



HAL
open science

Transdimensional inverse thermal history modeling for quantitative thermochronology

Kerry Gallagher

► **To cite this version:**

Kerry Gallagher. Transdimensional inverse thermal history modeling for quantitative thermochronology. *Journal of Geophysical Research*, 2012, 117, pp.B02408. 10.1029/2011JB008825. insu-00676497

HAL Id: insu-00676497

<https://insu.hal.science/insu-00676497>

Submitted on 5 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Transdimensional inverse thermal history modeling for quantitative thermochronology

Kerry Gallagher¹

Received 30 August 2011; revised 11 November 2011; accepted 22 December 2011; published 29 February 2012.

[1] A new approach for inverse thermal history modeling is presented. The method uses Bayesian transdimensional Markov Chain Monte Carlo and allows us to specify a wide range of possible thermal history models to be considered as general prior information on time, temperature (and temperature offset for multiple samples in a vertical profile). We can also incorporate more focused geological constraints in terms of more specific priors. The Bayesian approach naturally prefers simpler thermal history models (which provide an adequate fit to the observations), and so reduces the problems associated with over interpretation of inferred thermal histories. The output of the method is a collection or ensemble of thermal histories, which quantifies the range of accepted models in terms a (posterior) probability distribution. Individual models, such as the best data fitting (maximum likelihood) model or the expected model (effectively the weighted mean from the posterior distribution) can be examined. Different data types (e.g., fission track, U-Th/He, $^{40}\text{Ar}/^{39}\text{Ar}$) can be combined, requiring just a data-specific predictive forward model and data fit (likelihood) function. To demonstrate the main features and implementation of the approach, examples are presented using both synthetic and real data.

Citation: Gallagher, K. (2012), Transdimensional inverse thermal history modeling for quantitative thermochronology, *J. Geophys. Res.*, 117, B02408, doi:10.1029/2011JB008825.

1. Introduction

[2] Since the seminal publication of *Dodson* [1973] quantifying the relationship between geochronological ages and closure temperatures, an ongoing concern in thermochronology is reconstruction of thermal histories in various contexts, including tectonics, landscape evolution and resource exploration. Effort has focused on lower temperature systems such as noble gases (argon in feldspar, helium in apatite/zircon) and fission track analysis (apatite/zircon). Comprehensive sources for these techniques include *Reiners and Ehlers* [2005] and *Braun et al.* [2006]. In terms of thermal history information, advances have been made as a consequence of understanding the significance of fission track length distributions [e.g., *Gleadow et al.*, 1986; *Green et al.*, 1986] and multiple diffusion domains, either within single crystals ($^{40}\text{Ar}/^{39}\text{Ar}$ method) [*McDougall and Harrison*, 1999] or multiple grains (Helium method) as the grain is generally considered to be the diffusion domain for apatite [*Farley*, 2000, *Reiners and Farley*, 2001] and also probably for zircon [*Reiners et al.*, 2004].

[3] Extracting thermal history information is best treated as an inverse problem, given the complex relationship between the observations and the thermal history. The relationship is simulated through a mathematical (forward) model representing the time-temperature dependence of annealing or

diffusion, which is a simplification of complex physical processes operating over geological time. Furthermore, data are never perfect, being subject to human and technical vagaries and the complexity of the geological environment. This we refer to as noise and we follow the philosophy of *Scales and Snieder* [1998] in considering noise as that part of the data that we do not expect the model to explain. Also, there will be trade-off between noise and how well we expect to fit the data. Consequently, it is important to consider the resolution of the inferred thermal history, i.e., which features are over interpretation or even artifacts of the modeling process.

[4] As in many inverse problems, a continuous function (i.e., thermal history) is represented by a finite number of parameters, such as coefficients of Legendre polynomials [*Corrigan*, 1991], or more commonly discrete time temperature points [*Lutz and Omar*, 1991; *Gallagher*, 1995; *Willett*, 1997; *Ketcham*, 2005; *Harrison et al.*, 2005]. It is possible to place constraints on the range of thermal histories, typically defined as regions of time-temperature space through which acceptable models must pass. These constraints may include a range on the present-day temperature of a sample in a borehole or at the surface, requiring a sedimentary sample to be at near surface temperatures at the time of deposition, or during a period of erosion, or that a sample should be at elevated temperatures at a time of intrusion. If required, additional geological information can be incorporated into the modeling to further reduce the range of acceptable models.

[5] When solving the inverse problem (i.e., finding acceptable thermal histories), stochastic sampling methods have often been used, including basic Monte Carlo [*Lutz and*

¹Géosciences, Université de Rennes 1, Rennes, France.

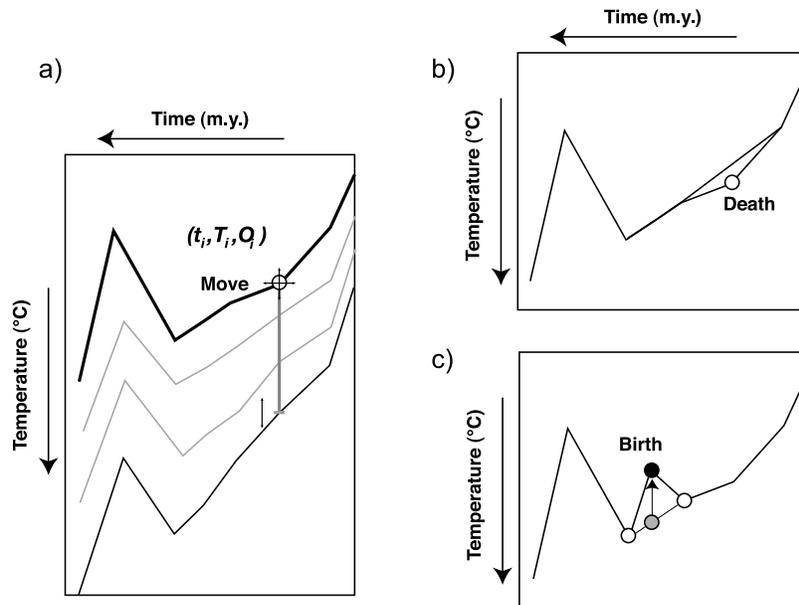


Figure 1. (a) Representation of a thermal history model. t_i-T_i is the i -th time temperature point (indicated by the open circle) and O_i is the temperature offset between the uppermost and lowermost sample in a vertical profile or borehole. By default the uppermost sample is used as the reference thermal history (thicker black line) and the thermal histories for all other samples are calculated using linear interpolation of the offset temperature to the appropriate elevation/depth for each sample. The arrows indicate the three moves possible for the three parameters. (b) Death of a time-temperature point. A point is selected at random and deleted. The two adjacent points are connected. (c) Birth of a time-temperature point. A point is selected at random and a new point is inserted (shaded circle) between it and the adjacent (younger) time point. The temperature value is interpolated and then a small random perturbation is made (shown by the arrow) to produce the new temperature (black circle). The same interpolation-perturbation procedure is used for the offset parameter.

Omar, 1991], genetic algorithms [Gallagher, 1995], and constrained random search [Willett, 1997]. These approaches make the implementation of the constraints as described above straightforward and are also relatively global when searching the model space. The issue then is how best to estimate those parts of the thermal history unconstrained by independent information, i.e., what is required to fit the thermochronological observations?

[6] We present a new transdimensional approach for finding thermal histories for multiple samples (in a vertical profile as in the work of Gallagher *et al.* [2005]). We do not specify the number of model parameters explicitly, but let the data determine the complexity of the thermal history, subject to user-defined constraints. The approach is formulated in a Bayesian framework, so thermal history constraints and additional information (e.g., kinetics of annealing or diffusion) are treated as information (prior probability). Model parameters are sampled, the appropriate forward problems are solved for a given thermal history and some measure of fit is calculated for each data type. This is repeated many times and the outcome is a collection of acceptable thermal history models, where acceptable can be quantified in terms of (posterior) probability. The solutions can be presented in terms of the probability of a thermal history passing through a particular region of the time-temperature space. It is also possible to select particular individual models from the probability distribution. The posterior probability distribution is produced by Markov

chain Monte Carlo sampling or MCMC, which although relatively straightforward to implement, does require tuning for specific problems [e.g., Gallagher *et al.*, 2009]. In the next section, we outline the basic methodology, although most of the mathematical details are in the appendix. In the subsequent section, we show some examples using both synthetic and real data, considering two of the common low temperature thermochronometers, apatite fission track and (U-Th)/He analysis, showing how various parameters can be tuned for a given problem. The methodology presented is readily applicable to other systems such as $^{40}\text{Ar}/^{39}\text{Ar}$, making method-specific modifications as required. Finally, we review the current and future possibilities of this approach.

2. Methodology

[7] The general problem is illustrated in Figure 1. We consider a series of samples from a vertical profile, in which the lower elevation samples are always at or above the temperature of the higher elevation samples. The thermal history model can be written as K discrete time (t) - temperature (T) points, with a temperature offset (O) between the uppermost and lowermost samples in the profile, i.e. $t_i, T_i, O_i, i = 1, K$ where the time-temperature-offset values are unknown, as is K . The uppermost sample is used to define a reference thermal history. The thermal histories for lower samples are then determined by linear interpolation using the elevation/depth of each sample relative to the upper sample and the

temperature offset (which can vary over time). Given observations, such as apatite fission track (AFT) data, or helium (AHe) ages, we want to identify thermal history models consistent with these data.

[8] To search the range of possible solutions for the unknown model parameters we use Bayesian Markov chain Monte Carlo [e.g., *Gilks et al.*, 1996; *Denison et al.*, 2002] which is a probabilistic sampling procedure. We implement the transdimensional form proposed by *Green* [1995, 2003] and recent descriptions of this methodology have been given by *Sambridge et al.* [2006] and *Gallagher et al.* [2009]. Examples of applications in Earth Sciences can be found in the work of *Malinverno* [2002], *Stephenson et al.* [2006b], *Charvin et al.* [2009], *Hopcroft et al.* [2007, 2009], *Bodin and Sambridge* [2009], and *Piana Agostinetti and Malinverno* [2010].

[9] The aim in the Bayesian approach is to construct the (posterior) probability distribution of the model parameter values, given the data. This relies on Bayes' rule [e.g., *Bernardo and Smith*, 1994] which, up to a constant of proportionality, can be written as

$$p(m|d) \propto p(m)p(d|m) \quad (1)$$

where $p(m|d)$ is the posterior probability density function (PDF) of the model parameters, m , given the data vector, d ; $p(m)$ is the prior PDF on the model, which represents what we consider reasonable for the values of the model parameters (e.g., a uniform distribution for temperature between 0 and 150°C, and for time between 300 and 0 Ma); $p(d|m)$ is the likelihood function which quantifies the probability of obtaining the data, d , given the model, m . This is a measure of the data fit and increases as the model fits the data better. Effectively the likelihood updates the prior information, transforming it to the posterior. If the prior and posterior distributions are the same, then we have learnt nothing new from the data.

[10] We choose a likelihood function for each data type and for computational reasons use the log of the likelihood. For AFT length and count data, we use the combined log likelihood function given by *Gallagher* [1995, equation A8], while for He age data we adopt a standard sums of squares measure as the log likelihood:

$$L = -\frac{1}{2} \sum_{i=1}^N \left(\frac{age_i^{obs} - age_i^{pred}}{\sigma_i} \right)^2 \quad (2)$$

where the superscripts *obs* and *pred* refer to observed and predicted and σ is the error on the observed age. Here the predicted age is a function of the model parameters, m .

[11] In general there is uncertainty in the observed thermochronometric data and also in the predictive models (for annealing or diffusion) as a result of measurement errors and simplifying assumptions, contributing to the noise referred to earlier. In the case of the observed data, it can be argued that we should not use the observed data directly (as these incorporate the measurement error which may be poorly known). Rather, we would like to model the unobserved true data (i.e., the data we would have recorded if there was no measurement error). This approach was adopted by *Jasra*

et al. [2006] for mixture modeling of geochronological data. In the case of the predictive models, these are calibrated from laboratory experimental data and extrapolated to geological timescales [e.g., *Laslett et al.*, 1987; *Green et al.*, 1989]. Both the fitting process and extrapolation can lead to uncertainty and the relative importance of physical mechanisms or controls may differ on different timescales.

[12] This uncertainty can be allowed for by sampling from a distribution of parameter or data values. For example, for AFT data we can resample the observed data (counts and lengths), using a bootstrap type of approach [*Efron and Tibshirani*, 1993]. To allow for uncertainty in the annealing models we could sample directly from uncertainties on the annealing model parameters [e.g., *Jones and Dokka*, 1990; *Stephenson et al.*, 2006a] or, in the absence of these uncertainties, we could sample the compositional or kinetic parameter (as implemented in the annealing models of, for example, *Carlson et al.* [1999], *Donelick et al.* [1999], and *Ketcham et al.* [1999, 2007]). For AHe data, we could sample using uncertainties (if available) on the diffusion parameters (D_0 , E) or resample from the data from a distribution based on the observed age and error. We can then use these sampled ages instead of the observed value in the likelihood function. We will consider two of these possibilities - resampling the AFT kinetic parameter and He age data to demonstrate these approaches. While these are not the same uncertainties as directly sampling the diffusion kinetics for example, this is one approach to deal with the general noise. Allowing for data/model uncertainties will then lead to greater uncertainties in the inferred thermal histories. We consider this a more conservative approach than not allowing at all for such uncertainties.

[13] MCMC is an iterative sampling approach in which the first part of the series (burn-in) represent an initial exploration of the model space (which is subsequently discarded). The second part of the series (post-burn-in) is used to approximate the posterior distribution for the model parameters. In MCMC we work with two models in a given iteration, the current (m_c) and the proposed model (m_p). The latter is generated by randomly perturbing the former and we repeat this many times, the first model being drawn randomly from the prior. At each iteration, we either replace the current model with the proposed model, or reject the proposed model. We collect the current model at each iteration and use them as an approximation of the posterior distribution. For this to be valid, the way we decide to accept the proposed model is critical. The acceptance criterion is

$$\alpha(m_p, m_c) = \min \left(1, \frac{p(m_p)p(d|m_p)q(m_c|m_p)}{p(m_c)p(d|m_c)q(m_p|m_c)} |J| \right) \quad (3)$$

The first two terms in the ratio represent the prior and likelihood functions, $q(m_p|m_c)$ is the proposal function which determines how we produce the proposed model given the current model and $|J|$ is the Jacobian, which allows for the fact that we are transforming probability distributions. We define these terms in the appendix.

[14] To produce the proposed model, we define a series of possible transformations on the current model: (1) move a time point, (2) move a temperature point, (3) move a offset

value, (4) remove a time-temperature-offset point (death), (5) add a time-temperature-offset point (birth), (6) move (sample) the kinetic parameter for an individual sample, (7) move (sample) an observed data value for an individual sample.

[15] The first five transformations directly influence the form of the thermal history as shown in Figure 1. For a given iteration we need to choose which transformation to make. Each is assigned a probability, p_i , such that $\sum p_i = 1$ and we randomly select one, based on these probabilities. The five move transformations involve simply a perturbation to a model parameter or data value, while the death and birth transformations involve a change in the model dimension. For the moves, we randomly select a time-temperature-offset point or individual sample, and modify the appropriate value in the current model. For a normal distribution proposal function, we have

$$q(m_p|m_c) = \frac{1}{\sigma_p^i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{m_c - m_p}{\sigma_p^i} \right)^2} \quad (4)$$

where σ_p^i is the standard deviation of the proposal function for the i -th transformation. To produce a proposed value, we generate a random number, u , drawn from a standard normal distribution (mean = 0, standard deviation = 1), and use the transform

$$m_p = m_c + u\sigma_p^i$$

Here we only change the one parameter selected from the current model, although it is possible to change several parameters at the same time (block update). The Jacobian term in equation (4) is always 1 for these moves.

[16] As discussed by *Gallagher et al.* [2009], σ_p^i controls the efficiency of sampling. If σ_p^i is too small (relative to the width of the posterior distribution) then the model perturbations are also small, the proposed model is very similar to the current model, the datafit is similar and we accept almost every proposed model. However, we tend to move around the model space slowly. Conversely if σ_p^i is large (relative to the posterior width), then many proposed models will be rejected. Again we move around the model space slowly. We need to choose an appropriate value of σ_p^i without knowing the width of the posterior distribution (which we are trying to estimate). This can be achieved by monitoring the acceptance rate of proposed models (i.e., the number of accepted models relative to the total number proposed). Typically we tune the proposal function scale (σ_p^i) to obtain an acceptance rate of around 20–50%. As a general guide, if the acceptance rate is too low, σ_p^i is too large and vice versa.

[17] For the birth or death transformations, we select a time-temperature point at random, and either add a new time-temperature-offset point either before or after it or delete that time-temperature-offset point (see Figure 1). The death step is trivial while the birth transformation we adopt is similar to that of *Hopcroft et al.* [2007] and *Charvin et al.* [2009] and is described in Appendix A, as is the derivation of the Jacobian term in equation (4). The acceptance rate for these two transdimensional transformations tends to be much lower than the values discussed above for moves. This is

because the change in dimension can propose a model with a likelihood value significantly different from the current model. Typical acceptance rates in this problem can be 0.1–10% and from experience better sampling occurs when the acceptance rates for birth and death are similar.

3. Application of Method to Synthetic and Real Data

[18] To demonstrate various features and applications of this methodology, we now consider some examples. The first two involve synthetic AFT and AHe data from a single sample, while the next two involve real data from multiple samples for a rotated fault block and a borehole.

3.1. Example 1

[19] The first example is shown in Figure 2a. This heating-cooling scenario was used to generate high quality synthetic apatite fission track data (300 track lengths and 50 track count ratios) and apatite He (AHe) ages for three different grain sizes. We used the annealing model of *Ketcham et al.* [2007] with a kinetic parameter, $D_{par} = 2.5 \mu\text{m}$. For the helium ages we used a spherical grain model with the diffusion parameters of *Farley* [2000]. The fission track data were sampled from predicted distributions so have some inherent noise and the kinetic parameter was fixed to the true value. The He ages are used as predicted (no resampling), with an error drawn from a normal distribution with a standard deviation equal to 2.5% of the predicted ages. The priors on the time-temperature parameters were set to 60 ± 60 m.y. and $70 \pm 70^\circ\text{C}$ and the proposal scale parameters were 5 m.y. and 5°C . The sampler was run for 100,000 iterations and the first 50,000 were discarded as burn-in.

[20] Figure 2a shows the best data-fitting solution in which the sampling algorithm was modified not to penalise more complex models (model 1). Figure 2a also shows the best data-fitting solution found using the transdimensional algorithm (model 2). Clearly the former is somewhat extreme in that much unwarranted structure is introduced. The penalised best fit model is much less complex, but also has structure not present in the true thermal history. The predictions are not significantly different and the data fit, as measured by the log likelihood, is almost the same (-1552.30 v -1552.62 , while the true solution gives a value of -1553.41 , as the fission track data are not perfect). On the basis of the difference in the likelihoods, the less complex model would be selected using, for example, the Bayesian Information Criterion (BIC) [*Schwarz*, 1978].

[21] We want to construct the probability distribution on the thermal history so we divide the time axis into 1 m.y. intervals and the temperature axis into 1°C intervals. By summing the number of time-temperature paths that pass through each time-temperature square (1 m.y. by 1°C), we can plot a contour map of the probability distribution over time and temperature. We also calculate the 95% credible intervals (the Bayesian form of confidence intervals) by identifying the values defining upper and lower 2.5% of the temperature distribution at each 1 m.y. time interval.

[22] We often want to examine an individual model. This may be the often too complex best fit model or, following the Bayesian philosophy, the expected model which is a

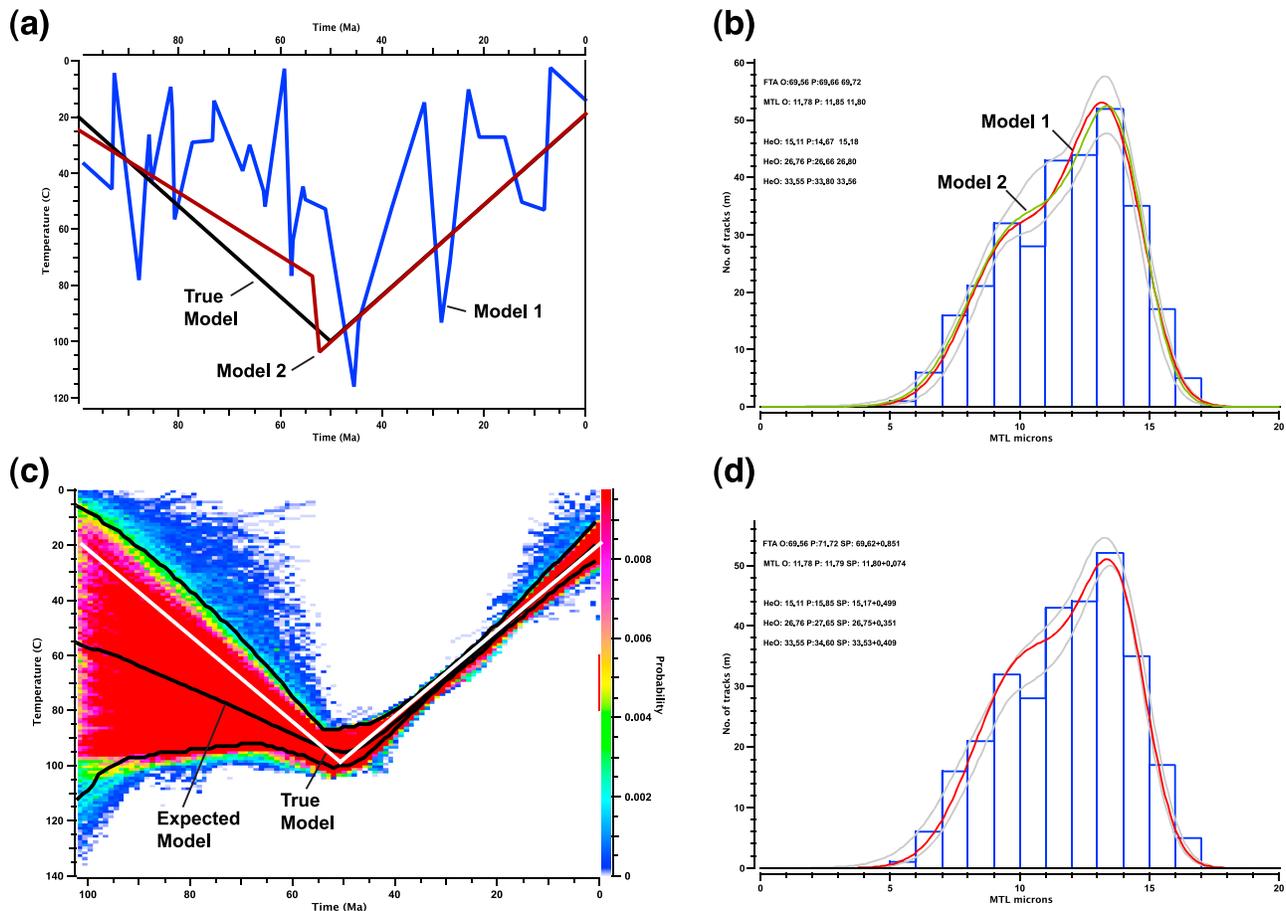


Figure 2. (a) Best data fitting (maximum likelihood) thermal histories for example 1 inferred using an approach which ignores the complexity of the thermal history (model 1) and transdimensional MCMC which penalises more complex models (model 2). The original thermal history used to generate the synthetic data is shown by the black line. (b) The observations (synthetic data) and model predictions for the two thermal histories in Figure 2a. The gray lines on length distribution are the 95% credible intervals for the predicted values. FTA is the fission track age (O: observed, P: predicted - model 1, model 2), MTL is the mean track length and He are the three helium ages (for grain radii of 50, 150, 300 μm) (c) The expected (weighted mean) thermal history (thick black line) and the probability distribution for the sampled thermal histories. The thinner black lines are the 95% credible intervals for the thermal history. The scale on the right indicates the probability. The true answer is shown as the white line. (d) As Figure 2b but for the expected thermal history model in Figure 2c. SP is the mean of the sampled predicted values (+ the standard deviation) for all sampled thermal histories. Note that this mean is not necessarily the same as the predicted value from the expected (mean) thermal history.

weighted average, where the weighting is the posterior probability. The expected model is defined as

$$E(m) = \int m p(m|d) dm \quad (5)$$

where $p(m|d)$ is the model posterior probability. As our sampled models are from the posterior, the integral is replaced by the average of the post-burn-in accepted models:

$$E(m) \approx \sum_{i=1}^N m_i \quad (6)$$

The expected model shown in Figure 2c and clearly we do not fit the data (log likelihood -1560.01) as well as the

maximum likelihood model (model 2). This is normal as the expected model reflects the range of models, both good and poor data fitting ones. However, the maximum likelihood model has added complexity (more time-temperature points) and so is penalised in the Bayesian approach which favors simpler models which fit the data reasonably well. Overall, the expected model is representative of the true solution (like the maximum likelihood model), where it is well resolved (i.e., the cooling episode), as shown by the credible intervals.

[23] To assess the performance of the MCMC sampler, it is useful to examine the acceptance rates and also the likelihood/posterior probabilities and the dimension of the thermal history models as a function of iteration. The acceptance rates were 51% and 37% for the time and temperature moves, and 6% for birth and death. The post-burn-in sampling

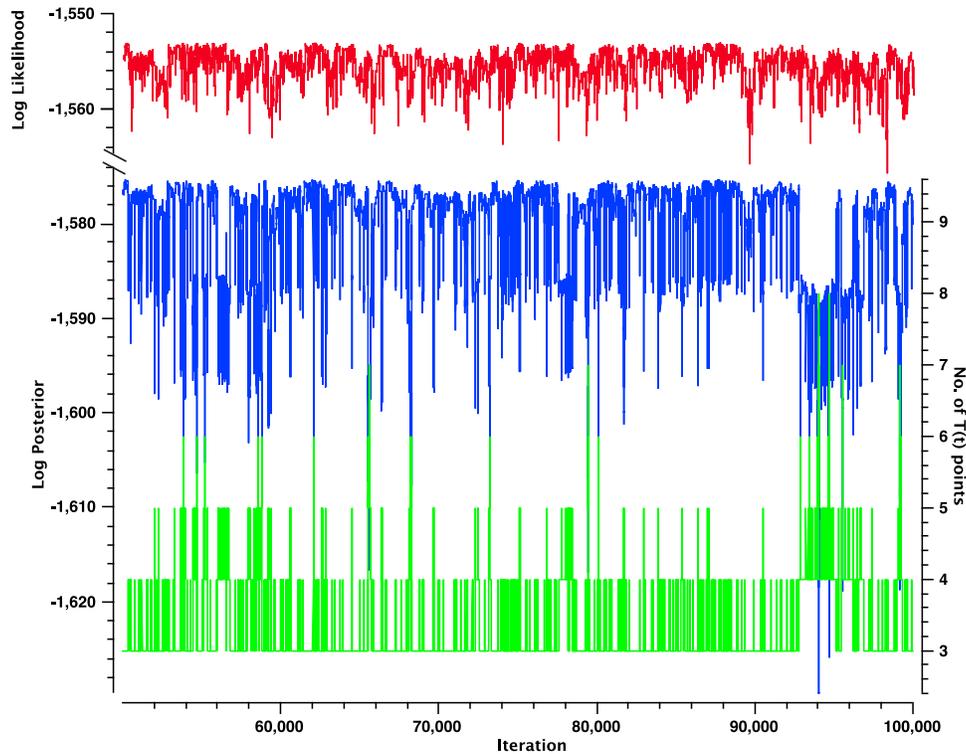


Figure 3. The log of the likelihood (data fit) function, log of the posterior probability and the number of time-temperature points as a function of post-burn-in iteration for the results shown in Figure 2.

chain (Figure 3) shows the dimension of the model varies regularly (up to a maximum of eight time-temperature points), but generally rests around 3–4. The likelihood tends to be higher for models with more time-temperature points, but the posterior tends to be, relatively, much lower than for models with fewer points (as seen around 94,000 iterations). As the acceptance criterion is based on the posterior ratio, this demonstrates how the Bayesian approach naturally prefers simpler models. Note that models with low posterior probabilities can be accepted, but at a rate proportional to the posterior probability. During the early iterations, the burn-in period, the likelihood tends to be low (poor data fitting models as a result of a random initial model), and generally we increase dimension to improve the data fit. We return to this in example 3.

3.2. Example 2

[24] The second example illustrates having the fission track composition (kinetic parameter) as a variable and resampling of the AHe ages. In both cases, we sample from a normal probability distribution centered on the input value, with standard deviation equal to the input error. Having drawn a proposed value, we use it in the thermal history modeling stage. Thus for the AFT kinetic parameter, we might use a slightly different annealing model (based on *Ketcham et al.* [2007]) between iterations. For the AHe ages, it is the observed value that changes in the likelihood function estimates for the current and proposed models.

[25] Here we use effectively the same data and priors as example 1, except the input kinetic parameter is $D_{\text{par}} = 2.0 \pm 0.3$ (original value = $2.5 \mu\text{m}$), while noise was added to the AHe ages, drawn from a normal distribution with standard

deviation equal to 2.5% of the input age. The input error was set to 5% of the age. Otherwise the inverse modeling was run as in example 1. The inferred thermal histories and the sampled values for the kinetic parameter, the predicted AFT age, mean track length (MTL) and three AHe ages (for three different grain sizes) are shown in Figure 4. Again, the maximum likelihood thermal history is more complex than the true solution. The maximum posterior solution (the model with the maximum posterior probability) is perhaps the most representative in this example. The expected model is also representative where the thermal history is well resolved, although the maximum temperature is a little underestimated. Relative to the previous example, the spread in the thermal history posterior distribution is greater. This is expected given the additional uncertainty associated with the kinetic parameter and the AHe ages.

[26] Considering the kinetic parameter (Figure 4b), the sampling produces a distribution with a peak around $2.4 \mu\text{m}$, clearly an improvement from the (incorrect) input value of $2.0 \mu\text{m}$. The predicted AFT data (Figures 4c and 4d) reflect the sampled range in the kinetic parameter, combined with the sampling range of thermal histories. Fixing the kinetic parameter to $2 \mu\text{m}$ leads to a similar form for the thermal history, except the maximum temperature is about 8°C lower than the true value. We anticipate a trade-off between the kinetic parameter and the maximum temperature, with lower D_{par} (F-rich apatite) leading to lower inferred maximum temperatures. This is reflected in the asymmetry seen in the posterior distribution around the maximum temperature as a result of the nonlinear temperature dependence of both annealing and diffusion.

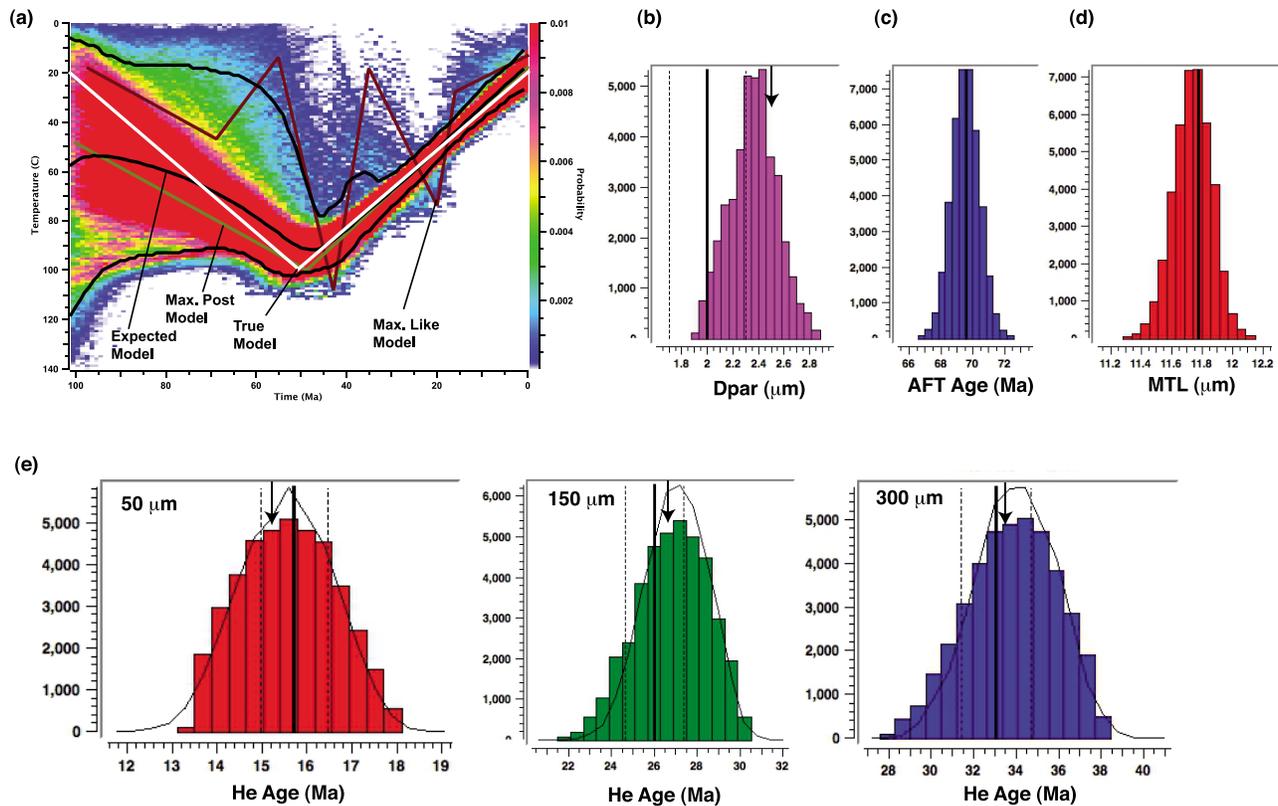


Figure 4. (a) Thermal history solutions (maximum likelihood, maximum posterior, and expected models) for example 2 together with the probability distribution. The original model solution is shown white line. (b) The post-burn-in sampling distribution of the kinetic parameter (D_{par}). The input value is shown by the vertical black line at $2 \mu\text{m}$, with the 1σ range as dashed lines, and the original value by the arrow ($2.5 \mu\text{m}$). The sampling is restricted to $\pm 3\sigma$. (c) The distribution on the predicted AFT age, with the input value shown as the vertical black line. (d) The distribution on the predicted mean track length (MTL), with the input value shown as the vertical black line. (e) The histograms are the distributions of the values sampled for observed AHe ages, the input value is shown by the vertical black line at $2 \mu\text{m}$, with the 1σ range as dashed lines, and the original value by the arrow. The thinner black lines in each panel are the distributions of the predicted values.

[27] The spread in the distribution of predicted values is about twice as large as that for example 1, when the observed AHe ages were fixed at their input values. This is because, as can be seen in Figure 4e, the sampling approach produces a distribution of “observed” ages, and then this spread in the data influences the distribution of the acceptable predicted ages.

3.3. Example 3

[28] For the next example we consider a series of AFT and AHe data from *Colgan et al.* [2006]. The data are from 16 granite outcrop samples (intrusion age of 108–115 Ma), currently at about the same elevation (~ 1500 m) from the footwall of a tilted fault block, the Pine Forest Range in northwest Nevada, USA. This is part of the Basin and Range province and the interpretation is that faulting has been active since 11–12 Ma and implies about 12% extension. Based on structural arguments, *Colgan et al.* [2006] estimate 7 to 9 km of slip on the range-front fault and 5 to 6.5 km of horizontal extension, with palaeodepths (below the pre-faulting land surface) of 1330 to 5370 m, with the samples closest to the fault plane having the largest palaeodepths. Given the

palaeodepth estimates and the fact that the samples are more or less at the same elevation today, this suite can be treated as a vertical profile in which the temperature difference between the deepest and shallowest (palaeodepth) has decreased over time. This is what we expect during exhumation of the footwall on a fault whose dip shallows with time (due to rotation) accompanied by erosion of the fault block crest.

[29] For the thermal history models, we follow the approach outlined by *Gallagher et al.* [2005] in which we assume the samples are at a fixed depth (=palaeodepth). We estimate a thermal history for the shallowest sample, and also the temperature offset between the shallowest and deepest samples, with linear interpolation to estimate thermal history for the samples in between. We assume a uniform prior on the temperature history (for the shallowest sample) of 80 ± 80 m.y. and $70 \pm 70^\circ\text{C}$, with a fixed surface temperature of 20°C (for all samples). This range of time is broadly equivalent to twice the maximum age of all the samples, and the temperature range spans the sensitivity of the two thermochronometric systems used. Given the known intrusion age, we impose a constraint on the structurally shallowest sample to be at a temperature of $160 \pm 10^\circ\text{C}$ between 110 ± 10 Ma.

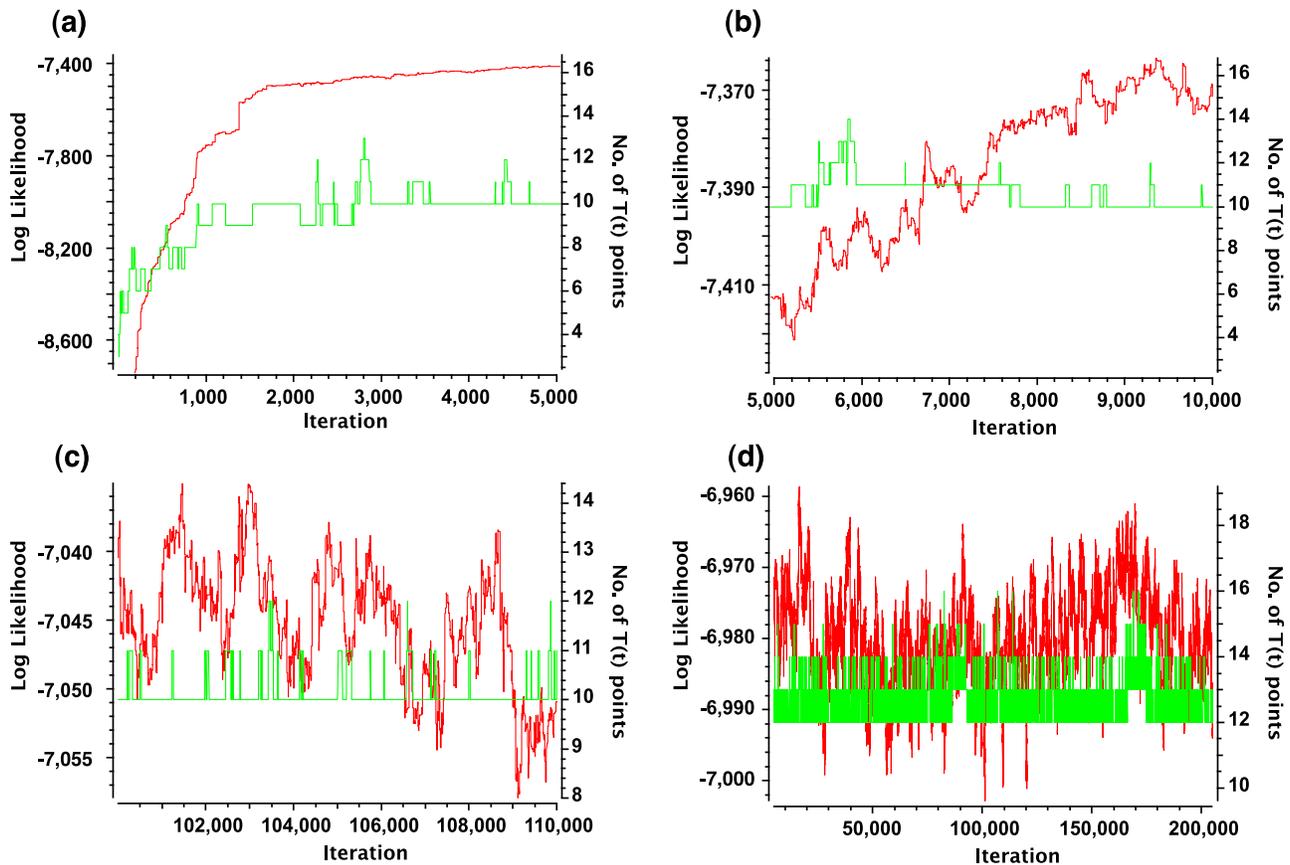


Figure 5. The log of the likelihood function and the number of time-temperature points at different stages of the sampling process. Note the different ranges on the Y scale for each plot. See the text for discussion.

After a series of trial runs, the temperature offset prior was set to $400 \pm 400^\circ\text{C}$. Using a smaller range for this parameter lead to values converging to the maximum limit of the prior. We return to the significance of this later. We use the observed AHe age errors in the resampling distribution as described earlier. For the AFT kinetic parameter (D_{par}), we use the mean of the measured values with an error equal to the standard deviation. The basic fission track data were the spontaneous and induced track counts, and measured (unprojected) lengths.

[30] Figure 5 shows the likelihood and the number of t-T points for various stages of a sampling run. The initial model is drawn from the prior and has just one time temperature point (in addition to the high temperature constraint and the present-day temperature). We do not expect such a simple model to fit the data well (log likelihood of -187728), so the number of model parameters increases rapidly as the sampling tries to identify parts of the time-temperature space that produce better data fits. In this case after about 2000 iterations the likelihood appears to flatten off on the scale of Figure 5a. Considering the same run for 5000 to 10000 iterations (Figure 5b), the likelihood function still increases but the sampling for the dimension of the thermal history is more stable. The tendency for the likelihood function to increase shows that the algorithm is still well in the burn-in stage of the sampling, that is we should discard these samples before making any inference. After 100000 iterations (Figure 5c), the likelihood function is also more stable as the

sampling has reached better regions of the model space (in terms of the data fit).

[31] In practice, we can use the samples from such exploratory runs as starting models for subsequent runs. For example, we could select either the maximum likelihood model or the last model sampled from a previous run and rerun a longer sampling chain for the final inference. In this way we can reduce the burn-in for the later runs, as the theory behind Bayesian MCMC states that once we start to sample from the target posterior distribution, all subsequently accepted models will be drawn from the same distribution [e.g., *Gilks et al.*, 1996]. In Figure 5d we show the sampling after running an exploratory chain for 500,000 iterations, then using the last model as a starting model for another chain of 600,000 models. We see that the sampling is fairly stable with the likelihood function showing no obvious trends with iteration and the dimension of the thermal history models varying regularly generally between 12 and 14, and sometimes reaching up to 16.

[32] To find reasonable values for the proposal scale parameters, it is typical to run several short chains and monitor the acceptance rate for all the variable parameters to identify reasonable scale parameter values. As mentioned earlier, there are no fixed values for these acceptance rates but it is typical to consider 20–50% as reasonable for most parameters, except the birth and death moves can be much lower than this. Increasing or decreasing the scale parameter for a proposal function tends to reduce or increase the

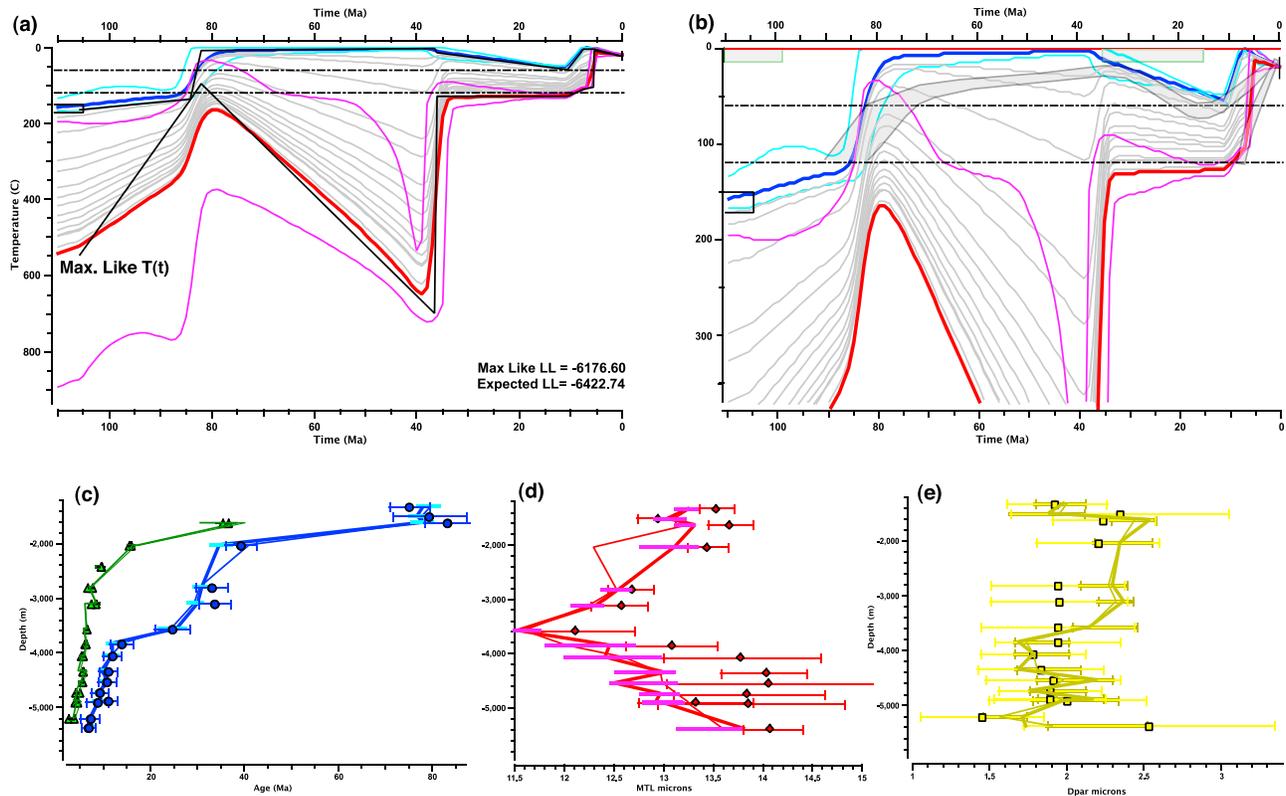


Figure 6. (a) The expected thermal history for the Pine Forest Range samples [see *Colgan et al.*, 2006]. The upper (reference) and lower thermal histories are shown as heavier lines, together with their 95% credible intervals. The credible intervals for the lower thermal history includes the uncertainty inferred for the temperature offset. The two dashed horizontal lines approximate the partial annealing zone (120 and 60°C) and the box at 100–120 Ma, and 160 ± 10°C indicates a constraint on the thermal history. The maximum likelihood thermal history is also indicated (just the upper and lowermost thermal histories as black lines), along with the calculated log likelihoods for the expected and maximum likelihood thermal histories. (b) As Figure 6a but with a smaller range on the temperature axis. The shaded bands indicates the general thermal histories inferred for shallow and deep samples, respectively, by *Colgan et al.* [2006]. The two boxes at 0°C on the time axis indicate the timing of intrusions (100–120 Ma) and Tertiary magmatism and burial (35–15 Ma) described by *Colgan et al.* [2006]. (c) Observed and predicted AFT (circles) and AHe ages (triangles). The predicted values for the maximum likelihood and expected models are shown as thick and thin lines (and individual AHe predicted ages are shown as x). Error bars (2 σ) are shown on the observed values and the lighter horizontal bars indicate the $\pm 1\sigma$ range on the predicted ages for all accepted models. (d) As Figure 6c but for the observed and predicted MTL values. The predicted values for the maximum likelihood and expected models are shown as thick and thin lines. Error bars (2 σ) are shown on the observed values and the lighter horizontal bars indicate the $\pm 1\sigma$ range on the predicted MTL values. (e) As Figure 6c but for the observed and predicted kinetic parameter (Dpar) values. The error bars represent $\pm 3\sigma$ of the observed values (and define the sampling range for these parameters).

acceptance rate, respectively. For the chain shown in Figure 5d, the proposal scale parameters were: $\sigma_{time} = 1$ Ma, $\sigma_{temperature} = 5^\circ\text{C}$, $\sigma_{offset} = 20^\circ\text{C}$, $\sigma_{kinetic} = 2$, $\sigma_{HeAge} = 2$ (the latter two have no units) and the acceptance rates for each move type were 39%, 34%, 48%, 45% and 41% respectively. The acceptance rates for birth and death moves were both 0.86%.

[33] Figure 6 shows the thermal histories and predictions for the maximum likelihood and expected models, together with the summary interpretation of *Colgan et al.* [2006]. This latter interpretation was based on modeling AFT data from individual samples (using the earlier annealing models described by *Ketcham et al.* [2000]) and also incorporating

regional geological information (for the lower temperature part of the thermal history). Overall, the trends in the AFT and AHe ages are well explained, but the mean track length tends to be about 1–1.5 μm too short for the deepest samples. However, the general trend of increasing mean length with depth is consistent with the observations. The fission track kinetic parameter sampling are constrained to be within the error bars shown on Figure 6e. The values generally stay close to the input values, except for the values between 2000 and 3800 m, which tend to be higher than the input values.

[34] Figure 7 illustrates the sampling of the kinetic parameter for the sample at ~ 2800 m. The sampling is reasonably stable, the distribution is skewed toward the upper

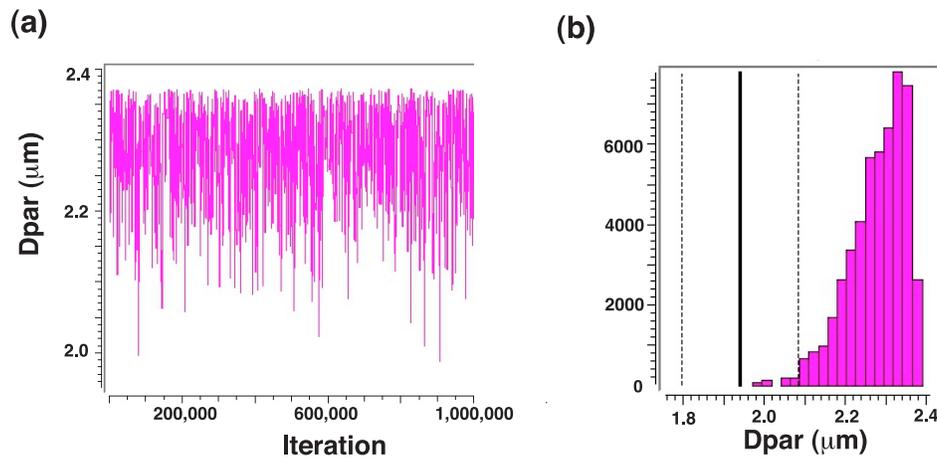


Figure 7. An example of the sampling of the kinetic parameter (D_{par}) for the Pine Forest Range sample at ~ 2800 m palaeodepth. (a) The sampled (accepted) values as a function of post-burn-in iteration (b) the distribution (histogram) of accepted values, together with the input values (the thicker vertical line in the mean measured value and the dashed lines represent $\pm 1\sigma$).

limit (about $2.4 \mu\text{m}$) and does not really want to stay at the input value of $1.94 \mu\text{m}$. Running the same thermal history with the kinetic parameters fixed to the input values leads to a worse fit to the data, as we expect with fewer free parameters. Relative to the variable kinetics model, this model underpredicts the FT age for the samples between 2000 and 3800 m, with slightly older ages around 4000 m and generally has lower predicted mean track lengths for all samples, by up to a micron. Comparing the maximum likelihood models using the BIC implies the model with the variable kinetic parameters is preferred. However, the thermal histories do not differ significantly, implying it is the overall thermal history that dominates the data fitting process. Thus, we consider this process of allowing the kinetic parameter to vary as more of a fine tuning process which allows for some uncertainty (or flexibility) in the annealing model parameters, rather than necessarily being a better estimation of the kinetic parameters.

[35] Figure 8 shows the result of sampling the AHe ages for the same sample at ~ 2800 m. One of the sample distributions have a peak slightly higher than the input value, but within

more or less 1σ . The predicted distribution is similar to sample distribution, but the peaks are slightly closer to the observed value. The spread in the predictions reflects the sampling of both the observed ages and the thermal histories. Fixing the AHe ages to the observed values does not lead to any significant differences in the inferred thermal histories (slightly less spread) as the final part of the thermal history (< 5 Ma) is well constrained by these data.

[36] The thermal history implies rapid cooling around 85–80 Ma followed by reheating, then a second period of rapid cooling around 40–35 Ma, with the deepest samples just about entering the typically defined fission track partial annealing zone (below 120 – 130°C), a third period of cooling starting around 10 Ma, then a final period of rapid cooling at 5 Ma to surface temperatures. The features of the thermal history that are well constrained from the data agree well with the form of the general thermal history inferred by *Colgan et al.* [2006]. As shown by *Gallagher et al.* [2005], modeling even synthetic data from individual samples can lead to overly complex thermal histories, compared to those inferred from multiple samples in a vertical profile. With real data,

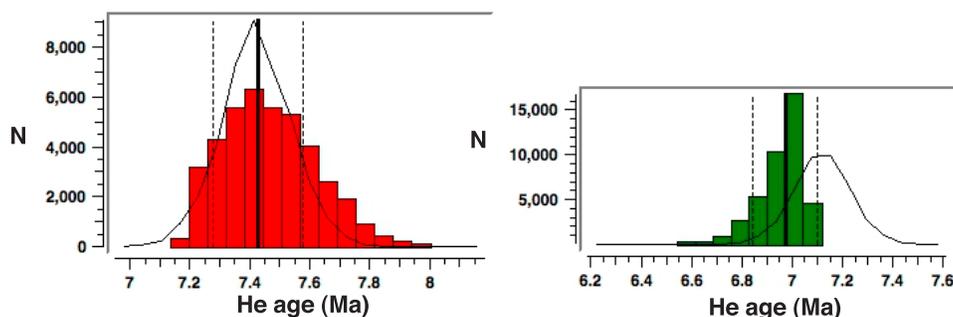


Figure 8. An example of the sampling distribution of accepted values (histogram), together with the input values (the thicker vertical line and the dashed lines represent $\pm 1\sigma$) for the two AHe samples for the Pine Forest Range sample at ~ 2800 m palaeodepth. The continuous lines represent the distribution on the predicted values for each sample.

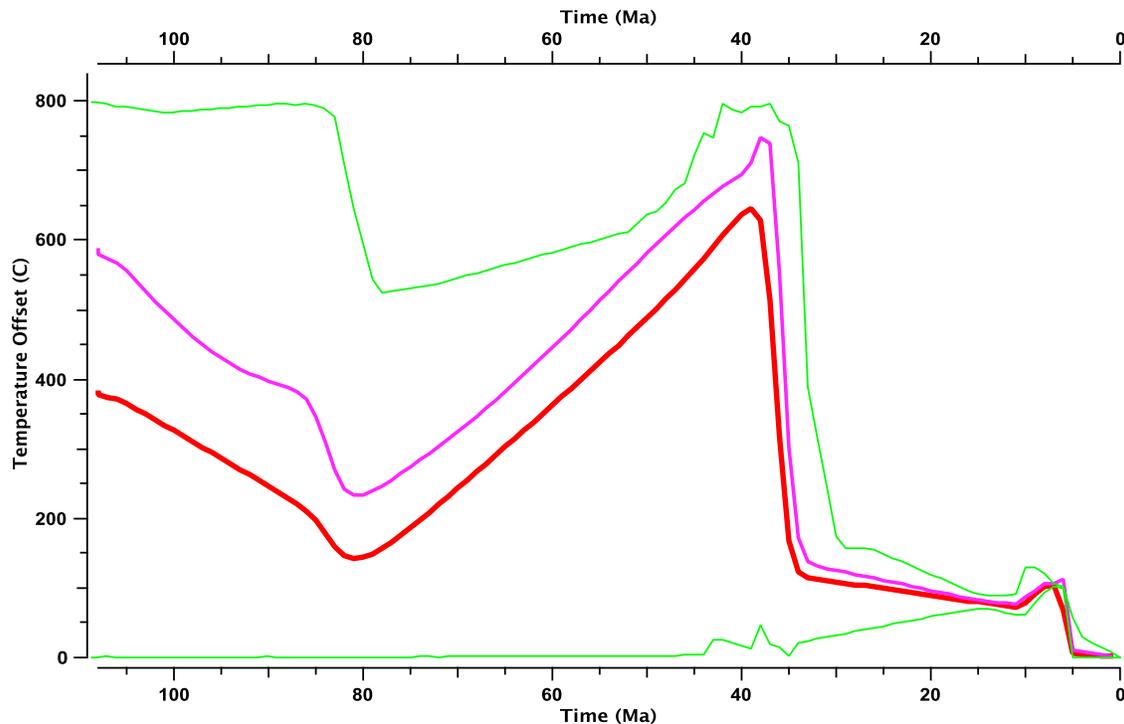


Figure 9. The inferred temperature offset (the temperature difference between the shallowest and deepest samples) for the Pine Forest Range samples. The thicker line is the expected value and inner lines mark the $\pm 1\sigma$ range while the outer thinner lines are the 95% credible ranges (which need not be symmetrical).

there will always be a compromise fit as the data are not perfect and so we suggest that the combined approach is preferable to avoid this problem.

[37] The inferred temperature offset is shown in Figure 9. The range in the temperature difference before rapid cooling around 30 Ma spans the prior range, showing that the data provide little information on this part of the thermal history. As only the four uppermost samples stay below 120°C during this time, we infer that the role of the offset parameter here is to ensure that the lower samples are hot enough to be completely degassed/annealed while allowing enough variation in the temperature to account for the top four samples. The large inferred offset value is a consequence of using linear interpolation between the upper and lower samples. For example if the deeper samples were affected by a thermal anomaly associated with the Oligocene-Miocene magmatism, they could all have been at 200°C , which cannot be modeled with the linear gradient assumption. In fact, setting all temperatures above 200°C to exactly 200°C leads to exactly the same values for the predicted ages and track lengths. After the rapid cooling, the offset is inferred to be just under 100°C (with a 95% credible range of $\sim 60\text{--}120^{\circ}\text{C}$), implying a temperature gradient of about $25^{\circ}\text{C}/\text{km}$. This stays more or less constant until the start of extension around 10–12 Ma, and as all samples are constrained to be at surface temperatures at the present-day, then the offset necessarily decreases, as we expect during rotational exhumation of the footwall.

3.4. Example 4

[38] This example also involves a vertical profile, but from an exploration well. These data are proprietary, but the

details are not important here. The 13 samples are from poorly dated sediments (known only to be lower/upper Cretaceous or Tertiary in age). We have only fission track data (but no compositional/kinetic parameter measurements) for these samples, and use projected track lengths for the modeling (following *Ketcham et al.* [2007]). It is clear that some of them have inherited fission tracks (indicated by ages older than the stratigraphic age). We have present-day temperatures for each sample which are incorporated as constraints.

[39] To allow for the lack of precise knowledge on the stratigraphic age, we include an additional prior on the temperature history for a range in stratigraphic age. We assign each sample one of three possible stratigraphic age ranges 140–100, 95–65 or 65–35 Ma and a temperature of $10 \pm 20^{\circ}\text{C}$ and the thermal history for each sample is required to pass through the specified prior range. We allow for inherited fission tracks by including an additional independent time-temperature point before the time of deposition. This point is drawn from the general prior for the thermal history (defined as a 2D uniform distribution with the mean \pm half widths given as 350 ± 350 Ma and $70 \pm 70^{\circ}\text{C}$), subject to the condition that it is prior to deposition. We allow the temperature offset to vary over time, with a prior of $30 \pm 30^{\circ}\text{C}$ and present temperatures are required to be within $\pm 5^{\circ}\text{C}$ of the measured values. These imply a present-day temperature offset of 46°C (temperature gradient of $23^{\circ}\text{C}/\text{km}$). Finally we allowed the fission kinetic parameter (D_{par}) to vary with a normal prior (mean of $2 \mu\text{m}$, standard deviation of 0.2).

[40] Exploratory runs were made to tune the proposal scale parameters and the final run was made with 60000 post burn-in iterations. The values adopted were: $\sigma_{\text{time}} = 10$ Ma, $\sigma_{\text{temperature}} = 5^{\circ}\text{C}$, $\sigma_{\text{offset}} = 5^{\circ}\text{C}$, $\sigma_{\text{kinetic}} = 1.5$ (no units) and

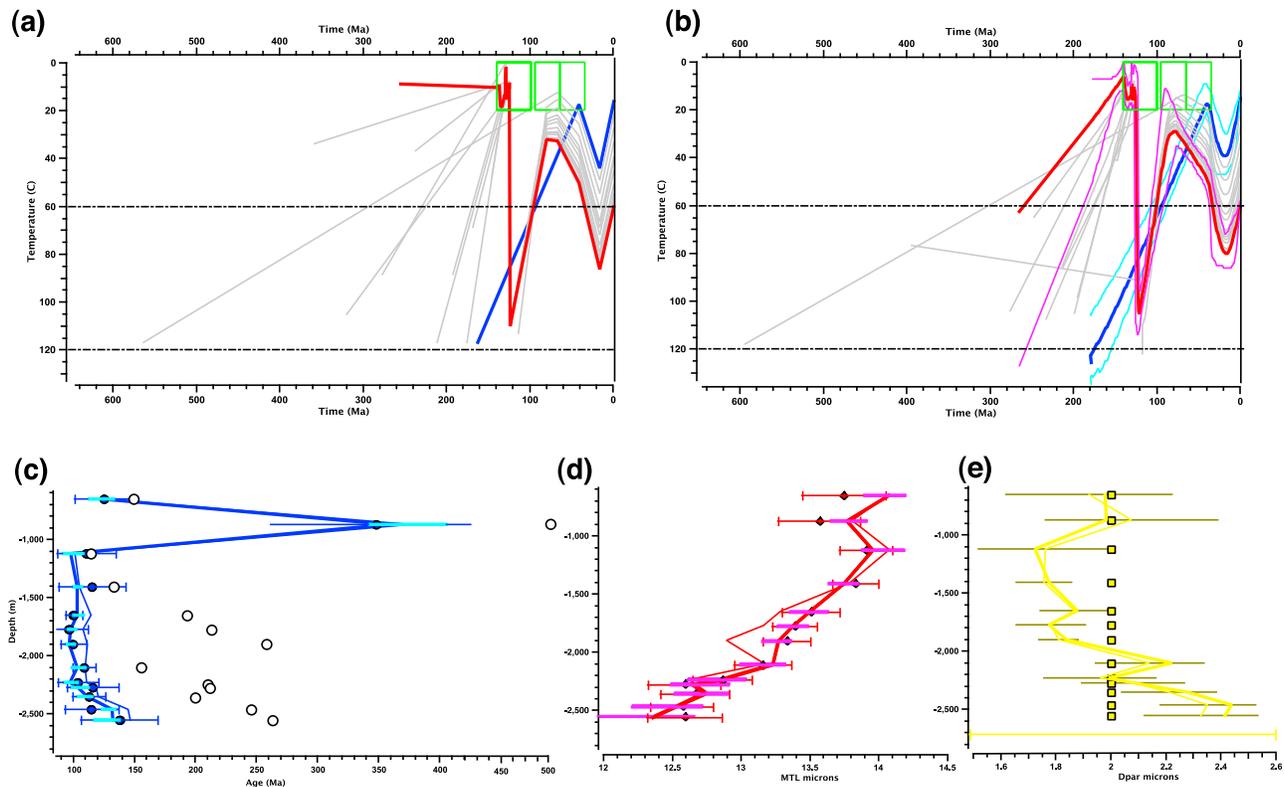


Figure 10. (a) The inferred maximum likelihood thermal history for a suite of borehole data. The three time-temperature boxes at 0–20°C are the input constraints on the stratigraphic age (and deposition temperature). The oldest point in each thermal history reflects a pre-depositional component to the thermal history (required as a consequence of the samples not being totally annealed post-deposition). (b) As Figure 10a but showing the expected thermal history (with the 95% credible intervals on the upper and lower thermal histories). (c) Observed and predicted AFT (filled circles) and the ages of the oldest preserved fission track for each sample predicted from the expected model are shown as open circles. The predicted values for the maximum likelihood and expected models are shown as thick and thin lines (and individual AHe predicted ages are shown as x). Error bars (2σ) are shown on the observed values and the lighter horizontal bars indicate the $\pm 1\sigma$ range on the predicted ages for all accepted models. (d) As Figure 10c but for the observed and predicted MTL values. The predicted values for the maximum likelihood and expected models are shown as thick and thin lines. Error bars are shown on the observed values and the lighter horizontal bars indicate the $\pm 1\sigma$ range on the predicted ages. (e) As Figure 10c but for the input and predicted kinetic parameter (D_{par}) values. Here there were no observed values so the input value was fixed at $2 \pm 0.2 \mu\text{m}$ for all samples. The error bar at the bottom shows the $\pm 3\sigma$ range used for sampling this parameter.

the acceptance rates for each move type were 37%, 45%, 45%, and 27% respectively, and 3% for both birth and death.

[41] The maximum likelihood and expected thermal history models (Figure 10) imply that all samples have an inherited track component as the oldest preserved fission track is inferred to be older than the stratigraphic age. The most extreme case of inherited tracks is the sample at 850 m, with a stratigraphic age of <100 Ma, but an AFT age of 350 Ma. While some of the deeper samples have AFT ages that are slightly younger than the stratigraphic age, the fact that the mean track lengths are often $<13 \mu\text{m}$ implies that the thermal history information is relevant to longer timescales than the post-depositional duration. Thus all samples require some pre-depositional thermal history component. It is clear from Figure 10 that the observed data are well explained with a period of cooling between 120 and 90 Ma (obviously only experienced by the samples that were deposited by that time), moderate reheating to a secondary temperature maximum

around 20 Ma, followed by 20°C cooling to the present-day temperatures, equivalent to about 500 m of erosion with the inferred temperature gradient of about 20°C/km at 20 Ma. The kinetic parameters generally change from the input values, implying a range of D_{par} values from 2.4 down to $1.7 \mu\text{m}$. Fixing the kinetic parameter to $2 \mu\text{m}$ leads to a poorer data fit. However, the form of the thermal history is similar (in terms of the timing of thermal events), although the absolute values of the temperature can differ. For example, at 120 Ma, the maximum temperature of the deepest sample for the fixed model is about 97°C compared to 105°C for the variable kinetics model. After the reheating to 20 Ma, the deepest sample in the fixed model has a temperature of 4°C lower than the variable model. In contrast the shallowest sample is $\sim 10^\circ\text{C}$ hotter. As we expect, this is manifested in the inferred temperature offset at this time for the two models (fixed = 26°C versus variable = 40°C, the latter being closer to the present-day value). These differences reflect

the compositional annealing properties inferred for the variable models. The deeper samples tend to have Dpar greater than the fixed values (so require higher temperatures to produce the same level of annealing) while it is the opposite for the shallower samples. Again, we suggest that variable kinetics gives us some additional freedom in fitting observations, and provides a way of allowing for unaccounted noise in the data and the predictive models.

4. Discussion/Concluding Comments

[42] We have presented a new approach to inverse thermal history modeling for multiple samples (in a vertical profile), with multiple data types, based on transdimensional Bayesian MCMC. The main features of this approach have been demonstrated with examples and these are as follows:

[43] 1. We do not need to specify the complexity of the thermal history. Rather, the data determine this and the Bayesian approach naturally avoids over complex models.

[44] 2. Thermal histories are characterized by a probability distribution which allows a visualization/quantification of the thermal history resolution and the uncertainty on model predictions.

[45] 3. We can select various individual models such as the maximum likelihood thermal history or the expected thermal history.

[46] 4. We can allow for uncertainty in annealing/diffusion model parameters. Here we sample the kinetic parameter for fission track data and the helium age for this type of data (although we could sample uncertainties on the annealing/diffusion model parameters if available).

[47] In the examples, we have only considered apatite fission track and (U-Th)/He data. However, the approach is generally applicable, and could be implemented for $^{40}\text{Ar}/^{39}\text{Ar}$, zircon fission track or U-Th/He, and vitrinite reflectance data. It requires requires the appropriate data-specific forward problem (e.g., annealing, diffusion, organic reactions) and likelihood function (how well we fit the data). Other often poorly constrained parameters, including those relevant to data directly (e.g., apatite composition), or the predictive model (diffusion or annealing model parameters) can be incorporated into the analysis and treated as unknowns characterized by a prior distribution. Here we use a single value for the fission track kinetic parameter for a given sample, but it is possible to split a sample in to sub-groups if the composition of grains varies significantly and treat these as a suite of samples. Finally, this transdimensional approach is readily extended to incorporate the transdimensional spatial modeling approach described by *Stephenson et al.* [2006a], allowing us to identify consistent thermal histories for regional data and to recognize spatial discontinuities in the thermal histories.

Appendix A: Transdimensional Markov Chain Monte Carlo (MCMC) Formulation for Thermal History Modeling

[48] The general theory has been described by *Green* [1995, 2003] and here we follow reasonably closely the approach developed by *Hopcroft et al.* [2007] and *Charvin et al.* [2009]. We define a series of time-temperature-offset points, $t_i, T_i, O_i, i = 1, k$ and for each data set we allow the

possibility of treating the kinetic parameters (c) as unknowns and/or resample the data values (d). The general acceptance function defined in the main text is

$$\alpha(m_p, m_c) = \min\left(1, \frac{p(m_p)p(d|m_p)q(m_c|m_p)}{p(m_c)p(d|m_c)q(m_p|m_c)}|J|\right) \quad (\text{A1})$$

where m_c, m_p are the current and proposed models, respectively, $p(m)$ and $p(d|m)$ are the prior and likelihood functions for model m (their product being proportional to the posterior, $p(m|d)$, through Bayes' rule). The proposal function is $q(m_c|m_p)$ and a requirement in MCMC is that the proposals are reversible, so we have the ratio of the reverse to forward proposals in the acceptance function.

A1. Priors

[49] We write the joint prior probability as

$$\begin{aligned} p(m) &= p(k, t, T, O, c, d) \\ &= p(k)p(t|k)p(T|k)p(O|k)p(c)p(d) \end{aligned} \quad (\text{A2})$$

We choose a uniform prior on k , the number of variable time-temperature points, up to a maximum of k_{\max} , so we have

$$p(k) = \frac{1}{k_{\max}} \quad (\text{A3})$$

We follow *Green* [1995] and *Denison et al.* [2002] and adopt a discrete uniform distribution on the time parameters, which weights each possible model over each dimension (k) equally. Dropping the conditional dependence on k for clarity, the prior on time is then given as

$$p(t) = \frac{k!}{(t_{\max} - t_{\min})^k} \prod_{j=1}^k (t^j - t^{j+1}) \quad (\text{A4})$$

where t_{\max} and t_{\min} are the maximum and minimum values assigned to the time parameters and typically $t_{\min} = t^{k+1} = 0$. This prior tends to avoid proposing points too close to each other.

[50] We assume uniform priors on the temperature and offset parameters parameters, and also that the time, temperature and offsets are independent. These uniform priors are written as

$$p(x) = \frac{1}{x_{\max} - x_{\min}} \quad (\text{A5})$$

where x_{\max} and x_{\min} are the specified upper and lower limits of the prior for a parameter x .

[51] For the kinetic parameters and age data resampling we assume normal prior distributions. For example, in the case of a kinetic parameter, c , is written as

$$p(c) = \frac{1}{\sigma_c \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{c-\bar{c}}{\sigma_c}\right)^2} \quad (\text{A6})$$

where \bar{c} and σ_c are the observed (or user specified) mean and standard deviation of the parameter. We use a similar form for the prior on the age data resampling prior.

A2. Moves/Proposals

[52] We define seven possible model transformations or moves: (1) move time, (2) move temperature, (3) move offset, (4) add a time-temperature point (birth), (5) remove a time-temperature point (death), (6) move kinetic parameter, and (7) move age data value. Each of these moves is selected with a probability p_i , such that $\sum p_i = 1$.

[53] For the first three moves, we randomly select a point with probability $1/k$, where k is the current number of variable time-temperature points and apply a random perturbation to the current value of that parameter. For example, the transform for the i -th time parameter can be written as

$$t_p^i = t_c^i + \sigma_{m,t} u_N \quad (\text{A7})$$

where t_p^i , t_c^i are the proposed and current values for time parameter i , $\sigma_{m,t}$ is the user defined proposal scale for moving time parameters and u_N is a random number drawn from a standard normal distribution, $N(0,1)$. Similar expressions are used for the temperature and offset parameter moves with their own specific proposal scales.

[54] For the kinetic parameter/age data resampling moves, we randomly select a value with probability $1/N$, where N is the number of data sets with a kinetic parameter, or the number of data available to resample (a single data set may have several AHe ages for example). The transform for the j -th kinetic parameter is then written as

$$c_p^j = c_c^j + \sigma_{m,c} u_N \quad (\text{A8})$$

where $\sigma_{c,t}$ is the proposal scale for move a kinetic parameter. A similar transform is made for the age data resampling.

[55] In these moves the proposal functions are symmetrical ($q(m_p|m_c) = q(m_c|m_p)$), so the ratios cancel out in the acceptance function (equation (A1)). For uniform priors, the ratios also cancel out, while for normal priors, we just use the appropriate ratio of the normal distributions. Finally, the Jacobian is always 1 for these moves.

[56] For the birth and death transformations, the acceptance criterion (equation (A1)) is a little more complex. Here we consider the birth move in detail, noting that for death the approach is effectively the same, and we end up using the inverse of the acceptance ratio for birth.

[57] The acceptance function can also be written as [e.g., Green, 2003]

$$\alpha(m_p, m_c) = \text{Min} \left(\frac{p(m_p)p(d|m_p)g(u')j_{m_p}}{p(m_c)p(d|m_c)g(u)j_{m_c}} \left| \frac{\partial(m_p, u')}{\partial(m_c, u)} \right| \right) \quad (\text{A9})$$

where the first three terms in both the denominator and numerator are the prior, likelihood, and proposal distributions used to transform between models (as described above) for the current and proposed models respectively and u are vectors of random numbers used in the transformations. The dimensions, v or v' , of these random number vectors are such that $v + k = v' + k'$, where k and k' are the dimensions the model before and after the transformation. The last term is the Jacobian written in terms of the models and the random numbers.

[58] As the random numbers u are all drawn from a uniform distribution, the ratio of the g terms would be 1 except

we also need to allow for how we decide to add or delete a point. For a birth, we select the new point for the time parameter according to the spacing between the existing points to encourage birth between the more widely spaced points, i.e., the probability of selecting time point, t_i , is given as

$$p(t^i) = \frac{(t^i - t^{i+1})}{(t_{\max} - t_{\min})} \quad (\text{A10})$$

while for the reverse death move we select a point to remove with probability $1/(k+1)$.

[59] The j terms in equation (A9) are the probability involved in selecting a particular move type (m). We define the probability of choosing birth as b_k and choosing death (which is required to reverse the birth move) as d_{k+1} . Thus for birth move, we have the ratio

$$\frac{j_{m_p}}{j_{m_c}} = \frac{d_{k+1}}{b_k} \frac{(t_{\max} - t_{\min})}{(t^i - t^{i+1})} \frac{1}{k+1} \quad (\text{A11})$$

where d and b refer to the probability of choosing a death or birth move, with the subscript being the number of time-temperature points. For a death move, we invert this ratio, and adjust the subscripts according, so we have

$$\frac{j_{m_p}}{j_{m_c}} = \frac{b_{k-1}}{d_k} \frac{(t^i - t^{i+1})}{(t_{\max} - t_{\min})} k \quad (\text{A12})$$

Note that the move proposal probabilities b_k and d_k tend to be equal, except in the special cases when $k=1$ (so we cannot have a death, so we set $d_1=0$, and $b_1=2b_k$) and $k=k_{\max}$ (so we cannot have a birth, so we set $b_{k_{\max}}=0$ and $d_{k_{\max}}=2d_k$).

[60] Having chosen an existing time point, say the i -th point, we then generate a new point at random between this point and the $i+1$ -th point (which is younger than point i). The proposed time is given as

$$t^* = t^i - \sigma_t u_1 (t^i - t^{i+1}) \quad (\text{A13})$$

where $u_1 = U(0,1)$, (where $U(a,b)$ indicates a uniform distribution with lower and upper limits of a and b). The new temperature and offset parameters are generated by linear interpolation between the existing values and then adding a small perturbation to that value (see Figure 1 in the paper). Then the proposed temperature and offset values are given as

$$\begin{aligned} T^* &= T^i - \frac{(T^i - T^{i+1})}{(t^i - t^{i+1})} (t^i - t^*) - \sigma_T u_2 \\ &= T^i - (T^i - T^{i+1}) \sigma_t u_1 - \sigma_T u_2 \end{aligned} \quad (\text{A14})$$

$$O^* = O^i - (O^i - O^{i+1}) \sigma_t u_1 - \sigma_O u_3 \quad (\text{A15})$$

where $u_2, u_3 = U(-0.5, 0.5)$ and the σ terms in equations (A13)–(A15) are the birth proposal scales (which do not need to be equal to the move proposal scales) for each parameter.

[61] With a uniform prior on k ($=1/k_{\max}$), we have the prior ratio for birth given as

$$\frac{p(k+1)}{p(k)} = 1 \quad (\text{A16})$$

The prior ratio given the proposed new time point, t^* , is given as

$$\frac{p(t_p)}{p(t_c)} = \left(\frac{k}{(t_{\max} - t_{\min})} \right) \cdot \frac{(t^i - t^*)(t^* - t^{i+1})}{(t^i - t^{i+1})} \quad (\text{A17})$$

For a uniform prior on the temperature and offset parameters we can write

$$\frac{p(x_p)}{p(x_c)} = \frac{(x_{\max} - x_{\min})^k}{(x_{\max} - x_{\min})^{k+1}} = \frac{1}{(x_{\max} - x_{\min})} \quad (\text{A18})$$

so we obtain the overall prior ratio as

$$\begin{aligned} \frac{p(m_p)}{p(m_c)} &= \left(\frac{k}{(t_{\max} - t_{\min})} \right) \cdot \frac{(t^i - t^*)(t^* - t^{i+1})}{(t^i - t^{i+1})} \\ &\times \frac{1}{(T_{\max} - T_{\min})(O_{\max} - O_{\min})} \end{aligned} \quad (\text{A19})$$

Note that we do not modify any kinetic parameters or data during a birth move, so these terms cancel out in the acceptance ratio and do not appear.

[62] The last term we need is the Jacobian, defined as (with t, T, O the vectors of the current model, and u the random numbers used to produce the proposed model with an extra dimension, t^*, T^*, O^*)

$$\begin{aligned} \left| \frac{\partial(m_p, u')}{\partial(m_c, u)} \right| &= \left| \frac{\partial(t, T, O, t^*, T^*, O^*)}{\partial(t, T, O, u_1, u_2, u_3)} \right| \\ &= (t_i - t_{i+1})\sigma_t\sigma_T\sigma_O \end{aligned} \quad (\text{A20})$$

Collecting all the terms together (priors – (A11), proposals – (A19), Jacobian = (A20)) and incorporating the likelihood ratio, we have the acceptance function for birth (adding a new point to k existing points) defined as

$$\begin{aligned} \alpha(m_p, m_c) &= \frac{p(d|m_p)}{p(d|m_c)} \frac{(t^i - t^*)(t^* - t^{i+1})}{(t^i - t^{i+1})(T_{\max} - T_{\min})(O_{\max} - O_{\min})} \\ &\cdot \frac{d_{k+1}}{b_k} \sigma_t\sigma_T\sigma_O \end{aligned} \quad (\text{A21})$$

For a death move (removing the i -th point from k points) we invert this function, and adjust the indices to obtain

$$\begin{aligned} \alpha(m_p, m_c) &= \frac{p(d|m_p)}{p(d|m_c)} \frac{(t^{i-1} - t^{i+1})(T_{\max} - T_{\min})(O_{\max} - O_{\min})}{(t^{i-1} - t^i)(t^i - t^{i+1})} \\ &\cdot \frac{b_{k-1}}{d_k} \frac{1}{\sigma_t\sigma_T\sigma_O} \end{aligned} \quad (\text{A22})$$

[63] **Acknowledgments.** I would like to thank Karl Charvin, Peter Hopcroft, John Stephenson, Malcolm Sambridge, and Thomas Bodin for many rambling but generally enlightening discussions on Bayesian transdimensional MCMC and Andy Carter, Nathan Cogné, Yuntao Tian, and Boris Avdeev for persevering with early versions of the software implementing the approach presented here (available by contacting the author). Thanks to Trevor Dumitru and an anonymous reviewer for useful comments on an earlier version of this manuscript and Joe Colgan for his data. Congratulations to Richie and the rest of the team...it was a long wait.

References

- Bernardo, J., and A. F. M. Smith (1994), *Bayesian Theory*, John Wiley, Chichester, U. K.
- Bodin, T., and M. Sambridge (2009), Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, *178*, 1411–1436, doi:10.1111/j.1365-246X.2009.04226.x.
- Braun, J., P. van der Beek, and G. Batt (2006), *Quantitative Thermochronology*, Cambridge Univ. Press, New York, doi:10.1017/CBO9780511616433.
- Carlson, W. D., R. A. Donelick, and R. A. Ketcham (1999), Variability of apatite fission-track annealing kinetics I: Experimental results, *Am. Mineral.*, *84*, 1213–1223.
- Charvin, K., K. Gallagher, G. Hampson, and R. Labourdette (2009), A Bayesian approach to infer environmental parameters from stratigraphic data 1: Methodology, *Basin Res.*, *21*, 5–25, doi:10.1111/j.1365-2117.2008.00369.x.
- Colgan, J. P., T. A. Dumitru, P. W. Reiners, J. L. Wooden, and E. L. Miller (2006), Cenozoic tectonic evolution of the basin and range province in northwestern Nevada, *Am. J. Sci.*, *306*, 616–654, doi:10.2475/08.2006.02.
- Corrigan, J. (1991), Inversion of fission track data for thermal history information, *J. Geophys. Res.*, *96*, 10,347–10,360, doi:10.1029/91JB00514.
- Denison, D. G. T., C. C. Holmes, B. K. Mallick, and A. F. M. Smith (2002), *Bayesian Methods for Nonlinear Classification and Regression*, Wiley, Chichester, U. K.
- Dodson, M. H. (1973), Closure temperature in cooling geochronological and petrological systems, *Contrib. Mineral. Petrol.*, *40*, 259–274, doi:10.1007/BF00373790.
- Donelick, R. A., R. A. Ketcham, and W. D. Carlson (1999), Variability of apatite fission-track annealing kinetics II: Crystallographic orientation effects, *Am. Mineral.*, *84*, 1224–1234.
- Efron, B., and R. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York.
- Farley, K. A. (2000), Helium diffusion from apatite: General behaviour as illustrated by Durango fluorapatite, *J. Geophys. Res.*, *105*, 2903–2914, doi:10.1029/1999JB900348.
- Gallagher, K. (1995), Evolving temperature histories from apatite fission-track data, *Earth Planet. Sci. Lett.*, *136*, 421–435, doi:10.1016/0012-821X(95)00197-K.
- Gallagher, K., J. A. Stephenson, R. W. Brown, C. C. Holmes, and P. Fitzgerald (2005), Low temperature thermochronology and strategies for multiple samples 1: Vertical profiles, *Earth Planet. Sci. Lett.*, *237*, 193–208, doi:10.1016/j.epsl.2005.06.025.
- Gallagher, K., K. Charvin, S. Nielsen, M. Sambridge, and J. Stephenson (2009), Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems, *Mar. Pet. Geol.*, *26*, 525–535, doi:10.1016/j.marpetgeo.2009.01.003.
- Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (1996), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.
- Gleadow, A. J. W., I. R. Duddy, P. F. Green, and J. F. Lovering (1986), Confined track lengths in apatite: A diagnostic tool for thermal history analysis, *Contrib. Mineral. Petrol.*, *94*, 405–415, doi:10.1007/BF00376334.
- Green, P. F., I. R. Duddy, A. J. W. Gleadow, P. R. Tingate, and G. M. Laslett (1986), Thermal annealing of fission tracks in apatite: 1. A qualitative description, *Chem. Geol. Isot. Geosci. Sect.*, *59*, 237–253, doi:10.1016/0168-9622(86)90074-6.
- Green, P. F., I. R. Duddy, G. M. Laslett, K. A. Hegarty, A. J. W. Gleadow, and J. F. Lovering (1989), Thermal annealing of fission tracks in apatite: 4. Quantitative modelling techniques and extension to geological timescales, *Chem. Geol. Isot. Geosci. Sect.*, *79*, 155–182, doi:10.1016/0168-9622(89)90018-3.
- Green, P. J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, *82*, 711–732, doi:10.1093/biomet/82.4.711.
- Green, P. J. (2003), Transdimensional MCMC, in *Highly Structured Stochastic Systems*, *Oxford Stat. Sci. Ser.*, vol. 27, edited by P. J. Green, N. Hjort, and S. Richardson, chap. 6, pp. 179–196, Oxford Univ. Press, Oxford, U. K.
- Harrison, T. M., M. Grove, O. M. Lobvera, and P. K. Zeitler (2005), Continuous thermal histories from inversion of closure profiles, *Rev. Mineral. Geochem.*, *58*, 389–409, doi:10.2138/rmg.2005.58.15.
- Hopcroft, P., K. Gallagher, and C. C. Pain (2007), Inference of past climate from borehole temperature data using Bayesian Reversible Jump Markov chain Monte Carlo, *Geophys. J. Int.*, *171*, 1430–1439, doi:10.1111/j.1365-246X.2007.03596.x.
- Hopcroft, P., K. Gallagher, and C. C. Pain (2009), A Bayesian partition modelling approach to resolve spatial variability in climate records from

- borehole temperature inversion, *Geophys. J. Int.*, *178*, 651–666, doi:10.1111/j.1365-246X.2009.04192.x.
- Jasra, A., D. A. Stephens, K. Gallagher, and C. C. Holmes (2006), Analysis of geochronological data with measurement error using Bayesian mixtures, *Math. Geol.*, *38*, 269–300, doi:10.1007/s11004-005-9019-3.
- Jones, S. M., and R. K. Dokka (1990), Modelling fission-track annealing in apatite—An assessment of uncertainties, *Nucl. Tracks Radiat. Meas.*, *17*(3), 255–260, doi:10.1016/1359-0189(90)90043-W.
- Ketcham, R. A. (2005), Forward and inverse modelling of low-temperature thermochronometry data, *Rev. Mineral. Geochem.*, *58*, 275–314, doi:10.2138/rmg.2005.58.11.
- Ketcham, R. A., R. A. Donelick, and W. D. Carlson (1999), Variability of apatite fission-track annealing kinetics: III. Extrapolation to geological timescales, *Am. Mineral.*, *84*, 1235–1255.
- Ketcham, R. A., R. A. Donelick, and M. B. Donelick (2000), AFTSolve: A program for multi-kinetic modeling of apatite fission-track data, *Geol. Mater. Res.*, *2*, 1–32.
- Ketcham, R. A., A. Carter, R. A. Donelick, J. Barbarand, and A. J. Hurford (2007), Improved modelling of fission-track annealing in apatite, *Am. Mineral.*, *92*, 799–810, doi:10.2138/am.2007.2281.
- Laslett, G. M., P. F. Green, I. R. Duddy, and A. J. W. Gleadow (1987), Thermal annealing of fission tracks in apatite: 2. A quantitative analysis, *Chem. Geol. Isot. Geosci. Sect.*, *65*, 1–13, doi:10.1016/0168-9622(87)90057-1.
- Lutz, T. M., and G. I. Omar (1991), An inverse method of modelling thermal histories from apatite fission-track data, *Earth Planet. Sci. Lett.*, *104*, 181–195, doi:10.1016/0012-821X(91)90203-T.
- Malinverno, A. (2002), Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.*, *151*(3), 675–688, doi:10.1046/j.1365-246X.2002.01847.x.
- McDougall, I., and T. M. Harrison (1999), *Geochronology and Thermochronology by the $^{40}\text{Ar}/^{39}\text{Ar}$ Method*, 2nd ed., Oxford Univ. Press, New York.
- Piana Agostinetti, N., and A. Malinverno (2010), Receiver function inversion by transdimensional Monte Carlo sampling, *Geophys. J. Int.*, *181*, 858–872.
- Reiners, P. W., and T. A. Ehlers (Eds.) (2005), *Low-Temperature Thermochronology: Techniques, Interpretations and Applications*, *Rev. Mineral. Geochem.*, *58*, 622 pp.
- Reiners, P. W., and K. A. Farley (2001), Influence of crystal size on (U-Th)/He thermochronology: Re 1 an example from the Bighorn Mountains, Wyoming, *Earth Planet. Sci. Lett.*, *188*, 413–420, doi:10.1016/S0012-821X(01)00341-7.
- Reiners, P. W., T. L. Spell, S. Nicolescu, and K. A. Zanetti (2004), He diffusion and (U-Th)/He thermochronometry: He diffusion and comparison with $^{40}\text{Ar}/^{39}\text{Ar}$ dating, *Geochim. Cosmochim. Acta*, *68*, 1857–1887, doi:10.1016/j.gca.2003.10.021.
- Sambridge, M., K. Gallagher, A. Jackson, and P. Rickwood (2006), Transdimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.*, *167*, 528–542, doi:10.1111/j.1365-246X.2006.03155.x.
- Scales, J., and R. Snieder (1998), What is noise?, *Geophysics*, *63*, 1122–1124, doi:10.1190/1.1444411.
- Schwarz, G. (1978), Estimating the dimension of a model, *Ann. Stat.*, *6*, 461–464, doi:10.1214/aos/1176344136.
- Stephenson, J., K. Gallagher, and C. Holmes (2006a), Low temperature thermochronology and modelling strategies for multiple samples 2: Partition modelling for 2D and 3D distributions with discontinuities, *Earth Planet. Sci. Lett.*, *241*, 557–570, doi:10.1016/j.epsl.2005.11.027.
- Stephenson, J., K. Gallagher, and C. Holmes (2006b), A Bayesian approach to calibrating apatite fission track annealing models for laboratory and geological timescales, *Geochim. Cosmochim. Acta*, *70*, 5183–5200, doi:10.1016/j.gca.2006.07.027.
- Willett, S. D. (1997), Inverse modelling of annealing of fission tracks in apatite 1: A controlled random search method, *Am. J. Sci.*, *297*, 939–969, doi:10.2475/ajs.297.10.939.

K. Gallagher, Géosciences, Université de Rennes 1, Campus de Beaulieu, F-35042 Rennes CEDEX, France. (kerry.gallagher@univ-rennes1.fr)