RESEARCH ARTICLE

Bayesian change-point modeling with segmented ARMA model

Farhana Sadia[1], Sarah Boyd[2], Jonathan M. Keith [ID][1] *

1 School of Mathematical Sciences, Monash University, Clayton, VIC, Australia, **2** Faculty of Information Technology, Monash University, Clayton, VIC, Australia

* jonathan.keith@monash.edu

Abstract

Time series segmentation aims to identify segment boundary points in a time series, and to determine the dynamical properties corresponding to each segment. To segment time series data, this article presents a Bayesian change-point model in which the data within segments follows an autoregressive moving average (ARMA) model. A prior distribution is defined for the number of change-points, their positions, segment means and error terms. To quantify uncertainty about the location of change-points, the resulting posterior probability distributions are sampled using the Generalized Gibbs sampler Markov chain Monte Carlo technique. This methodology is illustrated by applying it to simulated data and to real data known as the well-log time series data. This well-log data records the measurements of nuclear magnetic response of underground rocks during the drilling of a well. Our approach has high sensitivity, and detects a larger number of change-points than have been identified by comparable methods in the existing literature.

Introduction

A time series is a succession of measurements made over a time interval. Some time series can be divided into a sequence of individual segments, each with its own unique characteristic properties. Identifying segment boundaries and inferring dynamical properties of different segments is referred to as time series segmentation. Change-point detection methods can be used to segment time series data since the goal of such methods is to select a sequence of change-point locations such that the observations are, in some sense, homogeneous within segments and different between segments [1].

Statistical analysis of change-point problems has been the subject of intensive research in the past half-century and there has been a large amount of literature on this subject (for reviews, see [2–5]). Literature on Bayesian change-point modeling is as extensive as for the classical change-point model. The Bayesian change-point model was pioneered by Chernoff and Zacks [6], who estimated the mean of a normal distribution for each segment in a Bayesian framework. Smith [7] proposed a Bayesian change-point model for finite series with normal and binomial models. To detect change-points in multivariate time series, Harlé *et al.* [8] used a Bayesian approach where change-points are modeled using Bernoulli variables for the

change-point indicator vector. Successive generalizations and extensions of Bayesian methods for change-point problems include [9–13] and many others.

Different authors propose diverse models for each segment. For example, Punskaya *et al.* [14] suggested a Bayesian method for fitting piecewise linear regression models such as auto-regressive models. To explain regime switching behaviour in the conditional mean, Chan and Tong [15] proposed a new class of non-linear models, called the Smooth Transition Autoregressive (STAR) models. This model is an extension of the Threshold Autoregressive (TAR) model, introduced by Tong and Lim [16–18]. The TAR model is a piecewise linear model consisting of two or more linear sub-models. An indicator variable is used in the TAR model which represents a switch from one regime to another and takes a value zero or one, depending upon the values of a transition variable and a threshold parameter. This indicator variable implies abrupt jumps from one regime to the next. Chan and Tong [15] suggested the replacement of the indicator function with a smooth transition function in their STAR model, since sudden jumps from one regime to another may not be the best representation of the underlying mechanism generating observed data. Davis *et al.* [19] estimated structural breaks in a nonstationary time series by segmenting the series into blocks of distinct autore-gressive (AR) processes. They assumed the number of breakpoints, their locations and the orders of the respective AR models are unknown. To segment non-stationary time series data, Wood *et al.* [20] developed a Bayesian mixture of autoregressive models where compo-nents of the model are time series and mixture probabilities. The time series have constant but unknown parameters and those mixture probabilities depend on time. They assumed unknown lag of the AR processes as well as an unknown number of components in the mix-ture model. To estimate the number and locations of multiple change-points in the mean of a Gaussian AR(1) process, Chakar *et al.* [21] proposed a new approach where the unknown autocorrelation coefficient and the variance of an "innovation" term remain unchanged from one segment to the other. They firstly estimated the autocorrelation coefficient and then decorrelate the series. After that they applied a dynamic programming algorithm to the decorrelated series.

As noted above, some existing models for time series segmentation have used a segmented AR model [19, 22]. This paper proposes to segment time series data using a segmented ARMA model, an approach that is surprisingly absent from the existing literature. Results for the well-log time series data discussed below show that fitting an ARMA model in each segment poten-tially identifies a greater number of change-points than the AR model in both real and simu-lated data and thus provides higher sensitivity. Moreover, we find that this is true for both large and small values of the variance σ^2 of the innovation term. Detailed explanations and comparison with AR model are provided in S3 Appendix.

The Bayesian segmented ARMA change-point model presented here resembles the approach of Keith *et al.* [23]. The major modelling innovation here is the use of an ARMA model in all segments. The posterior marginal distribution of this model are difficult to analyse directly since they have nonstandard form. However, simulated sampling can be performed via a Markov chain Monte Carlo (MCMC) algorithm. Several MCMC algorithms are available in the literature including the Metropolis-Hastings algorithm [24, 25], Gibbs Sampler [26], the Reversible Jump MCMC algorithm [9]; Multiple-Try Metropolis algorithm [27] and Delayed Rejection Metropolis-Hastings algorithm [28]. The Bayesian segmented ARMA change-point model here uses a highly efficient sampling technique known as the Generalized Gibbs Sam-pler (GGS) for generating samples from a posterior distribution [29]. The dimension in this algorithm does not need to be fixed and it provides flexibility to sample from varying dimen-sional spaces. The Generalized Gibbs Sampler (GGS) has been applied to some very high dimensional problems (see [30, 31]). It has resulted in highly efficient sampling for these

problems. However, no systematic comparison of the advantages and disadvantages of the GGS versus the reversible jump sampler has been performed.

Methodology

Problem statement

We consider the problem of modeling a time series by segmenting the series into blocks of autoregressive moving average (ARMA) processes. Let $t = 1, \cdots, T$ be time points in the signal or time series, where T represents the total length of the signal. Let x_t represent a real valued signal at time point, t. Let, $\mathbf{X} = (x_t)_{t=1}^{T}$ represent the time series vector or the signal that we want to segment. The ARMA model is:

$$x_t = c + \epsilon_t + \sum_{i=1}^{a} \psi_i (x_{t-i} - c) + \sum_{i=1}^{m} \theta_i \epsilon_{t-i}.$$

where $\psi_1, \cdots, \psi_a$ and $\theta_1, \cdots, \theta_m$ are the parameters of the AR and MA sub-models, respectively; a and m denote the order of the AR and MA sub-models; c is the mean of the ARMA model; ϵ is white noise and x_t is the time series. The number of change-points and their locations are assumed unknown. Each segment in the time-series is assumed to be generated by an ARMA model with different mean, and the goal is to infer the most probable segment locations and model parameters that describe them.

Likelihood model

We start by writing down the likelihood function of a model in which the sequence within each segment is generated by an ARMA process. This determines the probability of generating the observed sequence for any given parameter values. For each position in the signal except the first, the probability of starting a new segment at that position is denoted by ϕ. Thus a time series with K segments that have starting positions $\mathbf{s} = (1 = s_1 < \cdots < s_K \leq T)$ is generated with probability:

$$p(K, \mathbf{s}|\phi) = \phi^{K-1}(1 - \phi)^{T-K}. \tag{1}$$

Here, $s_1 = 1$, indicating that the first segment always starts at the beginning of the signal. Let the right hand points of the segments be $\mathbf{d} = (d_1, \cdots, d_K)$ where $d_K = T$ so that the last segment always finishes at the end of the signal. Let X_k be the signal of the segment between positions s_k and d_k inclusive. Each segment is then assigned to one of N groups with probabilities $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_N)$ where π_n is the probability of assigning any segment to group n. We denote the group to which segment k is assigned by $g_k \in \{1, \cdots, N\}$ where $\mathbf{g} = (g_1, \cdots, g_K)$. Then the probability of a specific assignment of the K segments to the N groups is

$$p(\mathbf{g}|K, \boldsymbol{\pi}) = \prod_{k=1}^{K} \boldsymbol{\pi}_{g_k}. \tag{2}$$

Let b_n be the number of segments with $g_k = n$. The probability of a specific assignment of the K segments to the N groups can be alternatively defined as:

$$p(\mathbf{g}|K, \boldsymbol{\pi}) = \prod_{n=1}^{N} \boldsymbol{\pi}_n^{b_n}. \tag{3}$$

Each segment is then modeled by an ARMA model. We write the ARMA model in each segment as:

$$x_t = c_k + \epsilon_t + \sum_{i=1}^{a} \psi_i (x_{t-i} - c_k) + \sum_{i=1}^{m} \theta_i \epsilon_{t-i}. \tag{4}$$

Here, c_k is the mean signal level for segment k and $c_k \sim \mathcal{N}(\mu_{g_k}, \tau_{g_k}^2)$, where μ_{g_k} and $\tau_{g_k}^2$ are the mean and variance of the distribution of these means for the group g_k. Also, $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_T)$ is the vector of error terms and $\epsilon_t = x_t - \lambda_t$, where $\lambda_t = c_k + \sum_{i=1}^{a} \psi_i (x_{t-i} - c_k) + \sum_{i=1}^{m} \theta_i \epsilon_{t-i}$.

We suppose that $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$, where σ^2 is the variance of error terms and this applies for $t \in (s_k, \cdots, d_k)$. In our model only these segment means differ between segments; other parameters of the ARMA model, ψ and θ, are the same for all segments. This assumption is appropriate for the applications used in this paper but other data sets may need to allow all of these parameters to differ between segments. In this paper, when we define the order of the AR (a) and MA (m) submodels we only intend for a and m to take values 0 or 1. Note that when $t - s_k$ is less than a or m, the expression for λ_t includes terms from the previous segment. However, for the left end of the signal when we can not look back a or m order steps, λ_t includes only the c_k term. An alternative is to choose the initial values for each segment by initializing them from the stationary distribution for the ARMA process in that segment.

The probability density of the observed signal is a product over all segments, that is, the probability of the signal $\mathbf{X}$ conditioning on parameters K, $\mathbf{s}$, θ, ψ, $\mathbf{c}$ and σ^2 is expressed as a product of normal distributions with mean λ_t and variance σ^2 as follows:

$$
\begin{aligned}
p(\mathbf{X}|K, \mathbf{s}, \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{c}, \sigma^2) &= \prod_{t=1}^{T} p(x_t | K, s, \theta, \psi, c, \sigma^2, x_{<t}) \\
&= \prod_{t=1}^{T} \mathcal{N}(x_t | \lambda_t, \sigma^2).
\end{aligned} \tag{5}
$$

Here, $x_{<t}$ indicates the signal value at time points $1, 2, \cdots, t-1$. The joint distribution of $\mathbf{X}$, K, $\mathbf{s}$, $\mathbf{g}$ and $\mathbf{c}$ conditional on the other parameters is given by:

$$
\begin{aligned}
p(\mathbf{X}, K, \mathbf{s}, \mathbf{g}, \mathbf{c}|\phi, \boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\psi}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\tau}) &= \\
p(\mathbf{X}|K, \mathbf{s}, \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{c}, \sigma^2) &\times \\
p(K, \mathbf{s}|\phi) &\times \\
p(\mathbf{c}|\mathbf{g}, \boldsymbol{\mu}, \boldsymbol{\tau}) &\times \\
p(\mathbf{g}|K, \boldsymbol{\pi}) &
\end{aligned} \tag{6}
$$

Here, $p(\mathbf{c}|\mathbf{g}, \boldsymbol{\mu}, \boldsymbol{\tau})$ is the probability of the ARMA mean for all segments given by:

$$p(\mathbf{c}|\mathbf{g}, \boldsymbol{\mu}, \boldsymbol{\tau}) = \prod_{k=1}^{K} \mathcal{N}(c_k | \mu_{g_k}, \tau_{g_k}^2). \tag{7}$$

Fig 1 shows the parameters of this model and their conditional dependencies. A parameter at the head of an arrow is conditionally dependent on the parameter at the tail.

Prior distribution

Since the segmented ARMA model is presented here in a hierarchical Bayesian framework, we have to assign prior distributions for the unspecified parameters. A beta prior is assigned for ϕ
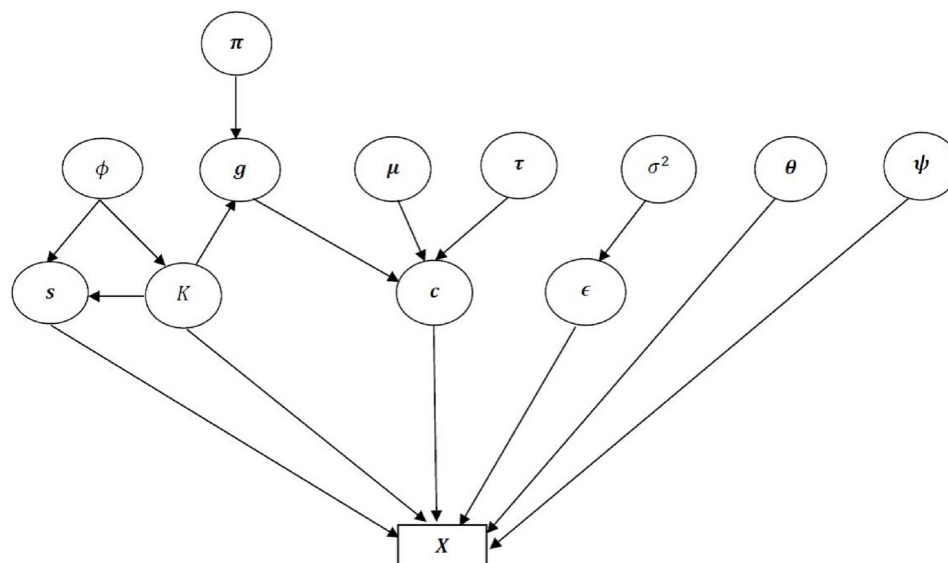
Fig 1. The conditional dependencies of the parameter.

https://doi.org/10.1371/journal.pone.0208927.g001

with parameters $a = 1.0$ and $b = 1.0$. Our computational algorithm is not sensitive to the choice of prior for ϕ, as we show in S4 Appendix. We assign a Dirichlet distribution to $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_N)$ with parameters $(\alpha_1, \cdots, \alpha_N) = (1.0, \cdots, 1.0)$ and $\sum_{n=1}^{N} \pi_N = 1$. Inferences are performed here using a weakly informative normal prior with mean 0.0 and variance 1.0 for the mean μ_n of the distribution of ARMA means in segment class N. We also assign an inverse gamma prior distribution with parameters $\alpha = 3.0$ and $\beta = 3.0$ for the variance (τ_n^2) of the distribution of $\mathbf{c}$ and the variance of the error terms (σ^2) of the ARMA model. The order of the AR model and the MA model will be considered fixed. Since we have no strong prior beliefs about the parameters of the ARMA model, we choose a uniform prior distribution for these on the interval $(-1,1)$ and assume they are independent of each other. The forms of these hyper-priors were chosen to reflect the degree of prior belief about their respective parameters. Note that in general the signal x_t can be shifted and scaled so that the above prior is appropriate.

Posterior distribution

Using Bayes' theorem the posterior distribution of parameters is:

$$
\begin{aligned}
p(K, \mathbf{s}, \mathbf{g}, \mathbf{c}, \phi, \boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\psi}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\tau} | \mathbf{X}) \\
\propto p(\mathbf{X}, K, \mathbf{s}, \mathbf{g}, \mathbf{c} | \phi, \boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\psi}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\tau}) p(\phi) p(\boldsymbol{\pi}) p(\boldsymbol{\theta}) p(\boldsymbol{\psi}) p(\sigma^2) p(\boldsymbol{\mu}) p(\boldsymbol{\tau})
\end{aligned}
\tag{8}
$$

See S1 Appendix for details of the calculation of the conditional posterior distribution of each parameter.

Sampling

The posterior distribution obtained in S1 Appendix is sampled using a Markov chain Monte Carlo technique known as the Generalized Gibbs Sampler, or GGS [29]. The GGS algorithm cycles through a sequence of steps in which parts of a sampled element are updated, while other parts are held constant. These different types of update are analogous to the coordinate updates of the conventional Gibbs sampler and are known as "move-types". This technique resembles a

conventional Gibbs sampler but can be applied in a transdimensional setting, where it provides an alternative to the reversible jump sampler. This technique allows the number of change-points to vary: it cycles through segments inserting and deleting change-points, and shifting change-point positions. See S2 Appendix for details of the GGS algorithm. The three main stages of this algorithm in Bayesian change-point segmented ARMA model are: (also see Fig 2)

- Iterate through the segments doing insertion and deletion updates, segment group assignments (g_k) and segment mean updates (c_k).

- Iterate through the groups updating group parameters (group mean μ_g and group variance τ_g^2).

- Update all the other parameters ($\boldsymbol{\pi}, \sigma^2, \phi, \boldsymbol{\theta}$ and $\boldsymbol{\psi}$).

Move types. Fig 2 illustrates the following defined move-types:

- (I, k): Decide whether to insert a new change-point in segment k, and at what position.

- (D, k): Decide whether to remove change-point k or move it to a new position (for each change-point except the first).

- c_k: Update mean signal level c_k in segment $k = 1, 2, 3, \cdots, K$.

- g_k: Update segment group assignments g_k in segment $k = 1, 2, 3, \cdots, K$.

- μ_g: Update group mean μ_g for group g.

- τ_g^2: Update group variance τ_g^2 for group g.

- $(\boldsymbol{\theta}, \boldsymbol{\psi})$: Update $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$.

- $(\boldsymbol{\pi}, \sigma^2, \phi)$: Update all other parameters, $\boldsymbol{\pi}, \sigma^2$ and ϕ.

There are K I-moves, $K - 1$ D-moves, K moves for updating segment group assignments, K moves for updating mean signal level, N moves for updating group mean, N moves for updating group variance, a moves to update parameters of the AR model, m moves to update parameters of the MA model and finally three moves for updating other parameters $\boldsymbol{\pi}, \sigma^2$ and ϕ. The total number of moves for a sequence with K segments is:

$$T(K) = 4K - 1 + 2N + a + m + 3. \qquad (9)$$

where N is the number of groups, a is the order of the AR model and m is the order of the MA model.

Insertion: Step (I, k). Each move-type of a GGS sampler involves drawing from a distribution over a subset of the target space. The insertion and deletion move-types mentioned in Fig 2 both involve selecting an element from a set containing only two elements. In this respect, these move-types resemble a Metropolis-Hastings update, in which a new element is proposed and either accepted or rejected with some probability. In fact the insertion and deletion move-types involve the same subsets: they differ only in which element of the set is regarded as the source or current element, and which is regarded as the proposed element. The two elements in these subsets each have non-zero probability of being selected, proportional to the target distribution multiplied by the probability of selecting the move type when that element is the current element (see [29] for details).

In an (I, k) insertion step (where $k \in \{1, \cdots, K\}$) the subset contains the current segmentation and a new segmentation with a change-point inserted somewhere in segment k. A new segment end-point position z is proposed between s_k and $d_k - 1$, inclusive. The location of z is selected from a uniform distribution over the set $\{s_k, \cdots, d_k - 1\}$. Then for the left segment
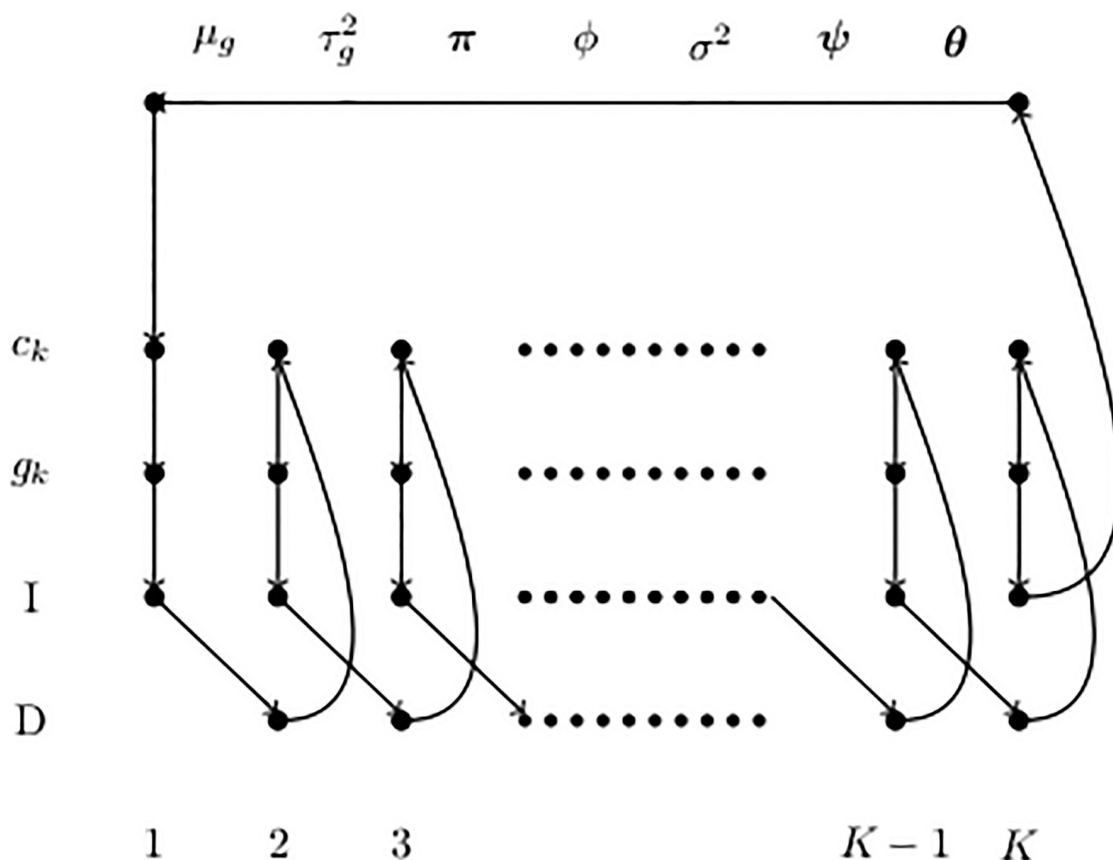
Fig 2. Order of move-types for the sampler. Note that I updates run from 1 to K whereas D updates run from 2 to K.

(from s_k to z) new values of g'_k and c'_k are proposed and for the right segment (from $z + 1$ to d_k) new values of g''_{k+1} and c''_{k+1} are proposed. Here, g'_k and g''_{k+1} are selected from a discrete distribution with parameters $\pi_1, \ldots, \pi_N$ and then c'_k and c''_{k+1} are selected from normal distributions with parameters $(\mu_{g'_k}, \tau^2_{g'_k})$ and $(\mu_{g''_{k+1}}, \tau^2_{g''_{k+1}})$ respectively. Choosing g'_k, g''_{k+1}, c'_k and c''_{k+1} in this manner results in cancellations such that the terms $p(\mathbf{c}|\mathbf{g}, \boldsymbol{\mu}, \boldsymbol{\tau})$ and $p(\mathbf{g}|K, \boldsymbol{\pi})$ in Eq 6 disappear when calculating the acceptance ratio.

After further cancellations, the new change-point at position $z + 1$ is rejected with a probability proportional to

$$P_1 = (1 - \phi)\prod_{t=s_k}^{d_k} p(\epsilon_t|0, \sigma^2)\frac{1}{(d_k - s_k)}\frac{1}{T(K)} \tag{10}$$

where

$$p(\epsilon_t, 0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{\epsilon_t^2}{2\sigma^2}\right]$$

and

$$\epsilon_t = x_t - \lambda_t = x_t - \left(c_k + \sum_{i=1}^{a}\psi_i(x_{t-i} - c_k) + \sum_{i=1}^{m}\theta_i\epsilon_{t-i}\right).$$

In this expression, $1/(d_k - s_k)$ is the probability of proposing the location of the new change-point and $1/T(K)$ is a correction factor used by the GGS to account for the number of move types available for the current segmentation with K segments.

Alternatively, the new change-point is accepted with probability proportional to:

$$P_0 = \phi \prod_{t=s_k}^{z} p(\epsilon_t'|0,\sigma^2) \times \prod_{t=z+1}^{d_k} p(\epsilon_t''|0,\sigma^2) \frac{1}{T(K+1)} \tag{11}$$

where

$$p(\epsilon_t'|0,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\epsilon_t'^2}{2\sigma^2}\right]$$

$$p(\epsilon_t''|0,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\epsilon_t''^2}{2\sigma^2}\right] \tag{12}$$

$$\epsilon_t' = x_t - \lambda_t' = x_t - \left(c_k' + \sum_{i=1}^{a}\psi_i(x_{t-i} - c_k') + \sum_{i=1}^{m}\theta_i\epsilon_{t-i}'\right)$$

and

$$\epsilon_t'' = x_t - \lambda_t'' = x_t - \left(c_k'' + \sum_{i=1}^{a}\psi_i(x_{t-i} - c_k'') + \sum_{i=1}^{m}\theta_i\epsilon_{t-i}''\right) \tag{13}$$

Here, Eq 12 applies for the left segment from s_k to z and Eq 13 applies for the right segment from $z + 1$ to d_k.

Thus the new change-point at $z + 1$ is accepted with probability $\frac{P_0}{P_0+P_1}$ or rejected with probability $\frac{P_1}{P_0+P_1}$. An alternative (which we have not implemented) is to use the Metropolis-Hastings acceptance probability $\min\{1, P_0/P_1\}$.

If a change-point is inserted, the move-type is updated to $(D, k+1)$, otherwise it remains (I, k). In either case, the move-type is then further updated as in Fig 2.

Deletion: Step (D, k). For each segment $k = 2, \cdots, K$, a (D, k) deletion step also involves selecting an element from a set containing only two elements: the current segmentation and a new segmentation with the change-point at the left end of segment k removed, so that segments $k - 1$ and k merge to form a new segment. Values of g_k and c_k are then chosen for the new merged segment: g_k is selected from a discrete distribution with parameters $\pi_1, \ldots, \pi_N$ and then c_k is selected from a normal distribution with parameters μ_{g_k} and $\tau_{g_k}^2$ respectively. Other parameters and the positions of other change-points are held constant.

The probability of accepting the deletion is proportional to:

$$P_1 = (1 - \phi) \prod_{t=s_{k-1}}^{d_k} p(\epsilon_t|0,\sigma^2) \frac{1}{(d_k - s_{k-1})} \frac{1}{T(K-1)}$$

where

$$p(\epsilon_t|0,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\epsilon_t^2}{2\sigma^2}\right]$$

and

$$\epsilon_t = x_t - \lambda_t = x_t - \left(c_k + \sum_{i=1}^{a}\psi_i(x_{t-i} - c_k) + \sum_{i=1}^{m}\theta_i\epsilon_{t-i}\right)$$

Here K is the number of segments in the current segmentation, that is, without making the deletion.

The probability of rejecting the deletion is proportional to:

$$P_0 = \phi \prod_{t=s_{k-1}}^{z} p(\epsilon_t'|0, \sigma^2) \times \prod_{t=z+1}^{d_k} p(\epsilon_t''|0, \sigma^2) \frac{1}{T(K)}.$$

Thus the deletion is accepted with probability $\frac{P_1}{P_0+P_1}$ or rejected with probability $\frac{P_0}{P_0+P_1}$. The conditional posterior distribution for $\mathbf{c}$ and $\boldsymbol{\epsilon}$ is proportional to:

$$\prod_{t=s_k}^{d_k} p(x_t|\lambda_t, \sigma^2) \times p(c_k|\mu_{g_k}, \tau_{g_k}^2).$$

The $\mathbf{c}$'s and the corresponding $\epsilon_{s_k}, \cdots, \epsilon_{s_{k+1}}$ can be updated one segment at a time with this conditional posterior distribution. Now, the conditional posterior distribution for $\boldsymbol{\theta}$ and ϵ is proportional to:

$$\prod_{t=1}^{T} p(x_t|\lambda_t, \sigma^2) \times p(\theta).$$

To update $\boldsymbol{\psi}$, similar procedures were used. The sampler iterates through the groups, updating group parameters ($\boldsymbol{\mu}$ and $\boldsymbol{\tau}$) with their respective conditional posterior distributions.

The sampler uses conventional Gibbs updates for other parameters $c_k, g_k, \theta_i, \phi_i, \sigma^2, \boldsymbol{\pi}, \phi, \mu_g$ and τ_g. We use the Slice sampler [32] where needed to draw from non-standard distributions.

Validation of the methodology

We have used a simulation-based method for testing the correctness of software for fitting Bayesian models using posterior simulation [33]. This validation technique is based on posterior quantiles. Consider the general Bayesian joint distribution $p(y|\Theta)p(\Theta)$, where $p(y|\Theta)$ presents the sampling distribution of the data, $p(\Theta)$ presents the proper prior distribution of the parameter vector Θ, and inferences are based on the posterior distribution, $p(\Theta|y)$. The validation method samples a parameter vector $\Theta^{(0)}$ from $p(\Theta)$. Then conditional on $\Theta^{(0)}$, this technique samples data y from $p(y|\Theta = \Theta^{(0)})$ and then simulates sampling from the posterior distribution, $p(\Theta|y)$ using the software to be validated. The resulting posterior sample of size L is denoted $(\Theta^{(1)}, \Theta^{(2)}, \cdots, \Theta^{(L)})$. Finally, for each coordinate of $\Theta^{(0)}$, denoted $\theta^{(0)}$, compute its posterior quantile $\hat{q}(\theta^{(0)})$, with respect to the posterior sample $(\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(L)})$. To perform the validation procedure, many replications are required, each drawing $\Theta^{(0)}$ from $p(\Theta)$ and y from $p(y|\Theta^{(0)})$. The simulation output is a collection of estimated posterior quantiles. From these quantiles calculate a test statistic $X_\theta^2 = \sum_{i=1}^{N_{rep}} \left(\Phi^{-1}(q_i)\right)^2$ where, $q_i = \frac{1}{L}\sum_{l=1}^{L} I_{\theta_i^{(0)} > \theta_i^{(l)}}$ is the posterior quantile for the ith replication, N_{rep} is the total number of replications, θ denotes component of Θ and Φ represents the standard normal CDF. If the software is implemented correctly, this test statistic follows a χ^2 distribution with N_{rep} degrees of freedom and also the posterior quantiles will be uniformly distributed. The posterior quantiles' deviation from uniformity can be quantified by calculating the associated p value, that is, p_θ for each X_θ^2. Extremely small p_θ values indicate an error in the software. As an exploratory tool, p_θ values can be transformed into a z_θ statistic ($z_\theta = \Phi^{-1}(p_\theta)$). If all $|z_\theta|$ statistics are not extreme, such as less than 2, the software may be considered validated.

To validate our method according to the above described technique, we generated the parameters $\tau^2, \sigma^2, \pi, \mu$ from the following prior distributions:

$$\tau^2 \sim Inv - gamma(3, 3)$$
$$\sigma^2 \sim Inv - gamma(3, 3)$$
$$\pi \sim U(0, 1)$$
$$\mu \sim N(0, 1)$$
$$\theta \sim U(-1, 1)$$
$$\psi \sim U(-1, 1)$$

We generated 20 sequences from an ARMA model with 20 different segment means where every sequence has length 100. We simulated 5000 draws from the posterior distribution of the model parameters. Then the quantiles of the posterior distributions for each parameter were determined and the whole procedure was repeated 20 times. From these quantiles we determined the absolute values of the z_θ statistic. The absolute z_θ statistics from this simulation are plotted in Fig 3:

In the above plot, since the z_θ statistic for each parameter is less than 2, we conclude the software is correctly written. More precisely, we find no evidence of software errors.

Bayesian software validation
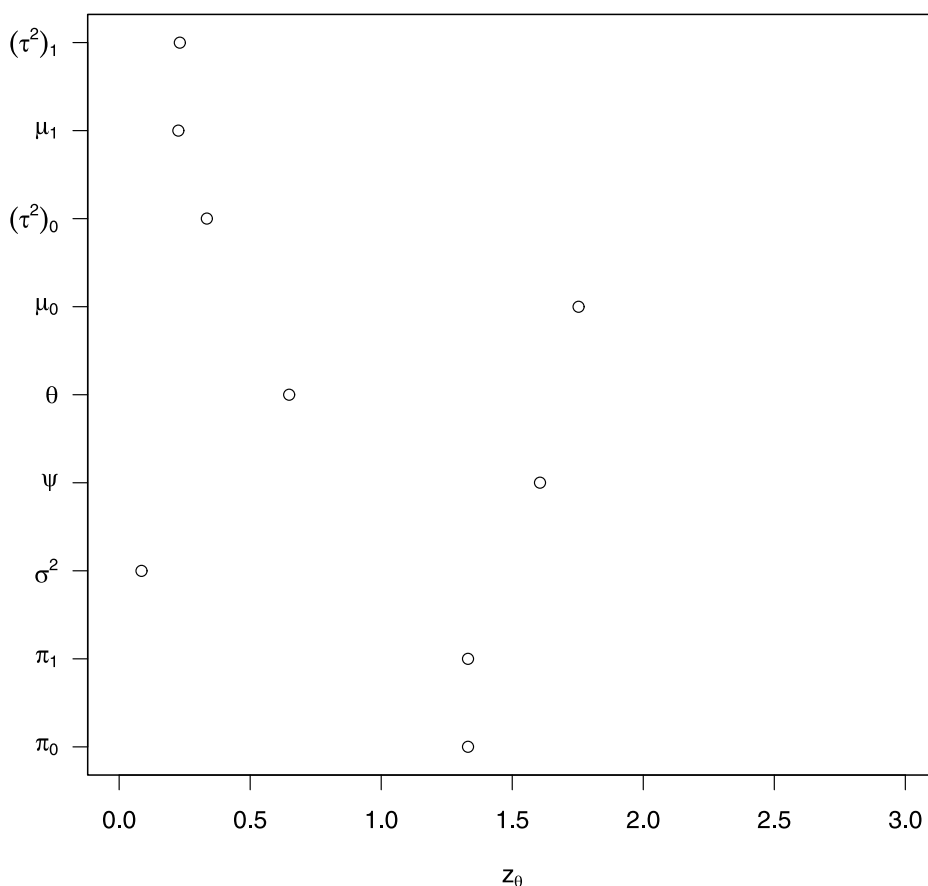
Fig 3. Absolute z_θ statistic plot. Each row shows a parameter of the segmented ARMA model and the $|z_\theta|$ statistics associated with these parameters are displayed as a circle in each row.
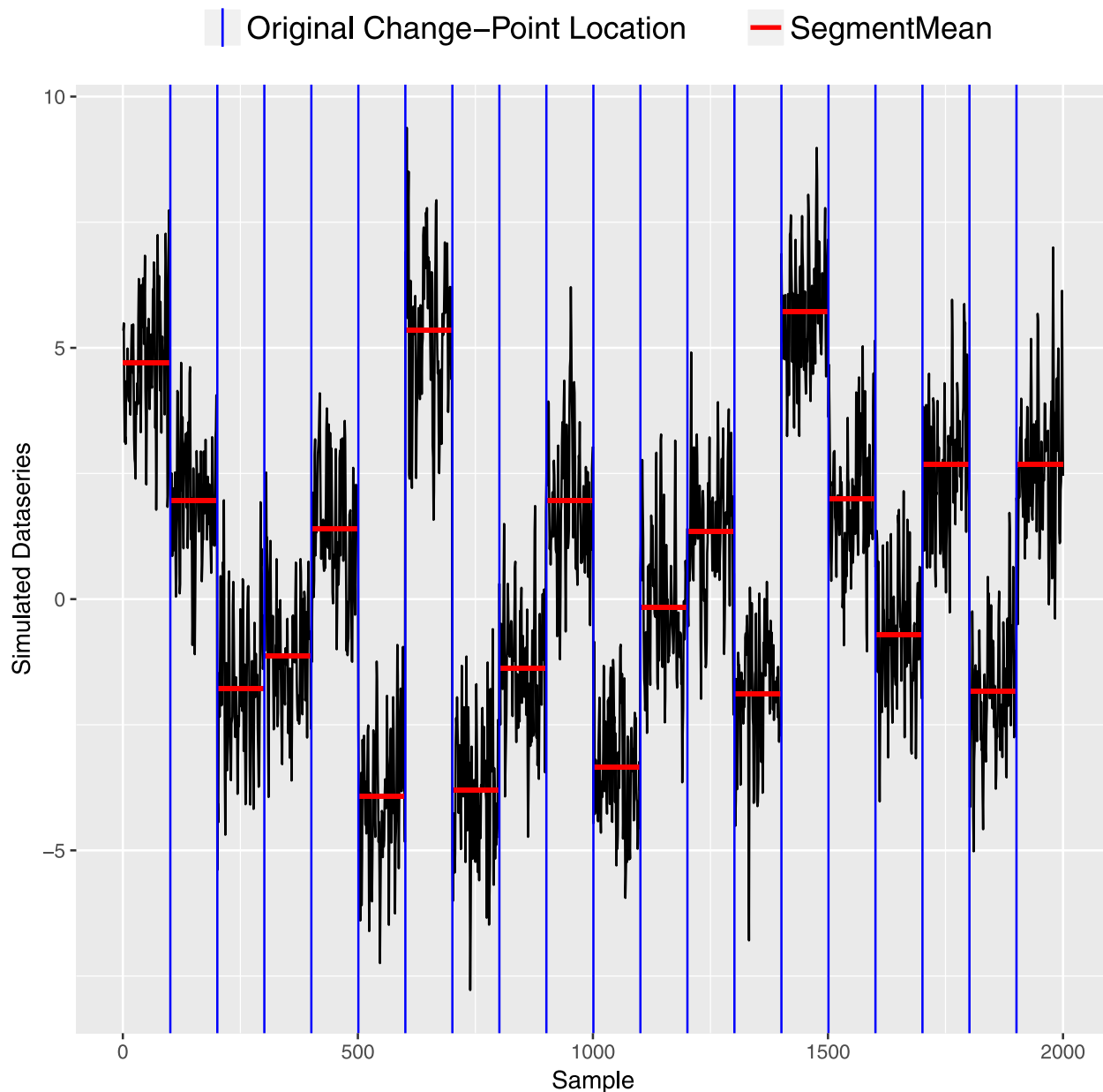
Fig 4. Simulated signal. The true change-point locations are shown as vertical blue lines and segment means are shown as horizontal red lines.

Illustrative examples

Simulation example

As a test of our method, we applied it to a simulated example in which the number and location of change-points, ARMA parameters, segment means and error variance are known. We analyzed 20 time series, each containing 100 observations, generated from the autoregressive moving average (ARMA(1,1)) model with parameter values $\psi = 0.22$ and $\theta = 0.60$. Each series was generated using $\sigma^2 = 0.96$ and 20 different segment means. The simulated ARMA data with the true segment means and the location of change-points is shown in Fig 4.
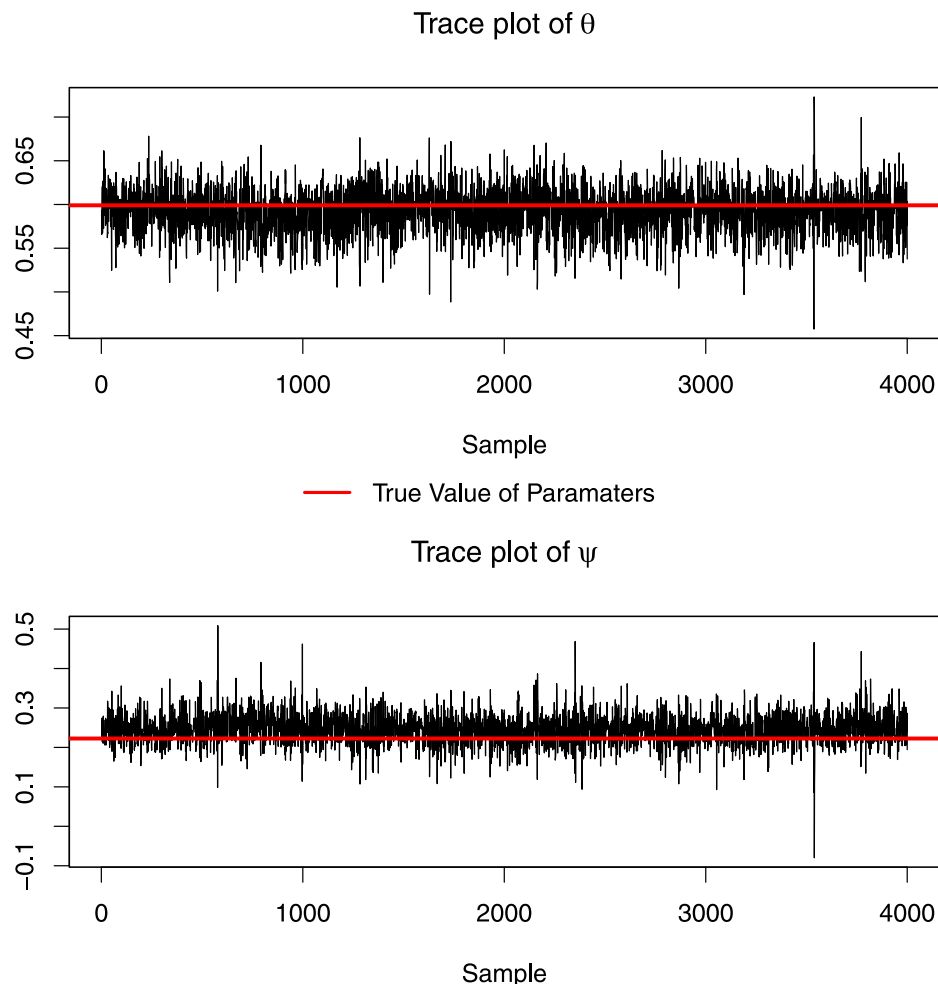
Fig 5. Trace plot of AR and MA parameters. Both parameters converged to the true values.

We executed 5,000 iterations of the MCMC estimation algorithm and the first 1,000 iterations were treated as a burn in period and discarded. The convergence of AR and MA parameters is evident in Fig 5. Both AR and MA parameters converge, display good mixing and are close to the true values of those parameters.

The top panel of Fig 6 presents the simulated signal with the true change-points (red vertical line). The middle plot shows the posterior distribution of occurrence of change-point locations, that is, the posterior probability of being change-points at each position. The height of the (red) 'spikes' indicates the posterior probability of a change-point being selected at each time point. The top two plots show the locations of estimated change-points and the true change-points are similar. This picture is clearer if we compare actual change-point locations and estimated change-point locations with the simulated data, as in the top and bottom plots of Fig 6. The blue lines indicate time points at which the posterior probabilities of change-points are greater than 0.5. Fig 6 (bottom) identifies 17 change-points out of 19 true change-points.

Fig 7 plots posterior estimators of mean signal level (c) at each position of the simulated signal. This plot clearly indicates 20 segments in the simulated signal detecting a change in mean even where change-points were not detected in Fig 6.
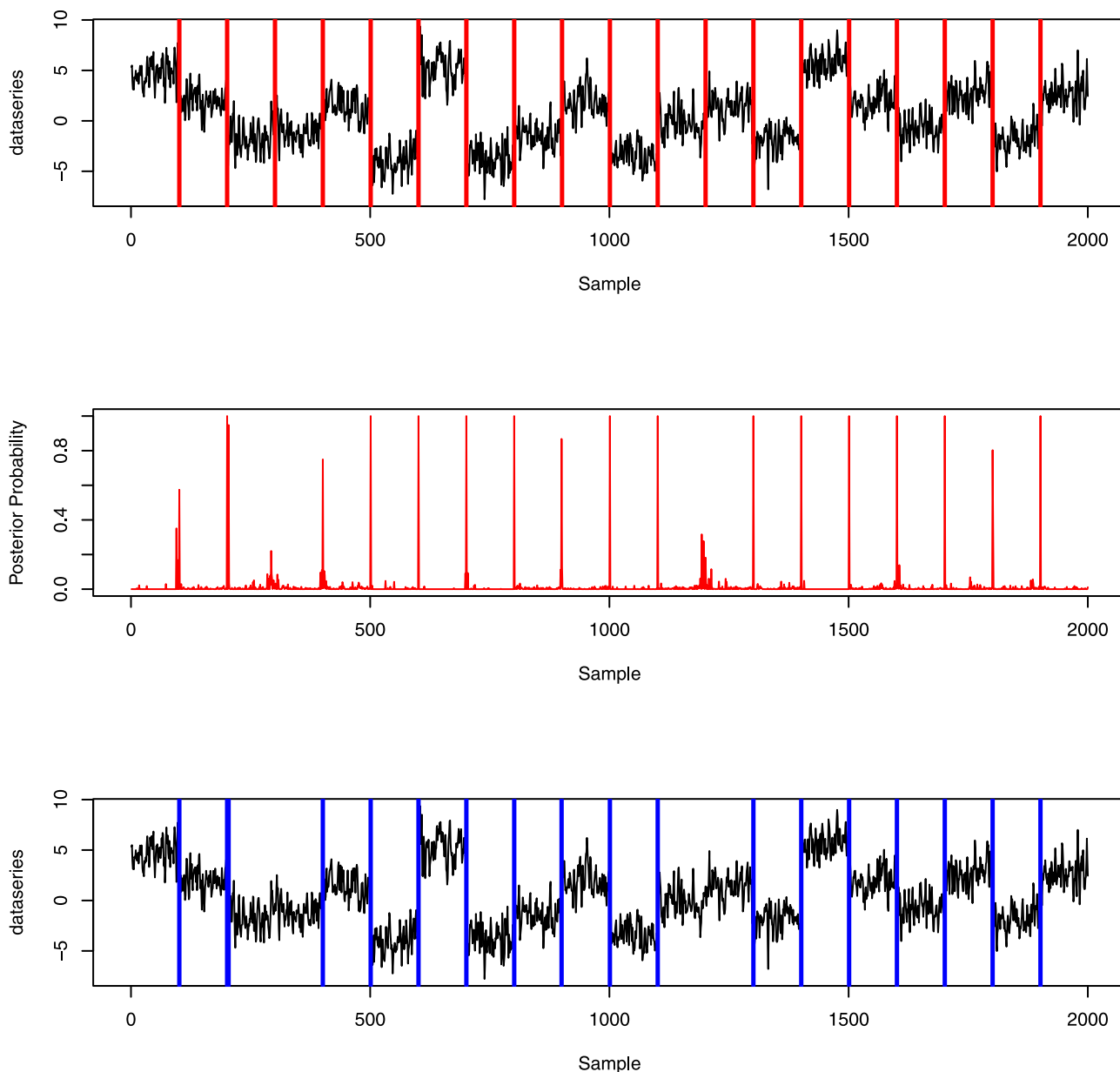
Fig 6. (Top) Segmented signal with the true change-point locations. (Middle) Posterior probabilities of occurrence of change-points. (Bottom) Estimated change-point locations (posterior probability greater than 0.5). The middle plot shows the location of peaks in the probability profile closely follows the true change-points locations but in some positions with low posterior probability. Using a threshold in posterior probability 0.5, we identify 17 change-points out of 19 which match the locations of the true change-points.

https://doi.org/10.1371/journal.pone.0208927.g006

Well-log data

We now apply the segmented ARMA model to identify change-points in a real data set. This data records 4050 measurements of nuclear magnetic response of underground rocks during the drilling of a well. During drilling, data were obtained at discrete time points by lowering a probe into a bore-hole in the Earth's surface. This geophysical data originates from Ó Rua-naidh and Fitzgerald [34] and has been previously analyzed in the context of change-point
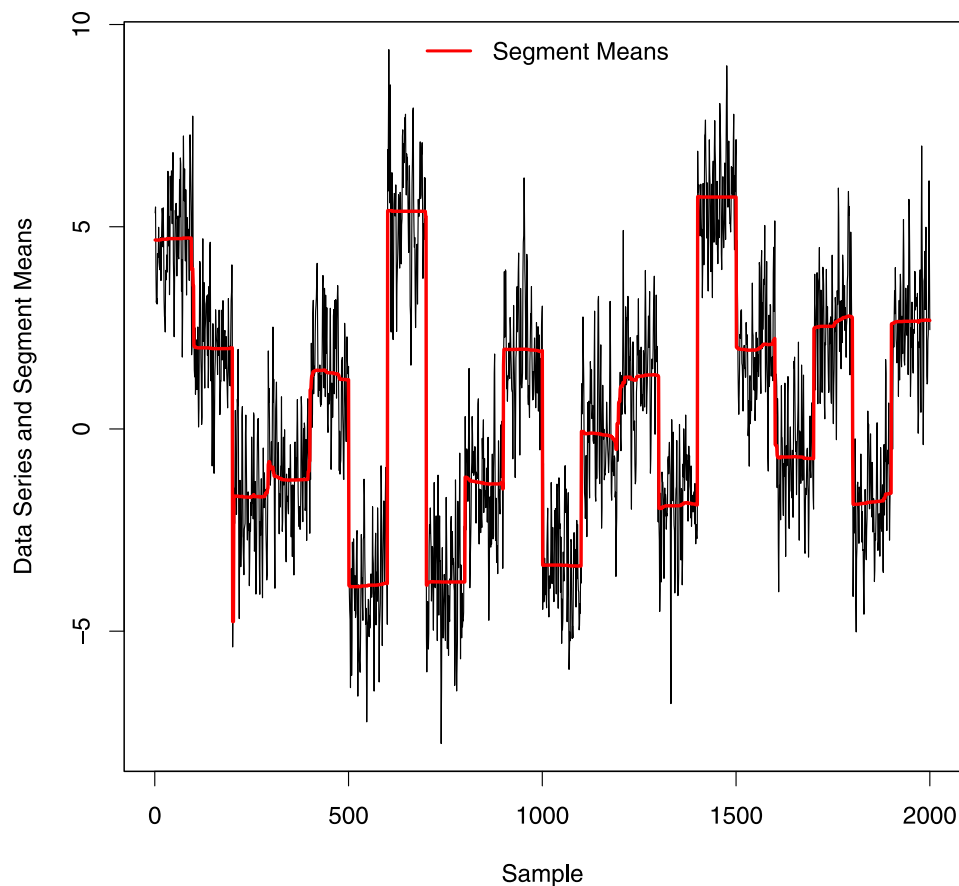
Fig 7. Segment means at each position of simulated data.

detection by many researchers, for example, by Fearnhead and Clifford [35], Fearnhead [36] and by Whiteley *et al.* [1]. A few outliers present in this data were removed by hand before analyzing the data as in Fearnhead [36]. The data is shown in Fig 8.

The piecewise constant signal of this data indicates information about the geophysical structure of the rocks in the well. Changes in mean occur at each time point whenever a new rock type is found. To identify change-points in this well-log data, Ó Ruanaidh and Fitzger-ald [34] used a Gibbs sampler to fit a Bayesian change-point model with a fixed number of change-points; Fearnhead and Clifford [35] used on-line Bayesian analysis of data with a hidden Markov model using particle filters; Fearnhead [36] considered an extension of the model described by Fearnhead and Clifford [35] in which they considered all the parameters of their model to be unknown and they used reversible jump MCMC to fit that model; Whiteley *et al.* [1] considered the same model used by Fearnhead [36] to analyze well-log data but here they used a block Gibbs sampler for generating samples from the posterior distribution.

To find change-points in this data, we attempted to fit an AR(1) model, an MA(1) model and an ARMA(1,1) model. We are the first to investigate this data using a segmented ARMA model. Each model was run for 5000 iterations and then tested for convergence of each param-eter using trace plots and an autocorrelation function (ACF) plot which represents the degree of correlation between all pairs of samples separated by progressively larger lags (number of
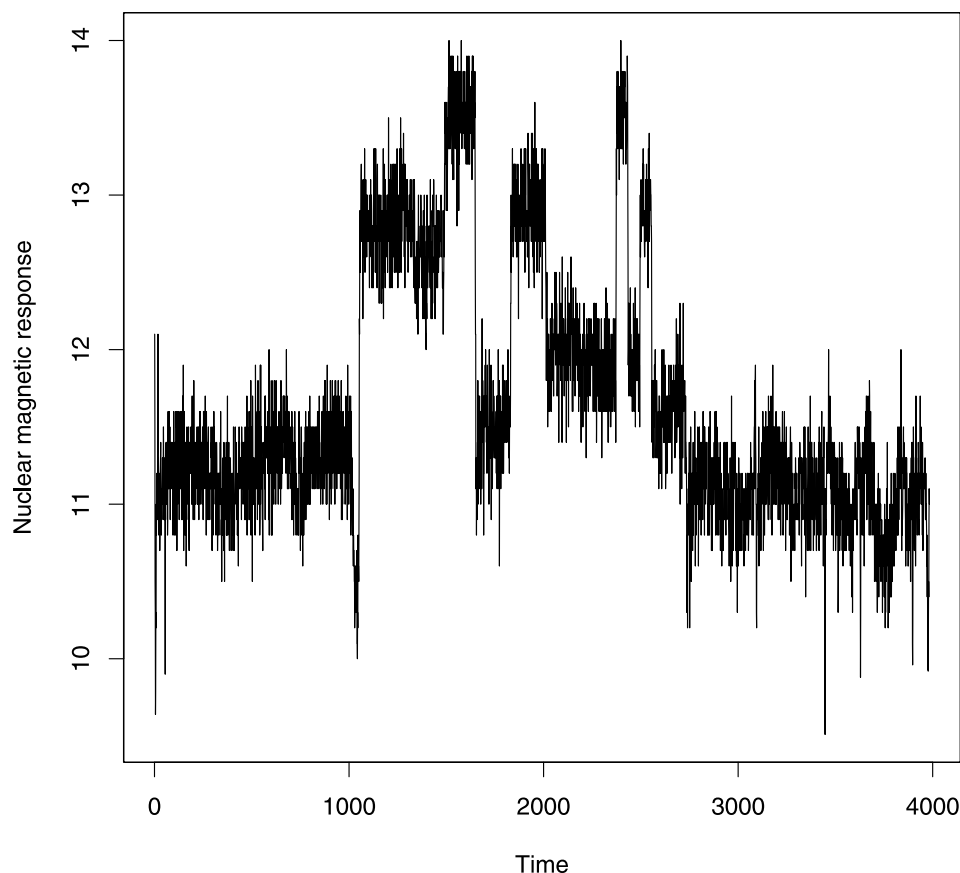
Fig 8. Well-log data. This data provides information about the rock structure of the well. Some change-points are present in the data reflecting the presence of a new rock.

samples). The change-point profiles and segment mean profiles for each model are given in Figs 9–11.

In our model, the initial segmentation was generated using a probability of starting a new segment $\phi = 0.1$. The initial segmentation was generated by throwing a uniform(0,1) random number for each sequence position except the first, making that position a change-point if the random number is less than ϕ at that position. Note that, this initial segmentation should not affect the stationary distribution of the Markov chain. The posterior probabilities of occurrence of change-points at each position of the input sequence are calculated using the uniform prior probability distribution for ϕ (the probability that any given sequence position is a change-point) and the likelihood probability ($p(K, \mathbf{s}|\phi) = \phi^{K-1}(1 - \phi)^{T-K-1}$) of generating a new segmentation with K change-points and $\mathbf{s} = (1 = s_1 < \cdots < s_K \leq T)$ starting positions.

All change-point profile plots (Figs 9–11) show almost the same change-points locations. But at some locations the AR(1) model gives comparatively smaller posterior probability than the MA(1) and ARMA(1,1) models. In addition, the AR(1) model identifies fewer change-points with high posterior probability (when the posterior probability is more than 0.5) than the other two models. These results are somewhat similar to the previous results found in the change-point literature [1, 34–36].

Since the three segmented ARMA models indicate many of the same change-point locations, we compare the change-points locations for which posterior probabilities are greater
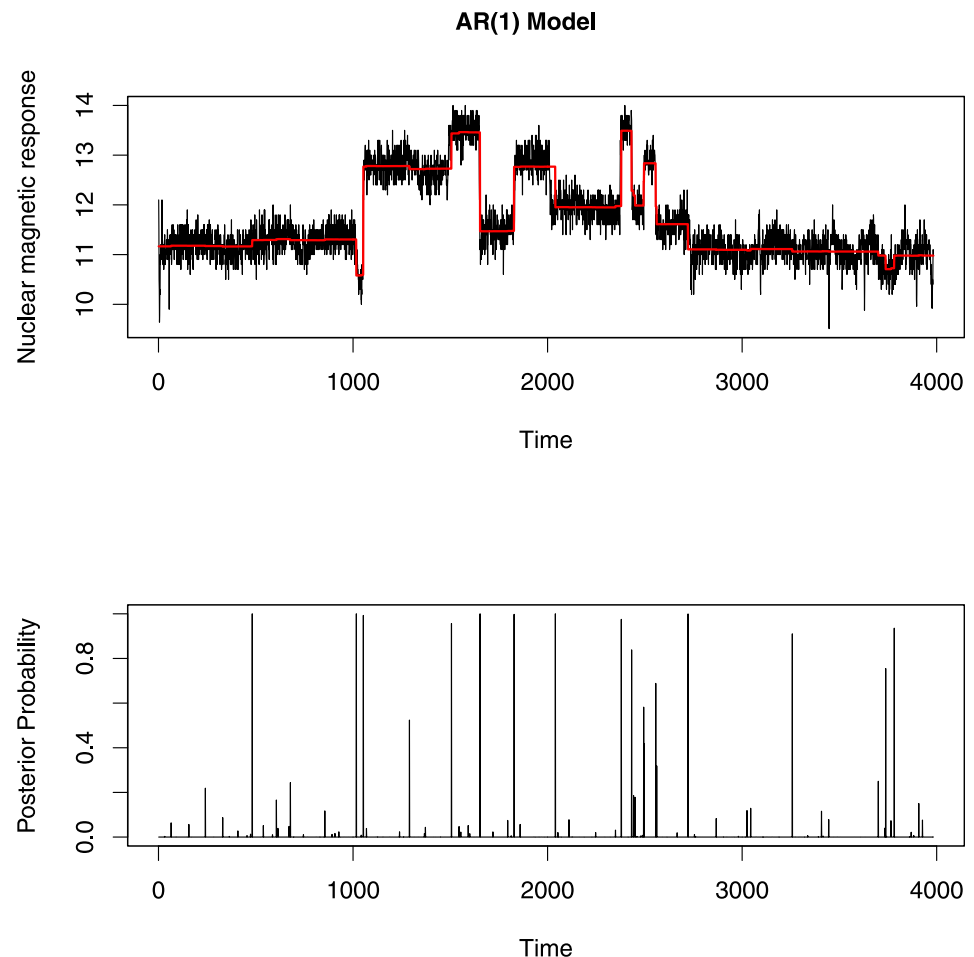
AR(1) Model

Fig 9. Top plot shows the posterior estimators of mean signal level (c) on each segment of AR(1) model. Bottom plot shows the posterior probability of a change-point at each position. These segment means and change-point positions indicate significant jumps in the original data.

than 0.5 in Fig 12. Among these three models, the ARMA(1,1) model identifies the largest number of change-points and matches more closely with the number and locations of change-points detectable to the eye. Moreover, the ARMA(1,1) model picks up small changes in mean with high posterior probability whereas the AR(1) and MA(1) models missed change-points at some time points where small jumps occured in the data.

To determine the best model among these three, we compare these models using the deviance information criterion (DICV). The DICV is defined as: $DICV = p_v + \overline{D(\theta)}$, where $\overline{D(\theta)}$ is the mean posterior deviance, $p_v = Var(D(\theta))/2$ and deviance $D(\theta) = -2\text{lnf}(y|\theta)$ (details of DICV are in [37, 38]). The DICV, $\overline{D(\theta)}$ and p_v of these three models are shown in Table 1.

Here, the segmented ARMA(1,1) model gives lower DICV than the other two models. The lower DICV of the ARMA(1,1) model supports the conclusion that the ARMA(1,1) model is the best of the three.

Discussion

In this paper, we have developed a Bayesian change-point segmented ARMA model to segment time series data. The novel features of our approach include: (1) It uses an ARMA model
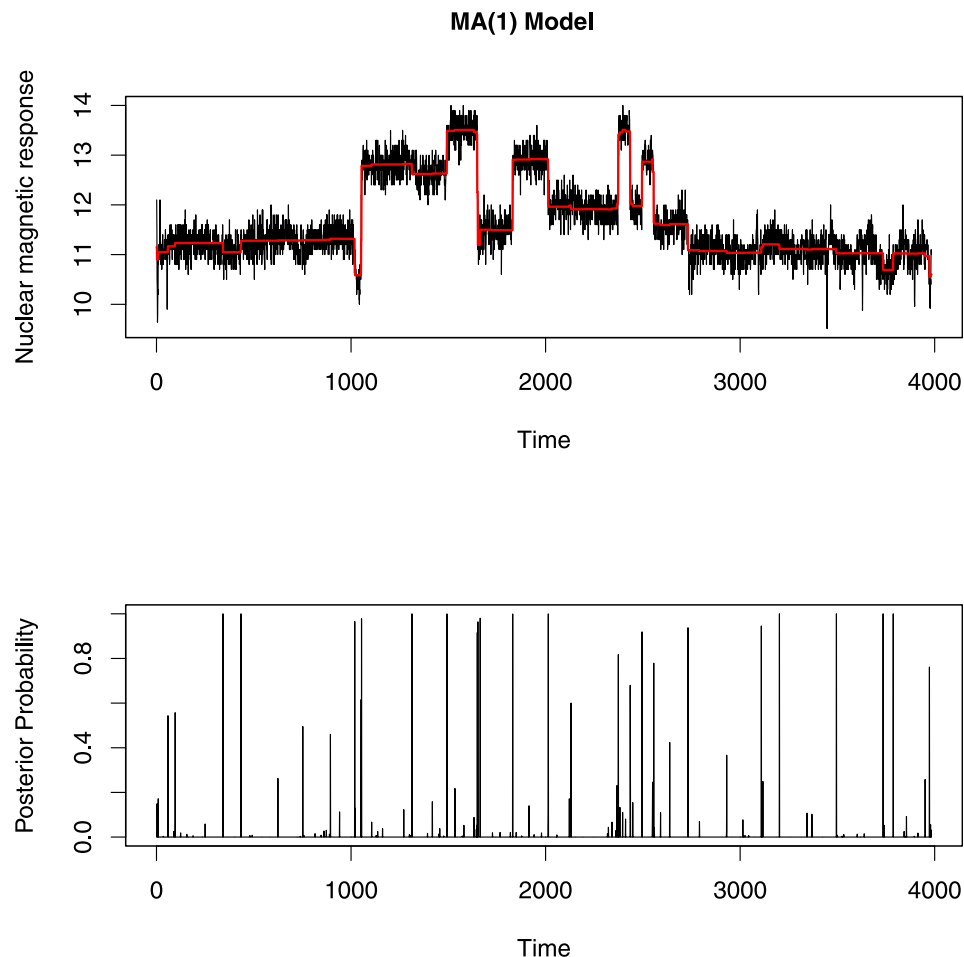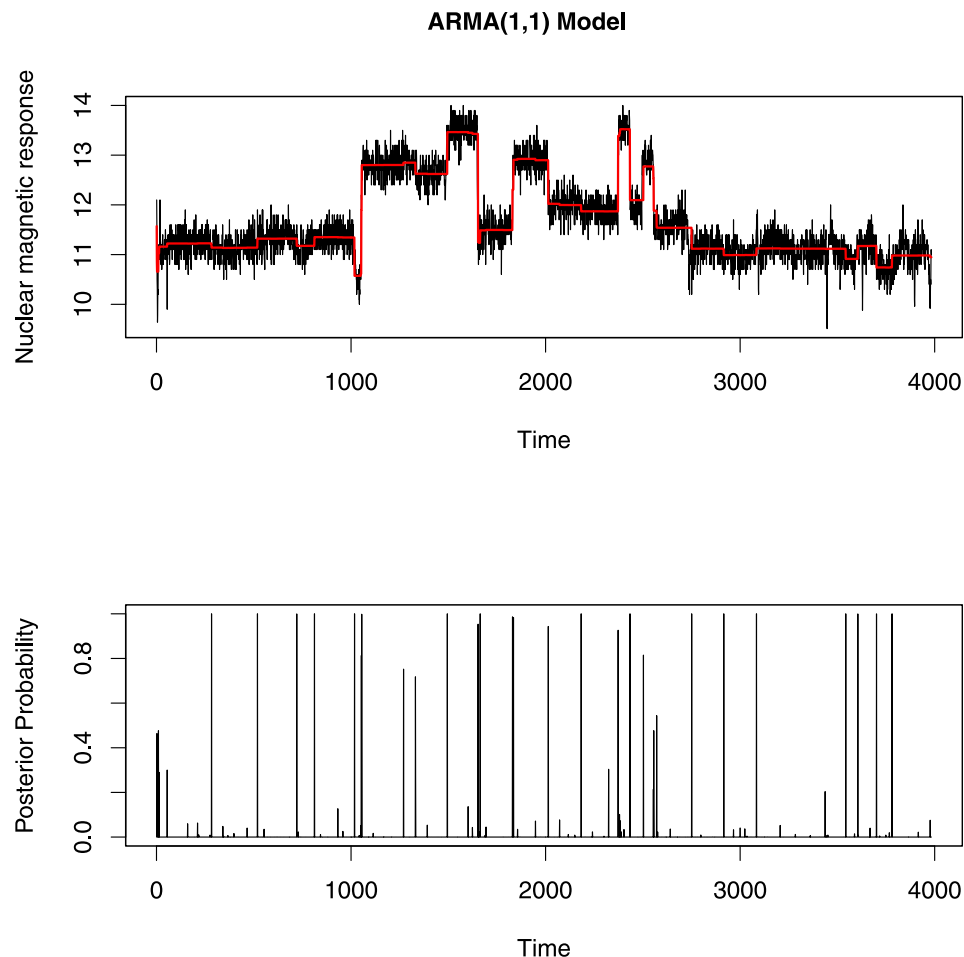
MA(1) Model

Fig 10. Top and bottom plot show the segment mean profiles and the change-point profiles for MA(1) model respectively. This model identifies more change-points than the AR(1) model.

in each segment (2) It uses a highly efficient sampling technique (GGS) to generate samples from a posterior distribution. Results for simulated data and real data show that this model achieves high detection accuracy.

The results we obtain for the well-log data seem reasonable when judged by eye. The posterior probabilities of change-point occurrences of the ARMA(1,1) model are somewhat similar to results previously found in the literature for this data. Ó Ruanaidh and Fitzgerald [34] assumed 13 change-points exist in this data (after removing outliers) but our ARMA (1,1) model identifies 27 change-points (considering posterior probability more than 0.5). They missed some small jumps in the data whereas the ARMA(1,1) model identifies small jumps as well as large jumps. Fearnhead and Clifford [35] inferred 16 change-points but since they did not remove the outliers of the data, the results cannot be directly compared. Fearnhead [36] found too many change-points in their piecewise constant model. They used a random walk sampler and found similar change-points to our ARMA(1,1) model. In addition to the change-points found by Fearnhead [36], the ARMA(1,1) model identifies some significant change-points between time points 1 and 1000 and after 2800 (shown in Figs 13 and 14).

ARMA(1,1) Model

Fig 11. Top and bottom plot show the segment mean profiles and the change-point profiles for ARMA(1,1) model respectively. These change-points and segment means are almost identical to those identified using the MA(1) model.

The results of the ARMA(1,1) model are similar to the results reported in Whiteley *et al.* [1] but in some time points ARMA(1,1) model shows a higher posterior probability of change-point occurrences.

Overall, this paper has presented a promising new direction for estimation of change-point models by assuming a segment-wise ARMA model. Most of the previous methods discussed in the literature use autoregressive (AR) models in each segment. Adding a moving average component helps to consider the dependence between residual terms which is an advantage over the segmented AR model. Our results obtained using simulated data demonstrate that when the data are generated via an ARMA process, the ARMA model finds more change-points than the AR model, without finding false positives. The ARMA model also finds more change-points in the well log data, but a question remains whether the additional change-points are false positives in this case. As the true locations of change in this data set are unknown, this question cannot be answered definitively. However, we compared the segmented ARMA (1,1) model, segmented AR(1) model and segmented MA (1) model using the deviance information criterion (DICV), and found that the ARMA (1,1) is favoured by this criterion, suggesting the additional change-points reflect a real feature of the data. Since this model assumes the same
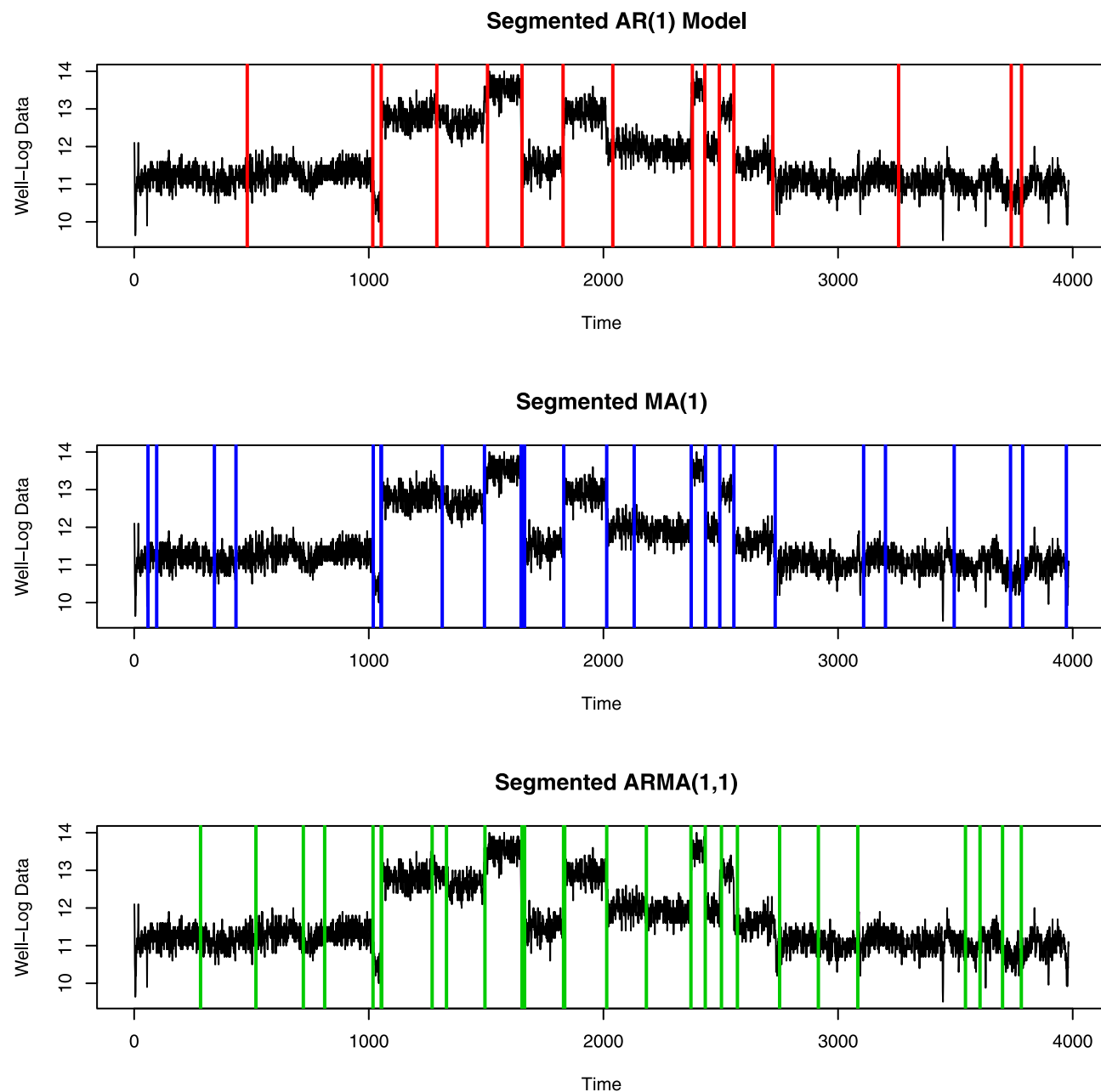
Segmented AR(1) Model

Segmented MA(1)

Segmented ARMA(1,1)

Fig 12. Estimated change-point locations with posterior probability greater than 0.5 for AR(1), MA(1) and ARMA(1,1) model.

https://doi.org/10.1371/journal.pone.0208927.g012

Table 1. DICV of Models.

Model	$\overline{D(\theta)}$	p_v	DICV
Segmented AR(1) Model	958.0043	181306.7	182264.7
Segmented MA(1) Model	509.5514	253045.4	253554.9
Segmented ARMA(1,1) Model	672.8849	169945.6	170618.5

https://doi.org/10.1371/journal.pone.0208927.t001
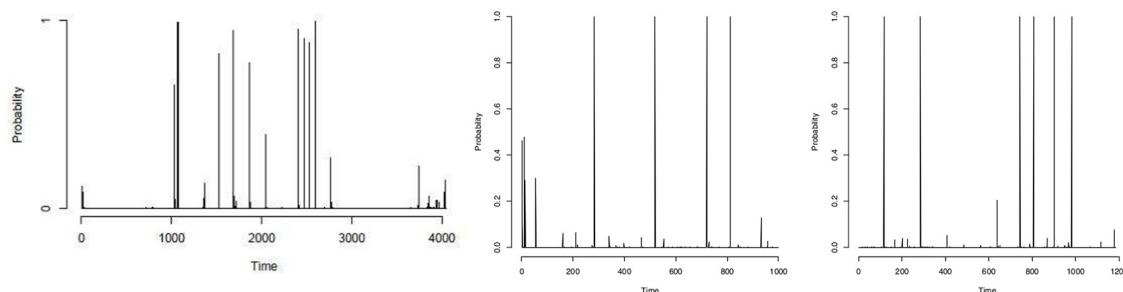
Fig 13. Comparison between the results of a random walk model [36] and our ARMA(1,1) model. (Left) change-point profiles of random walk model [36]; (centre) change-point profiles of ARMA(1,1) at time points 1 to 1000; (right) change-point profiles of ARMA(1,1) after time point 2800.

https://doi.org/10.1371/journal.pone.0208927.g013
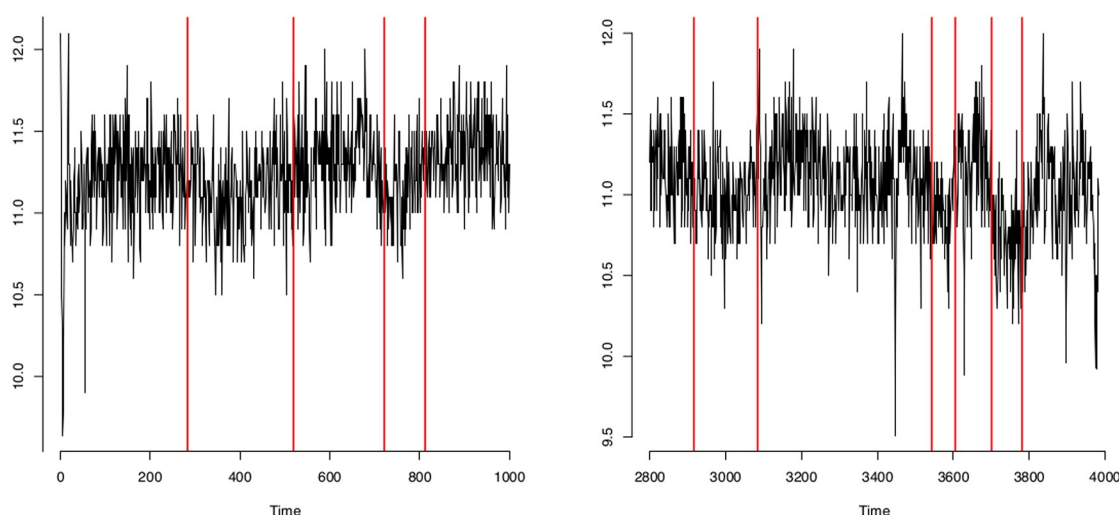
Fig 14. Change-point locations identified using ARMA(1,1) model displayed with the original well-log data at two different time points. (Left) change-point locations using ARMA(1,1) model in well-log data at time points 1 to 1000; (right) change-point locations using ARMA(1,1) model in well-log data after time point 2800.

https://doi.org/10.1371/journal.pone.0208927.g014

variance for all segments, it is not suitable for data sets in which different segments have different variance.

Supporting information

S1 Appendix. Details of posterior distribution.
(PDF)

S2 Appendix. Generalized gibbs sampling.
(PDF)

S3 Appendix. Supplementary material A.
(PDF)

S4 Appendix. Supplementary material B.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Farhana Sadia, Sarah Boyd, Jonathan M. Keith.

Data curation: Farhana Sadia.

Investigation: Farhana Sadia.

Methodology: Farhana Sadia, Sarah Boyd, Jonathan M. Keith.

Project administration: Jonathan M. Keith.

Software: Farhana Sadia, Sarah Boyd, Jonathan M. Keith.

Supervision: Jonathan M. Keith.

Validation: Farhana Sadia.

Writing – original draft: Farhana Sadia.

Writing – review & editing: Farhana Sadia, Sarah Boyd, Jonathan M. Keith.

References

1. Whiteley N, Andrieu C, Doucet A. Bayesian computational methods for inference in multiple change-points models; 2011.

2. Carlin BP, Gelfand AE, Smith AF. Hierarchical Bayesian analysis of changepoint problems. Applied statistics. 1992; p. 389–405. https://doi.org/10.2307/2347570

3. Reeves J, Chen J, Wang XL, Lund R, Lu QQ. A review and comparison of changepoint detection techniques for climate data. Journal of Applied Meteorology and Climatology. 2007; 46(6):900–915. https://doi.org/10.1175/JAM2493.1

4. Jensen U, Lütkebohmert C. Change-point models. Encyclopedia of Statistics in Quality and Reliability. 2007.

5. Ko SI, Chong TT, Ghosh P, et al. Dirichlet Process Hidden Markov Multiple Change-point Model. Bayesian Analysis. 2015; 10(2):275–296. https://doi.org/10.1214/14-BA910

6. Chernoff H, Zacks S. Estimating the current mean of a normal distribution which is subjected to changes in time. The Annals of Mathematical Statistics. 1964; p. 999–1018. https://doi.org/10.1214/aoms/1177700517

7. Smith A. A Bayesian approach to inference about a change-point in a sequence of random variables. Biometrika. 1975; 62(2):407–416. https://doi.org/10.1093/biomet/62.2.407

8. Harlé F, Chatelain F, Gouy-Pailler C, Achard S. Bayesian Model for Multiple Change-points Detection in Multivariate Time Series. arXiv preprint arXiv:14073206. 2014;.

9. Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika. 1995; 82(4):711–732. https://doi.org/10.1093/biomet/82.4.711

10. Chib S. Estimation and comparison of multiple change-point models. Journal of econometrics. 1998; 86(2):221–241. https://doi.org/10.1016/S0304-4076(97)00115-2

11. Keith JM. Segmenting eukaryotic genomes with the generalized Gibbs sampler. Journal of Computational Biology. 2006; 13(7):1369–1383. https://doi.org/10.1089/cmb.2006.13.1369 PMID: 17037964

12. Sofronov GY, Evans GE, Keith JM, Kroese DP. Identifying change-points in biological sequences via sequential importance sampling. Environmental Modeling & Assessment. 2009; 14(5):577–584. https://doi.org/10.1007/s10666-008-9160-8

13. Cribben I. Detecting Dependence Change Points in Multivariate Time Series with Applications in Neuroscience and Finance. Columbia University; 2012.

14. Punskaya E, Andrieu C, Doucet A, Fitzgerald WJ. Bayesian curve fitting using MCMC with applications to signal segmentation. Signal Processing, IEEE Transactions on. 2002; 50(3):747–758. https://doi.org/10.1109/78.984776

15. Chan KS, Tong H. On estimating thresholds in autoregressive models. Journal of time series analysis. 1986; 7(3):179–190. https://doi.org/10.1111/j.1467-9892.1986.tb00501.x

16. Tong H. On a threshold model. 29. Sijthoff & Noordhoff; 1978.

17. Tong H, Lim KS. Threshold autoregression, limit cycles and cyclical data. Journal of the Royal Statistical Society Series B (Methodological). 1980; p. 245–292. https://doi.org/10.1111/j.2517-6161.1980.tb01126.x

18. Tong H. Threshold models in non-linear time series analysis. vol. 21. Springer Science & Business Media; 2012.

19. Davis RA, Lee TCM, Rodriguez-Yam GA. Structural break estimation for nonstationary time series models. Journal of the American Statistical Association. 2006; 101(473):223–239. https://doi.org/10.1198/016214505000000745

20. Wood S, Rosen O, Kohn R. Bayesian mixtures of autoregressive models. Journal of Computational and Graphical Statistics. 2011; 20(1):174–195. https://doi.org/10.1198/jcgs.2010.09174

21. Chakar S, Lebarbier E, Lévy-Leduc C, Robin S, et al. A robust approach for estimating change-points in the mean of an AR(1) process. Bernoulli. 2017; 23(2):1408–1447. https://doi.org/10.3150/15-BEJ782

22. Albert JH, Chib S. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. Journal of Business & Economic Statistics. 1993; 11(1):1–15. https://doi.org/10.2307/1391303

23. Keith JM, Adams P, Stephen S, Mattick JS. Delineating slowly and rapidly evolving fractions of the Drosophila genome. Journal of Computational Biology. 2008; 15(4):407–430. https://doi.org/10.1089/cmb.2007.0173 PMID: 18435570

24. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. The journal of chemical physics. 1953; 21(6):1087–1092. https://doi.org/10.1063/1.1699114

25. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970; 57(1):97–109. https://doi.org/10.1093/biomet/57.1.97

26. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on pattern analysis and machine intelligence. 1984;(6):721–741. https://doi.org/10.1109/TPAMI.1984.4767596 PMID: 22499653

27. Liu JS, Liang F, Wong WH. The multiple-try method and local optimization in Metropolis sampling. Journal of the American Statistical Association. 2000; 95(449):121–134. https://doi.org/10.1080/01621459.2000.10473908

28. Mira A, et al. On Metropolis-Hastings algorithms with delayed rejection. Metron. 2001; 59(3-4):231–241.

29. Keith JM, Kroese DP, Bryant D. A generalized Markov sampler. Methodology and Computing in Applied Probability. 2004; 6(1):29–53. https://doi.org/10.1023/B:MCAP.0000012414.14405.15

30. Oldmeadow C, Mengersen Kerrie, Mattick JS, Keith JM. Multiple evolutionary rate classes in animal genome evolution. Molecular biology and evolution. 2009; 27(4):942–953. https://doi.org/10.1093/molbev/msp299 PMID: 19955480

31. Keith JM, Spring D. Agent-based Bayesian approach to monitoring the progress of invasive species eradication programs. Proceedings of the National Academy of Sciences. 2013; 110(33):13428–13433. https://doi.org/10.1073/pnas.1216146110

32. Neal RM. Slice sampling. Annals of statistics. 2003; p. 705–741. https://doi.org/10.1214/aos/1056562461

33. Cook SR, Gelman A, Rubin DB. Validation of software for Bayesian models using posterior quantiles. Journal of Computational and Graphical Statistics. 2006; 15(3). https://doi.org/10.1198/106186006X136976

34. Ó Ruanaidh JJK, Fitzgerald WJ. Numerical bayesian methods applied to signal processing; 1996.

35. Fearnhead P, Clifford P. On-line inference for hidden Markov models via particle filters. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2003; 65(4):887–899. https://doi.org/10.1111/1467-9868.00421

36. Fearnhead P. Exact and efficient Bayesian inference for multiple changepoint problems. Statistics and computing. 2006; 16(2):203–213. https://doi.org/10.1007/s11222-006-8450-8

37. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. vol. 2. Chapman & Hall/CRC Boca Raton, FL, USA; 2014.

38. Sturtz S, Ligges U, Gelman A, et al. R2WinBUGS: a package for running WinBUGS from R. Journal of Statistical software. 2005; 12(3):1–16. https://doi.org/10.18637/jss.v012.i03