RMetS

Royal Meteorological Society

Short Communication

A Bayesian approach to detecting change points in climatic records

Eric Ruggieri*

Department of Mathematics and Computer Science, Duquesne University, Pittsburgh, PA 15282, USA

ABSTRACT: Given distinct climatic periods in the various facets of the Earth's climate system, many attempts have been made to determine the exact timing of 'change points' or regime boundaries. However, identification of change points is not always a simple task. A time series containing N data points has approximately N^k distinct placements of k change points, rendering brute force enumeration futile as the length of the time series increases. Moreover, how certain are we that any one placement of change points is superior to the rest? This paper introduces a Bayesian Change Point algorithm which provides uncertainty estimates both in the number and location of change points through an efficient probabilistic solution to the multiple change point problem. To illustrate its versatility, the Bayesian Change Point algorithm is used to analyse both the NOAA/NCDC annual global surface temperature anomalies time series and the much longer $\delta^{18}O$ record of the Plio-Pleistocene. Copyright © 2012 Royal Meteorological Society

1. Background

The time series that record various aspects of the Earth's climate system are widely recognized as being non-stationary (Hays *et al.*, 1976; Imbrie *et al.*, 1992; Karl *et al.*, 2000; Tomé and Miranda, 2004; Raymo *et al.*, 2006; Beaulieu *et al.*, 2010; among others). Several methods have been implemented to solve the 'change point' problem for shorter climatic time series. For example, Karl *et al.* (2000) fixes the number of discontinuities and then uses both Haar (square) wavelets and a brute force minimization of the residual squared error for the placement of piecewise continuous line segments. Similar to this second approach, Tomé and Miranda (2004) automate the creation of a matrix of over-determined linear equations and consecutively solve this system for every possible combination of change points that satisfies their constraints. To deal with the exponentially increasing number of change point solutions associated with longer time series, dynamic programming change point algorithms have been developed that reduce the computational burden to a more manageable size (Ruggieri *et al.*, 2009). Branch and Bound techniques (Aksoy *et al.*, 2008) also aim to reduce the computational burden by screening and eliminating sub-optimal segmentations.

Alternatively, Seidel and Lanzante (2004) first identify change points by visual inspection and then refine their location so as to: (1) minimize the number of change points; (2) be consistent with previous research; and (3) have support from an iterative non-parametric statistical method (Lanzante, 1996). This iterative approach adds one change point at a time, testing each for statistical significance. In an attempt to minimize the *a priori* assumptions on the number and location of change points, Menne (2006) proposes a semi-hierarchic splitting algorithm to place the change points. Here, the placement of a change point splits the time series, but each splitting step is followed by a merge step to determine whether change points chosen earlier are still significant.

Each of these methods returns a single, 'optimal' solution. But if there are $\sim N^k$ possible placements of k change points in a time series of length N, how confident are we that this one solution is vastly superior to any other, especially one that may only differ by a single data point? A Bayesian approach to the change point problem can give uncertainty estimates not only for the location, but for the number of change points as well.

For computational reasons, Markov chain Monte Carlo (MCMC) (Barry and Hartigan, 1993; Lavielle and Lebarbier, 2001; Zhao and Chu, 2006) and Gibbs Sampling (Stephens, 1994; Khaliq *et al.*, 2007) approaches have dominated Bayesian solutions to the multiple change point problem. However, these techniques only approximate the posterior distribution of change point locations and leave open difficult questions of convergence.

* Correspondence to: E. Ruggieri, Department of Mathematics and Computer Science, Duquesne University, Pittsburgh, PA 15282, USA. E-mail: ruggierie@duq.edu

Bayesian change point algorithms that do not rely on MCMC procedures include Hannart and Naveau (2009) who use Bayesian Decision Theory to minimize a cost function for the detection of multiple change points and Beaulieu *et al.* (2010) who probabilistically locate multiple change points through a splitting algorithm akin to Menne (2006), but without the corresponding merge step. However, these approaches are limited to identifying changes in mean. Fearnhead (2006) developed a recursive algorithm similar to the Forward/Backward equations of a Hidden Markov Model which Seidou and Ouarda (2007) generalized to fit a regression model. With respect to the algorithm presented here, there are two main differences: the nature of the recursion and the prior distributions on the model parameters. Seidou and Ouarda (2007) require two training data sets and a prior distribution on the distance between adjacent change points (an implicit assumption on the number of change points in a time series). Our approach requires neither.

In what follows, we describe an exact Bayesian solution to the multiple change point problem that uses dynamic programming recursions to reduce the computational burden down to a point where a time series of any length can be analysed for an arbitrary number of change points. The key to dynamic programming is to break the multiple change point problem down into a set of progressively smaller sub-problems, the smallest of which (the placement of a single change point) can easily be solved. The full solution can then be obtained by efficiently piecing together the solutions to these sub-problems. The Bayesian Change Point algorithm can detect changes in the parameters of any regression model being used to describe a climatic time series, be it changes in the mean, trend, and/or variance of the climate signal. After describing its implementation, the Bayesian Change Point algorithm is used to analyse both the NOAA/NCDC global surface temperature anomalies time series and the $\delta^{18}O$ proxy record of the Plio-Pleistocene. For the latter, the goal is to show how the algorithm can be applied to very long time series and search for more than just changes in trend. The ability to provide uncertainty estimates in the number and timing change points is a key contribution of the Bayesian Change Point algorithm and a significant advantage over a frequentist approach.

2. Description of the algorithm

Given the dependent variable, Y, and m known predictor variables $X_1, \ldots, X_m$, linear regression methods are based upon the statistical model

$$Y = \sum_{l=1}^{m} \beta_l X_l + \varepsilon \qquad (1)$$

where β_l is the lth regression coefficient and ε is a random error term. For a time series, the predictors, X_l, are functions of time. Suppose that the change point model contains k change points and that the

regression model shown above applies between each consecutive pair of change points whose locations are $C = \{c_1, c_2, \ldots, c_k\}$. The time series is bounded by $c_0 = 1$, the first data point in the time series, and $c_{k+1} = N$, the final data point in the time series. To place a single change point, the probability of the data given the regression model $[f(Y_{i:j}) = f(Y_{i:j}|X)]$ must first be calculated for each and every possible sub-string of the data, $Y_{i:j} = \{Y_i, Y_{i+1}, \ldots, Y_{j-1}, Y_j\}, 1 \leq i < j \leq N$. For example, if the error terms, ε, are assumed to be independent, normally distributed random variables, then $f(Y_{i:j})$ is multivariate Normal. Let $d_{\min}$ be the minimum distance between adjacent change points, I be the identity matrix, and σ^2 be the residual variance. Then:

$$\begin{aligned} &\text{for } i = 1 \text{ to } N \\ &\quad \text{for } j = i + d_{\min} \text{ to } N \\ &\qquad Y_{i:j}|X \sim N(X\beta, \sigma^2 I) \\ &\quad \text{endfor} \\ &\text{endfor} \end{aligned}$$

Starting from one end of the time series, we can find the probability of any prefix of the data ($Y_{1:j}$, the first j data points in a time series) containing one change point by multiplying together the probabilities of two non-overlapping substrings (calculated above) and summing (marginalizing) over all possible placements of the change point. No further information about its location will be needed to solve the multiple change point problem. Let $P_k(Y_{1:j}) = P_k(Y_{1:j}|X)$ be the probability density of the first j observations containing k change points given the regression model. When $k = 1$ this gives, $P_1(Y_{1:j}) = \sum_{v=1}^{j-1} f(Y_{1:v}) \times f(Y_{v+1:j})$ Next, to find the probability density of a prefix of the time series with two change points, $P_2(Y_{1:j})$, multiply together the probability density of a prefix containing one change point, $P_1(Y_{1:v})$, and a non-overlapping substring which fills out the rest of the prefix, $f(Y_{v+1:j})$ (both previously calculated) and then marginalize over all possible placements of the second change point: $P_2(Y_{1:j}) = \sum_{v=1}^{j-1} P_1(Y_{1:v}) \times f(Y_{v+1:j})$. The process continues, $P_k(Y_{1:j}) = \sum_{v=1}^{j-1} P_{k-1}(Y_{1:v}) \times f(Y_{v+1:j})$, until $k = k_{\max}$, the maximum number of change points allowed.

Inferences about the parameters, including the locations of the change points, can be made by using Bayes Rule to sample from the exact posterior distribution of the quantities of interest. Sampling a set of solutions is straightforward and allows us to address questions related to uncertainty. First, a probabilistic sample of the number of change points is drawn, then their locations are recursively sampled, and finally, the parameters of the regression model can be sampled for each regime defined by the locations of the change points. The three steps of the Bayesian Change Point algorithm are detailed below:

[1] **Calculating the Probability Density of the Data $f(Y_{i,j}|\mathbf{X})$:**

Let $X = \{X_1, \ldots, X_m\}$ be the fixed set of regressors included in the regression model (this assumption can be relaxed by using a variable selection procedure to select only a subset of the regressors between each pair of consecutive change points). Our regression model assumes that the error term, ε, is an independent, mean zero normally distributed random variable. Therefore, the likelihood function for a substring of the data, $Y_{i:j}$, is $Y|\beta, \sigma^2, X \sim N(X\beta, \sigma^2 I)$, where I is the identity matrix. Conjugate priors were used for the vector of regression coefficients, β, and for the residual variance, σ^2. Specifically, β is multivariate Normal, $\beta|\sigma^2 \sim N(0, \sigma^2/k_0)$, where k_0 is a scale parameter relating the variance of the regression coefficients to the residual variance and $\sigma^2 \sim \text{Scaled} - \text{Inverse } \chi^2(v_0, \sigma_0^2)$, where v_0 and σ_0^2 act as pseudo data points (essentially, unspecified training data) – v_0 pseudo data points of variance σ_0^2. Let n be the number of data points in a substring. Then we can define the parameters for the posterior distribution on σ^2 and $\beta|\sigma^2$ as $v_n = v_0 + n$, $\beta^* = (X^T X + k_0 I)^{-1} X^T Y_{i:j}$, $s_n = (Y_{i:j} - X\beta^*)^T (Y_{i:j} - X\beta^*) + k_0 \beta^{*T}\beta^* + v_0\sigma_0^2$). β^* represents the posterior mean of the vector of regression coefficients, while v_n and s_n are the parameters for the posterior distribution on the residual variance, σ^2. To find the marginal probability of a substring of the data, $Y_{i:j}$, we integrate out the nuisance parameters, β and σ^2, $f(Y_{i:j}|X) = \int\int f(Y_{i:j}|\beta, \sigma^2, X) f(\beta|\sigma^2, X) f(\sigma^2|X) \mathrm{d}\beta \mathrm{d}\sigma^2$, which yields:

$$f(Y_{i:j}) = f(Y_{i:j}|X)$$
$$= \frac{(v_0\sigma_0^2/2)^{v_0/2} \, \Gamma(v_n/2)(k_0)^{m/2}}{\Gamma(v_0/2)(s_n/2)^{v_n/2}(2\pi)^{n/2}|X^T X + k_0 I|^{1/2}} \tag{2}$$

This quantity is calculated and then stored in memory for all possible substrings of the data, $Y_{i:j}$, with $1 \le i < j \le N$.

[2] **Forward Recursion [Dynamic Programming]:**

Let $P_k(Y_{1:j})$ be the density of the data $[Y_1 \ldots Y_j]$ with k change points. Define

$$P_1(Y_{1:j}) = \sum_{v < j} f(Y_{1:v}) f(Y_{v+1:j}) \tag{3}$$

and

$$P_k(Y_{1:j}) = \sum_{v < j} P_{k-1}(Y_{1:v}) f(Y_{v+1:j}) \tag{4}$$

for $k < j \Leftarrow N$.

[3] **Stochastic Backtrace via Bayes Rule**:

In order to have a completely defined partition function (or normalization constant), $f(Y_{1:N})$, two additional quantities need to be specified: (1) a prior distribution on the number of change points, $f(K = k)$; and (2) a prior distribution on the locations of the change points, $f(c_1, c_2, \ldots c_k|K = k)$, that is,

$$f(Y_{1:N}) = \sum_{k=0}^{k_{\max}} \sum_{c_1 \ldots c_k} f(Y_{1:N}|K = k, c_1 \ldots c_k)$$
$$\times f(K = k, c_1 \ldots c_k) \tag{5}$$

The inner summation is exactly the quantity calculated on the Forward Recursion step (Equation (4)). As for $f(K = k, c_1, \ldots, c_k)$, *a priori*, we place half the probability mass on zero change points $[f(K = 0) = 0.5]$ and the other half on a non-zero number of change points, divided uniformly across the possible values of k $[f(K = k) = 0.5/k_{\max}$, where $k_{\max}$ is the maximal number of allowed change points]. Additionally, we assume that all change point solutions with exactly k change points are equally likely, that is $f(c_1, \ldots, c_k|K = k) = 1/\binom{N}{k}$. This last assumption is a combinatorial prior that directly accounts for the greater number of solutions as the number of change points increases. Together, $f(K = 0) = 0.5$ and for $k > 0$,

$$f(K = k, c_1, \ldots, c_k) = \frac{0.5}{(k_{\max})\binom{N}{k}} \tag{6}$$

As this prior distribution is uniform, it does not have to be included in the Forward Recursion, but can instead be factored out and multiplied in at this point. However, use of a non-uniform prior must be incorporated in to the Forward Recursion step. With this quantity specified, we can now draw samples of the parameters of interest.

[3.1] **Sample a Number of Change Points, k**:

The Forward Recursion calculates the density of the entire data set, $Y_{1:N}$, given k change points, $P_k(Y_{1:N}) = f(Y_{1:N}|K = k)$. Using Bayes Rule, the posterior distribution on the number of change points given the data is:

$$f(K = k|Y_{1:N}) = \frac{P_k(Y_{1:N})f(K = k, c_1, \ldots, c_k)}{f(Y_{1:N})} \tag{7}$$

with $f(Y_{1:N})$ defined in Equation (5).

[3.2] **Sample the Locations of the Change Points, c_k**:

For $k = K, K - 1, \ldots, 1$, the posterior distribution on the location of the change points is:

$$f(c_k|c_{k+1}) = \frac{P_{k-1}(Y_{1:v})f(Y_{v+1:c_{k+1}})}{\sum_{v \in [k-1, c_{k+1})} P_{k-1}(Y_{1:v})f(Y_{v+1:c_{k+1}})} \tag{8}$$

When $k = 1$, $P_0(Y_{1:v}) = f(Y_{1:v})$ defined in (Equation (2))

[3.3] **Sample the Regression Parameters for the Interval between Adjacent Change Points c_k and c_{k+1}:**
The posterior distribution for the residual variance is:

$$\sigma^2|Y \sim \text{Scaled} - \text{Inverse}\chi^2(v_n s_n/v_n) \quad (9)$$

and the posterior distribution for the coefficients of the regression model is:

$$\beta|\sigma^2, Y \sim N(\beta^*, (X^T X + k_0 I)^{-1}\sigma^2) \quad (10)$$

where β^*, s_n and v_n are defined in Step 1: Calculating the Probability Density of the Data.

Calculating the probability density of the data is O(N^2), while the Forward Recursion and Stochastic Backtrace steps are both O(kN). Therefore, the algorithm has a total time complexity of O(N^2).

To run the Bayesian Change Point algorithm, a small number of parameters need to be set by the user, including k_{max}, d_{min}, and three hyperparameters needed to describe the prior distribution of the regression coefficients and the residual variance. A description of these variables and their values can be found in the Appendix.

3. Simulation of a homogeneous time series – no change points

One hundred data sets of size $N = 250$ were generated according to $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, where x_i represents the index of the data point, ε_i is a random error term of standard deviation 2, and β_1 [$\sim$Uniform (-10, 10)]

and β_2 [$\sim$Uniform (-0.10, 0.10)] represent the intercept and trend, respectively. With $k_{max} = 5$ and $d_{min} = 5$, the average posterior probability of zero change points for these one hundred data sets was 0.9996. A specific example, in which $Y_i = -6.25 + 0.05x_i + \varepsilon_i$ is shown in Figure 1. Here, all 500 independently sampled solutions contained zero change points. Had a change point been selected, its most probable location was position 6. Thus, the Bayesian Change Point algorithm is unlikely to produce a change point when none exists.

The ability to detect change points depends on the signal to noise ratio, the magnitude of the change, and the distance between adjacent change points. To highlight the algorithm's ability to detect change points, we analyse two climatic time series and compare our findings with previously detected change points in these series.

4. Results

4.1. NOAA/NCDC annual global surface temperature anomalies

A visual inspection of the NOAA/NCDC global surface temperature (combined land and ocean) anomalies from 1880 to 2010 (Quayle *et al.*, 1999) clearly shows that the increase in temperature is not constant, but that most of the warming takes place during two distinct periods, one beginning around 1910 and the other during the 1970s (Karl *et al.*, 2000; Seidel and Lanzante, 2004; Menne, 2006). This suggests three change points due to the fairly flat regime between the two warming periods. To fit the global surface temperature record, we assume that each temperature regime can be fit by a single linear trend, $Y = \beta_1 + \beta_2 X$ and set $k_{max} = 6$ as a reasonable maximum for the number of change points.
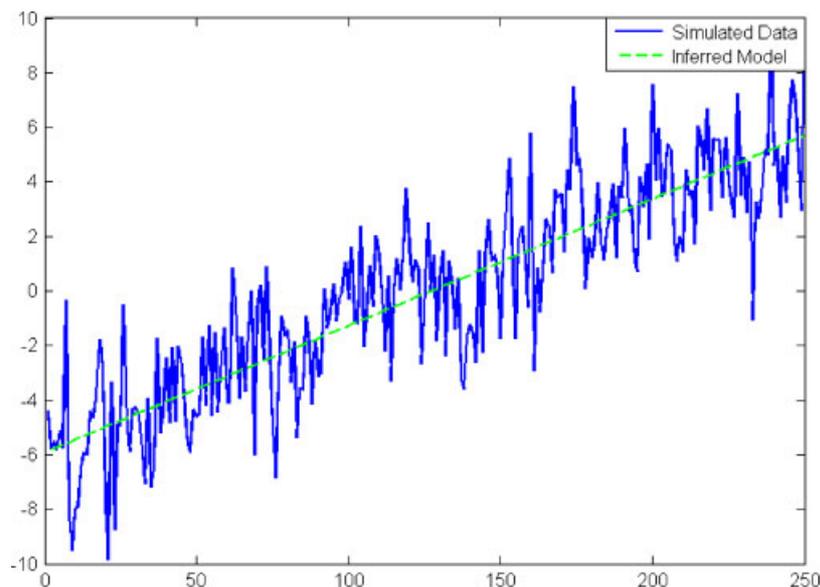
Figure 1. Simulation of a Homogeneous Time Series. A time series of size $N = 250$ was generated according to $Y_i = -6.25 + 0.05x_i + \varepsilon_i$, where x_i represents the index of the data point and ε_i is a random error term of standard deviation 2 (Solid line, Blue). The green (dotted) line represents the average model produced by the Bayesian Change Point algorithm. No change points were identified by any of the 500 sampled solutions. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

Research from Karl *et al*. (2000) and Tomé and Miranda (2004) suggest change points be separated by a minimum of 15 years (d_{min}). Our analysis of the temperature record reflects this constraint. However, the results are essentially unchanged if this constraint is eliminated.

The Bayesian Change Point algorithm indicates that there are most likely three change points in the NOAA/NCDC record (Table I). Five hundred solutions were independently sampled from the exact posterior distribution. Figure 2 shows the average of the 500 sampled solutions along with the locations of the change points, displayed as spikes at the bottom of the figure. The height of these 'spikes' is indicative of the probability of selecting a change point at a specific point in time. Tall thin 'spikes' represent relative certainty in the timing of a change point while shorter and wider 'spikes' indicate more uncertainty.

The timing of the change points matches well with previous studies (Table II). In all 500 sampled solutions, a change point was selected around 1906 (posterior probability of 59.4% for 1906–1907). A 95% credibility limit for this first change point (1902–1914) encompasses the dates of similar change points identified by previous studies (Table II). In 495 of the 500 samples (99.0%) a second change point was selected around 1945 [95% Credibility Limit 1944–1946], ending the trend of increasing temperature that began around 1910 [0.102 K/decade]. Only the change point identified by Karl *et al*. (2000) falls outside of this credibility limit (Table II). A third change point was sampled for 79.6% of solutions (398 of 500), with approximately one third of the probability mass around 1963 and the remaining two thirds

Table I. The posterior distribution of selecting a given number of change points for the NOAA/NCDC global surface temperature anomalies time series (1880–2010).

Number of change points	Posterior probability
0	0.0000
1	0.0006
2	0.2037
3	0.7954
4	0.0004
5	0.0000
6	0.0000

Table II. Comparing the placement of change points by the Bayesian change point algorithm to previous studies of the NOAA/NCDC annual global surface temperature anomalies time series (1880–2010)[a].

Study	Change points		
Karl *et al*. (2000)	1910	1941	1975
Seidel and Lanzante (2004)	n/a	1945	1977
Menne (2006)	1902	1945	1963
Ruggieri *et al*. (2009)	1906	1945	1963
Bayesian Change Point	1906	1945	1963 or 1976

[a] Seidel and Lanzante (2004) do not identify a change in the first decade of the 20th century. In this study, a third change point is selected for only 79.6% of the sampled solutions and it is located at either 1963 or 1976, indicating the importance of both changes in the record.
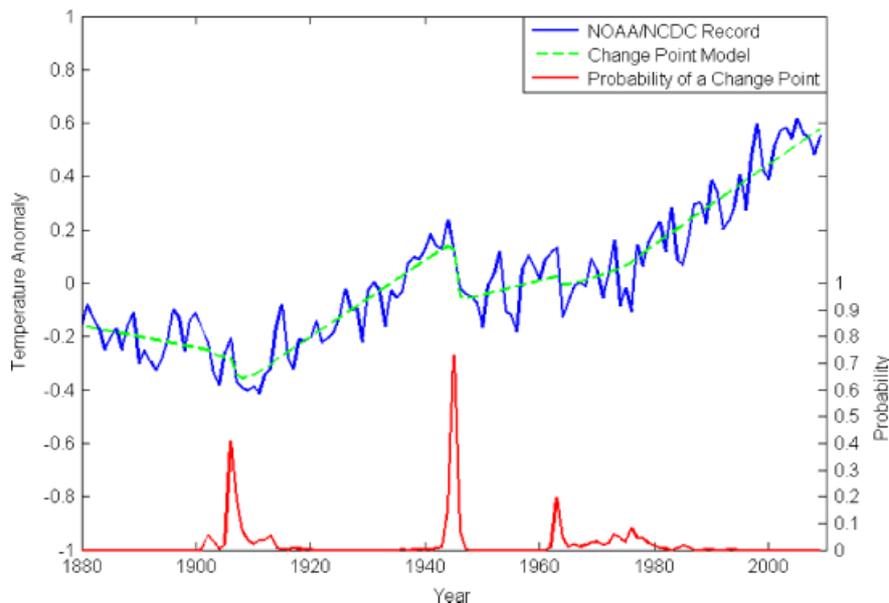
Figure 2. The NOAA/NCDC Global Surface Temperature Anomalies Time Series. The blue (solid) line represents the NOAA/NCDC global surface temperature anomaly time series, while the green (dashed) line represents the model predicted by the Bayesian Change Point algorithm. The height of the (red) 'spikes' at the bottom of the figure indicates the probability of a change point being selected at each time point. Tall, thin 'spikes' indicate relative certainty in the timing of a change point, while shorter and wider 'spikes' indicate more uncertainty in the timing of a change point. Global surface temperatures increased at a rate of 0.102 K/decade from 1906 to 1945 and at a rate of 0.145 K/decade from 1976 to present. An abrupt drop in temperatures in 1945 followed by relatively stable temperatures from 1946 to 1976 (+0.04 K/decade) separates these two climate regimes. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

around 1976. At this point global temperatures increase to the present day at a rate of 0.145 K/decade. The eruption of Mt Agung in 1963 (Ivanov and Evtimov, 2009) and a 1976 climate shift apparent in the Pacific Ocean (Trenberth and Hurrell, 1994) have been cited as abrupt changes in the climate system corresponding to these dates. The 95% credibility limit for this third change point (1963–1986) includes the all shifts identified by previous studies (Table II), but is much wider than for the first two change points due to its bimodal nature. Because the two modes are not well separated, it is not possible to determine a separate credibility interval for each of the modes.

Because of the presence of autocorrelation in the global temperature record, Karl *et al.* (2000), Seidel and Lanzante (2004), and Menne (2006) first sought to objectively identify change points, and then model the residual error within each regime using an autoregressive function. The Bayesian Change Point algorithm can be used without change in a similar manner. In this situation, there would be no impact on the posterior probability or credibility intervals of change points. However, a general caution is that ignoring positive autocorrelation can yield a higher false alarm rate (as a run of positive or negative residuals can be misinterpreted as a change in mean and/or trend) while ignoring negative autocorrelations may miss a true change point – effects which become more pronounced as the autocorrelation increases (Lund *et al.*, 2007). Additionally, in the presence of autocorrelated errors, regression coefficients (β) remain unbiased, but the standard error of these parameters may be underestimated. Taken together, the presence of autocorrelation may in some instances lead to an overly optimistic placement of change points. For the global temperature record, the first order autocorrelation of the residuals is 0.2475, implying that the impact on the results should be minimal.

4.2. δ^{18}O record of the Plio-Pleistocene

Geoscientists capitalize on the changes in the isotopic ratios of oxygen in the ocean to create a δ^{18}O ice volume proxy record from ocean sediment cores which quantifies the amount of ice on the Earth at a specific time in the past (Hays *et al.*, 1976; Imbrie *et al.*, 1992; Lisiecki and Raymo, 2005; among others). The gradual cooling of the Earth over the last 5 million years engendered the formation of more permanent ice sheets over the Northern Hemisphere around 2.7 million years ago (Ma), reflected by an obvious increase in the amplitude of the δ^{18}O at this time (Figure 3). Further cooling of the Earth likely contributed to the Mid-Pleistocene Transition (MPT) (Tziperman and Gildor, 2003; Raymo *et al.*, 2006) around 1 Ma. During the MPT, not only did the amplitude of the δ^{18}O proxy record increase, but the periodicity of the glacial cycles apparently changed from 41 to 100 kyr (Tziperman and Gildor, 2003).

Today, most theories of ice sheet dynamics are a variation of Milankovitch (1941) Theory, which loosely states that ice sheets respond linearly to the amount of solar insolation (i.e. energy) received at the top of the Earth's atmosphere at 65°N Latitude during the summer.
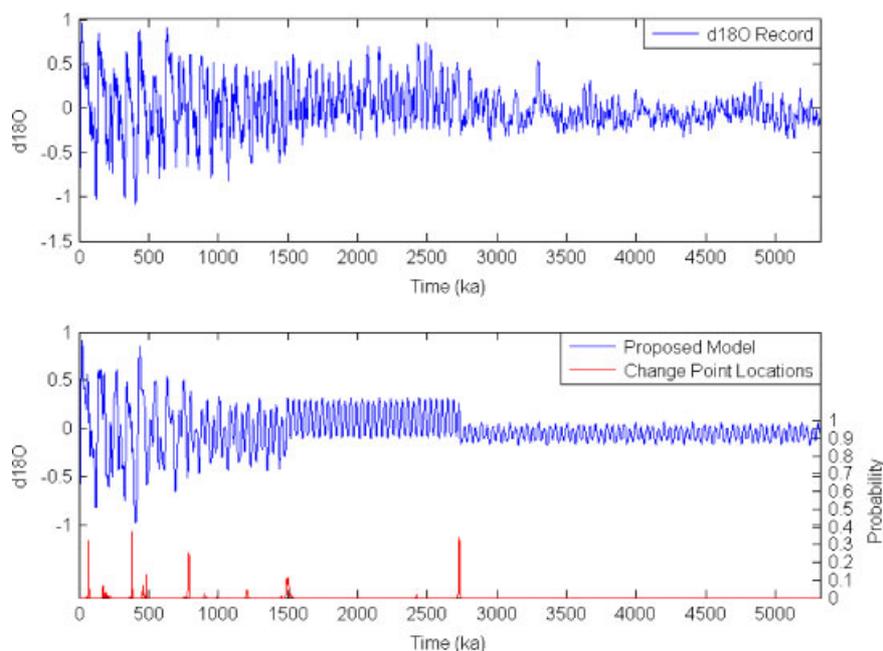
Figure 3. The δ^{18}O Record of the Plio-Pleistocene. The top panel displays the δ^{18}O proxy record of the Plio-Pleistocene (Lisiecki and Raymo, 2005) after removing the long-term cooling trend via an exponential function. The bottom panel displays the model proposed by the Bayesian Change Point algorithm along with the locations of the change points, indicated by the (red) 'spikes' at the bottom of the figure. Tall, thin 'spikes' represent relative certainty in the placement of a change point, while shorter and wider 'spikes' indicate greater uncertainty. An average of 7.16 change points is able to fit 71.6% of the variation in the δ^{18}O proxy record. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

To model glacial activity, sinusoidal functions will be used to approximate the three components of the Earth's orbital motion that impact its solar insolation budget: Precession [23 kyr], Obliquity [41 kyr], and Eccentricity [100 kyr]. For this analysis, we set $k_{max} = 15$ and $d_{min} = 50$ kyr, or roughly half a glacial cycle in the most recent past.

The Bayesian Change Point algorithm indicates that there are most likely seven change points in the $\delta^{18}O$ proxy record (Table III). Five hundred solutions were again sampled from the exact posterior distribution. Figure 3 shows the average of these 500 sampled solutions along with the locations of the change points, while Figure 4 displays the posterior mean of the regression coefficients for each of the three forcing functions at each point in time. The 41 kyr 'world' that existed prior to the MPT (Hays *et al*., 1976; Imbrie *et al*., 1992; Raymo

Table III. The posterior distribution of selecting a given number of change points for the $\delta^{18}o$ proxy record of the Plio-Pleistocene[a].

Number of change points	Posterior probability
5	0.0000
6	0.1399
7	0.5962
8	0.2257
9	0.0363
10	0.0019
11	0.0000

[a] The probability of selecting 0–4 or 12–15 change points is essentially zero and therefore not included on this table.

et al., 2006) saw an increase in the 41 kyr coefficient around 2.73 Ma (95% credibility limit 2.72–2.74 Ma), coinciding with the appearance of ice-rafted debris in the proxy record and the onset of Northern Hemisphere glaciation (Ruddiman, 2007). Beginning around 1.5 Ma (95% credibility limit 1.48–1.53 Ma), glacial suppression of North Atlantic deep water intensified (Raymo *et al*., 1990) and the dominance of the 41 kyr cycle began to erode (Muller and MacDonald, 1997) as there is a step-like increase in the regression coefficient of the 100 kyr sinusoid. This gradual transition from the '41 kyr world' to the '100 kyr world' is complete by 0.79 Ma [95% credibility limit 780–792 ka], a date commonly associated with the onset of 100 kyr variability (e.g. Mudelsee and Schulz 1997; Tziperman and Gildor, 2003; among others) and which roughly coincides with the Brunhes-Matuyama magnetic reversal [0.78 Ma]. However, the obliquity signal does not disappear from the proxy record at this time; it merely becomes overshadowed by a more dominant 100 kyr signal. More recently, change points were identified near the Mid-Brunhes event (430 ka), where there was a transition from weak to strong interglacials (EPICA, 2004) (∼470 ka, 95% credibility limit 453–487 ka) and near Marine Isotope Stage (MIS) 11 (380 ka, 95% credibility limit 375–383 ka), whose unusually long and warm interglacial during a time of low eccentricity poses a problem for Milankovitch Theory (Muller and MacDonald, 1995). The final two change points are located near MIS6 (∼185 ka, 95% credibility limit 168–224 ka) and MIS4 (∼71 ka, 95% credibility limit 66–74 ka). The multiple change points in the most recent part of the proxy record likely indicate that a single 100 kyr signal is insufficient to model

Figure 4. Regression Coefficients for the $\delta^{18}O$ Proxy Record of the Plio-Pleistocene. Sinusoidal functions approximating the three parameters of the Earth's orbital motion were used to model the $\delta^{18}O$ proxy record. The transition from the '41 kyr world' to the '100 kyr world' begins around 1.5 Ma with a gradual, step-like increase in the regression coefficient for the 100 kyr sinusoid and is complete by around 0.79 Ma. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

the glacial cycles during this time. The exact mechanism behind 100 kyr glacial cycles has been the subject of much debate (Tziperman and Gildor, 2003; Raymo *et al.*, 2006; Tziperman *et al.*, 2006; among others).

5. Discussion

An important contribution of the Bayesian Change Point algorithm is its ability to objectively assess the uncertainty surrounding the 'optimal' solution, both in the number and location of change points, a significant advantage over a frequentist solution. Thus, the novelty of the analyses is not the detection of new change points, but to provide a measure of uncertainty. The fact that the credibility limits encompass these previously detected change points helps to illustrate the accuracy and validity of the Bayesian Change Point method described in this paper. A downside to Bayesian approaches is often the computational complexity. However, the recursive nature of this algorithm allows the $\sim N^k$ calculations to be completed in a computationally efficient manner, $O(N^2)$. While the focus here was on a regression model, the Bayesian Change Point algorithm is applicable to a wide range of probabilistic models. For example, Liu and Lawrence (1999) use a multinomial distribution in the context of the Bayesian Change Point algorithm to model the frequency of nucleotides in a DNA sequence.

The linear model used to describe the global surface temperature time series is similar to the 'sloped steps' model described by Seidel and Lanzante (2004) in which there is no continuity constraint enforced between adjacent trend lines. A continuity constraint, (Karl *et al.*, 2000; Tomé and Miranda, 2004) can sometimes be a limitation as it does not allow abrupt changes in a climate record due to events such as volcanism (Menne, 2006; Ivanov and Evtimov, 2009).

The functions used to model both the global surface temperature and the $\delta^{18}O$ proxy record were almost certainly simplifications of the true climate phenomena. A more complex model would likely yield a better fit to both data sets, but the relative simplicity of these models gives a more straightforward physical interpretation.

The number of change points identified by the Bayesian Change Point algorithm can be sensitive to the choice of priors. Both the global surface temperature record and the $\delta^{18}O$ proxy record are sensitive to the priors on the variance, v_0 and σ_0^2, although this is not without precedence (See Appendix). On the other hand, neither of these time series are sensitive to alternative prior distributions on the number of change points (Equation 6), such as $f(K = k) = \frac{1}{k_{max} + 1}, k = 0, 1, \ldots, k_{max}$, although this may not always be the case. The proper choice of priors will depend on the nature of the time series being analysed.

Non-independence of residual errors is often a concern in time series analysis as it violates one of the assumptions of any linear regression model (Equation (1)).

The equations given in this paper are not applicable to autocorrelated residuals, although a generalization to include autoregressive models while non-linear, appears plausible.

Both the 5 Myr $\delta^{18}O$ proxy record of the Plio-Pleistocene (Lisiecki and Raymo, 2005) and the much shorter 130 year NOAA/NCDC annual global surface temperature anomalies time series show periods of markedly different climate activity. Given distinct climatic periods in these and other climate records, it is important to be able to break climatic time series into regimes and separately study each regime in order to better understand why and how our climate system has changed throughout history.

Matlab code to implement the Bayesian Change Point algorithm is available upon request from the author.

Appendix: Parameter Settings for the Bayesian Change Point Algorithm

There are five parameters that need to be set by the user in order to run the Bayesian Change Point algorithm.

1. k_{max} – *The maximum number of change points allowed.*
2. d_{min} – *The minimum distance between consecutive change points.* If a minimum distance is desried, it can be enforced in Step 1 of the Bayesian Change Point algorithm [calculating the Probability Density of the Data, $f(Y_{i:j}|X)$] by setting $f(Y_{i:j}|X) = 0$ for indices i and j such that $(j - i) < d_{min}$. As a general guideline, we recommended that the minimum distance between consecutive change points be at least twice as many data points as free parameters in the regression model. This helps to ensure that enough data is available to estimate the parameters of the model accurately.
3. k_0 – *The variance scaling hyperparameter for the multivariate Normal prior on β. $\beta|\sigma^2 \sim N(0, \sigma^2/k_0)$* implies that the variance of the regression coefficients is related to the residual variance, σ^2. A model that fits the data well will have a small residual variance, but may have large (relative to σ^2) regression coefficients. Therefore, we set k_0 to be small, yielding a wide prior distribution on the regression coefficients. For both the NOAA/NCDC global surface temperature anomaly time series and the $\delta^{18}O$ proxy record of the Plio-Pleistocene, k_0 was set to 0.01.
4. v_0 and σ_0^2 – *The hyperparameters for the Scaled-Inverse χ^2 prior on σ^2. v_0 and σ_0^2 act as pseudo data points – v_0 pseudo data point of variance σ_0^2* – that help to bound the likelihood function. The posterior

distribution on the number of change points can be sensitive to the choice of parameters for the prior distribution of the residual variance, v_0 and σ_0^2. Specifically, the number of change points, but not the distribution of their positions, can vary with the choice of these parameters, a phenomena previously noted by Fearnhead (2006). In the least squares/maximum likelihood setting, these parameters are equivalent to penalized regression (Silverman, 1985; Ciuperca *et al.*, 2003; Caussinus and Mestre, 2004) techniques. In general, the larger the product of these two parameters, the fewer the number of change points chosen by the algorithm as this product is essentially a penalty paid for introducing a new change point into the model.

As we can be sure that the variance of the residuals will not be larger than the overall variance of the data set, one option is to conservatively set the prior variance σ_0^2, equal to the variance of the data set being used and set v_0 to be <25% of the size of the minimum allowed sub-interval. For the simulation and the NOAA/NCDC global surface temperature anomaly time series, v_0 was chosen to be 1 and σ_0^2 as 0.05, while for the analysis of the $\delta^{18}O$ proxy record of the Plio-Pleistocence, v_0 was chosen to be 10 and σ_0^2 as 0.30. Larger numbers of change points can result from setting strongly informed prior distributions. A choice along these lines acts as if there were more data, and thus the algorithm is able to pick up more subtle changes.

References

Aksoy H, Gedikli A, Unal NE, Kehagias A. 2008. Fast segmentation algorithms for long hydrometeorological time series. *Hydrological Processes* **22**(23): 4600–4608.

Barry D, Hartigan JA. 1993. A Bayesian Analysis for Change Point Problems. *Journal of American Statistical Association* **88**(421): 309–319.

Beaulieu C, Ouarda T, Seidou O. 2010. A Bayesian normal homogeneity test for the detection of artificial discontinuities in climate series. *International Journal of Climatology* **30**: 2342–2357, DOI: 10.1002/joc.2056.

Caussinus H, Mestre O. 2004. Detection and correction of artificial shifts in climate series. *Journal of Applied Statistics* **53**(3): 405–425.

Ciuperca G, Ridolfi A, Idier J. 2003. Penalized maximum likelihood estimator for normal mixtures. *Scandinavian Journal of Statistics* **30**: 45–59.

EPICA Community Members. 2004. Eight glacial cycles from an Antarctic core. *Nature* **429**: 623–628.

Fearnhead P. 2006. Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing* **16**: 203–213, DOI: 10.1007/s11222-006-8450-8.

Hannart A, Naveau P. 2009. Bayesian multiple change points and segmentation: Application to homogenization of climatic series. *Water Resources Research* **45**: W10444.

Hays JD, Imbrie J, Shackleton NJ. 1976. Variations in the Earth's orbit: pacemakers of the ice ages. *Science* **194**: 1121–1132.

Imbrie J, Boyle EA, Clemens SC, Duffy A, Howard WR, Kukla G, Kutzbach J, Martinson DG, McIntyre A, Mix AC, Molfino B, Morley JJ, Peterson LC, Pisias NG, Prell WL, Raymo ME, Shackleton NJ, Toggweiler JR. 1992. On the structure of major glaciation cycles: 1. Linear responses to Milankovitch forcing. *Paleoceanography* **7**: 701–738.

Ivanov MA, Evtimov SN. 2009. 1963: The break point of the Northern Hemisphere temperature trend during the twentieth century. *International Journal of Climatology* **30**(11): 1738–1746, DOI: 10.1002/joc.2002

Karl TR, Knight RW, Baker B. 2000. The record breaking global temperatures of 1997 and 1998: Evidence for an increase in the rate of global warming? *Geophysical Research Letters* **27**(5): 719–722.

Khaliq MN, Ouarda TBMJ, St-Hilaire A, Gachon P. 2007. Bayesian change-point analysis of heat spell occurrences in Montreal, Canada. *International Journal of Climatology* **27**: 805–818, DOI: 10.1002/joc.1432.

Lanzante J. 1996. Resistant robust and nonparametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology* **16**: 1197–1226.

Lavielle M, Lebarbier E. 2001. An application of MCMC methods for the multiple change-points problem. *Signal Processing* **81**: 39–53.

Lisiecki LE, Raymo ME. 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* **20**: PA1003, DOI: 10.1029/2004 PA001071.

Lui JS, Lawrence CE. 1999. Bayesian inference on biopolymer models. *Bioinformatics* **15**(1): 38–52.

Lund R, Wang XL, Lu Q, Reeves J, Gallagher C, Feng Y. 2007. Changepoint detection in periodic and autocorrelated time series. *Journal of Climate* **20**: 5178–5190, DOI: 10.1175/JCLI4291.1

Menne MJ. 2006. Abrupt global temperature change and the instrumental record. *18th Conference on Climate Variability and Change*, AMS: Atlanta, GA.

Milankovitch M. 1941. Canon of Insolation and the Ice-Age Problem. Israel Program for Scientific Translations, Jerusalem 1969.

Mudelsee M, Schulz M. 1997. The mid-Pleistocene climate transition: onset of 100 ka cycle lags ice volume build-up by 280 ka. *Earth and Planetary Science Letters* **151**: 117–123.

Muller, RA, MacDonald GJ. 1995. Glacial Cycles and Orbital Inclination. *Nature* **377**: 107–108.

Muller RA, MacDonald GJ. 1997. Glacial cycles and astronomical forcing. *Science* **277**: 215–218.

Quayle RG, Peterson TC, Basist AN, Godfrey CS. 1999. An operational near-real-time global temperature index. *Geophysical Research Letters* **26**: 333–335.

Raymo ME, Ruddiman WF, Shackleton NJ, Oppo DW. 1990. Evolution of the Atlantic-Pacific $\delta^{13}C$ gradients over the last 2.5 m.y. *Earth and Planetary Science Letters* **97**: 353–368.

Raymo ME, Lisiecki LE, Nisancioglu KH. 2006. Plio-pleistocene ice volume, antarctic climate, and the global d18O record. *Science* **313**: 492–495.

Ruddiman WF. 2007. *Earth's Climate: Past and Future*, WH Freeman: New York.

Ruggieri E, Herbert T, Lawrence KT, Lawrence CE. 2009. Change point method for detecting regime shifts in paleoclimatic time series: application to $\delta^{18}O$ time series of the Plio-Pleistocene. *Paleoceanography* **24**: PA1204, DOI: 10.1029/2007 PA001568.

Seidel DJ, Lanzante JR. 2004. An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes. *Journal of Geophysical Research* **109**: D14108, DOI: 10.1029/2003JD004414

Seidou O, Ouarda TBMJ. 2007. Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows. *Water Resources Research* **43**: W07404.

Silverman BW. 1985. Penalized maximum likelihood estimation. *Encyclopedia of Statistical Sciences* **6**: 664–667.

Stephens DA. 1994. Bayesian retrospective multiple-changepoint Identification. *Journal of Applied Statistics* **43**(1): 159–178.

Tomé AR, Miranda PMA. 2004. Piecewise linear fitting and trend changing points of climate parameters. *Geophysical Research Letters* **31**: L02207, DOI: 10.1029/2003GL019100.

Trenberth KE, Hurrell JW. 1994. Decadal atmosphere-ocean variations in the Pacific. *Climate Dynamics* **9**: 303–319.

Tziperman E, Gildor H. 2003. On the mid-Pleistocene transition to 100-kyr glacial cycles and the asymmetry between glaciation and deglaciation times. *Paleoceanography* **18**(1): PA1001, DOI: 10.1029/2001 PA000627.

Tziperman E, Raymo ME, Huybers P, Wunsch C. 2006. Consequences of pacing the Pleistocene 100 kyr ice ages by nonlinear phase locking to Milankovitch forcing. *Paleoceanography* **21**: PA4206, DOI: 10.1029/2005 PA001241.

Zhao X, Chu P. 2006. Bayesian multiple changepoint analysis of hurricane activity in the eastern north pacific: a Markov chain Monte Carlo approach. *Journal of Climate* **19**: 564–578.