A Bayesian partition modelling approach to resolve spatial variability in climate records from borehole temperature inversion

Peter-O. Hopcroft, Kerry Gallagher, Chris-C. Pain

Geophys. J. Int. (2009) **178,** 651–666 doi: 10.1111/j.1365-246X.2009.04192.x

A Bayesian partition modelling approach to resolve spatial variability in climate records from borehole temperature inversion

Peter O. Hopcroft,[1]* Kerry Gallagher[2] and Christopher C. Pain[1]

[1]*Department of Earth Science & Engineering, Imperial College, London,* SW7 2AZ, *UK. E-mail: peter.hopcroft@bris.ac.uk*
[2]*Géosciences Rennes, Université de Rennes* 1, *Campus de Beaulieu, Rennes,* 35042, *France*

Accepted 2009 March 24. Received 2009 March 24; in original form 2008 June 17

SUMMARY

Collections of suitably chosen borehole profiles can be used to infer large-scale trends in ground-surface temperature (GST) histories for the past few hundred years. These reconstructions are based on a large database of carefully selected borehole temperature measurements from around the globe. Since non-climatic thermal influences are difficult to identify, representative temperature histories are derived by averaging individual reconstructions to minimize the influence of these perturbing factors. This may lead to three potentially important drawbacks: the net signal of non-climatic factors may not be zero, meaning that the average does not reflect the best estimate of past climate; the averaging over large areas restricts the useful amount of more local climate change information available; and the inversion methods used to reconstruct the past temperatures at each site must be mathematically identical and are therefore not necessarily best suited to all data sets. In this work, we avoid these issues by using a Bayesian partition model (BPM), which is computed using a trans-dimensional form of a Markov chain Monte Carlo algorithm. This then allows the number and spatial distribution of different GST histories to be inferred from a given set of borehole data by partitioning the geographical area into discrete partitions. Profiles that are heavily influenced by non-climatic factors will be partitioned separately. Conversely, profiles with climatic information, which is consistent with neighbouring profiles, will then be inferred to lie in the same partition. The geographical extent of these partitions then leads to information on the regional extent of the climatic signal. In this study, three case studies are described using synthetic and real data. The first demonstrates that the Bayesian partition model method is able to correctly partition a suite of synthetic profiles according to the inferred GST history. In the second, more realistic case, a series of temperature profiles are calculated using surface air temperatures of a global climate model simulation. In the final case, 23 real boreholes from the United Kingdom, previously used for climatic reconstructions, are examined and the results compared with a local instrumental temperature series and the previous estimate derived from the same borehole data. The results indicate that the majority (17) of the 23 boreholes are unsuitable for climatic reconstruction purposes, at least without including other thermal processes in the forward model.

Key words: Spatial analysis; Probability distributions; Heat flow.

1 INTRODUCTION

Borehole temperatures can be used to reconstruct surface temperature variations over past centuries (e.g. Lachenbruch & Marshall 1986; Mareschal & Beltrami 1992; Pollack & Huang 2000; Majorowicz *et al.* 2004). This method of climate reconstruction relies on the assumption that surface-air temperature variations are propagated into the ground by heat conduction. Accurate measurements of temperature–depth profiles are then used to constrain heat transfer at the surface in the past. Since the thermal diffusivity of the underlying rocks is low, of the order $10^{-6}\,m^2\,s^{-1}$, important information on the last 500–750 yr can be derived from temperatures measured in boreholes of around 500 m depth. However, this low thermal diffusivity also means that the resolution over time of the derived reconstructions is very low, and due to diffusion of the thermal signal with time, the resolution decreases rapidly with time before the present. Nonetheless, reconstructions derived from borehole temperatures are important because they are physically based, and so, no empirical calibration is required (in contrast to other types of proxy data, for example), and they are entirely independent

*Now at: BRIDGE, School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK.

of proxy and other physically based reconstruction techniques. Borehole data also provide important constraints on low-frequency climate variability (e.g. Pollack & Smerdon 2004; Hegerl *et al.* 2007), for which the reliability of other proxy data sources has been questioned (e.g. von Storch *et al.* 2004).

Large scale estimates of temperature changes derived from borehole temperatures consistently indicate around 1.0 °C of warming over the last 500 yr (Huang *et al.* 2000; Beltrami & Bourlon 2004; Pollack & Smerdon 2004; Harris & Chapman 2005), which is larger than many estimates from multiproxy data (e.g. Briffa & Osborn 2002), although more recent studies show good agreement with the borehole inferred amplitude (Esper *et al.* 2002; Moberg *et al.* 2005; Hegerl *et al.* 2007). However, there remains considerable uncertainty concerning the magnitude and timing of temperature change over this time period (Jansen *et al.* 2007), and further analysis of borehole data is therefore warranted.

The estimates derived from boreholes are calculated by averaging hundreds of individual reconstructions to reduce the effects of local perturbing factors, the thermal effects of which can be misidentified as due to changing climate. This averaging means that regional scale variations are not inferred. Furthermore, the net effect of the perturbing factors may not cancel upon averaging, and so, this could imply that the average of all of the reconstructions is not most representative of past climate. This issue has been tested in the context of a global climate model (González-Rouco *et al.* 2003, 2006), and the results support the borehole estimate, indicating that the influence of precipitation (particularly snow cover), vegetation and freezing and thawing of soil moisture are less important. However, other perturbing factors not tested in this approach include: site topography (see Hopcroft *et al.* 2009), subsurface heterogeneity, agricultural expansion and deforestation and underground fluid flow. These may all be significant, although it is plausible that some of these perturbing factors, for example, land use change, could be expected to produce systematic biases in the reconstructed temperature histories, whereas others, such as subsurface heterogeneity, may exert a more randomized influence. Finally, in deriving these individual reconstructions by inverting each individual borehole profile, the parameters of the inverse model must be kept the same for each site and, so, cannot necessarily be tailored to be optimal in each case (Huang *et al.* 1996, 2000; Beltrami *et al.* 2003).

In this work, we describe a novel method for inferring regional-scale past temperature variations from multiple borehole profiles, which additionally allows the identification of profiles resulting in inconsistent temperature histories. Our method does not rely on comparison or calibration with instrumental temperatures (e.g. Mann *et al.* 2003; Rutherford & Mann 2004) and explicitly accounts for uncertainty in both the borehole data and the inferred models. We use recent developments in Bayesian partition modelling (BPM; Denison *et al.* 2002; Stephenson *et al.* 2006), which allow us to consider spatial signals of temperature change in a flexible manner. In this, models of differing parametrizations can be compared robustly using a trans-dimensional Bayesian approach, which is calculated using Reversible jump-Markov chain Monte Carlo (RJ-MCMC; Green 1995, 2003; Sambridge *et al.* 2006). The BPM method is then used to infer the number and spatial distribution of independent ground temperature histories, which are supported by a given set of borehole profiles, given the estimated level of data noise (e.g. Backus & Gilbert 1970). Profiles that are influenced by non-climatic factors will be inferred to lie individually in separate partitions. Conversely, borehole profiles that are consistent with other neighbouring profiles will be inferred to lie in partitions with these other profiles. The ground-surface temperature (GST) histories relating to these larger partitions are then likely to be more robust, because the corrupted profiles are not included. Additionally, the spatial extent of these larger partitions will provide information on the regional extent of the inferred GST signal (e.g. Majorowicz *et al.* 2002; Beltrami *et al.* 2003).

The GST history in each partition is inferred jointly from all of the data contained in that partition, again using a trans-dimensional Bayesian approach computed with RJ-MCMC and as described previously in Hopcroft *et al.* (2007). The GST model parametrization itself is less likely to induce spurious GST histories, as demonstrated by Huang *et al.* (1996), since simpler models are automatically preferred by the Bayesian method and also because the approximate timing of past GST changes is also estimated in this method, by allowing the time points of the GST history to move along the time axis. Additionally, by relying on the natural parsimony of the Bayesian approach (e.g. Malinverno 2002), no temporal smoothing of the GST model parameters is required, meaning that the magnitude of temperature change is more objectively inferred from the data, and this is crucial given the level of uncertainty surrounding the magnitude of temperature change over the past millennium (Jansen *et al.* 2007). As in previous Bayesian borehole methods (e.g. Huang *et al.* 2000), uncertainty in both the data values and model parameters is explicitly accounted for. Here, however, the assumption that the inferred model parameters follow a Gaussian distribution is not required, allowing for more robust quantification of the uncertainty and any non-uniqueness in the posterior distribution.

In this paper, we provide three case studies to illustrate the BPM methodology. In the first synthetic case, 10 synthetic profiles have been calculated using three synthetic temperature reconstructions with largely differing trends. The BPM method is able to reconstruct the three spatial partitions applied to calculate the data and also to faithfully reconstruct the temperature evolution applied in each partition, and this is possible in the presence of data noise. In the second, more realistic example, 10 synthetic temperature profiles are calculated using surface-air temperature series at 10 gridpoints of a coupled ocean–atmosphere global climate simulation of last 250 yr. The BPM method is used to determine how much of the more realistic air temperature variations in space (and time) can be discerned using borehole data. The results show that the BPM is able to effectively divide up the geographical space according to the inferred GST history, and that for each gridpoint, the inferred GST reconstruction corresponds well with the applied air temperature series.

In both synthetic examples, only measurement errors are introduced, and so, there are no non-climatic signals present in the data. In the third case, a real data set of 23 borehole temperature profiles from the United Kingdom is examined to test whether the borehole data can be used to infer robust climatic signals, in the presence of potentially non-climatically perturbed data. The UK data are appropriate for this purpose since the small geographic coverage implies that the borehole sites used should have been subject to similar long-term air temperature variations, and because a long and reliable instrumental temperature record, the Central England Temperature (CET) series (Manley 1974; Parker *et al.* 1992), provides a good basis for comparison with the inferred GST histories.

This paper is structured as follows: in Section 2, the forward model used in this work is briefly described; in Sections 3 and 4, partition modelling and Bayesian inference are introduced; in Section 5, we describe the implementation of the BPM sampling algorithm used here, whereas a full mathematical description is given in the Appendix; in Section 6, the results from the two synthetic

cases are shown; and in Section 7, the results of the real data application are given. In Section 8, a discussion of the results from both synthetic and real data is given. This is followed by the conclusions, which can be drawn from the present study.

2 FORWARD MODEL

The forward model we use has been described in Hopcroft *et al.* (2007), and we only briefly describe this here. Heat flow is assumed to be purely vertical with no advective component. Thus thermal perturbations from the long term equilibrium at depth are the result of time-varying GST only. The relevant form of the heat conduction equation in 1-D is

$$\rho C \frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(k_c \frac{\partial T}{\partial z}\right), \tag{1}$$

where z is depth, t is time and ρ, c and k_c are the rock density, specific heat capacity and thermal conductivity, respectively. The present-day temperature–depth profile can then be expressed in terms of the equilibrium thermal conditions, the past surface temperature variations and random measurement error:

$$T(t_{\text{present}}, z) = T_{\text{eq}} + q_0 \int_0^{z\,\max} \frac{1}{k_c}\,\mathrm{d}z + T_s(z) + \epsilon_d, \tag{2}$$

where T_{eq} is the long-term equilibrium surface temperature, q_0 is the basal heat flux, $T_s(z)$ is the subsurface perturbation at depth z due to past surface temperature variations and ϵ_d is the measurement error. As in Hopcroft *et al.* (2007), eq. (1) is solved numerically by the method of finite elements. To estimate the background basal heat flow, a steady state solution is first derived, and this is then perturbed using a transient FE model to give the final temperature–depth values. These are then compared with the measured data through the likelihood function (see Section 5).

3 BAYESIAN PARTITION MODELLING USING VORONOI TESSELLATIONS

To divide up the space over which the boreholes are located, a Voronoi tessellation approach is used (e.g. Green 1995; Denison *et al.* 2002). This allows divisions of a specified space into regions defined by tessellation centres C_i. Each Voronoi tessellation region is formed by including in a tessellation, say R_1 associated with centre C_1, all the points in space which are closer to C_1 than to any other centre, C_i. This distance can be measured in a variety of manners (Denison *et al.* 2002), but in this work, the standard Euclidian distance metric is used:

$$D(\mathbf{x}_a, \mathbf{x}_b) = \sum_i (x_{ai} - x_{bi})^2. \tag{3}$$

Fig. 1 depicts an example Voronoi tessellation that has eight centres.

In this work, each partition has an independent GST history. Thus, the number of partitions, n_c can vary over the range $n_c = (1, n_b)$ where n_b is the number of boreholes. When the number of partitions is 1, all of the borehole profiles fall within the same Voronoi region and, so, are fitted with a common GST history. For $n_c = n_b$, each borehole is located in a separate Voronoi cell, and so, a different GST history is fitted to each individual borehole profile. In this case, we would expect to obtain the best fit to the observed data, as each borehole is treated independently. However, the aim of BPM is to find an appropriate and objective balance between the number of partitions (in a sense, a measure of the model complexity) and the quality of the fit to the observed data. The prior information plays a key role in achieving this balance (see Section A1). The results of this Bayesian partition method are conditional on the estimated data noise and the prescribed prior information, both of which are discussed in detail in Section 5.

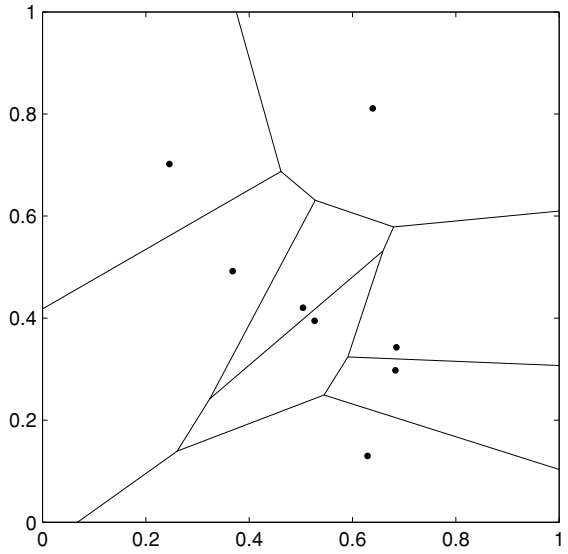

Figure 1. An example of a Voronoi tessellation with eight Voronoi centres (black dots). Any point lying within a particular Voronoi cell is closer to that cell's centre than any other centre. The boundaries of a cell are defined by the perpendicular bisectors of the line segments joining from neighbouring centres to the centre of that cell.

One of the key assumptions used in this method (see also Denison *et al.* 2002) is that the GST histories in different partitions (Voronoi cells) are independent. However, in reality, long-term temperature trends vary continuously and with spatial correlations across large areas. The BPM method is often used in cases for which the underlying pattern of interest is smoother or of a different form than the individual model realizations used (here, for example, the Voronoi partitions, Fig. 1). However, the predictive posterior model, which is found by averaging over all of the models in a sample (in proportion to their posterior probability values), will result in a generally smoothly varying function that will also respect any discontinuities evident in the data (e.g. Stephenson *et al.* 2004), as these will be present in the majority of the samples. Therefore, the actual parametrization of the partitioning is less important in the overall modelling scheme than, say, the chosen prior information. How this issue relates to the present borehole inversion problem will be discussed further, alongside the synthetic and real data examples.

4 BAYESIAN INFERENCE VIA MARKOV CHAIN MONTE CARLO

The parameters of interest are $\mathbf{C}$, the locations of the Voronoi centres, $\boldsymbol{\theta}$ the GST histories in each partition and $\mathbf{q}_0$ and $\mathbf{T}_{\text{eq}}$ the basal heat flux and surface equilibrium temperature at each borehole, respectively. These are encapsulated by a total model parameter vector:

$$\mathbf{m} = (\mathbf{C}, \boldsymbol{\theta}, \mathbf{q}_0, \mathbf{T}_{\text{eq}}). \tag{4}$$

where $\mathbf{C}_i$ are the coordinates of the Voronoi centres and $\boldsymbol{\theta}_i$ is a $2 \times k_i$ matrix, containing the time and temperature points, t_j, T_j, $j = 1, k_i$, in the GST history i. $\boldsymbol{\theta}_i$ are defined for each partition, and q_0^j and T_{eq}^j are defined for each borehole.

In Bayesian formalism, no distinction is made between data and model parameters, and so, in theory, all of the parameters of the simulation could be incorporated into **m**; including, for example, the thermal conductivity values and even the temperature data themselves (e.g. Gallagher 1990). In this work, we limit **m** to the Voronoi centres and those parameters that have been of most interest in similar work, namely the GST histories and the background equilibrium thermal conditions (Huang *et al.* 2000; Beltrami *et al.* 2003). Casting the problem in a probabilistic framework using Bayesian inference (e.g. Mosegaard & Tarantola 1995; Tarantola 2005) allows for uncertainty in all of these model parameters, which can be quantified through the associated conditional posterior probability density functions (pdf). The joint pdf of the four model parameter vectors conditioned on the data and prior information is given by Bayes' law and is termed the posterior (e.g. Bernardo & Smith 1994; Sivia & Skilling 2006):

$$p(\mathbf{m} \mid \mathbf{d}, \wp) = \frac{p(\mathbf{m} \mid \wp) \times p(\mathbf{d} \mid \mathbf{m}, \wp)}{p(\mathbf{d} \mid \wp)}, \tag{5}$$

where p is probability and $p(a \mid b, c)$ means the probability of a given b and c. **m** and **d** are the model and data vectors and $\wp$ is the theory or hypothesis underlying the model formulation and incorporates the prior constraints. In words, Bayes' Law can be expressed as

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}. \tag{6}$$

The goal of Bayesian inference is to determine the posterior probability distribution. To do this, we resort to Markov chain Monte Carlo methods (e.g. Gilks *et al.* 1996), which allow us to sample the posterior up to a constant of proportionality [i.e. we neglect the evidence term, $p(\mathbf{d} \mid \wp)$].

In this work, we employ RJ-MCMC (Green 1995), as it allows inference on both model parameters and model dimensionality (see e.g. Malinverno 2002). Like the more well-known Metropolis–Hastings, RJ-MCMC constitutes a two stage process of proposing a model probabilistically and then accepting or rejecting this proposed model. The proposal is made by drawing from a probability distribution $q(\mathbf{m}' \mid \mathbf{m})$ such that a new proposed model $\mathbf{m}'$ is conditional only on the current model **m**. The new model is then accepted with a probability min $[1, \alpha(\mathbf{m}', \mathbf{m})]$. If the model is accepted, the current model **m** is replaced by $\mathbf{m}'$, which becomes the current model for the next iteration. If it is not accepted, the current model **m** is retained for another iteration. This process is then iterated many times so that after a period of initial exploration of the model space (referred to as the burn-in), a series of samples of the model parameters is collected. It can be shown that for any proposal distribution, the stationary probability distribution sampled in this way will be a good estimate of the true probability distribution (Gilks *et al.* 1996).

To ensure convergence of the sampling distribution to the true distribution, each transformation of the model needs to satisfy detailed balance (see Green 1995, 2003). This is accounted for by the exact form of the acceptance probability, α. When the number of dimensions does not change during a proposal, α can be written in the same form as the well-known Metropolis–Hastings sampler. However, the more general form of acceptance probability is

$$\begin{aligned}\alpha &= \min\left[1, \frac{p(\mathbf{m}' \mid \wp)}{p(\mathbf{m} \mid \wp)} \frac{p(\mathbf{d} \mid \mathbf{m}', \wp)}{p(\mathbf{d} \mid \mathbf{m}, \wp)} \frac{q(\mathbf{m} \mid \mathbf{m}')}{q(\mathbf{m}' \mid \mathbf{m})} \left|\frac{\partial(\mathbf{m}', \mathbf{u}')}{\partial(\mathbf{m}, \mathbf{u})}\right|\right] \\ &= \min\big(1, (\text{prior ratio}) \\ &\quad (\text{likelihood ratio})\,(\text{proposal ratio})|\text{Jacobian}|\big),\end{aligned} \tag{7}$$

where **u** and $\mathbf{u}'$ are vectors of random numbers, which are used to transform the current model **m** to the proposed model $\mathbf{m}'$, which may or may not be of a different dimension. In most practical fixed dimension cases, the Jacobian is 1 and can be ignored. For variable dimension models, the form of the Jacobian depends on the form of the transformation between dimensions. The derivation of the correct values for the Jacobian as well as the prior, proposal and likelihood ratios are given in the Appendix, where our implementation of the RJ-MCMC algorithm is described in detail.

As implied earlier, we might anticipate that the RJ-MCMC algorithm will tend to a solution with n_b partitions, such that each borehole profile is individually fitted, and we achieve the lowest possible misfit to all the data. However, Bayesian inference is naturally parsimonious and tends to avoid unnecessarily complex models (Jefferys & Berger 1992). To illustrate this, consider the evidence (see eq. 5) for two models m_1 and m_2, where m_1 is the simpler of the two; for example, it could be formulated with a smaller number of free parameters. As a result of the normalization of the pdf, the evidence for the simpler model m_1 is larger, as it is spread over a smaller region in data space (Denison *et al.* 2002; MacKay 2003). We do not calculate the evidence, instead the RJ-MCMC algorithm samples models of different dimensionality in proportion to their relative evidence values (Green 2003; Sambridge *et al.* 2006). Thus, the method will tend to favour the simplest model that can adequately fit the data. This therefore relies on the *a priori* estimates of the noise on the data. Here we fix the estimates of the data noise *a priori* at realistic values, which are given alongside the examples in Sections 6 and 7.

5 THE BAYESIAN PARTITION MODEL FOR BOREHOLE TEMPERATURES

The prescribed prior information has been designed to be as general as possible while providing sufficient information for constraining the model. The number and positions of the partitions are constrained using a uniform prior probability distribution, with the limits chosen to prevent partitions centres from being outside of the spatial domain and to prevent more partitions than boreholes in the model. Apart from this, the uniform distribution implies no preference for a certain number of partitions or a certain set of locations of the partitions.

The prior distributions used to constrain the GST histories are given in Hopcroft *et al.* (2007) and also in the Appendix. Briefly, these priors imply no change of GST over time and also no correlation between different times. This latter point means that no explicit temporal smoothing of the temperature history is used. Additionally, for simplicity, we impose the same prior uncertainty at all times in the past, which is similar to, for example, Huang *et al.* (2000). Uniform prior constraints are used for the number of points used to parametrize each GST history and also the heat flux and pre-reconstruction equilibrium surface temperature at each site.

To sample from the posterior pdf of the model parameters **m** (see eq. 4) using RJ-MCMC as described previously, a set of probabilistic proposal functions is required. Here six such proposals functions are used, and these mainly focus on the partition structure. One of these proposals is randomly selected at each iteration of the RJ-MCMC algorithm. The proposed model $\mathbf{m}'$ is then accepted according the acceptance probability (eq. 7). The six proposal options we select from are:

(1) Birth: add a new Voronoi centre with the position found by drawing from the prior distribution on **C**;

(2) Death: delete one Voronoi centre chosen randomly from the current set;

(3) Move an existing Voronoi centre to a position drawn from the prior on **C**;

(4) Perturb one Voronoi centre by drawing from a normal distribution centred on its current position;

(5) Update the heat flux ($\mathbf{q}_0$) and equilibrium surface temperature ($\mathbf{T}_{\rm eq}$) for one borehole site;

(6) Propose a new GST history ($\boldsymbol{\theta}_i$) for partition *i*.

The appropriate acceptance probability α for each proposal type is given in the Appendix.

Since the GST history in each different partition is unknown, a secondary RJ-MCMC sampling algorithm is additionally used to approximate the conditional posterior of the GST history conditioned on the data within that partition $p(\boldsymbol{\theta}_i \mid \mathbf{d}_{s(i)}, \wp)$, where s*(i)* indicates the borehole profiles in a partition *i*. This distribution is then used as the basis for the proposal distribution each time a new GST history is required. This occurs when option (6) is selected or because one of options (1)–(4) causes a new configuration of the data with respect to the partition structure. The details of the secondary sampler are the same as the method used in Hopcroft *et al.* (2007), with some minor modifications described in Appendix.

The likelihood term for a particular borehole quantifies the fit of the simulated data calculated using the GST history to the measured data. In this work, the likelihood term is assumed to be a multivariate Gaussian distribution with uncorrelated parameters (see also for example Malinverno 2002) and is a function of the calculated temperature profile (using finite elements) and the measured temperature profile:

$$p(\mathbf{d} \mid \mathbf{m}, \mathbf{k}, \wp) = \frac{1}{[(2\pi)^n \det \mathbf{C}_{\rm d}]^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{d}_{\rm sim} - \mathbf{d}_{\rm obs})^T \mathbf{C}_{\rm d}^{-1} \times (\mathbf{d}_{\rm sim} - \mathbf{d}_{\rm obs})\right], \quad (8)$$

where there are n data points, $\mathbf{d}_{\rm sim}$ and $\mathbf{d}_{\rm obs}$ are the simulated and observed subsurface temperatures, respectively, and $\mathbf{C}_{\rm d}$ is the data covariance matrix. By assuming the errors on the measured data in each borehole are statistically independent of other boreholes, the joint likelihood for an ensemble of borehole profiles can be simplified; thus,

$$\begin{aligned} p(\mathbf{d}_{j=1,n_b} \mid \mathbf{m}, \mathbf{k}, \wp) &= p(\mathbf{d}_1 \mid \mathbf{m}_{v(1)}, \mathbf{k}_{v(1)}, \wp) \\ &\quad \times p(\mathbf{d}_2 \mid \mathbf{m}_{v(2)}, \mathbf{k}_{v(2)}, \wp) \ldots \\ &\quad \times p(\mathbf{d}_{n_b} \mid \mathbf{m}_{v(n_b)}, \mathbf{k}_{v(n_b)}, \wp) \\ &= \prod_{j=1}^{n_b} p(\mathbf{d}_j \mid \mathbf{m}_{v(j)}, k_{v(j)}, \wp), \end{aligned} \quad (9)$$

where d_j is the data for borehole j, $\mathbf{m}$ is the current model and $\mathbf{m}_{v(j)}$ is the relevant set of model parameters for borehole j, where $v(j)$ gives the Voronoi partition in which borehole j is located.

The estimated noise on each data point in a borehole profile is quantified by the correlation matrix of eq. (8). In all the work presented, this matrix is assumed to be diagonal, so that the data errors are assumed to be uncorrelated. The diagonal values of $\mathbf{C}_{\rm d}$ are given given in the descriptions of the three case studies in Sections 6 and 7. In each case the diagonal value is chosen so that $\sigma_{\rm d}$, the standard deviation of the data error is equal to either 0.05 or 0.1 K, corresponding to medium and high levels of data noise, respectively.

6 SYNTHETIC CASES

6.1 Three partition synthetic example

In this first example, an arbitrary partition structure is set up on a square of dimensions 1 × 1, by selecting three points as the centres. 10 points in the square were selected as the locations of the synthetic borehole profiles (Fig. 2). The GST histories in each partition have been designed to give markedly different trends over the 600 yr time period also shown in Fig. 2. The first (Partition1) shows cooling to year 300, followed by 1 °C of warming; Partition2 shows nearly linear cooling of 0.5 °C and the third, Partition3, shows the opposite trend to Partition1. The temperature profiles were calculated at 2.5 m depth intervals, to a maximum depth of 500 m, with 600 × 1 yr time steps, using the finite element model of Section 2. Each profile was degraded with normally distributed random noise, with a standard deviation of 0.1 K. The model setup (Fig. 2) shows how

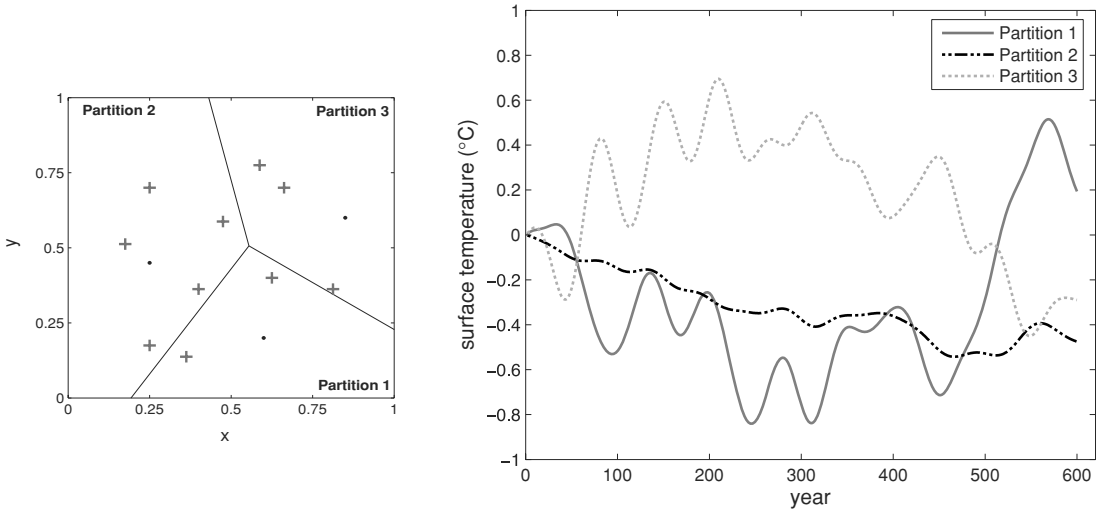

Figure 2. Left-hand panel: true model setup for the synthetic test case. Borehole locations are given by crosses, the partition boundaries show how the profiles are grouped according to the applied surface GST history. Right-hand panel: temperature history applied in each partition in the true forward model for the synthetic data case 1.

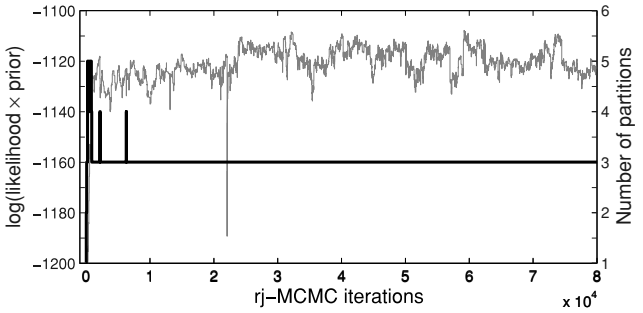

Figure 3. The posterior values and number of Voronoi centres with RJ-MCMC iterations for the synthetic test case.

the three different GST histories have been applied to the 10 synthetic profiles. For simplicity, all of the profiles were initialized with a heat flux of 60 mW m^{-2} and an equilibrium surface temperature of 9.0 °C.

The BPM algorithm was run for 80 000 iterations, with 3500 iterations in each run of the secondary RJ-MCMC algorithm, the first 500 of which were discarded each time as burn-in. The model was initialized with one partition, and the secondary algorithm was used to generate an initial GST history for this partition, with the $\mathbf{q}_0$ and $\mathbf{T}_{eq}$ values initialized to the secondary RJ-MCMC sample means. The algorithm then progresses as described in the previous section.

The resulting samples of all of the model parameters are used to generate posterior probability distributions. The algorithm was found to converge on the true partition structure (i.e. profiles were correctly grouped together) after an initial 872 iterations, and for the analysis of the posterior, the first 5000 iterations were discarded as burn-in. This relatively short period of burn-in is most likely due to the small number of partitions in the true model.

In Fig. 3, the evolution of the number of partitions with RJ-MCMC iterations together with the likelihood × prior value (proportional to the posterior probability value) is shown. Although the algorithm has sampled four partitions, the preferred model has three partitions (with a high probability). This is because the introduction of extra partitions results in lower values for the posterior. This is the natural parsimony feature of Bayesian inference as described in Section 4, such that the improved data fit (greater likelihood) is counteracted by the reduction in the value of the prior term as the number of parameters increases. In Fig. 4, the 2-D posterior distribution of the inferred partition boundaries is shown. It is clear that this distribution is concentrated around the boundaries of the true partition model (Fig. 2). Fig. 3 also shows the conditional posterior distributions for the three temperature histories derived from the mode partition structure, which, in this case, is equal to the true model setup. The true GST histories are shown for comparison by the grey lines. For each partition, the posterior mean GST (black lines), as inferred by the secondary algorithm, matches well with the true GST history. We have thus shown that the BPM method is able to distinguish between applied GST histories in the presence of considerable data noise. The second case now described has been designed to test how the BPM method performs with synthetic data, forced with realistic air temperature histories as derived from a global climate model simulation.

6.2 Synthetic data derived from a global climate model

In this second example, the reconstruction method is tested by using a set of temperature series extracted from a coupled ocean–atmosphere general circulation model (GCM) simulation (e.g. González-Rouco *et al.* 2003, 2006; Mann & Schmidt 2003). The aim of using the BPM methodology is to assess how much of realistic spatial temperature variation can be inferred from such borehole data and whether realistic variations in air temperatures can be distinguished by the BPM algorithm. To approach this, the simulated air temperature series at a set of gridpoints of a GCM simulation are used to calculate a set of realistic synthetic borehole temperature–depth profiles. These profiles are then subject to random noise before using the BPM algorithm.

The GCM model simulations used here have been described in Tett *et al.* (2007) and constitute three climate simulation runs of the UK Met Office HadCM3 model. This model has been described in detail elsewhere (see Tett *et al.* 2007, and references therein). Briefly, however, HadCM3 is a GCM, which is able to simulate a stable climate with no flux adjustments and includes 3-D atmospheric and oceanic components, a comprehensive radiation scheme, a land surface scheme, a sea ice model and a cloud scheme. The experiments described by Tett *et al.* (2007) were designed to simulate the global climate evolution since 1550, as forced by changes in solar forcing,

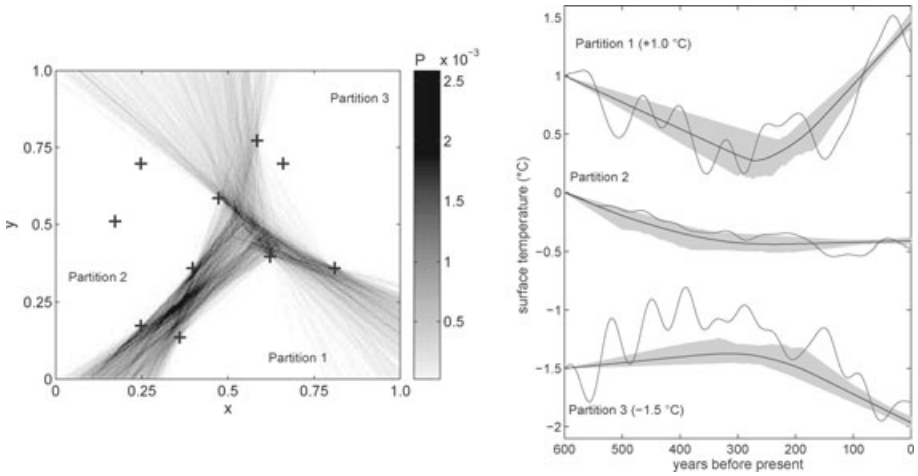

Figure 4. Left-hand panel: the posterior probability distribution of the partition boundaries conditioned on the borehole profiles, their locations and the prior for the first synthetic test case. Right-hand panel: Reconstructed GST histories for the mode partition setup. Posterior mean and true GST histories are shown by the black and grey lines, respectively, and the posterior 95 per cent credible intervals are indicated by the grey shaded areas.

volcanic aerosol forcing and greenhouse gas and ozone variations. Additionally, the surface vegetation was prescribed according to vegetation reconstructions (from 1750 to 1999). Because of the complexity of the HadCM3 model and the experimental setup, the simulation should produce a physically plausible global climate evolution, which provides an ideal test for the BPM methodology.

The issue of air–ground coupling is beyond the scope of this work, and so, we use air temperatures (rather than soil temperatures as in González-Rouco *et al.* 2006) as the forcing for calculating the synthetic borehole profiles. The case study is set up at 10 locations, drawn from across the Northern Hemisphere in the HadCM3 model. These 10 locations are shown in Fig. 6(a) by stars. The mean annual surface-air temperature series (at 2 m) at each site from HadCM3 ALL simulation (which is 250 yr long) are each used to drive a 1-D FE model (Section 2). For simplicity, the thermal conductivity, basal heat flux and equilibrium surface temperatures are the same as used in the previous example and take values of 2.0 W m^{-1} K^{-1}, 60 mW m^{-2} and 9.0 °C, respectively. The air temperature series have then been calculated as deviations from their respective initial 30 yr mean, and the equilibrium surface temperature value for each site has then been added to each air temperature series. Although using the same reference temperature at each site is unrealistic, its value has no effect on determination of the spatial and temporal surface temperature changes. The resultant synthetic borehole profiles have been degraded with Gaussian random noise of standard deviation $\sigma_d = 0.05$ K to mimic real data.

The evolution of the RJ-MCMC algorithm likelihood × prior and the number of partitions for the 130 000 iterations are shown in Fig. 5, and the posterior distribution of the partition boundaries is shown in Fig. 6(a). The inferred temperature histories are compared with the true models in Figs 6(b)–(f). These show the inferred temperature histories in terms of the posterior means and 95 per cent credible intervals in each partition compared with the original input series. The largest grouping has five boreholes and relates to five input series with less long-term variability. The remaining four partitions show stronger departures over the same time period, which are all faithfully recovered by the algorithm. The temperature trends in partitions II and IV are similar, and so, it might be expected that they fall in the same partition. However, the physical spacing of the boreholes means that it is difficult for a Voronoi tessellation to encompass all of the data without including other boreholes that show differing trends (this could be remedied in this case by using a spherical space rather than the flat Cartesian space). The issue of the flexibility of the Voronoi tessellation is discussed further in Section 8.

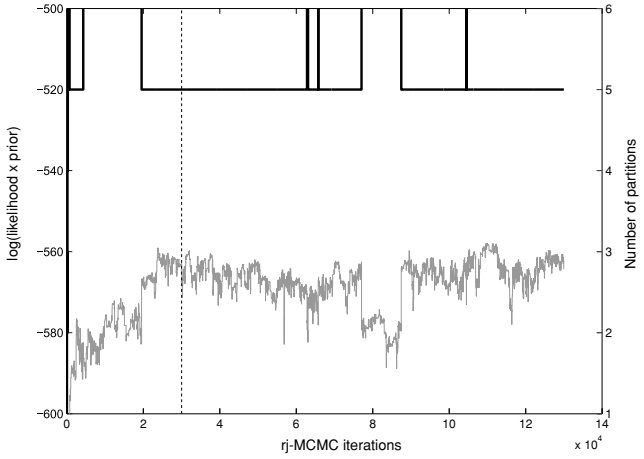

Figure 5. The evolution of the posterior and the number of partitions for the GCM derived synthetic case. The burn-in period is assumed to end at 30 000 iterations and is shown by the vertical dashed line. Note that the posterior is generally lower when the algorithm samples six partitions rather than five, again demonstrating the parsimonious nature of the Bayesian partition model used here.

The applied temperature histories at the 10 synthetic borehole profiles are all different; however, the BPM algorithm posterior distribution indicates that the 10 synthetic temperature-depth profiles can be robustly fit using five GST histories. The inferred temperature histories for these five partitions are reliable reconstructions of the 10 GCM air temperature series. The BPM method has therefore allowed a simple overall spatial parametrization to be automatically constructed, which reflects the underlying spatial climate trends, as can be best inferred from borehole data. This example also demonstrates that borehole data can be useful in determining realistic spatial variations in the surface temperature response over the past few centuries, in agreement with previous studies (Majorowicz *et al.* 2002; Beltrami *et al.* 2003; Stevens *et al.* 2008).

Since only 10 locations from across the Northern Hemisphere have been used, the variations in the surface temperature histories, as reconstructed by the BPM algorithm, cannot be interpreted as representative of either natural variability or the GCM simulation variability. An expanded test, perhaps using the GCM gridpoints corresponding to the locations in the Huang & Pollack (1998) database (as in González-Rouco *et al.* 2006), could be used required for this purpose. Furthermore, the real benefit of the BPM algorithm would be more apparent in a larger data set where actual spatial variations in climate can be more clearly defined because of the denser spatial coverage of the borehole data.

7 UK BOREHOLE DATA SET

7.1 Data

In this real data case, 22 boreholes have been taken from the Huang & Pollack (1998) database, and one other has been provided by the British Geological Survey (Rollin 1987). The locations and relative depths are shown in Fig. 7. Previous GST reconstructions from the 22 boreholes have been derived by Huang *et al.* (2000) and are used in hemispheric and global estimates of surface temperature changes. The 22 individual reconstructions are shown in Fig. 8(a). The authors emphasize that these individual reconstructions are not necessarily optimal, as the methods employed have been designed for consistency in analysing 837 borehole profiles rather than for deriving optimal individual reconstructions. However, there is clearly a large spread in the GST reconstructions, which is inconsistent with what is known regarding palaeoclimate for the region.

In Fig. 8(b), the average reconstruction for this ensemble is compared with the CET record (Manley 1974; Parker *et al.* 1992). The Huang *et al.* (2000) average appears to underestimate the long term variations in the CET record. Jones (1999) showed that the average of 26 borehole GST reconstructions from the British Isles (from Pollack *et al.* 1998 using an earlier compilation of Huang & Pollack (1998)) shows excellent agreement with the CET record. However, these 26 sites also include four located in Ireland, all of which show much larger warming over the 500 yr (up to 3.5 °C in one case). Including these four reconstructions reduces the pre-1750 average reconstructed temperatures (see Fig. 8b). However, 3.5 °C of warming over 500 yr may be unrealistically large, and the

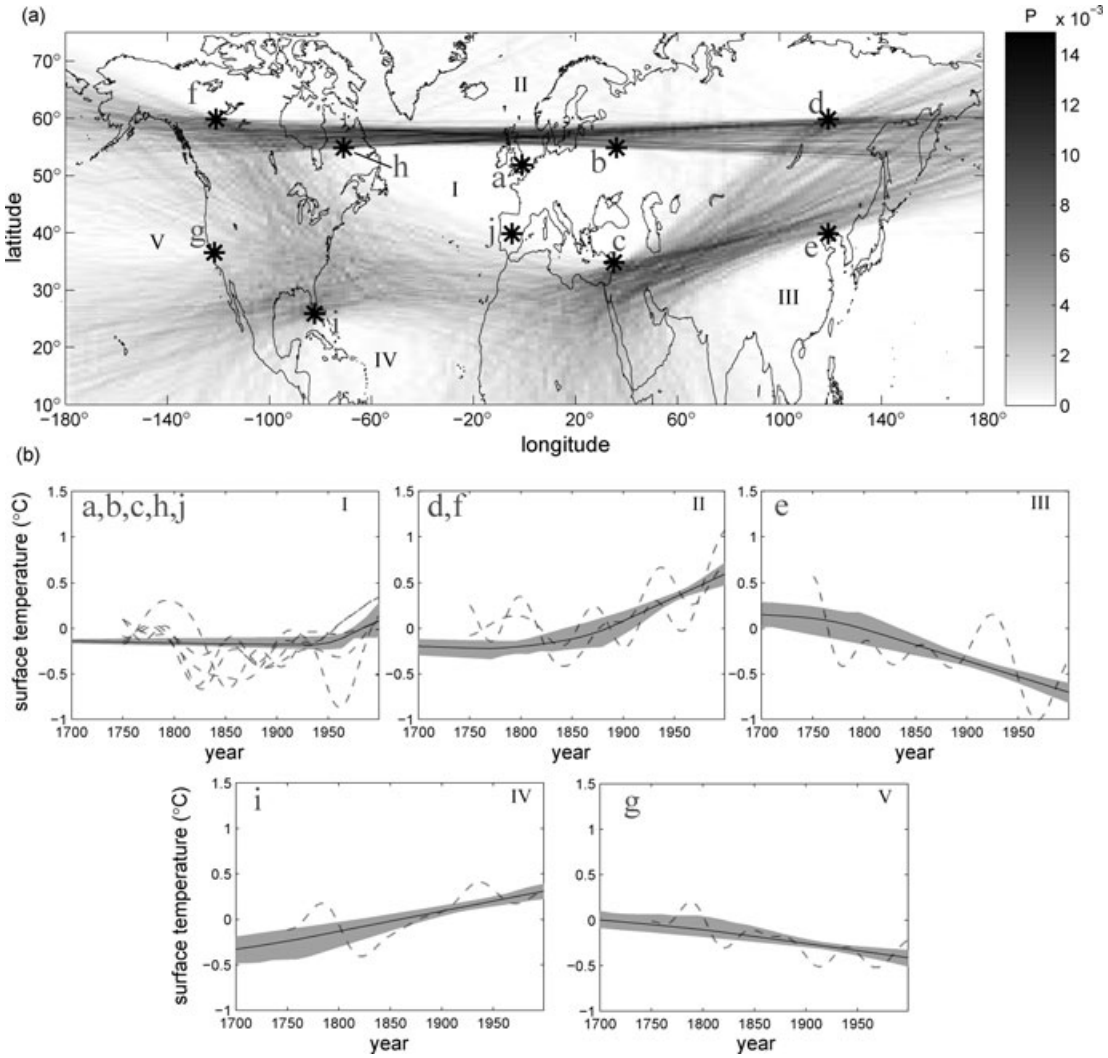

Figure 6. (a) The posterior distribution of the partition boundaries as sampled by the BPM algorithm for the GCM derived synthetic case after 130 000 iterations. (b) The posterior mean and 95 per cent credible limits (solid line and shaded region respectively) for the five partitions in the mode setup of (a). The dashed lines are 30 yr smoothed temperature histories from the GCM data used for the true model.

variability between the two averages shown provides evidence that this averaging process may be inappropriate.

Table 1 summarizes the 23 data sets (Rollin 1987) used in this final example, giving the maximum depth of temperature measurements, the range of depths of thermal conductivity measurements, the borehole longitude and latitude and the date of logging. The heat flux and equilibrium surface temperatures given in the table are the starting values used in the algorithm. As shown in Table 1, there is a reasonable amount of thermal conductivity data from each borehole. However, for most of the boreholes, the conductivity measurements have been made at different depths than the temperature measurements. In this case, the thermal conductivity data have been interpolated to 1 m depth intervals. In the forward model, the 1 m interpolated values are then interpolated onto the temperature depth grid. For depth ranges where no conductivity measurements are available, the nearest 20-m-average value (from the same borehole) has been used. This then reflects a best estimate of a suitable value. More detailed geological investigation of each site may also yield suitable conductivity values, but this is beyond the scope of the present work. Since the range of dates at which the 23 boreholes have been logged is small (1979–1987), we do not take account of this in our model; thus, all boreholes are assumed to be logged at 1987. Due to the low resolution of the GST reconstructions, this assumption should have a small overall effect on the reconstructed temperatures.

The noise values for the UK data are again assumed to be Gaussian uncorrelated and with a standard deviation of 0.1 K. This corresponds to fairly high level of data noise. The spatial domain chosen for the BPM algorithm is a square with southwest and northeast vertices at (longitude, latitude) (-6.6°, 1.78°) and (50.4°, 58.16°), respectively. The starting values given in Table 1 were found by running the BPM algorithm, starting with one partition and updating only the heat flow and equilibrium temperature values at each site and taking the average values of the resultant samples.

7.2 Sampling and posterior results

The BPM algorithm was run for 1 million iterations using the 23 UK data sets. The starting values for the GST histories and equilibrium surface temperature and basal heat fluxes are found by an initial run of the secondary RJ-MCMC algorithm for each borehole data set. To save on computation, the algorithm was initialized arbitrarily with

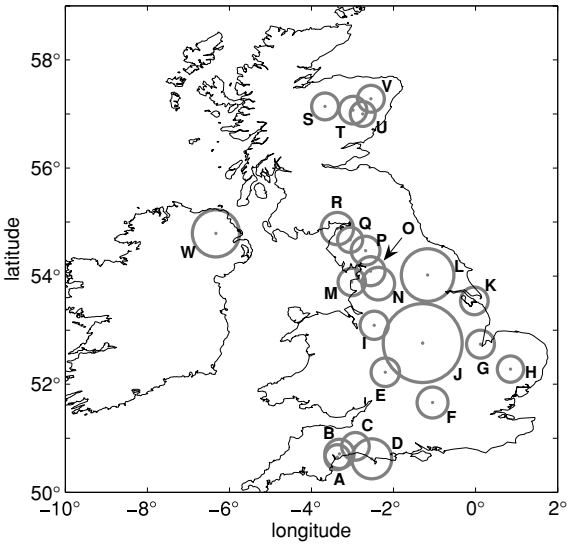

Figure 7. Locations of the 23 boreholes used in this study as labelled in table 1. The maximum temperature measurement depth at each borehole is indicated proportionally by the diameter of the circle marker. The deepest is J (Morley) at 823 m and the shallowest is U (Mount) at 261 m.

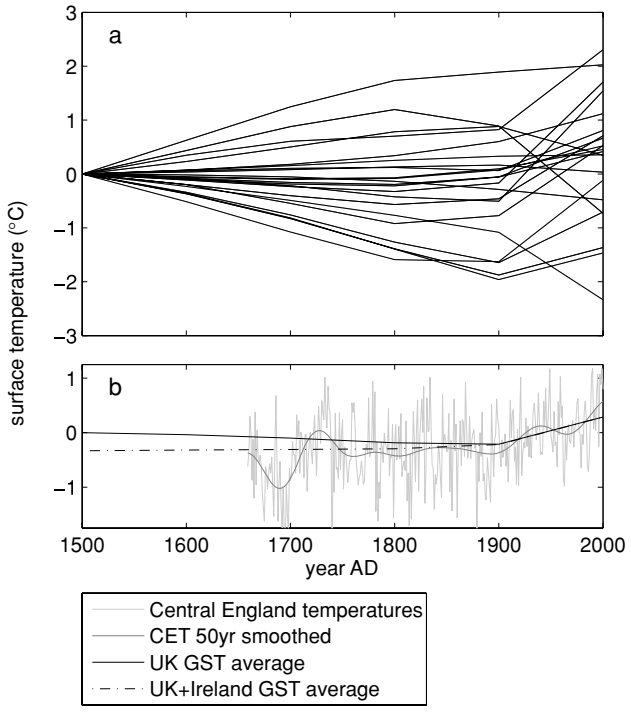

Figure 8. (a) Ground surface temperature reconstructions from Huang *et al.* (2000) derived from 22 UK boreholes from (Huang & Pollack 1998), locations shown in Fig. 7. (b) The average GST reconstruction of the 22 UK GST histories compared with the Central England temperature series and the average GST reconstruction for 26 UK and Ireland boreholes as in Jones (1999). The series have been shifted to match the CET record average for the years 1900–2000.

four partitions. The algorithm quickly loses memory of this initial state and moves to a different configuration, eventually converging on 10 or 11 partitions. Fig. 9 shows the posterior probability and the number of partitions with iterations. The vertical dashed line indicates where burn-in is assumed to end and is placed at 600 000 iterations. We note that all of the larger partitions in the model are present by 350 000 iterations (see below). The evolution of the number of partitions shows that the model eventually settles on 10 or 11 partitions with a slight preference for the latter. The algorithm therefore infers that given the 23 borehole data sets, 10 or 11 independent GST histories are required to achieve an adequate fit to the data, conditional on the assumption of conductive 1-D heat transfer.

The posterior distribution of the partition boundaries as inferred by the algorithm is shown in Fig. 10. The 11 partitions are labelled I–XI, with the largest partition II centred on the English Midlands. Of the 11 partitions, three are inferred to contain more than two boreholes, with partitions II, VIII and IX containing six, three and three boreholes, respectively. The inferred GST histories for the 11 partitions are shown in Fig. 11. The GST reconstructions are labelled according to the partitions in Fig. 10 and show the posterior mean (solid line) and the 95 per cent credible regions (shaded grey). The majority of the partitions indicate warming. However, the magnitude is variable, covering a range up to 2.25 °C (±0.25 including the credible limits). Similarly, the timing of a temperature minimum (where there is one) is inferred to occur between 1500 and 1950. Overall, this ensemble of reconstructions shows a similar level of variability as the equivalent reconstructions shown in Fig. 8(a). This large range of the inferred GST histories for the past few hundred years indicates that these data provide a poor constraint on past climate.

If we examine only those partitions inferred to contain three boreholes or more (e.g. similar to Pollack & Smerdon 2004), then the GST history becomes slightly simplified. These partitions have been inferred to contain more data, and so, more confidence can be placed in the signal of past climate as opposed to that of other non-climatic heat transfer processes. This is because to within the noise values, the boreholes can be fit using the same GST history. These three reconstructions all show the same trend of a cool period constrained to between 1650 and 1800, followed by 0.5–1.3 °C warming to 1987. The spread between the three reconstructions, however, is still large and is illustrated alongside the CET record (Manley 1974; Parker *et al.* 1992) and the average UK reconstruction of Huang *et al.* (2000) in Fig. 12. It is evident that two of the three BPM posterior means (VII and IX) show very poor agreement with the CET record, with VIII indicating around three times the CET warming magnitude at around 1.5 °C and IX around double at 1.0 °C. Furthermore, partition VIII shows around 1.5 times the warming rate indicated by partition X. It is encouraging that the single largest partition (II) shows very good agreement with the CET record. However, no variation is inferred prior to 1600–1700, probably because of the limited depths of the boreholes used in this study.

Previously, we have shown that the mean GST derived from the individual reconstructions of Huang *et al.* (2000) is possibly not robust for this data set, and the BPM results all indicate a differing trend before 1900. That the algorithm infers differing amplitudes for the inferred GST histories in the larger partitions indicates that systematic errors in the data relating to non-climatic influence must be present. For the boreholes in partitions VIII and IX, the GST history has a much larger amplitude. This could be caused by some combination of fluid flow or surface topography. Taken alongside the divergent trends shown in Fig. 11, the BPM algorithm results suggest that only a fraction of the data examined can provide a robust signal of past climate, since the majority of the boreholes used here lead to conflicting GST histories. Comparison with the Huang *et al.* (2000) average also indicates that averaging of individual noisy

Table 1. Borehole summary information for the UK data set. For each borehole, this table gives the depth of each borehole used in the study, the depth range of available thermal conductivity measurements, the longitude and latitude of the borehole, the year of logging and the starting values found by the BPM algorithm for the geothermal heat flux and the equilibrium surface temperature.

	Borehole	Depth (m)	Cond. z-range	Long. (°)	Lat. (°)	Log date	$\mathbf{q}_0$ (mW m^{-2})	$\mathbf{T}_{eq}$(°C)
A	Withycombe	262	16–262	−3.37	50.65	1983	52.9	9.69
B	Venn	307	18–305	−3.32	50.71	1984	57.1	9.85
C	Chard	289	37–289	−2.93	50.85	1984	50.4	9.14
D	Seabarn	420	18–415	−2.53	50.62	1979	61.2	9.41
E	Worcester	298	33–298	−2.2	52.22	1984	43.4	9.60
F	Chalgrove	323	91–322	−1.05	51.66	1984	51.7	10.48
G	Tydd	294	51–295	0.12	52.74	1984	56.6	8.80
H	Stowlangtoft	277	13–276	0.85	52.28	1984	35.9	9.25
I	Crewe	296	53–296	−2.47	53.09	1984	58.6	9.15
J	Morley	823	10–830	−1.29	52.76	1987	56.6	8.94
K	Cleethorpes	290	80–290	−0.03	53.54	1985	72.2	9.28
L	Shipton	549	9–547	−1.17	54.02	1987	59.9	8.71
M	Thornton	287	42–288	−3.02	53.89	1984	52.2	9.07
N	Clitheroe	341	81–301	−2.37	53.86	1985	45.1	9.06
O	Wray*	303	100–301	−2.56	54.09	1986	40.2	7.61
P	Shap	301	10–300	−2.68	54.47	1983	73.5	6.19
Q	Skiddaw	265	10–281	−3.06	54.67	1983	125.4	7.55
R	Silloth	340	48–340	−3.37	54.88	1982	53.1	8.48
S	Cairngorn	282	10–296	−3.67	57.14	1983	75.9	4.86
T	Ballater	290	10–298	−2.99	57.07	1983	75.6	7.25
U	Mount	261	10−260	−2.75	57.0	1983	69.9	6.74
V	Bennachie	282	10–294	−2.55	57.28	1983	81.5	6.46
W	Ballymacilroy	493	61–500	−6.33	54.79	1980	59.5	9.02

*Not previously selected by Huang & Pollack (1998).

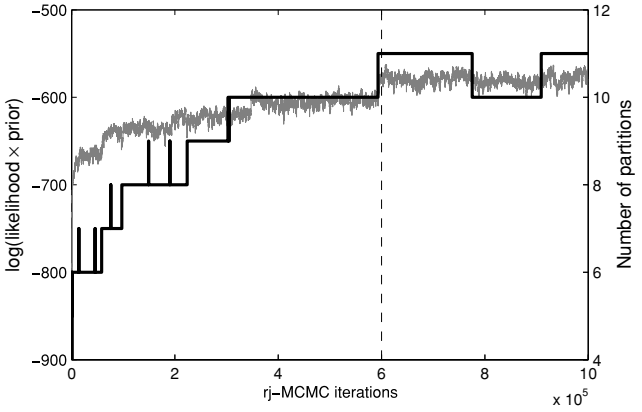

Figure 9. The evolution with RJ-MCMC iterations of the posterior and the number of partitions. The algorithm eventually converges on a model with 10 or 11 partitions. The three largest partitions (II, VIII, IX) are present in all of samples from iteration number 350 000 to the end.

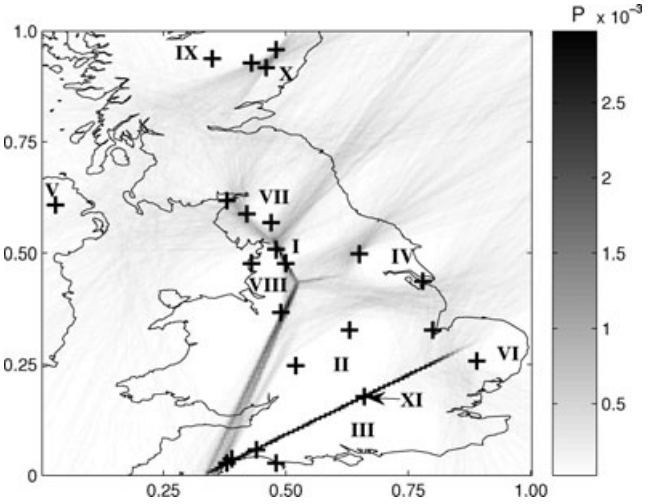

Figure 10. The posterior probability distribution of the partition boundaries for the UK real data case, shown on the model space rather than in longitude/latitude coordinates. Borehole locations are given by crosses. The algorithm infers 10–11 partitions in the data, which are labelled I to XI.

reconstructions does not necessarily produce the same result as joint inversion of a smaller (plausibly more robust) subset drawn from the same profiles. This is important since the hemispheric and global borehole temperature estimates are derived in this way, although the borehole data used may be more suitable for climatic reconstructions than the small set of 23 examined here.

8 DISCUSSION AND CONCLUSIONS

In this work, we have demonstrated how spatial and temporal trends in GSTs can be jointly inferred from collections of borehole temperature profiles. We have demonstrated the efficacy of the BPM algorithm with realistic synthetic data and have shown how the algorithm is able to correctly group borehole profiles according to the applied GST history. This method therefore allows the data to determine the optimum grouping for joint inversion, which takes account of the relative locations of the profiles.

In the first synthetic case, the Bayesian Partition algorithm demonstrates the natural parsimony, which allows the RJ-MCMC algorithm to preferentially sample simple models, which, in this case, correspond well with the true model used. In other synthetic experiments (not shown), the algorithm has converged on structures for which partitions present in the true model have been merged with

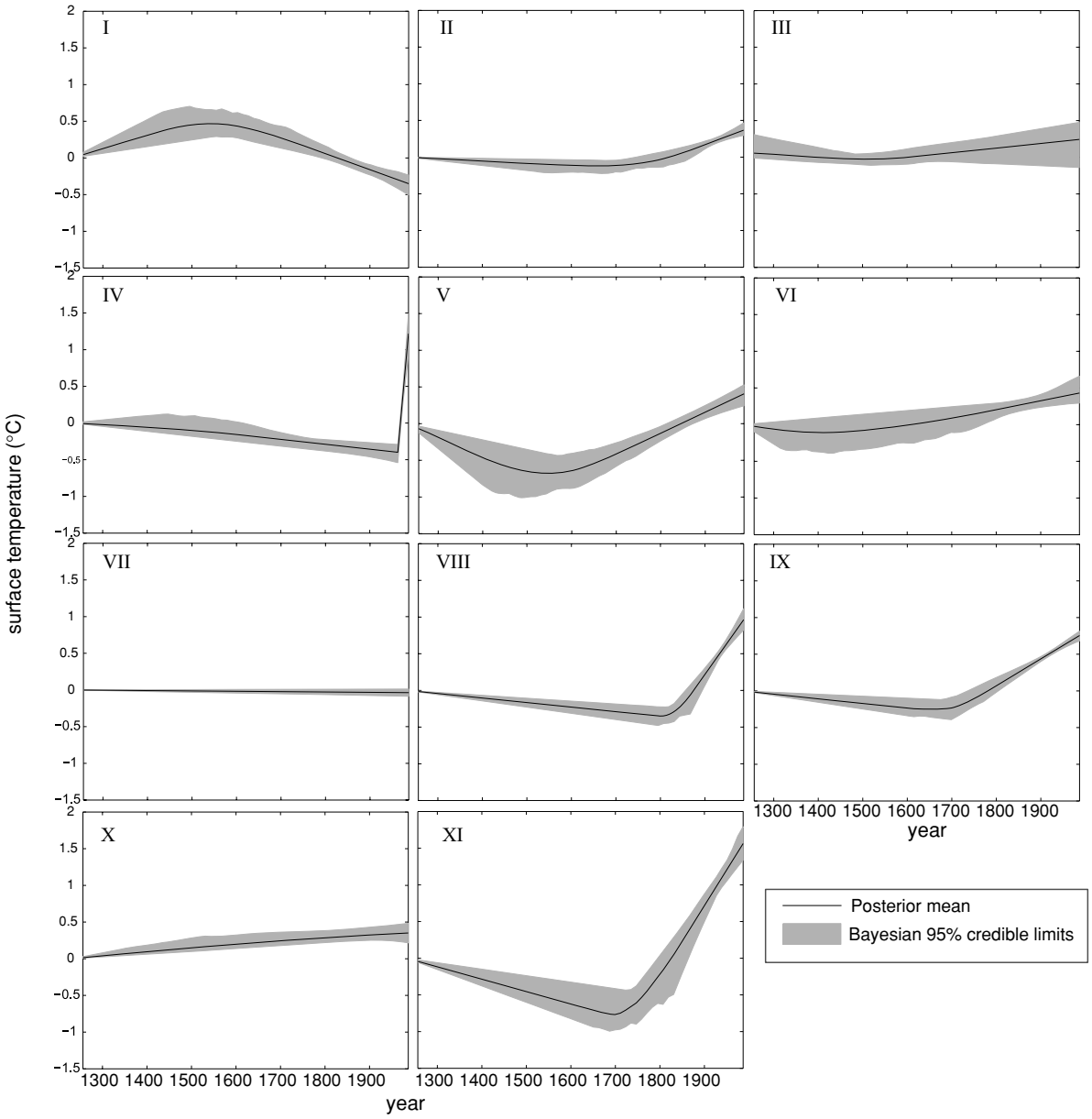

Figure 11. The posterior distributions of the GST histories as inferred by the secondary RJ-MCMC algorithm for the mode setup shown in Fig. 10, which has 11 partitions. Note that many of the partitions show warming over the last few centuries but the magnitude of this warming is variable.

neighbouring partitions, where the data can be reasonably fit by the same GST history. The tendency toward simpler models is a crucial feature for this work, as it allows more useful models of the underlying trends in past GST rather than simply optimizing the fit separately at each borehole, which can lead to a more confusing and possibly misleading set of GST reconstructions.

A second case study using surface air temperatures from 10 grid-points of an GCM simulation of the last 250 yr provides a more realistic test for the BPM method. This example demonstrated that realistic variations in air temperature variations from across the globe can be distinguished from each other by the BPM algorithm when the data are artificially degraded with moderate measurement errors. This implies that analysis of an expanded set of real borehole data from across the globe could reveal variations in warming magnitude from region to region. The algorithm leads to a posterior partition structure that accurately reflects the applied temperature series, and the inferred GST histories for each partition agree well with these true models in each case.

A first case study with real data has been made using a data set of 23 boreholes located in the United Kingdom. We have demonstrated that an average of 22 original reconstructions is inappropriate and that including four extra reconstructions from Ireland leads to a different average GST history. This latter average reconstruction compares favourably with the CET record but includes temperature histories with up to 3.5 °C warming since 1500. Using the BPM, we have shown that a previously large range of GST histories can be somewhat reduced by examining the support from the data for independent temperature histories. This is also aided by the use of measured thermal conductivity data (Rollin 1987). However, the preliminary results for this data set indicate a posterior on the partition structure with high probability of 10 or 11 independent GST trends, whereas examination of instrumental data covering

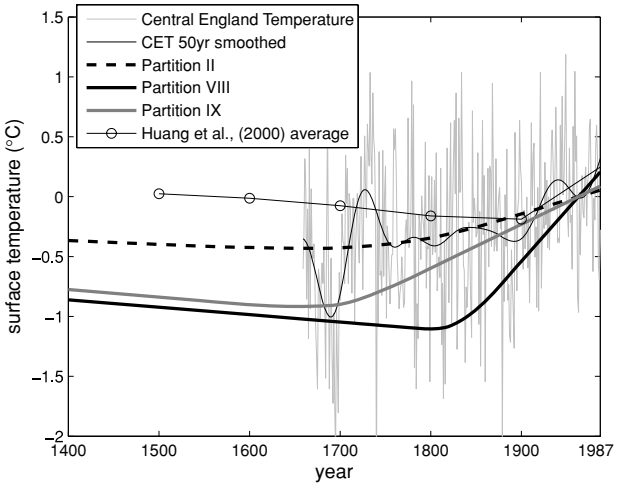

Figure 12. Comparison of the posterior mean GST histories for the three largest partitions (II, VIII and IX) compared with the Central England temperature record and the average GST reconstruction for the UK derived from Huang *et al.* (2000). The CET and BPM GST histories are shown as deviations from the 1950–1979 mean whereas the Huang *et al.* (2000) average has been shifted to have the same 20th Century mean as the CET record as plotted here. The GST reconstructions for the three partitions show poor agreement with each other. The largest partition (II) shows good agreement with the CET record. The Huang *et al.* (2000) average shows a contrasting temperature trend before 1900 compared with all of the other temperature series.

the period 1850–present from the Meteorological Office Historical Instrumental Station data set (UKMO 2006) indicates that the long term changes in surface temperature are very similar across the study area. The posterior mean GST histories for the three largest partitions, which would be expected to be more robust indicators of past climate, show a large range in the amplitude and timing of the inferred GST histories, with the amplitude of the posterior means ranging over 0.5–1.3 °C warming by AD 1987 and with a temperature minimum somewhere between 1650 and 1800. The largest partition agrees very well with the CET instrumental record, but the remaining partitions demonstrate that other thermal effects, which could not be accounted for in the 1-D models used here, are present. The mismatch between the ensemble average GST history derived from Huang *et al.* (2000) and those derived in the largest partitions in this work indicates that for this data set, averaging does not always cancel non-climatic errors. Overall, we have found that the majority of the 23 temperature–depth profiles examined (including those in two of the three larger partitions) are sufficiently affected by these other heat transfer processes so that extracting a meaningful climatic signal is difficult. The majority of the UK data set therefore requires geophysical site analysis and the use of a more advanced forward modelling scheme before it can be used for climatic reconstruction purposes.

8.1 Future work

The Bayesian Partitioning algorithm is fairly computationally burdensome, as it relies on two different RJ-MCMC samplers. For example, in the GCM case, 788 groupings of the data were evaluated during the 130 000 RJ-MCMC iterations. Multiplying this by the number of secondary sampler iterations (=5000) amounts to 4 million secondary sampler forward model evaluations. The total number of individual forward evaluations will be larger since many of these configurations that have been evaluated, contain more than one profile. In the real data case, the number of data groupings evaluated by the BPM model is of the order of 2000, meaning that the total number of forward evaluations is of the order of 10 million, and again the total number of forward evaluations will be larger, as there are 23 data sets in the UK case compared with 10 in the two synthetic cases. The computational cost of reading data from a previously sampled configuration is not negligible compared with the cost of running the secondary sampler, but the computation speed of the model does gradually increase, as more of the model space is sampled and stored for future iterations. The real data case example took 2 weeks on a Linux 3.3 GHz Pentium 4 for 1 000 000 iterations. To apply this method to a larger data set, this algorithm would therefore benefit from parallelization: where proposal choices lead to more than one partition being updated in a single iteration, the calculations for each partition could be sent to separate processors and calculated simultaneously.

The Voronoi tessellation method provides a convenient mathematical method of dividing the data into separate geographical regions. However, there are drawbacks due to its simplicity. For example, for geometrical reasons, more partitions or tessellations can be required than there are independent trends in the data. Andersen *et al.* (2003) provide a method for avoiding these problems by parametrizing the vertices of the tessellations and using the notion of coloured polygons, that is, similarly coloured polygons, which can be geographically separate, would contain data with the same inferred GST history. Future work involving the methods presented here could benefit from such a formulation.

The methods presented here could be developed to incorporate proxy palaeoclimate data together with the borehole temperatures. For example, by implementing a Bayesian calibration method for the proxy data, the borehole data could serve as a prior constraint (through the heat conduction forward model) on low-frequency variations of the climate reconstructions from proxy data.

ACKNOWLEDGMENTS

Discussions on MCMC methods with Chris Holmes, John Stephenson and Karl Charvin over the course of this work have been invaluable. We are grateful to Simon Tett and Philip Brohan for kindly supplying the HadCM3 data from the EU SO&P project. We also thank two anonymous reviewers and the editor for many detailed and insightful comments on an our work. POH was supported through a studentship grant from the Environmental Mathematics and Statistics programme of NERC and EPSRC to KG. The posterior GST samples and the computer code for inverting individual boreholes are available upon request from POH.

REFERENCES

Andersen, K.E., Brooks, S.P. & Hansen, M.B., 2003. Bayesian inversion of geoelectrical resistivity data, *J. R. Stat. Soc. Ser. B (Statistical methodology),* **65,** 619–642.

Backus, G.E. & Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross Earth data. *Phil. Trans. R. Soc. Lond., A,* **266,** 123–192.

Beltrami, H. & Bourlon, E., 2004. Ground warming patterns in the Northern Hemisphere during the last five centuries, *Earth planet. Sci. Lett.,* **227**(3-4), 169–177.

Beltrami, H., Gosselin, C. & Mareschal, J.C., 2003. Ground surface temperatures in Canada: spatial and temporal variability, *Geophys. Res. Lett.,* **30**(10), 1499, doi:10.1029/2003GL017144.

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

Briffa, K.R. & Osborn, T.J., 2002. Paleoclimate—blowing hot and cold, *Science,* **295**(5563), 2227–2228.

Chouinard, C. & Mareschal, J.-C., 2007. Selection of borehole temperature depth profiles for regional climate reconstructions, *Clim. Past,* **3,** 297–313.

Curtis, A. & Wood, R., (eds) 2004. *Geological Prior Information, Geological Society London Special Publication,* Vol. 239, Geological Society London.

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

Esper, J., Cook, E.R. & Schweingruber, F.H., 2002. Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability, *Science,* **295**(5563), 2250–2253.

Gallagher, K., 1990. Some strategies for estimating present day heat flow from exploration wells, with examples, *Explor. Geophys.,* **21**(4), 145–159.

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

González-Rouco, J.F., Beltrami, H., Zorita, E. & von Storch, H., 2006. Simulation and inversion of borehole temperature profiles in surrogate climates: spatial distribution and surface coupling, *Geophys. Res. Lett.,* **33,** L01703.

González-Rouco, J.F., von Storch, H. & Zorita, E., 2003. Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years, *Geophys. Res. Lett.,* **30**(21), 2116, doi:10.1029/2003GL018026.

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

Green, P.J., 2003. Trans-dimensional Markov chain Monte Carlo, in *Highly Structured Stochastic Systems,* pp. 179–198, eds Green, P.J., Hjort, N.L. & Richardson, S., Oxford University Press, Oxford.

Harris, R.N. & Chapman, D.S., 2005. Borehole temperatures and tree rings: seasonality and estimates of extratropical Northern Hemispheric warming, *J. geophys. Res.,* **110**(F04003), doi:10.1029/2005JF000303.

Hegerl, G.C., Crowley, T.J., Allen, M., Hyde, W.T., Pollack, H.N., Smerdon, J. & Zorita, E., 2007. Detection of human influence on a new, validated 1500-year temperature reconstruction, *J. Clim.,* **20**(4), 650–666.

Hopcroft, P.O., Gallagher, K. & Pain, C.C., 2007. Inference of past climate from borehole temperature data using Bayesian Reversible Jump Markov chain Monte Carlo, *Geophys. J. Int.,* **171**(3), 1430–1439.

Hopcroft, P.O., Gallagher, K., Pain, C.C. & Fang, F., 2009. Three-dimensional simulation and inversion of borehole temperatures for reconstructing past climate in complex settings, *J. geophys. Res.,* doi:10.1029/2008JF001165.

Huang, S.P. & Pollack, H.N., 1998. Global Borehole Temperature Database for Climate Reconstruction, IGBP PAGES/World Data Center-A for Paleoclimatology Data Contribution Series No. 1998−044, NOAA/NGDC Paleoclimatology Program, Boulder CO, USA.

Huang, S.P., Pollack, H.N. & Shen, P.Y., 2000. Temperature trends over the past five centuries reconstructed from borehole temperatures, *Nature,* **403**(6771), 756–758.

Huang, S.P., Shen, P.Y. & Pollack, H.N., 1996. Deriving century-long trends of surface temperature change from borehole temperatures, *Geophys. Res. Lett.,* **23**(3), 257–260.

Jansen, E. *et al.*, 2007. Palaeoclimate, in *Climate Change 2007: The Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change,* eds Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. & Miller, H.L., CUP, Cambridge, UK and New York, USA.

Jefferys, W.H. & Berger, J.O., 1992. Ockham's razor and Bayesian analysis, *Am. Scientist,* **80,** 64–72.

Jones, P.D., 1999. Classics in physical geography revisited, *Progr. Phys. Geogr.,* **23**(3), 425–428.

Lachenbruch, A.H. & Marshall, B.V., 1986. Changing climate—geothermal evidence from permafrost in the Alaskan Arctic, *Science,* **234**(4777), 689–696.

MacKay, D.J.C., 2003. *Information Theory, Inference, and Learning Algorithms,* Cambridge University Press, Cambridge.

Majorowicz, J.A., Šafanda, J. & Skinner, W., 2002. East to west retardation in the onset of the recent warming across Canada inferred from inversions of temperature logs, *J. geophys. Res.,* **107**(B10), 2227, doi:10.1029/2001JB000519.

Majorowicz, J.A., Šafanda, J. & Skinner, W., 2004. Past surface temperature changes as derived from continental temperature logs—Canadian and some global examples of application of a new tool in climate change studies, *Adv. Geophys.,* **47,** 113–174.

Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.,* **151**(3), 675–688.

Manley, G., 1974. Central England temperatures—monthly Means 1659 to 1973, *Q. J. R. Meteorol. Soc.,* **100**(425), 389–405.

Mann, M.E., Rutherford, S., Bradley, R.S., Hughes, M.K. & Keimig, F.T., 2003. Optimal surface temperature reconstructions using terrestrial borehole data, *J. geophys. Res.,* **108**(D7), 4203, doi:10.1029/2002JD002532.

Mann, M.E. & Schmidt, G.A., 2003. Ground vs. surface air temperature trends: implications for borehole surface temperature reconstructions, *Geophys. Res. Lett.,* **30**(12), 1607, doi:10.1029/2003GL017170.

Mareschal, J.C. & Beltrami, H., 1992. Evidence for recent warming from perturbed geothermal gradients: examples from eastern Canada, *Clim. Dyn.,* **6,** 135–143.

Moberg, A., Sonechkin, D.M., Holmgren, K., Datsenko, N.M. & Karlén, W., 2005. Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, *Nature,* **433**(7026), 613–617.

Mosegaard, K. & Tarantola, A., 1995. Monte-Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100**(B7), 12 431–12 447.

Parker, D.E., Legg, T.P. & Folland, C.K., 1992. A new daily central England temperature series, 1772–1991, *Int. J. Climatol.,* **12**(4), 317–342.

Pollack, H.N. & Huang, S.P., 2000. Climate reconstruction from subsurface temperatures, *Ann. Rev. Earth Planet. Sci.,* **28,** 339–365.

Pollack, H.N., Huang, S.P. & Shen, P.Y., 1998. Climate change record in subsurface temperatures: a global perspective, *Science,* **282**(5387), 279–281.

Pollack, H.N. & Smerdon, J.E., 2004. Borehole climate reconstructions: spatial structure and hemispheric averages, *J. geophys. Res.,* **109**(D11), 11106, doi:10.1029/2003JD004163.

Rollin, K., 1987. *Catalogue of Geothermal data for the Land Area of United Kingdom: Investigations of the Geothermal Potential of the UK,* British Geological Survey.

Rutherford, S. & Mann, M.E., 2004. Correction to 'Optimal surface temperature reconstructions using terrestrial borehole data', *J. geophys. Res.,* **109,** D11107.

Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.,* **167**(2), 528–542.

Scales, J.A. & Tenorio, L., 2001. Prior information and uncertainty in inverse problems, *Geophysics,* **66**(2), 389–397.

Sivia, D.S. & Skilling, J., 2006. *Data Analysis: A Bayesian Tutorial,* 2nd rev. edn, Oxford University Press, Oxford.

Stephenson, J., Gallagher, K. & Holmes, C.C., 2004. Beyond kriging: dealing with discontinuous spatial data fields using adaptive prior information and Bayesian partition modelling, in *Geological Prior Information,* Vol. 239, pp. 195–209, eds Curtis, A. & Wood, R., Geological Society London, Special Publications.

Stephenson, J., Gallagher, K. & Holmes, C.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.

Stevens, B., González-Rouco, J. & Beltrami, H., 2008. North American climate of the last millennium: underground temperatures and model comparison, *J. geophys. Res.,* **113**(F1), F01008.

von Storch, H., Zorita, E., Jones, J.M., Dimitriev, Y., González-Rouco, J.F. & Tett, S.F.B., 2004. Reconstructing past climate from noisy data, *Science,* **306,** 679–682.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* SIAM, Philadelphia.

Tett, S.F.B. *et al.*, 2007. The impact of natural and anthropogenic forcings on climate and hydrology since 1550, *Clim. Dyn.,* **28,** 3–34.

UKMO 2006. *MIDAS Land Surface Stations data (1853-current),* UK Meteorological Office, British Atmospheric Data Centre, available at http://badc.nerc.ac.uk/data/ukmo-midas.

APPENDIX A: BPM SAMPLING ALGORITHM

A1 Prior information

In any Bayesian formulation, prior information for all model parameters must be specified. In inverse problems, the prior information acts to ensure that the solutions have acceptable or physically reasonable values (Scales & Tenorio 2001; Curtis & Wood 2004). In this work, the prior information is chosen to correspond with what is usually known in a general case.

Treating the prior probability hierarchically, the terms in $\boldsymbol{\theta}$ (the GST histories in each partition) can be separated from those in $\mathbf{C}$ (the partition centres):

$$p(\mathbf{C}, \boldsymbol{\theta} \mid \wp) = p(\boldsymbol{\theta} \mid \mathbf{C}, \wp) p(\mathbf{C} \mid \wp). \tag{A1}$$

Here a continuous uniform prior is employed for the positions of the Voronoi centres, and a discrete uniform prior is used for the number of the Voronoi centres (Denison *et al.* 2002):

$$p(\mathbf{C}) = \frac{n_c!}{(\Delta x \Delta y)^{n_c}} \times p(n_c), \tag{A2}$$

where $\Delta x = x_{\max} - x_{\min}$ and $\Delta y = y_{\max} - y_{\min}$ are the dimensions of the spatial domain (rectangular) and are both equal to 1.0 in all the examples presented. A uniform probability distribution over the range $(1, n_b)$ is used as the prior on the number of partitions, that is, $p(n_c) = 1/n_{\mathrm{b}}$. A more conservative choice would be a Poisson distribution with an expected value. However, to demonstrate the efficacy of the partition modelling method, the more general prior term is used.

The prior information on the GST model parameters $\boldsymbol{\theta}_i$ is the same as used by Hopcroft *et al.* (2007). The prior distribution on the past temperature values is given by

$$p(\mathbf{T}_i \mid \wp) = \frac{1}{\left[(2\pi)^{k_i} \det \mathbf{C}_\wp\right]^{1/2}} \exp\left[-\frac{1}{2}\left(\mathbf{T}_i^{\mathrm{GST}} - \mathbf{T}^\wp\right)^T \mathbf{C}_\wp^{-1} \times \left(\mathbf{T}_i^{\mathrm{GST}} - \mathbf{T}^\wp\right)\right], \tag{A3}$$

where $\mathbf{T}_i^\wp \in \mathbf{R}^{k_i}$ gives the most likely values at each of the k_i time points. In this work, $\mathbf{T}^\wp$ takes the same value at all time points so that the prior is biased towards temperatures histories with no change (Huang *et al.* 2000). The values used for $\mathbf{T}^\wp$ and $\mathbf{C}_\wp$ are given alongside the specific examples described in Sections 6 and 7. When combining the prior for multiple partitions, the GST histories in each partition are assumed to be independent so that the prior term for the GST for the n_c partitions is given by the product of the individual prior terms:

$$p(\mathbf{T} \mid \wp) = \prod_{i=1}^{n_c} p(\mathbf{T}_i \mid \wp), \tag{A4}$$

where there are n_c partitions or independent GST histories.

The prior distribution for the time positions of the nodes used to parametrize the GST histories is a uniform order statistic distribution (e.g. Green 1995; Hopcroft *et al.* 2007), which introduces a gentle bias toward equal spacing of the nodes over the given reconstruction length. The prior on the time points of the past temperatures in region i is

$$p(\mathbf{t}_i \mid \wp) = \frac{k_i! \times t_i^1 \times \left(t_i^2 - t_i^1\right) \times \cdots \times \left(L - t_i^{k_i}\right)}{L^{k_i}}. \tag{A5}$$

Note that we have tested the effect of this prior term (not shown) using, instead, a uniform constraint on the locations of the GST time points. This showed that the prior in eq. (A5) has little effect on the posterior distribution of the GST history. As before, the terms for each partition are independent and, so, can be combined as a product:

$$p(\mathbf{t} \mid \wp) = \prod_{i=1}^{n_c} p(\mathbf{t}_i \mid \wp). \tag{A6}$$

The prior on the number of points in the GST history is uniform over the range (2–20), hence the prior probability of a GST model of dimension k is $p(k) = 1/19$, the combined prior for n_c partitions is $p(\mathbf{k}) = \prod_{i=1}^{n_c} p(k_i)$.

Combining the prior terms for the temperatures T_i, the time points T_i and the number of (t, T) points in each partition k, gives us the prior used in the acceptance term α. This then takes the form

$$\pi(\boldsymbol{\theta} \mid \wp) = \prod_{i=1}^{n_c} p(\mathbf{T}_i \mid \wp) p(\mathbf{t}_i \mid \wp) p(k_i). \tag{A7}$$

The prior distributions for the heat flow and equilibrium surface temperatures $(\mathbf{q}_0, \mathbf{T}_{\mathrm{eq}})$ are set to uniform over the range (5, 150) mW m^{-2} and $(-5, 15)$ °C. The prior terms relating to these two quantities will always cancel in the acceptance term α (provided no values outside of the specified ranges are accepted), and so, these terms are omitted in subsequent equations.

A2 RJ-MCMC algorithm proposal distributions

Each possible proposal is selected with equal probability (=1/6), except when the number of partitions is one or at the maximum, in which case, we do not allow death or birth, respectively. When the Voronoi centres are changed using one of the proposals (1)–(4) (see Section 5), the overall structure is recalculated. Often, the majority of the tessellation boundaries will remain unchanged relative to the borehole locations. However, each time a new data configuration is arrived at (i.e. a different grouping of the boreholes in one or more Voronoi polygons) a proposal for a new GST history must be made. In Fig. A1, this is illustrated for the case of a birth from two partitions to three—there is no obvious choice for the GST histories $\boldsymbol{\theta}_i'$ of the new partitions. Moreover, the proposal probability for the

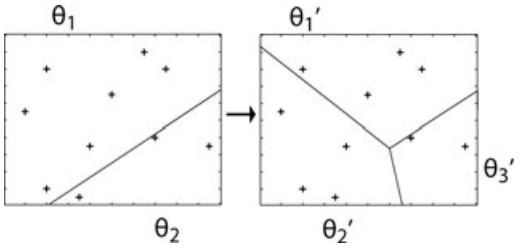

Figure A1. Schematic of a birth step for a Voronoi centre, where crosses denote borehole locations and the lines indicate partition boundaries. Appropriate choices for the new GST histories $\boldsymbol{\theta}_i$' are not obvious as the new temperature histories are now conditioned on a different set of borehole data. A secondary RJ-MCMC algorithm is used and samples from the resultant conditional posteriors are used to as proposals for the $\boldsymbol{\theta}_i$'.

reverse proposal (i.e. the corresponding reduction in the number of partitions) is also required to satisfy detailed balance of the Markov chain.

When a proposal for $\boldsymbol{\theta}_i$ is required in this way or because option (5) is selected, a secondary RJ-MCMC algorithm is run (e.g. Stephenson *et al.* 2006) unless the proposed partition has been previously sampled. This secondary algorithm gives an approximation to the conditional posterior of the GST history in a partition, conditional on the set of borehole profiles enclosed by that partition. One of the GST histories sampled by the secondary RJ-MCMC, $\boldsymbol{\theta}_s$ is then randomly selected and set as the proposal $\boldsymbol{\theta}'_i$. This means that the proposal function for $\boldsymbol{\theta}'_i$ has no dependence on the current state of $\boldsymbol{\theta}_i$, and thus, the proposal term is greatly simplified. The sampled models are all stored for cases in which the algorithm proposes to move back to a previously sampled state, in which case a new $\boldsymbol{\theta}_i$ is drawn randomly from the stored samples.

The secondary RJ-MCMC algorithm is as described in Hopcroft *et al.* (2007) and, so, does not need to be described in detail here. However, many of the details of the secondary RJ-MCMC algorithm are similar to those used in the main algorithm. The priors on the GST histories and the likelihood formulation are both as described above for the main RJ-MCMC. The only difference is that to constrain the model space when inverting a number of borehole profiles simultaneously, the prior on the GST temperatures is modified so that the standard deviation varies linearly from 0.1 to 1.0 K over the reconstruction time length. This prior structure prevents large changes in the GST values compensating for poor proposals of the basal heat flux or equilibrium surface temperature. The narrower prior range in the past also makes sense because the model setup assumes no changes in GST prior to onset of the reconstruction, an assumption common to most GST reconstruction methods. This problem in joint inversion has also been identified by Chouinard & Mareschal (2007) using a singular value decomposition inversion method. The authors resorted to inverting reduced temperature profiles only (i.e. subtracting a calculated equilibrium temperature–depth profile from the measured data).

In both the main and secondary algorithms, the heat flow and equilibrium surface temperature are updated using a bi-variate Gaussian proposal distribution, with an assumed correlation between the two parameters for each borehole of 75 per cent. The standard deviations for each parameter are 4×10^{-5} W m^{-2} and 4×10^{-2} K, respectively. The values sampled by the secondary algorithm for these two parameters are not 'returned' to the main algorithm except at initialization when the initial values of $\boldsymbol{\theta}$, $\mathbf{q}_0$ and $\mathbf{T}_{\rm eq}$ are found. This is to ensure adequate acceptance rates, since perturbing too many parameters at each proposal will lead to very low percentage of acceptances.

A3 Proposal ratio terms for the acceptance probability

The proposal ratio used in the main RJ-MCMC algorithm can be simplified by separating out the four model parameters:

$$\frac{q(\mathbf{m} \mid \mathbf{m}')}{q(\mathbf{m}' \mid \mathbf{m})} = \frac{q(\mathbf{C}, \boldsymbol{\theta}, q_0, T_{\rm eq} \mid \mathbf{C}', \boldsymbol{\theta}', q'_0, T'_{\rm eq})}{q(\mathbf{C}', \boldsymbol{\theta}', q'_0, T'_{\rm eq} \mid \mathbf{C}, \boldsymbol{\theta}, q_0, T_{\rm eq})} = \frac{q(\mathbf{C} \mid \mathbf{C}')}{q(\mathbf{C}' \mid \mathbf{C})} \cdot \frac{q(\boldsymbol{\theta} \mid \boldsymbol{\theta}', \mathbf{C}')}{q(\boldsymbol{\theta}' \mid \boldsymbol{\theta}, \mathbf{C})} \cdot \frac{q(\mathbf{q}_0, \mathbf{T}_{\rm eq} \mid \mathbf{q}'_0, \mathbf{T}'_{\rm eq})}{q(\mathbf{q}'_0, \mathbf{T}'_{\rm eq} \mid \mathbf{q}_0, \mathbf{T}_{\rm eq})}. \quad \text{(A8)}$$

The proposal ratio for centres given a perturb, move or GST update is equal to unity. For a birth step it is given by

$$\frac{q(\mathbf{C} \mid \mathbf{C}')}{q(\mathbf{C}' \mid \mathbf{C})} b = \frac{1/(n_c + 1)}{[1/(\Delta x \Delta y)]}. \quad \text{(A9)}$$

This accounts for choosing a C_i point at random on a spatial domain size ($\Delta x \times \Delta y$). The $1/n_{c+1}$ term accounts for choosing that point for death (following the birth, so that the proposal is reversible).

In the proposal ratio, the choices of new temperature histories $\boldsymbol{\theta}_i$ must be taken into account. The proposal term can be written as

$$\frac{q(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}'_i, \mathbf{C}')}{q(\boldsymbol{\theta}'_i \mid \boldsymbol{\theta}_i, \mathbf{C})} = \frac{p(\boldsymbol{\theta}_i)}{p(\boldsymbol{\theta}'_i)}. \quad \text{(A10)}$$

Therefore, the probability of selecting a particular GST model must be quantified each time, so that detailed balance of the main RJ-MCMC algorithm is satisfied. In the proposal a single GST history $\boldsymbol{\theta}'_i$ is randomly selected from those sampled. The probability of selecting each parameter (k_i, t^{k_i}, T^{k_i}) in $\boldsymbol{\theta}'_i$ is then calculated using 24 yr and 0.1 K increments for time and temperature, respectively. This is conditional on the remaining model parameters being fixed at the values in the selected temperature history $\boldsymbol{\theta}'_i$. The proposal probability $p(\boldsymbol{\theta}'_i)$ is then given by the product of these individual probabilities.

The third term of eq. (A8) is equal to 1, as the proposal distribution is a bivariate Gaussian centred on the current values with a fixed covariance matrix (it does not change between iterations). This term can therefore be discarded in subsequent equations.

The overall proposal ratio (eq. A8) is then given by the products of the proposal probabilities in each partition before and after a proposal is made, multiplied by the proposal ratio for **C** the Voronoi centres and the proposal ratio for the proposal type (birth, death, move, perturb, etc.). Thus, for a birth move, the proposal ratio is given by

$$\frac{q(\mathbf{m} \mid \mathbf{m}')}{q(\mathbf{m}' \mid \mathbf{m})} = \frac{(\Delta x \Delta y)}{n_c + 1} \times \frac{d_{n_C+1}}{b_{n_C}} \times \frac{\prod_{i=0}^{n_c} p(\boldsymbol{\theta}_i)}{\prod_{i=0}^{n_c+1} p(\boldsymbol{\theta}'_i)}, \quad \text{(A11)}$$

where there are n_c partitions originally and n_{c+1} partitions in the proposed model and b_{n_c} and d_{n_c+1} both take the value 1/6 unless the number of partitions reaches the maximum value (or minimum in the case of a death move). To keep the number of partitions in the model in the range $(1, n_b)$, birth (death) moves are not allowed when the number of partitions reaches the maximum (minimum) of this range. The proposal ratio for the move type will therefore be different from 1. In the case that $\mathrm{n_c} = 1$, the move type proposal ratio for a birth is given by $d_2/b_1 = \frac{1/6}{1/5}$, and in the case that $n_c = n_b$, the move type proposal ratio for a death is given by $b_{n_b-1}/d_{n_b} = \frac{1/6}{1/5}$.

A4 Calculating the overall RJ-MCMC acceptance term

When the number of centres changes through a birth or death move, the Jacobian needs to be calculated. However, for the choices of birth and death used here, the Jacobian turns out to equal 1.

The prior ratios for the Voronoi centres **C** will cancel except in the cases of a birth or death. For a birth proposal type the prior ratio on the Voronoi centres is given by

$$\frac{p(\mathbf{C}')}{p(\mathbf{C})} = \frac{n_c + 1}{\Delta x \Delta y}. \quad \text{(A12)}$$

This term will then cancel with the first term of the proposal ratio given by eq. (A11). The same terms appear inverted in the death

move equations and will therefore cancel out in a similar manner. Combining the prior, proposal and likelihood ratios, the acceptance term for a birth step is

$$\alpha = \min\left[1, \frac{\prod_{i=1}^{n_c+1} p(\mathbf{T}'_i \mid \wp) \cdot p(\mathbf{t}'_i \mid \wp) \cdot p(k'_i)}{\prod_{i=1}^{n_c} p(\mathbf{T}_i \mid \wp) \cdot p(\mathbf{t}_i \mid \wp) \cdot p(k_i)} \times \prod_{j=1}^{n_b} \frac{p(\mathbf{d}_j \mid \mathbf{m}'_{v(j)}, k'_{v(j)}, \wp)}{p(\mathbf{d}_j \mid \mathbf{m}_{v(j)}, k_{v(j)}, \wp)} \cdot \frac{d_{n_c+1}}{b_{n_c}} \times \frac{\prod_{i=1}^{n_c} p(\boldsymbol{\theta}_i)}{\prod_{i=1}^{n_c+1} p(\boldsymbol{\theta}'_i)}\right]. \tag{A13}$$

The acceptance term for a death move is similar to that for a birth. In general, each ratio term in the birth move acceptance term is inverted. The proposal probabilities for a death move are different if the algorithm reaches the maximum number of points. For the death step in which one of the Voronoi centres is deleted (with all the other remaining constant), the proposal ratio is given by the inverse of that for the birth move,

$$\frac{q(\mathbf{C} \mid \mathbf{C}')}{q(\mathbf{C}' \mid \mathbf{C})} d = \frac{b_{n_c-1}}{d_{n_c}} \times \frac{[1/(\Delta x \Delta y)]}{1/(n_c)}. \tag{A14}$$

In general, the ratio b_{n_C-1}/d_{n_C} is equal to 1 but takes the value $\frac{1/6}{1/5}$ when the number of centres reaches the maximum. This proposal ratio is needed to satisfy detailed balance as it takes account of not being able to choose a birth move at this stage. The general death move acceptance probability is given by

$$\alpha = \min\left[1, \frac{\prod_{i=1}^{n_c-1} p(\mathbf{T}'_i \mid \wp) p(\mathbf{t}'_i \mid \wp) p(k'_i)}{\prod_{i=1}^{n_c} p(\mathbf{T}_i \mid \wp) p(\mathbf{t}_i \mid \wp) p(k_i)} \times \prod_{j=1}^{n_b} \frac{p(\mathbf{d}_j \mid \mathbf{m}'_{v(j)}, k'_{v(j)}, \wp)}{p(\mathbf{d}_j \mid \mathbf{m}_{v(j)}, k_{v(j)}, \wp)} \frac{b_{n_c-1}}{d_{n_c}} \times \frac{\prod_{i=1}^{n_c} p(\boldsymbol{\theta}_i)}{\prod_{i=1}^{n_c-1} p(\boldsymbol{\theta}'_i)}\right]. \tag{A15}$$

For a move or perturb, or a GST history update, the proposal probability ratio of eq. (A14) is always equal to 1 and, so, can be omitted. The acceptance term is then calculated according to

$$\alpha = \min\left[1, \prod_{i=1}^{n_c} \frac{p(\mathbf{T}'_i \mid \wp) p(\mathbf{t}'_i \mid \wp) p(k'_i) p(\boldsymbol{\theta}_i)}{p(\mathbf{T}_i \mid \wp) p(\mathbf{t}_i \mid \wp) p(k_i) p(\boldsymbol{\theta}'_i)} \times \prod_{j=1}^{n_b} \frac{p(\mathbf{d}_j \mid \mathbf{m}'_{v(j)}, k'_{v(j)}, \wp)}{p(\mathbf{d}_j \mid \mathbf{m}_{v(j)}, k_{v(j)}, \wp)}\right]. \tag{A16}$$

For a heat flux and $\mathbf{T}_{\mathrm{eq}}$ update the prior and proposal terms dependent on $\boldsymbol{\theta}$, $\mathbf{T}$, $\mathbf{t}$ and k can be omitted so that the acceptance term in this case is the ratio of the likelihood values, since the prior and proposal terms will always cancel:

$$\alpha = \min\left[1, \prod_{j=1}^{n_b} \frac{p(\mathbf{d}_j \mid \mathbf{m}'_{v(j)}, k'_{v(j)}, \wp)}{p(\mathbf{d}_j \mid \mathbf{m}_{v(j)}, k_{v(j)}, \wp)}\right]. \tag{A17}$$