read aloud, first our offending sentence, then the editor's suggested alternative and then, occasionally, an essay by the editor's friend. My co-author would idly tap his keyboard in consideration, possibly changing a hyphen here or a comma there, before moving to the next sentence. Occasionally, with a gasp of satisfaction, he would find another log and change it to ln.

Oxford blues

COMMAS were, in fact, of some concern. It was clear that our editor had no time for that noble literary construction the Oxford comma, for they had been ruthlessly blue-pencilled, except for a few late ones that we put down to luck and one that was, unaccountably, inserted. Here is an example:

> "... this would normally involve using suitable graphical displays, numerical summaries and diagnostics, and the fitting of models that reflect the geological environment."

The Oxford comma is the one before the "and". Of course we are taught that a comma should not precede an and or an or. But in this case it does serve a purpose of separating the first three activities, to which the verb "using" applies, from the other activity (the fitting of models) to which the verb "fitting" applies. Without it, the meaning is less clear and the grammar dubious. You may well say why not write something more interesting, but that is not the point.

An Oxford comma, incidentally, should not be confused with an Oxford *Street* comma, which would come between Dickens and Jones and Marks and Spencer and Waring and Gillow. Alas, these last names are rarely heard in Oxford Street nowadays, of whom the late Sir Maurice Kendall memorably recalled:

> Waring and Gillow,
> Slept on one pillow,
> Not for economy,
> But for reasons of bonhomie.

SIR MAURICE'S speech was on the importance of being a pair — and on the relative importance of choosing a partner whose name starts later in the alphabet. Examples included Neyman and Pearson, Durbin and Watson, Box and Cox, and of course Kendall and Stuart. Waring and Gillow were the exception, which may also explain their demise. But, while becoming a double act was a sound recipe for success, MGK (as he was known) warned us against collaborations of more than two, unless you came first, for fear of vanishing into *et al.* He would therefore be astonished to learn that the latest edition of Kendall and Stuart's *Advanced Theory of Statistics* is called *Kendall's Theory of Statistics* by Stuart and Ord.

Chicago style

OUR EDITOR, presumably out of fairness of mind, had in addition inserted a number of what can only be described as Chicago commas, such as:

> In statistical science, it is standard practice to use ...,
>
> For some data sets, it is not obvious whether ...

and

> With the external detector method, we could ...

so that the overall comma density, ρ_c, was practically unchanged. However, the main rule of Chicago English is that "however" is invariably

replaced by "but", which is undoubtedly a good thing.

Prepositions were another contentious matter. A preposition, so the saying goes, is something you should never end a sentence with. But it is impossible to explain naturally the meaning of a confidence interval without ending with a preposition — it is the range of values of the parameter that the data are consistent with. Our school teacher, who was not acquainted with the finer concepts of statistical inference, would say that ending a sentence with a preposition was something up with which he will not put, thereby avoiding a double-fault. This must be an extreme position, otherwise why would Shakespeare have written:

> My pulse as yours doth temperately keep
> time,
> And makes as healthful music – it is not
> madness
> That I have uttered, bring me to the test
> And I the matter will re-word, which
> madness
> Would gambol from.

?

It is of course possible to end a sentence with *more* than two prepositions. I believe the record is

five, held by the child who said "What did you bring that book I didn't want to be read to out of up for?"

A basic unit

OUR EDITOR'S QUESTION came as a shock for another reason, for we had just been pondering a more basic question: "What are the units of *track density*?" Tracks per square centimetre, you will no doubt say, writing it as t/cm^2. Of course "tracks" is short for "number of tracks" and is not really part of the *units*, so the official symbol would be cm^{-2}. But no one *thinks* of tracks per square cm, as shown by the following typical conversation in the lab:

> KERRY: Those zircons from China are too dense to count.
> CHERRY: We must have over-dosed them. What was row dee?
> KERRY: One point three.

For Kerry/Cherry, you can substitute Andy/Sandy, Barry/Thierry, Casey/Tracy, Danni/Yanni, or Hari/Shari as necessary. Kerry says "One point three" while looking at some computer output displaying the number 0.13287936E+07, which would no doubt later become 1.33×10^6 cm^{-2} in a report.

Our editor had also doggedly inserted $\times 10^6$ cm^{-2} wherever we gave a numerical track density. But Kerry does *not* say "One point three times ten to the six". The natural unit of track density is surely *millions* of tracks per square cm, that is, 10^6 cm^{-2}, rather than cm^{-2}. This unit needs a name: I propose *Nancy*, with international symbol Ny. The above track density would then be 1.33 Ny. As a side effect, this would also help the computer to avoid printing unreadable columns of 8-digit numbers followed by an exponent.

Units of zeta

WELL, what *are* the units of zeta?

A phone call to a nearby authority confirmed that the FT community is suspiciously silent on the matter. Clearly a name is required here also. My colleague proposed *Tony*, and I heartily agreed, with international symbol Ty. Zeta historians will note that the symbol ζ was first proposed by Fleischer and Hart (1972), long before Hurford and Green (1983) firmly established the viability of zeta calibration and promoted its routine use. There is therefore a *prima facie* case for *Bobs* or *Harts*. *Greens* would be another possibility, but they would undoubtedly be confused with *Grays* by future generations of dating undergraduates. No doubt

there will be much further debate during which someone may discover that Archimedes used ζ to calibrate the hydrothermal displacement in a domestic basin.

IN THE MEAN TIME let us try to work out what a Tony is. The modern fission track age equation is usually written as

$$t = \lambda_d^{-1} \log(1 + g\,\lambda_d\,\zeta\,\rho_d\,\rho_s/\rho_i)$$

where g is the geometry factor, taking the value $\frac{1}{2}$ for the external detector method and 1 for the population method. Now t is a time and λ_d is a rate (reciprocal time) so that $\lambda_d t$ is dimensionless. Multiplying both sides of the equation by λ_d shows that the term inside the log( ) must therefore be dimensionless. Also ρ_s/ρ_i is dimensionless, as is the geometry factor g. Therefore $\lambda_d \zeta \rho_d$ must be dimensionless. But λ_d is reciprocal time and ρ_d is reciprocal area, so ζ must have the dimension of area $\times$ time (or L^2T, as we used to say). The *units* of zeta may therefore be taken to be "square centimetre years", or $cm^2 \times a$, which may also be written as $Ma \times cm^2/10^6$, or in other words Ma/Ny. Therefore a Tony is a million years per Nancy, or, if this is easier to remember: one Nancy Tony equals one Ma.

WHILE on the subject of the FTA equation and units, let us resolve to express λ_d as 1.55125×10^{-4} Ma^{-1} instead of 1.55125×10^{-10} a^{-1}, and ρ_d in Nancys (millions of tracks per square cm). Then the age t will automatically be in Ma, which is what we want, and we need never again worry about whether or not we should have multiplied ρ_d by 10^6 or about writing decimal numbers like 0.0000000000155. Thus Kerry's λ_d and ρ_d would be 1.55125×10^{-4} and 1.33 instead of 1.55125×10^{-10} and 1.33×10^6, respectively.

Real Age

READERS ARE invited to join the Campaign for Real Age, in which the fission track age equation is written with the factor μ_s/μ_i included, i.e., as

$$t = \lambda_d^{-1} \log\left(1 + g\,\lambda_d\,\zeta\,\rho_d\,(\rho_s/\rho_i)\,(\mu_i/\mu_s)\right)$$

or, in the linearised version,

$$t = g\,\zeta\,\rho_d\,(\rho_s/\rho_i)\,(\mu_i/\mu_s)\,,$$

where μ_s and μ_i are the mean (equivalent isotropic) lengths of spontaneous and induced tracks. This is natural, because ρ_s/ρ_i is a ratio of *areal* densities of intersections of tracks with a surface, whereas it is the ratio of *volume* densities of fissioned ^{238}U and

^{235}U atoms, τ_s/τ_i say, that might be regarded as the more fundamental quantity. The ρs are related to the τs by the equations

$$\rho_s = \tfrac{1}{2}\tau_s\mu_s \quad \text{and} \quad \rho_i = \tfrac{1}{2}g\tau_i\mu_i$$

so that $g(\rho_s/\rho_i)(\mu_i/\mu_s) = \tau_s/\tau_i$ and the simple fission track age equation is really

$$t = \zeta\,\rho_d\,\tau_s/\tau_i\,.$$

FLEISCHER AND HART (1972) explicitly expressed ζ as a product of several factors, including the ratio of etching efficiencies of induced and spontaneous tracks and the ratio of their etchable ranges. So this real age equation effectively extracts the latter ratio from ζ. This formally changes the *meaning* of zeta, but not the units. With much relief I will leave it to others — such as Hurford (1988) or Van den haute et al (1998) — to discuss what zeta actually *is*.

INCIDENTALLY, I have noticed a recent disturbing practice of presenting papers at conferences *without* citing the age equation. Indeed, I have even seen presentations containing no equation at all! My advice to the young tracker is to program the equation into your logo or header so that it is displayed at the top of each slide. Then you can safely ignore it and still maintain credibility. After

all, at least we *have* an equation, and a more interesting one than most, so let us flaunt it!

What is a Tony worth?

HOW SHOULD we interpret a measurement of one Tony?

The obvious way is to put $\zeta = 1$ in the FTA equation, along with standard values of the other terms. A problem with a direct interpretation is that the value of the induced track density, ρ_i, depends on the neutron fluence used, which in turn is measured by the dosimeter track density, ρ_d, which in turn will depend on the standard glass used. The value is also specific to the mineral and to the analyst. Also, as noted above, the real FTA equation has another factor μ_i/μ_s inside the parentheses. So we assume that $\mu_s = \mu_i$, which will be appropriate for an age standard, but not for a sample that may have been annealed. Then the condition $g\rho_s/\rho_i = 1$ means that the concentrations of fissioned ^{238}U and ^{235}U atoms are the same, i.e. $\tau_s = \tau_i$. This is the natural condition, rather than $\rho_s = \rho_i$. Equally naturally we take $\rho_d = 1$ Nancy. Then, substituting all of these values in the FTA equation gives an age of $t = 1$ Ma.

THUS, eventually, we get the meaning of **one Tony:**

> **the number of square centimetre years required to produce an age of 1 Ma when dating an age standard with equal concentrations of fissioned ^{238}U and ^{235}U atoms and a dosimeter track density of one Nancy.**

Actually, the age you get is 0.99992 Ma, but 1 Ma is close enough for Geology.

Sources

Fleischer, R.L. and Hart, H.R. (1972) Fission track dating: techniques and problems. In *Calibration of Hominoid Evolution*. (eds W.W. Bishop, J.A. Miller and S Cole) Scottish Academic Press, Edinburgh, pp.135–170.

Hurford, A.J. (Tony) (1998) Zeta: the ultimate solution to fission- track analysis calibration or just an interim measure? In *Advances in Fission-Track Geochronology.* (eds Peter Van den haute and Frans De Corte) Kluwer Academic Publishers..

Hurford, A.J. and Green, P.F. (1983) The zeta age calibration of fission-track dating. *Isotope Geoscience,* **1,** 285–317.

Kendall, M.G. (1971) After dinner speech at University College London.

Naeser, N.D. (Nancy) (1997) Editorial comments on draft chapter.

Naeser, C.W. (Chuck) (1997) Supplementary editorial comments on draft chapter.

Shakespeare, W.S. (1605) Hamlet, *III, iv,* 140. The New Shakespeare. Cambridge University Press.

Van den haute, P., De Corte, F., Jonkheere, R. and Bellemans, F. (1998) The parameters that govern the accuracy of fission-track age determinations: a re-appraisal. In *Advances in Fission-Track Geochronology.* (eds Peter Van den haute and Frans De Corte) Kluwer, Academic Publishers.

First published in *On Track* in June 1998, by *Tracker,* Your man at the microscope with the Hercynian granitoid and the natural glasses.

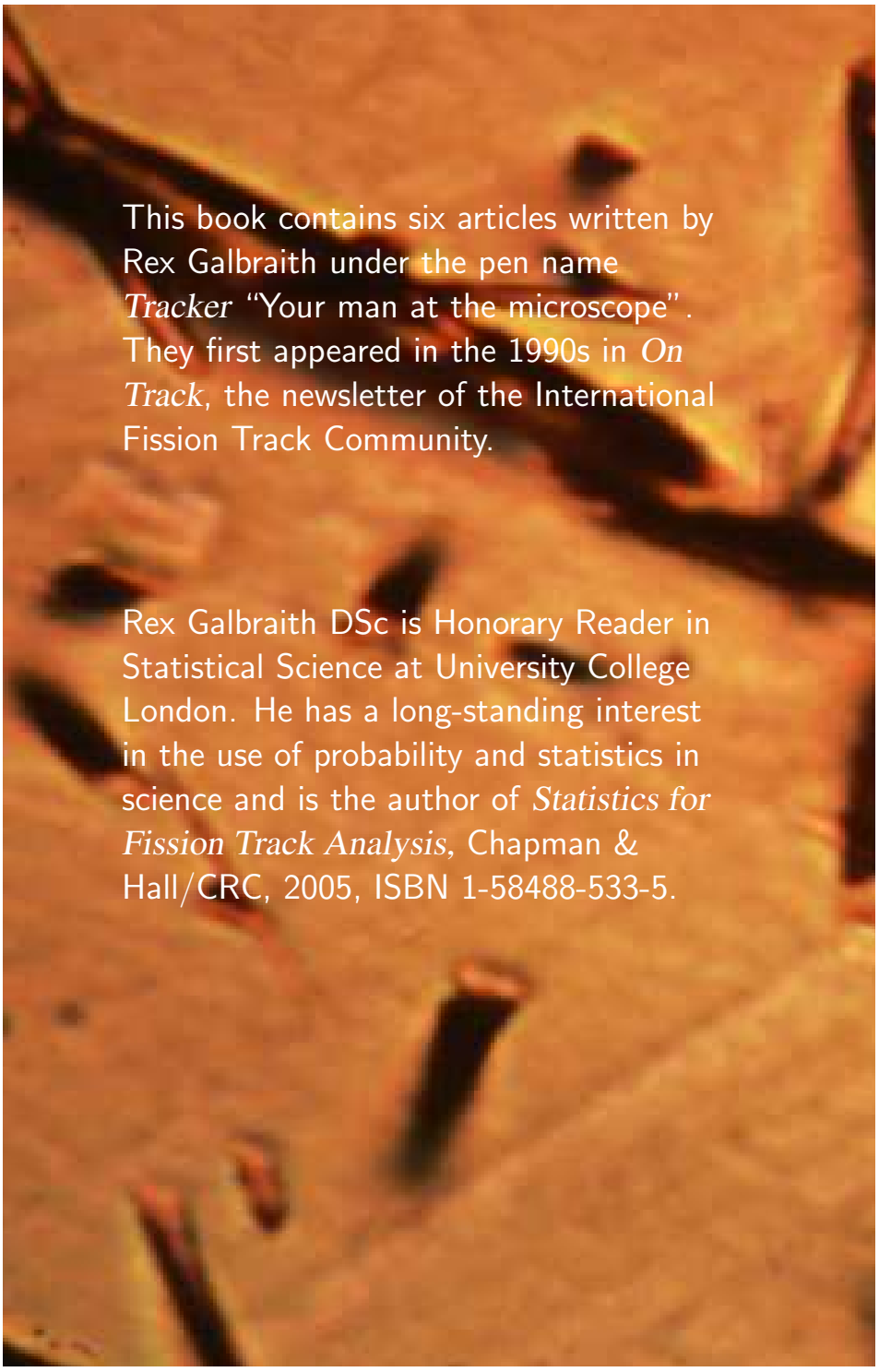
Trackers from the Nineties

Alain Chamboudet
Alasdair Duncan
Amanjit Sandhu
Ana-Voica Bojar
Andrea O'Sullivan
Andy Carter
Andy Gleadow
Ann Blythe
Annia Fayon
Antoine De Wispelaere
Arnaud Meyer
Asaf Raza
Barbara Ventura
Barry Kohn
Bart Kowallis
Bernhard Fugenschuh
Bob Fleischer
Bob Geigengack
Bob McCorkell
Bob Walter
Bob Zimmerman
Burford Price
Carlo Sanders
Carlos Saenz
Caroline Vercoutere
Casey Ravenurst
Charlotte Cederbom
Cherry Lewis
Chuck Naeser
Dale Issler
Dan Kendrick
Danny Stockli
Dave Belton
Dave Coyle
Dave Fisher
Dave Foster
Dave Upton
Dennis Arne
Di Seward
Dick Van der Wateren
Dieter Storzer
Dominik Hungerbuhler
Don Miller
Doug Walker
E Mark
Ed Sobel
Edmund Stump
Elizabeth Miller
Ewald Hejl
Frank Bellemans
Frank Lisker
Frans De Corte
Fred Walgenwitz
Ian Duddy
Geoff Laslett
Geoff Ruiz
Gerard Poupeau
Gerco Stapel

Gerd Sekyra
Gian Zuffa
Giulio Bigazzi
Giulio Viola
Glen Stockmal
Gloria Toro
Gomaa Omar
Gunter Wagner
Gurinder Singh
Hardev Singh Virk
Hari Saini
Heimo Schierl
Helen Gibson
Henning Qvale
Hideki Iwano
Hiroto Ohira
Hisotoshi Ito
Istvan Dunkl
James Talbot
Jaime Perez De Armas
Jan Burchart
Jan-Peter Stiberg
Jean-Jaques Braun
Jeff Corrigan
Jenny Winkler
Jocelyn Barbarand
Joelle Carpena
John Dekker
John Garver
John Murphy
John Westgate
Jon Linn
Joong Kim
Jorge Clavero
Jorge Mora
Jos Terken
Joseph Vance
Julio Hadler
KD Bal
Keith Mahon
Kerry Gallagher
Kerry Hegarty
Kevin Crowley
Kevin Hill
Kiyoshi Wadatsumi
Kirsten Graefe
Kirsten Hansen
Kit Johnston
Koichiro Watanabe
Lothar Ratchbacher
Lisel Currie
Liu Shunsheng
Luis Barbero
Manfred Brix
Manuel Grivet
Marc Jolivet
Marcos Zentilli
Maria Balestrieri
Mark Brandon
Mark Cloos
Mark Harrison
Martin Dodson
Martin Olesch
Martin Wagner
Mary Roden-Tice
Marylin Moore
Masako Ono
Masao Hayashi

Masao Kasuya
Masao Suzuki
Massimo Oddone
Massmilliano Zattin
Mat Roden
Matthais Giger
Maurice Brunel
Maurice Pagel
Max Coleman
Max Rohrman
Mehdi Sohrabi
Meinert Rahn
Mel Giles
Mel Mitchell
Michael Steckler
Michael Steinman
Michel Rebetez
Mike de Wit
Mike Hulver
Mike Krochmal
Mike Summerfield
Minoru Aizawa
Nancy Naeser
Nand Lal
Nelson Eby
Nick Petford
Noriko Hasebe
Olivier Dorighel
Osamu Himeno
Pat Kelly
Paul Andreissen
Paul Fitzgerald
Paul Green
Paul O'Sullivan
Pete George
Peter Kamp
Peter Van den haute
Peter Zeitler
Phil Armstrong
Phil Shane
Rajinder Prashad
Rasoul Sorkhabi
Ray Donelick
Raymond Jonckheere
Rebecca Harman
Renato Pagliuca
Rex Galbraith
Rich Ketchman
Rich Weiland
Richard Spikings
Rizwana Waraich
Rod Brown
Roy Dokka
Russ Evarts
Ruth Sidall
Ruyji Yamada
SR Hashemi-Nezhad
Saeed Durrani
Sandy Grist
Sandra Guglielmetti
Sara Foland
Satoshi Koshimizu
Sean Willet
Sergio Iriarte
Shari Kelly
Shauhua Sun
Simon Parand
Steffan Boettcher

Steve Bergman
Steve Carpenter
Stuart Thomson
Sue Baldwin
Sue Marshallsea
Susumu Nishimura
Suzanne Baldwin
Taka Tagami
Takaaki Matsuda
Teruyyki Honda
Tetsu Yoshoika
Thane McCulloh
Thierry Calmus
Thomas Puch
Thorsten Shaefer
Tilmann Mark
Tim Lutz
Tohru Danhara
Tohru Yamashita
Tom Redfield
Tomaki Sumii
Tony Hurford
Tracy Paul
Trevor Dumitru
Ulrich Glasmacher
VP Perelygin
Victor Maksaev
Wayne Noble
Wes Wallace
Will Maze
William Carlson
Yanni Gunnell
Yoshiro Ganzawa
Yun Pan
Yunlong Zhao
Zehra Yegingil
Zhai Pengji



This book contains six articles written by Rex Galbraith under the pen name *Tracker* "Your man at the microscope". They first appeared in the 1990s in *On Track*, the newsletter of the International Fission Track Community.

Rex Galbraith DSc is Honorary Reader in Statistical Science at University College London. He has a long-standing interest in the use of probability and statistics in science and is the author of *Statistics for Fission Track Analysis*, Chapman & Hall/CRC, 2005, ISBN 1-58488-533-5.