

# Particle Markov Chain Monte Carlo for Multiple Change-point Problems

Nick Whiteley, Christophe Andrieu

Department of Mathematics,

University of Bristol,

University Walk,

Bristol BS8 1TW, UK.

Email: {Nick.Whiteley,C.Andrieu}@bris.ac.uk

Arnaud Doucet

The Institute of Statistical Mathematics,

4-6-7 Minami-Azabu, Minato-ku,

Tokyo 106-8569, Japan.

Email: Arnaud@ism.ac.jp

## Abstract

Multiple change-point models are a popular class of time series models which allow the description of temporal heterogeneity in data. We develop efficient Markov Chain Monte Carlo (MCMC) algorithms to perform Bayesian inference in this context. Our so-called Particle MCMC (PMCMC) algorithms rely on an efficient Sequential Monte Carlo (SMC) technique for change-point models, developed in [13], to build high-dimensional proposals. The construction of the new algorithms differs significantly from the PMCMC schemes proposed in [1]. We demonstrate the performance of our algorithms on various examples.

## 1 Introduction

Time series data often exhibit temporal heterogeneity. In multiple change-point problems, the task is to segment a sequence of observations  $y_1, y_2, \dots, y_T$  by choosing a sequence of change-point locations  $0 < \tau_1 < \tau_2 < \dots < \tau_k < T$  such that the observations are homogeneous within segments and heterogeneous across segments. Statistical analysis of change-point problems has a long history; as of 1992 the literature on this topic was “enormous” [5]. In the Bayesian approach to multiple change-point problems, a joint prior distribution is placed over the number and locations of change-points [3], [4]. An observation model then describes the distribution of the data, given the change-points. Generally, both the prior over the change-points and the observation model depend on an unknown parameter  $\theta$ . The values taken by this parameter can dramatically influence the properties of the model and inferences drawn about change-point locations. Therefore it is very important to treat this parameter as an unknown. In a Bayesian framework, this can be achieved easily by placing a prior over  $\theta$ .

However, performing Bayesian inference for multiple change-points and the parameter  $\theta$  is a challenging problem. Even when  $\theta$  is assumed known, exact computation of the posterior distribution of the change-points is intractable for large data sets. This issue is typically tackled using MCMC techniques. Unfortunately, MCMC algorithms which update change-point locations in a one-at-a-time manner [16], [19], or condition on latent variables associated with each segment [7], can be slow mixing due to the strong correlations between the change-point locations and latent variables. Alternatives, which sample all the change-points in one block with the latent variables integrated out, have a cost per MCMC iteration which is of the order  $T^2$  [12], i.e. quadratic in the length of the data record. For real-world data sets of several thousand observations, this  $T^2$  method can be prohibitively expensive.

PMCMC algorithms, recently introduced in [1], are a class of MCMC algorithms which allow SMC a.k.a. particle filtering techniques [9] to be used in building high-dimensional proposals within a MCMC scheme. The SMC method proposed in [13] is an efficient algorithm for sampling change-points. However,

its structure differs very significantly from the SMC algorithms treated in [1]. The main contribution of this paper is to show how this very algorithm can be used within MCMC schemes to obtain efficient samplers for multiple-change point models. We derive two original PMCMC algorithms, whose cost per MCMC iteration is of the order  $N T$  where  $N$  is the number of particles in the SMC approximation and whose experimental performance compares remarkably well to  $T^2$  algorithms for  $N \ll T$ .

The rest of this paper is structured as follows. Section 2 specifies the change-point model of interest. Section 3 describes existing MCMC methods for change-point models. Section 4 describes the exact and SMC sampling techniques of [13]. In section 5 we introduce the new PMCMC algorithms and a simulation study is documented in section 6.

## 2 Multiple Change-point Model

The change-point model we consider is essentially the same as that treated in [13], except that the parameter  $\theta$ , valued in some space  $\Theta$ , is here assumed unknown and is assigned a suitable prior  $p(\theta)$ . It is important to note that the parameter  $\theta$  is global, in the sense that it is common to all segments defined by any change-point configuration (specific examples are given in section 6). From now on by convention we fix  $\tau_0 = 0$ ,  $\tau_{k+1} = T$  and for some generic sequence  $\{z_n\}$  we adopt the notation  $z_{i:j} := (z_i, z_{i+1}, \dots, z_j)$ .

The sequence  $\tau_0, \tau_1, \tau_2, \dots$  is increasing and is assumed Markov with homogeneous transition probabilities given for  $j = 1, 2, \dots$  by  $p_\theta(\tau_j | \tau_{j-1}) = h_\theta(\tau_j - \tau_{j-1})$ . We are concerned with the space of all change-point configurations over  $\{1, \dots, T-1\}$  which we denote by the disjoint union  $\mathcal{T}_T := \bigcup_{k=0}^{T-1} \mathcal{T}_{T,k}$ , where  $\mathcal{T}_{T,k} = \{\tau_{1:k} \in \{1, \dots, T-1\}^k : \tau_1 < \dots < \tau_k\}$ . The joint prior probability of exactly  $k$  change-points and their locations is then given by

$$p_\theta(\tau_{1:k}) = [1 - H_\theta(T - \tau_k)] \prod_{j=2}^k h_\theta(\tau_j - \tau_{j-1}),$$

where  $H_\theta$  is the c.d.f. associated with  $h_\theta$ , i.e.  $H_\theta(n) = \sum_{i=1}^n h_\theta(i)$ .

A collection of change-points  $\tau_{1:k}$  defines  $k+1$  segments where for  $j = 1, 2, \dots, k+1$ , the  $j^{\text{th}}$  segment consists of the observations  $y_{\tau_{j-1}+1:\tau_j}$ . Similarly to [3], [13], we assume that given the location of a change-point and  $\theta$ , the observations before that change-point are independent of those after. We also assume that, for any pair of consecutive change-point locations  $\tau_{j-1}$  and  $\tau_j$  we are able to evaluate the likelihood  $p_\theta(y_{\tau_{j-1}+1:\tau_j} | \tau_{j-1:j})$ . As in [13], this likelihood arises from summing over a number of possible models for the data in each segment and/or integrating out local latent variables associated with each segment under conjugate priors.

We are interested in the joint posterior distribution

$$p(\theta, \tau_{1:k} | y_{1:T}) \propto p_\theta(y_{1:T} | \tau_{1:k}) p_\theta(\tau_{1:k}) p(\theta). \quad (1)$$

As advocated in [7], [8], [13], for computational purposes it is fruitful to consider an alternative parameterization of the change-point model. For each  $n = 1, 2, \dots, T$ , let  $X_n$  be the random variable valued in  $E_n := \{0, \dots, n-1\}$  which is the location of the latest change-point prior to time  $n$ . It is straightforward to verify that there is a one-to-one correspondence between paths  $x_{1:T}$  and change-point configurations  $\tau_{1:k}$ . The above prior on  $\tau_{1:k}$  then corresponds to the sequence  $X_{1:T}$  being Markov with, for  $0 \leq x_{n-1} \leq x_n < n$ , transition probabilities  $f_n^\theta(x_n | x_{n-1}) := p_\theta(x_n | x_{n-1})$  equal to

$$\begin{aligned} & \frac{1 - H_\theta(n - x_{n-1} - 1)}{1 - H_\theta(n - x_{n-1} - 2)} && \text{if } x_n = x_{n-1}, \\ & \frac{H_\theta(n - x_{n-1} - 1) - H_\theta(n - x_{n-1} - 2)}{1 - H_\theta(n - x_{n-1} - 2)} && \text{if } x_n = n - 1, \\ & 0 && \text{otherwise,} \end{aligned}$$

and initial distribution specified by the convention  $f_1^\theta(0|0) = p_\theta(x_1 = 0) = 1$ . For  $n > x_n$  the predictive likelihood is denoted

$$g_n^\theta(x_n) := p_\theta(y_n | y_{1:n-1}, x_n).$$

In order to simplify presentation, throughout this paper we assume that for all  $\theta, n, x_n$  and  $x_{n-1}$ ,  $f_n^\theta(x_n | x_{n-1}) > 0$  and  $g_n^\theta(x_n) > 0$ . This assumption is satisfied for the majority of change-point models of practical interest and relaxing it requires only cosmetic changes to the proposed algorithms.

The reparameterization described above can be exploited in order to approximate  $p_\theta(y_{1:T})$  and sample from approximations of  $p_\theta(\tau_{1:k}|y_{1:T})$  using a SMC algorithm. In section 5 we will see that these two tasks are central to the construction of new MCMC algorithms targeting  $p(\theta, \tau_{1:k}|y_{1:T})$ . First, we review some existing approaches to posterior sampling in multiple-change point models using standard MCMC methods.

### 3 MCMC for Multiple Change–point models

The design of MCMC algorithms for posterior sampling in change–point models has been a topic of interest for some years. In [5], a Gibbs sampler was devised for a Bayesian change–point model in which the number of change–points was fixed to 1. This Gibbs sampler was extended to models with multiple change–points in [19]. The latter approach involves one-at-a-time sampling of the change-point locations and associated latent variables from their respective full conditional distributions. It is well known that one-at-a-time Gibbs samplers can suffer from slow mixing which arises from strong correlations in the target distribution. Empirical evidence shows that similar algorithms which employ Metropolis–Hastings steps to sample one-at-a-time in change–point models [16] suffer from similar problems [12].

For multiple change–point models, [7] presented a Gibbs sampling algorithm in which the entire change–point configuration  $\tau_{1:k}$  is sampled from its full conditional distribution, given  $\theta$  and the latent variables associated with each segment. This is an instance of blocking in a Gibbs sampler: a technique which often improves mixing significantly [17]. One step further is to sample  $\tau_{1:k}$  from  $p_\theta(\tau_{1:k}|y_{1:T})$ , i.e. with the latent variables associated with each segment integrated out. This can improve mixing even further and is achievable using the exact forward-backward recursions of [11], [12], [13]. However, these schemes have a computational cost per MCMC iteration which grows quadratically in  $T$ , the length of the data record. In modern applications,  $T$  may be of the order of thousands and so the cost of these methods can be prohibitive. In the next section we review the exact forward-backward methods for sampling from  $p_\theta(\tau_{1:k}|y_{1:T})$ , computing  $p_\theta(y_{1:T})$ , and the corresponding SMC methods.

## 4 Forward Filtering and Backward Sampling

### 4.1 Exact Methods

References [11], [12] and [13] proposed methods for obtaining exact samples from  $p_\theta(\tau_{1:k}|y_{1:T})$  and computing  $p_\theta(y_{1:T})$ . We focus on the algorithm of [13], which involves two steps. In the *forward filtering* pass, for  $n = 1, \dots, T$ , each filtering distribution  $p_\theta(x_n|y_{1:n})$  is computed recursively and stored. The likelihood  $p_\theta(y_{1:T})$  can be obtained from quantities computed in this filtering pass. In the *backward sampling* pass, a sequence of change-point locations  $\tau_{1:k}$  is obtained by sampling back through the stored filtering distributions, re-weighted appropriately. We refer to [13] for specific details.

In the change-point model of section 2, recall that the support of  $p_\theta(x_n|y_{1:n})$  is  $E_n = \{0, \dots, n-1\}$  which obeys the recursion

$$E_1 = \{0\}, \quad E_{n+1} = E_n \cup \{n\}, \quad n = 1, 2, \dots \quad (2)$$

from which it is clear that the support of the filtering distributions is increasing with time. It follows that the cost of computing  $p_\theta(y_{1:T})$  and the storage requirements of sampling from  $p_\theta(\tau_{1:k}|y_{1:T})$  grow quadratically in  $T$ . In many practical problems,  $T$  may be of the order of thousands and so these exact methods can be prohibitively expensive. In [11] and [12], it is suggested to reduce computational complexity by performing a deterministic truncation of certain quantities which arise in the filtering recursion, but this introduces a bias which is difficult to quantify.

### 4.2 SMC Method

Reference [13] also proposed an algorithm which employs SMC ideas in order to approximate the filtering distributions. This allows the cost per filtering iteration to be upper-bounded uniformly in time and is of order  $N$ . The storage requirements of obtaining a sample from the corresponding approximation of  $p_\theta(\tau_{1:k}|y_{1:T})$  and computing the corresponding approximation of  $p_\theta(y_{1:T})$  then grow only linearly with  $T$ .

SMC methods are a class of stochastic algorithms which allow approximation of a sequence of probability distributions [9]. Standard SMC methods yield a collection of  $N$  weighted samples, termed *particles*

$\{X^i, W(X^i)\}_{i=1}^N$  which define a random probability distribution,

$$\sum_{i=1}^N W(X^i) \delta_{X^i}, \quad \sum_{i=1}^N W(X^i) = 1, \quad (3)$$

where in full generality, the weight  $W(\cdot)$  is some possibly random function of the particle location  $X^i$ .

The structure of the change-point model is significantly different from that of the state-space models to which standard SMC algorithms are usually applied. For each  $n$ ,  $|E_n|$  is finite; in standard state-space models, the filtering distribution is usually defined on some continuous space. The SMC algorithm of [13] exploits the fact that  $|E_n|$  is finite and avoids the random proposal/importance sampling step present in standard SMC. Furthermore, it has the desirable property of avoiding particle duplication. This is a further development of the ingenious optimal resampling particle filter introduced in [10]. We refer to [13] for a detailed discussion of efficiency and applicability of this method relative to alternative SMC schemes for sampling change-points; see for example [8].

In this paper, we take a non-standard approach to describing the SMC algorithm of [13] and the random probability distributions it generates. For brevity, we treat the variant of [13] in which resampling is applied at every iteration, but other variants can be considered in the same framework. Our system of notation is central to the understanding of the new MCMC algorithms described in the next section. The first step in this non-standard approach is to move away from the system of indexing displayed in (3) via the trivial equality

$$\sum_{i=1}^N W(X^i) \delta_{X^i} = \sum_{X \in \mathbb{S}} W(X) \delta_X,$$

where  $\mathbb{S} := \{X^1, X^2, \dots, X^N\}$  is the random support. In standard SMC algorithms,  $\mathbb{S}$  is propagated from one iteration to the next by resampling followed by random proposal of new support points. In the resampling step, particles with large weights are duplicated and those with small weights are discarded, according to some stochastic rule. The stochasticity of the proposal and resampling steps inevitably contributes to the variance of (3) and associated estimators.

In the case of the change-point model outlined above, the state space  $|E_n|$  is finite and so the random proposal step can be avoided. Empirical studies, see for example [10], show that this leads to a dramatic decrease in Monte Carlo error. In expressing the algorithm of [13], for each  $n$ , we consider  $\hat{p}(x_n | y_{1:n})$ , an approximation of the filtering distribution, as having random support  $\mathbb{S}_n \subset E_n$ , corresponding to the non-zero members of a collection of random weights  $\mathbf{W}_n^\theta := \{W_n^\theta(x_n)\}_{x_n \in E_n}$ . We write the approximation  $\hat{p}_\theta(x_n | y_{1:n})$  in the form

$$\sum_{x_n \in \mathbb{S}_n} W_n^\theta(x_n) \delta_{x_n}, \quad \sum_{x_n \in \mathbb{S}_n} W_n^\theta(x_n) = 1, \quad (4)$$

where now the members of each  $\mathbb{S}_n$  are not obtained by a random proposal mechanism and are thus written in lower case. We now describe how the random support  $\mathbb{S}_n$  is propagated from one iteration to the next in the algorithm of [13]. At each  $n$  we specify this via a collection of binary-valued random variables,  $\mathbf{S}_n := \{S_n(x_n)\}_{x_n \in E_n}$ , which we refer to as survival indicator variables. The propagation of the random support occurs according to

$$\mathbb{S}_1 = \{0\}, \quad \mathbb{S}_{n+1} = \{x_n \in \mathbb{S}_n : S_n(x_n) = 1\} \cup \{n\}, \quad n = 1, 2, \dots \quad (5)$$

where the probability law of  $\mathbf{S}_n$  is such that, for  $n \leq N$ ,  $|\mathbb{S}_n| = n$ , and for  $n > N$ ,  $|\mathbb{S}_n| = N + 1$  with probability (w.p.) 1 (c.f. the support of the exact filtering distributions which grows deterministically over time (2)). Note that from (5), once a support point is lost, it is never subsequently recovered and the weights associated with it need never be computed. Discarding support points in this manner implies that the computational cost per SMC filtering iteration is bounded uniformly in time. Thus in the SMC forward filtering algorithm below it is implicit that, for any  $n$  and  $x_n$ , once  $S_n(x_n)$  is set to zero, all subsequent survival indicators and weights associated with this point are also set to zero. We return to this issue in section 5.

---

## SMC Forward Filtering

At time  $n = 1$

- Set  $\bar{w}_1^\theta(0) = g_1^\theta(0)$ ,  $W_1^\theta(0) = 1$ ,  $\mathbb{S}_1 = \{0\}$  and  $S_1(0) = 1$ .

At time  $n = 2, 3, \dots, T$

- If  $n - 1 \leq \bar{N}$  set  $C_{n-1} = \infty$ , otherwise set  $C_{n-1}$  to the unique solution of

$$\sum_{x_{n-1} \in \mathbb{S}_{n-1}} 1 \wedge C_{n-1} W_{n-1}^\theta(x_{n-1}) = N.$$

- Set  $\mathcal{I}_{n-1} = \{x_{n-1} \in \mathbb{S}_{n-1} : W_{n-1}^\theta(x_{n-1}) \leq 1/C_{n-1}\}$ .
- Maintain the  $L_{n-1} = n - 1 - |\mathcal{I}_{n-1}|$  support points in  $\mathbb{S}_{n-1}$  which have weights strictly superior to  $1/C_{n-1}$ . If  $|\mathcal{I}_{n-1}| > 0$  resample  $N - L_{n-1}$  times from  $\mathcal{I}_{n-1}$  using the stratified resampling mechanism.
- For each  $x_{n-1}$  having survived the previous resampling steps, set  $S_{n-1}(x_{n-1}) = 1$  and otherwise  $S_{n-1}(x_{n-1}) = 0$ .
- Set

$$\bar{w}_n^\theta(n-1) = g_n^\theta(n-1) \sum_{x_{n-1} \in \mathbb{S}_{n-1}} f_n^\theta(n-1|x_{n-1}) W_{n-1}^\theta(x_{n-1}).$$

- For  $x_{n-1} \in \mathbb{S}_{n-1}$ , set  $\bar{w}_n^\theta(x_{n-1}) = 0$  if  $S_{n-1}(x_{n-1}) = 0$  and otherwise

$$\bar{w}_n^\theta(x_{n-1}) = g_n^\theta(x_{n-1}) f_n^\theta(x_{n-1}|x_{n-1}) \frac{W_{n-1}^\theta(x_{n-1})}{1 \wedge C_{n-1} W_{n-1}^\theta(x_{n-1})}.$$

- Update the support:

$$\mathbb{S}_n = \{x_{n-1} \in \mathbb{S}_{n-1} : S_{n-1}(x_{n-1}) = 1\} \cup \{n\}$$

- For  $x_n \in \mathbb{S}_n$ ,

$$W_n^\theta(x_n) \propto \bar{w}_n^\theta(x_n), \quad \sum_{x_n \in \mathbb{S}_n} W_n^\theta(x_n) = 1.$$

---

The SMC forward filtering scheme of [13] is presented above using our system of notation. A component of the SMC forward filtering algorithm is the stratified resampling procedure [6]. This is also described below. As noted in [13], in this context the stratified resampling procedure has the property of assigning either 1 or 0 offspring to each particle, thus avoiding duplicates. An important point to note about the forward SMC filtering algorithm is that if  $T < N$ , the resampling operation is never applied, in which case the algorithm below performs exact filtering as described in section 4.1. The forward SMC filtering algorithm yields an approximation of the filtering distributions  $\{p_\theta(x_n|y_{1:n})\}_{n=1}^T$  of the form (4) and an estimate of the likelihood  $p_\theta(y_{1:T})$  given by

$$\hat{p}_\theta(y_{1:T}) := \prod_{n=1}^T \left( \sum_{x_n \in \mathbb{S}_n} \bar{w}_n^\theta(x_n) \right). \quad (6)$$


---

## Stratified Resampling

Input:  $\mathcal{I}_{n-1}$  and the corresponding weights.

- Normalise the weights by setting for  $x_{n-1} \in \mathcal{I}_{n-1}$

$$\widehat{W}_{n-1}^\theta(x_{n-1}) \propto W_{n-1}^\theta(x_{n-1}), \quad \sum_{x_{n-1} \in \mathcal{I}_{n-1}} \widehat{W}_{n-1}^\theta(x_{n-1}) = 1$$

and construct the corresponding c.d.f,

$$Q_{n-1}^\theta(x_{n-1}) = \sum_{\{x'_{n-1} \in \mathcal{I}_{n-1} : x'_{n-1} \leq x_{n-1}\}} \widehat{W}_{n-1}^\theta(x'_{n-1}).$$

- Sample  $U_1$  uniformly on  $[0, 1/(N - L_{n-1})]$ , then set  $U_p = U_{p-1} + \frac{1}{N - L_{n-1}}$  for  $p = 2, \dots, N - L_{n-1}$ .
- Each particle in  $\mathcal{I}_{n-1}$  is assigned  $O_{x_{n-1}} \in \{0, 1\}$  offspring where

$$O_{x_{n-1}} = \#\{U_p : Q_{n-1}^\theta(x_{n-1} - 1) \leq U_p \leq Q_{n-1}^\theta(x_{n-1})\}.$$


---

Given the sequence of approximate filtering distributions  $\{\hat{p}_\theta(x_n|y_{1:n})\}_{n=1}^T$ , we can obtain an approximate sample from  $p_\theta(\tau_{1:k}|y_{1:T})$  using a backward sampling recursion [13]. Again, if  $T < N$ , the forward filtering is exact and the below scheme then yields an exact sample from  $p_\theta(\tau_{1:k}|y_{1:T})$ .

---

### SMC Backward Sampling

- Sample  $\tau_1$  from the distribution defined by  $\{W_T^\theta(x_T)\}_{x_T \in \mathbb{S}_T}$  and set  $k = 1$ .
  - While  $\tau_k > 0$ , sample  $\tau_{k+1}$  from the distribution proportional to  $f_{\tau_{k+1}}^\theta(\tau_k|\tau_{k+1})W_{\tau_k}^\theta(\tau_{k+1})$  on  $\mathbb{S}_{\tau_k}$  and set  $k = k + 1$ .
  - Output the sampled sequence  $\tau_{1:k}$  with the indices reversed.
- 

## 5 Particle MCMC

In this section we introduce the two new PMCMC algorithms. Each PMCMC algorithm is the analogue of a standard MCMC algorithm which uses the exact forward filtering/backward sampling methods of section 4.1 to compute  $p_\theta(y_{1:T})$  and sample exactly from  $p_\theta(\tau_{1:k}|y_{1:T})$  at cost  $T^2$  per MCMC iteration. In section 5.3 we establish the validity of new PMCMC algorithms: even though they employ SMC approximations of the exact forward filtering and backward sampling schemes, they are valid MCMC algorithms for sampling from  $p(\theta, \tau_{1:k}|y_{1:T})$ .

### 5.1 Particle Marginal Metropolis Hastings

A standard marginal Metropolis-Hastings (MMH) algorithm samples from  $p(\theta, \tau_{1:k}|y_{1:T})$  using the joint proposal given by

$$q((\theta^*, \tau_{1:k}^*) | (\theta, \tau_{1:k})) = q(\theta^* | \theta) p_{\theta^*}(\tau_{1:k}^* | y_{1:T}) .$$

In this scenario the proposed  $\tau_{1:k}^*$  is “adapted” to the proposed  $\theta^*$  and the resulting acceptance ratio is given by

$$\frac{p_{\theta^*}(y_{1:T}) p(\theta^*) q(\theta | \theta^*)}{p_\theta(y_{1:T}) p(\theta) q(\theta^* | \theta)} . \quad (7)$$

If  $T$  is large, it is too expensive to compute the likelihood terms appearing in this ratio and sample from  $p_{\theta^*}(\tau_{1:k}^* | y_{1:T})$ . We propose the following particle MMH (PMMH) sampler.

---

### PMMH Sampler

Initialisation,  $i = 0$

- Set  $\theta(0)$  arbitrarily.
- Run the SMC forward filtering and backward sampling algorithms with parameter  $\theta(0)$ , yielding an approximation  $\hat{p}_{\theta(0)}(y_{1:T})$  and a change-point configuration  $\tau_{1:k}(0)$ .

For iteration  $i \geq 1$

- Sample  $\theta^* \sim q(\theta^* | \theta(i-1))$
- Run the SMC forward filtering and backward sampling algorithms with parameter  $\theta^*$ , yielding an approximation  $\hat{p}_{\theta^*}(y_{1:T})$  and a change-point configuration  $\tau_{1:k}^*$ .
- Set  $(\theta(i), \tau_{1:k}(i), \hat{p}_{\theta(i)}(y_{1:T})) = (\theta^*, \tau_{1:k}^*, \hat{p}_{\theta^*}(y_{1:T}))$  with probability

$$1 \wedge \frac{\hat{p}_{\theta^*}(y_{1:T}) p(\theta^*) q(\theta(i-1) | \theta^*)}{\hat{p}_{\theta(i-1)}(y_{1:T}) p(\theta(i-1)) q(\theta^* | \theta(i-1))} ,$$

otherwise set  $(\theta(i), \tau_{1:k}(i), \hat{p}_{\theta(i)}(y_{1:T})) = (\theta(i-1), \tau_{1:k}(i-1), \hat{p}_{\theta(i-1)}(y_{1:T}))$ .

---

## 5.2 Particle Gibbs Sampler

It is possible to implement a block Gibbs algorithm which samples from  $p(\theta, \tau_{1:k}|y_{1:T})$  using draws from  $p_\theta(\tau_{1:k}|y_{1:T})$  and  $p(\theta|y_{1:T}, \tau_{1:k})$ . If  $T$  is large, it is too expensive to sample exactly from  $p_\theta(\tau_{1:k}|y_{1:T})$ , so we propose the following particle Gibbs (PG) sampler.

---

### PG Sampler

Initialisation,  $i = 0$

- Set  $\theta(0), X_{1:T}(0)$  arbitrarily.

For iteration  $i \geq 1$

- With parameter  $\theta(i-1)$  and given  $\tau_{1:k}(i-1)$ , run the conditional SMC forward filtering algorithm and then run the SMC backward sampling algorithm to obtain a change-point configuration  $\tau_{1:k}(i)$ .
  - Sample  $\theta(i) \sim p(\theta|y_{1:T}, \tau_{1:k}(i))$ .
- 

This PG algorithm relies on the conditional forward filtering algorithm which is now described.

---

### Conditional SMC Forward Filtering

Input: change-point locations  $\tau_{1:k}$ , and  $\theta$ .

At time  $n = 1$

- Set  $\bar{w}_1^\theta(0) = g_1^\theta(0), W_1^\theta(0) = 1, \mathbb{S}_1 = \{0\}, S_1(0) = 1$  and  $\kappa = 0$ .

At time  $n = 2, \dots, T$

- If  $n-1 \leq N$  set  $C_{n-1} = \infty$ , otherwise set  $C_{n-1}$  to the unique solution of

$$\sum_{x_{n-1} \in \mathbb{S}_{n-1}} 1 \wedge C_{n-1} W_{n-1}^\theta(x_{n-1}) = N.$$

- If  $\tau_{\kappa+1} < n-1$  set  $\kappa = \kappa + 1$ .
- Set  $\mathcal{I}_{n-1} = \{x_{n-1} \in \mathbb{S}_{n-1} : W_{n-1}^\theta(x_{n-1}) \leq 1/C_{n-1}\}$ .
- If  $\tau_\kappa \notin \mathcal{I}_{n-1}$ , maintain the  $L_{n-1} = n-1 - |\mathcal{I}_{n-1}|$  support points in  $\mathbb{S}_{n-1}$  which have weights strictly superior to  $1/C_{n-1}$  (which includes  $\tau_\kappa$ ), then if  $|\mathcal{I}_{n-1}| > 0$  resample  $N - L_{n-1}$  times from  $\mathcal{I}_{n-1}$  using the stratified resampling mechanism.
- If  $\tau_\kappa \in \mathcal{I}_{n-1}$ , maintain the  $L_{n-1} = n-1 - |\mathcal{I}_{n-1}|$  support points in  $\mathbb{S}_{n-1}$  which have weights strictly superior to  $1/C_{n-1}$  (which does not include  $\tau_\kappa$ ), then resample  $N - L_{n-1}$  times from  $\mathcal{I}_{n-1}$  using the conditional stratified resampling mechanism.
- For each  $x_{n-1}$  having survived the previous resampling steps, set  $S_{n-1}(x_{n-1}) = 1$  otherwise  $S_{n-1}(x_{n-1}) = 0$ .
- Set

$$\bar{w}_n^\theta(n-1) = g_n^\theta(n-1) \sum_{x_{n-1} \in \mathbb{S}_{n-1}} f_n^\theta(n-1|x_{n-1}) W_{n-1}^\theta(x_{n-1}).$$

For  $x_{n-1} \in \mathbb{S}_{n-1}$ , set  $\bar{w}_n^\theta(i) = 0$  if  $S_{n-1}(x_{n-1}) = 0$  and otherwise

$$\bar{w}_n^\theta(x_{n-1}) = g_n^\theta(x_{n-1}) f_n^\theta(x_{n-1}|x_{n-1}) \frac{W_{n-1}^\theta(x_{n-1})}{1 \wedge C_{n-1} W_{n-1}^\theta(x_{n-1})}.$$

- Update the support:

$$\mathbb{S}_n = \{x_{n-1} \in \mathbb{S}_{n-1} : S_{n-1}(x_{n-1}) = 1\} \cup \{n\}$$

- For  $x_n \in \mathbb{S}_n$ ,

$$W_n^\theta(x_n) \propto \bar{w}_n^\theta(x_n), \quad \sum_{x_n \in \mathbb{S}_n} W_n^\theta(x_n) = 1.$$


---

A component of the conditional SMC forward filtering procedure is the following conditional stratified resampling scheme.

---

### Conditional Stratified Resampling

Input:  $\mathcal{I}_{n-1}$ , the corresponding weights and an index  $\kappa$ .

- Normalise the weights by setting for  $x_{n-1} \in \mathcal{I}_{n-1}$

$$\widehat{W}_{n-1}^\theta(x_{n-1}) = \frac{W_{n-1}^\theta(x_{n-1})}{\sum_{x'_{n-1} \in \mathcal{I}_{n-1}} W_{n-1}^\theta(x'_{n-1})}$$

and construct the corresponding c.d.f,

$$Q_{n-1}^\theta(x_{n-1}) = \sum_{\{x'_{n-1} \in \mathcal{I}_{n-1} : x'_{n-1} \leq x_{n-1}\}} \widehat{W}_{n-1}^\theta(x'_{n-1}).$$

- Sample  $U^*$  uniformly on  $[Q_{n-1}^\theta(\kappa - 1), Q_{n-1}^\theta(\kappa)]$ , set  $U_1 = U^* - \frac{\lfloor (N - L_{n-1})U^* \rfloor}{N - L_{n-1}}$  and  $U_p = U_{p-1} + \frac{1}{N - L_{n-1}}$  for  $p = 2, \dots, N - L_{n-1}$ .
- Each particle is assigned  $O_{x_{n-1}} \in \{0, 1\}$  offspring where

$$O_{x_{n-1}} = \# \{U_p : Q_{n-1}^\theta(x_{n-1} - 1) \leq U_p \leq Q_{n-1}^\theta(x_{n-1})\}.$$


---

### 5.3 Validity of the Algorithms

We first establish an expression for the distribution of the random variables  $\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_{T-1}$  generated through the SMC forward filtering algorithm. By construction, we have

$$\mathbf{S}_n | (\mathbf{W}_n^\theta = \mathbf{w}_n^\theta) \sim r_\theta^N(\cdot | \mathbf{w}_n^\theta). \quad (8)$$

This density is parameterized by  $N$  because for all  $n \geq N$ ,

$$\sum_{x_n \in \mathbb{S}_n} S_n(x_n) = N, \quad \text{w.p. 1.}$$

We will not need an explicit expression for the density (8), but from the definition of the optimal resampling mechanism [10], we know that it has the following marginal property for all  $x_n \in \{0, 1, \dots, n-1\}$

$$r^N(S_n(x_n) = 1 | \mathbf{w}_n^\theta) = 1 \wedge c_n w_n^\theta(x_n), \quad (9)$$

which implies that

$$r^N(S_n(x_n) = 1 | w_n^\theta(x_n) = 0) = 0,$$

so combined with Eq. (8) we see that for any  $n > 0$  and  $x_n \in \{1, \dots, n-1\}$ , conditional on the event that  $W_{n-1}^\theta(x_n) = 0$ , at any subsequent  $n$ ,  $W_n^\theta(x_n) = 0$  w.p. 1. Thus the corresponding subsequent survival indicators and weights need never be simulated or stored. To summarize the law of the forward filtering algorithm, we can write the distribution of  $\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_{T-1}$  on  $\prod_{n=1}^{T-1} \{0, 1\}^n$  as

$$\psi_\theta^N(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{T-1}) = \prod_{n=1}^{T-1} r_\theta^N(\mathbf{s}_n | \mathbf{w}_n^\theta). \quad (10)$$

The weights  $\mathbf{W}_n^\theta$  being just a deterministic function of  $\mathbf{S}_1, \dots, \mathbf{S}_{n-1}$ , it is not necessary to introduce them as arguments of  $\psi_\theta^N$ .

The key to establishing the validity of the PMCMC algorithms is to define the following joint probability density for  $\theta, \tau_{1:k}$  and  $\mathbf{S}_1, \dots, \mathbf{S}_{T-1}$  on  $\Theta \times \mathcal{T}_T \times \prod_{n=1}^{T-1} \{0, 1\}^n$ ,

$$\begin{aligned} \pi^N(\theta, \tau_{1:k}, \mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{T-1}) &:= p(\theta, \tau_{1:k} | y_{1:T}) \psi_\theta^N(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{T-1}) \\ &\cdot \frac{\prod_{j=1}^k \prod_{n=\tau_{j-1}+1}^{\tau_j} \mathbb{I}[s_n(\tau_{j-1}) = 1]}{\prod_{j=1}^k \prod_{n=\tau_{j-1}+1}^{\tau_j} r_n(s_n(\tau_{j-1}) = 1 | \mathbf{w}_n^\theta)} \\ &\cdot \frac{\prod_{n=\tau_k+1}^{T-1} \mathbb{I}[s_n(\tau_k) = 1]}{\prod_{n=\tau_k+1}^{T-1} r_n(s_n(\tau_k) = 1 | \mathbf{w}_n^\theta)}. \end{aligned} \quad (11)$$

By construction (11) admits  $p(\theta, \tau_{1:k} | y_{1:T})$  as a marginal. We show in the following theorems that the PMMH and PG algorithms are just standard MCMC updating schemes targeting (11). Furthermore, the convergence to  $p(\theta, \tau_{1:k} | y_{1:T})$  of the distribution of the samples of  $\theta, \tau_{1:k}$  they generate is inherited from the corresponding standard MCMC schemes. Proofs are given in the appendix.

We first deal with the PMMH algorithm and consider the following assumption.

- (A1) The standard MMH sampler of target density  $p(\theta, \tau_{1:k} | y_{1:T})$  is irreducible and aperiodic (and hence converges for almost all starting points).

We have the following result.

**Theorem 1**

1. For any  $N \geq 1$ , the PMMH algorithm is a Metropolis-Hastings sampler on the space  $\Theta \times \mathcal{T}_T \times \prod_{n=1}^{T-1} \{0, 1\}^n$  with target density  $\pi^N(\theta, \tau_{1:k}, \mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{T-1})$  and using the proposal

$$q(\theta^* | \theta) \cdot \psi_{\theta^*}^N(\mathbf{s}_1^*, \dots, \mathbf{s}_{T-1}^*) \cdot w_T^{\theta^*}(\tau_k^*) \prod_{j=1}^{k^*} \left\{ \frac{f_{\tau_j^*+1}^{\theta^*}(\tau_j^* | \tau_{j-1}^*) w_{\tau_j^*}^{\theta^*}(\tau_{j-1}^*)}{\sum_{x \in \mathcal{S}_{\tau_j}} f_{\tau_j^*+1}^{\theta^*}(\tau_j^* | x) w_{\tau_j^*}^{\theta^*}(x)} \right\}. \quad (12)$$

2. If additionally (A1) holds, the PMMH sampler generates a sequence  $\{\theta(i), \tau_{1:k}(i)\}$  whose marginal distributions  $\mathcal{L}^N((\theta(i), \tau_{1:k}(i)) \in \cdot)$  satisfies

$$\|\mathcal{L}^N((\theta(i), \tau_{1:k}(i)) \in \cdot) - p(\cdot | y_{1:T})\|_{tv} \rightarrow 0 \text{ as } i \rightarrow \infty.$$

We now turn to the PG sampler and consider the following assumption on the corresponding standard Gibbs sampler.

- (A2) The standard block Gibbs sampler which samples from  $p_\theta(\tau_{1:k} | y_{1:T})$  and  $p(\theta | \tau_{1:k}, y_{1:T})$  is irreducible and aperiodic (and hence converges for almost all starting points).

We have the following result.

**Theorem 2**

1. For any  $N \geq 2$ , the PG sampler defines a transition kernel on the space  $\Theta \times \mathcal{T}_T \times \prod_{n=1}^{T-1} \{0, 1\}^n$  of invariant density  $\pi^N(\theta, \tau_{1:k}, \mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{T-1})$ . One PG iteration is equivalent to sampling from the sequence of conditional distributions  $\pi^N(\mathbf{s}_1, \dots, \mathbf{s}_{T-1} | \theta, \tau_{1:k})$ ,  $\pi^N(\tau_{1:k} | \theta, \mathbf{s}_1, \dots, \mathbf{s}_{T-1})$  and  $\pi^N(\theta | \tau_{1:k})$ .
2. If additionally (A2) holds, then the PG sampler generates a sequence  $\{\theta(i), \tau_{1:k}(i)\}$  whose marginal distributions  $\mathcal{L}^N((\theta(i), \tau_{1:k}(i)) \in \cdot)$  satisfies

$$\|\mathcal{L}^N((\theta(i), \tau_{1:k}(i)) \in \cdot) - p(\cdot | y_{1:T})\|_{tv} \rightarrow 0 \text{ as } i \rightarrow \infty.$$

## 6 Simulation Study

### 6.1 Well-Log Data

We consider a piecewise constant model which was used in [12] to analyze well-log data. The prior distribution  $h_\theta$  is geometric with unknown success probability  $p$ , over which a flat Beta hyper-prior is placed. Given change-points  $\tau_{j-1:j}$  the observations  $y_{\tau_{j-1}+1:\tau_j}$  in the  $j^{\text{th}}$  segment are i.i.d.  $\mathcal{N}(\mu_j, \sigma^2)$ , where  $\mu_j$  is the mean associated with segment  $j$ . Given  $\tau_{1:k}$  and hyper parameters  $\eta$  and  $\alpha$ , the segment means,  $\{\mu_j\}_{j=1}^{k+1}$  are i.i.d.  $\mathcal{N}(\eta, \sigma^2 \alpha^2)$ . Finally, uninformative, improper priors are placed over the hyper-parameters:  $p(\eta) \propto 1$ ,  $p(\sigma) \propto 1/\sigma$  and  $p(\alpha) \propto 1/\alpha$ . The parameter vector is then  $\theta = [\eta \ p \ \sigma \ \alpha]^T$ . We refer to [12] for the expression of the likelihood of observations in one segment, given values of  $\eta$ ,  $\alpha$  and  $\sigma$ , with the corresponding value of  $\mu_j$  integrated out.

The data, originating from [18], are shown in figure 1 and consist of  $T = 4050$  measurements of the nuclear magnetic response of underground rocks. As in [12], a small number of outlying observations were manually removed from the data set. For comparison with PMCMC, we considered the block Gibbs sampler which samples from  $p_\theta(\tau_{1:k} | y_{1:T})$  (using the exact method of [13]),  $p(\sigma, p, \{\mu_j\}_{j=1}^{k+1} | y_{1:T}, \theta, \tau_{1:k})$  and

$p(\eta, \alpha | y_{1:T}, \theta, \tau_{1:k}, \sigma, p, \{\mu_j\}_{j=1}^{k+1})$ . In [12] it was shown that for the same data set and model, an algorithm of this form significantly outperformed the MCMC scheme of [16].

The above block Gibbs sampler was compared with the PG algorithm obtained by substituting in place of the draw from  $p_\theta(\tau_{1:k} | y_{1:T})$  a conditional SMC forward filtering/backward sampling step. Figure 2 shows autocorrelation plots for each component of  $\theta$ , for the block Gibbs and PG algorithm with  $N = 50$ . For this data record  $T = 4050$ , so the computational cost of the PG algorithm is an order of magnitude less than that of the block Gibbs sampler. The auto-correlation plots indicate that the block Gibbs sampler is extremely efficient for this model, with the autocorrelation dropping to zero well before lag 10. The performance of the PG algorithm is almost identical to that of the block Gibbs. Using other values of  $N$  resulted in auto-correlation curves which were too similar to be displayed clearly on the same axes. For  $N < 50$  it was found that there was a small difference in the corresponding approximations of the posterior marginal for the parameter  $p$  as the PG scheme did not fully explore its support (not shown).

Figure 1 shows estimated marginal posterior probabilities of change-point occurrences, from the block Gibbs and the PG algorithm with  $N = 50$ . These posterior probabilities are indistinguishable to the eye and the results are very similar to those reported in [12] with any discrepancies likely to be attributable to which outlying observations were removed ([12] does not state how this was performed).

Figure 2 also shows approximations of the marginal posterior distributions for each of the four parameters obtained with the PG algorithm. Identical histograms were obtained from the block Gibbs output. The modes of these posterior marginals coincide with the results reported in [12].

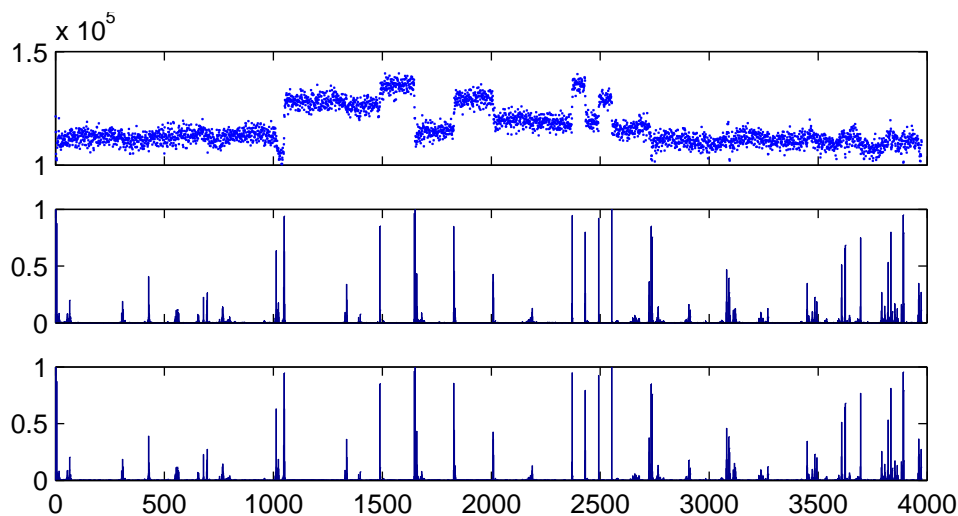


Figure 1: Well-log example. Top: well-log data. Approximate marginal posterior probabilities of a change-point at each time index obtained from PG sampler with  $N = 50$  (middle) and the Block Gibbs sampler (bottom).

## 6.2 Coal Mining Disasters

We consider a piece-wise constant Poisson intensity model for the classic Coal Mining Disaster data set of [15]. This data set consists of the dates of 191 disasters between 1851 and 1962. The data were analysed in [14] via a reversible-jump MCMC algorithm for a continuous time model and in [12] using the exact sampling methods.

Following [12] we discretize time and form observations by counting the number of disasters each week. This yields  $T = 5844$  observations. Given change-points  $\tau_{j-1:j}$  the observations  $y_{\tau_{j-1}+1:\tau_j}$  in the  $j^{\text{th}}$  segment are i.i.d.  $\mathcal{P}o(\lambda_j)$ , where  $\lambda_j$  is the intensity associated with segment  $j$ . Given  $\tau_{1:k}$  the segment intensities,  $\{\lambda_j\}_{j=1}^{k+1}$  are i.i.d.  $\mathcal{G}(1, 200/7)$ . As opposed to choosing the prior distribution on inter-change-point times, in [12] a Poisson prior was placed over the number of change-points. The intensity parameter of this Poisson prior was treated as fixed and chosen to give a mean of 3 change-points over the duration of the observation record. Given the number of change-points  $k$ , the change-point locations were a-priori given by the even order statistics of  $2k + 1$  uniform draws on  $\{1, \dots, T - 1\}$ . We investigate a different approach via the

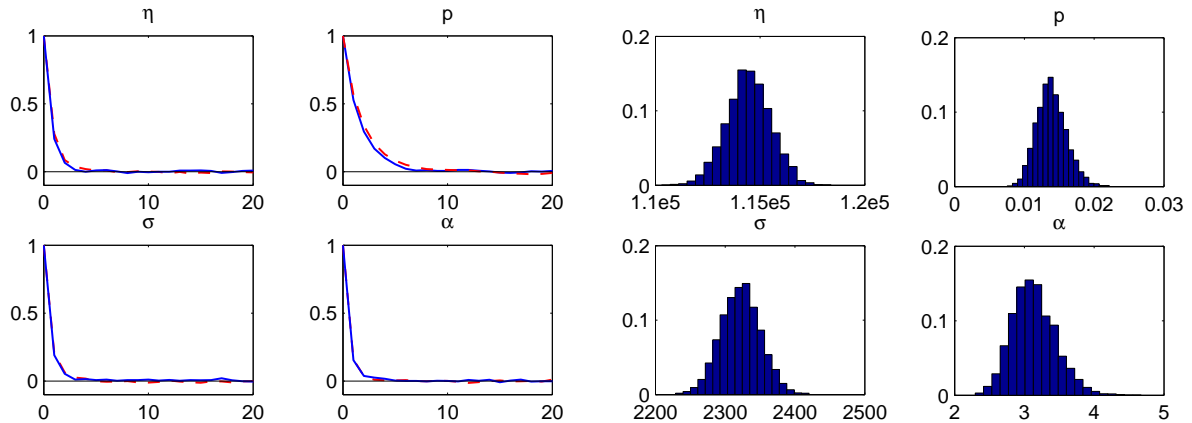


Figure 2: Well-log example. The left pane shows auto-correlation plots. Solid: Block Gibbs. Dashed: PG with  $N = 50$ . The right pane shows approximate posterior marginals obtained from the PG algorithm. Identical histograms were obtained using the Block Gibbs sampler.

distribution on inter-change-point times, with parameters to be inferred from the data. In our model the prior density  $h_\theta$  is negative binomial with parameters  $r$  and  $p$ , over which gamma and beta hyper-priors were placed, respectively. The parameters of the hyper-priors were chosen so as to provide some penalisation of very short segments whilst still allowing flexibility: for the gamma hyper-prior on  $r$  the shape parameter was 10 and the scale parameter 1; for the beta prior on  $p$  the parameters were  $(1, 10)$ . In this model the parameter is  $\theta = [r \ p]^T$ .

We compared the PMMH algorithm with a standard MMH algorithm employing exact computation of  $p_\theta(y_{1:T})$  over 30,000 MCMC iterations. Random walk proposals were made over  $r$  and  $p$  with standard deviations of 1 and 0.005, respectively. These values were chosen after a preliminary run and resulted in an average acceptance probability of 0.48 for the MH algorithm.

Figure 3 shows approximations of the marginal and joint posterior distributions of the two parameters from the output of the PMMH algorithm with  $N = 200$ . Identical histograms were obtained using the standard MMH algorithm. The figure also shows auto-correlation plots for the two parameters. In these plots, the solid line corresponds to the exact MMH algorithm. The dashed and dash-dot lines correspond to the PMMH algorithm with  $N = 200$  and  $N = 50$ , respectively. In terms of auto-correlation, there is very little difference between the exact MMH and PMMH with  $N = 200$ . For the parameter  $r$ , the difference between the dash-dot, dashed and solid lines is more noticeable, but still not very large. Figure 4 shows the acceptance rate in the PMMH algorithm against  $N$ . Only as the number of particles falls below 200 does the acceptance rate fall significantly. For this data set,  $N = T = 5844$ , corresponds to the exact MMH algorithm; with  $N$  an order of magnitude less than this value the performance of the PMMH algorithm is very good.

Figure 4 also shows the approximate marginal posterior distributions over the number of change-points and over the location of change-points, given that there are 2, obtained from the PMMH algorithm with  $N = 200$ . Again, identical results were obtained with the MMH algorithm. There are some similarities and interesting differences between these results and those obtained in [12] under the Poisson change-point prior with fixed parameter. Firstly, the marginal over the number of change-points shown in figure 4 exhibits a stronger mode at 2 than that obtained [12], but is otherwise similar. Secondly, the conditional posterior for the change-point locations, given there are 2, exhibits two strong modes in the same locations as found in [12]. However the results in figure 4, obtained using the PMMH algorithm with the negative binomial inter-change-point prior, exhibit a third, weaker mode around time 4900. This third mode was not reported in [12]. In [12] it was not reported what kernel bandwidth was used, so precise comparisons are difficult, for even with a relatively large bandwidth, the third mode shown in 4 was evident. This shows how the choice of prior can values of parameters can affect inferences drawn about change-point locations.

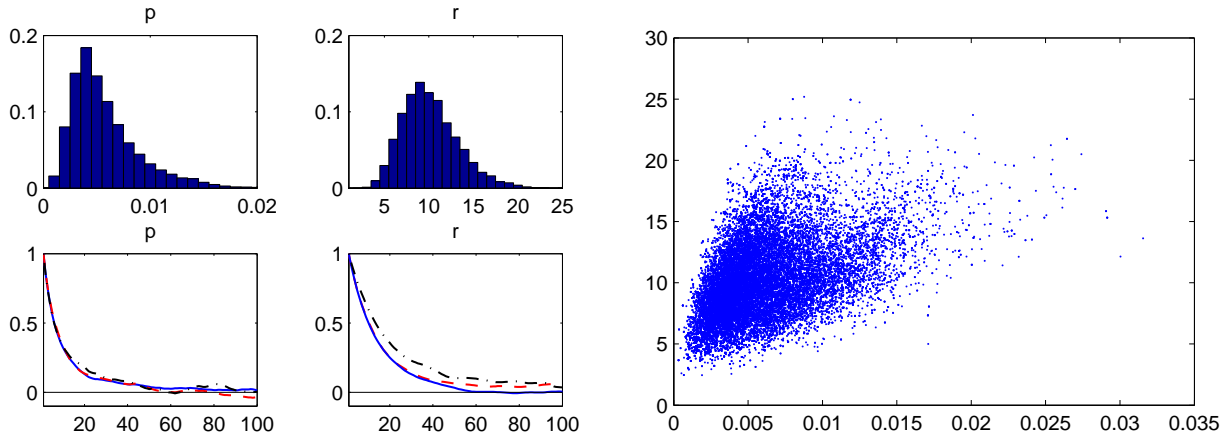


Figure 3: Coal Mining disasters example. Left: Approximate posterior marginals using PMMH and auto-correlation plots for the two parameters. In auto-correlation plots, solid: Exact MMH. dashed: PMMH with  $N = 200$ , dash-dot: PMMH with  $N = 50$ .

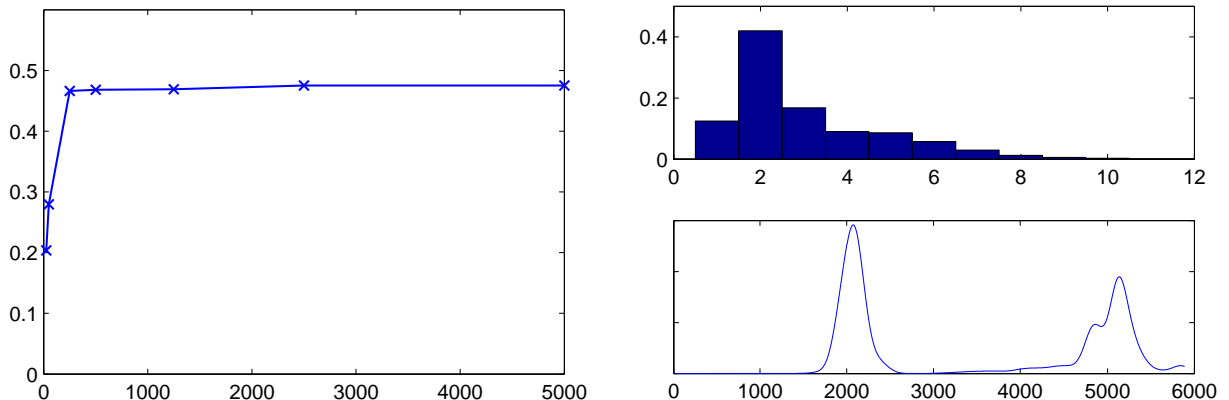


Figure 4: Coal Mining Disasters example. Left: Acceptance rate vs.  $N$ . Right top: Approximate posterior over number of change-points from PMMH algorithm. Right bottom: kernel smoothed posterior marginal for change-point locations conditional on 2 change-points.

## 7 Discussion

We have proposed PMCMC algorithms for multiple change-point models which rely on the efficient SMC method proposed in [13] to approximate the filtering distributions and likelihood. These PMCMC algorithms have a cost per MCMC iteration of the order  $N T$ , where  $N$  is the number of particles in the SMC algorithm, compared to  $T^2$  for the “exact” MCMC algorithms relying on the exact filtering distributions and likelihood. We have demonstrated experimentally that these PMCMC algorithms perform remarkably well compared to the “exact” algorithms even for  $N$  an order of magnitude smaller than  $T$ . This is an attractive feature for performing Bayesian inference in multiple change-point models with long data records.

## 8 Appendix

### Proof of Theorem 1.

1. By construction, the PMMH algorithm makes proposals from (12): the left-hand term is the proposal distribution for the new parameter  $\theta^*$ ; the middle term is the law of the SMC forward filtering algorithm with this parameter; the right-hand term is the conditional probability of obtaining a sequence of change-points from the SMC backward sampling algorithm. Furthermore (12) is a distribution over the same space as (11).

From the definition of the SMC forward filtering algorithm, for two consecutive change-point locations

$\tau_j > \tau_{j-1} > 0$ , on the event  $\prod_{n=\tau_{j-1}+1}^{\tau_j} S_n(\tau_{j-1}) = 1$ , we have the expansion of weight  $w_{\tau_j}^\theta(\tau_{j-1})$  given in (14). By using analogous expansions of  $w_T^\theta(\tau_k)$ ,  $w_{\tau_1}^\theta(0)$  on the corresponding events and by using (13), (6) and (9) it follows that the acceptance probability given in the PMMH Sampler algorithm is precisely that of an MH sampler targeting the extended target distribution (11) and proposing from (12).

2. Proof of the second component of the theorem is a direct consequence of the arguments in Theorem 1 in [2] and **(A1)**.

**Proof of Theorem 2.**

1. For the first step stated in the theorem, it is straightforward to check that running the conditional SMC forward filtering algorithm is equivalent to sampling from  $\pi^N(\mathbf{s}_1, \dots, \mathbf{s}_{T-1} | \theta, \tau_{1:k})$ .

Now consider the second step. Recalling the definition of the extended target distribution (11), using again on the event  $\prod_{n=\tau_{j-1}+1}^{\tau_j} S_n(\tau_{j-1}) = 1$ , the expansion (14), and using (13) and (9) we conclude that

$$\pi^N(\tau_{1:k} | \theta, \mathbf{s}_1, \dots, \mathbf{s}_{T-1}) = w_T^\theta(\tau_k) \prod_{j=1}^k \frac{f_{\tau_{j+1}}^\theta(\tau_j | \tau_{j-1}) w_{\tau_j}^\theta(\tau_{j-1})}{\sum_{x \in \mathcal{S}_{\tau_j}} f_{\tau_{j+1}}^\theta(\tau_j | x) w_{\tau_j}^\theta(x)},$$

where the right-hand term is the conditional probability of sampling the change-point configuration  $\tau_{1:k}$  using the SMC backward sampling algorithm, given the outcome of the forward sampling.

For the third step it is direct that  $p(\theta | y_{1:T}, \tau_{1:k}) = \pi^N(\theta | \tau_{1:k})$ .

2. We focus on establishing irreducibility and aperiodicity of the transition probability of the PG algorithm. We denote by  $\mathcal{L}_G$  the law of the standard Gibbs sampler to which assumption **(A2)** applies and  $\mathcal{L}_{PG}^N$  the law of the PG sampler using  $N$  particles.

Let  $A \times B \times C \in \mathcal{B}(\Theta) \times \mathcal{B}(\mathcal{T}_T) \times \mathcal{B}(\prod_{n=1}^{T-1} \{0, 1\}^n)$  be such that  $\pi^N(\theta \in A, \tau_{1:k} \in B, \mathbf{s}_1, \dots, \mathbf{s}_{T-1} \in C) > 0$ . It follows from (11) that  $p((\theta, \tau_{1:k}) \in A \times B | y_{1:T}) > 0$  and then from irreducibility of the corresponding block Gibbs sampler (assumption **(A2)**) there exists a finite  $i$  such that  $\mathcal{L}_G((\theta(i), \tau_{1:k}(i)) \in A \times B) > 0$ .

From the specification of the conditional SMC forward filtering scheme, for any  $\theta \in \Theta$ ,  $N \geq 2$ , given any  $\tau_{1:k}$  and for any time step, any particle which has positive weight immediately before resampling has a positive probability of surviving that resampling step. It follows that for any  $n = 1, \dots, T-1$ , any point in the support of  $p_\theta(x_n | y_{1:n})$  has positive probability of being assigned a positive weight at time  $n$ . It follows from the definitions of  $X_{1:T}$  that any point in the support of  $p_\theta(\tau_{1:k} | y_{1:n})$  has positive probability of being selected in the backward sampling. Then the  $A \times B$  from above is marginally an accessible set of the PG sampler for the same  $i$ : i.e.  $\mathcal{L}_{PG}^N((\theta(i), X_{1:T}(i)) \in A \times B) > 0$ . Furthermore, as the conditional forward SMC filtering corresponds to drawing from  $\pi^N(\mathbf{s}_1, \dots, \mathbf{s}_{T-1} | \theta, \tau_{1:k})$ ,

$$\mathcal{L}_{PG}^N((\theta(i+1), \tau_{1:k}(i+1), \mathbf{s}_1(i+1), \dots, \mathbf{s}_{T-1}(i+1)) \in A \times B \times C) > 0$$

and irreducibility follows. Furthermore, aperiodicity of the PG sampler holds by contradiction: if the PG sampler were periodic, then the Gibbs sampler would be too; this violates **(A2)**.

## References

- [1] Andrieu, C., Doucet, A. and Holenstein, R. (2010) Particle Markov chain Monte Carlo methods (with discussion). *J. Roy. Stat. Soc. B*, **72**, 1-33.
- [2] Andrieu, C. and Roberts, G.O. (2009) The pseudo-marginal approach for efficient computation. *Annals Statist.*, **37**, 697-725.
- [3] Barry, D. and Hartigan, J. A., (1992) Product partition models for change point problems. *Ann. Statist.*, **20**, 260–279.
- [4] Barry, D. and Hartigan, J. A., (1993) A Bayesian analysis for change point problems. *J. Am. Stat. Soc.*, **88**(421), 309–319.

---


$$\begin{aligned}
& p_{\theta}(\tau_{1:k}, y_{1:T}) \\
&= \left\{ \left[ \mathbb{I}[\tau_k > \tau_{k-1} + 1] \left( \prod_{n=\tau_k+1}^T g_n^{\theta}(\tau_k) f_n^{\theta}(\tau_k | \tau_k) \right) + \mathbb{I}[\tau_k = \tau_{k-1} + 1] g_{\tau_k+1}^{\theta}(\tau_k) f_{\tau_k+1}^{\theta}(\tau_k | \tau_{k-1}) \right] \right\} \\
&\cdot \prod_{j=2}^k \left\{ \left[ \mathbb{I}[\tau_j > \tau_{j-1} + 1] \left( \prod_{n=\tau_{j-1}+1}^{\tau_j} g_n^{\theta}(\tau_{j-1}) f_n^{\theta}(\tau_{j-1} | \tau_{j-1}) \right) + \mathbb{I}[\tau_j = \tau_{j-1} + 1] g_{\tau_{j-1}+1}^{\theta}(\tau_{j-1}) f_{\tau_{j-1}+1}^{\theta}(\tau_{j-1} | \tau_{j-2}) \right] \right\} \\
&\cdot \left\{ \mathbb{I}[\tau_1 > 1] \left( \prod_{n=1}^{\tau_1} g_n^{\theta}(0) f_n^{\theta}(0 | 0) \right) \right\}. \tag{13}
\end{aligned}$$


---

$$\begin{aligned}
& w_{\tau_j}^{\theta}(\tau_{j-1}) \\
&= \left[ \mathbb{I}[\tau_j > \tau_{j-1} + 1] \left( \frac{g_{\tau_j}^{\theta}(\tau_{j-1}) f_{\tau_j}^{\theta}(\tau_{j-1} | \tau_{j-1})}{\sum_{x \in \mathbb{S}_{\tau_j}} \bar{w}_{\tau_j}^{\theta}(x)} \prod_{n=\tau_{j-1}+2}^{\tau_j-1} \frac{g_n^{\theta}(\tau_{j-1}) f_n^{\theta}(\tau_{j-1} | \tau_{j-1})}{1 \wedge c_n w_n^{\theta}(\tau_{j-1})} \frac{1}{\sum_{x \in \mathbb{S}_n} \bar{w}_n^{\theta}(x)} \right) + \mathbb{I}[\tau_j = \tau_{j-1} + 1] \right] \\
&\cdot \frac{g_{\tau_{j-1}+1}^{\theta}(\tau_{j-1}) \sum_{x \in \mathbb{S}_{\tau_{j-1}}} f_{\tau_{j-1}+1}^{\theta}(\tau_{j-1} | x) w_{\tau_{j-1}}^{\theta}(x)}{1 \wedge c_{\tau_{j-1}+1} w_{\tau_{j-1}+1}^{\theta}(\tau_{j-1})} \frac{1}{\sum_{x \in \mathbb{S}_{\tau_{j-1}+1}} \bar{w}_{\tau_{j-1}+1}^{\theta}(x)}. \tag{14}
\end{aligned}$$


---

- [5] Carlin, B.P., Gelfand, A.E. and Smith A.F.M., (1992) Hierarchical Bayesian Analysis of change-point problems. *Appl. Statist.*, **41**, 389–405.
- [6] Carpenter, J., Clifford, P. and Fearnhead, P. (1999) An improved particle filter for non-linear problems. *IEE Proc. F*, **146**, 2–7.
- [7] Chib, S. (1998) Estimation and comparison of multiple change-point models. *J. Econometr.*, **86**, 221–241.
- [8] Chopin, N. (2007) Dynamic detection of change points in long time series. *Ann. Inst. Statist. Math.*, **59**, 349–366.
- [9] Doucet, A., de Freitas, J.F.G. and Gordon, N.J. (eds.) (2001) *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.
- [10] Fearnhead, P. and Clifford, P. (2003) On-line inference for hidden Markov models via particle filters. *J. Roy. Stat. Soc. B*, **65**, 887–899.
- [11] Fearnhead, P. (2005) Exact Bayesian curve fitting and signal segmentation. *IEEE Trans. Sig. Proc.*, **53**(6), 2160–2166.
- [12] Fearnhead, P. (2006) Efficient and exact Bayesian inference for multiple changepoint problems. *Statistics and Computing*, **16**, 203–213.
- [13] Fearnhead, P., and Liu, Z. (2007). On-line inference for multiple changepoint problems. *J. R. Statist. Soc. B*, **69**(4), 589–605.
- [14] Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**(4), 711–732.
- [15] Jarrett R. G. (1979). A note on the intervals between coal-mining disasters. *Biometrika*, **66**, 191-193.
- [16] Lavielle M. and Lebarbier E. (2001). An application of MCMC methods for the multiple change-points problem. *Signal Processing* **81**, 39–53.
- [17] Liu, J. S., Wong, W. H. and Kong, A. (1994). Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes. *Biometrika*, **81**(1), 27–40.

- [18] Ó Ruanaidh J. J. K. and Fitzgerald W. J. 1996. *Numerical Bayesian Methods Applied to Signal Processing*. New York: Springer.
- [19] Stephens, D. A. (1994) Bayesian retrospective multiple-change-point identification. *Appl. Statist.*, **43**, 159—178.