Special Issue Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.2067

Published online 30 August 2016 in Wiley Online Library

Online Steady State Detection Based on Rao-Blackwellized Sequential Monte Carlo

Yuxing Hou,[a] Jianguo Wu[b] and Yong Chen[a]*†

Online detection of whether a data stream has reached the steady state is known to be an important problem in many applications such as process control, data reconciliation and fault detection. This paper introduces a novel online steady state detection algorithm under the Bayesian framework based on a multiple change-point state space formulation and the sequential Monte Carlo methods. A Rao-Blackwellization technique is proposed to substantially reduce the variance of Monte Carlo estimation and greatly enhance the computational efficiency. In addition, a resampling scheme called the Optimal Resampling is used for eliminating duplicate samples and the robustness of steady state detection is significantly improved by using the information of the particles more efficiently. Numerical studies based on simulated signals and application to a real data set are used to evaluate the performance of the proposed method and compare with other existing methods from the literature. The proposed method is shown to establish a more robust performance than other methods. And it is much more computationally efficient than the standard sequential Monte Carlo method. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: online steady state detection; multiple change-point model; sequential Monte Carlo; rao-Blackwellization; optimal resampling

1. Introduction

Steady state of a system in many applications is one of the most important requirements to evaluate the performance of the process or to trigger the next actions in the process control. Timely detection of whether a data stream reaches the steady state (i.e., mean and variance unchanged) has been found critical in various fields, such as process control considered by Mahuli *et al*.,[1] Cao and Rhinehart,[2] Jiang *et al*.,[3] and Wu *et al*.,[4] data reconciliation considered by Narasimhan *et al*.,[5] Bagajewicz and Jiang,[6] Bhat and Saraf,[7] Korbel *et al*.,[8] fault detection and diagnosis (FDD) considered by Kim *et al*.,[9] and process optimization considered by Mhamdi *et al*.[10] We can categorize the steady state detection problem into two types: *offline* and *online*. Most of the well-developed methods in the literature correspond to the offline detection problems arising from discrete-event simulations, where it is usually very difficult to start the simulation directly from the steady state because the steady state of the system is typically unknown. Data collected during the transient period (or *warm-up period*) prior to the steady state causes estimation bias, which is called the *initialization bias* in the steady state parameter estimation which is considered by Kelton and Law,[11] Gallagher *et al*.,[12] Fishman[13] and Hoad *et al*.[14] To solve this problem, usually the simulation is first run long enough to guarantee the simulation output has reached the steady state. Then *offline* methods are typically used to identify the starting point of the steady state in the simulation outputs so that the data from the transient period can be removed. These methods can be further classified into five different types as proposed by Robinson[15]: graphical methods, heuristic approaches, statistical methods, initialization bias tests, and hybrid methods.

Compared to the offline methods which have been extensively investigated in the simulation literature, studies for *online* steady state detection are limited. The main challenge of online detection is that it has to be done *in real-time*, which justifies the need for detection procedures to timely update estimations as the latest observations become available. The existing online methods typically utilize a moving data window, based on which some test statistics are developed to decide if the signal has entered the steady state. Examples of such methods include polynomial interpolation test (PIT) which is proposed by Savitzky and Golay,[16] Roux *et al*.,[17] variance ratio test (VRT) which is proposed by Crow *et al*.,[18] Cao and Rhinehart,[2] slope detection method (SDM) which is proposed by Holly *et al*.,[19] Bethea and Rhinehart,[20] Wu *et al*.,[4] and *t*-test which is proposed by Narasimhan *et al*.[21] However, the performance of all these methods is highly dependent on the selection of the data window size. Either too small or too large the size may significantly increase either the *false alarm rates* (FAR) or detection delays. Also, the appropriate window size is very sensitive to noise

[a]*Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City, Iowa, USA*
[b]*Department of Industrial, Manufacturing and Systems Engineering, University of Texas, El Paso, TX, USA*
**Correspondence to: Yong Chen, Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City, Iowa, 52242, USA.*
†*E-mail: yong-chen@uiowa.edu*

levels as well as types of signals. Usually one specific window size cannot perform consistently well for various types of signals with different noise levels. Therefore, it is urgent to develop methods with more flexibility and robustness to handle various situations.

Naturally the online steady state detection problem can be approached from the Bayesian perspective, where the prior information of model parameters and transition point to the steady state, combined with the latest observations, are used to dynamically update the knowledge of the current state of the process. The main disadvantage of the Bayesian method is that it usually leads to intractable posterior distributions. However, recent advance in the field of computational statistics makes the computations of posterior distributions feasible. The most common computational Bayesian methods are the *Markov Chain Monte Carlo* (MCMC) algorithms, such as Liu,[22] Robert and Casella.[23] However, MCMC algorithms are typically not appropriate for online applications because of its fast increasing computational cost over time when data size gets larger and larger. On the other hand, the *sequential Monte Carlo* (SMC) methods, including the *particle filtering* (PF) techniques, can be used, such as Doucet *et al*.,[24] Arulampalam *et al*.[25] In contrast to the standard MCMC approaches, the sequential structure of the SMC methods allows for updating estimations sequentially for each newly arriving observation in a computationally efficient way, which is very useful for online inference.

Recently, a PF method has been proposed by Wu *et al*.[26] for online steady state detection. In their method, the targeting signal is approximated by a multiple change-point model and the PF techniques are used to estimate the posterior distribution of the latest change-point and other model parameters (slope, intercept and noise variance). Some improvement strategies including the stratified sampling, partial Gibbs resample-move techniques, and timeliness improvement strategy are developed to overcome the particle degeneracy and reduce the computational cost. Such PF method is shown to be much more effective and robust in steady state detection compared with moving data window based methods. However, the major disadvantage of the PF method is that it requires a large number of particles to achieve accurate posterior distribution approximations, due to the high dimension of model parameters, which leads to high computational cost and may limit its applications in many online detection problems that require quick responses. In this paper, we propose a novel SMC method using a Rao-Blackwellization technique, combined with a resampling method called the *Optimal Resampling*, to substantially reduce the computational cost and improve the detection robustness on noisy data. The main contribution of our algorithm is its significant improvement in computational efficiency by taking advantage of the Rao-Blackwellization technique, while still achieving comparable or even better detection performance compared with the PF method. The significant reduction of computational costs makes our method a much more preferred method for many online applications where quick steady state detection is critical.

The rest of the paper is organized as follows. A multiple change-point model formulation of this problem is introduced in Section 2. In Section 3, we give a detailed description of the proposed SMC algorithm. The numerical examples and application to real signals are presented in Section 4. And a summary is given in Section 5.

2. Piecewise linear model with multiple change-points

Given a noisy signal $\mathbf{y}_{0:N}=(y_0, y_1, \ldots, y_N)$, this paper targets detecting the steady state of the signal using a multiple change-point model. The basic idea is that we approximate any signal, linear or nonlinear, using a piecewise linear model that allows for local linear representations of the signal, as shown in Figure 1. The parameters (slope, intercept and noise variance) are assumed to be independent across different line segments. When the latest line segment is sufficiently 'flat', the signal is considered to be in the steady state.

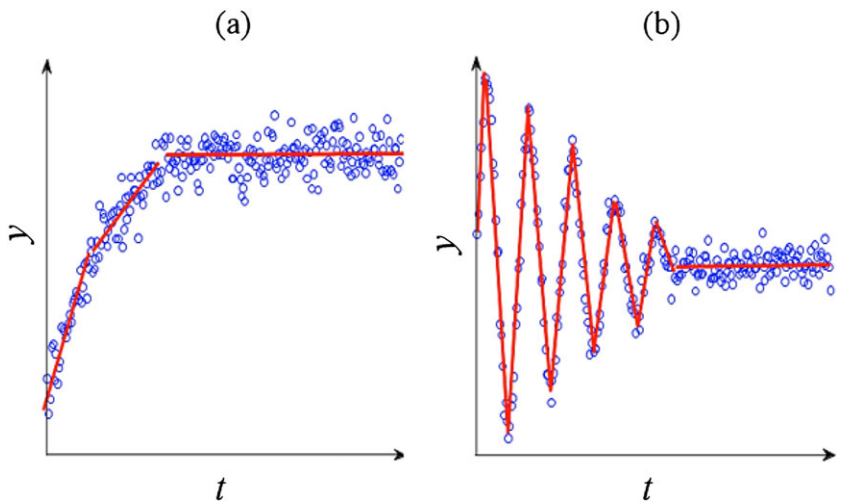


Figure 1. Illustration of approximating nonlinear signals using piecewise linear model: (a) signal generated using exponential function and noise; (b) oscillating nonlinear function

Suppose at time t the model parameters are $\xi_t = (\beta_{0t}, \beta_{1t}, \sigma_t^2)$, where β_{0t} is the intercept of the current line segment, β_{1t} is the slope, and σ_t^2 is the unknown variance of Gaussian noise. The model parameters ξ_t are subject to change at m unknown change-points $\mathbf{C}_m = (c_1, c_2, \ldots, c_m)$ while remaining constant otherwise. This multiple change-point model can be written as

$$\xi_t = \begin{cases} \boldsymbol{\theta}_1 & \text{if } 0 \le t < c_1 \\ \boldsymbol{\theta}_2 & \text{if } c_1 \le t < c_2 \\ \vdots & \vdots \\ \boldsymbol{\theta}_m & \text{if } c_{m-1} \le t < c_m \\ \boldsymbol{\theta}_{m+1} & \text{if } c_m \le t \le N \end{cases} \tag{1}$$

where $\boldsymbol{\theta}_i \in .\mathscr{R}^3$, $i = 1, 2, \ldots, m+1$ are the values of model parameters at $m+1$ different line segments.

Within the Bayesian framework, we assign appropriate priors to the change-points between two segments, as well as other model parameters. Their posterior distributions can then be sequentially updated and the steady state can be inferred based on the posterior distributions of model parameters (e.g., slope) of the current line segment. In model (1), the $\boldsymbol{\theta}_i$'s and c_i's are unknown. We assign a prior distribution $q_{\boldsymbol{\theta}}(\cdot)$ for $\boldsymbol{\theta}_i$'s. Let τ_t be the latest change-point up to time t. It is easy to see that $\tau_t = c_{i-1}$ iff $c_{i-1} \le t < c_i$. Therefore τ_t's and c_i's contain equivalent information. We assign a prior transition probability for τ_t given τ_{t-1} as $P(\tau_t | \tau_{t-1})$. If a change occurs at time t, we have $\tau_t = t$, otherwise $\tau_t = \tau_{t-1}$. In this paper we assume $P(\tau_t = t | \tau_{t-1}) = p$, which corresponds to a geometric distribution with probability p for the random duration of each line segment.

To facilitate application of a sequential Monte Carlo (SMC) method for online inference of model parameters, we first reformulate the piecewise linear model in (1) into a state space model, where at each time t the distribution of the observation y_t depends on a hidden state vector denoted by $\mathbf{x}_t$. Defining the state vector as $\mathbf{x}_t = (\xi_t, \tau_t)$, the state space model is given as

$$\begin{aligned} \mathbf{x}_t = (\xi_t, \tau_t) | \mathbf{x}_{t-1} &= \begin{cases} (\boldsymbol{\theta}^*, t) & \text{with probability p (change occurs at t)} \\ \mathbf{x}_{t-1} = (\xi_{t-1}, \tau_{t-1}) & \text{with probability 1-p (no change at t)} \end{cases} \\ y_t &= \beta_{0t} + \beta_{1t} t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \end{aligned} \tag{2}$$

where $\boldsymbol{\theta}^* \sim q_{\boldsymbol{\theta}}(\cdot)$ and is independent of ξ_{t-1}, and ε_t is the Gaussian noise. The state transition of the state space model is illustrated in Figure 2. It can be seen that the state space model in (2) equivalently represents the piecewise linear model in (1). Based on (2), at any time t, with probability p a change occurs ($\tau_t = t$) and a new line segment is started. Since $\xi_t, t \ge 0$ contain all the information on $\boldsymbol{\theta}_i$'s in (1) and $\tau_t, t \ge 0$ contain all the information on c_i's in (1), the inference of the parameters in the piecewise linear model in (1) is equivalent to the inference of the state vector $\mathbf{x}_t = (\xi_t, \tau_t)$ in (2). Consequently, we can focus on the state space model in (2) and the inference of (ξ_t, τ_t) to develop the steady state detection method.

If $\tau_t = 0$ for all t, that is, if there is no change-point, the state space model in (2) reduces to a linear Gaussian system, in which the *Kalman filter* (KF), proposed by Kalman,[27] provides a closed-form solution for efficient state estimations. However, because of the existence of unknown change-points, the state space model is a nonlinear model which cannot be solved directly by KF. On the other hand, the PF techniques are common and effective for nonlinear state space models. Recently, a PF method has been proposed by Wu *et al.*[26] for online steady state detection. A disadvantage of the method in Wu *et al.*[26] is that it requires a large number of particles because each parameter in the state vector $\mathbf{x}_t$ needs to be sampled. In the following section, we propose a more efficient sequential Monte Carlo algorithm based on a variance reduction method called *Rao-Blackwellization*.

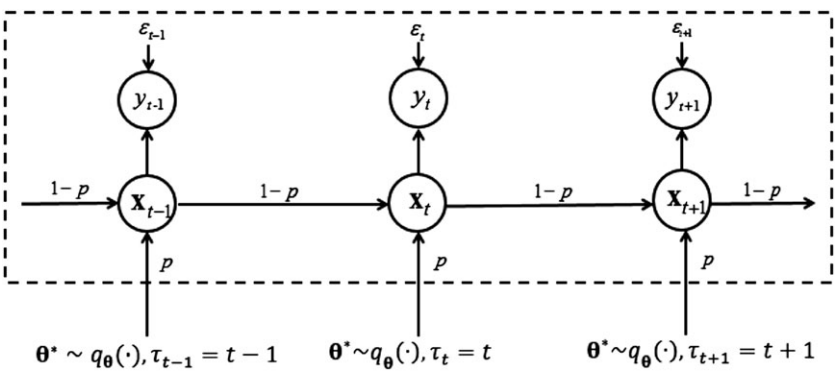


Figure 2. Illustration of the state space model

3. Rao-blackwellzied sequential Monte Carlo method for online steady state detection

3.1. Review of standard SMC algorithm for state space model

In this section, the general framework of standard SMC algorithm for state space model is reviewed. The main idea for SMC method is to use a recursive importance sampling strategy to get a set of properly weighted samples, which are used to approximate the desired posterior density of the state vector in a state space model. For the state space model in Eq. (2), let $\mathbf{x}_{0:t}=\{\mathbf{x}_0,\mathbf{x}_1,\ldots,\mathbf{x}_t\}$ be the set of all state vectors up to the current time t, and $\mathbf{y}_{0:t}=(y_0,y_1,\ldots,y_t)$ be the observations up to the current time t. Online inference of $\mathbf{x}_{0:t}$ is of our central interest; that is, at time t, we want to estimate the posterior density $p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$. Usually $p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$ is of an intractable density form where direct simulations cannot be implemented.

In such case, we often resort to the *Importance Sampling* (IS) approach in which we bypass the intractable distribution by considering some *importance distribution* where direct simulations can be easily performed. Let $\left\{\mathbf{x}_{0:t}^{(i)}\right\}_{i=1}^{n}$ be the n samples (particles) generated from an importance distribution $\pi(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$, where $\mathbf{x}_{0:t}^{(i)}$ denotes the i -th sample/particle. By associating the *importance weight*

$$w_t^{(i)}=\frac{p\left(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{0:t}\right)}{\pi\left(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{0:t}\right)} \tag{3}$$

to the sample $\mathbf{x}_{0:t}^{(i)}$, the posterior density of interest can be approximated as

$$\hat{p}\left(\mathbf{x}_{0:t}|\mathbf{y}_{0:t}\right)=\sum_{i=1}^{n}w_t^{(i)}\delta\left(\mathbf{x}_{0:t}-\mathbf{x}_{0:t}^{(i)}\right) \tag{4}$$

where δ denotes the Dirac function and the weights are normalized such that $\sum_{i=1}^{n}w_t^{(i)}=1$. Therefore $\hat{p}\left(\mathbf{x}_{0:t}|\mathbf{y}_{0:t}\right)$ is a discrete weighted approximation of the true posterior $p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$. The pairs $\left\{\mathbf{x}_{0:t}^{(i)},w_t^{(i)}\right\}_{i=1}^{n}$ are a collection of *properly weighted sample* with respect to the posterior distribution $p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$. An important observation is that $\mathbf{x}_t^{(i)}$ is also properly weighted by $w_t^{(i)}$ with respect to the marginal posterior distribution $p(\mathbf{x}_t|\mathbf{y}_{0:t})$.

To implement Monte Carlo techniques for the online estimation problem, $\hat{p}\left(\mathbf{x}_{0:t}|\mathbf{y}_{0:t}\right)$ with respect to $p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$ needs to be sequentially computed. Since the state equation in our system follows the Markovian structure, we can implement the importance sampling recursively, which forms the basis of SMC methods, as in Liu and Chen,[28] Doucet *et al.*[29] Based on the Bayes' theorem, weights can be sequentially updated as follows

$$\begin{aligned}w_t^{(i)}&=\frac{p\left(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{0:t}\right)}{\pi\left(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{0:t}\right)}\propto\frac{p\left(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{0:t-1}\right)p\left(y_t|\mathbf{x}_{0:t}^{(i)},\mathbf{y}_{0:t-1}\right)}{\pi\left(\mathbf{x}_{0:t-1}^{(i)}|\mathbf{y}_{0:t-1}\right)\pi\left(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)},\mathbf{y}_{0:t}\right)}\\&=w_{t-1}^{(i)}\frac{p\left(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)},\mathbf{y}_{0:t-1}\right)p\left(y_t|\mathbf{x}_{0:t}^{(i)},\mathbf{y}_{0:t-1}\right)}{\pi\left(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)},\mathbf{y}_{0:t}\right)}\end{aligned} \tag{5}$$

Applying (5) to the state space model in (2), we have

$$w_t^{(i)}\propto w_{t-1}^{(i)}\frac{f_{t-1}\left(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)}\right)g_{t-1}\left(y_t|\mathbf{x}_t^{(i)}\right)}{\pi\left(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)},\mathbf{y}_{0:t}\right)} \tag{6}$$

where $f_{t-1}(\mathbf{x}_t|\mathbf{x}_{t-1})$ is the probability density function (pdf) of $(\mathbf{x}_t|\mathbf{x}_{t-1})$ and $g_{t-1}(y_t|\mathbf{x}_t)$ is the pdf of y_t given $\mathbf{x}_t$.

If we want to estimate the expectation of a function of $\mathbf{x}_{0:t}$, say $m(\mathbf{x}_{0:t})$, conditioning on $\mathbf{y}_{0:t}$, we have

$$E(m(\mathbf{x}_{0:t})|\mathbf{y}_{0:t})=\int m(\mathbf{x}_{0:t})p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})d\mathbf{x}_{0:t}\simeq\sum_{i=1}^{n}w_t^{(i)}m\left(\mathbf{x}_{0:t}^{(i)}\right) \tag{7}$$

There are two important issues regarding the design and implementation of the SMC algorithm. One is the selection of the importance density $\pi\left(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)},\mathbf{y}_{0:t}\right)$ in (6). It is often convenient to choose it to be the prior $\pi\left(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)},\mathbf{y}_{0:t}\right)=f_{t-1}\left(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)}\right)$, which greatly simplifies the weights update in (6) as

$$w_t^{(i)}\propto w_{t-1}^{(i)}g_{t-1}\left(y_t|\mathbf{x}_t^{(i)}\right) \tag{8}$$

The other important issue is the use of *resampling*. The standard SMC algorithm suffers from the *degeneracy* phenomenon, where after a few iterations, all but one of the weights are very close to zero, as mentioned in Doucet *et al.*[29] This degeneracy indicates the

majority of computational work for updating samples and weights is wasted. The resampling scheme is used to reduce such effect by eliminating particles with low weights and concentrating on those with large weights, so that all of the resampled particles can contribute significantly to the importance sampling estimates. A simple and common choice of resampling procedure is the *multinomial resampling*, which involves generating a new set of particles $\left\{\mathbf{x}_t^{(j)}\right\}_{j=1}^{n}$ by resampling with replacement n times from $\left\{\mathbf{x}_t^{(i)}\right\}_{i=1}^{n}$ according to their weights $\left\{w_t^{(i)}\right\}_{i=1}^{n}$, and then assigning equal weights $1/n$ to the new set of particles $\left\{\mathbf{x}_t^{(j)}\right\}_{j=1}^{n}$.

The above procedure of the standard SMC algorithm with resampling is usually called the PF algorithm and shown in Algorithm 1. In the following sections, we propose two improvements of the standard SMC algorithm. In section 3.2, a variance reduction method called Rao-Blackwellization is used to substantially lower the computational cost of the standard SMC algorithm. In section 3.3, we introduce a new resampling scheme to achieve better detection estimation of the steady state.

Initialization, for $i = 1, \ldots, n$:
- ♦ Sample particles $\mathbf{x}_0^{(i)} = \left(\xi_0^{(i)}, \tau_0^{(i)}\right)$, where $\xi_0^{(i)} = \boldsymbol{\theta}_0^{(i)} \sim q_{\boldsymbol{\theta}}(\cdot)$ and $\tau_0^{(i)} = 0$.
- ♦ Assign $w_0^{(i)} = 1/n$ to each particle $\mathbf{x}_0^{(i)}$.

For $t = 1, \ldots, T$:
- ♦ For $i = 1, \ldots, n$:
  - ○ Draw samples $\mathbf{x}_t^{(i)}$ from the importance distribution $f_{t-1}\left(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)}\right)$.
- ♦ Updating weights:
  - ○ Compute weights $w_t^{(i)}$ according to Eq. (8) and normalize the weights such that $\sum_{i=1}^{n} w_t^{(i)} = 1$.
- ♦ Resampling: resample $\left\{\mathbf{x}_t^{(i)}, w_t^{(i)}\right\}_{i=1}^{n}$ to generate n equally weighted particles $\left\{\tilde{\mathbf{x}}_t^{(j)}, \frac{1}{n}\right\}_{j=1}^{n}$, and set $\left\{\mathbf{x}_t^{(i)}, w_t^{(i)}\right\}_{i=1}^{n} = \left\{\tilde{\mathbf{x}}_t^{(j)}, \frac{1}{n}\right\}_{j=1}^{n}$.

Algorithm 1. Standard SMC Algorithm

3.2. Rao-blackwellized SMC algorithm

In this section, a Rao-Blackwellized version of the SMC algorithm is proposed. Its main idea is to marginalize out the parameters associated with each line segment and achieves substantial variance reduction of Monte Carlo estimates. The parameters of each line segment can be integrated out and estimated by taking advantage of the results from Bayesian linear regression. With Rao-Blackwellization proposed by Casella and Robert,[30] Doucet *et al.*,[24] the algorithm is much more efficient than the standard SMC algorithm.

To make proper inference of the state vector $\mathbf{x}_t$ which consists of four parameters, the common way for standard SMC algorithm is to obtain estimates based on their joint posterior distribution, namely $p\left(\boldsymbol{\beta}_t, \sigma_t^2, \tau_t | \mathbf{y}_{0:t}\right)$, where $\boldsymbol{\beta}_t = (\beta_{0t}, \beta_{1t})$ in Wu *et al.*[26] Although such method is effective and robust to detect the steady state, it is at the expense of high computational cost since a large number of particles have to be generated at each time step t to approximate the joint distribution of four variables. However, under appropriate choice of priors, $\boldsymbol{\beta}_t$ and σ_t^2 of each line segment can actually be integrated out, resulting in a particularly efficient algorithm with reduced variance of the estimates. More specifically, the unknown posterior density of interest can be factorized as follows

$$p\left(\boldsymbol{\beta}_t, \sigma_t^2, \tau_t | \mathbf{y}_{0:t}\right) = p\left(\boldsymbol{\beta}_t, \sigma_t^2 | \tau_t, \mathbf{y}_{0:t}\right) P(\tau_t | \mathbf{y}_{0:t}) \tag{9}$$

in which the conditional posterior density $p\left(\boldsymbol{\beta}_t, \sigma_t^2 | \tau_t, \mathbf{y}_{0:t}\right)$ can be solved analytically if we use the conjugate priors for $\boldsymbol{\beta}_t$ and σ_t^2. Consequently, estimating the joint distribution $p\left(\boldsymbol{\beta}_t, \sigma_t^2, \tau_t | \mathbf{y}_{0:t}\right)$ requires sampling from only a one dimensional distribution $P(\tau_t | \mathbf{y}_{0:t})$, which can dramatically reduce the number of particles needed to reach a given estimation accuracy. In the remainder of this section, we will first discuss how to obtain the analytical solution for the conditional joint posterior density $p\left(\boldsymbol{\beta}_t, \sigma_t^2 | \tau_t, \mathbf{y}_{0:t}\right)$. Then we will discuss how to sample τ_t based on another application of the Rao-Blackwellization.

To derive the conditional joint posterior density $p\left(\boldsymbol{\beta}_t, \sigma_t^2 | \tau_t, \mathbf{y}_{0:t}\right)$, we first factorize it as

$$\begin{aligned} p\left(\boldsymbol{\beta}_t, \sigma_t^2 | \tau_t, \mathbf{y}_{0:t}\right) &= p\left(\boldsymbol{\beta}_t | \sigma_t^2, \tau_t, \mathbf{y}_{0:t}\right) p\left(\sigma_t^2 | \tau_t, \mathbf{y}_{0:t}\right) \\ &= p\left(\boldsymbol{\beta}_t | \sigma_t^2, \tau_t, \mathbf{y}_{\tau_t:t}\right) p\left(\sigma_t^2 | \tau_t, \mathbf{y}_{\tau_t:t}\right) \end{aligned} \tag{10}$$

The last equation in (10) is due to the assumption that the model parameters between different line segments are independent. Using the conjugate prior, we assume that $\boldsymbol{\beta}_t | \sigma_t^2$ follows a normal distribution with mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\sigma_t^2 \boldsymbol{\Sigma}_0$, where $\boldsymbol{\Sigma}_0$ is a 2×2 positive definite matrix. The prior distribution of σ_t^2 is chosen to be the inverse gamma density with shape parameter a_0 ($a_0 > 1$) and scale parameter $b_0 (b_0 > 0)$. Due to the conjugacy of normal and inverse gamma distributions and based on the results from Bayesian linear regression in O'Hagan,[31] we have

$$\left(\boldsymbol{\beta}_t|\sigma_t^2,\tau_t,\mathbf{y}_{\tau_t:t}\right)\sim \mathrm{N}\left(\boldsymbol{\mu}_{\tau_t,t},\sigma_t^2\boldsymbol{\Sigma}_{\tau_t,t}\right)$$
$$\left(\sigma_t^2|\tau_t,\mathbf{y}_{\tau_t:t}\right)\sim \mathrm{IG}\left(a_{\tau_t,t},b_{\tau_t,t}\right)$$

and the parameters $\boldsymbol{\mu}_{\tau_t,t},\boldsymbol{\Sigma}_{\tau_t,t},a_{\tau_t,t},b_{\tau_t,t}$ are updated according to the following equations

$$\begin{aligned}\sum\nolimits_{\tau_t,t}&=\left(\sum\nolimits_0^{-1}+\mathbf{D}_{\tau_t,t}'\mathbf{D}_{\tau_t,t}\right)^{-1}\\ \boldsymbol{\mu}_{\tau_t,t}&=\sum\nolimits_{\tau_t,t}\left(\sum\nolimits_0^{-1}\boldsymbol{\mu}_0+\mathbf{D}_{\tau_t,t}'\mathbf{y}_{\tau_t:t}\right)\\ a_{\tau_t,t}&=a_0+\frac{t-\tau_t+1}{2}\\ b_{\tau_t,t}&=b_0+\frac{1}{2}\left(\mathbf{y}_{\tau_t:t}'\mathbf{y}_{\tau_t:t}+\boldsymbol{\mu}_0'\sum\nolimits_0^{-1}\boldsymbol{\mu}_0-\boldsymbol{\mu}_{\tau_t,t}'\sum\nolimits_{\tau_t,t}^{-1}\boldsymbol{\mu}_{\tau_t,t}\right)\end{aligned}\tag{11}$$

where $\mathbf{D}_{\tau_t,t}=\begin{bmatrix}1 & \tau_t\\ 1 & \tau_t+1\\ \vdots & \vdots\\ 1 & t\end{bmatrix}$, matrix $\mathbf{S}'$ is the transpose of matrix $\mathbf{S}$.

When approaching the steady state, the signal can be characterized by a sustained 'flat' line segment. For this reason, the posterior knowledge of $\boldsymbol{\beta}_t$, especially that of the slope parameter β_{1t}, is very important for steady state detection. The marginal posterior distribution of $\boldsymbol{\beta}_t$ after integrating out σ_t^2 can be computed based on Lemma 1 as follows (the proof is included in Appendix A).

Lemma 1:

$\left(\boldsymbol{\beta}_t|\tau_t,\mathbf{y}_{\tau_t:t}\right)$ follows a bivariate non-standardized student's t-distribution with degrees of freedom $2a_{\tau_t,t}$, location parameter $\boldsymbol{\mu}_{\tau_t,t}$, and scale matrix $\frac{b_{\tau_t,t}}{a_{\tau_t,t}}\boldsymbol{\Sigma}_{\tau_t,t}$, which is denoted as $\left(\boldsymbol{\beta}_t|\tau_t,\mathbf{y}_{\tau_t:t}\right)\sim t_2\left(\boldsymbol{\mu}_{\tau_t,t},\frac{b_{\tau_t,t}}{a_{\tau_t,t}}\boldsymbol{\Sigma}_{\tau_t,t},2a_{\tau_t,t}\right)$.

To approximate the desired posterior density $p\left(\boldsymbol{\beta}_t,\sigma_t^2,\tau_t|\mathbf{y}_{0:t}\right)$ using (9), we also need to generate Monte Carlo samples of τ_t, the latest change-point at the current time t, to approximate $P(\tau_t|\mathbf{y}_{0:t})$. Intuitively, we can generate each sample $\tau_t^{(i)}$ from its prior transition probability $P\left(\tau_t^{(i)}|\tau_{t-1}^{(i)}\right)$ based on the idea of the standard SMC algorithm. However, a more efficient way to sample $\tau_t^{(i)}$ can be done as follows, which is based on the application of Rao-Blackwellization to τ_t that is similar to Chopin[32]:

- For each particle $\tau_{t-1}^{(i)}$ and its associated weight $w_{t-1}^{(i)}$, $i=1,\ldots,n$, create its two possible descendants at time t, with weights:

$$\begin{aligned}\tau_t^{(i,1)}&=\tau_{t-1}^{(i)},\ \ w_t^{(i,1)}=w_{t-1}^{(i)}P\left(\tau_t^{(i)}=\tau_{t-1}^{(i)}|\tau_{t-1}^{(i)}\right)p\left(y_t|\tau_t^{(i)}=\tau_{t-1}^{(i)},\mathbf{y}_{\tau_t^{(i)}:t-1}\right);\\ \tau_t^{(i,2)}&=t,\ \ w_t^{(i,2)}=w_{t-1}^{(i)}P\left(\tau_t^{(i)}=t|\tau_{t-1}^{(i)}\right)p\left(y_t|\tau_t^{(i)}=t,\mathbf{y}_{\tau_t^{(i)}:t-1}\right)\end{aligned}\tag{12}$$

The predictive density $p\left(y_t|\tau_t^{(i)},\mathbf{y}_{\tau_t^{(i)}:t-1}\right)$ in (12) can be calculated based on the following Lemma 2 (see Appendix B for the proof).

Lemma 2:
Denote $\mathbf{X}_t=[1\quad t]$, then

(a)

$$\left(y_t|\tau_t,\mathbf{y}_{\tau_t:t-1}\right)\sim t_1\left(\mathbf{X}_t\boldsymbol{\mu}_{\tau_t,t-1},\frac{b_{\tau_t,t-1}}{a_{\tau_t,t-1}}\left(1+\mathbf{X}_t\boldsymbol{\Sigma}_{\tau_t,t-1}\mathbf{X}_t'\right),2a_{\tau_t,t-1}\right),\ \text{if } \tau_t=\tau_{t-1};$$

(b)

$$\left(y_t|\tau_t,\mathbf{y}_{\tau_t:t-1}\right)\sim t_1\left(\mathbf{X}_t\boldsymbol{\mu}_0,\frac{b_0}{a_0}\left(1+\mathbf{X}_t\boldsymbol{\Sigma}_0\mathbf{X}_t'\right),2a_0\right),\ \text{if } \tau_t=t.$$

where $\boldsymbol{\mu}_{\tau_t,t-1},\boldsymbol{\Sigma}_{\tau_t,t-1},a_{\tau_t,t-1},b_{\tau_t,t-1}$ are parameters associated with the line segment $\mathbf{y}_{\tau_t:t-1}$ that can be obtained using equations in (11), $t_1(\cdot)$ the univariate non-standardized student's t-distribution.

The resulting set of $2n$ particles after Rao-Blackwellzation of $\tau_t^{(i)}$, $i = 1, \ldots, n$ can be resampled according to the weights $w_t^{(i,1)}$, $w_t^{(i,2)}$, $i = 1, 2, \ldots, n$ to obtain n resampled particles to avoid an exponentially increasing number of particles. By doing exact calculations on weights of all possible values of $\tau_t^{(i)}$, i.e., $\tau_t^{(i)} = \tau_{t-1}^{(i)}$ and $\tau_t^{(i)} = t$, the randomness inherent in the simulation of $\tau_t^{(i)}$ based on $P\left(\tau_t^{(i)} | \tau_{t-1}^{(i)}\right)$ is removed, which leads to further variance reduction.

At time t, similar to the standard SMC algorithm, the resulted $2n$ particles can be used to approximate the true posterior distribution $P(\tau_t|\mathbf{y}_{0:t})$. Suppose $\left\{\tau_t^{(i,j)}, w_t^{(i,j)}\right\}_{i=1}^{n}, j = 1, 2$ are the $2n$ properly weighted samples with respect to $P(\tau_t|\mathbf{y}_{0:t})$ after the Rao-Blackwellization of $\tau_t^{(i)}$, then the posterior density $p(\boldsymbol{\beta}_t|\mathbf{y}_{0:t})$ can be approximated as

$$\hat{p}(\boldsymbol{\beta}_t|\mathbf{y}_{0:t}) = \sum_{i=1}^{n}\sum_{j=1}^{2} w_t^{(i,j)} p\left(\boldsymbol{\beta}_t | \tau_t^{(i,j)}, \mathbf{y}_{\tau_t^{(i,j)}:t}\right) \tag{13}$$

where $p\left(\boldsymbol{\beta}_t | \tau_t^{(i,j)}, \mathbf{y}_{\tau_t^{(i,j)}:t}\right)$ can be obtained from Lemma 1.

It is well-known that the student's t-distribution in Lemma 2 can be well-approximated by normal distribution with the same mean and variance when its degrees of freedom $2a_{\tau_t,t-1}$ (or $2a_0$) ≥ 30 to reduce the computational cost, as proposed by Li and Moor.[33] In Section 4, we will use both the exact calculations for the pdf of student's t-distribution and their normal approximations to study the detection performance.

In summary, the proposed Rao-Blackwellized version of our SMC algorithm is given in Algorithm 2 as follows.

Initialization:
♦ Sample particles $\left\{\tau_0^{(i)}, w_0^{(i)}\right\}_{i=1}^{n}$ with $\tau_0^{(i)} = 0$ and $w_0^{(i)} = \frac{1}{n}$.
For $t = 1, \ldots, T$:
♦ For $i = 1, \ldots, n$:
○ Create two possible descendants $\tau_t^{(i,1)} = \tau_{t-1}^{(i)}, \tau_t^{(i,2)} = t$ of $\tau_{t-1}^{(i)}$, calculate their associated weights $w_t^{(i,1)}$, $w_t^{(i,2)}$ based on Eq. (12) and normalize the weights such that $\sum_{i=1}^{n}\sum_{j=1}^{2} w_t^{(i,j)} = 1$.
♦ Parameter estimation: estimate $p(\boldsymbol{\beta}_t|\mathbf{y}_{0:t})$ based on Eq. (13).
♦ Resampling: resample $\left\{\tau_t^{(i,j)}, w_t^{(i,j)}\right\}_{i=1}^{n}, j = 1, 2$ to generate n equally weighted particles $\left\{\tilde{\tau}_t^{(k)}, \frac{1}{n}\right\}_{k=1}^{n}$, and set $\left\{\tau_t^{(i)}, w_t^{(i)}\right\}_{i=1}^{n} = \left\{\tilde{\tau}_t^{(k)}, \frac{1}{n}\right\}_{k=1}^{n}$.

Algorithm 2. Rao-Blackwellized SMC Algorithm

3.3. *Optimal resampling*

A simple resampling method for the Rao-Blackwellized SMC algorithm discussed in previous section is the multinomial resampling method. However, the resampled particles using multinomial resampling suffer from a significant loss of diversity (with many duplicate particles) because the particles with significant importance weights $w_t^{(i)}$ are repeatedly selected many times. Due to the discrete nature of the change points τ_t in our model, having duplicate particles is wasteful as they contain exactly the same information as a single particle with its weight equal to the sum of the weights of the duplicate particles.

Therefore, at each time t we will combine duplicate particles into a *single* particle so that we only have distinctive particles before the resampling step. Then we will apply the *Optimal Resampling* (OR) method, which is proposed by Fearnhead and Clifford,[34] for the set of distinctive particles. Suppose at time $t-1$ we have n_d distinctive particles after resampling. Then after Rao-Blackwellization of $\tau_t^{(i)}$ at time t, there will be $2n_d$ particles $\left\{\tau_t^{(i,j)}, w_t^{(i,j)}\right\}_{i=1}^{n_d}$, $j = 1, 2$, where $\tau_t^{(i,1)} = \tau_{t-1}^{(i)}, \tau_t^{(i,2)} = t$ and $\sum_{i=1}^{n_d}\sum_{j=1}^{2} w_t^{(i,j)} = 1$. By combining the duplicate particles $\left\{\tau_t^{(i,2)} = t, w_t^{(i,2)}\right\}_{i=1}^{n_d}$, we will have n_d+1 distinctive particles $\left\{\tau_t^{(i)}, w_t^{(i)}\right\}_{i=1}^{n_d+1}$, where $\tau_t^{(i)} = \tau_t^{(i,1)}, w_t^{(i)} = w_t^{(i,1)}$, $i = 1, 2, \ldots, n_d; \tau_t^{(n_d+1)} = t, w_t^{(n_d+1)} = \sum_{i=1}^{n_d} w_t^{(i,2)}$. The following Algorithm 3, which is a special case of the OR method, can be used to resample n_d distinctive particles from the n_d+1 distinctive particles $\left\{\tau_t^{(i)}, w_t^{(i)}\right\}_{i=1}^{n_d+1}$.

• Step 1: calculate the unique solution c to the equation: $\sum_{i=1}^{n_d+1}\min\left(1, cw_t^{(i)}\right) = n_d$.

• Step 2: for $i = 1, \ldots, n_d + 1$, if $w_t^{(i)} \geq 1/c$ then particle $\tau_t^{(i)}$ is kept with its original weight $w_t^{(i)}$. Assume we keep q particles.

• Step 3: apply the stratified sampling algorithm in Appendix C to resample $n_d - q$ particles from the remaining $n_d + 1 - q$ particles, each resampled particle is assigned a weight 1/c.

Algorithm 3. OR Algorithm

The stratified sampling scheme used in the last step ensures each distinctive particle is resampled at most once so that the resulting n_d particles are distinctive. It is shown in Fearnhead and Clifford[34] that the resampled n_d samples $\left\{\widetilde{\tau}_t^{(i)}, \widetilde{w}_t^{(i)}\right\}_{i=1}^{n_d}$ are still properly weighted with respect to the desirable posterior distribution $P(\tau_t|\mathbf{y}_{0:t})$. The OR method has the computational complexity of $O(n)$. And it is optimal since it minimizes the expected squared error $E\left[\sum_{i=1}^{n_d+1}\left(W_t^{(i)} - w_t^{(i)}\right)^2\right]$, where $W_t^{(i)}$ is the random weight of a particle after resampling ($W_t^{(i)}$ equals $w_t^{(i)}$, 1/c or 0), which is proved in Fearnhead and Clifford.[34]

Note that when t is small (e.g., $t = 1, 2\ldots$), only t distinctive particles are needed, corresponding to $\tau_t = 1, \ldots, t$, respectively. When t is large, we need to limit the maximum number of distinctive particles after resampling at each time t. Let n_d denote this maximum limit and n_t ($n_t \leq n_d + 1$) be the number of distinctive particles after Rao-Blackwellization of $\tau_t^{(i)}$ at time t, with the properly weighted samples $\left\{\tau_t^{(i)}, w_t^{(i)}\right\}_{i=1}^{n_t}$ with respect to $P(\tau_t|\mathbf{y}_{0:t})$, the posterior density $p(\boldsymbol{\beta}_t|\mathbf{y}_{0:t})$ at time t can be approximated as

$$\hat{p}\left(\boldsymbol{\beta}_t|\mathbf{y}_{0:t}\right) = \sum_{i=1}^{n_t} w_t^{(i)} p\left(\boldsymbol{\beta}_t|\tau_t^{(i)}, \mathbf{y}_{\tau_t^{(i)}:t}\right) \tag{14}$$

The Rao-Blackwellized SMC algorithm with the OR method is summarized in Algorithm 4.

Set $n_t = 1$, $\tau_1^{(1)} = 1$ and $w_1^{(1)} = 1$.
For $t = 2, \ldots, T$:
♦ For $i = 1, \ldots, n_t$:
○ Create two possible descendants $\tau_t^{(i,1)} = \tau_{t-1}^{(i)}, \tau_t^{(i,2)} = t$ of each $\tau_{t-1}^{(i)}$, calculate their associated weights $w_t^{(i,1)}$, $w_t^{(i,2)}$ based on Eq. (12) and normalize the weights such that $\sum_{i=1}^{n_t}\sum_{j=1}^{2} w_t^{(i,j)} = 1$.
♦ Combine duplicate particles into $n_t + 1$ distinctive particles, the resulted samples are $\left\{\tau_t^{(i)}, w_t^{(i)}\right\}_{i=1}^{n_t+1}$, where $\tau_t^{(i)} = \tau_t^{(i,1)}, w_t^{(i)} = w_t^{(i,1)}$, $i = 1, 2, \ldots, n_t$; $\tau_t^{(n_t+1)} = t, w_t^{(n_t+1)} = \sum_{i=1}^{n_t} w_t^{(i,2)}$. Set $n_t = n_t + 1$.
♦ Parameter estimation: estimate $p(\boldsymbol{\beta}_t|\mathbf{y}_{0:t})$ based on Eq. (14).
♦ Resampling: if $n_t = n_d + 1$, using the OR method in Algorithm 3 to obtain $\left\{\widetilde{\tau}_t^{(i)}, \widetilde{w}_t^{(i)}\right\}_{i=1}^{n_d}$; set $\left\{\tau_t^{(i)}, w_t^{(i)}\right\}_{i=1}^{n_t-1} = \left\{\widetilde{\tau}_t^{(i)}, \widetilde{w}_t^{(i)}\right\}_{i=1}^{n_d}$ and $n_t = n_d$.

Algorithm 4. Rao-Blackwellized SMC Algorithm with Optimal Resampling

3.4. *Steady state detection rule*

As the steady state can be characterized by a sustained 'flat' line segment in the piecewise linear model of signals, we can develop our detection rule based on the slope parameter of the last line segment of the piecewise linear model. Therefore, we define the steady state detection index as $\pi_t \equiv \Pr(|\beta_{1t}| \leq s_0|\mathbf{y}_{0:t})$ where s_0 is the slope threshold. At time t, the detection index can be estimated as

$$\hat{\pi}_t = \sum_{i=1}^{n_t} w_t^{(i)} \Pr\left(|\beta_{1t}| \leq s_0|\tau_t^{(i)}, \mathbf{y}_{\tau_t^{(i)}:t}\right) \tag{15}$$

Based on Lemma 1, the marginal posterior distribution of β_{1t} follows a non-standardized student's t-distribution. However, computing cumulative probability for the non-standardized student's t-distribution is time-consuming. Note that the conditional posterior distribution of β_{1t} given σ_t^2 is

$$\beta_{1t}\,|\sigma_t^2, \tau_t^{(i)}, \mathbf{y}_{\tau_t^{(i)}:t} \sim N\left(\mu_{\tau_t^{(i)},t}^{(2)}, \sigma_t^2 \boldsymbol{\Sigma}_{\tau_t^{(i)},t}^{(2,2)}\right)$$

where $\mu^{(2)}_{\tau_t^{(i)},t}$ is the second element of $\boldsymbol{\mu}_{\tau_t^{(i)},t}$, $\Sigma^{(2,2)}_{\tau_t^{(i)},t}$ is the second diagonal element of $\boldsymbol{\Sigma}_{\tau_t^{(i)},t}$, and $\boldsymbol{\mu}_{\tau_t^{(i)},t}$ and $\boldsymbol{\Sigma}_{\tau_t^{(i)},t}$ are obtained using (11). If we replace σ_t^2 with its posterior mean value $E\left(\sigma_t^2|\tau_t^{(i)},\mathbf{y}_{\tau_t^{(i)}:t}\right)=\frac{b_{\tau_t^{(i)},t}}{a_{\tau_t^{(i)},t}-1}$, where $a_{\tau_t^{(i)},t}$, $b_{\tau_t^{(i)},t}$ are obtained using (11), the marginal posterior distribution of $\beta_{1t}^{(i)}$ can be approximated by the normal distribution as

$$\beta_{1t}|\tau_t^{(i)},\mathbf{y}_{\tau_t^{(i)}:t}\tilde{\sim}N\left(\boldsymbol{\mu}^{(2)}_{\tau_t^{(i)},t},\frac{b_{\tau_t^{(i)},t}}{a_{\tau_t^{(i)},t}-1}\boldsymbol{\Sigma}^{(2,2)}_{\tau_t^{(i)},t}\right)$$

Based on this approximation, the corresponding probabilities of each particle can be calculated much faster. From our experience $\hat{\pi}_t$ typically increases quickly and becomes close to one after the data reach the steady state. So the detection result is pretty robust to small errors in approximating $\hat{\pi}_t$.

At each time t, if $\hat{\pi}_t$ is larger than a given threshold value, π_0, we stop the algorithm and consider that a steady state is detected at time t. Similarly, since $\hat{\pi}_t$ typically increases quickly and becomes close to one after the data reach the steady state, the detection results are robust to small change of π_0. In this paper, we set π_0 to be 0.9 for all the examples.

4. Numerical Study

4.1. Simulated signals and model parameter setup

Simulations based on artificially generated signals are first conducted to evaluate the performance of our proposed steady state detection algorithm. A signal is generated based on the superposition of a bias functions and the noise. This paper uses four types of bias functions: linear, quadratic, exponential and oscillating, as shown in Table I. These are the most commonly tested functions for off-line steady state detection algorithms in the discrete-event simulation literature. For the bias direction of the first three functions, without loss of generality, we use the negative bias which represents the biased data starting below the steady state mean, as mentioned in Hoad *et al.*[14] A linearly decreasing function is chosen for the amplitude of the oscillating signals. For the noises, we use three types of autoregressive model: random Gaussian error (AR(0)), AR(1) and AR(2), as shown in the following Table II. The signal, $y(t)$, is generated based on the addition of the bias function $B(t)$ in Table I and the noise r_t in Table II, namely $y(t)=B(t)+r_t$, $t=1,\ldots,N$.

For all signals T_0 is the true transition point to the steady state. In selecting the prior parameters, a non-informative prior with $\boldsymbol{\mu}_0=[0\quad 0]'$ and $\boldsymbol{\Sigma}_0=\begin{bmatrix}10000 & 0\\ 0 & 10000\end{bmatrix}$ is used for $\boldsymbol{\beta}$. For the noise variance σ^2, its prior parameters a_0 and b_0 are set to be 10 and 0.1, respectively. For the prior transition probability p, any value between 0.1 and 0.5 can work well and here we use $p=0.2$. The maximum number of distinctive particles after resampling is chosen to be $n_d=16$ for all simulations.

Table I. Four types of bias functions

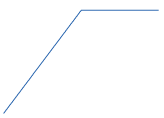
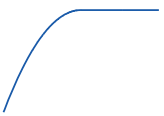
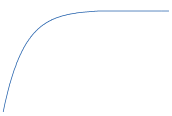
Signal	*Bias function*	*Shape*
Linear	$B(t)=\begin{cases}\frac{t}{T_0}h, & t=1,\ldots,T_0\\ h, & t=T_0+1,\ldots,N\end{cases}$	
Quadratic	$B(t)=\begin{cases}h\left(1-\frac{(t-T_0)^2}{(T_0-1)^2}\right), & t=1,\ldots,T_0\\ h, & t=T_0+1,\ldots,N\end{cases}$	
Exponential	$B(t)=\begin{cases}h\left(1-10^{-\frac{1-t}{T_0-1}}\right), & t=1,\ldots,T_0\\ y(T_0), & t=T_0+1,\ldots,N\end{cases}$	
Oscillating	$B(t)=\begin{cases}h\frac{T_0-t}{T_0-1}sin\left(\frac{\pi t}{f}\right), & t=1,\ldots,T_0\ \left(f=\frac{T_0}{10}\right)\\ 0, & t=T_0+1,\ldots,N\end{cases}$	

Table II. Equations and parameter values for three types of noises

Types	*Equation*	*Parameter values*
AR(0)	$r_t = \varepsilon_t$	$\varepsilon_t \sim N(0, \sigma_t^2)$
AR(1)	$r_t = \varphi_1 r_{t-1} + \varepsilon_t$	$\varphi_1 = 0.6$
AR(2)	$r_t = \varphi_2 r_{t-1} + \varphi_3 r_{t-2} + \varepsilon_t$	$\varphi_2 = -0.25, \varphi_3 = 0.5$

4.2. *Illustration of the steady state detection*

We illustrate our proposed algorithm on steady state detection by using a simulated linear signal with noise level $\sigma = 0.06$, $h = 1$ and $T_0 = 200$. The slope threshold s_0 is set to be 0.0025. Figure 3 shows the corresponding detection process. It can be seen that $\hat{\pi}_t$ jumps abruptly to large values close to 1 shortly after T_0 (the true steady state transition point). Besides, at time $t = 100$, 200 and 500, the estimate of the posterior probability of the latest change-point, $\hat{p}(\tau_t | \mathbf{y}_{0:t})$, is almost concentrated near the true change-points: 1 ($t \leq 200$) or 200 ($t > 200$). Processing each signal with 500 time steps by the proposed algorithm only takes an average of 0.9 seconds in MATLAB 2014b running on a 3.40 GHz Intel processor, which is much lower than that of the PF method in Wu *et al.*[26] (12 seconds for 500 time steps). Meanwhile, by sampling only one variable based on the Rao-Blackwellization method and employing the efficient OR resampling algorithm, our method uses at most 16 particles for each time step, which is substantially lower than the 1000 particles used for each time step by the PF method of Wu *et al.*[26] In addition, using normal approximations for calculating the pdf of student's *t*-distribution leads to further reduction of computational time to an average of 0.6 seconds and similar detection performance. This example shows that our algorithm can detect the change-point with timeliness and high computational efficiency.

Many signals in practice have a decaying variance with a fixed mean. When the signal enters the steady state, the variance is small and stable. To see how well our algorithm performs to detect the transition to steady state for such signals, we simulate the signal with zero mean and the noise amplitude as follows

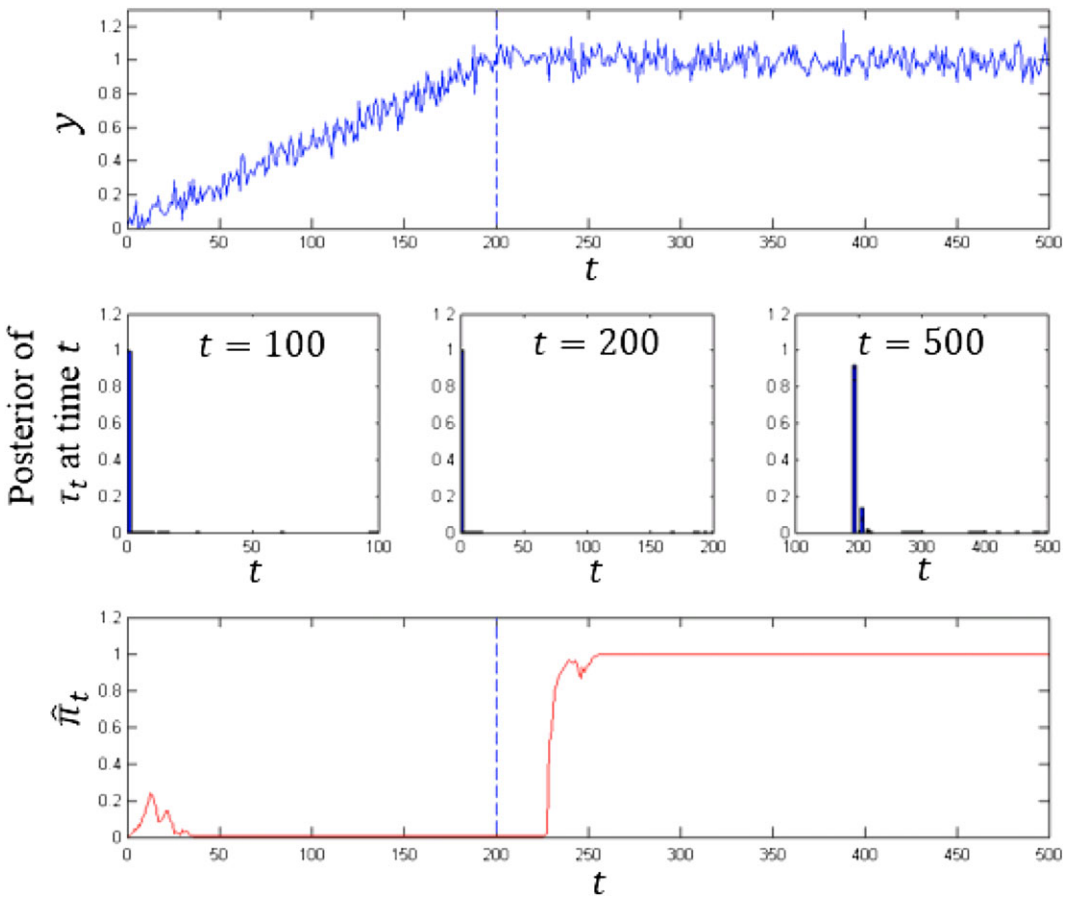


Figure 3. Illustration of steady state detection for linear signal: simulated signal (top), estimated posterior probability of the latest change-point at different time steps (middle), and the detection index $\hat{\pi}_t$ (bottom)

$$\sigma(t) = \begin{cases} 30^{(T_0-t)/(T_0-1)}\sigma_0, \text{if } t \le T_0 \\ \sigma_0 \qquad\qquad\qquad\quad , \text{if } t > T_0 \end{cases}$$

where $\sigma_0 = 0.1$ and $T_0 = 300$. The following Figure 4 shows that our detection algorithm is also well-performed for such a signal.

4.3. Performance evaluation

Following Wu *et al.*,[26] we consider the *false alarm rates* (FAR) and deviation of the estimated steady state transition point $^{\tau}$ from the true transition point T_0 (also called *detection bias*) to evaluate the performance of the proposed online steady state detection algorithm. The FAR is the probability that $\hat{\tau} < T_0$. It is considered because in some situations the cost of early detection is higher than that of delayed detection. The FAR usually serves as an auxiliary evaluation metric. The main evaluation metric in this paper is the *weighted standard detection error* (WSDE), as defined in Wu *et al.*[26]

$$WSDE = \sqrt{\frac{1}{N_s}\sum_{i=1}^{N_s} w(\hat{\tau}_i)(\hat{\tau}_i - T_0)^2} \qquad w(\hat{\tau}_i) = \begin{cases} w \in (0,1], \hat{\tau}_i \ge T_0 \\ 1, \qquad\quad \hat{\tau}_i < T_0 \end{cases} \tag{16}$$

where $\hat{\tau}_i$ denotes the estimated change-point in the i^{th} replication and N_s is the total number of replications used for each type of signals in the simulation. When $w = 1$, WSDE is the root mean square deviation (RMSD) of $\hat{\tau}_i$ from T_0. When $w < 1$, larger penalty is given to early detection than late detection in WSDE.

4.4. Comparison with existing online methods

In this section, we compare the proposed method in this paper with three other existing online steady state detection methods: the PF method proposed by Wu *et al.*,[26] the SDM method proposed by Holly *et al.*,[19] Bethea and Rhinehart,[20] Wu *et al.*,[4] and the VRT method proposed by Crow *et al.*,[18] Cao and Rhinehart.[2] The PF method proposed by Wu *et al.*[26] is the most recently developed online steady state detection method in the literature. It is based on the standard SMC, or the PF algorithm, and incorporates several improvement strategies such as the partial Gibbs resample-moves technique. The other two methods, the SDM and VRT, both incorporate a moving data window based on which they estimate either the slope or the variance to determine if the signal enters the steady state. In this section, we refer to the proposed Rao-Blackwellized SMC method in this paper as RBSMC method.

Each type of signal is generated by combining one of bias functions in Table I and one of noise autoregressive types in Table II. For the bias function, we set $h = 1$ and test two values of T_0: $T_0 = 200$ and $T_0 = 300$. Three noise amplitudes $\sigma_t = 0.06, 0.1, 0.14$ are used for

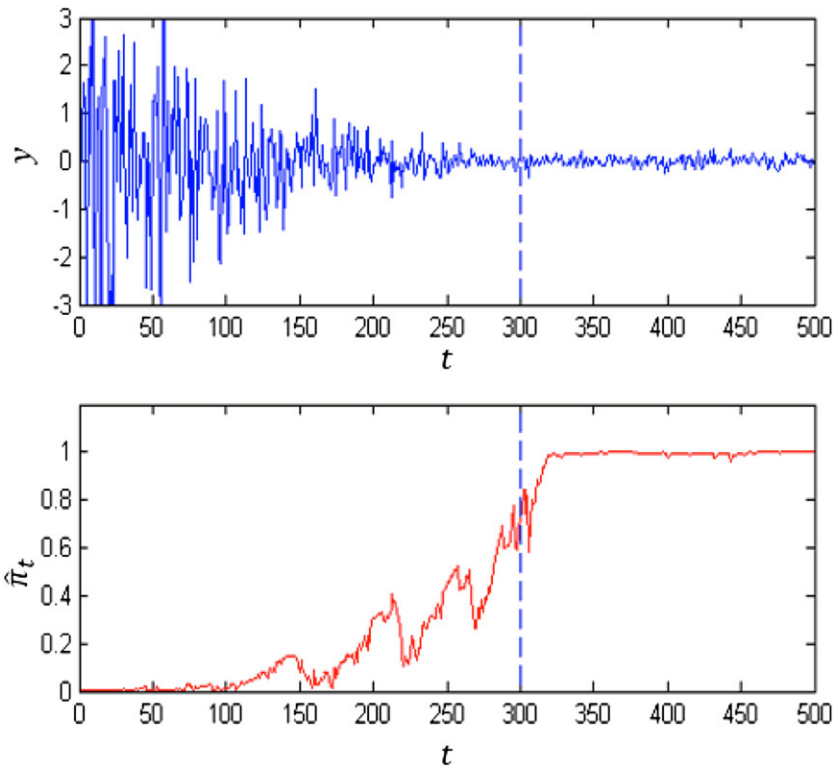


Figure 4. Steady state detection for a signal with zero mean and exponentially decaying variance: simulated signal (top), and the detection index $\hat{\pi}_t$ (bottom)

AR(0) and $\sigma_t = 0.06$, 0.1 are used for AR(1) and AR(2). As in Wu *et al.*,[26] we use $N_s = 500$ replications for each type of signal and the detection parameters (window size, thresholds) of all methods are selected to minimize the WSDE with $w = 1$. For our proposed method, we set slope threshold $s_0 = 0.0021$ for all signals.

Figure 5 shows the WSDE as functions of the penalty weight w for each noise autoregressive type. As we can see, the proposed RBSMC method and PF are much more accurate than SDM, VRT in terms of the WSDE in all penalty weights. Besides, the proposed RBSMC method is slightly better comparing with the PF method in most of the penalty weights.

The detailed results for the AR(0) noise with $w = 1$ are shown in Table III. From Table III, it can be seen that the two SMC-based methods (RBSMC and PF) offer the most competitive detection results in terms of the WSDE and FAR with signals of various bias

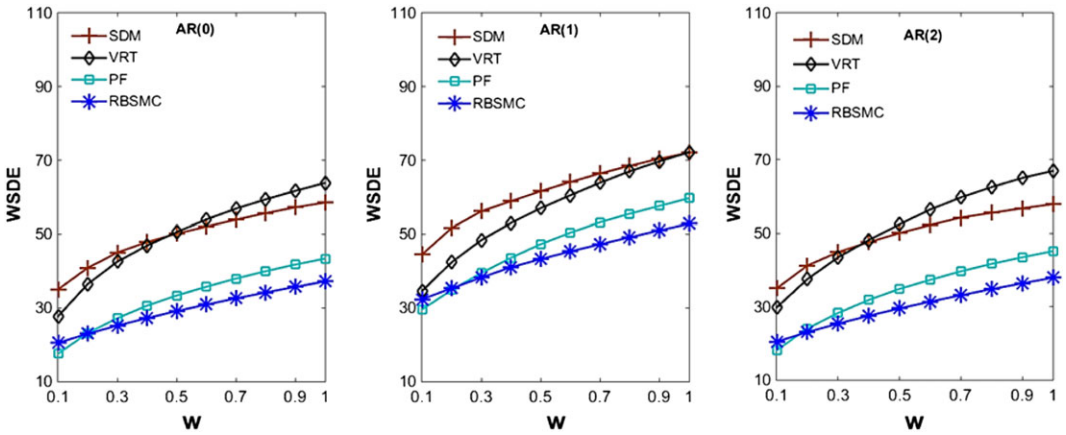


Figure 5. The weighted standard detection error (WSDE) of the proposed RBSMC, PF, SDM, VRT as a function of the penalty weight w for AR(0), AR(1) and AR(2)

Table III. Comparison of RBSMC, PF, SDM and VRT for w = 1 and noise type AR(0). The detection parameters are (1) RBSMC, $S_0 = 0.0021$; (2) PF, $S_0 = 0.0022$; (3) SDM, window size m = 50, threshold = 8×10^{-5}; (4) VRT, m = 98, threshold = 0.6

Signal		σ	WSDE				FAR			
			RBSMC	PF	SDM	VRT	RBSMC	PF	SDM	VRT
Linear	$h = 1, T_0 = 200$	0.06	52.6	42.0	59.8	78.6	0	0	0	0
		0.10	55.7	53.9	60.7	66.1	0	0	0	0
		0.14	59.6	64.5	57.6	60.4	0	0	0	0
	$h = 1, T_0 = 300$	0.06	46.7	40.9	58.0	70.9	0	0	0	0
		0.10	49.9	53.1	55.9	56.6	0	0	0	0.04
		0.14	53.5	70.4	58.6	121.6	0.01	0.01	0.03	0.75
Quadratic	$h = 1, T_0 = 200$	0.06	17.6	12.1	33.6	37.6	0.01	0.11	0	0
		0.10	22.8	21.2	31.9	26.7	0.03	0.04	0.02	0.04
		0.14	26.3	33.7	28.2	18.5	0.06	0.06	0.14	0.36
	$h = 1, T_0 = 300$	0.06	20.0	33.8	22.4	16.3	0.89	1	0.11	0.33
		0.10	24.4	28.1	23.9	37.5	0.81	0.93	0.45	0.83
		0.14	29.9	22.4	34.5	72.8	0.71	0.62	0.62	0.93
Exponential	$h = 1, T_0 = 200$	0.06	23.3	16.6	45.9	44.4	0.01	0.12	0	0
		0.10	28.1	26.3	40.5	23.2	0.07	0.08	0.03	0.30
		0.14	31.7	35.0	34.8	26.1	0.12	0.06	0.17	0.73
	$h = 1, T_0 = 300$	0.06	37.6	61.7	35.5	27.2	0.93	1	0.04	0.51
		0.10	45.4	49.4	32.2	67.3	0.85	0.98	0.38	0.96
		0.14	51.1	40.1	55.7	107.8	0.85	0.88	0.82	0.99
Oscillating	$h = 1, T_0 = 200$	0.06	23.1	27.1	94.9	74.4	0	0.01	1	0
		0.10	29.3	27.6	90.2	61.5	0	0.04	0.99	0
		0.14	38.7	26.6	94.0	54.1	0	0.06	1	0
	$h = 1, T_0 = 300$	0.06	9.4	23.0	156	63.6	0.39	0.04	1	0
		0.10	19.5	25.7	156	49.8	0.21	0.2	0.99	0
		0.14	32.6	29.4	152	40.6	0.07	0.55	1	0
Overall			37.3	39.1	59.9	60.3	0.25	0.28	0.38	0.28

functions. Both of them have consistently better performance than the moving-window based methods (SDM and VRT). The advantage of the SMC-based methods over the moving-window based methods can be understood intuitively as follows: The SMC-based methods are based on the piecewise linear model in (1) with multiple unknown change-points. Therefore they behave as methods with *adaptive* "window" sizes. Compared with the moving-window based methods with a *fixed* "window" size, SMC-based methods are very flexible in adjusting their "window" sizes based on the observed signals to give much more robust detection performance.

Comparing between the RBSMC and PF methods, the RBSMC has slightly better overall performance in terms of smaller WSDE (37.3 vs. 39.1) and smaller FAR (0.25 vs. 0.28), and using normal approximations for calculating the pdf of student's *t*-distribution in RBSMC leads to the same overall performance (see Appendix D for detailed results). Most importantly, the main advantage of the RBSMC method to the PF method is in the substantial saving of computational cost. This significant saving of computational cost makes the RBSMC method a much more preferred method for many online applications where quick processing of the data in real time is critical.

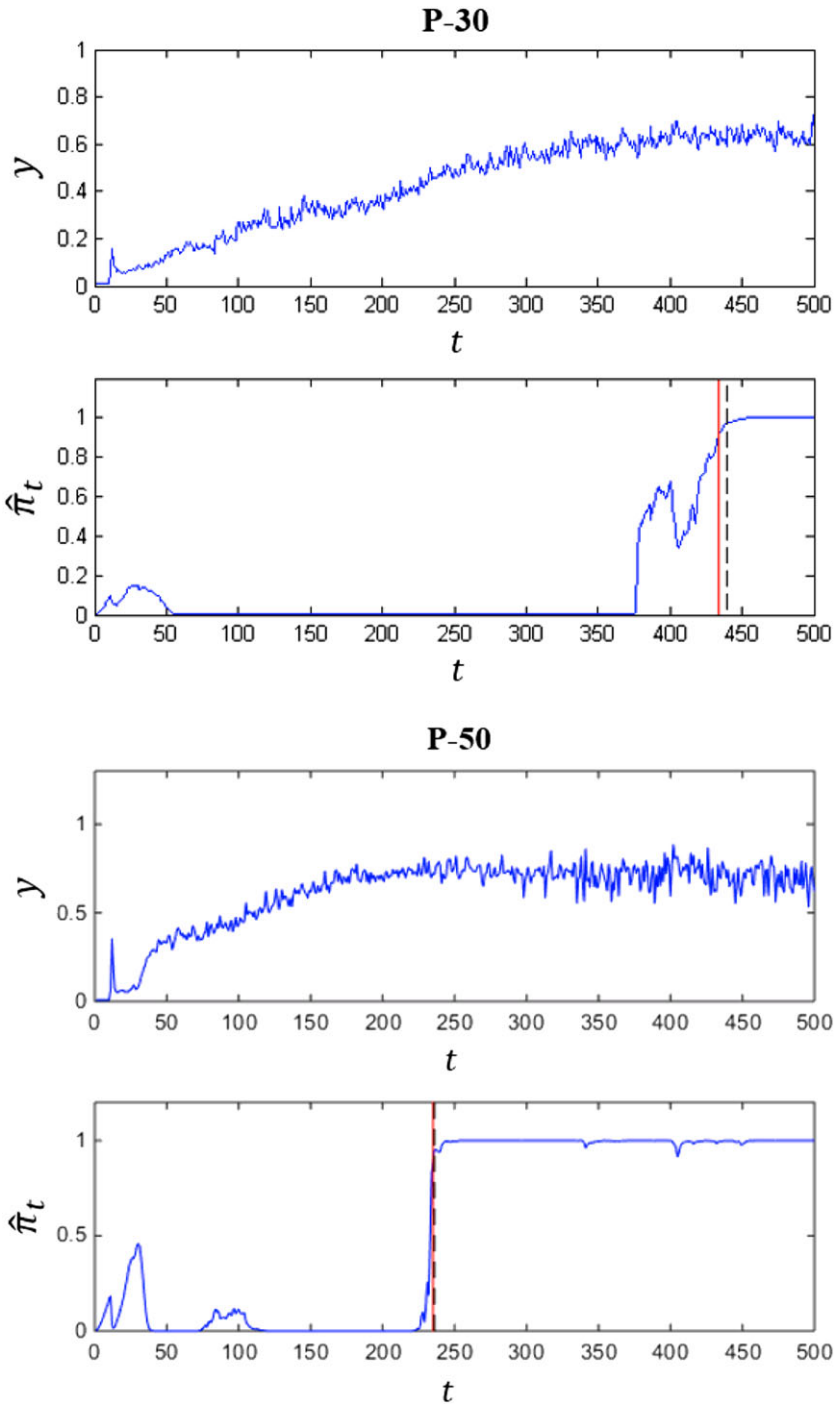


Figure 6. Steady state detection for the CNP signal with ultrasonic power 30 W and 50 W in the dispersion of 30 g Al_2O_3

4.5. *Application to steady state detection in micro/nanoparticle dispersion process*

In this subsection, we apply the RBSMC algorithm to real signals called *cavitation noise power* (CNP) signals from the ultrasonic-cavitation based nanoparticle dispersion process. Currently, the micro/nanoparticle research has attracted intense scientific interests because of its potential applications in biomedical, optical and electrical fields. In these applications, micro/nanoparticles need to be dispersed evenly into the base materials before use to improve the material properties. However, the particles often cluster together as a result of high surface energy and large surface-to-volume ratio. The ultrasonic cavitation method can be used for effective dispersion of micro/nanoparticles. The dispersion process can be monitored by detecting the steady state of CNP signals based on the fact that the steady state of CNP signals corresponds to the maximum dispersion extent at the ultrasonic power level. Please refer to Wu *et al.*[4] for details on the ultrasonic cavitation process, the experimental setup used to collect the data, and the method to obtain the CNP signals.

Figure 6 shows the detection results for the CNP signals with ultrasonic power 30 W and 50 W, respectively, in the dispersion of 30 g Al_2O_3. The offline method EWMA-MSER in Wu *et al.*[4] is used as a benchmark method to evaluate the proposed method. It can be seen that the detection results (red solid line) of the proposed RBSMC method are quite close to those of the offline method (black dashed line).

5. Summary

In this paper, we study the problem of online steady state detection using a multiple change-point model and sequential Monte Carlo methods. A piecewise linear model is used to approximate the signal. Within the Bayesian framework, the posterior densities of model parameters can be sequentially updated given the latest observations. The stopping criterion for detecting the steady state is established based on the fact that the steady state can be characterized by a sustained 'flat' line segment in the piecewise linear model.

The main contribution of our proposed algorithm is its high computational efficiency based on the Rao-Blackwellization technique. By solving analytically for the conditional distribution of the model parameters, estimating the joint posterior distribution of four variables in the state vector requires sampling from only a one dimensional posterior distribution. This leads to substantial variance reduction of the Monte Carlo estimates and the use of significantly smaller number of particles than the standard SMC algorithm, while achieving comparable or better estimation accuracy. In addition, by applying the so-called Optimal Resampling method and eliminating duplicate particles, the robustness and timeliness of steady state detection is significantly improved by using the information of the particles more efficiently.

The performance of our proposed method is evaluated through both artificially simulated signals and a real data example from the ultrasonic-cavitation based nanoparticle dispersion process. Results demonstrate the robustness of the proposed algorithm for various types of signals with different levels of noises, and much faster computational time compared to the standard PF method. Note that in this paper, only non-informative priors are selected for our change-point model parameters. However, in practice, we may have better prior knowledge for the signals, such as slopes, bias shapes, levels of noises, and steady state transition locations. Therefore, under the Bayesian framework we can easily use more informative priors of model parameters in our steady state detection algorithm to take advantage of the prior knowledge on the process and further improve the detection performance. Note that in our current method, we only consider the problem with univariate observations. A potential future work is to develop methods for multivariate data. Additionally, the change point model used in this paper can be applied to develop a Bayesian statistical process control (SPC) method for short-run production processes, where we can conduct on-line inference of process parameters to determine whether they have shifted into the out-of-control range. This is our ongoing research.

Acknowledgements

This paper is based upon works supported by the National Science Foundation under Grants CMMI-1335454 and IIS-1343974, and by the DDDAS program of AFOSR.

References

1. Mahuli SK, Rhinehart RR, Riggs JB. Experimental demonstration of non-linear model-based in-line control of pH. *Journal of Process Control* 1992; **2**:145–153.
2. Cao S, Rhinehart RR. An efficient method for on-line identification of steady state. *Journal of Process Control* 1995; **5**:363–374.
3. Jiang T, Chen B, He X, Stuart P. Application of steady-state detection method based on wavelet transform. *Computers and Chemical Engineering* 2003; **27**:569–578.
4. Wu J, Zhou S, Li X. Acoustic emission monitoring for ultraonic cavitation based dispersion process. *Journal of Manufacturing Science and Engineering* 2013; **135**:031015.
5. Narasimhan S, Mah RSH, Tamhane AC, Woodward JW, Hale JC. A composite statistical test for detecting changes of steady states. *AIChE Journal* 1986; **32**:1409–1418.
6. Bagajewicz MJ, Jiang Q. Comparison of steady state and integral dynamic data reconciliation. *Computers and Chemical Engineering* 2000; **24**:2367–2383.
7. Bhat SA, Saraf DN. Steady-state identification, gross error detection, and data reconcilation for industrial process units. *Industrial & Engineering Chemistry Research* 2004; **43**:4323–4336.
8. Korbel M, Bellec S, Jiang T, Stuart P. Steady state identification for on-line data reconcilation based on wavelet transform and filtering. *Computers and Chemical Engineering* 2014; **63**:206–218.
9. Kim M, Yoon SH, Domanski PA, Payne WV. Design of a steady-state detector for fault detection and diagnosis of a residential air conditioner. *International Journal of Refrigeration* 2008; **31**:790–799.

10. Mhamdi A, Geffers W, Flehmig F, Marquardt W. On-line optimization of MSF desalination plants. *Desalination and Water Resources: Thermal Desalination Processes* 2010; **1**:136–162.
11. Kelton WD, Law AM. A new approach for dealing with the startup problem in discrete event simulation. *Naval Research Logistics Quarterly* 1983; **30**:641–658.
12. Gallagher MA, Bauer KW, Maybeck PS. Initial data truncation for univariate output of discrete-event simulations using the Kalman Filter. *Management Science* 1996; **42**:1996.
13. Fishman GS. *Discrete-event simulation, modelling, programming, and analysis*. Springer-Verlag: New York, 2001.
14. Hoad K, Robinson S, Davies R. Automating warm-up length estimation. *Journal of the Operational Research Society* 2010; **61**:1389–1403.
15. Robinson S. A statistical process control approach to selecting a warm-up period for a discrete-event simulation. *European Journal of Operational Research* 2007; **176**:332–346.
16. Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 1964; **36**:1627–1639.
17. Roux GACL, Santoro BF, Sotelo FF, Teissier M, Joulia X. Improving steady-state identification. *18th European Symposium on Computer Aided Process Engineering* 2008; **25**:459–464.
18. Crow EL, Davis FA, Maxfiled MW. *Statistics Manual: with examples taken from ordnance development*. Dover Publications: New York, 1960.
19. Holly W, Cook R, Crowe C. Reconciliation of mass flow rate measurements in a chemical extraction plant. *The Canadian Journal of Chemical Engineering* 1989; **67**:595–601.
20. Bethea RM, Rhinehart RR. Applied engineering statistics. Marcel Dekker, Inc.: New York, NY, 1991.
21. Narasimhan S, Kao CS, Mah R. Detecting changes of steady states using the mathematical theory of evidence. *AIChE Journal* 1987; **33**:1930–1932.
22. Liu JS. Monte Carlo strategies in scientific computing. Springer: New York, NY, 1998.
23. Robert CP, Casella G. Monte Carlo statistical methods. Springer: New York, NY, 2004.
24. Doucet A, de Freitas N, Gordon N. Sequential Monte Carlo methods in practice. Springer: Berlin Heidelberg, New York, 2001.
25. Arulampalam MS, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-Gaussion Bayesian tracking. *IEEE Transactions on Signal Processing* 2002; **50**:174–188.
26. Wu J, Chen Y, Zhou S, Li X. Online Steady-state Detection for Process Control Using Multiple Change-point Models and Particle Filters. *IEEE Transactions on Automation Science and Engineering* 2015; **13**:688–700.
27. Kalman RE. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 1960; **82**:35–45.
28. Liu JS, Chen R. Sequential Monte Carlo Methods for Dynamic Systems. *Journal of the American Statistical Association* 1998; **93**:1032–1044.
29. Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* 2000; **10**:197–208.
30. Casella G, Robert CP. Rao-Blackwellization of sampling schemes. *Biometrika* 1996; **83**:81–94.
31. O'Hagan A. Bayesian Inference. Kendall's Advanced Theory of Statistics, Vol. **2B**. Hodder Headline Group: London, 1994.
32. Chopin N. Dynamic detection of change points in long time series. *Annals of the Institute of Statistical Mathematics* 2007; **59**:349–366.
33. Li B, Moor DB. A corrected normal approximation for the Student's *t* distribution. *Computational Statistics and Data Analysis* 1999; **29**:213–216.
34. 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**:887–899.
35. Carpenter J, Clifford P, Fearnhead P. An improved particle filter for non-linear problems. *IEE Proceedings Radar Sonar and Navigation* 1999; **146**:2–7.

Appendix A

Proof of Lemma 1

The joint posterior distribution of $\boldsymbol{\beta}_t$ and σ_t^2 is:

$$\begin{aligned}
&p(\boldsymbol{\beta}_t, \sigma_t^2|\tau_t, \mathbf{y}_{\tau_t:t}) = p(\boldsymbol{\beta}_t|\sigma_t^2, \tau_t, \mathbf{y}_{\tau_t:t})p(\sigma_t^2|\tau_t, \mathbf{y}_{\tau_t:t}) \\
&\propto (\sigma_t^2)^{-1} \exp\left\{-\frac{1}{2\sigma_t^2}(\boldsymbol{\beta}_t-\boldsymbol{\mu}_{\tau_t,t})'\boldsymbol{\Sigma}_{\tau_t,t}^{-1}(\boldsymbol{\beta}_t-\boldsymbol{\mu}_{\tau_t,t})\right\}(\sigma_t^2)^{-(a_{\tau_t,t}+1)} \exp\left\{-\frac{b_{\tau_t,t}}{\sigma_t^2}\right\} \\
&= (\sigma_t^2)^{-(a_{\tau_t,t}+1+1)} \exp\left\{-\frac{1}{\sigma_t^2}\left(b_{\tau_t,t}+\frac{1}{2}(\boldsymbol{\beta}_t-\boldsymbol{\mu}_{\tau_t,t})'\boldsymbol{\Sigma}_{\tau_t,t}^{-1}(\boldsymbol{\beta}_t-\boldsymbol{\mu}_{\tau_t,t})\right)\right\}
\end{aligned}$$

Let $A = b_{\tau_t,t} + \frac{1}{2}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_{\tau_t,t})'\boldsymbol{\Sigma}_{\tau_t,t}^{-1}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_{\tau_t,t})$

The marginal posterior distribution of $\boldsymbol{\beta}_t$ can be obtained after integrating out σ_t^2:

$$\begin{aligned}
&p(\boldsymbol{\beta}_t|\tau_t, \mathbf{y}_{\tau_t:t}) = \int p(\boldsymbol{\beta}_t, \sigma_t^2|\tau_t, \mathbf{y}_{\tau_t:t})d\sigma_t^2 \propto \int (\sigma_t^2)^{-(a_{\tau_t,t}+1+1)} \exp\left\{-\frac{1}{\sigma_t^2}A\right\}d\sigma_t^2 \\
&= \int \frac{A^{a_{\tau_t,t}+1}}{\Gamma(a_{\tau_t,t}+1)}(\sigma_t^2)^{-(a_{\tau_t,t}+1+1)} \exp\left\{-\frac{1}{\sigma_t^2}A\right\}d\sigma_t^2 \frac{\Gamma(a_{\tau_t,t}+1)}{A^{a_{\tau_t,t}+1}} \propto A^{-(a_{\tau_t,t}+1)} \\
&= \left(b_{\tau_t,t}+\frac{1}{2}(\boldsymbol{\beta}_t-\boldsymbol{\mu}_{\tau_t,t})'\boldsymbol{\Sigma}_{\tau_t,t}^{-1}(\boldsymbol{\beta}_t-\boldsymbol{\mu}_{\tau_t,t})\right)^{-(a_{\tau_t,t}+1)} \\
&\propto \left(1+\frac{1}{2a_{\tau_t,t}}\frac{(\boldsymbol{\beta}_t-\boldsymbol{\mu}_{\tau_t,t})'\boldsymbol{\Sigma}_{\tau_t,t}^{-1}(\boldsymbol{\beta}_t-\boldsymbol{\mu}_{\tau_t,t})}{\frac{b_{\tau_t,t}}{a_{\tau_t,t}}}\right)^{-\left(\frac{2a_{\tau_t,t}+2}{2}\right)}
\end{aligned}$$

Therefore $(\boldsymbol{\beta}_t|\tau_t, \mathbf{y}_{\tau_t:t}) \sim t_2\left(\boldsymbol{\mu}_{\tau_t,t}, \frac{b_{\tau_t,t}}{a_{\tau_t,t}}\boldsymbol{\Sigma}_{\tau_t,t}, 2a_{\tau_t,t}\right)$.

Appendix B

Proof of Lemma 2

a If $\tau_t = \tau_{t-1}$, we are making posterior inferences based on the observations $\mathbf{y}_{\tau_{t-1}:t-1}$.

Since $\left(\boldsymbol{\beta}_{t-1}|\sigma^2_{t-1}, \tau_t = \tau_{t-1}, \mathbf{y}_{\tau_{t-1}:t-1}\right) \sim N\left(\boldsymbol{\mu}_{\tau_t,t-1}, \sigma^2_{t-1}\boldsymbol{\Sigma}_{\tau_t,t-1}\right)$, we have $\left(\mathbf{X}_t\boldsymbol{\beta}_{t-1}|\sigma^2_{t-1}, \tau_t = \tau_{t-1}, \mathbf{y}_{\tau_{t-1}:t-1}\right) \sim N\left(\mathbf{X}_t\boldsymbol{\mu}_{\tau_t,t-1}, \sigma^2_{t-1}\mathbf{X}_t\boldsymbol{\Sigma}_{\tau_t,t-1}\mathbf{X}_t^{'}\right)$, and the conditional distribution of y_t would be $\left(y_t|\sigma^2_{t-1}, \tau_t = \tau_{t-1}, \mathbf{y}_{\tau_{t-1}:t-1}\right) \sim N\left(\mathbf{X}_t\boldsymbol{\mu}_{\tau_t,t-1}, \sigma^2_{t-1}\left(1 + \mathbf{X}_t\boldsymbol{\Sigma}_{\tau_t,t-1}\mathbf{X}_t^{'}\right)\right)$, The joint posterior distribution of y_t, σ^2_{t-1} is:

$$
\begin{aligned}
&\mathrm{p}\left(\mathbf{y}_t, \sigma^2_{t-1}|\tau_t = \tau_{t-1}, \mathbf{y}_{\tau_{t-1}:t-1}\right) = \mathrm{p}\left(\mathbf{y}_t|\sigma^2_{t-1}, \tau_t = \tau_{t-1}, \mathbf{y}_{\tau_{t-1}:t-1}\right)\mathrm{p}\left(\sigma^2_{t-1}|\tau_t = \tau_{t-1}, \mathbf{y}_{\tau_{t-1}:t-1}\right)\\
&\propto\left(\sigma^2_{t-1}\right)^{-1/2}\exp\left\{-\frac{\left(y_t - \mathbf{X}_t\boldsymbol{\mu}_{\tau_t,t-1}\right)^2}{2\sigma^2_{t-1}\left(1 + \mathbf{X}_t\boldsymbol{\Sigma}_{\tau_t,t-1}\mathbf{X}_t^{'}\right)}\right\}\left(\sigma^2_{t-1}\right)^{-\left(a_{\tau_t,t-1}+1\right)}\exp\left\{-\frac{b_{\tau_t,t-1}}{\sigma^2_{t-1}}\right\}\\
&= \left(\sigma^2_{t-1}\right)^{-\left(a_{\tau_t,t-1}+\frac{1}{2}+1\right)}\exp\left\{-\frac{1}{\sigma^2_{t-1}}\left(b_{\tau_t,t-1} + \frac{\left(y_t - \mathbf{X}_t\boldsymbol{\mu}_{\tau_t,t-1}\right)^2}{2\left(1 + \mathbf{X}_t\boldsymbol{\Sigma}_{\tau_t,t-1}\mathbf{X}_t^{'}\right)}\right)\right\}
\end{aligned}
$$

Let $B = b_{\tau_t,t-1} + \frac{\left(y_t - \mathbf{X}_t\boldsymbol{\mu}_{\tau_t,t-1}\right)^2}{2\left(1+\mathbf{X}_t\boldsymbol{\Sigma}_{\tau_t,t-1}\mathbf{X}_t^{'}\right)}$ The posterior predictive distribution of y_t can be obtained after integrating out σ^2_{t-1}:

$$
\begin{aligned}
&p\left(y_t|\tau_t = \tau_{t-1}, \mathbf{y}_{\tau_{t-1}:t-1}\right) = \int p\left(y_t, \sigma^2_{t-1}|\tau_t = \tau_{t-1}, \mathbf{y}_{\tau_{t-1}:t-1}\right)d\sigma^2_{t-1}\\
&\propto\int\left(\sigma^2_{t-1}\right)^{-\left(a_{\tau_t,t-1}+\frac{1}{2}+1\right)} exp\left\{-\frac{1}{\sigma^2_{t-1}}B\right\}d\sigma^2_{t-1}\\
&= \int\frac{B^{a_{\tau_t,t-1}+\frac{1}{2}}}{\Gamma\left(a_{\tau_t,t-1}+\frac{1}{2}\right)}\left(\sigma^2_{t-1}\right)^{-\left(a_{\tau_t,t-1}+\frac{1}{2}+1\right)} exp\left\{-\frac{1}{\sigma^2_{t-1}}B\right\}d\sigma^2_{t-1}\frac{\Gamma\left(a_{\tau_t,t-1}+\frac{1}{2}\right)}{B^{a_{\tau_t,t-1}+\frac{1}{2}}} \propto B^{-\left(a_{\tau_t,t-1}+\frac{1}{2}\right)}\\
&= \left(b_{\tau_t,t-1} + \frac{\left(y_t - \mathbf{X}_t\boldsymbol{\mu}_{\tau_t,t-1}\right)^2}{2\left(1 + \mathbf{X}_t\boldsymbol{\Sigma}_{\tau_t,t-1}\mathbf{X}_t^{'}\right)}\right)^{-\left(a_{\tau_t,t-1}+\frac{1}{2}\right)}\\
&\propto\left(1 + \frac{1}{2a_{\tau_t,t-1}}\frac{\left(y_t - \mathbf{X}_t\boldsymbol{\mu}_{\tau_t,t-1}\right)^2}{\frac{b_{\tau_t,t-1}}{a_{\tau_t,t-1}}\left(1 + \mathbf{X}_t\boldsymbol{\Sigma}_{\tau_t,t-1}\mathbf{X}_t^{'}\right)}\right)^{-\left(\frac{2a_{\tau_t,t-1}+1}{2}\right)}
\end{aligned}
$$

Therefore $\left(y_t|\tau_t = \tau_{t-1}, \mathbf{y}_{\tau_{t-1}:t-1}\right) \sim t_1\left(\mathbf{X}_t\boldsymbol{\mu}_{\tau_t,t-1}, \frac{b_{\tau_t,t-1}}{a_{\tau_t,t-1}}\left(1 + \mathbf{X}_t\boldsymbol{\Sigma}_{\tau_t,t-1}\mathbf{X}_t^{'}\right), 2a_{\tau_t,t-1}\right)$

b If $\tau_t = t$, y_t is independent of $\mathbf{y}_{0:t-1}$ and we are making posterior inferences based on the prior information. That is to say, $p\left(y_t|\tau_t = t, \mathbf{y}_{\tau_t:t-1}\right) = p\left(y_t\right)$. Therefore, we have $\left(y_t|\tau_t = t, \mathbf{y}_{\tau_t:t-1}\right) \sim t_1\left(\mathbf{X}_t\boldsymbol{\mu}_0, \frac{b_0}{a_0}\left(1 + \mathbf{X}_t\boldsymbol{\Sigma}_0\mathbf{X}_t^{'}\right), 2a_0\right)$.

Appendix C

Let $w_t^{(i)}$, $i = 1, 2, \ldots, n_d + 1 - q$ be the weights of remaining particles to be resampled, a total of $n_d - q$ particles need to be resampled. With the unique solution c, we apply the stratified sampling algorithm of Carpenter *et al.*[35]:

- Initialize: simulate s as the realization of a uniform random variable on [0,1/c], and set i = 1.
- While $i \le n_d + 1$ - q do

    $s = s$-$w_t^{(i)}$.

    If $s < 0$ do

        Resample particle i, assign it with weight 1/c;

        $s = s + 1/c$.

    End

    $i = i + 1$.

End

Appendix D

RBSMC detection results for AR(0) noise with both exact calculations and normal approximations for the pdf of student's *t*-distribution:

Signal		σ	WSDE		FAR	
			Exact	Approximation	Exact	Approximation
Linear		0.06	52.6	52.6	0	0
	$h = 1, T_0 = 200$	0.10	55.7	55.0	0	0
		0.14	59.6	59.6	0	0
		0.06	46.7	45.9	0	0
	$h = 1, T_0 = 300$	0.10	49.9	49.5	0	0
		0.14	53.5	54.5	0.01	0.01
		0.06	17.6	17.4	0.01	0.01
	$h = 1, T_0 = 200$	0.10	22.8	22.8	0.03	0.02
Quadratic		0.14	26.3	27.5	0.06	0.06
		0.06	20.0	21.3	0.89	0.93
	$h = 1, T_0 = 300$	0.10	24.4	23.7	0.81	0.77
		0.14	29.9	28.7	0.71	0.75
		0.06	23.3	24.5	0.01	0.01
	$h = 1, T_0 = 200$	0.10	28.1	28.9	0.07	0.06
Exponential		0.14	31.7	32.5	0.12	0.11
		0.06	37.6	38.9	0.93	0.92
	$h = 1, T_0 = 300$	0.10	45.4	44.0	0.85	0.87
		0.14	51.1	52.1	0.85	0.86
		0.06	23.1	23.4	0	0
	$h = 1, T_0 = 200$	0.10	29.3	28.6	0	0
Oscillating		0.14	38.7	40.0	0	0
		0.06	9.4	9.8	0.39	0.39
	$h = 1, T_0 = 300$	0.10	19.5	20.8	0.21	0.19
		0.14	32.6	31.4	0.07	0.05
	Overall		37.3	37.3	0.25	0.25

Authors' biographies

Yuxing Hou received the BS degree in Mechanical Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2012 and the MS degree in Statistics from University of Iowa, Iowa City, IA, USA, in 2015. He is currently pursuing the PhD degree in Industrial Engineering at University of Iowa. His research interest is process monitoring and statistical quality control.

Jianguo Wu is an Assistant Professor in the Department of Industrial, Manufacturing and Systems Engineering at University of Texas-El Paso, TX. He received the BS degree in Mechanical Engineering from Tsinghua University, Beijing, China, in 2009, the MS degree in Mechanical Engineering from Purdue University in 2011, and MS degree in Statistics in 2014 and PhD degree in Industrial and Systems Engineering in 2015, both from University of Wisconsin-Madison. His research interests are focused on statistical modeling, monitoring and analysis of advanced manufacturing processes, engineering and service systems for quality control, fault diagnosis and prognosis through integrated application of metrology, engineering domain knowledge, and advanced data analytics.

Yong Chen is currently a professor in the Department of Mechanical and Industrial Engineering at the University of Iowa. He received the BE degree in Computer Science from Tsinghua University, China, in 1998, and the master's degree in Statistics and PhD degree in Industrial and Operations Engineering, both from the University of Michigan in 2003. His research interests include reliability modeling, novelty detection, process monitoring, diagnosis and prognosis, and maintenance decision-making. He received best paper awards from IIE Transactions in 2010 and 2004. He has served as an associate editor for Technometrics, an editorial board member of Journal of Quality Technology, and an associate professor for Naval Research Logistics. He is a member of the INFORMS, the ASA, and the IMS.