

Inference of abrupt changes in noisy geochemical records using transdimensional changepoint models

Kerry Gallagher¹, Thomas Bodin², Malcolm Sambridge², Dominik Weiss³,
Malin Kylander⁴, David Large⁵

1 – Géosciences, Université de Rennes 1, Rennes, 35042, France

2 – Research School of Earth Sciences, Australian National University, Canberra, ACT
0200, Australia

3 – Dept. of Earth Science and Engineering, Imperial College London, London SW7 2AZ,
England

4 – Dept. of Geological Sciences, Stockholm University, 106 91 Stockholm, Sweden

5 – Dept of Chemical and Environmental Engineering, University of Nottingham,
University Park, Nottingham, NG7 2RD, England

Corresponding author :

Kerry Gallagher, kerry.gallagher@univ-rennes1.fr, +33 (0)2 23 23 6081

1 **Abstract**

2
3 We present a method to quantify abrupt changes (or changepoints) in data series,
4 represented as a function of depth or time. These changes are often the result of climatic or
5 environmental variations and can be manifested in multiple datasets as different responses,
6 but all datasets can have the same changepoint locations/timings. The method we present
7 uses transdimensional Markov chain Monte Carlo to infer probability distributions on the
8 number and locations (in depth or time) of changepoints, the mean values between
9 changepoints and, if required, the noise variance associated with each dataset being
10 considered. This latter point is important as we generally will have limited information on
11 the noise, such as estimates only of measurement uncertainty, and in most cases it is not
12 practical to make repeat sampling/measurement to assess other contributions to the
13 variation in the data. We describe the main features of the approach (and describe the
14 mathematical formulation in supplementary material), and demonstrate its validity using
15 synthetic datasets, with known changepoint structure (number and locations of
16 changepoints) and distribution of noise variance for each dataset. We show that when using
17 multiple data, we expect to achieve better resolution of the changepoint structure than when
18 we use each dataset individually. This is conditional on the validity of the assumption of
19 common changepoints between different datasets. We then apply the method to two sets of
20 real geochemical data, both from peat cores, taken from NE Australia and eastern Tibet.
21 Under the assumption that changes occur at the same time for all datasets, we recover
22 solutions consistent with those previously inferred qualitatively from independent data and
23 interpretations. However, our approach provides a quantitative estimate of the relative
24 probability of the inferred changepoints, allowing an objective assessment of the
25 significance of each change.

26
27 **Keywords** : Transdimensional changepoint models, Geochemical data, Bayesian
28 modelling, climate change
29

30

31 **1. Introduction**

32 A major issue in the interpretation of geochemical data (represented as depth or time series)
33 is the detection of changes in trends. Reliable identification of short-term variations
34 superimposed on long-term trends is critical for answering questions about uniformity in
35 the rates of geological processes. The research effort towards understanding climate change
36 often focusses on the inference of rapid or abrupt changes in the mean signal over time.
37 Such environmental changes are recorded by geochemical proxies (e.g. McDermott 2004,
38 Carslaw et al. 2006, Kylander et al. 2007, 2009, 2010, Gutjahr et al 2008, Cloy et al. 2008,
39 Yasahuru et al. 2008, Ruggieri et al 2009, Large et al 2009, Cole et al. 2009, Palmer et al.
40 2010, Burton et al. 2007, 2010). Ideally, recognition of signals from data should based on
41 sound qualitative interpretation, but also involve quantitative inference from the
42 observations, allowing for possibly unknown noise levels in the data.

43 Here we recognise that the definition of abrupt or rapid is subjective. The definition
44 and identification of change also depends on the form of the trend we expect between
45 changes. We define abrupt changes as statistically significant variation in the trend over a
46 scale of one or two samples of the total dataset. Furthermore, we note that the conversion of
47 depth to time (or age) generally involves calibration of a depth-age relationship, which
48 itself will have uncertainty (e.g. Thompson and Goldstein 2006, Kylander et al. 2009,
49 2010). For data collected from one borehole for example, these uncertainties will not
50 change the positions of the underlying changes, only affecting the inference of absolute
51 timing, but may become important for the inference of simultaneous changes in data from
52 different locations.

53 To provide a brief overview of some approaches for inferring abrupt changes in
54 geochemical records, we draw on a selection of published work, including some of our
55 own. Large et al. (2009) described geochemical data including C, N, H concentrations and

56 C, H and O isotopes from a 6 m core taken in the Hongyuan peatland, eastern Qinghai–
57 Tibetan Plateau. The core material was dated back to 9.6 kiloyears using ^{14}C . The
58 interpreted palaeoenvironmental history was linked to the climate variations in northwest
59 Pacific, the El Niño-Southern Oscillation, movement of the Intertropical Convergence Zone
60 and the East Asian Monsoon. This interpretation was based on a qualitative visual
61 inspection and comparison to other proxy data and interpretations.

62 Ruggieri et al. (2009) developed a method to infer Milankovitch–type cycles from
63 geochemical ($\delta^{18}\text{O}$) data,. They allow for discrete changes (change points) between which
64 the trend (defined by superimposed sine functions) can change abruptly. They apply their
65 model to 2 sets of benthic $\delta^{18}\text{O}$ isotope data with time ranges going back to 2500 and 5000
66 kiloyears. As these authors state, an important limitation of their method is that they do not
67 include the number of change points as a parameter to be inferred directly. Instead they
68 examine the variation of the data fit as a function of the number of change points, and try to
69 identify the upper limit such that adding more change points makes little difference to the
70 data fit. One additional limitation of this approach is that the inference will depend on the
71 errors inherent in the data, although the data fit function adopted by Ruggieri et al (2009)
72 does not incorporate a data error term explicitly. In general, we expect a more complex
73 model (i.e. more change points) given more precise data. Often however, we do not have
74 reliable estimates of the data errors and then choosing a suitable model becomes an issue.

75 Tomé and Miranda (2004), looking for changes in linear trends, fit gradients to time
76 series, subject to constraints on the minimum distance between change points, and the
77 magnitude of the changes in trends. The approach requires a user to specify a range for
78 number of change points, and finds best fitting functions for each value of the number of
79 change points in turn. Although the authors state that they can examine a series of the sums
80 of squares of residuals for each model (with different numbers of change points) to choose

81 an appropriate value, they do not explicitly do this. Rather they seem to favour visual
82 inspection to select a preferred model, which is required to have a constant distance
83 between the changepoints.

84 Finally, Kylander et al. (2007) analysed rare earth elements and lead isotopes from
85 samples in an Australian peatland as inorganic proxies for climatic variations, reflected in
86 atmospheric dust. The changepoint modelling approach described in Denison et al. (2002)
87 was adapted to allow common changepoints for multiple datasets. Kylander et al. (2007)
88 used Eu abundance and Pb isotopes to infer the changepoints, under the assumption that
89 trends in the data between the changepoints could be expressed as a simple linear
90 regression function (a constant value or constant slope). In contrast to previous approaches,
91 the number of changepoints was a parameter to be inferred directly. Moreover, the problem
92 is formulated in a Bayesian framework, and provides probability distributions on the
93 number of changepoints and on the changepoint locations.

94 In this paper, we present a general approach to infer an unknown number of
95 changepoints when the errors, or the noise variances, are unknown for one or more datasets.
96 The data can be irregularly spaced in depth (or time) and there is no requirement for
97 different datasets to be sampled at the same depths (time). We specify nothing about the
98 spacing between changepoints, and the approach we present generalises to any linear
99 function between changepoints. A key difference in the approach we present here to that of
100 Denison et al. (2002) is that we decouple the estimates of the data noise variance and the
101 model parameters. A related approach to deal with poorly constrained errors was described
102 by Malinverno and Briggs (2004) applied to seismic traveltimes inversion for 1D velocity
103 structure (and see also Malinverno and Parker, 2006).

104 We begin with a general discussion of changepoint modelling and the role of
105 different sources of variability in the data. Then we describe the approach we adopt,

106 although most of the mathematical details are given in the supplementary material. The
107 method we present includes not only estimation of the number and location of
108 changepoints, the regression functions between changepoints, but also the distribution of
109 the data noise variance (if it is unknown or considered unrepresentative). We conclude with
110 some examples of the application of the method to synthetic and real data and a brief
111 discussion/summary.

112

113

114 **2. Changepoint modelling and noise**

115 The general problem can be stated as follows (and see figure 1) ; given one or more
116 sets of (geochemical) data, $f(x_i)$, $i = 1, N$, at positions x_i , representing a depth or time series,
117 with either known or unknown levels of noise, can we identify the underlying trends or
118 signal (e.g. the mean or a more general regression function) and the locations of discrete
119 changes in the trends ? Often ,we may want to infer the same changepoint locations, but
120 with different signals in each dataset.

121 In general we do not know how many changes are appropriate and ideally we
122 should estimate this from the data. The problem becomes a transdimensional inverse
123 problem (that is we do not specify in advance the number of unknown parameters, e.g.
124 Sambridge et al. 2006). Furthermore, there is a trade-off between the level of noise in the
125 data, and how well we expect to fit the data. Here we broadly follow the philosophy of
126 Scales and Snieder (1999) in considering noise as that part of the data that we do not
127 wish/expect the model to explain. The spread in a geochemical dataset can then be divided
128 into the variation (σ_{GP}^2) due to time/depth varying geological processes (which we are
129 interested in understanding) and the variation due to geological (σ_{GN}^2) and analytical (σ_{AN}^2)
130 noise. Here geological noise might arise from spatial or temporal variations in small scale

131 or short term processes, local geological/biological/atmospheric variability and analytical
 132 noise (or more commonly, errors) typically can arise from factors such as instrumental drift
 133 or calibration from imperfect standards. In general we can express the total variance, σ_T^2 as

134

$$135 \quad \sigma_T^2 = \sigma_{GP}^2 + \sigma_{GN}^2 + \sigma_{AN}^2 \quad (1)$$

136

137 When we are interested in identifying trends and changes in trends, we would like to
 138 choose a model that adequately explains the variation due to geological process, and then
 139 the residuals between the model predictions and the observed data reflect the 2 sources of
 140 noise as defined above. We want good control on the noise (or noise variance) as this will

141 directly influence how well we should fit our data. Intuitively, we can see that if we
 142 consider a series of scattered data with lower noise variance than is appropriate then we
 143 will tend to fit many changepoints. An extreme case would be if we assume no noise, then
 144 we will fit the data perfectly (with a changepoint between each data point). On the other

145 hand if we regard the data as more noisy than they really are, then we will tend to fit a
 146 model that has too few changepoints.

147

148 **2.1 Bayesian formulation of the change-point modelling problem**

149 Underpinning the Bayesian approach is that unknowns are expressed in terms of probability
 150 density functions (e.g. Tarantola and Valette 1982). A common form of Bayes' rule is

$$151 \quad p(m|d) \propto p(d|m)p(m) \quad (2)$$

152 where $p(m|d)$ is the probability density function (PDF) of the unknown model parameter
 153 vector, m , containing the unknowns, given the data vector, d ; $p(d|m)$ is the likelihood

154 function which is effectively the probability of the data, d , being observed given the
 155 model, m . The likelihood increases as the model fits the data better relative to the data
 156 noise and the form of the likelihood depends on the statistical character of the noise on the
 157 data. Finally, $p(m)$ is the prior PDF on the model (that is what we think we know about m
 158 before we have the data). The aim of Bayesian inference is to try and estimate the posterior
 159 PDF, $p(m|d)$ as this characterises all we need to know about the distribution of model
 160 parameter values, given the prior and the information contained in the data (incorporated
 161 through the likelihood). Useful references for Bayesian inference are by Box and Tiao
 162 (1973), Lee (1989), Bernardo and Smith 1994) and Gelman *et al* (2004).

163

164 **2.1.1 Model parameters**

165 In a changepoint problem, the unknown model parameters are the number of
 166 changepoints (n), their locations (c), the parameters of a regression function between the
 167 changepoints (A) and the noise level (σ) for each dataset being considered. Thus we can
 168 write a general model vector, m , as

$$169 \quad m = (n, c, A, \sigma)$$

170 in which c , A , and σ can all be vectors.

171 We write the unknown locations of changepoints as c_i , $i = 1, n$ (note that n itself is
 172 also an unknown). We refer to the region between each changepoint as a partition, and a
 173 predictive regression function, $f_i(x)$, is defined whose parameters (A) depend on the data in
 174 that partition. Thus $f_i(x)$ refers to the regression function within the partition at the left of
 175 change point c_i (if these are considered on a horizontal axis). As there are $n+1$ partitions for
 176 n change points, we define c_0 as the location of the first data point so the regression
 177 function f_1 in the first partition is defined between c_0 and c_1 .

178 Given vectors of independent and dependent (observed) variables, \mathbf{x} and \mathbf{d}_{obs} , such
 179 as depth (or time) and observed data respectively, the linear regression function between
 180 changepoints can be written as

$$181 \quad f(x) = \sum_{i=1}^M \alpha_i G_i(x) \quad (3)$$

182 where G_i represents a specified basis function and α_i represents an unknown coefficient,
 183 which can be thought of as weights on each basis function. For example, the common
 184 straight line relationship given by $d_{pred} = \alpha_1 + \alpha_2 x$, is written as a vector-matrix equation,

$$185 \quad \mathbf{d}_{pred} = \mathbf{G}^t \mathbf{A}$$

186 where the superscript t represents the matrix transpose, and

$$187 \quad \mathbf{G} = \begin{pmatrix} 1 & x_1 \\ \cdot & \cdot \\ 1 & x_i \\ \cdot & \cdot \\ 1 & x_k \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \quad (4)$$

188

189 A model with a constant value (normally the mean) is written as

$$190 \quad \mathbf{G} = \begin{pmatrix} 1 \\ \cdot \\ 1 \\ \cdot \\ 1 \end{pmatrix}, \quad \mathbf{A} = \alpha_1 \quad (5)$$

191 As we can separate the basis functions from the coefficients, this is a linear problem (linear
 192 in terms of the unknown coefficients). Here, we use only constant value regression functions
 193 between changepoints. However the approach we present generalises to any linear function,
 194 allowing for more gradual transitions between states (e.g. linear drift).

195

196 **2.1.2 Data likelihood**

197 As commonly assumed in many geochemical problems, we make an implicit assumption
 198 that the noise associated with the data is normally distributed with a mean of $\bar{\varepsilon}$ and a
 199 variance, σ^2 , i.e. if we have individual errors represented by ε_i , then we can write the
 200 distribution of the errors as

$$201 \quad p(\varepsilon_i) = N(\bar{\varepsilon}, \sigma^2) \quad (6)$$

202 and the variance of this noise distribution may be poorly (or not) known. This is equivalent
 203 to the residuals between the observations and the predictions being normally distributed
 204 with mean of zero and a variance of σ^2 . If we have reliable and representative estimates of
 205 the noise variances, we can use the values in the likelihood instead of σ . In practice, the
 206 noise term determines the uncertainty in fitting the data which, due to geological
 207 complexity, is usually greater than the reported analytical precision. Given this assumption,
 208 a Gaussian likelihood function for observed data (using $d = d_{obs}$) in the i -th partition, given
 209 the predictions with a particular set of model parameters (\mathbf{m}) is

$$210 \quad p(d_i | \mathbf{m}) = \prod_{j=1}^{k_i} \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2} \left(\frac{d_{i,j} - f_i(x)}{\sigma} \right)^2} \quad (7)$$

211 where the subscript j refers to the k_i data in the i -th partition (i.e. the region bounded by
 212 changepoints c_i and c_{i-1} , with c_0 defined the lowest value of the data locations) and $d_{i,j}$ is the
 213 j -th observation in partition i .

214 More generally, for n partitions and N_d different datasets, we have the joint likelihood
 215 function,

216

$$217 \quad p(d_1, d_2 \dots d_{N_d} | \mathbf{m}) = \prod_{l=1}^{N_d} \prod_{i=1}^n \prod_{j=1}^{k_i} \frac{1}{(2\pi\sigma_l^2)^{\frac{1}{2}}} e^{-\frac{1}{2} \left(\frac{d_{i,j,l} - f_{i,l}(x)}{\sigma_l} \right)^2} \quad (9)$$

218

219 Given we subsequently assume a constant value in a partition, an appropriate form to chose
 220 values for the predictive function (f) is a normal distribution centred on the mean value
 221 (and a variance equal to the variance) of the data in that partition. In this case, values for
 222 the regression model parameters (A) can be drawn from this distribution while the mean is
 223 the most probable value from the posterior distribution in a Bayesian formulation,
 224 equivalent to the maximum likelihood estimate. We could use a similar approach for any
 225 linear function with unknown coefficients, using standard least squares inverse methods
 226 (e.g. Menke 1989) to find the maximum likelihood values and the covariance matrix for the
 227 coefficients for a given partition, and draw samples for the parameters.

228

229 **2.1.3 Prior distributions**

230 In a Bayesian formulation, we need to specify prior distributions on all unknown
 231 parameters. The priors reflect what we consider reasonable to assume about the possible
 232 values for each parameter. Bayes's rule lets us use the information about the model
 233 parameters contained in the data to update our prior information (i.e. to produce the
 234 posterior distribution). If the posterior distribution is the same as the prior, then the data
 235 have told us nothing we did not already know.

236 We use the law of hierarchical probability to write the priors as

$$\begin{aligned}
 237 \quad p(m) &= p(n)p(c, A, \sigma | n) \\
 &= p(n)p(c | n) \prod_l^{N_d} p(A_l | c, n) p(\sigma_l | c, n)
 \end{aligned}
 \tag{10}$$

238

239 where N_d is the number of datasets. The choice of priors in our formulation is
 240 straightforward (usually uniform between a specified minimum and maximum value) and
 241 are given in Supplementary material, SM2.

242

243 3. Markov chain Monte Carlo method for solving the changepoint problem

244 In using the Bayesian formulation described above, our goal is to generate a collection, or
 245 ensemble, of values approximating the posterior distribution, whose form we do not know
 246 in advance. As we also do not know the number of changepoints, the problem becomes
 247 what is known as transdimensional, where the number of model parameters itself becomes
 248 an unknown. To solve this problem we use a generalised version of Markov chain Monte
 249 Carlo (MCMC) sampling, known as Reversible Jump MCMC (Green 1995, 2003). A
 250 general introduction to MCMC methods is given by Gilks et al. (1996), a review of
 251 transdimensional Markov chains is given by Sisson (2005) and Gallagher et al. (2009)
 252 present an overview of the general methodology and its application to Earth Science
 253 problems. Specific applications to Earth Science problems have been presented by
 254 Malinverno (2002), Malinverno and Leaney (2005), Stephenson et al (2006), Jasra et al.
 255 (2006), Sambridge et al. (2006), Hopcroft et al. (2009), Charvin et al. (2008) , Bodin and
 256 Sambridge (2009) and Piana Agostinetti and Malinverno (2010). We give brief overview of
 257 MCMC in the Supplementary material, SM1, while here we describe the aspects important
 258 for the transdimensional changepoint problem. The mathematical details for the MCMC
 259 implementation we adopt for the changepoint problem are given in Supplementary
 260 material, SM2 and SM3.

261 MCMC is an iterative method, and at each iteration, we consider 2 sets of model
 262 parameters, the current and proposed models (m_c and m_p). The procedure for a given
 263 iteration can be described as follows

- 264 (i) Randomly perturb the current model to produce the proposed model
- 265 (ii) Randomly accept or reject the proposed model (in terms of replacing the
 266 current model), according to the acceptance criterion ratio (see equation
 267 A1.2 in Supplementary material, SM1) .

268 In principle, after many iterations, the MCMC sampler should converge to a stable
269 configuration (that is sampling according to the posterior distribution) and the final stage of
270 MCMC is to use the sampling to infer characteristics and uncertainties for the model.

271

272 (i) Model perturbations/moves

273 As stated above, the sampling should converge to the target posterior distribution.
274 However, the efficiency of the method does depend on choosing a reasonable proposal
275 function to avoid moving too slowly around the model space as of result of the
276 perturbations being either too small or too big (see figure 2 of Gallagher et al. 2009). The
277 scale of model parameter perturbations can be tuned to achieve a reasonable balance
278 between accepting and rejecting proposed models.

279 For the changepoint problem, we define 5 types of model perturbation or move :

- 280 1. Change the location of a changepoint
- 281 2. Change the regression function (mean estimate) in a partition between 2
282 changepoints
- 283 3. Change the value of noise for a dataset (if appropriate for that dataset)
- 284 4. Add a new changepoint (birth)
- 285 5. Remove an existing changepoint (death)

286

287 Each type of move has specified probability of being selected (which forms the jump
288 proposal, R , referred to in Supplementary Material), and these probabilities need to sum to
289 unity. In our problem, these are set to 0.2, 0.15, 0.15, 0.25 and 0.25 in the order listed
290 above. The birth and death probabilities need to be modified to avoid having more/less
291 changepoints than the maximum/minimum values (n_{\max}/n_{\min}). We do this by setting the
292 birth (death) probability to 0.5 (0.0) for a model with n_{\min} changepoints, and the death

293 (birth) probability to 0.5 (0.0) for a model with n_{\max} changepoints. Having selected a
 294 perturbation type for a particular iteration, all other parameters are kept constant.

295

296 (ii) Acceptance criterion

297 For the purposes of describing how we accept or reject the proposed model, we use
 298 a simplified form of acceptance criterion ratio appropriate for 2 models with the same
 299 number of model parameters (and the full expressions for the transdimensional case are
 300 given in the Supplementary material). This can be written as a ratio of probabilities, given as

$$301 \quad \alpha(m_p, m_c) = \text{Min} \left[1, \frac{p(m_p)p(d|m_p)q(m_c|m_p)}{p(m_c)p(d|m_c)q(m_p|m_c)} \right] \quad (11)$$

302 where $\text{Min}[1, Z]$ means we take the minimum of 1 and Z. The terms $p(m)$ and $p(m/d)$ are the
 303 prior and likelihood probabilities for a particular model (and so define the posterior
 304 probability, at least up to the constant of proportionality) . We have already introduced the
 305 concepts of the prior and likelihood functions (and where appropriate, the problem specific
 306 details are given in Supplementary material, 2).

307 The proposal probability, $q(m_p/m_c)$ determines how we move from the current
 308 model to proposed model (step (i) above). The theory underlying MCMC requires us to be
 309 able to reverse such a move (so we need to also include the reverse proposal probability in
 310 the ratio). If we consider prior distributions to be uniform (i.e. all models have the same
 311 probability), and if we have proposal distributions that are symmetric, then we can write
 312 equation 14 above as

$$313 \quad \alpha(m_p, m_c) = \text{Min} \left[1, \frac{p(d|m_p)}{p(d|m_c)} \right] \quad (12)$$

314 This is just a ratio of the likelihoods (i.e. the probability of the proposed and current models
 315 producing the observed data). Thus if the proposed model fits the observed data better than

316 the current model (so it has a higher likelihood), then the likelihood ratio is > 1 , and then
317 $\alpha(\cdot)$ is set to 1. Conversely, if the current model fits the data better than the proposed
318 model, the ratio is < 1 , and then $\alpha(\cdot)$ is set to $p(\mathbf{d}/\mathbf{m}_p)/p(\mathbf{d}/\mathbf{m}_c)$, which itself is always > 0 .
319 The final step in an iteration requires us to generate a uniform random number, u , between
320 0 and 1, and compare this to $\alpha(\cdot)$. If $u \leq \alpha(\cdot)$ we accept the proposed model (and this
321 becomes the current model for the next iteration), otherwise we reject it (and we retain the
322 current model for the next iteration). From this, we can see that (given the assumptions
323 about flat priors and symmetrical proposal functions) we will always accept a proposed
324 model fits the data better as u is always ≤ 1 . If we consider a proposed model that fits the
325 data almost as well as the current model (say $\alpha(\cdot) = 0.95$), then, on average, we will accept
326 the proposed model 95% of the time. For a proposed model considerably worse than the
327 current model in terms of data fit (say $\alpha(\cdot) = 0.05$), then, on average, we will only accept
328 the proposed model 5% of the time. If the posterior distribution resembles a normal
329 distribution, we can see that this process will tend to concentrate the sampling under and
330 around the peak (the higher probability region), but also allows us to sample less good
331 models (out in the tails of the distribution). In fact, the number of accepted samples for
332 each model is proportion to the posterior probability of that model and the ensemble of
333 accepted models then is a good approximation of the posterior distribution.

334

335 For the first 3 moves described in (i) above, the number of model parameters is
336 constant and the acceptance ratio is given by equation 11. For the birth and death moves,
337 the dimensions of the current and proposed model are different and it is necessary to use
338 the acceptance ratio given in Supplementary Material 1 (equation A1.2). Intuitively, we
339 might expect that models with more parameters will tend to provide a better fit to the
340 observed data, and then that the sampler would tend always to increase the number of

341 model parameters towards the maximum. However, to demonstrate how MCMC operates
 342 during transdimensional moves, we can consider again a simplified form of the acceptance
 343 criterion as

$$344 \quad \alpha(m_p, m_c) = \text{Min} \left[1, \frac{p(m_p)p(d|m_p)}{p(m_c)p(d|m_c)} \right] \quad (13)$$

345
 346 If all the parameters are independent and we consider a proposed model with 1 more
 347 parameter more than the current model (all other parameters being the same), then, with the
 348 prior on the extra parameter given as $p(m_{p,n+1})$ we can write this as

$$349 \quad \begin{aligned} \alpha(m_p, m_c) &= \text{Min} \left[1, \frac{p(m_{p,+1})p(m_c)p(d|m_p)}{p(m_c)p(d|m_c)} \right] \\ &= \text{Min} \left[1, \frac{p(m_{p,+1})p(d|m_p)}{p(d|m_c)} \right] \end{aligned} \quad (14)$$

350
 351 If both models fit the observations equally well (the likelihoods have the same value), then
 352 as $p(m_{p,n+1}) < 1$, we have

$$353 \quad \alpha(m_p, m_c) = p(m_{p,+1}) \quad (15)$$

354
 355 Thus, in this special case, the acceptance probability is equal to prior probability on the
 356 extra model parameter. In other words, even when the fit to the observations is as good as
 357 the current model, the proposed model (with more parameters) is less likely to be accepted,
 358 by a factor equal to the prior probability of the extra parameter.

359 More specifically, when we propose an increase in the number of changepoints
 360 (birth), an increase in likelihood function will tend to encourage acceptance of the proposed
 361 model. However the decrease in the prior ratio will tend to discourage acceptance due to

362 the increased dimensionality of the space. Overall, the algorithm always prefers a large
 363 partition rather than two small partitions with similar mean values (which would have
 364 similar likelihood values). This is an example of a property of Bayesian inference referred
 365 to as ‘natural parsimony’, which means that given a choice between a simple and complex
 366 models that provide similar fits to data, the simpler one will be favoured (e.g. Jeffreys and
 367 Burger 1992, Mackay 1992, Bretthorst 1993, O’Ruanaidh and Fitzgerald 1996, Sivia 1996,
 368 Jaynes 2003).

369

370 (iii) Calculation of model and uncertainties

371 Typically the MCMC sampling is run for many (10^4 - 10^6) iterations, and includes an
 372 initial period during which the samples are not yet from the target posterior distribution.
 373 This is known as burn-in and these samples are discarded before making inference from the
 374 posterior distribution. Gilks et al. (1996) show examples of these characteristic behaviours
 375 as a guide for their recognition and we discuss this later with the examples. The post-burn-
 376 in samples should then provide a good approximation to the posterior distribution for the
 377 model parameters, i.e. $p(\mathbf{m}|\mathbf{d})$. This can be visualised for one model parameter by plotting a
 378 histogram. We can also calculate the expected model as a average, i.e.

379

$$380 \quad \bar{m} = \frac{1}{N} \sum_{i=1}^N m_i \quad (16)$$

381

382 which is effectively a weighted mean, in which the weighting is the posterior probability
 383 for each model. Similarly the variance and co-variance of model parameters are given by
 384 standard formulae. Finally, we can readily calculate the 95% credible intervals by ordering
 385 the samples for a particular variable, and simply identifying the upper and lower 2.5% of
 386 the distribution as the 95% credible interval.

387 Rather than choosing the best data fitting model, which tends to be overly complex,
388 our preferred final solution is given by the expected model (equation 16) with 95% credible
389 intervals around the regression function parameters, and the distributions on the number
390 and locations of changepoints. When a large number of models are added together, their
391 partitions overlap so the average model is continuous and smooth. An advantage of this is
392 that we can produce a model that contains the features common to the majority of sampled
393 models, but also can be more complex (yet smoother) than any individually sampled model.

394

395 **4. Examples of changepoint modelling**

396

397 We first use synthetic data to demonstrate that we can recover the known signal and noise
398 terms. In this example, we discuss how to assess whether the MCMC sampler has
399 performed adequately. Subsequently, we apply the method to 2 sets of real geochemical
400 data, from peat cores in northeast Australia (Kylander et al. 2007) and eastern Tibet (Large
401 et al. 2009).

402

403 ***4.1 Synthetic data***

404 The synthetic data are shown in figures 2a,b,c. We randomly selected 4 changepoints and
405 different mean value functions in each partition to produce 150 irregularly distributed
406 samples for each dataset and added noise with different levels to each. We used these data,
407 assuming unknown noise variance, to infer the distributions on the number and locations of
408 changepoints and the noise variance.

409 As stated earlier, we tune the proposal functions to achieve adequate sampling of
410 the model parameters. Among the more common ways to assess if these input parameters
411 are appropriate are to examine the rate of acceptance (typically around 30%, although 10-

412 60% may be adequate, e.g. Brooks et al., 2003), and also the behaviour of the likelihood or
 413 model parameters as a function of iteration (they should show no long term trend, and
 414 ideally resemble white noise). However, for birth and death we can not readily control the
 415 acceptance rates, which can be much lower (<5%) in this problem.

416 We choose a proposal function scale (θ in equation A2.6) that is proportional to the
 417 range (maximum and minimum values) of particular parameter. We use 0.2, 0.05 and 0.025
 418 of the prior ranges for the regression function (here the mean), the changepoint locations
 419 and the noise, respectively. We make exploratory runs in which we monitor the acceptance
 420 rates over 10^4 - 10^5 iterations, and adjust the scaling parameters accordingly. In this problem,
 421 if the acceptance rate is too high, the scaling parameters are too small, and vice-versa.
 422 Having tuned the proposal functions, we run the chain for 5×10^5 iterations, with a burn-in
 423 of 2.5×10^5 iterations.

424 In figure 3, we show the log likelihood (data fit), the number of partitions and the
 425 sampling for the 3 noise parameters. During the early part of sampling (figure 3a,b), there
 426 is clearly structure in the chain. The initial log-likelihood is about -1.2×10^4 , but even over
 427 the initial 5000 iterations we see that the sampler quickly arrives in better regions of the
 428 model space, and the log-likelihood increases rapidly, even though the number of
 429 changepoints is decreasing. Over the same sampling period, the noise values have not
 430 equilibrated either, the blocky structure, indicative of relatively poor movement around the
 431 model space (or mixing). In contrast, the post-burn parts of the chain show that the
 432 sampling appears to have reached equilibrium. There are no significant trends in the
 433 sampling (they look like white noise), and the number of partitions is sampled between 4
 434 and 7.

435 As described in section 3, we use the post-burn-in samples to calculate the expected
 436 (or average) changepoint structure, the 95% credible intervals around it, the probability of

437 having changepoints over the range of the samples and also the mean and distribution on
438 the noise values for each dataset. In figures 2d,e,f we show the 3 datasets, with the mean
439 estimated noise value as error bars, together with the expected changepoint structure, and in
440 figure 4 we show the distributions on the noise parameters. It is clear, by comparison with
441 figure 2, the changepoint structure has been recovered well, with no spurious features, and
442 also the mean noise is in good agreement with the original values. In figure 5 we show the
443 distribution on the inferred number of changepoints, demonstrating that the inference leads
444 to about 80% probability there are 4. This is conditional on all the model assumptions (a
445 finite number of discrete changes with constant mean values in each partition), although
446 these are appropriate in this example.

447 To demonstrate the influence of different datasets, we ran each dataset
448 independently, using the same parameters as the joint run and the results are shown in
449 figure 6. Again the main changepoint structure is recovered, although we see that some small
450 scale artifacts have been introduced for individual datasets. This leads to a slightly different
451 distribution on the number of changepoints, although all 3 datasets still have 4 as the most
452 probable, with the frequency of 5 changepoints, relative to 4, being higher than in the joint
453 model (25% compared with 98% dataset 1, 65% dataset 2, and 30% dataset 3). In practice, it
454 is unlikely that we would be primarily interested in the absolute number of changepoints,
455 but rather in where changes are inferred to occur. If we have multiple datasets, and expect
456 them to have the same changepoint structure, then we recommend modelling them together
457 for consistency in the changepoint structure.

458

459 **4.2 Real data examples**

460 In the real data examples described below, we follow a similar approach as described
461 above. We monitored the acceptance rate on all variable parameters, and where relevant we
462 adjusted the proposal function scales to achieve an acceptance rate of 20-40%.

463

464 **4.2.1 Lynch's Crater, Australia**

465 The first real data example we consider is from Kylander et al (2007), who undertook
466 geochemical analyses on a 13m section equivalent to ~50Kyr from a peatbog at Lynch's
467 crater, north-east Australia. They reported Pb isotope and Rare Earth Element data, used as
468 proxies for climate change, and in particular for variations in air-transported mineral dust
469 sources. Here we follow that paper, and consider the $^{206}\text{Pb}/^{207}\text{Pb}$ isotope ratios and the
470 Europium anomaly, $(\text{Eu}/\text{Eu}^*)_{\text{PAAS}}$, which is a measure of Eu^{2+} fractionation from Eu^{3+}
471 relative to the adjacent ions, Sm^{3+} and Gd^{3+} (Eu^* being the geometric mean of these two).

472 We first use this example to demonstrate the inference of changepoints either
473 representing the noise with the analytical errors or estimating the noise variance directly
474 from data. In terms of analytical errors, those for the Pb isotope ratio were determined from
475 a long-term series (nearly 2 years) measurements of the NBS 981 standard. In the absence
476 of an equivalent estimate from repeat measurements, we assumed 10% of the observed
477 value for the Eu anomaly. The mean noise variances on Pb and the Eu anomaly are
478 3.86×10^{-4} and 0.123, respectively.

479 In figure 7 we show the inferred changepoint structure using these specified errors
480 in the data likelihood, and also the case in which we also infer the noise variances, for each
481 dataset in terms of a probability distribution (figure 8). Clearly, the structure in the first
482 case is dominated by the Pb isotope ratios (which have relatively small analytical errors).
483 The mean number of changepoints is 106 (± 4 , 1σ) and it is difficult to make meaningful

484 sense of these results. In the second case, where we estimate the noise variance, the mean
485 number of changepoints is $6 (\pm 1, 1\sigma)$, and correspondingly, the mean of the noise variances
486 on the Pb isotope data and the Eu anomaly are 0.00913 and 0.0341, respectively. In this
487 case, the inferred noise variance on the Eu anomaly is about 3-4 times smaller than the
488 assumed 10%, while for the Pb isotopes, it is about 24 times larger than the analytical level.
489 In terms of the inferred changepoint locations, the major peaks around 150 and 620m
490 correspond to the two most significant changes inferred by Kylander et al. (2007). The
491 shallowest changepoint is related to a change to warm, wet conditions while the second is
492 change from humid to arid. The 3rd peak around 820m in our results also corresponds to a
493 lesser change inferred by the earlier work. We refer the reader to Kylander et al. (2007) and
494 Muller et al. (2008), for a more detailed discussion of the environmental significance of
495 these changes.

496 We also ran the 2 datasets individually, assuming the errors were unknown, and the
497 results were very similar in terms of the error distributions. The Pb isotope data however
498 only required one significant changepoint (around 820m), while the Eu anomaly data
499 produced essentially the same result as the joint modeling.

500

501 **4.2.2 Hongyuan, Eastern Tibet**

502 Large et al. (2009) presented a series of geochemical and physical property measurements
503 from a 6m deep section, equivalent to ~10Kyr, of the Hongyuan peatbog in eastern Tibet.
504 The aim of this study was to assess the relative influences of the Indian and east Asian
505 monsoons, and to relate this to other inferences of climate variations in China. Here, we use
506 the C, N, H and $\delta^{13}\text{C}$ analyses, together with the bulk density and carbon density to make
507 quantitative inference of changepoints. In this case, we have no specific information
508 concerning the errors for each dataset, so we also need to infer the noise variance.

509 The changepoint and noise variance distributions are shown in figures 9 and 10. The
510 noise levels are lower than the standard deviation of each dataset (~2-3 times lower), except
511 for the $\delta^{13}\text{C}$ dataset, for which the inferred noise variance is similar in magnitude to the
512 variation in data. The summary diagram of Large et al. (2009) (their figure 7) compares
513 their data to previous studies, and in particular of inferred periods of cold, dry (permafrost)
514 periods relative to warmer, wetter periods. Thus our inferred changepoints should
515 correspond to times when these conditions switch. Apart from the relatively low amplitude
516 probability changepoint inferred around 200 cm and the recent variations (<50cm,
517 attributed to disturbance as a consequence of Yak grazing by Large et al, 2009), the
518 changepoints agree well with those inferred by a qualitative comparison of regional
519 datasets from China by Large et al. (2009) (see figure 9).

520 Although we do not show the results here, we also ran these 6 datasets individually.
521 As we might now expect, the details of the changepoint location structure differs between
522 each dataset. Also, the mean values of the estimated noise levels were lower (by between
523 10 and 60%) than for the joint model. This latter result is also not unexpected as the joint
524 modelling tends to compromise (increase) the noise variance to accommodate common
525 changepoints for multiple datasets. While there is clearly common information, it is not
526 easy to identify reliably the changepoints by considering these datasets individually. Again
527 we recommend joint modelling of multiple datasets if we anticipate a common changepoint
528 structure for a particular problem.

529

530 **5. Summary**

531 Changepoints can be defined as abrupt changes in trends (such as the mean, gradient or any
532 function) over depth or time. In this paper, we have presented a new approach to
533 changepoint modelling, applicable to multiple datasets with common changepoint

534 locations, allowing for unknown noise variance in each dataset. The approach is based on
535 Bayesian transdimensional Markov chain Monte Carlo and we estimate the changepoint
536 structure in terms of distributions for number and location of changepoints, the regression
537 function parameters and the noise variance on multiple datasets. Here, we have considered
538 the regression function in terms of a constant value between 2 changepoints, but the
539 approach generalises to any linear function of the data. In any transdimensional problem,
540 the solution (i.e. the number and location of changepoints) is strongly influenced by the
541 assumed noise variance. Our approach, in allowing us to estimate the noise variance
542 directly from the data, is particularly useful when we do not have reliable estimates of the
543 data error/noise, or perhaps only consider analytical errors (i.e. we neglect natural variation
544 due to geological complexity) and so implicitly assume the data are more precise than is
545 perhaps advisable. Furthermore, the Bayesian approach we adopt is naturally parsimonious
546 and avoids inferring unwarranted complexity when finding the changepoint structure. Thus
547 we expect to favour models with fewer changepoints, while still achieving an adequate fit
548 to the observed data.

549 Using synthetic data, we have demonstrated that we can recover the changepoint
550 structure and the noise variances reliably. When dealing with multiple datasets, we assume
551 that all datasets contain the same changepoint locations, but the response, or regression
552 functions, and noise variances are different. The approach we present can be generalised
553 readily to allow for different noise variance between partitions, if required. Additionally,
554 the different datasets can be irregularly spaced in depth (or time) and there is no need for
555 the data to be sampled at the same depths (time). The details of the solutions will depend on
556 which datasets are used (i.e. singly or jointly) and we recommend using joint modelling if
557 the assumption of common changepoints is considered valid. This assumption is perhaps
558 best assessed from the understanding of geochemical behaviour in different environmental

559 systems. Certainly, the results are more coherent and generally easier to interpret than by
560 combining results from individual dataset modelling. Applications of the method to real
561 datasets from NE Australia and eastern Tibet provide results in agreement with previous
562 qualitative interpretations based on visual inspection. However, our approach is preferable
563 as it is more objective, explicitly incorporates the noise variance (either known or unknown),
564 allows us to assess quantitatively the relative importance of the inferred changepoint
565 structure, and we obtain probability distributions on all parameters. Finally, directions for
566 future work would be to consider transdimensional regression functions (for example we
567 estimate the order of a polynomial which could be different between partitions) and to
568 allow for uncertainty in depth to age conversions (which will be important when comparing
569 records from different locations).

Acknowledgements

We thank the French Australian Science and Technology (FAST) program for support during this work. We thank 2 anonymous referees for comments on an earlier version.

REFERENCES

- Bernardo, J. & Smith, A.F.M., 1994. *Bayesian Theory*, John Wiley and Sons, Ltd., Chichester.
- Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, 178, 1411-1436, doi:10.1111/j.1365-246X.2009.04226.x
- Box, G.E.P. and Tiao, G.C. 1973. *Bayesian Inference in Statistical Inference*, Wiley & Sons, New York, pp 588.
- Bretthorst, G. L. (1993). An introduction to model selection using probability theory as logic. In Heidbreder, G., editor, *Maximum Entropy and Bayesian Methods*, pages 1-42. Kluwer Academic, Dordrecht.
- Brooks, S.P., Giudici, P., & Roberts, G.O., 2003. Efficient construction of reversible jump Markov chain Monte Carlo proposal distribution, *J. R. Stat. Soc., Ser. B90 J. Stat. Method.* 65, 3-39.
- Burton, G.R., Rosman, K.J.R., Candelone, J.P., Burn, L.J., Boutron, C.F. & Hong, S.M., 2007. The impact of climatic conditions on Pb and Sr isotopic ratios found in Greenland ice, 7–150 ky BP. *Earth Planet. Sci. Lett.* 259, 557–566.
- Burton, K.W., Gannoun, A. & Parkinson, I.J., 2010. Climate driven glacial–interglacial variations in the osmium isotope composition of seawater recorded by planktic foraminifera. *Earth Planet. Sci. Lett.* 295 58–68
- Carslaw, D.C., Ropkins, K & M.C. Bell., 2006. Change-Point Detection of Gaseous and Particulate Traffic-Related Pollutants at a Roadside Location. *Env. Sci. Tech.* 40, 22. 6912-6918.
- Charvin, K., Gallagher, K., Hampson, G. & Labourdette, R., 2009. A Bayesian approach to infer environmental parameters from stratigraphic data 1: Methodology, *Basin Research*, 21, 5-25.
- Cloy, J.M., Farmer, J.G., Graham, M.C., MacKenzie, A.B. & Cook, G.T., 2008. Historical records of atmospheric Pb deposition in four Scottish ombrotrophic bogs: an isotopic comparison with other records from western Europe and Greenland. *Glob. Biogeochem. Cy.* 22, GB2016.
- Cole, J.M., Goldstein, S.L., deMenocal, P.B., Hemming, S. R., & Grousset, F.E., 2009, Contrasting compositions of Saharan dust in the eastern Atlantic Ocean during the last deglaciation and African Humid Period. *Earth Planet. Sci. Lett.*, 278, 257-266
- Denison, D.G.T., Holmes, C.C., Mallick, B.K. & Smith, A.F.M., 2002. *Bayesian Methods for Nonlinear Classification and Regression*, Wiley, Chichester, 2002.
- Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M. and Stephenson, J., 2009. Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems, *J. Mar. Petrol. Geol.*, 26, 525-535.
- Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. 2004. *Bayesian Data Analysis. Texts in Statistical Science*, Chapman & Hall. Pp. 668
- Gilks, W.R., Richardson, S. & Spiegelhalter, D.J., 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

- Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82, 711-732.
- Green, P.J., 2003. Transdimensional MCMC, in: P.J. Green, N. Hjort and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, Oxford Statistical Sciences Series, Ch.6, pp. 179-196.
- Gutjahr, M., Frank, M., Stirling, C.H., Keigwin, L. & Halliday, A.N., 2008. Tracing the Nd isotope evolution of North Atlantic Deep and Intermediate waters in the western North Atlantic since the Last Glacial Maximum from Blake Ridge sediments. *Earth Planet. Sci. Lett.* 266, 61–77.
- Hopcroft, P., Gallagher, K. & Pain, C.C., 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.
- Jasra, A. Stephens, D.A., Gallagher, K. & Holmes. C.C., 2006. Analysis of geochronological data with measurement error using Bayesian mixtures *Math. Geol.*, 38, 269-300.
- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, U. K.
- Jefferys, W. H. and Berger, J. O. (1992). Ockham's razor and Bayesian analysis. *American Scientist*, 80:64-72.
- Kylander, M.E., Muller, J., Weust, R., Gallagher, K., Garcia-Sanchez, R., Coles, B. J. & Weiss, D.J., 2007. Rare earth element and Pb isotope variations in a 52 kyr peat core from Lynch's Crater (NE Queensland, Australia): Proxy development and application to paleoclimate in the Southern Hemisphere, *Geochim. Cosmochim. Acta*, 71, 942-960.
- Kylander, M.E., Weiss, D.J. & Kober, B., 2009. Two high resolution terrestrial records of atmospheric Pb deposition from New Brunswick, Canada, and Loch Laxford, , Scotland. *Sci Tot Environ* 407, 1644-1657.
- Kylander, M.E., Klaminder J., Bindler R., & Weiss D.J. 2010. Natural lead isotope variations in the atmosphere. *Earth Planet. Sci. Lett.*, 290, 44-53.
- Large, D.J., Spiro, B., Ferrat, M., Shopland, M., Kylander, M. Gallagher, K., Li, X., Chengde, S., Shen, C., Possnert, G., Zhang, G., Darling, W. & Weiss, D. 2009. The influence of climate, hydrology, and permafrost on Holocene peat accumulation at 3500m on the Eastern Qinghai-Tibetan Plateau, *Quat. Sci. Rev.* 28, 3303-3314.
- Lee, P.M., 1989. *Bayesian Statistics : An Introduction*, Edward Arnold, pp 294.
- MacKay, D. J. C. 1992. Bayesian interpolation. *Neural Comp.*, 4, 415-447.
- Malinverno, A. 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophys. J. Int.*, 151, 675-688.
- Malinverno, A. & Briggs, V., 2004. Expanded uncertainty quantification in inverse problems : Heirachical Bayes and empirical Bayes, *Geophysics*, 69, 1005-1016.
- Malinverno, A. and Leaney, W. S. 2005. Monte Carlo Bayesian look-ahead inversion of walkaway vertical seismic profiles. *Geophys. Prosp.*, 53, 689-703.
- Malinverno, A. & Parker, R.L, 2006. Two ways to quantify uncertainty in geophysical inverse problems, *Geophysics*, 71, W15.
- McDermott, F., 2004. Palaeo-climate reconstruction from stable isotope variations in

speleothems: a review, *Quat. Sci. Rev.* 23 901–918

- Menke, W., 1989. *Geophysical Data Analysis : Discrete Inverse Theory Revised Edition*, International Geophysics Series 45, Academic Press, pp 289.
- Muller, J., M. Kylander, R.A.J. Wüst, D. Weiss, A. Martinez-Cortizas, A.N. LeGrande, T. Jennerjahn, H. Behling, W.T. Anderson, and G. Jacobson, 2008. Possible evidence for wet Heinrich phases in tropical NE Australia: The Lynch's Crater deposit. *Quaternary Sci. Rev.*, 27, 468-475, doi:10.1016/j.quascirev.2007.11.006.
- O' Ruanaidh, J. J. K. and Fitzgerald, W. J. (1996). *Numerical Bayesian Methods Applied to Signal Processing*. Springer-Verlag, New York.
- Palmer M.R., Brummer, G.J, Cooper, M.J., Elderfield, H., Greaves, M.J., Reichart, G.J., Schouten, S. & Yu, J.M., 2010. Multi-proxy reconstruction of surface water pCO₂ in the northern Arabian Sea since 29 ka. *Earth Planet. Sci. Letts* 295 49–57
- Piana Agostinetti, N. & Malinverno A., 2010. Receiver function inversion by transdimensional Monte Carlo sampling, *Geophys. J. Int.* 181, 858-872.
- Ruggieri, E. Herbet, T., Lawrence, K.T. & Lawrence, C.E., 2009. Change point method for detecting regime shifts in paleoclimatic time series : Application to $\delta^{18}\text{O}$ time series of the Plio-Pleistocene, *Palaeocean.*, 24, PA1204, doi:10.1029/2007/PA001568
- Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Transdimensional inverse problems, *Model Comparison and the Evidence*. *Geophys. J. Int.* 1, 167, 528-542.
- Scales, J., & Snieder, R., 1998. What is noise?, *Geophysics*, 63, 1122-1124.
- Sisson, S.A., 2005, *Transdimensional Markov Chains: A Decade of Progress and Future Perspectives*, *J. Am. Stat. Assoc.*, 100, 1077-1090.
- Sivia, D. S. (1996). *Data Analysis: A Bayesian Tutorial*. Clarendon Press, Oxford.
- Stephenson, J., Gallagher, K. & Holmes, C., 2006. Low temperature thermochronology and modelling strategies for multiple samples 2 : Partition modelling for 2D and 3D distributions with discontinuities, *Earth Plan. Sci Letts*, 241, 557-570.
- Tarantola, A. & Valette, B., 1982. Inverse problems = quest for information, *J. Geophys.*, 50, 159-170.
- Thompson, W.G., & Goldstein, S.L., 2006. A radiometric calibration of the SPECMAP timescale, *Quat. Sci. Rev.* 25, 3207–3215.
- Tomé, A.R. and Miranda, P.M.A. 2004. Piecewise linear fitting and trend changing points of climate parameters, *Geophys. Res. Letts.* 31, L02207, doi :10.1029/2003GL019100.
- Yasuhara, M. Cronin, T.M., deMenocal, P.B., Okahashi, H. & Linsley, B.K. 2008, Abrupt climate change and collapse of deep-sea ecosystems. *Proc. Nat. Acad. Sciences.*, 105, 1556-1560.

Figure Captions

Figure 1. An example of the changepoint problem. We have a set of noisy data (dots), with a common noise variance (the noise is Gaussian, and 1σ value is shown as the error bar in the bottom right). The underlying function from which the data were generated is shown by the solid line. The function is discontinuous, with 4 changepoints at $x = 2, 5, 6,$ and 8 . In a real problem, the model parameters are the number of changepoints (n), the locations of the changepoints ($x_i, i = 1, n$) and the values of the function in each region (in this case, this is just the mean value of the data between each changepoint).

Figure 2

(a,b,c). The 3 synthetic datasets (grey dots), with 4 common changepoints. The noise scale (σ from equation 7) is given in the top left, with an error bar $\pm 1\sigma$ shown just below. The true regression function for each dataset is shown by the solid line.

(d,e,f) Changepoint structure inferred for the 3 synthetic dataset. The solid line is the inferred function (relative to the lefthand axis), and the lighter dashed lines represent the 95% credible intervals on this function. The continuous lines represent the probability of a changepoint (relative to the right hand axis). The error bars are drawn using the mean value of the noise variances for each data set (see figure 4).

Figure 3.

- (a) Log-likelihood (LL) and the number of changepoints (n) for the initial 5000 iterations.
- (b) The sampled values for the 3 noise terms over the initial 5000 iterations.
- (c) As (a), but during the post-burn-in sampling
- (d) As (b), but during the post-burn-in sampling

Figure 4. Inferred distributions of the noise variances for each data set. The true values are shown as the heavy vertical lines.

Figure 5. Inferred distribution on the number of changepoints for the data in figure 3.

Figure 6. As figure 2,d,e,f,, but each data set was modelled individually.

Figure 7. Inferred changepoint model for the data from Kylander et al. (2007). Their inferred climatic-related variations are indicated by the grey bar bars at the base of each graph.

- (a) Pb isotope data using the analytical errors for the noise term.
- (b) Eu anomaly data, using 10% of the observed value as the noise term
- (c) As (a), but we infer the noise variance (mean value shown as the error bars)
- (d) As (b), but we infer the noise variance (mean value shown as the error bars)

Figure 8. Inferred distributions of the noise variance for the two datasets of Kylander et al (2007).

Figure 9. Inferred changepoint model for the data from Large et al. (2009). Their inferred climatic-related variations are indicated by the grey bar bars at the base of each graph.

- (a) H, (b) N, (c) C, (d) Total Carbon, (e) $\delta^{13}\text{C}$, (f) density

Figure 10. Inferred distributions of the noise variance for the 6 datasets of Large et al (2009). The sequence of graphs is the same as in figure 9.