Appendices to *Adaptive sequential Monte Carlo for multiple changepoint analysis*

Nicholas A. HEARD and Melissa J. M. TURCOTTE

A Exact simulation

The sequential Monte Carlo (SMC) algorithm for changepoint problems introduced in the article requires changepoint samples $\tilde{\boldsymbol{\tau}}_n$ for each update interval $(t_{n-1}, t_n]$ to be drawn from the local posterior distribution based on the data observed in $(t^*_{n-1}, t_n]$, $\pi_{(t_{n-1},t_n]}(\tilde{\boldsymbol{\tau}}_n|y((t^*_{n-1}, t_n]))$. For $t^*_{n-1} < t_n$, it will typically not be possible to simulate directly from $\pi_{(t_{n-1},t_n]}(\tilde{\boldsymbol{\tau}}_n|y((t^*_{n-1}, t_n]))$, and it is suggested that reversible jump Markov chain Monte Carlo (RJMCMC) simulation be used to provide approximate samples from $\pi_{(t_{n-1},t_n]}(\tilde{\boldsymbol{\tau}}_n|y((t^*_{n-1}, t_n]))$. This approximation prevents the SMC algorithm being exact. It is therefore important to investigate the magnitude of inferential bias caused by this approximation, to see if it is sufficiently small to justify use of the algorithm in place of higher variance, exact SMC methods.

A.1 Alternative prior

A simple way to allow exact sampling of changepoints from $\pi_{(t_{n-1},t_n]}(\tilde{\boldsymbol{\tau}}_n|y((t^*_{n-1}, t_n]))$ (and thus enable a comparison) is to modify the prior on the changepoints. The alternative prior (2) in the main article generalizes the Poisson process prior (1) by assuming there is a constant intensity ν until a changepoint occurs, at which point the intensity drops to 0 for s units of time, after which the process resumes with constant intensity ν. Thus changepoints are forced to be at least a distance s apart. The Poisson process prior (1) is simply the case $s = 0$, whereas crucially *any* choice $s > 0$ implies a bounded possible number of changepoints in any update interval $(t_{n-1}, t_n]$. In particular, if s is chosen such that $\forall n, s \geq t_n - t_{n-1}$, this means that there can be at most one changepoint in any update interval. So *a priori*, if $s \geq t_n - t_{n-1}$ then

$$\begin{aligned} p_{(t_{n-1},t_n]}(\tilde{k}_n = 0) &= e^{-\nu(t_n - t_{n-1})}, \\ p_{(t_{n-1},t_n]}(\tilde{k}_n = 1) &= 1 - e^{-\nu(t_n - t_{n-1})}, \\ p_{(t_{n-1},t_n]}(\tilde{\tau}_1|\tilde{k}_n = 1) &= \mathbb{I}_{(t_{n-1},t_n]}(\tilde{\tau}_1)/(t_n - t_{n-1}). \end{aligned} \tag{A.1}$$

A.2 Rejection sampling

To perform exact sampling from $\pi_{(t_{n-1},t_n]}(\tilde{\boldsymbol{\tau}}_n|y((t^*_{n-1},t_n]))$ when adopting the prior (A.1), rejection sampling can be deployed. First, proposed values of $\tilde{\boldsymbol{\tau}}_n$ are sampled from the prior, and then are accepted as exact draws from the target with probability

$$\frac{f\left(y((t^*_{n-1},t_n])|\tilde{\boldsymbol{\tau}}_n\right)}{\sup_{\tilde{\boldsymbol{\tau}}'_n} f\left(y((t^*_{n-1},t_n])|\tilde{\boldsymbol{\tau}}'_n\right)}.$$

For Poisson process data, the supremum in the denominator is easily calculated through an exhaustive search of the $2(r_n+1)$ possible candidates for the maximum likelihood value of $\tilde{\boldsymbol{\tau}}_n$, where r_n is the number of events in $(t_{n-1},t_n]$: the case $\tilde{k}_n=0$, and then the cases where $\tilde{k}_n=1$ and $\tilde{\tau}_1$ is either equal to t_{n-1} or at the left or right limit of one of the r_n event times. It is easily shown that the likelihood is maximized at one of these positions, since between event times the likelihood function $f(y((t^*_{n-1},t_n])|\tilde{\boldsymbol{\tau}}_n)$ is log-concave.

A.3 Reanalyzing the coal data

The coal data from Section 3.3.1 of the main paper are now reanalyzed assuming the prior model (A.1) for changepoints, with a minimum distance of $s=1$ year between any two changepoints. Independent conjugate priors (9) are assumed for the intensities in each segment. For comparison, the yearly sequence of posterior distributions under this model are sampled by SMC Algorithm 1 using proposals for each update interval from one of the following:

- RJMCMC from $\pi_{(t_{n-1},t_n]}(\tilde{\boldsymbol{\tau}}_n|y((t^*_{n-1},t_n]))$ (SMC_CP);
- exact sampling from $\pi_{(t_{n-1},t_n]}(\tilde{\boldsymbol{\tau}}_n|y((t^*_{n-1},t_n]))$ (SMC_CP_REJ);
- exact sampling from the prior $p_{(t_{n-1},t_n]}(\tilde{\boldsymbol{\tau}}_n)$ (SMC_CP_P).

Recall that SMC_CP_P is considered in the main paper as a fixed-cost approximation to the method of Martin et al. (2013)

A.3.1 Results

Figure 1 shows the results of the comparison, averaged over 5,000 runs. From the top panel, a first observation is that the estimated online intensity and number of changepoints are unaffected by switching prior from (1) to (A.1), enforcing a space between changepoints of $s=1$. However, under (A.1) it is possible to sample exactly from the target proposal distributions by rejection sampling.

Rejection sampling turned out to be a very efficient method for sampling from the local posterior distributions $\pi_{(t_{n-1},t_n]}(\tilde{\boldsymbol{\tau}}_n|y((t^*_{n-1},t_n]))$ in this example: the overall acceptance rate of proposals from the prior, averaged across update intervals and runs, was 58%. This suggests considerable similarity between these posterior distributions and the prior, which is perhaps unsurprising since the coal data are fairly sparse with each one year interval containing less than two event times on average.

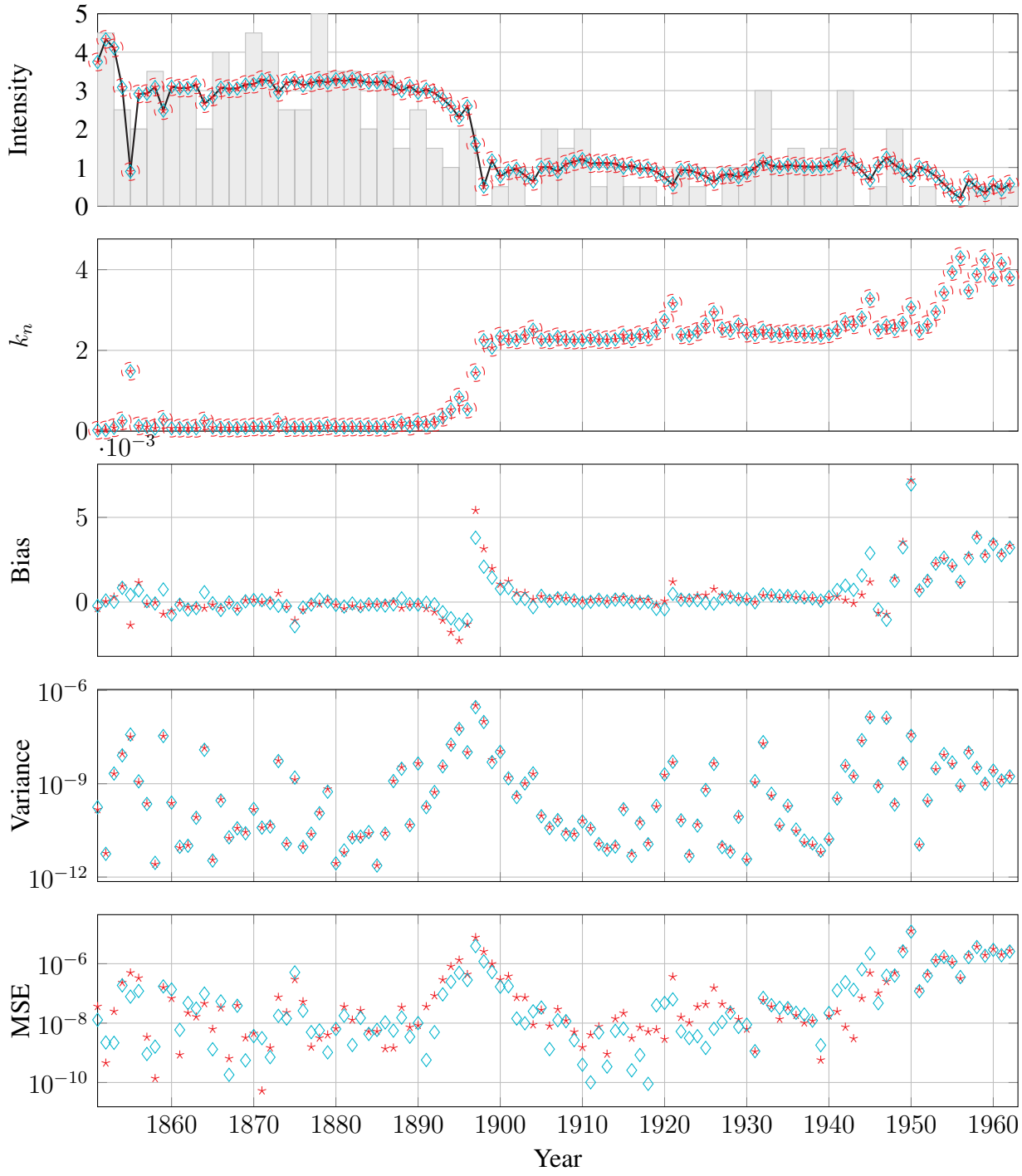


Figure 1: Comparison of sequential analyses of the conjugate posteriors for the coal-mining disaster data under the changepoint prior (A.1), using SMCMC (——), SMC_CP (⋆) and SMC_CP_REJ (◇). Rows 1-2: estimated online intensity and number of changepoints, along with the estimates from the SMC_CP method under the original Poisson process prior and the conjugate intensity prior (◯). Row 3: Bias. Rows 4-5: Variance, mean square error (log-scale).

The RJMCMC and rejection sampling SMC methods both track the SMCMC intensity and number of changepoint functions almost perfectly. Both SMC methods will have a small bias due to the approximate normalization of SMC weights by summing across the sample, and as expected the RJMCMC method shows a slightly higher bias than the rejection sampling approach (see the middle panel of Figure 1). But in both cases the biases, and mean square errors (MSEs) are tiny. This provides convincing evidence that the bias of the algorithm from using RJMCMC rather than exact sampling to sample from the update intervals, is negligible.

B Simulated inhomogeneous Poisson process data

The coal disaster data analyzed in the main paper and Section A.3 contain only 191 events over 112 years, and so it is unsurprising that most of the posterior mass is centered around a low number of changepoints (the prior mean) in that example. Also, given the similarity noted in the previous section between the prior distribution and the posteriors, it follows that proposals from the prior distribution will not be as dissimilar to the true target distribution as they might have been, which would explain the strong performance of that approach (SMC_CP_P) in the main paper.

B.1 Simulation details

To investigate this claim, a more challenging simulated inhomogeneous Poisson process data set was generated synthetically, with 500 changepoints occurring on a data interval of $[0, 10{,}000]$. Intensities for each segment were generated independently from $\Gamma(1, 0.01)$, and then finally event data were simulated on $[0, 10{,}000]$ according to those intensities. The resulting data set had 981,552 event times.

B.2 Results

The data were analyzed using the same parameters as used for simulation, but with the Poisson process prior rate for changepoints taking value $\nu = 1 \times 10^{-4}$, implying a prior expectation of just one changepoint over the interval $[0, 10{,}000]$. For a sequential analysis, $[0, 10, 000]$ was divided into 100 equally sized sub-intervals; note that within each update interval, the probability of no changepoints is $e^{-0.01} \approx 0.99$, whereas on average each interval contains five changepoints.

Figure 2 shows the results from sequential analysis, comparing against SMCMC as an approximation to the true posterior; the SMC_CP approach of Algorithm 1; and the special case of Algorithm 1 with $t^*_{n-1} = t_n$ implying independent proposal draws from the prior (SMC_CP_P). Unlike the coal data analysis, the distance between the posterior distribution and the prior demonstrates the limitations of SMC_CP_P. The effective sample size (bottom panel) drops down to near 1 at every update interval, as the rare particles containing changepoints get retained and replicated at the expense of all others. Even with this resampling, the estimated number of changepoints k_n cannot keep up with the true posterior (see second panel of Figure 2). The online estimates of the intensity and the time passed since the most recent changepoint at t_n fails to track the true posterior. In contrast to these results, SMC_CP tracks the SMCMC estimates perfectly in each case.

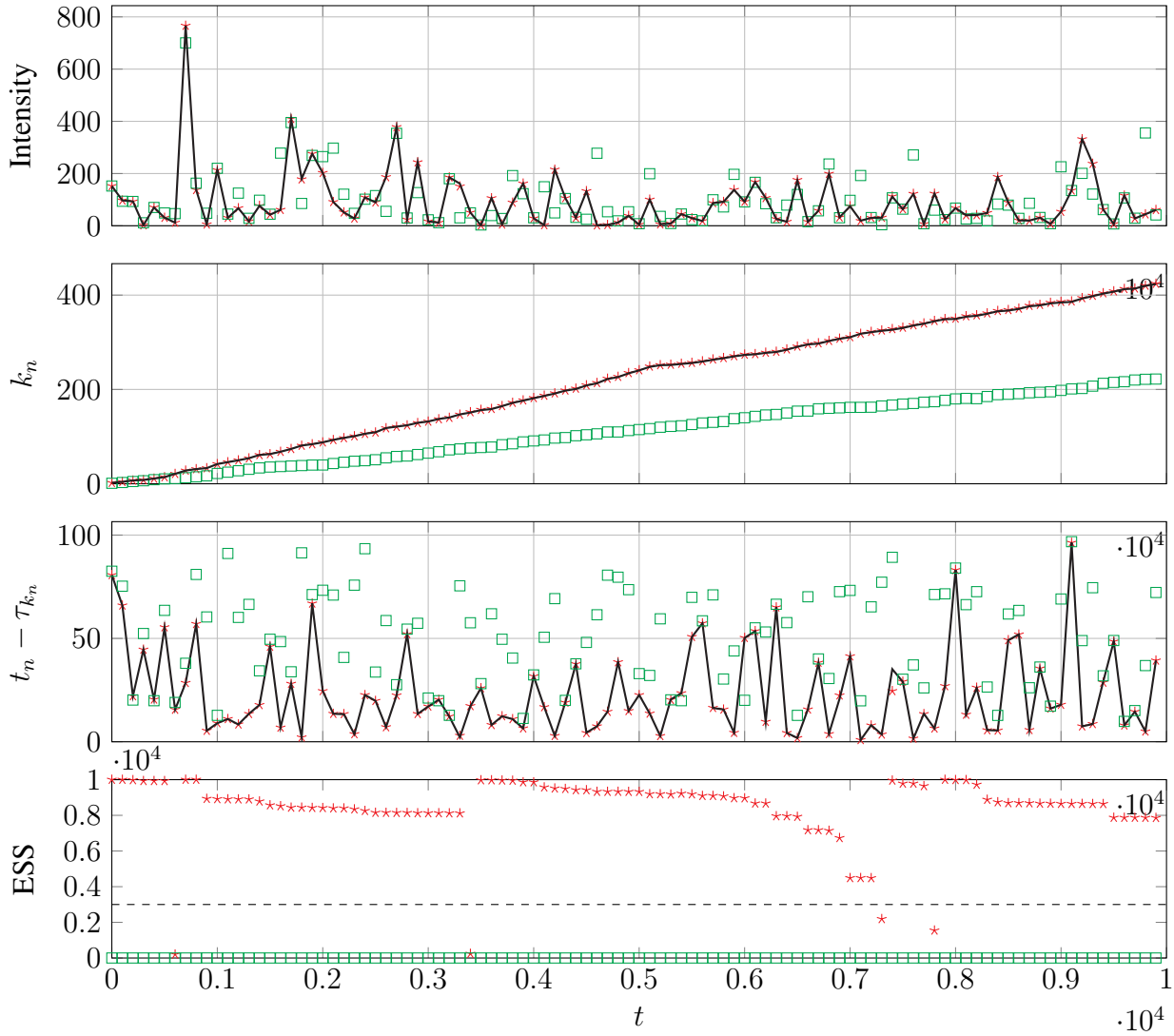


Figure 2: Results for the simulated inhomogeneous Poisson process using SMCMC (——), SMC_CP (⋆) and SMC_CP_P (▫). From top: Online estimated intensity; Online estimated number of changepoints; Online estimated time since last changepoint; Effective sample size, threshold 3,000 (- - -).

C Other applications in regression

To demonstrate the effectiveness of the SMC algorithm on applications other than Poisson process event data, two regression examples are briefly illustrated here.

C.1 Piecewise constant Gaussian regression

First suppose the underlying stochastic process $\theta(t)$ under consideration is bivariate. The pair $\theta(t) = (\theta_1(t), \theta_2(t))$ will represent, respectively, then mean and variance of a univariate data stream, so $\theta(t) \in \Theta = \mathbb{R} \times \mathbb{R}^+$. Second, suppose $0 < x_1 < x_2 < \ldots$ are a sequence of observation

times in $\mathbb{R}^+$, and that at time x_j a corresponding mark with Gaussian error is observed, denoted z_j, such that

$$z_j \sim \mathrm{N}(\theta_1(x_j), \theta_2(x_j)).$$

In the notation of the main paper, the data process observed, $y(t)$ is this marked point process $(x_1, z_1), (x_2, z_2)$ for event times x_j occurring by time t. So for a time interval $I \subset \mathbb{R}^+$,

$$y(I) = \{(x_j, z_j) | x_j \in I\}.$$

Applying the changepoint model to these processes implies that between neighboring changepoints $[\tau_i, \tau_{i+1})$, the bivariate process $\theta(t)$ is constant and equal to $\theta_i = (\theta_{i,1}, \theta_{i,2})$ and hence the observed data stream has constant mean and variance.

To complete the specification of the Gaussian regression model, changepoints are again assumed to follow a Poisson process (1) with rate ν, and the mean and variance for each segment follow independent, conjugate priors:

$$\begin{aligned}\theta_{i,1} &\sim \mathrm{N}(0, \theta_{i,2} v), \\ \theta_{i,2} &\sim \Gamma(\alpha, \beta),\end{aligned}$$

for $v, \alpha, \beta > 0$.

At time t_n with k_n changepoints, let $(z_{i_1}, \ldots, z_{i_{n_i}})$ be the marks of the $n_i \geq 0$ data points which, $\forall j = 1, \ldots, n_i$, $x_{i_j} \in [\tau_i, \tau_{i+1})$, and let $n_\cdot = \sum_{i=0}^{k_n} n_i$ be the total number of observations by time t_n. Then the marginal likelihood of the observed process data is

$$\begin{aligned} f(y([0, t_n]) | \boldsymbol{\tau}_n) =& \pi^{-n_\cdot/2} \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^{k_n+1} \\ & \prod_{i=0}^{k_n} \frac{\Gamma(\alpha + n_i/2)}{\sqrt{v + n_i} \{\beta + \sum_{j=1}^{n_i} z_{i_j}^2/2 - v(\sum_{j=1}^{n_i} z_{i_j})^2/(2(vn_i + 1))\}^{\alpha + n_i/2}}. \end{aligned}$$

C.1.1 Simulation

To illustrate Algorithm 1 on piecewise constant regression, Gaussian marks for $n_\cdot = 1000$ equally spaced events at times $x_i = i$, $i = 1, \ldots, n_\cdot$ were simulated from a time varying model. (Note that equally spaced events are not a requirement for this model or that in Section C.2.) Four changepoints in the mean and variance of the Gaussian $\theta(t)$ occurred at times $(225, 325, 525, 725)$, and the corresponding mean and variance parameter pairs were

$$((-5, 1), (0, 0.25), (5, 4), (0, 4), (2, 2.25)).$$

A scatter plot of the simulated data can been seen in Figure 3, along with posterior distribution online mean estimates of the regression function $\theta_1(t_n)$ from sequential analysis using Algorithm 1 and full sequential SMCMC (SMCMC). The mean regression function can be seen to track well the mean from SMCMC. Figure 3 also plots the online estimates of the posterior expectation of the error variance, $\theta_2(t_n)$, along with a scatter plot of the squared residuals of the data; the online estimates of the expected number of changepoints, the time passed since the last changepoint and

the online expected number of changepoints, k_n; and finally, the effective sample size (ESS).

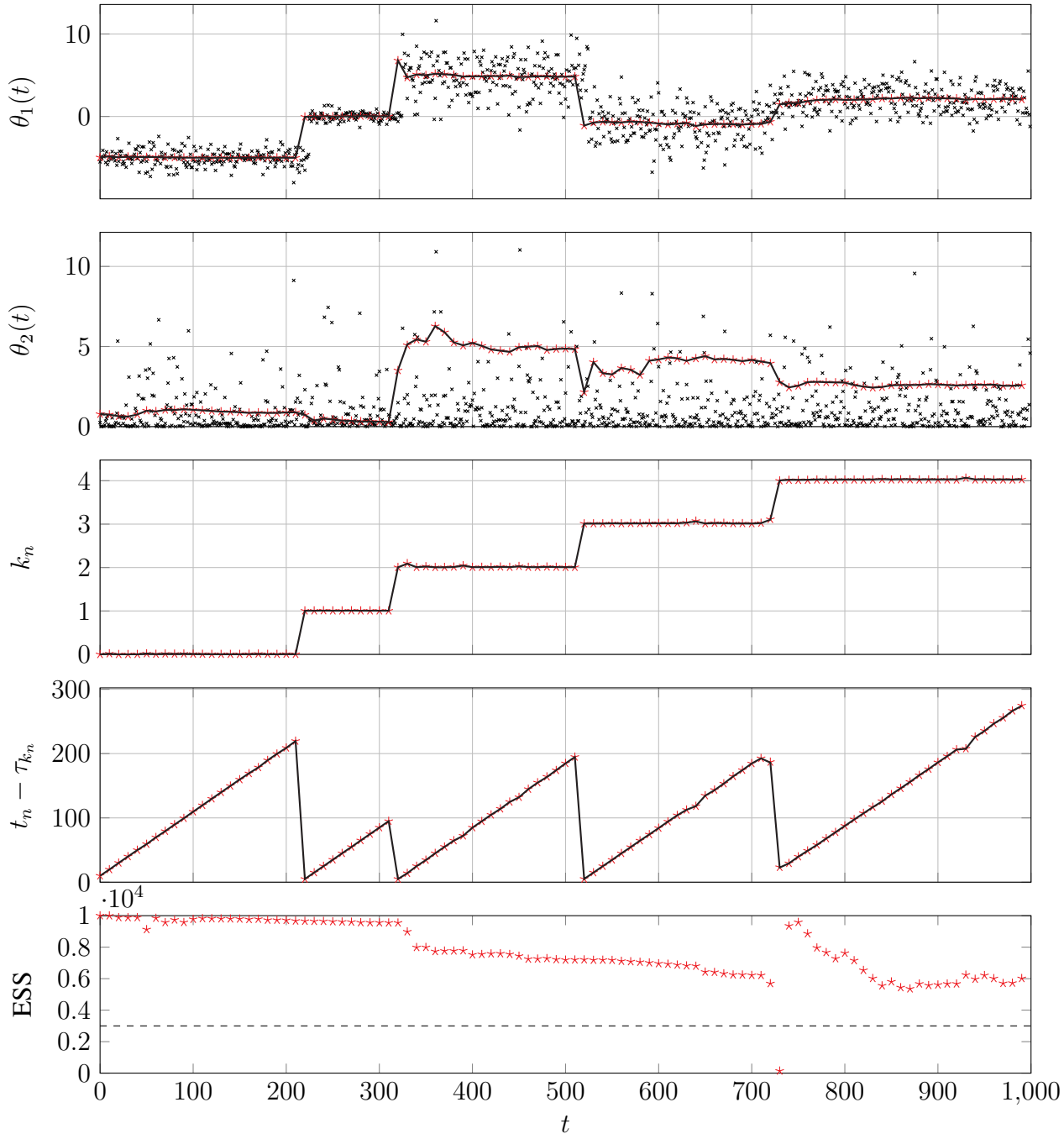


Figure 3: Results for the Gaussian piecewise regression model using SMCMC (—), SMC_CP (⋆). From top: Online estimated regression function; Online estimated regression error variance; Online estimated number of changepoints; Online estimated time since last changepoint; Effective sample size, threshold 3,000 (- - -).

C.2 Poisson regression

Following on from the bivariate estimation problem of piecewise constant Gaussian regression, a similar, univariate model for Poisson regression is now considered. Let $\theta(t) \in \Theta = \mathbb{R}^+$ represent time-varying value of both the mean and variance of a stream of integer count data.

Again suppose $0 < x_1 < x_2 < \ldots$ are a sequence of observation times in $\mathbb{R}^+$, and that at time x_j a corresponding mark is observed, denoted z_j, such that

$$z_j \sim \text{Poisson}(\theta(x_j)).$$

Following the notation of Section C.1, if θ_i is the Poisson distribution parameter for observations in $[\tau_i, \tau_{i+1})$, then each of these segment parameters are assumed to have independent, conjugate priors,

$$\theta_i \sim \Gamma(\alpha, \beta),$$

for $\alpha, \beta > 0$.

At time t_n, then the marginal likelihood of the observed process data is

$$f(y([0, t_n])|\boldsymbol{\tau}_n) = \prod_{i=0}^{k_n} \frac{\beta^\alpha \Gamma(\alpha + \sum_{j=1}^{n_i} z_{i_j})}{\Gamma(\alpha)(\beta + n_i)^{\alpha + \sum_{j=1}^{n_i} z_{i_j}}}.$$

C.2.1 Simulation

To illustrate Algorithm 1 on piecewise constant regression, Poisson marks for $n. = 500$ equally spaced events at times $x_i = i$, $i = 1, \ldots, n.$ were simulated from a time varying model. Three changepoints in the Poisson parameter $\theta(t)$ occurred at times $(100, 150, 350)$, and the corresponding Poisson parameters were (1.0,3.0,5.0,2.0).

A scatter plot of the simulated Poisson regression data can been seen in Figure 4, along with the posterior distribution online mean estimates of the regression function from sequential analysis using Algorithm 1 and SMCMC. As with the Gaussian example, the online mean regression function can be seen to follow the SMCMC curve closely. Figure 3 also plots online estimates of the expected number of changepoints, the time passed since the last changepoint, the online expected number of changepoints and the ESS performance.

References

Martin, J., Jasra, A., and McCoy, E. (2013). Inference for a class of partially observed point process models. *Annals of the Institute of Statistical Mathematics*, 65(3):413–437.

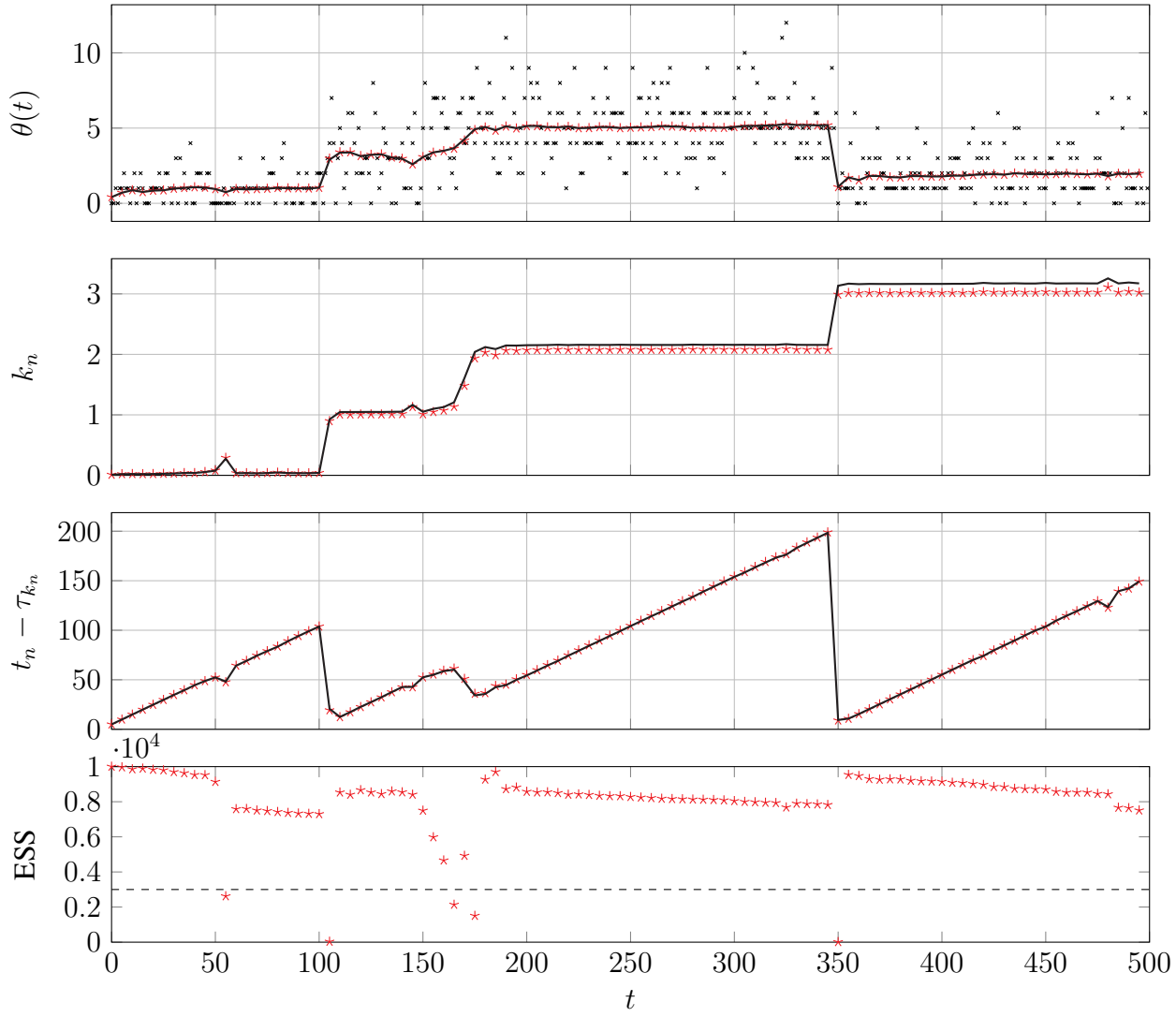


Figure 4: Results for the Poisson regression model using SMCMC (——), SMC_CP (⋆). From top: Online estimated regression function; Online estimated regression error variance; Online estimated number of changepoints; Online estimated time since last changepoint; Effective sample size, threshold 3,000 (- - -).