Available online at www.sciencedirect.com
ScienceDirect
Geochimica et Cosmochimica Acta 70 (2006) 5183–5200
Geochimica
www.elsevier.com/locate/gca

A Bayesian approach to calibrating apatite fission track annealing models for laboratory and geological timescales

John Stephenson [a], Kerry Gallagher [a,*], Chris Holmes [b]

[a] Department of Earth Science and Engineering, Imperial College London, South Kensington, London SW7 2AS, UK
[b] Department of Statistics, Oxford University, Oxford OX1 3TG, UK

Received 27 February 2006; accepted in revised form 26 July 2006

Abstract

We present a new approach for modelling annealing of fission tracks in apatite, aiming to address various problems with existing models. We cast the model in a fully Bayesian context, which allows us explicitly to deal with data and parameter uncertainties and correlations, and also to deal with the predictive uncertainties. We focus on a well-known annealing algorithm [Laslett, G.M., Green, P.F., Duddy, I.R., Gleadow. A.J.W., 1987. Thermal annealing of fission tracks in apatite. 2. A quantitative-analysis. *Chem. Geol.*, **65** (1), 1–13], and build a hierachical Bayesian model to incorporate both laboratory and geological timescale data as direct constraints. Relative to the original model calibration, we find a better (in terms of likelihood) model conditioned just on the reported laboratory data. We then include the uncertainty on the temperatures recorded during the laboratory annealing experiments. We again find a better model, but the predictive uncertainty when extrapolated to geological timescales is increased due to the uncertainty on the laboratory temperatures. Finally, we explictly include a data set [Vrolijk, P., Donelick, R.A., Quenq, J., Cloos. M., 1992. Testing models of fission track annealing in apatite in a simple thermal setting: site 800, leg 129. In: Larson, R., Lancelet, Y. (Eds.), *Proceedings of the Ocean Drilling Program, Scientific Results*, vol. 129, pp. 169–176] which provides low-temperature geological timescale constraints for the model calibration. When combined with the laboratory data, we find a model which satisfies both the low-temperature and high-temperature geological timescale benchmarks, although the fit to the original laboratory data is degraded. However, when extrapolated to geological timescales, this combined model significantly reduces the well-known rapid recent cooling artifact found in many published thermal models for geological samples.

1. Introduction

Fission track analysis is widely used to constrain the low-temperature thermal history of rocks in the upper part of the Earth's crust, relying on the thermally activated shortening, or annealing, of the meta-stable fission tracks. Both zircon and apatite are used for analysis, and the latter is generally considered to be better calibrated in terms of annealing. To date, these calibrations rely on empirical model formulations, typically aiming to capture the temperature and time dependence of track shortening in terms of a linear dependence on time, and exponential in temperature (or log in time, and linear in temperature). Furthermore, they are generally based on laboratory experiments which are performed on timescales of less than 1 year, and typically much less (e.g., Green et al., 1986; Carlson et al., 1999; Barbarand et al., 2003a). A common criticism of these empirically calibrated models is the lack of real physics. However, a more practical problem is the uncertainty introduced through the extrapolation of the laboratory-calibrated models (<1 year) to geological timescales (10^6–10^8 years).

A consequence of this extrapolation when modelling thermal histories from apatite fission track data with certain annealing models (e.g., the widely used Laslett et al., 1987) is the introduction of rapid recent cooling from around 60 °C to surface temperatures. This apparent artifact is particularly pronounced when dealing with samples

* Corresponding author. Fax: +0 20 7594 7444.
E-mail addresses: john.stephenson@imperial.ac.uk (J. Stephenson), kerry@imperial.ac.uk (K. Gallagher).

0016-7037/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.gca.2006.07.027

collected from surface outcrop, showing relatively old ages (150–250 Ma) and short mean track lengths (11–12 μm). A likely cause of this is that the annealing model does not adequately capture subtle low-temperature annealing behaviour over geological timescales.

In this contribution, we address this extrapolation problem explicitly by calibrating the model with both laboratory and geological timescale data. In doing this, we adopt a Bayesian approach to quantify the annealing model, and its uncertainties, focussing on the well-known Laslett et al. (1987) annealing model for Durango apatite, although the methodology we present is more generally applicable. The method takes into account uncertainty in the thermal history based on the field (geological) data, as well as that inherent when calibrating the annealing model solely on laboratory data, including the uncertainty on the laboratory temperature measurements.

The remainder of this paper is structured so that after briefly describing fission track and Bayesian models in Sections 2 and 3 we gradually build up the complexity of the model, conditioning it on an increasing amount of data. Throughout this work we focus on assessing how well our models honour the constraints, and in particular, how applicable the annealing models are when extrapolated to geological time scales.

2. Fission track annealing models

A typical fission track analysis consists of measuring a series of confined horizontal track lengths, and counting the number of surface intersecting tracks in apatite or zircon. As the rate of formation of natural or spontaneous tracks (from ^{238}U) is effectively constant, and newly formed tracks are assumed to have a more or less constant initial length, their length distribution can be used to infer the thermal history of the host rock if a quantitative annealing model is used, e.g., Gallagher (1995).

In order to quantify the annealing behaviour, the approach to date has been to anneal artificial, or induced, tracks (from ^{235}U) over a range of different temperature and time conditions (Green et al., 1986; Carlson et al., 1999; Barbarand et al., 2003a). Wendt et al. (2002) undertook a series of experiments to assess the role of pressure. They inferred some pressure and stress influence on annealing, although this influence remains controversial (e.g., Kohn et al., 2003) and we do not consider this further.

When considering fission track length data and a thermally activated annealing process, the model has typically been formulated in terms of the reduced track length ratio, r,

$$r = \frac{l}{l_0}, \tag{2.1}$$

where l is the measured track length on an annealed sample, and l_0 is the length of an unannealed track. Although there is some evidence of very short term room temperature annealing from around 17 μm over 2–3 weeks (Donelick et al., 1990), the effective value of l_0 is inferred to be stable at around 16–16.3 μm (Green et al., 1986), with some dependence on the observer and sample preparation conditions, as well as the track orientation (Barbarand et al., 2003b).

To model annealing under different temperature and time conditions, we need some annealing function A, to relate time and temperature to the observed reduced track length, i.e.,

$$r = A(T, t, \boldsymbol{\beta}), \tag{2.2}$$

where T refers to isothermal temperature in kelvin, t is the time in seconds at temperature T, and $\boldsymbol{\beta}$ is a vector of parameters to be inferred from observations.

Historically, the most popular annealing model is the fanning model of Laslett et al. (1987), which we shall refer to as the Laslett model in this paper. There are other annealing models which are based on the same general formulation (i.e., annealing has a linear dependence on time and exponential on temperature), with a set of empirical parameters to be determined from laboratory annealing experiments, i.e., $\boldsymbol{\beta}$ as mentioned above (e.g., Crowley et al., 1991; Laslett and Galbraith, 1996; Ketcham et al., 1999). The model presented by Carlson (1990) was formulated in a more obvious chemical kinetic way in terms of the unknown parameters, in the hope that some of these would be amenable to direct measurement to provide more insight into the mechanisms of annealing. The various models have been reviewed in detail by Ketcham et al. (1999), who also discuss the relevance of the empirical models in providing insights into the underlying physical processes. Briefly, they suggest that the observed annealing behaviour may reflect defect elimination via a distribution of activity energies, rather than a single value as proposed by Carlson (1990). The detail of the distribution changes as annealing progresses and also is likely to be a function of composition. However, the calibration of a multiple activation energy model is a non-trivial task, requiring knowledge of the distribution of defects, or the positions and nature of atoms along a track. If the positions of all atoms in a newly formed track could be specified, then some form of molecular dynamics simulation may be a feasible way forward (e.g., Chartier et al., 2001).

The original Laslett model is formulated in two basic functions, given as

$$g(r) = c_0 + c_1 T(\ln(t) + c_2), \tag{2.3}$$

where the transformation of r is given by

$$g(r) = \frac{((1 - r^b)/b)^a - 1}{a}. \tag{2.4}$$

Thus, the vector of unknown annealing parameters in (2.2) is given by $\boldsymbol{\beta} = \{c_0, c_1, c_2, a, b\}$, whose order we shall retain for the remainder of this paper, to facilitate comparison to the original work of Laslett et al. (1987). The first of the functions (RHS of (2.3)) captures the common Arrhenius type relationship between temperature, time and the degree or rate of reaction, whilst the second is a variation on the Box–Cox transform, used in this case under the assumption that the transformed data will have a constant residual variance (Box and Cox, 1964). The Laslett model and the associated parameters were originally constrained using a series of laboratory annealing experiments on the Durango apatite, performed on newly formed (i.e., induced) fission tracks, under different time–temperature conditions (see Green et al. (1986) for a description of the experiments and data). The model parameters were subsequently estimated using a maximum likelihood optimisation technique. The final model parameters, $\boldsymbol{\beta}_{\mathrm{Las}}$, reported in Laslett et al. (1987) are given in Table 1, and the initial track length used in the model was fixed at 16.3 μm. These values will be used as a performance benchmark for evaluating the Bayesian approach we present in this paper.

This model produces non-linear behaviour in the rate of annealing, which increases with the degree of annealing. When extrapolated to geological timescales (Green et al., 1989), the model predicts that total annealing (equivalent to a reduced track length of around 0.4–0.5) will occur around 100–120 °C, depending on the heating duration. This behaviour is broadly similar to the high-temperature sensitivity observed in geological samples from exploration wells in the Otway Basin, SE Australia (Green et al., 1989).

However, when used to model cooling histories for rock samples currently on the Earth's surface, a variety of studies have inferred that the model does not capture the low-temperature, long timescale annealing behaviour well. Perhaps the most convincing demonstration of this is based on the data of Vrolijk et al. (1992), who used volcanogenic sediments from an Ocean Drilling Program (ODP) well which have a maximum burial depth of a few 100 m. Using a well-constrained forward thermal history simulation, the Laslett model predicts mean track lengths up to 1 μm longer than observed. This seems to be a result of the Laslett model underpredicting the degree of annealing at low temperatures (60 °C) for geological timescales. A corollary of this is that when using the model to constrain thermal histories directly from observed data (e.g., Gallagher, 1995), the inferred thermal histories tend to reside at temperatures > 60 °C until very recently. This is particularly acute for geological samples with old apatite fission track ages (200 m.y.) and short mean track lengths (<12 μm). This is the rapid recent cooling mentioned earlier.

One recent approach to dealing with this problem is to introduce a shorter initial track length l_0 of around 14.5–15 μm into the model formulation (e.g., Allen et al., 2002; Kohn et al., 2002; Gunnell et al., 2003). This is justified heuristically, in that the inferred thermal histories that lack the rapid recent cooling are more in accord with independent geological and geomorphological evidence. Furthermore, Green (1988) effectively invoked an initial track length of around 14.5–15 μm for natural (or spontaneous tracks) to reconcile the relationship between track length and track density observations for laboratory and geological data sets. This rescaling was invoked to account for natural long-term annealing. Laslett and Galbraith (1996) proposed a revised model formulation in which the initial track length was estimated from the data. However, the validity of this approach has been questioned by Ketcham et al. (1999) in that it is not clear what this estimated 'instantaneous' initial length represents in terms of a measurable parameter even over short laboratory timescales.

In experimental rock deformation, it has been pointed out (Paterson and Wong, 2005) that, when extrapolating laboratory data to geological timescales, the rate determining deformation mechanism may change depending on the timescale of interest. Given the very rapid low-temperature annealing observed by Donelick et al. (1990) over 2–3 weeks, then a period of longer term (at least 10 years) stability where the mean track length is around 16.3 μm (A.J. Hurford, pers. comm., 2006), this may also be the case for fission track annealing over such timescales. Unfortunately, at the moment, the physical mechanism of annealing of fission tracks in natural samples is not well understood at any timescale.

Therefore, the current situation is that we have laboratory timescale annealing data, and limited geological constraints. However, we do not know the mechanism of annealing nor whether it changes depending on the timescale. Given that all previous studies have assumed a common model over all timescales, we will follow that assumption and retain the form of the original model from Eqs. (2.3) and (2.4). However, we develop a formal probabilistic approach, allowing us to include explicitly all of the

Table 1

Values for the parameters, $\boldsymbol{\beta}$, for the Laslett et al. (1987) model, the ranges used for the uniform distribution prior on each parameters, and the Bayesian models conditioned only on the laboratory annealing data (MAP1), conditioned on the laboratory annealing and allowing for uncertainty in the laboratory temperature data (MAP2), and conditioned on laboratory and geological data (Vrolijk et al., 1992), and allowing for uncertainty in the laboratory temperature data (MAP3)

Model	c_0	c_1	c_2	a	b	LL
Laslett	−4.87	0.000168	28.12	0.35	2.7	133.3
Prior upper limit	−2.83	0.0005	32.61	0.61	4.4	—
Prior lower limit	−12.05	0.0001	23.81	−0.27	−1.1	—
MAP1	−4.20	0.000140	27.90	0.42	3.1	137.9
MAP2	−3.38	0.000103	28.24	0.52	3.5	138.2
MAP3	−4.30	0.000164	24.46	0.46	2.1	129.4

LL is the negative log-likelihood as defined in Laslett et al. (1987).

available data when inferring the model parameters. Thus, we will use constraints from both laboratory and geological studies in the model estimation process. In this case, we might expect to compromise the quality of the fit to the laboratory data when including geological data. This may reflect a change in the annealing mechanism, or merely the statistical uncertainty arising from the data acquisition and fitting processes.

3. A Bayesian formulation for annealing models

In this paper, we have opted for a Bayesian approach, allowing us to incorporate a range of model parameters as well as data from a variety of sources within a fully probabilistic framework. This allows us to access the level of uncertainty directly in our approach, something which previously has only been advanced in a relatively ad hoc fashion.

For example, in an earlier study, Jones and Dokka (1990) have considered the influence of uncertainties in the model parameters, c_0 and c_1 (see Eq. (2.3)). They implemented a simple Monte-Carlo simulation, and used normal distributions with a standard deviation equal to the standard error on the mean estimates of the two parameters. Also, they adopted a normal distribution (with an unspecified variance) on $g(\mathrm{r})$, T and $T\ln(t)$, thus estimating a perturbed value for c_0. It is not clear from their paper whether they allow for covariance between the model parameters, nor is there a robust estimate of the quality of the data fit for the simulations. However, they conclude that a relative uncertainty of 8% is appropriate for the reduced track length from two laboratory timescale simulations, equivalent to about twice the typical measurement error.

Laslett and Galbraith (1996) used a revised formulation of the original Laslett model, but applied to different laboratory data. The revised model allows for the initial track length to be estimated, and explicitly deals with the variance in the observed data (rather than assuming a constant variance on a transform of the data). They used a maximum likelihood approach and constructed profile likelihoods and confidence regions for some of the parameters. Unfortunately, they did not consider the original data of Green et al. (1986), and so a direct comparison to the original model of Laslett et al. (1987) is not possible.

Our new approach differs significantly from these studies. Here, we use Markov chain Monte Carlo (MCMC) to sample the parameter space, and deal directly with the appropriate data likelihood as a measure of the data fit for both laboratory and field (geological timescale) data. This approach of sampling the model parameter space allows us to readily estimate probability distributions on the estimated parameters. One of our aims is to compare the results of this probabilistic approach to the results of Laslett et al. (1987), and so we use the same basic model formulation, and the same Durango dataset.

3.1. Bayes' theory

We shall now give some brief background to the Bayesian approach and how it can be implemented using MCMC. For more discussion on the Bayesian approach, we refer the reader to Lee (1989), Bernardo and Smith (1994), and Gilks et al. (1996), and to Denison et al. (2002) for a detailed description of the Bayesian approach to regression problems. A detailed exposition of the philosophical distinctions between Bayesian and the traditionally more familiar Frequentist approaches is beyond the scope of this paper, but Bernardo and Smith (1994) give a useful overview. However, a key difference between Bayesian and Frequentist methods is that in the Bayesian framework we assume there is a range of plausible model parameterisations, rather than a single 'true' set of parameters $\boldsymbol{\theta}$. Then, we condition our inference on the observed data, rather than on a single model. The Bayesian approach then allows us to directly characterise the uncertainty in our models, and their resultant predictive densities.

Fundamental to Bayes' theory is the updating of our belief in (or prior knowledge of) the model parameters using observed data. The updated knowledge, conditional on the data likelihood, is then referred to as the posterior. Qualitatively this can be written as

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior} \tag{3.1}$$

or formally

$$p(\boldsymbol{\theta}|\mathscr{D}) \propto p(\mathscr{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{3.2}$$

where $\mathscr{D}$ represents the set of available data, $\boldsymbol{\theta}$ a vector of unknown model parameters to be determined (e.g., the annealing model parameters or an unknown thermal history) and $a|b$ means a conditional on b, so $p(a|b)$ implies the probability of a given b.

In our case, the prior knowledge can take the form of geological or physical constraints on the thermal history (e.g., the present day current temperature, the surface temperature at the time of deposition) or information from previous analyses (e.g., ranges of permissible parameters in the annealing model). The priors may also express our intuitive belief in the uncertainty in a set of parameters. The choice of prior distribution is necessarily subjective, but is based on the belief regarding the likely values of the parameter in question. One issue that commonly arises is the assumption of a relatively uninformative prior, e.g., defining a maximum and minimum range and assuming the all values in this range are equally likely. Often this assumption is fine, but it is important to bear in mind that a uniform prior does not remain uniform if we transform the parameter. Thus, assuming a uniform distribution on x certainly does not imply a uniform distribution on $\log(x)$. Finally, the form of the likelihood is determined by assumptions on the model in question, e.g., Gaussian residuals or Poissonian counting statistics.

In this study, we will consider three different structures for the Bayesian formulation, increasing the complexity with each structure to allow for three different types of data, laboratory timescale track length data, laboratory timescale temperature data, and finally geological timescale track length and thermal history data. The relationship between the data and the unknown parameters in the Bayesian approach can be represented graphically, and in Appendix A, we give a graphical representation for the formulation of the probability models discussed subsequently.

3.2. Markov chain Monte Carlo

Unfortunately, due to the intractability of the constant of proportionality (or normalisation) and the high number of parameters, we cannot solve Eq. (3.2) directly. This leads us to consider Markov chain Monte Carlo, a technique synonymous with Bayesian statistics, which allows us to draw samples from any density up to the proportionality constant. In practice, we deal with ratios of probabilities to draw samples, and so the constant of proportionality cancels out. A key advantage of MCMC sampling is the immediate access to the level of uncertainty, both in the parameters (directly given by the range of parameters we have sampled) and in any predictions we make based on these parameters.

Although there are a variety of MCMC algorithms, they are essentially refinements of the Metropolis–Hastings algorithm, which is the simplest and most general. Here, we give a brief overview, and more detail on MCMC algorithms, we refer the reader to Gilks et al. (1996) and Denison et al. (2002).

Algorithms proceed iteratively from an initial set of parameters $\boldsymbol{\theta}_0$, and probabilistically propose a transition or update from samples $\boldsymbol{\theta}_i$ to $\boldsymbol{\theta}_{i+1}$. We require a pre-defined proposal function $q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)$, which is itself a probability density function. We can see that the new model parameter set, $\boldsymbol{\theta}_{i+1}$, is then conditional on the current state of the parameter set, $\boldsymbol{\theta}_i$. These proposed updates are then accepted or rejected according to a transition probability designed to guarantee that those accepted samples are from the desired stationary distribution $p(\boldsymbol{\theta}|\mathscr{D})$, i.e., we can sample from the posterior distribution, without needing to specify its form.

For example, if $\boldsymbol{\theta}$ comprised a single parameter θ, we could choose a normal distribution for the proposal function, centred on θ_i, such that $p(\theta_{i+1}|\theta_i) \sim \mathscr{N}(\theta_i, v)$ where v is a variance set at the start of the run.

Once a proposal is made, the step is accepted if

$$u < \min\left\{1, \frac{p(\boldsymbol{\theta}_{i+1})p(\mathscr{D}|\boldsymbol{\theta}_{i+1})q(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1})}{p(\boldsymbol{\theta}_i)p(\mathscr{D}|\boldsymbol{\theta}_i)q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)}\right\}, \tag{3.3}$$

where u is a random number drawn from uniform distribution between 0 and 1, and $p(\boldsymbol{\theta})$ and $p(\mathscr{D}|\boldsymbol{\theta})$ are the prior and data likelihood, respectively, which are proportional to the posterior, $p(\boldsymbol{\theta}|\mathscr{D})$, as defined in Eq. (3.2). As we take the ratios of these probabilities, we do not need to know the constant of proportionality (or normalising constant) as it is the same for both $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_{i+1}$, and cancels out. If the proposal function is symmetrical (i.e., $q(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1}) = q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)$), these terms also cancel and we can write the acceptance criterion in terms of just the ratio of the posterior distributions, $p(\boldsymbol{\theta}_{i+1}|\mathscr{D})/p(\boldsymbol{\theta}_i|\mathscr{D})$. In practice, we always accept $\boldsymbol{\theta}_{i+1}$ if it fits the data better than $\boldsymbol{\theta}_i$, otherwise, it is accepted with a probability depending on the posterior ratio, i.e.,

$$u < \min\left\{1, \frac{p(\boldsymbol{\theta}_{i+1}|\mathscr{D})}{p(\boldsymbol{\theta}_i|\mathscr{D})}\right\}. \tag{3.4}$$

The choice for the proposal functions is arbitrary, in that the theoretical application of algorithm does not depend on it. However, the performance, in terms of computational efficiency, does depend on it, and generally some exploratory runs are made to tune the proposal functions appropriately. For example, we need to specify a suitable value of v in the example above to ensure that proposed transitions are neither too small so that movement between different parameter sets is slow, nor too large where few proposed parameter sets are accepted. The choice of proposal density is typically made by trying to keep the number of accepted transitions to around 25%.

As $\boldsymbol{\theta}$ is generally a vector of parameters, we have the option of either proposing and updating each member individually in turn using univariate proposal densities (whilst keeping all other parameters the same, and conditioning on already updated parameters), or 'blocking' together groups of parameters and updating each group in turn. In practice, we are able to combine the transition stage into a hybrid sampler, so that different parameter groups are updated with different proposal schemes. The choice of updating scheme is again largely heuristic.

3.2.1. Maximum a posterior (MAP) samples and credible intervals

While we will consider the range of possible values for $\boldsymbol{\theta}$, in order to compare our derived parameters to those of the original parameters, we will generally make use of the maximum a posterior (MAP) sample. This is simply the sample which has the highest posterior probability (from Eq. (3.2)). It is analogous to the maximum likelihood estimate if prior belief is not taken into account. At various stages, we will be able to integrate analytically over some parameters, a process known as marginalisation. In this case the MAP will be the sample which maximises this marginalised posterior probability.

Bayesian methodology additionally allows us to define 95% credible intervals. These will be used throughout our explanation to demonstrate the level of uncertainty of our annealing and thermal history model predictions. In general, the credible intervals should not be confused with the more frequently used confidence intervals. In certain circumstances, these can be numerically equivalent but the interpretation of their meaning is different. For example, when estimating a mean, the 95% confidence interval represents the interval in which the true value will fall 95% of the time, if we could run many repetitions of the data collection process. In contrast, given the mean estimate (from one realisation of the data), the 95% credible interval represents the region in which we have a probability of 0.95 of containing the true mean value (see Appendix B, Bernardo and Smith, 1994).

4. Constraining the model with laboratory data

We now establish the formal framework for our Bayesian model of fission track analysis. We use the annealing model formulation of Laslett et al. (1987) as described in Section (2), and also the laboratory data of Green et al. (1986), on which the original annealing model was calibrated. Here, we want to demonstrate how the Bayesian approach allows us to quantify the uncertainty in model parameters, and how this influences the uncertainty in the model predictions. We will begin by describing the calibration data and how they were obtained, followed by appropriate likelihood and prior densities (as required by Eq. (3.2)).

The data set described in Green et al. (1986) consists of $n_L = 77$ measurements of reduced track lengths, r, for Durango apatite maintained at different isothermal temperatures (95–398 °C) for different periods of time from 20 min to 500 days. Our aim at this stage is to calibrate the parameterised annealing model $A(t, T, \boldsymbol{\beta})$ using these data and to assess the uncertainty in the modelling process. We shall refer to this data set as $L = \{L^{(t)}, L^{(T)}, L^{(r)}\}$, referring to the time, temperature, and track length reduction, respectively, for the original data set of Green et al. (1986).

At this point, we note the reported possible measurement error in the laboratory temperatures $L^{(T)}$, which is estimated by Green et al., 1986 to be between ± 1 and ± 3 °C. We shall build this information into the model subsequently. For the moment, we fix $L^{(T)}$ to the the reported temperature values. For all subsequent discussion, we fix $L^{(t)}$ to the reported laboratory times.

4.1. Bayesian formulation for the annealing model

Initially, therefore we are only concerned with the annealing model parameters $\boldsymbol{\beta}$ and the laboratory track length data $L^{(r)}$. The first task is to construct a valid likelihood function for this model and then specify the prior distributions on the unknown parameters.

As described earlier, the Laslett annealing model formulation has two steps. The first is a generalised Box–Cox transform of the annealed lengths $g(r)$ (see Eq. 2.4 and Box and Cox, 1964), followed by the fitting of a second, function of time and temperature, $B(T, t, c_0, c_1, c_2)$, which is linear in the unknown parameters (see Eq. (2.3)). The key assumption is that the errors on the transformed lengths $g(r)$ are independently normally distributed, with a constant variance, σ^2 (Laslett et al., 1987). However, we note that later work by Laslett and Galbraith, 1996 allowed for the observed variance to be included in the model formulation. Furthermore, the validity of this assumption has been questioned by Ketcham et al. (1999), who pointed out that the observations are not consistent with a constant variance (of the transformed data). Standard error propagation gives the ratio of the transformed variance to the variance of the reduced track length as

$$\frac{\sigma^2_{g(r)}}{\sigma^2_r} = \left|\frac{\mathrm{d}g(r)}{\mathrm{d}r}\right|, \tag{4.1}$$

where

$$\frac{\mathrm{d}g(r)}{\mathrm{d}r} = -r^{b-1}((1-r^b)/b)^{a-1}. \tag{4.2}$$

This shows that the variance ratio increases as r increases (or $g(r)$decreases). As stated by Ketcham et al. (1999), this implies that the longer track lengths potentially contain relatively little information or resolution on the model parameters.

Accepting there are limitations to the assumption of constant variance of the transformed data, we make this to allow direct comparison to the original model of Laslett et al. (1987). Then the posterior distribution is written as

$$p(\boldsymbol{\beta}, \sigma^2 | L^{(r)}) \propto p(L^{(r)} | \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}, \sigma^2) \tag{4.3}$$

and as $\boldsymbol{\beta}$ and σ^2 are a priori independent, we can write this as

$$p(\boldsymbol{\beta}, \sigma^2 | L^{(r)}) \propto p(L^{(r)} | \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}) p(\sigma^2). \tag{4.4}$$

Considering now the terms to the right of the proportionality sign, we can write the full likelihood of the data, using independent normal distributions on the transformed lengths, as,

$$p(L^{(r)} | \boldsymbol{\beta}, \sigma^2) = \prod_{k=1}^{n_L} \mathcal{N}\left(B\left(L_k^{(t)}, L_k^{(T)}, c_0, c_1, c_2\right), \sigma^2\right) J(a, b, L^{(r)}) \tag{4.5}$$

where $J(a, b, L^{(r)})$ is a Jacobian term, introduced to account for the Box–Cox transformation of the data, and is given by

$$J(a, b, L^{(r)}) = \prod_{k=1}^{p} \left| \frac{\mathrm{d}g\left(L_k^{(r)}, a, b\right)}{\mathrm{d}L_k^{(r)}} \right|, \tag{4.6}$$

which can be readily calculated for different values of $\boldsymbol{\beta}$.

The next requirement in the Bayesian formulation is the definition of the prior distribution over the annealing model parameters $\boldsymbol{\beta}$ and σ^2, $p(\boldsymbol{\beta})$ and $p(\sigma^2)$. Initially, we choose uniform priors over the components of $\boldsymbol{\beta}$, so that $p(\boldsymbol{\beta}) = U(\boldsymbol{\beta}_{\min}, \boldsymbol{\beta}_{\max})$, and thus have constant probability for each parameter existing between some upper and lower boundary. The uniform distributions are selected using the parameters of Laslett et al. (1987) as a guideline with a broad range (see Table 1).

For the prior on the regression variance, σ^2, we take an inverse gamma distribution so that

$$p(\sigma^2) = IG(\eta_1, \eta_2) \propto (\sigma^2)^{-(\eta_1+1)} \exp(-\eta_2/\sigma^2) \tag{4.7}$$

with the fixed parameters $\eta_1 = \eta_2 = 0.01$. This distribution guarantees a positive value, and also provides a diffuse prior, expressing our uncertainty in the value of σ^2. The inverse gamma is chosen as it is the conjugate prior for the normal distribution (see Appendix B for a brief discussion of conjugate priors). This choice enables us to later integrate out, or marginalise (see Bernardo and Smith (1994) for more details on marginalising parameters), σ^2 from the expression of the posterior distribution (see below). In practice, we do not need to sample σ^2, so we can effectively ignore it, treating it as a *nuisance* parameter, leading to more efficient sampling algorithms. We now have all we require to begin sampling the posterior distribution from Eq. (4.4).

4.2. Sampling strategy

We now describe how we generate new samples from the desired stationary posterior distribution $p(\boldsymbol{\beta}, \sigma^2|L^{(r)})$ using MCMC. Fortunately, as all subsequent work does not require samples of σ^2 for predictions and, we can integrate it out from the posterior, giving

$$p(\boldsymbol{\beta}|L^{(r)}) = \int p(\boldsymbol{\beta}, \sigma^2|L^{(r)})\, \mathrm{d}\sigma^2 \tag{4.8}$$

$$\propto (SS^{(\boldsymbol{\beta})}/2 + \eta_2)^{-(\eta_1+n_L/2)} p(\boldsymbol{\beta}) J(a, b, L^{(r)}), \tag{4.9}$$

where $SS^{(\boldsymbol{\beta})}$ is the residual sum of squares given by $\sum_{k=1}^{n_L}[g(L_k^{(r)}) - B(L_k^{(t)}, L_k^{(T)}, c_0, c_1, c_2)]^2$. For more details on the derivation of these expressions, see Appendix B.

After numerous attempts using Metropolis–Hastings MCMC with single parameter updates, it became clear that there is a high level of correlation between the annealing model parameters. Consequently, we chose to design a proposal function to account for these correlations directly and used a block update to state $\boldsymbol{\beta}_{i+1}$, so that all parameters were updated simultaneously and more efficiently. We used a multivariate normal distribution, centred on $\boldsymbol{\beta}_i$, with a full (positive definite) covariance matrix, Σ, giving

$$q(\boldsymbol{\beta}_{i+1}|\boldsymbol{\beta}_i) = \mathcal{N}_p(\boldsymbol{\beta}_i, \Sigma), \tag{4.10}$$

where $\mathcal{N}_p$ is a $p \times p$ dimension multivariate normal distribution (in the Laslett model $p = 5$).

The fixed values contained in the 5×5 matrix Σ, were derived by taking the empirical covariance of the independently sampled parameters (using a individual parameter updating scheme). This matrix was then scaled until a proposal acceptance rate of around 25% was achieved.

The MCMC sampler was then used to draw 10^5 samples from $p(\boldsymbol{\beta}|L)$, before taking every 10th sample to encourage sample independence. We experimented with longer sampling runs, but it was clear the chain was effectively stationary after 10^5 samples. The annealing model samples and posterior distributions can be found in Fig. 1. For comparison purposes, we have summarised the Laslett annealing model parameters $\boldsymbol{\beta}_{\mathrm{Las}}$, along with the values $\boldsymbol{\beta}_{\mathrm{MAP1}}$, from the MAP model sampled during the MCMC runs in Table 1. The MAP model has a slightly higher log-likelihood than the original Laslett model, $\boldsymbol{\beta}_{\mathrm{Las}}$ although the parameterisation is the same. As can be inferred from Fig. 1, parameter c_2 is relatively uncorrelated with the others, but we show examples of the strong correlation between the other parameters in Fig. 2. These parameters have absolute correlation coefficients greater than 0.87. The regions plotted in this figure broadly correspond to the 95% credible regions for each pair of parameters.

We show the quality of fit to the laboratory annealing data, $L^{(r)}$, in Fig. 3, which shows no significant outliers or trends in the residuals. The 95% credible intervals (as indicated by the displayed error bars in Fig. 3(a)), demonstrate increasing levels of uncertainty as track length reduce, whilst still generally capturing the desired 1:1 relationship. Although a wide range of parameters have been sampled (see Fig. 1), the quality of fit to the data remains consistent and is effectively the same as that obtain by Laslett et al. (1987).

5. Incorporating uncertainty in the laboratory temperature data

Given there were errors in the recorded laboratory temperatures of ±1 to ±3 °C (Green et al., 1986), it is desirable to include this knowledge in the model calibration. This is straightforward in the Bayesian framework by using a hierarchal model (see Appendix A). We then have three sets of parameters; $\boldsymbol{\beta}$ which as before is the vector of annealing model parameters; $\boldsymbol{T}$ as the $n_L \times 1$ vector of true (but unknown) laboratory temperatures, and λ as a parameter used to estimate the error in temperature measurements.

Having integrated out the regression variance, σ^2, as described earlier, we can write the posterior distribution for the laboratory data (cf. (4.4)) as

$$p(\boldsymbol{\beta}, \boldsymbol{T}, \lambda|L^{(r)}) \propto p(L^{(r)}|\boldsymbol{\beta}, \boldsymbol{T})p(\boldsymbol{\beta})p(\boldsymbol{T}|\lambda)p(\lambda), \tag{5.1}$$

where we have opted for a hierarchal prior over $p(\boldsymbol{T}, \lambda)$. As we can simply substitute $\boldsymbol{T}$ for $L^{(T)}$ in the laboratory data likelihood, all that remains is to define a suitable form for the prior density $p(\boldsymbol{T}—\lambda)p(\lambda)$. The temperature data may have both random and systematic errors, reflected in the precision and accuracy, respectively. Here, we do not deal with systematic errors, which may arise if a thermistor is incorrectly calibrated for example. In principle, it would not be difficult to deal with systematic errors, in that we could specify a prior on the magnitude of this, relative to the observed temperature. In the absence of information concerning such errors, we do not pursue this. Here, we make the assumption that the errors arise in the precision of the temperature measurements and are normally distributed. We assume the distributions are centred on the originally measured temperatures $L^{(T)}$, with constant variance λ.

For $p(\lambda)$ we opt for a prior conjugate to the normal distribution, i.e., the inverse gamma with $p(\lambda) = IG(\epsilon_1, \epsilon_2)$. This gives up to a constant of proportionality (which we do not need to know for MCMC sampling)

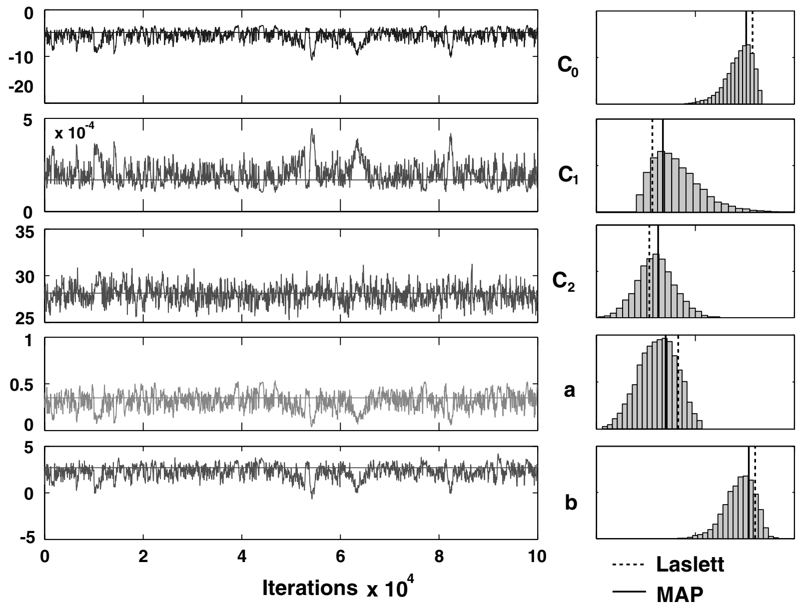


Fig. 1. Post-burn in MCMC samples and posterior distributions for each of the annealing model parameters in $\boldsymbol{\beta}_{\mathrm{MAP1}}$.

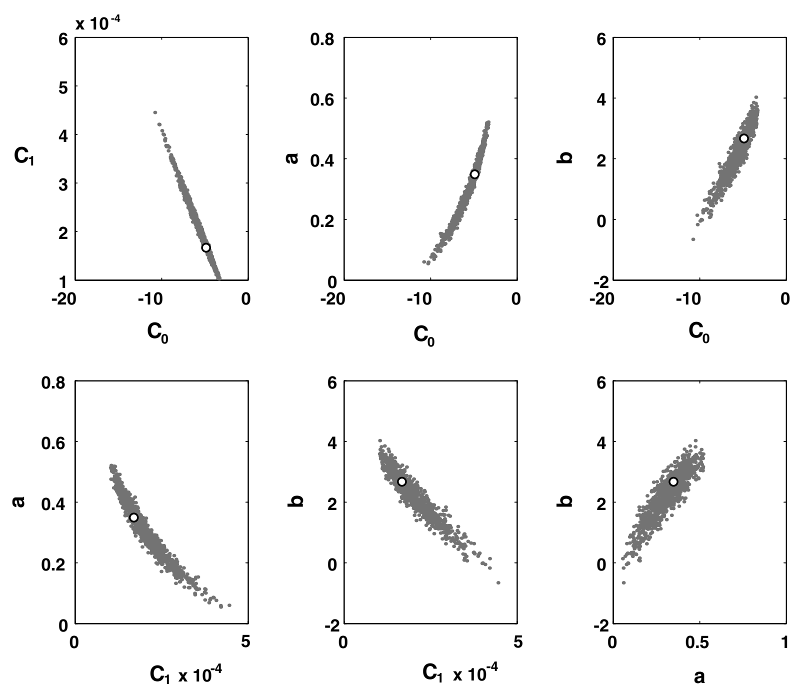


Fig. 2. Correlations between a selection of annealing model parameters $\boldsymbol{\beta}$, inferred from MCMC sampling. The MAP model is identified for each parameter pair by the white circle.

$$p(\boldsymbol{T}|\lambda)p(\lambda) \propto \lambda^{-(\epsilon_1+1)} \exp(-\epsilon_2/\lambda) \prod_{k=1}^{n_L} \mathcal{N}\left(L_k^{(T)}, \lambda\right). \tag{5.2}$$

The values of ϵ_1 and ϵ_2 are set to values of 0.5 and 6, respectively. These values produce a distribution for λ with a mode of 4 and so provide a realistic constraint over the sampling range of temperature error permitted, assumed to be up to ±10 °C. This range is somewhat larger than that quoted by Green et al. (1986), but allows for a greater flexibility in the sampling.

5.1. Sampling the laboratory temperatures

Having added $\boldsymbol{T}$ and λ to our model, we must now determine how these will be sampled. As for σ^2 previously, we choose a conjugate form for the prior distribution on λ and so can integrate it out analytically. We then obtain,

$$p(\boldsymbol{T}|L^{(r)}, \boldsymbol{\beta}) \propto p(L^{(r)}|\boldsymbol{\beta}, \boldsymbol{T}) \int p(\boldsymbol{T}|\lambda)p(\lambda)\,\mathrm{d}\lambda \tag{5.3}$$

$$\propto p(L^{(r)}|\boldsymbol{\beta}, \boldsymbol{T})(SS^{(T)} + \epsilon_2)^{-(\epsilon_1+n_L/2)}, \tag{5.4}$$

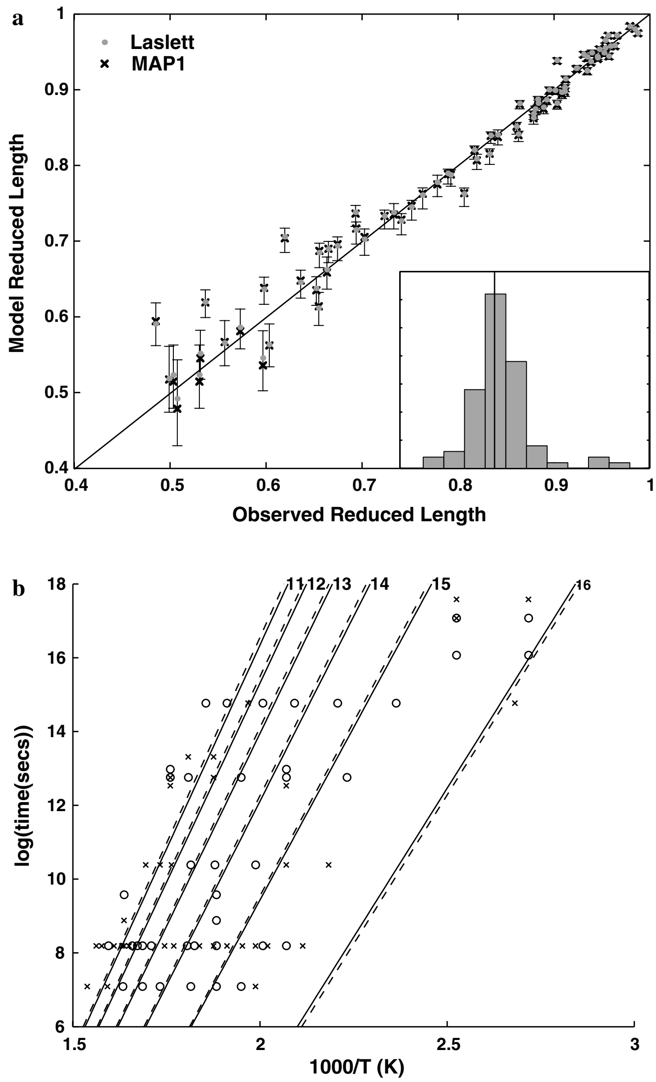


Fig. 3. (a) Length reduction prediction compared with measurements for $\boldsymbol{\beta}_{\mathrm{MAP1}}$ (based on laboratory annealing data) and Laslett parameters sets, with error bars representing 95% credible intervals. The inset shows the distribution of residuals (predicted – observed) for the reduced tack lengths. The vertical line in the inset indicates a residual of zero. (b) Arrhenius plot showing the observed data and predicted contours of constant track length (indicated by the number at the top in μm) for $\boldsymbol{\beta}_{\mathrm{MAP1}}$ (solid line) and Laslett (dashed line) models. The circles indicate those values where the predicted length is less than the observed value, and the crosses indicate where the predicted length is greater than the observed value.

where $SS^{(T)} = \sum_{k=1}^{n_L}(L_k^{(T)} - T_k)^2$, i.e., the residual sum of squares over our temperature predictions (see Section B.2).

Furthermore, as Eq. (2.3) can be written as a linear basis function model, with a matrix having n_L rows, so that the kth row is given by $\boldsymbol{E}_k = [1, T_k \ln L_k^{(t)}, T_k]$, we can also marginalise c_0, c_1 and c_2 (see Section B.2). As the $p(\boldsymbol{\beta})$ term is constant, it can been dropped and we still maintain the proportionality. This gives a final density of

$$p(\boldsymbol{T}|L^{(r)}, a, b) \propto |V|^{-1/2}\left(\eta_2^*\right)^{-(n/2-k/2+\eta_1)}\left(SS^{(\lambda)} + \epsilon_2\right)^{-(\epsilon_1+n/2)} \tag{5.5}$$

with the terms as defined below

$$V = (\boldsymbol{E}'\boldsymbol{E})^{-1}, \tag{5.6}$$

$$\eta_2^* = \eta_2 + (\boldsymbol{f}\,'\boldsymbol{f} - \hat{\boldsymbol{c}}'V^{-1}\hat{\boldsymbol{c}})/2, \tag{5.7}$$

$$\hat{\boldsymbol{c}} = (\boldsymbol{E}'\boldsymbol{E})^{-1}\boldsymbol{E}'\boldsymbol{f}, \tag{5.8}$$

where $\boldsymbol{f}$ refers to a $(n_L \times 1)$ vector, containing the Box–Cox transforms $g(a, b, L^{(r)})$.

Using the marginalised density $p(\boldsymbol{T}|L^{(r)}, a, b)$, we are in a position to use Metropolis–Hastings MCMC to sample $\boldsymbol{T}$. To maintain simplicity and algorithm efficiency, we choose to use a full block update on the temperature parameters (but note these are independent). We draw new values via

$$\boldsymbol{T}_{i+1} \sim \mathcal{N}(\boldsymbol{T}_i, v_{\boldsymbol{T}}\boldsymbol{I}), \tag{5.9}$$

where $\boldsymbol{I}$ is a $(n_L \times n_L)$ identity matrix, and $v_{\boldsymbol{T}}$ a scalar parameter controlling the size of each Metropolis–Hastings step. This was again tuned through multiple runs choosing different values of $v_{\boldsymbol{T}}$ to provide an acceptance rate of around 25–30%.

Using this sampling strategy, we ran the algorithm for 10^6 iterations, before thinning the chain by taking every 100 samples. The values for the MAP model, $\boldsymbol{\beta}_{\mathrm{MAP2}}$, are summarised in Table 1. In Fig. 4, we plot a histogram of the lab temperature residuals $T_i^k - L^k$, for each of the 77 measurements. These are centred around 0 with an average standard deviation of 7.8 °C, and have a 95% credible interval spanning −16.2 to 16.7 °C and a 90% interval spanning −11.3 to 11.6 °C. This range could be reduced by changing the prior values ϵ_1 and ϵ_2 if required. However, the sampled temperature values ($\boldsymbol{T}$) for the MAP model are also plotted in Fig. 4. These are generally within 3 °C of the reported values, similar to the resolution cited by Green et al. (1986).

Having incorporated the uncertainties in the laboratory temperatures, we can see the increase in predictive uncertainty by looking at the credible intervals over the predicted reduced track lengths in Fig. 5a. As before, however, the

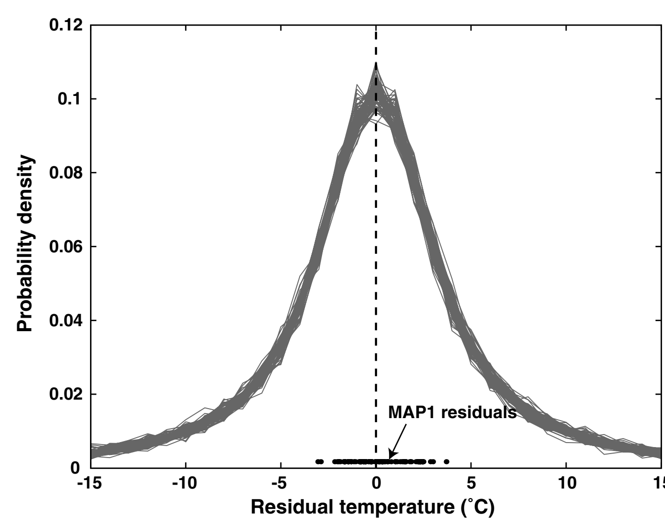


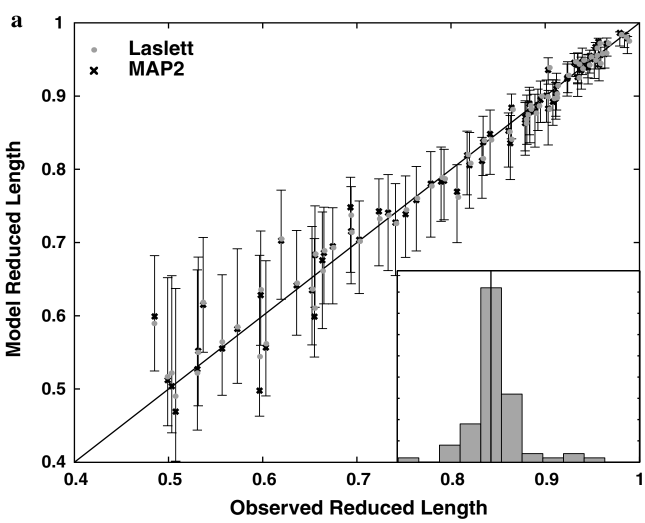
Fig. 4. Distribution of the residuals from the MCMC temperature sampling. The individual residuals for the MAP2 model are also shown on the x-axis.

distribution of the predictions is still centred over the expected 1:1 relationship. Despite the large increase in uncertainty, the MAP and Laslett predictions are still close together and show strong correlation to the original measurements. Thus, despite greatly increasing the parameter space by adding $\boldsymbol{T}$, our sampling methodology over $\boldsymbol{\beta}$ remains consistent with the observed data.

6. Extrapolation of revised laboratory models to geological times

Having looked at the consistency of the Bayesian annealing models with the laboratory data, we will now analyse how well the new $\boldsymbol{\beta}$ samples extrapolate to geological timescales. In Ketcham et al. (1999), a high-temperature benchmark is given, based on reported fission track data from volcanogenic sediments in the Otway Basin, SE Australia. Citing an article by Paul Green in an informal publication (OnTrack newsletter, No. 11) from 1995, they state that Durango apatite (0.4wt% Cl) are close to being fully annealed for (present day) temperatures in the range 95–100 °C. From Green's article, it seems also that by 111 °C, apatites with 0.7% weight Cl are totally annealed (have zero fission track age). The total annealing temperature depends on the definition of the length at which the age goes to zero. Ketcham et al. (1999) define the total annealing temperature (T_{F}) as that required to reduce the track length to below $r = 0.41$, the shortest observed reduced track length over a given time interval. We have not formally used this information in the annealing model calibration process, as the thermal histories of the Otway Basin samples are not particularly well constrained. However, to demonstrate the behaviour of these models relative to this high-temperature benchmark, we show in Fig. 6 the 95% credible intervals for the amount of annealing tracks experience over 30 Ma, at different temperatures, for the $\boldsymbol{\beta}_{\mathrm{MAP}}$ (the parameters maximising Eq. (4.9) and also $\boldsymbol{\beta}_{\mathrm{MAP}}$ using the parameters maximising Eqs. (4.9) and (5.5) (Fig. 6, i.e., the laboratory annealing data neglecting and including the uncertainty on the laboratory temperatures, respectively)). We also show the predicted curve using the Laslett parameters $\boldsymbol{\beta}_{\mathrm{Las}}$.

Fig. 6 shows the wide variety of solutions that exist within the credible intervals, giving a range of 96–115 and 96–133 °C for the 95% credible intervals. These overlap the range of temperatures described in Ketcham



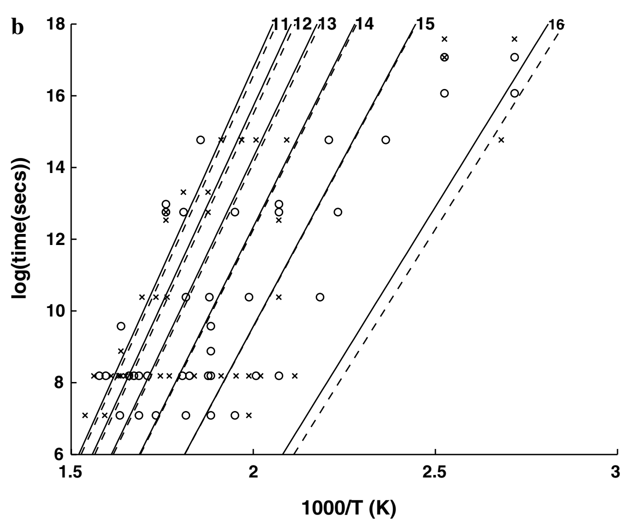


Fig. 5. (a) Length reduction prediction compared with measurements for $\boldsymbol{\beta}_{\mathrm{MAP2}}$ (based on laboratory annealing and laboratory temperature data) and Laslett parameters sets, with error bars representing 95% credible intervals. The inset shows the distribution of residuals (predicted – observed) for the reduced tack lengths. The vertical line in the inset indicates a residual of zero. (b) Arrhenius plot showing the observed data and predicted contours of constant track length (indicated by the number at the top in μm) for $\boldsymbol{\beta}_{\mathrm{MAP2}}$ (solid line) and Laslett (dashed line) models. The circles indicate those values where the predicted length is less than the observed value, and the crosses indicate where the predicted length is greater than the observed value.

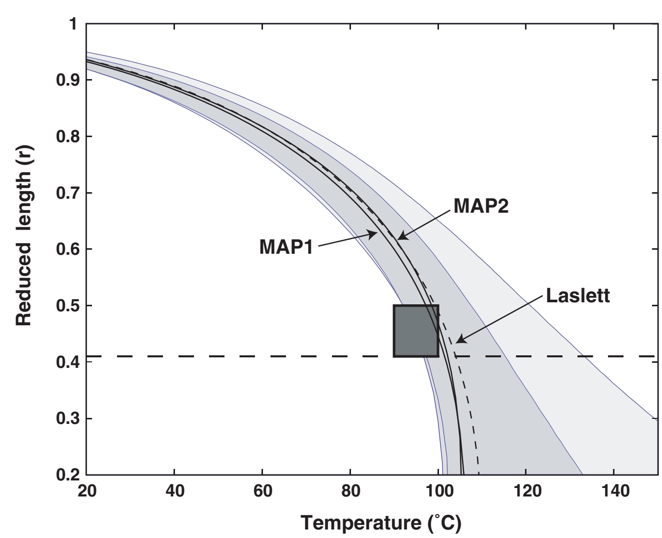


Fig. 6. Predictions and 95% credible intervals of the reduced track lengths with Laslett, MAP1 and MAP2 models extrapolated to geological timescale (30 Ma), and compared to expected range of results (shaded box, after Ketcham et al., 1999). MAP1 ignores the temperature uncertainties while they have been explicitly included for MAP2. The credible intervals for the MAP2 model are broader than for MAP1.

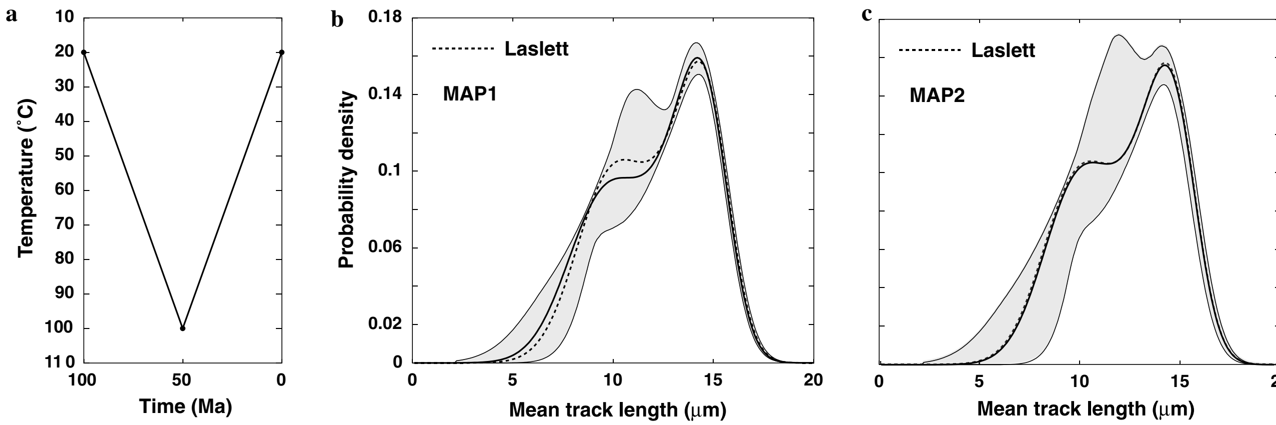


Fig. 7. (a) Simple thermal history used in forward model, (b) 95% credible intervals of annealing model predictions using thermal history in (a), with Laslett (dotted line) and MAP1 predictions, (c) as (b), but with MAP2 predictions

et al., 1999 although are on the higher end of their range. Also, the predictions from Laslett and MAP samples are fairly similar, particularly when we consider the temperature uncertainties.

With the uncertainty measures available from the Bayesian models, we can also use the posterior samples of the model parameters to predict the range in track length distributions using the simple thermal history shown in Fig. 7(a). The 95% credible intervals for this posterior predictive density are found in Fig. 7(b) and (c), with the predicted distribution using the Laslett and MAP models overlain.

This clearly demonstrates that the range of $\boldsymbol{\beta}$ samples permitted by the laboratory data provide a broad range of track length distributions. All of this added variability implies greater flexibility in the annealing model parameters than previously considered, and therefore greater uncertainty in the predictive modelling of fission track data.

7. Bayesian annealing model calibrated from laboratory and geological data

In order to assess the influence of including low-temperature geological constraints when estimating the model parameters with the laboratory annealing data, we use the data given in Vrolijk et al. (1992), from sample MB7, obtained courtesy of Ray Donelick. This data set comprises an analysis of a deep-sea drill core, which include volcanogenic sediments with apatite similar in composition to Durango apatite, and appears to have experienced an exclusively low-temperature (<25 °C) history since deposition some 120 Ma. It is these data that Ketcham et al., 1999 refer to as their low-temperature benchmark and were used by Ketcham et al. (1999) to test whether laboratory defined annealing models could be extrapolated to geological time scales, though they did not formally use this information in the estimation of the annealing model parameters. The data are summarised in Fig. 8, showing the geologically derived thermal history discussed by Vrolijk et al. (1992) in (a), and the measured track length distribution in (b). The data are high quality, consisting of 195 track length (TINT) measurements, yielding an average track length of 14.5 μm, and 45 single grain counts, giving a central age of 125 Ma. We refer to these new data as V, and include this dataset into our Bayesian model (see Appendix A), modifying our posterior predictions of the annealing model parameters, and sampling the range of possible thermal histories shown in Fig. 8(a).

7.1. Adpating the laboratory-based Bayesian model

To incorporate the geological timescale data of Vrolijk et al. (1992), V and the associated thermal history, we write the full posterior distribution for the new model as

$$p(L, V|\boldsymbol{\beta}, \boldsymbol{\phi}, \lambda, \boldsymbol{T}, \sigma^2) \propto p(V|\boldsymbol{\phi}, \boldsymbol{\beta})p(L|\boldsymbol{\beta}, \boldsymbol{T}, \sigma^2)p(\boldsymbol{T}|\lambda)p(\lambda)p(\boldsymbol{\beta})p(\boldsymbol{\phi})p(\sigma^2), \quad (7.1)$$

where we have added $p(V|\boldsymbol{\phi}, \boldsymbol{\beta})$, which is the likelihood of the Vrolijk data, conditioned on the annealing model parameters, and the geological thermal history $\boldsymbol{\phi}$ for these data, which requires a new prior term $p(\boldsymbol{\phi})$. The remainder of the terms have not changed. In the sections to follow we shall describe these terms, and how we sample from them. The remaining terms are as previously described.

The likelihood term for the geological data, $p(V|\boldsymbol{\phi}, \boldsymbol{\beta})$, is fully described in Gallagher (1995), and takes into account the observed individual crystal spontaneous and induced track counts, and how well the annealing model, $A(t, \boldsymbol{T}, \boldsymbol{\beta})$, describes the observed track lengths. It is amended slightly to allow dependence on the annealing parameters which we allow to change during the sampling.

We parameterise the thermal history, $\boldsymbol{\phi}$, in terms of time–temperature nodes, joined with linear segments as described in Stephenson et al. (2006) (using a Bayesian approach) and Gallagher (1995) (using optimised maximum likelihoods), Although other parameterisations are available (e.g., splines), by using nodes we can easily include a priori knowledge, e.g., the temperature or temperature

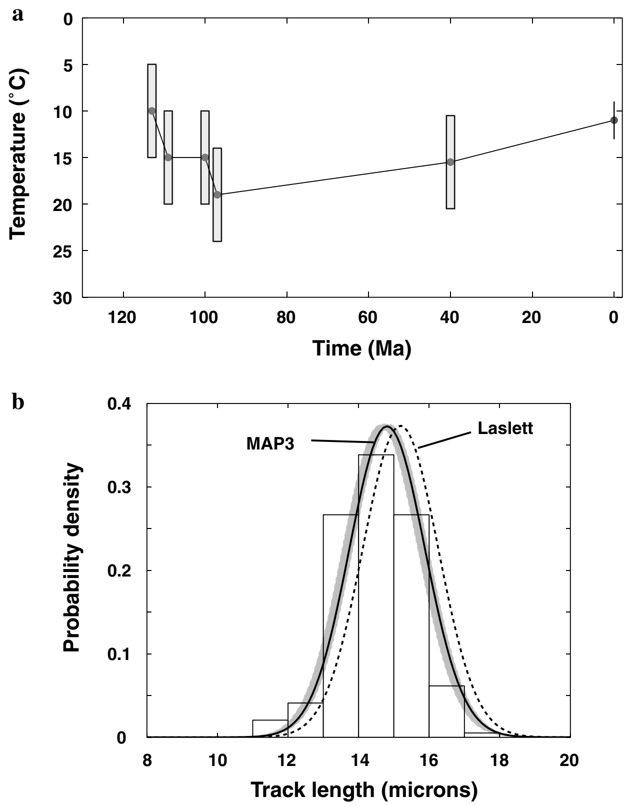


Fig. 8. (a) Geologically constrained thermal history, with uniform prior boundaries. (b) The observed track length distribution, together with the predictions from the Laslett and MAP3 models. Both taken from Vrolijk et al. (1992).

range at a given time. This is also the method used in the majority of work published on thermal histories, allowing easy comparisons. Here, we fix the number of nodes to be 6, and place uniform priors on each time and temperature node, with widths driven by any priori geological knowledge. This includes uncertainty on whether the observed rifting was Jurassic or Cretaceous, consistent with the thermal histories described by Vrolijk et al. (1992) (see Fig. 8(a)). In this case the prior information we have is relatively well defined, and will therefore influence the posterior distribution, leading to more geologically relevant samples from $\boldsymbol{\beta}$.

We note in passing that the Bayesian framework allows us to have the number of nodes as another parameter, using Reversible-Jump MCMC (Green, 1995) to sample across this variable dimensional parameter space. This would provide a fully parsimonious thermal history to be derived, with its complexity governed by the data, rather than subjective opinion or iterative solutions.

7.2. Sampling the geological timescale models

Again, we use Metropolis–Hastings sampling and the required marginal distribution, after removing terms that do not directly influence the geological data (and their prediction), is given by

$$p(\boldsymbol{\phi}|L^{(r)}, V, \boldsymbol{\beta}, \boldsymbol{T}, \lambda, \sigma^2) = p(\boldsymbol{\phi}|V, \boldsymbol{\beta}) \tag{7.2}$$

$$\propto p(V|\boldsymbol{\phi}, \boldsymbol{\beta})p(\boldsymbol{\phi}) \tag{7.3}$$

and this forms the acceptance probability in the MCMC sampler. Although correlations are likely to exist for individual time–temperature nodes, in practice it was not necessary to include these specifically in the sampling strategy. As such we initially use a univariate normal proposal distribution for each parameter in $\boldsymbol{\phi}$, centred on the current value for the parameter and one value of the variance (v) for time parameters, and another for temperature parameters. From repeated experiments, it was found that a value of $v_{time} = 2$ Ma and $v_{\mathrm{temp}} = 1$ °C provide reasonable proposal acceptance rates. In subsequent runs, we used a block update over all of the available thermal history parameters.

After a run of 10^6 iterations, we use our sampled values of $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, to run the forward model and compare these to the observed data, as in Fig. 8(b). Along with the 95% credible intervals, we plot the the MAP result, and the distribution produced using $\boldsymbol{\beta}_{\mathrm{Las}}$. The values of $\boldsymbol{\beta}_{\mathrm{MAP3}}$ are given in Table 1. The MAP result has clearly improved on the track length density prediction made by $\boldsymbol{\beta}_{\mathrm{Las}}$. With this result we can see an MAP mean length of 14.77, with 95% credible intervals giving a range of 14.56–14.86, compared to the observed mean of 14.5 μm. $\boldsymbol{\beta}_{\mathrm{Las}}$ predicts a longer mean length of 15.14 μm, as noted previously by Ketcham et al. (1999).

An additional check on this model is to examine the quality of fit to the laboratory data as described in Section 4. In Fig. 9 we compare laboratory measured annealed lengths, with those predicted using the samples from $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$. The numerical fit to the observed lengths is not as good as the original Laslett model based solely on the laboratory data, and there is a perhaps a tendency to underpredict the degree of annealing for shorter lengths, and the two sets of model predictions diverge at the longest lengths more than the earlier results. There are no lengths less than 11.3 μm (or reduced length of 0.69) in the geological dataset. The predictions of the two models in the Arrhenius plot (Fig. 9(b)) crossover in the range of the data. However, the general trend of the residuals is still fairly symmetrical, particularly when the 95% credible interval range is considered relative to the 1:1 line. As we are deal with probabilities, there is no need to specify an explicit weighting between the geological and laboratory timescale datafits. The degraded quality of the fit to shorter lengths relative to the earlier models (MAP1, MAP2) reflects the information mapped probabilistically from the geological timescale data into the model parameter estimates. However, the now degraded fit to the laboratory data implies that the 2 datasets may be inconsistent. Clearly, we need to consider more datasets (both laboratory and geological), with a greater range in the track length measurements to assess this appropriately.

Finally, the residuals of the temperature data ($L_k^{(T)} - T_k$) for the 77 laboratory temperature measurements have a 95% credible interval of about ±13 °C, and exhibit no obvi-

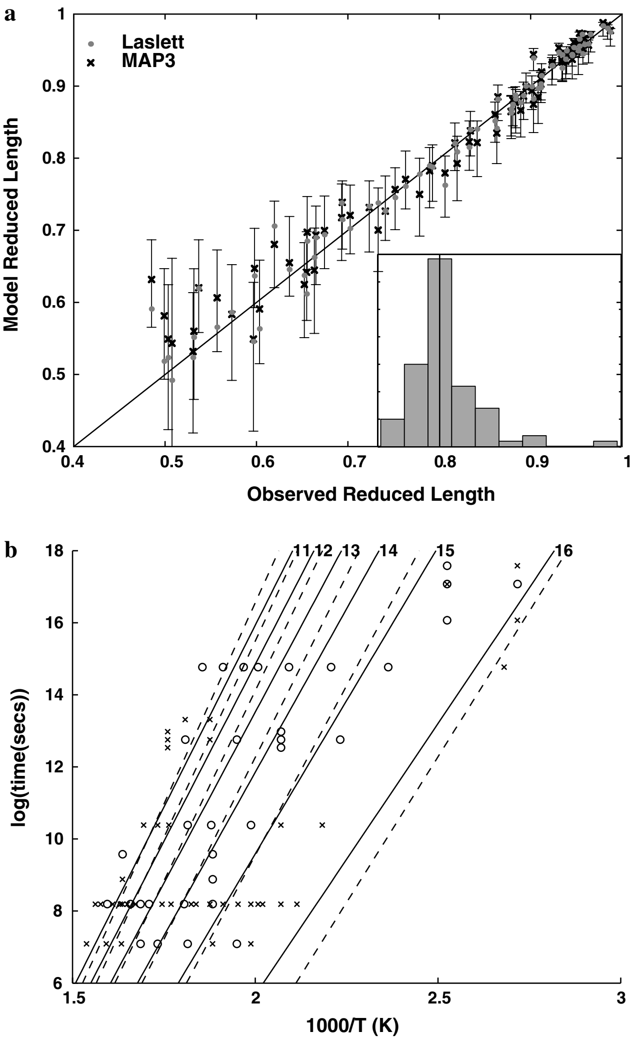


Fig. 9. (a) Length reduction prediction compared with measurements for $\boldsymbol{\beta}_{\mathrm{MAP3}}$ (based on laboratory annealing and temperature data, and the geological timescale fission track data of Vrolijk et al., 1992) and Laslett parameters sets, with error bars representing 95% credible intervals. The inset shows the distribution of residuals (predicted – observed) for the reduced tack lengths. The vertical line in the inset indicates a residual of zero. (b) Arrhenius plot showing the observed data and predicted contours of constant track length (indicated by the number at the top in μm) for $\boldsymbol{\beta}_{\mathrm{MAP3}}$ (solid line) and Laslett (dashed line) models. The circles indicate those values where the predicted length is less than the observed value, and the crosses indicate where the predicted length is greater than the observed value.

ous bias. The distributions of these residuals are not shown, but are similar to those shown in Fig. 4. Again the majority of the residuals for the MAP model were <±3 °C.

7.3. Extrapolation to geological times

When we extend the track length prediction out to geological times (as described in Section 6), we can compare the new MAP and credible range with the temperature range for the high-temperature benchmark given by Ketcham et al. (1999). The MAP fully annealed temperature lies at 96 °C, as compared to original Laslett model prediction of 104 °C. The reduction in uncertainty is also clearly apparent on comparing Fig. 6 with that of Fig. 10, with new 95% credible intervals of 88–112 °C. Having increased the amount of data that the model is conditioned on, the level of uncertainty, and range of credible models in the forward prediction is greatly reduced.

Finally, we return to the problem of rapid recent cooling mentioned in the introduction. As an example, we use sample (94-70) from northeast Brazil (Harman, 2000). This sample has the characteristics of those which, when modelled with the original Laslett model, tend to imply recent cooling, typically from around 60 °C. This sample has an apatite fission track age of nearly 200 Ma (based on 20 crystals), and a mean track length of 11.6 μm (based on 65 track length measurements). Similar data and thermal history features have been described from Australia (Kohn et al., 2002), Ireland (Allen et al., 2002), and India (Gunnell et al., 2003). Their approach to alleviate this rapid cooling was to reduce the initial track length in the Laslett model.

Here, we have used the approach described by Gallagher (1995), sampling 5000 thermal histories, with a genetic algorithm, using both the original Laslett model, and the MAP model parameters determined from both the laboratory and geological data, as described above, using an initial track length of 16.3 μm in both cases. We show the results of this exercise in Fig. 11. Here, we can clearly see the rapid, recent cooling in the inferred thermal histories using the Laslett model. This cooling starts from around 60 °C at 10 m.y. (the youngest time allowed by the bounds specified on the input thermal history). The revised model parameters imply a cooling history in which the rate also increases. However, the change in rate starts around 50–70 Ma, and is <0.5 °C per m.y. The cooling rates at higher temperatures are similar for both models. Therefore, the incorporation of the low-temperature, geological timescale

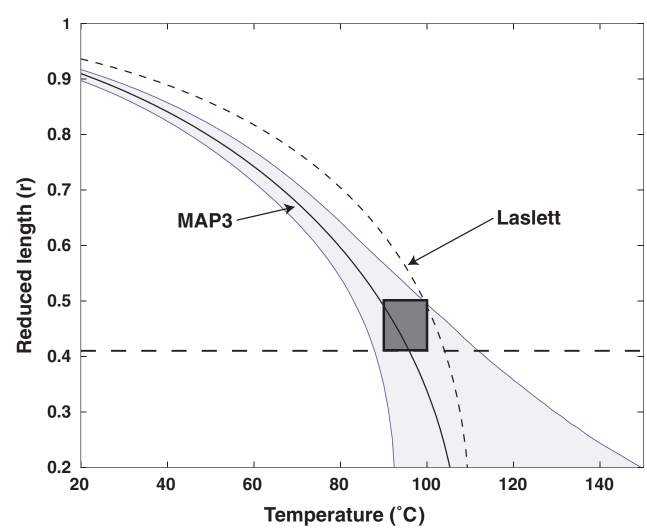


Fig. 10. Predicted reduced track length and 95% credible intervals, for MAP3 extrapolated to geological times (30 Ma).

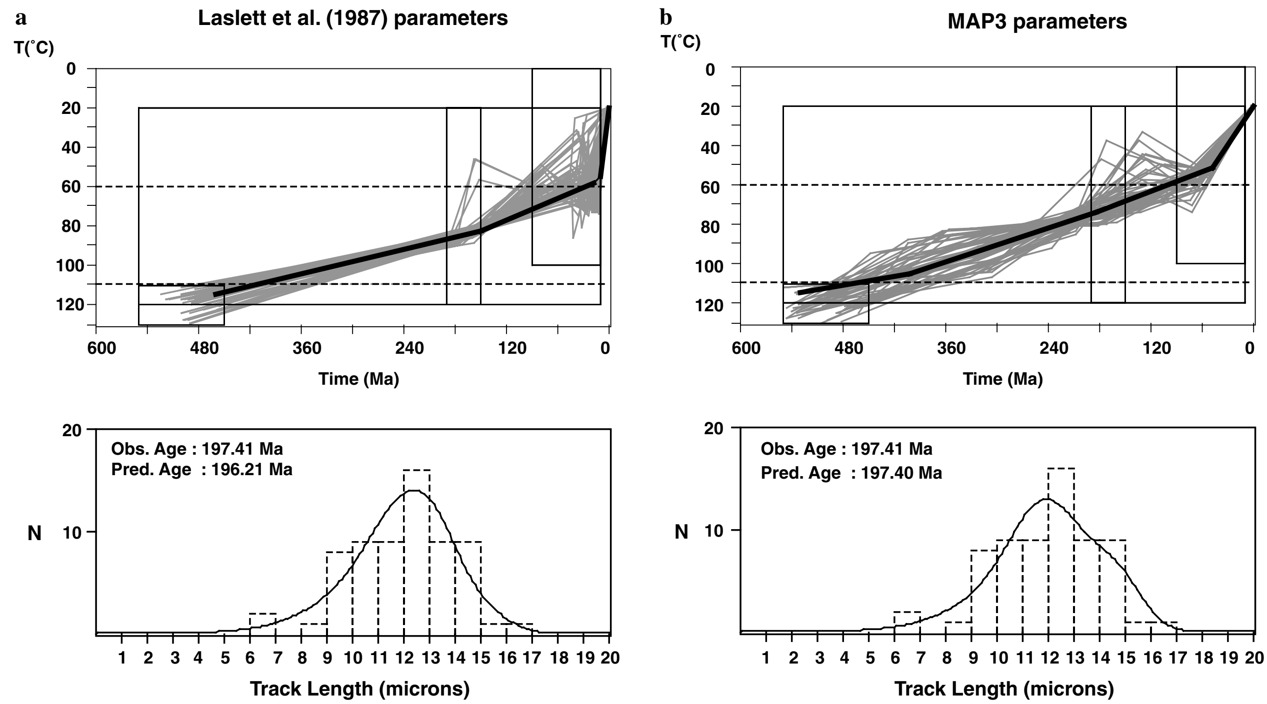


Fig. 11. Inferred thermal histories from sample 94-70 from northern Brazil using (a) the Laslett et al. (1987) model, and (b) MAP3 model. The grey boxes indicate the range of time–temperature parameters sampled. The maximum likelihood thermal history is shown as a heavy line, and the other thermal histories are effectively within the 95% confidence interval, as defined by the difference in the log likelihoods from the best model (see Gallagher, 1995).

constraints appears to result in an annealing model (if we choose the MAP sample) that reduces this apparent artifact considerably. Of course, we do not know the 'true' geological thermal history in this case, but the lower rate of cooling intuitively seems more reasonable.

8. Conclusions

In this paper, we have presented a fully Bayesian framework and methodology for analysing fission track data and thermal history models on both laboratory and geological timescales. Rather than accept a single fixed set of annealing model parameters, we opt to continually sample from their posterior densities whilst also sampling the thermal history. This allows us to guide our choice of annealing model using all the available conditioning data, rather than solely base our results on the fit to laboratory derived data.

To this end, we have gradually built up a hierarchal model, increasing the amount of data and constraints that it is conditioned on. The most complete model we presented incorporates uncertainty in the annealing model and laboratory track length dat and temperature measurements, whilst also conditioning on a geologically constrained low-temperature data set. Using a methodology incorporating MCMC, we have been able to access the true uncertainty in the modelling procedure, and assess the quality of our results. This is clearly shown in the decrease of uncertainty, as the amount of data included in the model increases.

The results of this study demonstrate that by using all of the available data, we are able to extend the model robustly out to geological time scales, maintaining a good fit to the expected range of laboratory temperatures, while also improving the fit to a low-temperature benchmark. The fit to the original laboratory track length data is degraded when we condition jointly on the laboratory and geological timescale data. This may indicate that these two datasets are incompatible, and/or that the nature of the annealing process is different over these two timescales. However, when we applied the final MAP model to geological data whose thermal history is unknown, the apparent rapid recent cooling seen in many thermal history models is significantly reduced, without the need to reduce the initial track length.

In principle, it is straightforward to incorporate the unknown thermal history of a geological sample in the Bayesian model formulation rather than using just the MAP model. However, given the low quality of prior information, this may not be such a useful approach unless the other prior information (e.g., on the annealing model is fairly tight). The general Bayesian approach as presented here does readily provide a framework for the predictive uncertainty when modelling geological data, by allowing for the uncertainty inherent in the laboratory calibration (e.g., see Fig. 7). In modelling more than one geological data set, we can use the same general formulation, sample the same annealing model for all data, or alternatively we could allow the annealing model to vary between samples. This reflects the underlying philosophy of the Bayesian approach, i.e., that we condition on the data, rather than a single model. Given that the output of this approach can be in terms of the probability distributions on all unknown parameters, including the annealing model and the laboratory and geological thermal histories, then this allows a fully probabilistic assessment of the model results.

Although we have focussed on the Laslett et al. (1987) model, the general approach is readily applicable to other data sets (e.g., Carlson et al., 1999; Barbarand et al., 2003a), and model formulations (e.g., Ketcham et al., 1999). Furthermore, well-designed geological case studies over a range of time and temperature scales would provide more constraints for the extrapolation process, and we fully concur with the statement of Ketcham et al. (1999) that there is a great need for a more extensive database of reliable geological case studies. Low-temperature studies are of particular interest for understanding long-term denudation and landscape evolution, and also for improving our understanding of low-temperature helium diffusion in apatite over geological timescales. In future work, we also plan to extend the Bayesian methodology to allow uncertainty in the number of control points when defining the thermal history. This will be achieved using reversible-jump MCMC, which allows trans-dimensional jumps in model space (Green, 1995). Additionally, this work can be fully incorporated into the Bayesian methodology described in Stephenson et al., 2006 for incorporating multiple samples in three dimensions.

Acknowledgments

We thank Ray Donelick for providing the MB7 data in good spirit, and Richard Ketcham, two anonymous reviewers and A.E., Rick Ryerson, for useful critique of an earlier version of this manuscript.

Associate editor: F.J. Ryerson

Appendix A. Graphical representation for the Bayesian model formulation

The relationships between the data and the unknown parameters can be represented using a graphical model, constructed using probability theory (see Jordan, 2004; Spiegelhalter et al., 1996). A directed acyclic graph (DAG) for the models we consider in this paper is shown in Fig. A.1 (note that here we have ignored the data relevant to the duration of the laboratory experiments, $L^{(t)}$). In this representation the quantities in the model are represented as nodes on the graph which interact in different ways. These quantities can be data (shown as a square), or variables (shown as a circle). The interactions between the different quantities are shown as arrows, and the direction of an arrow indicates the conditional dependence of one element on another. The node from which the arrow starts is known as the parent, while the node at the end of the arrowhead is the descendant. If there is no direct connection between two nodes then they are independent or conditionally independent (that is, they may be indirectly linked by dependence on an intermediate node). Thus, if

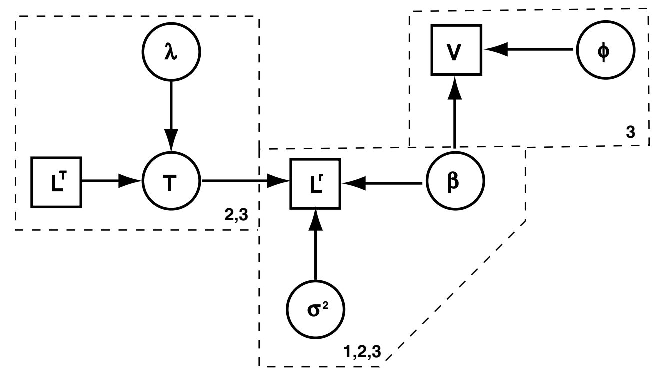


Fig. A.1. Graphical representation of the model structure used in this paper. Data are represented by squares, and variables by circles, and the arrows indicate dependence (with a parent influencing a descendant in the direction of the arrow). The terms in boxes and circles directly represent those defined in the main text. The dashed lines define the groups of data/variables used at different stages of the Bayesian model formulation, as indicated by the numbers in the bottom right of each region. See text in Appendix A for further explanation.

we know the values of the parents of a given node, n, then no other node except the descendants of n can provide information on the value of the variable at n. Quantities that act only as parents have no conditional dependence, except (if appropriate) through the prior distribution. This let us readily write down the full joint distribution of the quantities. Then we can express this in terms of the conditional probabilities and then derive the appropriate form of Bayes theorem for the parameters and data.

The first model we consider in Section 4.1 is shown in the DAG by the sub-region on the graph identified by 1, and we do not need to consider any arrows that cross the boundary of this region. Here, the quantities are $\boldsymbol{\beta}$, σ^2 and the laboratory track length data, $L^{(r)}$. Using the arrows to infer the dependence, we can write the full joint distribution in terms of the conditional and prior distributions, i.e.,

$$p(L^{(r)}, \boldsymbol{\beta}, \sigma^2) = p(L^{(r)}|\boldsymbol{\beta}, \sigma^2)p(\boldsymbol{\beta})p(\sigma^2). \tag{A.1}$$

The first term on the righthand side is just the data likelihood, and from Bayes theorem, we can write the posterior for the unknown parameters as

$$p(\boldsymbol{\beta}, \sigma^2|L^{(r)}) \propto p(L^{(r)}|\boldsymbol{\beta}, \sigma^2)p(\boldsymbol{\beta})p(\sigma^2). \tag{A.2}$$

The second model we consider in Section 5 is represented by the subregions identified by 2. Here, we see that the unknown laboratory temperatures (T) are conditional on the reported laboratory temperature data and the unknown variance, λ. Again we can write the full joint distribution of all quantities, and this is

$$p(L^{(\boldsymbol{T})}, \boldsymbol{T}, \lambda, L^{(r)}, \boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{T}|L^{(\boldsymbol{T})}, \lambda)p(\lambda)p(L^{(r)}|\boldsymbol{\beta}, \sigma^2)p(\boldsymbol{\beta})p(\sigma^2). \tag{A.3}$$

Again, we use Bayes theorem (and note we use the observed temperatures, $L^{(\boldsymbol{T})}$, in the prior for $\boldsymbol{T}$) to obtain

$$p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{T}, \lambda|L^{(r)}) \propto p(L^{(r)}|\boldsymbol{\beta}, ,\sigma^2, \boldsymbol{T}, \lambda)p(\boldsymbol{\beta})p(\sigma^2)p(\boldsymbol{T}|\lambda)p(\lambda). \tag{A.4}$$

The final model we considered in Section 7 incorporates the geological timescale data, V, with a relatively well-specified thermal history, $\boldsymbol{\phi}$. Again, we write the joint distribution in terms of the conditional and prior distributions,

$$p(L^{(r)}, V, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{T}, \lambda) = p(V|\boldsymbol{\beta}, \boldsymbol{\phi})p(L|\boldsymbol{\beta}, \sigma^2, \boldsymbol{T}, \lambda)p(\boldsymbol{\beta})p(\sigma^2)p(\boldsymbol{T}|\lambda)p(\lambda)p(\boldsymbol{\phi}) \tag{A.5}$$

and following the same procedure as above, we have

$$p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{T}, \lambda, \boldsymbol{\phi}|L^{(r)}, V) \propto p(L^{(r)}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{T}, \lambda)p(V|\boldsymbol{\beta}, \boldsymbol{\phi})p(\boldsymbol{\beta})p(\sigma^2)p(\boldsymbol{T}|\lambda)p(\lambda)p(\boldsymbol{\phi}). \tag{A.6}$$

As described in the text, under the assumption of an appropriate form for the priors of σ^2 and λ, we can integrate out these parameters to simplify the sampling procedure.

If we wished to model uncertainty in the laboratory times, $L^{(t)}$, this is readily achieved by inserting nodes to a new region similar to those in region 2,3.

Finally, if we wanted to include additional geological data from additional well-constrained studies, or directly model some geological data for which we do not a well-constrained thermal history, then the DAG is readily expanded by adding additional regions with nodes similar to, but independent of, region 3.

Appendix B. Marginalising conditional posteriors for efficient sampling

In this appendix, we will give some more details on how parameters are marginalised to increase efficiency, as well as more information on the likelihood function for the field data.

For several of the parameters in the posterior density (Eq. (5.1)), we are able to marginalise out various nuisance parameters making the sampling algorithm far more efficient. This is made possible by choosing an appropriate form for the prior, i.e., one that is conjugate to the likelihood. Then the posterior distribution has the same form as the prior and the integrals become analytically solvable up to proportionality (see Bernardo and Smith, 1994). For the normally distributed likelihood function, the conjugate prior is the inverse gamma distribution, as found over both λ and σ^2, which enables us to remove both of these parameters from our conditional posterior densities.

B.1. Conditional posterior over $\boldsymbol{\beta}$

With $\boldsymbol{\beta}$, we are able to marginalise the constant variance parameter σ^2. This is also performed in Laslett et al. (1987) using a result from Box and Cox (1964), though they do not consider prior densities, so the form of the posterior will be slightly different. Note that we do not need to consider the later dependence of $\boldsymbol{\beta}$ on the field collected data, as we assume independence between these likelihoods. In the following, we use L for the data in general, and distinguish $\{L^{(t)}, L^{(T)}, L^{(r)}\}$ only where it is required. We begin by expanding the model so that,

$$p(\boldsymbol{\beta}|L, \sigma^2, \boldsymbol{T}, \lambda) = p(\boldsymbol{\beta}|L, \boldsymbol{T}\sigma^2) \tag{B.1}$$

$$\propto p(L|\boldsymbol{\beta}, \sigma^2, \boldsymbol{T})p(\boldsymbol{\beta})p(\sigma^2) \tag{B.2}$$

$$\propto p(\boldsymbol{f}|\sigma^2, \boldsymbol{T})J(a, b, L^{(r)})p(\sigma^2)p(\boldsymbol{\beta}), \tag{B.3}$$

where $\boldsymbol{f}$ is the $(n_L \times 1)$ vector, containing the Box–Cox transformed data, $g(a, b, L^{(r)})$. As $p(\sigma^2)$ is given by an inverse gamma distribution, we can also represent the dependence of $p(\boldsymbol{f}|\sigma^2, \boldsymbol{T})p(\sigma^2)$ on σ^2 as following the form of an inverse gamma distribution. This allows us to easily marginalise it as a nuisance parameter

$$p(\boldsymbol{\beta}|L, \boldsymbol{T}) = \int p(\boldsymbol{\beta}|L, \boldsymbol{T}, \sigma^2)\,\mathrm{d}\sigma^2 \tag{B.4}$$

$$\propto p(\boldsymbol{\beta})J(a, b, L^{(r)})\int (\sigma^2)^{-(\eta_1+1+n_L/2)} \times \exp\left(-\frac{(SS^{(\boldsymbol{\beta})}) + 2\eta_2}{2\sigma^2}\right)\mathrm{d}\sigma^2 \tag{B.5}$$

$$\propto p(\boldsymbol{\beta})J(a, b, L^{(r)})(SS^{(\boldsymbol{\beta})}/2 + \eta_2)^{-(\eta_1+n_L/2)} \times \int IG(\eta_1 + n_L/2, SS^{(\boldsymbol{\beta})}/2 + \eta_2)\,\mathrm{d}\sigma^2 \tag{B.6}$$

$$\propto p(\boldsymbol{\beta})J(a, b, L^{(r)})(SS^{(\boldsymbol{\beta})}/2 + \eta_2)^{-(\eta_1+n_L/2)}, \tag{B.7}$$

where $\mathrm{SS} = \sum_{k=1}^{n_L}[g(r_k) - B(L_k^{(t)}, L_k^{(T)}, c_0, c_1, c_2)]^2$ is the residual sum of squares for the fit to the transformed data.

B.2. Conditional posterior over $\boldsymbol{T}$

As $\boldsymbol{T}$ is of such high dimension ($n_L = 77$), it is essential that we remove as many parameters as possible to maintain sampling efficiency. After removing non-relevant terms from the full posterior distribution, the conditional posterior for $\boldsymbol{T}$ becomes

$$p(\boldsymbol{T}|L, \boldsymbol{\beta}, \sigma^2, \lambda) \propto p(L|\boldsymbol{\beta}, \sigma^2, \boldsymbol{T})p(\sigma^2)p(L|\lambda)p(\lambda) \tag{B.8}$$

$$\propto p(\boldsymbol{f}^{\ (L)}|\sigma^2, \boldsymbol{T})p(\sigma^2)p(L|\lambda)p(\lambda), \tag{B.9}$$

we can marginalise both $\boldsymbol{c} = \{c_0, c_1, c_2\}'$ and σ^2 so that

$$p(\boldsymbol{T}|L, \lambda) \propto p(\sigma^2)p(L|\lambda)p(\lambda)\int p(\boldsymbol{T}|L, \boldsymbol{\beta}, \sigma^2, \lambda)\,\mathrm{d}\boldsymbol{c}\,\mathrm{d}\sigma^2 \tag{B.10}$$

$$\propto p(L|\lambda)p(\lambda)\int (\sigma^2)^{-(n_L/2+\eta_1+1)} \times \exp\left\{\frac{-(\boldsymbol{f} - \boldsymbol{E}\boldsymbol{c})'(\boldsymbol{f} - \boldsymbol{E}\boldsymbol{c}) + 2\eta_2}{2\sigma^2}\right\}\mathrm{d}\boldsymbol{c}\,\mathrm{d}\sigma^2, \tag{B.11}$$

where we have collected terms from the inverse gamma and normal distributions. This latter integral can be solved analytically, provided we rearrange terms using the following identity,

$$(\boldsymbol{f} - \boldsymbol{E}\boldsymbol{c})'(\boldsymbol{f} - \boldsymbol{E}\boldsymbol{c}) + 2\eta_2 = (\boldsymbol{c} - \hat{\boldsymbol{c}})\boldsymbol{V}^{-1}(\boldsymbol{c} - \hat{\boldsymbol{c}}) + 2\eta_2^*, \tag{B.12}$$

where $\hat{\boldsymbol{c}} = (\boldsymbol{E}'\boldsymbol{E})^{-1}\boldsymbol{E}'\boldsymbol{f}$ is the ordinary least squares estimate of $\boldsymbol{c}$ and $\boldsymbol{V} = (\boldsymbol{E}'\boldsymbol{E})^{-1}$. The term η_2^* captures the terms not comprising $\boldsymbol{c}$ and so

$$\eta_2^* = \eta_2 + \{\boldsymbol{f}'\boldsymbol{f} - \hat{\boldsymbol{c}}'\boldsymbol{V}^{-1}\hat{\boldsymbol{c}}\}/2 \tag{B.13}$$

This allows us to write (up to proportionality)

$$p(\boldsymbol{T}|L,\lambda) \propto p(L|\lambda)p(\lambda)\int (\sigma^2)^{-(n_L/2+\eta_1+1)} \exp\left(\frac{\eta_2^*}{\sigma^2}\right) \mathrm{d}\sigma^2 \tag{B.14}$$

$$\times \int \exp\left\{\frac{(\boldsymbol{c}-\hat{\boldsymbol{c}})\boldsymbol{V}^{-1}(\boldsymbol{c}-\hat{\boldsymbol{c}})}{2\sigma^2}\right\} \mathrm{d}\boldsymbol{c} \tag{B.15}$$

$$\propto p(L|\lambda)p(\lambda)|\boldsymbol{V}|^{-1/2}$$

$$\times \int (\sigma^2)^{-(n_L/2-k/2+\eta_1+1)} \exp\left(\frac{\eta_2^*}{\sigma^2}\right) \mathrm{d}\sigma^2 \tag{B.16}$$

$$\propto p(L|\lambda)p(\lambda)|\boldsymbol{V}|^{-1/2}(\eta_2^*)^{-(n_L/2-k/2+\eta_1)}$$

$$\times \int IG(n_L/2-k/2+\eta_1, \eta_2^*)\,\mathrm{d}\sigma^2 \tag{B.17}$$

$$\propto p(L|\lambda)p(\lambda)|\boldsymbol{V}|^{-1/2}(\eta_2^*)^{-(n_L/2-k/2+\eta_1)}. \tag{B.18}$$

The final step is to note that we can also marginalise λ in exactly the same manner, so that

$$p(\boldsymbol{T}|L,a,b) \propto |\boldsymbol{V}|^{-1/2}(\eta_2^*)^{-(n_L/2-k/2+\eta_1)} \int p(L|\lambda)p(\lambda)\,\mathrm{d}\lambda \tag{B.19}$$

$$\propto |\boldsymbol{V}|^{-1/2}(\eta_2^*)^{-(n_L/2-k/2+\eta_1)}$$

$$\times \int \lambda^{\eta_1+1+n_L/2}$$

$$\times \exp\left\{-\frac{(\boldsymbol{L}-\boldsymbol{T})'(\boldsymbol{L}-\boldsymbol{T})+2\eta_2}{2\lambda}\right\} \tag{B.20}$$

$$\propto |\boldsymbol{V}|^{-1/2}(\eta_2^*)^{-(n_L/2-k/2+\eta_1)}(SS^{(\lambda)}+2\eta_2)^{-(\eta_1+n_L/2)}, \tag{B.21}$$

where $SS^{(\lambda)} = (\boldsymbol{L}-\boldsymbol{T})'(\boldsymbol{L}-\boldsymbol{T})$, and is the residual sum of squares between the observed and predicted temperatures.

References

Allen, P., Carter, A., Cunningham, M., Densmore, A., Galewsky, J., Gallagher, K., Lazzaretti, E., Naylor, D., Phillips, A., 2002. The post-variscan thermal and denudational history of ireland. In: Dore, A. (Ed.), *Exhumation of the North Atlantic Margin: Timing, Mechanisms, and Implications for Petroleum Exploration*. Geological Society of London, Special Publication, vol. 196, pp. 371–399.

Barbarand, J., Carter, A., Hurford, T., Wood, I., 2003a. Compositional and structural control of fission track annealing in apatite. *Chem. Geol.* **198**, 107–137.

Barbarand, J., Hurford, T., Carter, A., 2003b. Variation in apatite fission-track length measurement: implications for thermal history modelling. *Chem. Geol.* **198**, 77–106.

Bernardo, J.M., Smith, A.F.M., 1994. *Bayesian Theory*. Wiley, Chichester.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. *J. R. Stat. Soc. Ser. B (Methodological)* **26** (2), 211–252.

Carlson, W., 1990. Mechanisms and kinetics of apatite fission-track annealing. *Am. Mineral.* **75**, 1120–1139.

Carlson, W.D., Donelick, R.A., Ketcham, R.A., 1999. Variability of apatite fission-track annealing kinetics: I. experimental results. *Am. Mineral.* **84** (9), 1213–1223.

Chartier, A., Meis, C., Gale, J., 2001. Computational study of cs immobilization in the apatites: ca10(po4)6f2, ca4la6(sio4)6f2 and ca2la8(sio4)6o2. *Phys. Rev. B* **64**, 085110.

Crowley, K., Cameron, M., Schaefer, R., 1991. Experimental studies of annealing etched fission tracks in fluorapatite. *Geochim. Cosmochim. Acta* **55**, 449–1465.

Denison, D.G.T., Holmes, C.C., Mallick, B., Smith, A.F.M., 2002. *Bayesian Methods for Nonlinear Classification and Regression*. Wiley, Chichester.

Donelick, R.A., Roden, M.K., Mooers, J.D., Carpenter, B.S., Miller, D.S., 1990. Etchable length reduction of induced fission tracks in apatite at room-temperature (approximately 23-degrees-c)—crystallographic orientation effects and initial mean lengths. *Nucl. Tracks Radiat. Meas.* **17** (3), 261–265.

Gallagher, K., 1995. Evolving temperature histories from apatite fission-track data. *Earth Planet. Sci. Lett.* **136** (3–4), 421–435.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Green, P., 1988. The relationship between track shortening and fission track age reduction in apatite: combined influences of inherent instability, annealing anisotropy, length bias, and system calibration. *Earth Planet. Sci. Lett.* **89**, 335–352.

Green, P., Duddy, I., Laslett, G., Hegarty, K., Gleadow, A., Lovering, J., 1989. Thermal annealing of fission tracks in apatite 4. Quantitative modelling and extension to geological timescales. *Chem. Geol.* **60**, 5117–5131.

Green, P.F., Duddy, I.R., Gleadow, A.J.W., Tingate, P.R., Laslett, G.M., 1986. Thermal annealing of fission tracks in apatite. 1. a qualitative description. *Chem. Geol.* **59** (4), 237–253.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** (4), 711–732.

Gunnell, Y., Gallagher, K., Carter, A., Widdowson, M., Hurford, A.J., 2003. Denudation history of the continental margin of western peninsular india since the early mesozoic—reconciling apatite fission-track data with geomorphology. *Earth Planet. Sci. Lett.* **215** (1–2), 187–201.

Harman, R., 2000. Long term denudation of the cratons of north eastern brazil. Ph.D. thesis, Imperial College, London, 365pp.

Jones, S.M., Dokka, R.K., 1990. Modeling fission-track annealing in apatite—an assessment of uncertainties. *Nucl. Tracks Radiat. Meas.* **17** (3), 255–260.

Jordan, M.I., 2004. Graphical models. *Stat. Sci.* **19** (1), 140–155.

Ketcham, R.A., Donelick, R.A., Carlson, W.D., 1999. Variability of apatite fission-track annealing kinetics: Iii. Extrapolation to geological time scales. *Am. Mineral.* **84** (9), 1235–1255.

Kohn, B., Belton, D., Brown, R., Gleadow, A., Green, P., Lovering, J., 2003. Comment on: Experimental evidence for the pressure dependence of fission track annealing in apatite. *Earth Planet. Sci. Lett.* **215**, 299–306.

Kohn, B.P., Gleadow, A.J.W., Brown, R.W., Gallagher, K., O'Sullivan, P.B., Foster, D.A., 2002. Shaping the australian crust over the last 300 million years: insights from fission track thermotectonic imaging and denudation studies of key terranes. *Aust. J. Earth Sci.* **49** (4), 697–717.

Laslett, G.M., Galbraith, R.F., 1996. Statistical modelling of thermal annealing of fission tracks in apatite. *Geochim. Cosmochim. Acta* **60** (24), 5117–5131.

Laslett, G.M., Green, P.F., Duddy, I.R., Gleadow, A.J.W., 1987. Thermal annealing of fission tracks in apatite. 2. A quantitative-analysis. *Chem. Geol.* **65** (1), 1–13.

Lee, P., 1989. *Bayesian Statistics: An Introduction*. Edward Arnold, London.

Paterson, M.S., Wong, T.-F., 2005. *Experimental Rock Deformation: The Brittle Field*, second completely revised and updated edition. Springer, Berlin.

Spiegelhalter, D., Best, N., Gilks, W., Inskip, H., 1996. Hepatitis b: a case study in mcmc methods. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, p. 2143.

Stephenson, J., Gallagher, K., Holmes, C., 2006. Low temperature thermochronology and strategies for multiple samples 2: partition modelling for 2d/3d distributions with discontinuities. *Earth Planet. Sci. Lett.* **241** (3–4), 557–570.

Vrolijk, P., Donelick, R.A., Quenq, J., Cloos, M., 1992. Testing models of fission track annealing in apatite in a simple thermal setting: site 800, leg 129. In: Larson, R., Lancelet, Y. (Eds.), *Proceedings of the Ocean Drilling Program, Scientific Results*, vol. 129, pp. 169–176.

Wendt, A., Vidal, O., Chadderton, L., 2002. Experimental evidence for the pressure dependence of fission track annealing in apatite. *Earth Planet. Sci. Lett.* **201**, 593–607.