



## Online detection of steady-state operation using a multiple-change-point model and exact Bayesian inference

Jianguo Wu, Yong Chen & Shiyu Zhou

To cite this article: Jianguo Wu, Yong Chen & Shiyu Zhou (2016) Online detection of steady-state operation using a multiple-change-point model and exact Bayesian inference, IIE Transactions, 48:7, 599-613, DOI: [10.1080/0740817X.2015.1110268](https://doi.org/10.1080/0740817X.2015.1110268)

To link to this article: <http://dx.doi.org/10.1080/0740817X.2015.1110268>



Accepted author version posted online: 15 Dec 2015.  
Published online: 15 Dec 2015.



[Submit your article to this journal](#)



Article views: 66



[View related articles](#)



[View Crossmark data](#)

# Online detection of steady-state operation using a multiple-change-point model and exact Bayesian inference

Jianguo Wu<sup>a</sup>, Yong Chen<sup>b</sup> and Shiyu Zhou<sup>c</sup>

<sup>a</sup>Department of Industrial, Manufacturing & Systems Engineering, University of Texas at El Paso, TX, USA; <sup>b</sup>Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City, IA, USA; <sup>c</sup>Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI, USA

## ABSTRACT

The detection of steady-state operation is critical in system/process performance assessment, optimization, fault detection, and process automation and control. In this article, we propose a new robust and computationally efficient online steady-state detection method using multiple change-point models and exact Bayesian inference. An average run length approximation is derived that can provide insight and guidance in the application of the proposed algorithm. An extensive numerical analysis shows that the proposed method is much more accurate and robust than currently available methods.

## ARTICLE HISTORY

Received 22 October 2014  
Accepted 21 September 2014

## KEYWORDS

steady-state detection;  
multiple change-point  
model; Bayesian inference;  
average run length

## 1. Introduction

The problem of detecting whether a noisy signal or time series data is steady (true process value and noise variance stay unchanged) arises in various application areas, such as process performance assessment and optimization (Ricker and Lee, 1995; Mhamdi *et al.*, 1999; Hoad *et al.* 2009; Marchetti *et al.* 2014), data reconciliation (Bhat and Saraf, 2004; Schladt and Hu, 2007), fault detection and diagnosis (Chen and Howell, 2001; Kim *et al.*, 2008), and process automation and control (Mahuli *et al.*, 1992; Cao and Rhinehart, 1995; Yao *et al.*, 2009; Wu *et al.*, 2013). In these application areas, a steady state of the system is one of the most important requirements to evaluate or model the process or to trigger the next action in the process control. For example, in a discrete-event simulation, there are biases in the initial outputs of the simulation. In order to accurately assess the process performance, the steady state of the outputs needs to be identified (Kelton and Law, 1983; Cash *et al.*, 1992; White, 1997; Spratt, 1998; Robinson, 2002; Hoad *et al.*, 2009). In the chemical industry, process modeling, control, and optimization need to be performed under steady-state operation conditions (e.g., temperature, flow rate, pressure, pH value; Cao and Rhinehart (1995)). Most of the fault detection and diagnosis methodologies on cooling systems are based on the assumption of steady-state operation (Li, 2004; Kim *et al.*, 2008). In process industries (Aguado *et al.*, 2008; Yao *et al.*, 2009), batch operations are not stable in the start-up period due to non-stabilized incoming materials or machine conditions, which cannot guarantee product quality. Therefore, the steady state needs to be identified to avoid expensive quality inspection and scrap costs. The detection of a steady-state condition can also be used as a stopping criterion in iterative numerical methods, such as nonlinear regression, optimization, and neural network training (Natarajan and Rhinehart, 1997). The procedure is stopped when the

objective function (e.g., sum of squared error) reaches a steady state.

During the past several decades, over 40 offline steady-state detection algorithms have been developed, most of which are used to remove initial bias in discrete-event simulations (Franklin, 2009). These methods are classified into five categories by Robinson and Davies (Hoad *et al.*, 2009):

- graphical methods, including time-series inspection (Gordon, 1977), CUSUM plots (Nelson, 1992), etc.;
- heuristic approaches, such as the Marginal Standard Error Rules (White, 1997; White, *et al.*, 2000);
- statistical methods;
- initialization bias tests (e.g., batch means-based tests (Cash *et al.*, 1992));
- hybrid methods.

These methods can be used to identify the steady-state period when the full signal is available; however, they are not suitable for real-time detection; i.e., process control and real-time optimization (Mhamdi *et al.*, 1999; Marchetti *et al.*, 2014), which is based only on the observations up to the current time and requires high timeliness.

To our best knowledge, there are only several online steady-state detection algorithms currently in existence; they can be summarized as follows.

1. A linear regression is performed over a moving data window and the fitted slope is monitored. When the fitted line is sufficiently “flat” or the absolute value of the slope is below a threshold, the process signal is considered steady (Holly *et al.*, 1989; Bethea and Rhinehart, 1991; Wu *et al.*, 2013).
2. A *t*-test is performed to compare the means of two recently adjacent moving data windows with pooled standard deviations. The signal is considered steady if

there is no significant difference (Narasimhan *et al.*, 1987).

3. The standard deviation of a moving data window is monitored, and the steady state is detected if the standard deviation is below a threshold (Kim *et al.*, 2008).
4. An  $F$ -test is performed on the ratio of two variances calculated using two methods, the mean squared deviation and the mean of squared differences of successive data. In the transient state the ratio is very large and when it is below a threshold close to one, the process is considered to be steady (Crow *et al.*, 1960; Cao and Rhinehart, 1995).

Although these methods may perform well on certain types of signals, they are not sufficiently flexible to allow them to be applied to various signals with different characteristics using only one set of detection parameters (e.g., window size, threshold). For example, all of these methods use a moving data window for detection. A long data window may delay the detection of signals that have small noise amplitudes or a large changing rate before the steady state, whereas a short window might result in a high false alarm rate for a noisy signal or a signal with a small changing rate. Therefore, it is highly desirable to develop a more robust and flexible method that can handle various situations.

Recently, Wu *et al.* (2015) developed a more robust steady-state detection algorithm based on multiple change-point models and particle filtering techniques. Compared with the above-mentioned methods, their method is more robust in terms of detection bias and false alarm rate. However, their method has a relatively high computational cost, which may limit its applications in many cases that need rapid detection. Decreasing the number of particles or samples could decrease the computational load; however, it will also decrease the detection stability and accuracy. Therefore, a fast and also accurate method is highly desirable. In this article, we develop a new online steady-state detection algorithm by using piecewise linear functions and exact Bayesian inference. The piecewise linear model is used to approximate the noisy signals and the exact online Bayesian inference is used to calculate the sequential posterior distributions of the latest change-point and model parameters. The exact Bayesian inference is implemented with a recursive message-passing algorithm (Adams and MacKay, 2007; Fearnhead and Liu, 2007) to calculate the posterior distribution of the latest change-point. The steady-state can be detected when the probability of the slope amplitude of the latest linear segment is below a certain threshold.

The rest of this article is organized as follows. Section 2 introduces the modeling of multiple change-points in process signals and the advantages of the proposed detection method for steady-state conditions. In Section 3, the online exact Bayesian inference of the change-points and other model parameters are introduced, a detection probability index is proposed, and various computational issues and approximations are addressed. The approximated formula for the average run length of the proposed method for a steady-state time series is also derived in this section. A numerical illustration, comparisons, and applications are shown in Section 4. Section 5 presents our conclusions and discussions.

## 2. Piecewise linear model for the detection of steady-state conditions

In the proposed method, the target signal is sequentially fitted using a multiple change-point model, or specifically a piecewise linear model. The rationale is that any signal can be closely fitted or approximated by a piecewise linear model with an appropriate number of change-points at appropriate locations. In Fig. 1, we illustrate how a piecewise linear model can be used to approximate nonlinear signals, such as exponential and oscillating sinusoidal signals. In the detection of a steady-state condition, we seek to use only the observations in the latest line segment to test whether the signal is steady. When the latest line segment meets the steady-state criterion (see Section 3.3), we claim that the signal is steady.

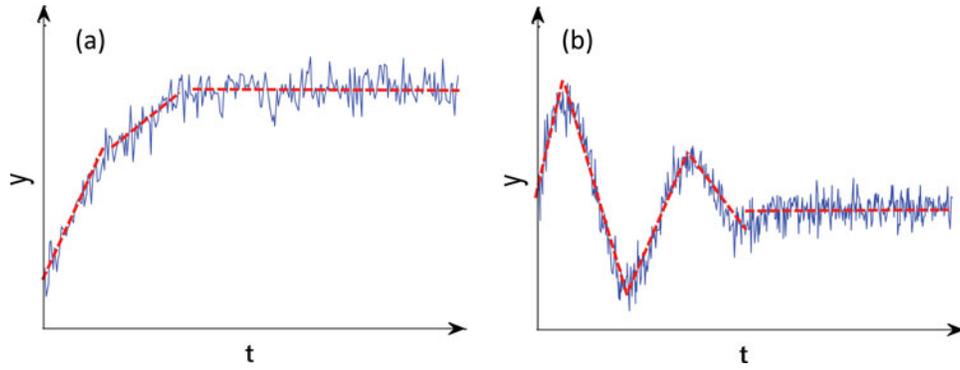
The proposed method has one key advantage over moving window-based methods. For the moving window-based methods, the window may contain both transient state and steady-state observations or may contain oscillating signals with unchanging means, which may influence the effectiveness of the testing procedure. For example, for oscillating signals, the slope detection method or  $t$ -test method may totally fail, due to the linear regression slope or the mean of the signals in the moving window being temporarily zero. The proposed method finds the location of the latest change-point and then uses the observations in the latest line segment to test the level of stability. Therefore, it is expected to be more robust than the moving window-based methods.

In this method, the main challenge is how to sequentially and efficiently estimate the latest change-point and the parameters for the latest line segment (e.g., slope, intercept, noise variance). We propose to use the Bayesian inference approach, as illustrated in Fig. 2, where the posterior distributions of the Latest Change-Point (LCP) and model parameters for the latest line segment are sequentially updated; i.e., re-estimated when a new data point is obtained. For example, at time  $t_1$ , the posterior of an LCP is concentrated around the starting time  $t = 1$  and conditioning on this LCP, the posterior distribution of the other model parameters (e.g., slope, noise variance) is estimated using all observations between the LCP and  $t_1$ . At time  $t_2$ , with the emergence of the new linear segment starting at  $t_1$ , the center of the posterior of the LCP jumps to the location around  $t_1$  and the posterior is almost zero at the locations far before  $t_1$ . Therefore, the observations between  $t_1$  and  $t_2$  are mainly used for the estimation of other parameters of the model. In the following sections, the online inference of the LCP and model parameters and steady-state detection algorithm will be introduced in detail.

## 3. Online Bayesian inference and detection of steady-state conditions

### 3.1 Bayesian formulation of the multiple change-point model

Since the detection of steady-state conditions is mainly based on the observations in the latest line segment, it is critically important to obtain the posterior distribution of the LCP. To facilitate



**Figure 1.** Illustration of approximating the nonlinear signals using the piecewise linear model: (a) an exponential signal in the transient state and (b) a sinusoidal oscillating signal in the transient state.

online updating, we first formulate it into a multiple-change-point model in the Bayesian framework. Suppose the given time series with  $n$  observations is  $\{y_t, t = 1, 2, \dots, n\}$ , with  $y_t$  being the observation at time step  $t$ . The corresponding line segment parameter at time  $t$  is defined as  $\xi_t = (a_t, b_t, \sigma_t^2)$  where  $a_t$  is the slope of the line segment,  $b_t$  is the intercept, and  $\sigma_t^2$  is the variance of the signal noise. The multiple-change-point model can be expressed as

$$\xi_t = \begin{cases} \theta_1 & \text{if } 1 \leq t < C_1 \\ \theta_2 & \text{if } C_1 \leq t < C_2 \\ \vdots & \vdots \\ \theta_m & \text{if } C_{m-1} \leq t < C_m \\ \theta_{m+1} & \text{if } C_m \leq t \leq n \end{cases}, \quad (1)$$

where  $\theta_i \in \mathbb{R}^2 \times \mathbb{R}^+$ ,  $m$  and  $\{C_i, i = 1, 2, \dots, m\}$  are the number and locations of the change-points. The parameters are assumed to be independent across different segments. The observation  $y_t$  is modeled as

$$y_t = a_t t + b_t + \varepsilon_t, \quad (2)$$

where  $\varepsilon_t$  is the independent and identically distributed (i.i.d.) Gaussian noise and  $\varepsilon_t \sim N(0, \sigma_t^2)$ .

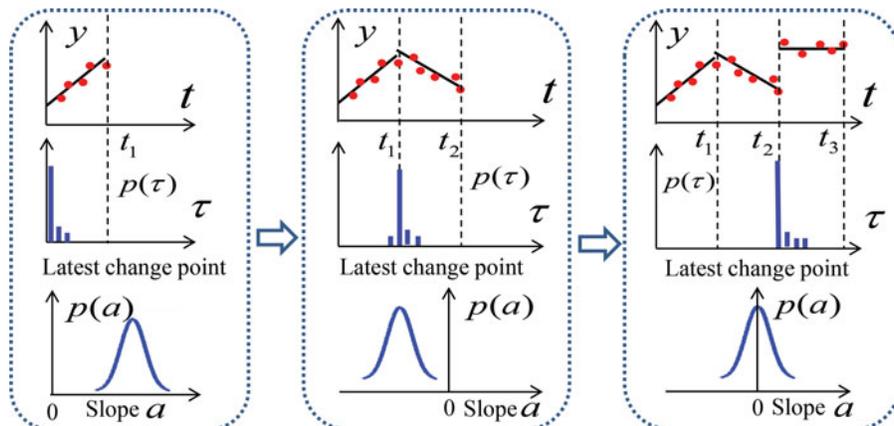
In the Bayesian formulation, the priors for the change-points (number  $m$  and locations  $\{C_i, i = 1, \dots, m\}$ ) and line segments (slope, intercept, noise variance) need to be specified. For

the change-points, it is natural to directly select a joint prior distribution  $P(m, C_1, \dots, C_m)$ . Alternatively, we could specify a prior for the duration of the line segment, which is equivalent to placing a certain joint prior on the number and locations of the change-points (Fearnhead and Liu, 2007). The second approach has an advantage in formulating the model into a Markov transition process, which is commonly used to facilitate online Bayesian inference of multiple-change-point models (Chopin, 2007; Fearnhead and Liu, 2007). Let  $\tau_t$  denote the LCP at time  $t$ . Suppose the prior probability mass function and cumulative distribution function for the length of the line segments is  $g(\cdot)$  and  $G(\cdot)$ , respectively. Then the prior transition probability of the change-point can be expressed as

$$P(\tau_t = j | \tau_{t-1} = i) = \begin{cases} \frac{1 - G(t-i)}{1 - G(t-i-1)} & \text{if } j = i \\ 1 - \frac{1 - G(t-i)}{1 - G(t-i-1)} & \text{if } j = t \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

where  $i \leq t - 1$ .

The most popular and simplest prior for the change-point is a geometric prior applied on the segmental duration, which corresponds to a Markov transition process with a constant prior transition probability  $p$ . Other common priors include the Poisson distribution, gamma distribution, etc., which are often



**Figure 2.** Illustration of the linear model and Bayesian inference at times  $t_1$ ,  $t_2$ , and  $t_3$ ; top: the observations and fitted line; middle: the posterior distribution of the LCP; bottom: the posterior distribution of the latest slope.

used in studies on speech segmentation (Wang, 1997). For simplicity, we select the geometric distribution for the segmental duration; i.e., we assume  $P(\tau_t = t | \tau_{t-1} = i) \equiv p$ ,  $i \leq t - 1$  in this article. For the model parameter  $\xi$ , a conjugate prior is used, which makes the posterior of both the LCP  $\tau$  and model parameter  $\xi$  analytically tractable, as shown in the following subsection.

### 3.2 Exact calculation of the posterior distribution of the LCP

Let  $y_{s:t} = (y_s, y_{s+1}, \dots, y_t)^T$ . For  $j = 1, 2, \dots, t + 1$ , the posterior for the LCP at time step  $t + 1$  can be calculated recursively as

$$P(\tau_{t+1} = j | y_{1:t+1}) \propto P(\tau_{t+1} = j, y_{t+1} | y_{1:t}) \\ = P(y_{t+1} | \tau_{t+1} = j, y_{j:t}) P(\tau_{t+1} = j | y_{1:t}), \quad (4)$$

$$P(\tau_{t+1} = j | y_{1:t}) = \sum_{i=1}^{\min(j,t)} P(\tau_{t+1} = j | \tau_t = i) \\ \times P(\tau_t = i | y_{1:t}), \quad (5)$$

$$P(y_{t+1} | \tau_{t+1} = j, y_{j:t}) = \begin{cases} \frac{P(y_{j:t+1} | \tau_{t+1} = j)}{P(y_{j:t} | \tau_{t+1} = j)} & \text{if } j \leq t \\ P(y_{t+1}) & \text{if } j = t + 1 \end{cases}. \quad (6)$$

Note that the equality in Equation (4) is due to the fact that for  $j \leq t$ ,  $(y_{t+1} | \tau_{t+1} = j)$  is independent of  $y_{1:j-1}$  and so the distribution of  $(y_{t+1} | \tau_{t+1} = j, y_{1:t})$  is identical to the distribution of  $(y_{t+1} | \tau_{t+1} = j, y_{j:t})$ . Define  $P(s, t) = P(y_{s:t} | y_{s:t})$  is in the same linear segment) and substitute Equations (5) and (6) into Equation (4), then we obtain

$$P(\tau_{t+1} = j | y_{1:t+1}) \propto \begin{cases} \frac{P(j, t+1)}{P(j, t)} P(\tau_{t+1} = j | \tau_t = j) P(\tau_t = j | y_{1:t}) & \text{if } j \leq t \\ P(t+1, t+1) \sum_{i=1}^t P(\tau_{t+1} = j | \tau_t = i) P(\tau_t = i | y_{1:t}) & \text{if } j = t + 1 \end{cases}. \quad (7)$$

From Equation (7) we can see that the distribution of the LCP at time  $t + 1$  can be recursively calculated based on the previously calculated distribution at time  $t$ . The terms  $P(j, t+1)/P(j, t)$  and  $P(t+1, t+1)$  are the only terms that incorporate information on the newest observation  $y_{t+1}$  into the updating of the posterior distribution of the LCP. Therefore, they play a decisive role in the detection of the change-point. For example, if the latest true change-point is at time step  $t + 1$  with  $y_{t+1}$  significantly different from  $y_{1:t}$ , then  $P(j, t+1)/P(j, t)$  is expected to be much smaller than  $P(t+1, t+1)$  and thus more weight is put on time step  $t + 1$  as the LCP. The calculation of  $P(s, t)$  is the key part in the above recursion equation. By using conjugate priors for model parameters, it can be calculated analytically and involves no complex numerical integrations. Let  $\beta$  be the parameter of the linear segment thus,  $\beta = (a, b)^T$ . The joint prior distribution for  $\beta$  and noise variance  $\sigma^2$  can be assigned using Gaussian and inverse gamma distribution as

$$\sigma^2 \sim IG\left(\frac{\nu}{2}, \frac{\gamma}{2}\right), \quad \beta | \sigma^2 \sim N(\beta_0, \sigma^2 \Sigma), \quad (8)$$

where  $\gamma$ ,  $\beta_0$ , and  $\Sigma$  are known parameters. Define

$$X_{st} = \begin{bmatrix} s & 1 \\ s+1 & 1 \\ \vdots & \vdots \\ t & 1 \end{bmatrix}.$$

The term  $P(s, t)$  in Equation (7) can then be calculated by integrating out  $\beta$  and  $\sigma^2$ :

$$P(s, t) = \pi^{-(t-s+1)/2} \left( \frac{|M_{st}|}{|\Sigma|} \right)^{\frac{1}{2}} \\ \times \frac{\gamma^{\nu/2}}{(H_{st})^{(t-s+1+\nu)/2}} \frac{\Gamma((t-s+1+\nu)/2)}{\Gamma(\nu/2)}, \quad (9)$$

where

$$M_{st} = (X_{st}^T X_{st} + \Sigma^{-1})^{-1}, \quad N_{st} = (\Sigma^{-1} \beta_0 + X_{st}^T y_{s:t}),$$

$$H_{st} = y_{s:t}^T y_{s:t} + \gamma + \beta_0^T \Sigma^{-1} \beta_0 - N_{st}^T M_{st} N_{st}. \quad (10)$$

The derivation of Equation (9) is shown in Appendix A.

### 3.3 Detection of steady-state conditions

In this article, the term ‘‘steady state’’ refers to the condition where the true signal value and noise variance stay unchanged. Let  $\mu(t)$  and  $\sigma^2(t)$  denote the true signal value and noise variance, respectively. Mathematically, the signal  $y(t)$  is steady in the time period  $[t_1, t_2]$  if  $\mu(t) \equiv c_1$  and  $\sigma^2(t) \equiv c_2$  for  $t \in [t_1, t_2]$ , where  $c_1$  and  $c_2$  are constants, and  $t_1 < t_2$ . It is natural to use the ‘‘flatness’’ of the current line segment to decide whether the signal is steady. Specifically, when the amplitude of the slope  $|a_t|$  is sufficiently small, the signal can be claimed to be steady. Note that here we do not need to consider the noise variance in the evaluation of extent of stability, as the noise of the current line segment is statistically steady (based on the definition of the multiple-change-point model, the model parameters are steady between two adjacent change-points).

Naturally, we can use the estimated slope amplitude of the latest linear segment  $|\hat{a}_t|$  as the detection index. In this article, we use a more flexible and stable detection index  $P_t$ , which is defined as the probability of  $|a_t|$  being less than a slope threshold  $s_0$  given observations  $y_{1:t}$ :

$$P_t = \Pr(|a_t| < s_0 | y_{1:t}) = \sum_{i=1}^t P(|a_t| < s_0 | \tau_t = i, y_{i:t}) \\ \times P(\tau_t = i | y_{1:t}). \quad (11)$$

Once  $P_t$  is greater than or equal to a probability index threshold  $\alpha$ , the signal is claimed to be steady. In the above equation, the probability of the LCP  $P(\tau_t = i | y_{1:t})$  can be calculated using Equation (7) and  $P(|a_t| < s_0 | \tau_t = i, y_{i:t})$  can be calculated based on Lemma 1 as follows (see Appendix B for proof).

**Lemma 1.** Suppose  $\sigma^2 \sim IG(v/2, \gamma/2)$  and  $\beta|\sigma^2 \sim N(\beta_0, \sigma^2 \Sigma)$ . Define:

$$\mathbf{X}_{st} = \begin{bmatrix} s & 1 \\ s+1 & 1 \\ \vdots & \vdots \\ t & 1 \end{bmatrix},$$

then

1.  $(\beta_t|\tau_t = s, y_{s:t})$  follows a bivariate  $t$  distribution with degrees of freedom  $d_{st} = t - s + v + 1$ , mean  $\mu_{st} = \mathbf{M}_{st} \mathbf{N}_{st}$  and covariance matrix  $(d_{st}/(d_{st} - 2)) \Sigma_{st}$  ( $d_{st} > 2$ ) where  $\Sigma_{st} = H_{st} \mathbf{M}_{st} / d_{st}$ ,  $\mathbf{M}_{st}$ ,  $\mathbf{N}_{st}$  and  $H_{st}$  are defined in Equation (10). Denote it as  $(\beta_t|\tau_t = s, y_{s:t}) \sim t_2(d_{st}, \mu_{st}, \Sigma_{st})$ .
2. Suppose  $\Sigma_{st} = \mathbf{K}_{st} \mathbf{R}_{st} \mathbf{K}_{st}^T$  where  $\mathbf{R}_{st}$  is the correlation matrix and  $\mathbf{K}_{st}$  is a diagonal matrix with positive diagonal entries  $k_{st}^{(i,i)}$ , then

$$(\mathbf{K}_{st}^{-1} \beta_t | \tau_t = s, y_{s:t}) \sim t_2(d_{st}, \mathbf{K}_{st}^{-1} \mu_{st}, \mathbf{R}_{st}),$$

and

$$\left( \frac{a_t}{k_{st}^{(1,1)}} | \tau_t = s, y_{s:t} \right) \sim t_1(d_{st}, (\mathbf{K}_{st}^{-1} \mu_{st})^{(1)}, 1),$$

or  $((a_t - \mu_{st}^{(1)})/k_{st}^{(1,1)} | \tau_t = s, y_{s:t})$  follows the standard  $t$  distribution with degrees of freedom  $d_{st}$ . Here  $(\mathbf{K}_{st}^{-1} \mu_{st})^{(i)}$  and  $\mu_{st}^{(i)}$  denote the  $i$ th element of the vector  $\mathbf{K}_{st}^{-1} \mu_{st}$  and  $\mu_{st}$ , respectively.

Based on Lemma 1, the detection index  $P_t$  can be calculated as

$$P_t = \sum_{i=1}^t \left[ \Psi_{d_{it}} \left( \frac{s_0 - \mu_{it}^{(1)}}{k_{it}^{(1,1)}} \right) - \Psi_{d_{it}} \left( \frac{-s_0 - \mu_{it}^{(1)}}{k_{it}^{(1,1)}} \right) \right] \times P(\tau_t = i | y_{1:t}), \quad (12)$$

where  $\Psi_{d_{it}}(\cdot)$  is the cumulative distribution function (CDF) of a standard  $t$  distribution with degrees of freedom  $d_{it}$ . The filtered observation  $\hat{y}_t$  (used in Section 4 for the purpose of illustration) at time  $t$  can be calculated as

$$\hat{y} = \sum_{i=1}^t (\mu_{it}^{(1)} t + \mu_{it}^{(2)}) P(\tau_t = i | y_{1:t}). \quad (13)$$

In the development of a decision rule, we need to specify or tune the thresholds for both the slope amplitude and the probability detection index. However, in the application, the probability index often increases rapidly at a certain time after the signals enter into a steady-state condition (see Section 3.5 for a detailed analytical explanation). In addition, the numerical results in Section 4.2 show that although the average run length of a steady-state time series is sensitive to the slope threshold  $s_0$ , it is not sensitive to the probability index threshold  $\alpha$ . Therefore, to simplify the algorithm, we set  $\alpha$  to be 0.9 and only tune the slope threshold in application.

In Bayesian inference, informative priors are often preferable if prior knowledge or historical data are available. However, when we know very little about the data and we just want the data to “speak” for themselves, the non-informative priors would then be a better choice. In the detection of

steady-state conditions, we often face nonlinear signals that may need multiple linear segments with both increasing and decreasing trends for their approximation. Also, the amplitude of the slopes and line durations may significantly vary between different segments. In such cases a non-informative prior for  $\beta$  is recommended. We can assign flat priors for  $\beta$  with zero mean ( $\beta_0 = 0$ ) and large variance; i.e., large value for the diagonal entries of  $\Sigma$  and to describe the uncertainty of the slope and intercept. For the signal noise, typically we can select an informative prior. The noise amplitude for each signal in most cases is fixed and we can roughly estimate it based on the historical data or prior knowledge. For the prior transition probability  $p$ , we found that it has very little influence on the detection results in the range [0.05, 0.5], which will be shown in Section 4.2. Therefore, we can arbitrarily select a value from [0.05, 0.5] in application. The steady-state detection process is summarized in Algorithm 1.

---

**Algorithm 1. Steady-state detection algorithm using exact calculation**

1. Specify  $v, \gamma, \beta_0, \Sigma, p$ , and  $s_0$ .
  2. Set  $P_1 = 0$  and  $P(\tau_1 = 1 | y_1) = 1$ .
  3. For  $t = 2, 3, \dots, n$ .
    - For  $i = 1, 2, \dots, t$ 
      - Calculate the un-normalized  $P(\tau_t = i | y_{1:t})$  based on Equation (7).
      - End
    - Calculate the normalized  $P(\tau_t | y_{1:t})$ :
- $$P(\tau_t = i | y_{1:t}) = \frac{P(\tau_t = i | y_{1:t})}{\sum_{j=1}^t P(\tau_t = j | y_{1:t})}.$$
- Calculate the probability index  $P_t$  based on Equation (12).
  - If  $P_t > 0.9$ , the signal is steady and stop.
- End
- 

After the posterior of the LCPs  $P(\tau_t | y_{1:t})$  is calculated for all time steps  $t = 1, 2, \dots, n$ , we can easily backwards reconstruct the trajectories of all change-points for the purpose of illustration using Algorithm 2.

---

**Algorithm 2 Simulation of change-point trajectories in  $N$  realizations**

1. Count[ $i$ ] = 0 for  $i = 1, 2, \dots, n$ .
  2. For  $rep = 1 : N$ 
    - Simulate  $t_1$  from  $P(\tau_n | y_{1:n})$ . Set Count[ $t_1$ ] = Count[ $t_1$ ] + 1 and  $k = 1$ .
    - While  $t_k > 1$ 
      - Simulate  $t_{k+1}$  from the support  $\{1, 2, \dots, t_k - 1\}$  with the discrete probability proportional to  $P(\tau_{t_k-1} | y_{1:t_k-1}) P(\tau_{t_k} = t_k | \tau_{t_k-1})$ . Set  $k = k + 1$  and Count[ $t_k$ ] = Count[ $t_k$ ] + 1.
  - End
  - End
  3. Calculate the frequency  $f_i = \text{Count}[i]/N$  for  $i = 1, 2, \dots, n$ .
-

### 3.4 Computational issues and approximations

Algorithm 1 uses the exact calculation for the posterior distribution and is expected to be very accurate. However, there may be various computational issues that may limit its applications. The most important issue is that the computational cost and memory cost in the recursive calculation of  $P(\tau_t|y_{1:t})$  significantly increase with time  $t$ . For example, at time  $t$ , we have to calculate  $t$  discrete probabilities. The calculation of each probability  $P(\tau_t = t|y_{1:t})$  also rapidly increases with  $j$  in the recursion. A natural way to reduce the computational cost is to approximate the calculated  $P(\tau_t|y_{1:t})$  using another probability mass function with fixed size of support  $m < t$ . In practice when  $t$  is large,  $P(\tau_t|y_{1:t})$  is almost equal to zero in many locations. Setting  $P(\tau_t|y_{1:t})$  to be zero at these locations can reduce the computational cost in the calculation of  $P(\tau_{t+1}|y_{1:t+1})$  and  $P_t$ .

In this article, we use the following strategy: at each time step  $t$  ( $t > m$ ), we only calculate the probability  $P(\tau_t|y_{1:t})$  at  $m$  certain locations that are very likely to be the LCP. The specific steps are as follows.

1. Select  $m - 1$  locations from  $\{1, 2, \dots, t - 1\}$  using weighted sampling without replacement (Wong and Easton, 1980). The weight for location  $i$  is  $P(\tau_{t-1} = i|y_{1:t-1})$ .
2. Normalize the weights of the selected locations.
3. Calculate  $P(\tau_t|y_{1:t})$  at  $t$  and these  $m - 1$  selected locations.
4. Set  $P(\tau_t|y_{1:t}) = 0$  at other locations.

In other words, this method is to select location  $\tau_t = t$  and other  $m - 1$  locations from the support of  $P(\tau_{t-1}|y_{1:t-1})$  of size  $m$  to be the support for  $P(\tau_t|y_{1:t})$ . In this strategy, the computation is almost balanced at large time step  $t$ .

Another computational issue is the calculation of  $P(j, t + 1)/P(j, t)$ , which can be expressed as

$$\frac{P(j, t + 1)}{P(j, t)} = \frac{1}{\sqrt{\pi}} \left( \frac{|M_{j,t+1}|}{|M_{j,t}|} \right)^{\frac{1}{2}} \left( \frac{H_{jt}}{H_{j,t+1}} \right)^{\frac{t-j+v-1}{2}} \times \frac{1}{\sqrt{H_{j,t+1}}} \frac{\Gamma((t-j+v+2)/2)}{\Gamma((t-j+v+1)/2)}. \quad (14)$$

In the above equation, the calculation of  $\Gamma((t-j+v+1)/2)$  may be a problem as  $t-j$  increases to a very large value. For example, in MATLAB,  $\Gamma(172)$  becomes infinite due to the precision issue. One way to solve this issue is to compute the difference of the natural logarithm of the gamma function  $\Gamma((t-j+v+2)/2)$  and  $\Gamma((t-j+v+1)/2)$  and then calculate the exponential function of this difference. Another method that is more preferable in terms of the computational cost is to use Stirling's series to approximate the ratio of the gamma function (Tricomi and Erdélyi, 1951):

$$\frac{\Gamma(z+z_1)}{\Gamma(z+z_2)} = z^{z_1-z_2} \left[ 1 + \frac{(z_1-z_2)(z_1+z_2-1)}{2z} + O(|z|^{-2}) \right].$$

The gamma ratio in Equation (14) can thus be approximated as

$$\frac{\Gamma((t-j+v+2)/2)}{\Gamma((t-j+v+1)/2)} \approx \sqrt{\frac{t-j+v+1}{2}}$$

$$\times \left[ 1 - \frac{1}{4(t-j+v+1)} \right]. \quad (15)$$

This approximation has high accuracy and can be quickly calculated.

The calculation of  $P_t$  in Equation (12) involves many CDFs  $\Psi_d(\cdot)$  of the  $t$  distribution, which can also be approximated to reduce the computational cost. The first method is to use a normal approximation. It is well known that the Student's  $t$  distribution can be closely approximated by a normal distribution with the same mean and variance when  $d \geq 30$  (Li and Moor, 1998). Therefore, for  $d \geq 30$ ,  $\Psi_d(x) \approx \Phi(x/\sqrt{d/(d-2)})$  where  $\Phi(\cdot)$  is the CDF of the standard normal distribution.

### 3.5 Approximation of the average run length for a steady-state time series

The Average Run Length (ARL) is an important performance criterion used to evaluate a detection scheme, and it is commonly used in statistical process control charts. Here ARL is defined as the average number of observations required for the algorithm to claim a steady-state on multiple steady-state time series (no initial transient state) with the same mean and noise variance. It can also provide insight and guidance on understanding and tuning the algorithm in applications. In this subsection, we develop an approximation of the ARL for a steady-state time series as follows. Suppose the detection probability index is approximated using the normal CDF as

$$P_t \approx \sum_{i=1}^t \left[ \Phi \left( \frac{s_0 - \mu_{it}^{(1)}}{k_{it}^{(1,1)} \sqrt{d_{it}/(d_{it}-2)}} \right) - \Phi \left( \frac{-s_0 - \mu_{it}^{(1)}}{k_{it}^{(1,1)} \sqrt{d_{it}/(d_{it}-2)}} \right) \right] P(\tau_t = i|y_{1:t}).$$

Assume  $y_1, y_2, \dots, y_n$  is a steady-state time series and  $y_i \sim \text{i.i.d. } N(0, \sigma^2)$ . In the detection process, it is observed that the posterior  $P(\tau_t|y_{1:t})$  is almost focused on  $t = 1$  or  $P(\tau_t = 1|y_{1:t}) \approx 1$  for linear signals (see Section 4.1 for details). Therefore,

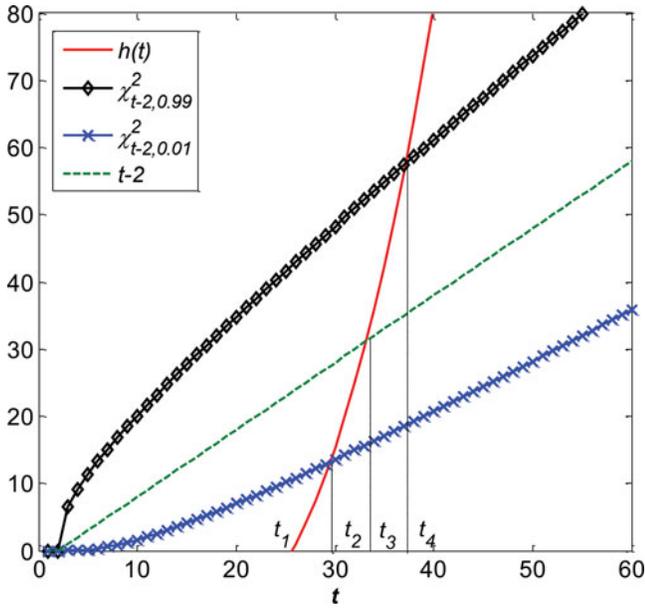
$$P_t \approx \left[ \Phi \left( \frac{s_0 - \mu_{1t}^{(1)}}{k_{1t}^{(1,1)} \sqrt{d_{1t}/(d_{1t}-2)}} \right) - \Phi \left( \frac{-s_0 - \mu_{1t}^{(1)}}{k_{1t}^{(1,1)} \sqrt{d_{1t}/(d_{1t}-2)}} \right) \right].$$

When  $\Sigma^{(i,i)} \rightarrow \infty$  and  $\beta_0 = 0$  (flat prior for  $\beta$ ),  $\mu_{1t} \approx (\mathbf{X}_{1t}^T \mathbf{X}_{1t})^{-1} \mathbf{X}_{1t} y_{1:t}$ , which is the ordinary least square estimator for  $\beta_t$  and therefore  $\mu_{1t} \approx 0$ . Also,  $\Sigma_{1t} \approx [y_{1:t}^T (\mathbf{I} - \mathbf{P}) y_{1:t} + \gamma] (\mathbf{X}_{1t}^T \mathbf{X}_{1t})^{-1} / d_{1t}$  where  $\mathbf{P}$  is the projection matrix  $\mathbf{P} = \mathbf{X}_{1t} (\mathbf{X}_{1t}^T \mathbf{X}_{1t})^{-1} \mathbf{X}_{1t}^T$ . Therefore,

$$k_{1t}^{(1,1)} \approx \sqrt{\frac{12}{[y_{1:t}^T (\mathbf{I} - \mathbf{P}) y_{1:t} + \gamma] t (t^2 - 1) d_{1t}}},$$

and

$$P_t \approx 2\Phi \left( s_0 \sqrt{\frac{t(t^2-1)(t-2+v)}{12[y_{1:t}^T (\mathbf{I} - \mathbf{P}) y_{1:t} + \gamma]}} \right) - 1.$$



**Figure 3.** Illustration of the change of  $h(t)$ ,  $\chi_{t-2,0.99}^2$ ,  $\chi_{t-2,0.01}^2$ , and  $E(\chi_{t-2}^2)$  with  $t$ ; the parameters were set to  $s_0 = 0.003$ ,  $\sigma = 0.1$ ,  $v = 20$ ,  $\gamma = 0.2$ , and  $\alpha = 0.9$ .

Suppose the detection is stopped when  $P_t \geq \alpha$ ; that is,

$$\frac{y_{1:t}^T (\mathbf{I} - \mathbf{P}) y_{1:t}}{\sigma^2} \leq \frac{s_0^2 t (t - 2 + v) (t^2 - 1)}{12z_{(1+\alpha)/2}^2 \sigma^2} - \frac{\gamma}{\sigma^2}, \quad (16)$$

where  $z_{(1+\alpha)/2}$  is the quantile for a standard normal distribution with  $\Phi(z_{(1+\alpha)/2}) = (1 + \alpha)/2$ . Denote:

$$Y_t = \frac{y_{1:t}^T (\mathbf{I} - \mathbf{P}) y_{1:t}}{\sigma^2} \text{ and } h(t) = \frac{s_0^2 t (t - 2 + v) (t^2 - 1)}{12z_{(1+\alpha)/2}^2 \sigma^2} - \frac{\gamma}{\sigma^2}.$$

It is well known that  $Y_t \sim \chi_{t-2}^2$ . Based on Equation (16), it is almost impossible to obtain the exact analytical form for the ARL since it involves multiple integrals and conditional distributions. For example, the ARL can be expressed as

$$ARL = \sum_{t=1}^{\infty} t \Pr(Y_t \leq h(t) | Y_i) h(i) \text{ for } i < t,$$

where the analytical form of the conditional probability  $\Pr(Y_t \leq h(t) | Y_i) h(i)$  for  $i < t$  is hard to derive. Here we use an approximation method as follows.

Suppose we select  $s_0 = 0.003$ ,  $\sigma = 0.1$ ,  $v = 20$ ,  $\gamma = 0.2$ , and  $\alpha = 0.9$ . **Figure 3** shows the function  $h(t)$ , the 0.99 and 0.01 quantiles of the  $\chi_{t-2}^2$  distribution, and the mean of  $\chi_{t-2}^2$ . From it we can see that  $h(t) \leq 0$  for  $t \leq t_1$  and then it increases much more rapidly (polynomial of fourth order) than other three curves. For  $t \leq t_1$ ,  $Y_t$  is always larger than  $h(t)$  and a steady-state will not be claimed. At  $t = t_2$ ,  $P(Y_t \leq h(t)) = 0.01$  and in the interval  $[t_1, t_2]$ ,  $P(Y_t \leq h(t)) \ll 0.01$ , due to a rapid decrease in the probability density function of the  $\chi_{t-2}^2$ . Similarly, at  $t = t_4$ ,  $P(Y_t \leq h(t)) = 0.99$  and in the interval  $(t_4, \infty)$ ,  $P(Y_t \leq h(t)) \gg 0.99$ . Therefore, it is highly probable that the stopping time will be in the time interval  $[t_2, t_4]$ . Since the width of the interval is small, we use  $t_4$  as the ARL:

$$ARL \approx \arg \min_t \{t | \chi_{t-2,0.99}^2 \leq h(t)\}. \quad (17)$$

We found that this approximation is very close to the simulated ARL under different values of  $s_0$ ,  $\sigma$ ,  $v$ ,  $\gamma$ , and  $\alpha$ , which will be shown in Section 4.2.

Let the Run Length (RL) denote the number of observations required to claim a steady-state for a single run. From the above analysis we see that  $t_1 + 1$  is approximately the lower bound of the distribution for RL. Also,  $t_1$  only depends on the algorithm parameters; i.e., hyper-parameters of the prior distributions and detection thresholds. It is independent of signal noise. However, the distribution of RL is highly dependent on both the signal noise and all other algorithm-related parameters, as seen from Equation (16). For example, the higher the signal noise, the lower the rate of increase of  $h(t)$  and thus the less probable it is that  $Y(t) \leq h(t)$  for a certain  $t > t_1$ . In other words, we need more steady-state observations to claim a steady-state when the signal is noisier. On the other hand, the higher the slope threshold, the larger the rate of increase of  $h(t)$  and thus fewer observations are needed to claim a steady state.

## 4. Case studies for illustration, comparison, and application

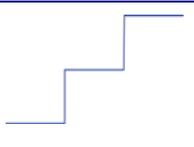
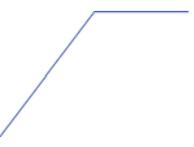
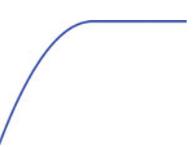
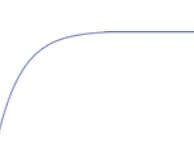
### 4.1 Illustration

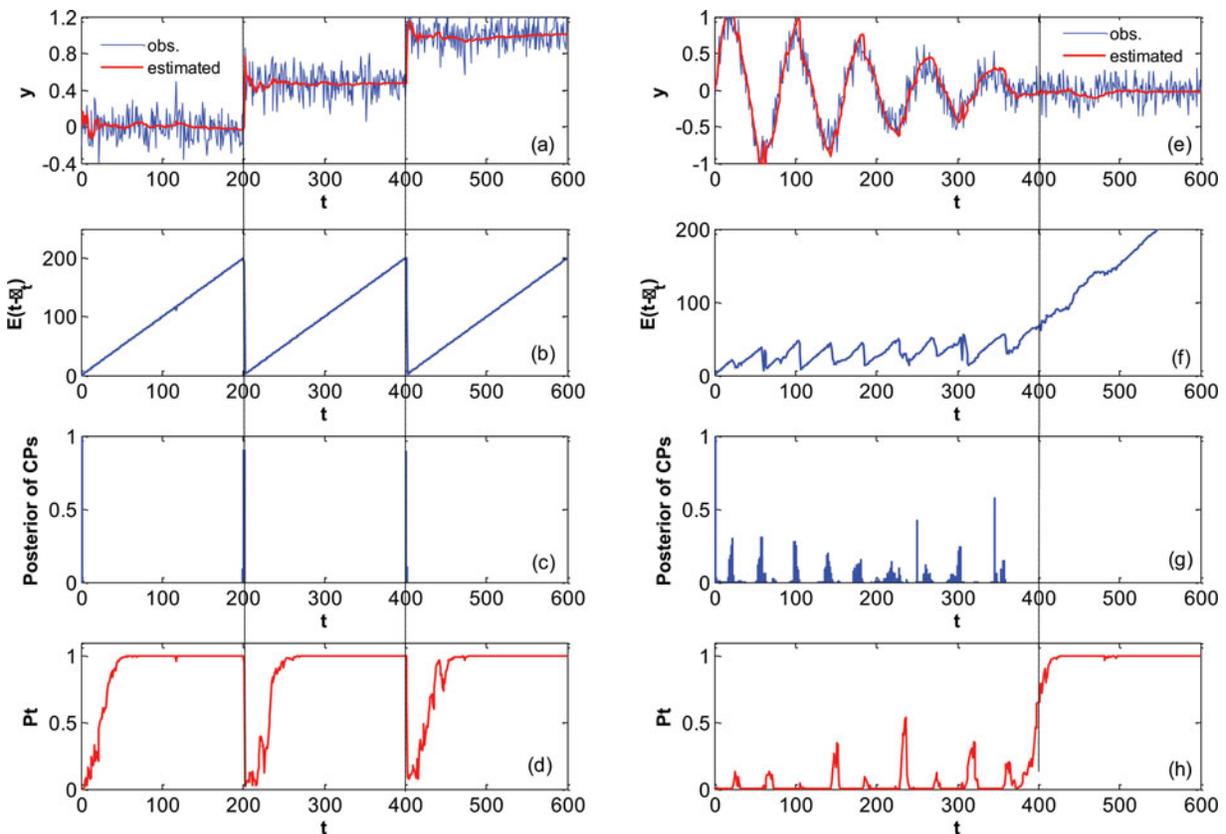
Simulated signals are used to illustrate the detection process of the steady-state algorithm and compare it with other existing algorithms. They were generated using bias functions and noise, where the bias functions consisted of an initial transient state and a steady-state. A total of five bias functions are used in this article: step function, linear function, quadratic function, exponential function, and oscillating function, as shown in **Table 1**. Note that the step function is not strictly a bias function, as it contains a piecewise steady-state period. The final four bias functions are often used to test the offline heuristic truncation algorithms used in discrete-event simulations (Cash *et al.*, 1992; Spratt, 1998; White *et al.*, 2000; Hoard *et al.*, 2009). For simplicity, only the negative bias scenarios (i.e., increasing before steady-state) are considered for the linear, quadratic, and exponential functions.

The step function and oscillating functions are used to illustrate the detection process for signals with fixed noise amplitudes. For the step function,  $h_1 = 0$ ,  $h_2 = 0.5$ ,  $h_3 = 1$ ,  $T_1 = 200$ , and  $T_2 = 400$ . For the oscillating function,  $h = 1$ ,  $T_0 = 400$ , and  $f = 30$  (a total of 10 peaks and troughs). For both signals, the number of observations  $n = 600$  and noise  $\sigma_y = 0.14$ . The priors for the steady-state detection algorithm were set to be  $\beta_0 = 0$ ,  $\Sigma = 1 \times 10^4 \mathbf{I}$ ,  $v = 20$ ,  $\gamma = 0.2$ , and  $p = 0.2$ , where  $\mathbf{I}$  is the  $2 \times 2$  identity matrix. The size of the support  $m$  for the posterior  $P(\tau_t | y_{1:t})$  was set to 50 and the slope threshold  $s_0$  was set to 0.003.

**Figure 4** shows the steady-state detection process for the step signal and oscillating signal. **Figures 4(a) and 4(e)** show the observations and estimated signals, where the estimated values are very close to the true values. In addition, the estimations become increasingly smooth as the length of the linear segments grows. **Figures 4(b) and 4(f)** show the sequentially estimated durations of the latest linear segments, and **Figs. 4(c) and 4(g)** show the simulated frequencies of all change-points using Algorithm 2. They are used to capture the jump of the center of

**Table 1.** Five bias functions and their shapes

Signal	Function	Shape
Step	$y(t) = \begin{cases} h_1, & t = 1, \dots, T_1 \\ h_2, & t = T_1 + 1, \dots, T_2 \\ h_3, & t = T_2, \dots, n \end{cases}$	
Linear	$y(t) = \begin{cases} \frac{t}{T_0} h, & t = 1, \dots, T_0 \\ h, & t = T_0 + 1, \dots, n \end{cases}$	
Quadratic	$y(t) = \begin{cases} h \left[ 1 - \frac{(t-T_0)^2}{(T_0-1)^2} \right], & t = 1, \dots, T_0 \\ h, & t = T_0 + 1, \dots, n \end{cases}$	
Exponential	$y(t) = \begin{cases} h \left[ 1 - 10^{\frac{1-t}{T_0-1}} \right], & t = 1, \dots, T_0 \\ y(T_0), & t = T_0 + 1, \dots, n \end{cases}$	
Oscillating	$y(t) = \begin{cases} h \frac{T_0-t}{T_0-1} \sin\left(\frac{\pi t}{T_0}\right), & t = 1, \dots, T_0 \\ 0, & t = T_0 + 1, \dots, n \end{cases}$	



**Figure 4.** Illustration of the steady-state detection using step function (a-d) and oscillating function (e-f): (a) and (e): simulated  $y$  and estimated values  $\hat{y}$  using Equation (13). The dotted vertical lines indicate the starting point of the steady-state; (b) and (f): the estimated duration of the latest linear segment; (c) and (g): simulated posterior (frequencies) of change-points using Algorithm 2; (d) and (h): the probability index  $P_t$ .

the posterior  $P(\tau_t | y_{1:t})$ . For the step signal, there are immediate jumps at the mean shift locations. The reason for this behavior is that  $P(t, t)$  in Equation (7) is significantly larger than  $P(j, t)/P(j, t - 1)$  for a mean shift or jump change at time  $t$ , and thus the change can be immediately detected. Also, the posterior of the LCP is focused almost at the starting time of the linear segment, which is the assumption in the approximation of the ARL, and therefore the estimated duration of the latest linear segment is almost equal to the true value. For the oscillating signal, there are nine jumps, which correspond to the movement of the posterior at nine of the peaks and troughs of the oscillating signal. The last trough has disappeared into the signal noise and is difficult to detect. The probability detection indexes are shown in Figs. 4(d) and 4(h), from which we can see that the detection index rises sharply around the starting point of the steady-state operation. This can be explained using Equation (16), with  $h(t)$  being a fourth-order polynomial in  $t$  that rapidly increases with  $t$  shortly after the transition to the steady-state condition.

In many applications, the signals have a decaying variance while the mean remains unchanged. In such cases, the slope detection method and  $t$ -test on the mean of two adjacent moving windows may totally fail. To see the effectiveness of our algorithm, we use a signal with a zero mean and noise amplitude of

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

where  $T_0 = 300$  and  $\sigma_0 = 0.1$ . The detection results shown in Fig. 5 illustrate that the steady-state can be effectively detected with a small detection delay.

Figure 6 shows the computational cost of each time step using three different values of the support:  $m = 50, 100,$  and  $150$ . As we can see, the computational cost per step linearly increases with  $t$  when  $t < m$ . When  $t \geq m$ , the computational cost per step is fully controlled. Without setting a fixed value of  $m$ , the total computational cost would increase in a quadratic manner.

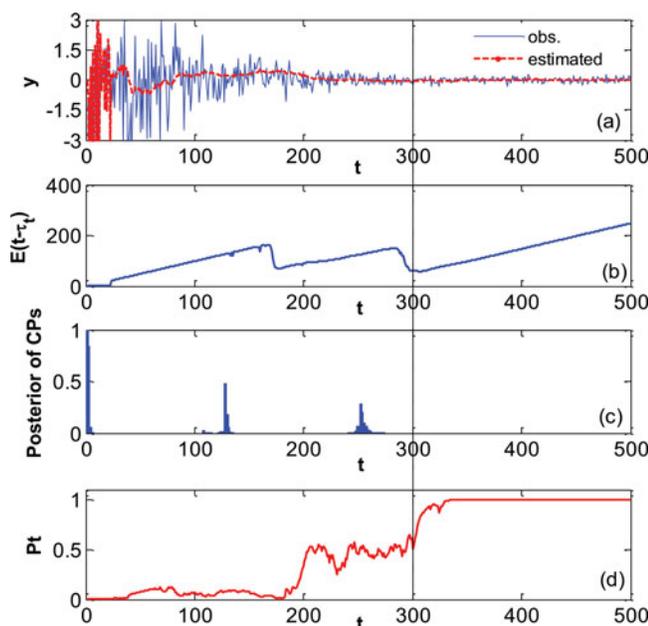


Figure 5. The detection of the steady-state for a signal with a zero mean and an exponentially decreasing variance.

In application,  $m$  may be set as low as 10 to further reduce the execution time.

### 4.2 ARL

To see how accurate the ARL approximation is and how ARL is influenced by the algorithm and the signal parameters, we calculated the ARL using both Monte Carlo simulation and approximation method. In the simulation,  $\Sigma = 1 \times 10^4 I$ ,  $\beta_0 = 0$ , where  $I$  is the identity matrix. The signals were generated using a zero mean and Gaussian noise. The other signal parameters and detection parameters were set to  $\alpha = 0.9$ ,  $s_0 = 0.003$ ,  $\sigma = 0.1$ ,  $p = 0.2$ ,  $v = 20$ , and  $\gamma = 0.2$  for both simulation and approximation except for the changing parameters. The simulation was repeated 500 times for each parameter setting.

Figure 7 shows the ARL as a function of different parameters calculated using both the simulation and approximation methods. As we can see, the approximated ARL is almost identical to the simulated one in all cases. The maximum approximation errors for Figs. 7(a) to 7(f) are 6.48, 5.3, 5.0, 1.88, 3.57, and 3.85%, respectively. We also observe that the simulated ARL is slightly higher than the calculated one in most cases. The reason for this behavior is that the posterior of the LCP is not completely focused on  $t = 1$ —i.e., there is more than one support however, in the approximation we assume that  $t = 1$  is the only support.

The ARL is not very sensitive to  $\alpha$ , as shown in Fig. 7(a); therefore,  $\alpha$  is not treated as a tuning parameter and we suggest using  $\alpha = 0.9$  in actual applications. The transition prior probability  $p$  has almost no influence on the ARL in the range [0.05, 0.5], as shown in Fig. 7(d). Therefore, in practice, we could arbitrarily select a value within that range. The ARL is also insensitive to the hyper-parameters  $v$  and  $\gamma$  of the noise prior. In the application we can roughly predict the noise amplitude and then select  $v$  and  $\gamma$  based on that value. The sole tuning parameter is the slope threshold  $s_0$ , which significantly influences the ARL.

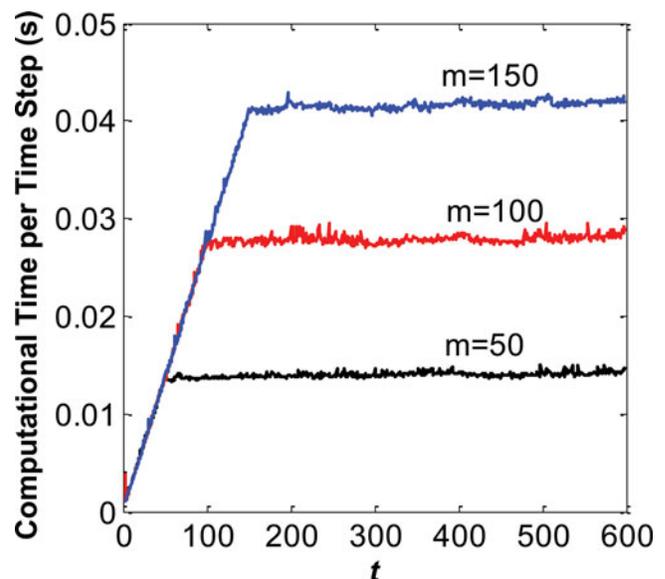
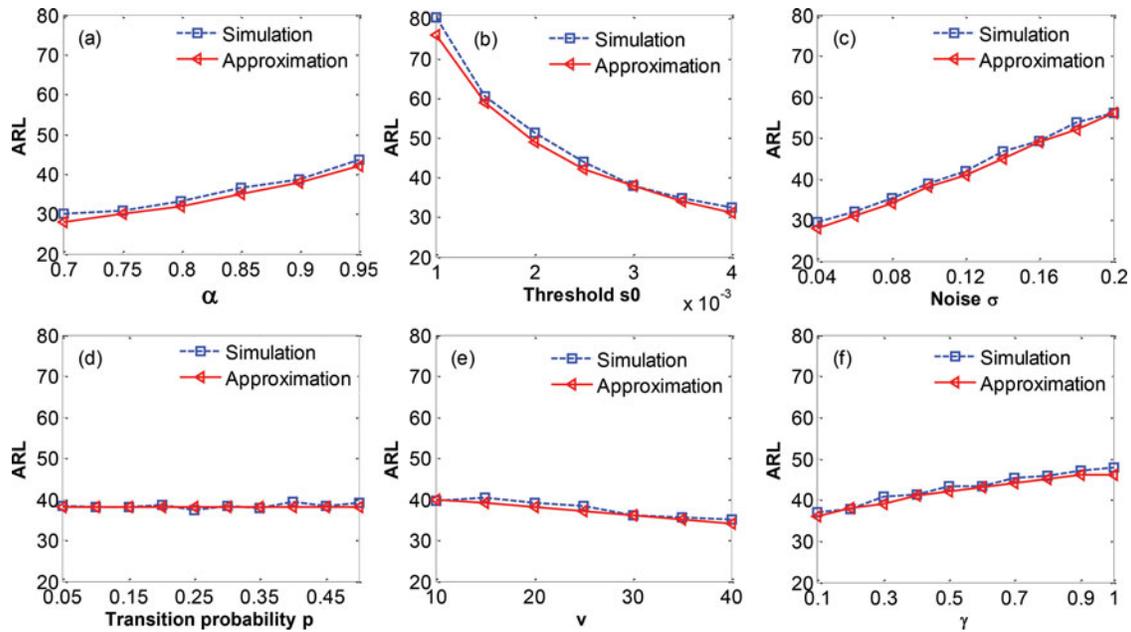


Figure 6. Computational cost for each time step for support sizes  $m = 50, 100,$  and  $150$ .



**Figure 7.** ARL calculated using Monte Carlo simulation and approximation Equation (17). The parameters were set as  $\alpha = 0.9$ ,  $s_0 = 0.003$ ,  $\sigma = 0.1$ ,  $p = 0.2$ ,  $\nu = 20$ , and  $\gamma = 0.2$  for all calculations except the changing parameter.

Figure 7(c) shows that the ARL increases almost linearly with the amplitude of the signal noise. This is an advantage of this algorithm, as it can automatically adjust the ARL to reduce the false alarm rate when the signal noise is large or reduce the detection delay when the noise is small.

### 4.3 Evaluation and comparison with other methods

For statistical monitoring schemes, the performance is typically evaluated in terms of one of two criteria, the False Alarm Rate (FAR) and the detection delay. Usually a required FAR (or detection delay) is specified and the corresponding detection delay (or FAR) is used to evaluate or compare different detection schemes. However, in the detection of a steady-state condition, the FAR does not capture the closeness of the detected time to the true value. In practice, the closeness in the false alarm is also important since it determines the amount of initial bias. Therefore, we develop another evaluation metric, called the Weighted Standard Detection Error (WSDE), which is defined as

$$WSDE = \sqrt{\frac{1}{N} \sum_{i=1}^N w(\hat{T}_i) (\hat{T}_i - T_0)^2}, \quad (18)$$

where  $\hat{T}_i$  is the detected time for the steady-state,  $N$  is the number of replications, and  $w(\hat{T}_i)$  is the penalty weight. The cost of a false detection is often higher than that for the detection delay and therefore the weight is selected as

$$w(\hat{T}_i) = \begin{cases} w \in (0, 1] & \text{if } \hat{T}_i \geq T_0 \\ 1 & \text{if } \hat{T}_i < T_0 \end{cases}$$

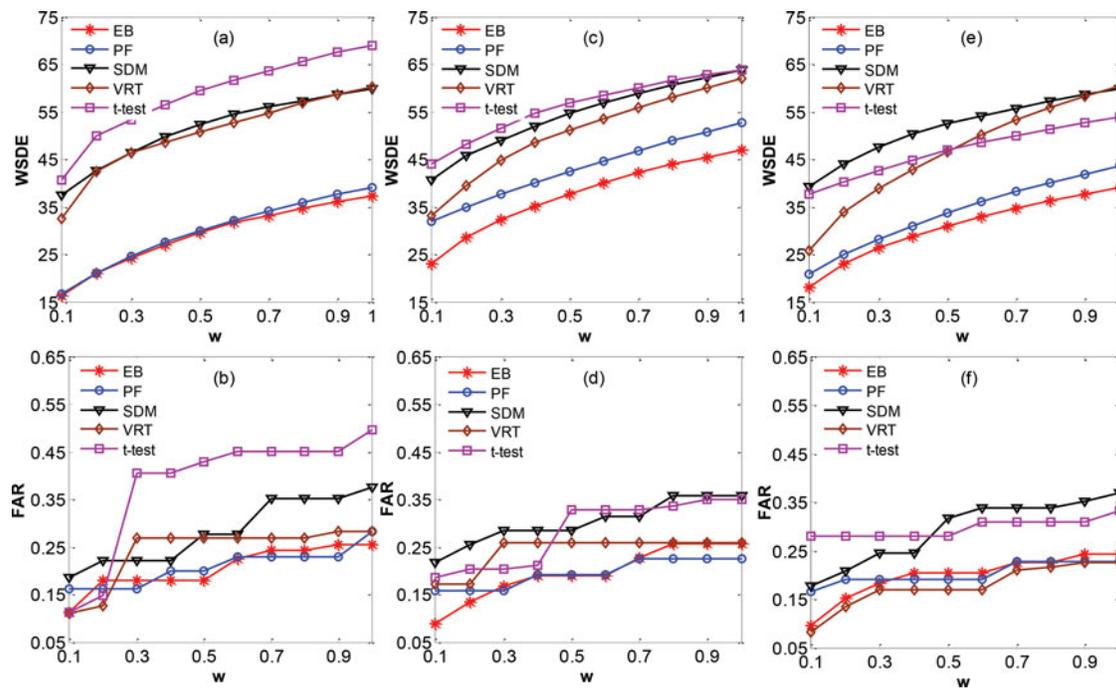
and  $w$  can be treated as the penalty ratio of the detection delay over the FAR. If  $w = 1$  for both FAR and detection delay, then only the detection deviation is considered. A lower  $w$  denotes a higher relative penalty for a false detection.

The proposed method (EB: Exact Bayesian inference) is compared with four other online methods (see the Introduction): the Particle Filter-based method (PF; Wu (2015)), the Slope Detection Method (SDM; Holly *et al.* (1989); Bethea and Rhinehart (1991); Wu *et al.* (2013)), the Variance Ratio Test (VRT) method (Crow *et al.*, 1960; Cao and Rhinehart 1995), and the  $t$ -test method (Narasimhan *et al.*, 1987). Linear, quadratic, exponential, and oscillating signals were used to create opportunities for comparisons. For each type of signal, two sets of signal parameters were used: either  $h = 1$ ,  $T_0 = 200$ , or  $h = 1$ ,  $T_0 = 300$  to simulate different levels of severity in the initial bias. To test the algorithm for both Gaussian/non-Gaussian noise, we used three kinds of autoregressive noise: no auto-correlation (AR(0)), first-order autoregressive correlation (AR(1)), and second-order autoregressive correlation (AR(2)), as shown in Table 2. Three noise amplitudes  $\sigma_t = 0.06$ ,  $0.1$ ,  $0.14$  were used for AR(0) and  $\sigma_t = 0.06$  and  $0.1$  were used for AR(1) and AR(2) correlations. In the simulation,  $N = 500$  signals (replications) were generated for each set of signal parameters. For each set of penalty weight  $w$  and noise type, the detection parameters (window size and threshold for SDM, VRT, and  $t$ -test, slope threshold  $s_0$  for PF and EB) that minimized the overall WSDE of all generated signals were selected (e.g.,  $4 \times 2 \times 3 \times 500$  for AR(0)). The support size  $m = 10$  for EB and the other algorithm parameters were the same as in Section 4.1.

Figure 8 shows the WSDE and FAR as functions of  $w$  for each noise autoregressive type. Here FAR is only used to provide extra detection information. Note that the WSDE and FAR

**Table 2.** Noise types and their parameters

Auto-correlation type	Equation	Parameter
AR(0)	$\psi_t^{(0)} = \epsilon_t$	$\epsilon_t \sim N(0, \sigma_t^2)$
AR(1)	$\psi_t^{(1)} = \phi_1 \psi_{t-1}^{(1)} + \epsilon_t$	$\phi_1 = 0.4$
AR(2)	$\psi_t^{(2)} = \phi_2 \psi_{t-1}^{(2)} + \phi_3 \psi_{t-2}^{(2)} + \epsilon_t$	$\phi_2 = -0.25, \phi_3 = 0.5$



**Figure 8.** The WSDE and FAR of the proposed method (EB), PF, SDM, VRT, and  $t$ -test as a function of penalty weight  $w$  for (a)-(b): AR(0); (c)-(d): AR(1); and (e)-(f): AR(2).

are calculated over all signals for each penalty weight  $w$ . As we can see, the proposed method EB and PF are more accurate than SDM, VRT, and  $t$ -test in terms of the overall WSDE at all penalty weights. EB is slightly better than the PF method in terms of WSDE. The main advantage of EB compared with the PF method is its computational cost. The computation time ( $m = 10$ ) for 500 observations is about 1.5 seconds using MATLAB running on a Q9550 2.83 GHz Intel processor, which is much lower than the PF method (12 seconds for 500 observations with 1000 particles). The value of the FAR of the proposed method is also lower than other methods in most cases.

Table 3 shows the detailed detection results for various bias signals and noise amplitudes in the case of  $w = 1$  and Gaussian noise. The proposed method is much more robust than other methods in terms of the overall WSDE in handling different bias signals using only one set of detection parameters. The maximum WSDE among the different types of signal is also lower than other methods. Note that in order to minimize the overall WSDE, the optimal slope threshold may not guarantee that the proposed method outperforms other methods in all signals. From the table we can also see that decreasing the rate of change in the signal (e.g., changing  $T_0$  from 200 to 300 for linear signals) or increasing the signal noise would result in a higher FAR. This is intuitive, as the initial bias (deviation from steady state) will be more easily drowned out by the noise and thus more difficult to detect. That is, the main contributions to the FAR are made by the quadratic and exponential signals.

#### 4.4 Application in the micro/nanoparticle dispersion process

In this section, we use Cavitation Noise Power (CNP) signals obtained from a micro/nanoparticle dispersion process (Wu

*et al.*, 2013) to illustrate a real application of the proposed steady-state detection procedure. Micro/nanoparticles research is currently an area of intense scientific interest, due to a wide variety of potential applications in the biomedical, optical, and electronics fields. In actual application, micro/nanoparticles often cluster together due to a high surface energy and a large surface-to-volume ratio, which may result in the loss of their size-dependent properties. Therefore, they have to be dispersed before use. Ultrasonic cavitation is an effective method to disperse micro/nanoparticles. In the dispersion process, it has been reported that the steady state of CNP signals corresponds to the maximum extent of dispersion at that ultrasonic power level. Therefore, the dispersion process can be monitored by detecting the steady state of the CNP signals.

Figure 9 shows the detection of CNP signals in the dispersion of 30 g  $Al_2O_3$  particles under values of the ultrasonic power of 30 W (Fig. 9(a)) and 40 W (Fig. 9(b)). We used the same detection parameters as used in the simulation except for the slope threshold, which was set to 0.001. The offline method EWMA-MSER (Wu *et al.*, 2013) was used as a benchmark to evaluate the proposed method. The offline method detected steady-state times of 418 and 293 seconds for 30 and 40 W CNP signals, respectively. In contrast, the detection results for the proposed method are 434 and 329 seconds, or the detection delays are 16 and 35 seconds, respectively, both of which are satisfactory compared with the typical results on WSDE values listed in Table 3.

## 5 Discussion and conclusion

In this article, we have developed a new online steady-state detection method using the multiple-change-point model and EB inference method. Signals were formulated as piecewise linear models and the posterior of the LCP was recursively

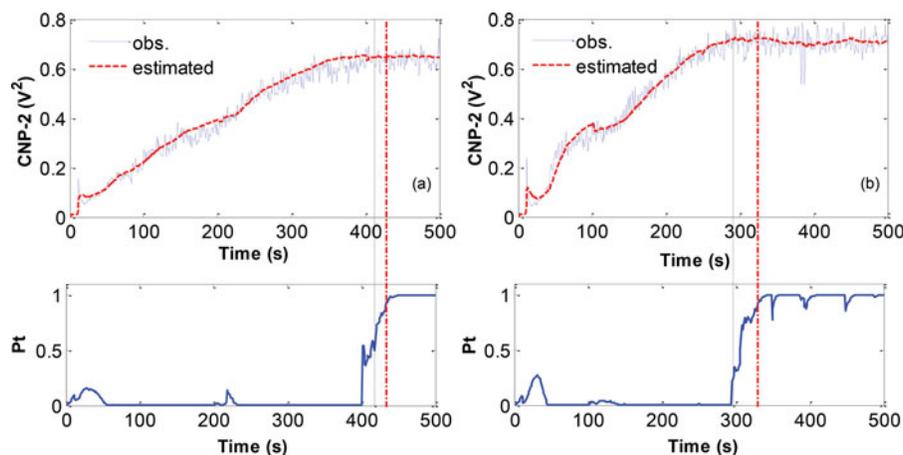
**Table 3.** Comparison of EB, PF, SDM, VRT, and  $t$ -test for  $w = 1$  and noise type AR(0). The detection parameters are (i) EB,  $s_0 = 0.0020$ ; (ii) PF,  $s_0 = 0.0022$ ; (iii) SDM, window size  $L = 50$ , threshold  $= 8 \times 10^{-5}$ ; (iv) VRT,  $L = 98$ , threshold  $= 0.6$ ; (v)  $t$ -test,  $L = 28$ , threshold  $= 0.9$

Signal	$\sigma_y$	WSDE					FAR					
		EB	PF	SDM	VRT	$t$ -test	EB	PF	SDM	VRT	$t$ -test	
Linear	$T_0 = 200$	0.06	38.0	42.0	59.8	78.6	79.0	0	0	0	0	0
		0.10	47.8	53.9	60.7	66.1	65.8	0	0	0	0	0
		0.14	55.4	64.5	57.6	60.4	81.4	0	0	0	0	0.02
	$T_0 = 300$	0.06	37.5	40.9	58.0	70.9	92.1	0	0	0	0	0
		0.10	38.3	53.1	55.9	56.6	65.0	0	0	0	0.04	0.04
		0.14	47.2	70.4	58.6	121.6	122.0	0.02	0.01	0.03	0.75	0.44
Quad.	$T_0 = 200$	0.06	15.3	12.1	33.6	37.6	39.1	0	0.11	0	0	0.04
		0.10	25.7	21.2	31.9	26.7	32.7	0.06	0.04	0.02	0.04	0.16
		0.14	37.2	33.7	28.2	18.5	40.8	0	0.06	0.14	0.36	0.58
	$T_0 = 300$	0.06	34.0	33.8	22.4	16.3	46.4	1	1	0.11	0.33	0.2
		0.10	32.8	28.1	23.9	37.5	49.8	0.98	0.93	0.45	0.83	0.72
		0.14	34.6	22.4	34.5	72.8	70.0	0.84	0.62	0.62	0.93	0.84
Exp.	$T_0 = 200$	0.06	18.4	16.6	45.9	44.4	58.0	0.06	0.12	0	0	0.02
		0.10	24.7	26.3	40.5	23.2	48.6	0.1	0.08	0.03	0.30	0.22
		0.14	40.7	35.0	34.8	26.1	47.5	0.08	0.06	0.17	0.73	0.7
	$T_0 = 300$	0.06	54.0	61.7	35.5	27.2	41.0	1	1	0.04	0.51	0.28
		0.10	54.9	49.4	32.2	67.3	65.1	0.96	0.98	0.38	0.96	0.8
		0.14	52.6	40.1	55.7	107.8	103.6	0.94	0.88	0.82	0.99	0.98
Osc.	$T_0 = 200$	0.06	23.3	27.1	94.9	74.4	60.0	0	0.01	1	0	1
		0.10	31.6	27.6	90.2	61.5	70.1	0	0.04	0.99	0	0.98
		0.14	35.1	26.6	94.0	54.1	82.8	0	0.06	1	0	0.94
	$T_0 = 300$	0.06	21.7	23.0	156	63.6	73.2	0	0.04	1	0	1
		0.10	24.7	25.7	156	49.8	71.5	0	0.2	0.99	0	0.98
		0.14	24.2	29.4	152	40.6	72.5	0.1	0.55	1	0	0.98
Overall		37.3	39.1	59.9	60.3	69.0	0.25	0.28	0.38	0.28	0.50	

calculated using a recursive message-passing algorithm. The slope and intercept of the current linear segment conditioning on the LCP and observations were proved to follow nonstandard bivariate Student  $t$  distribution. Based on this finding, a probability index was developed to detect the steady state.

A fixed support size strategy for the posterior of the LCP was proposed using weighted sampling without replacement to control and balance the computational cost of each time step. Other approximation strategies for the gamma ratio and probability index were also proposed for further reduction of the computational cost. The computational cost of the proposed method is significantly lower than the PF method. An accurate approximation formula for the ARL on the steady-state observations was derived to provide insight and guidance on understanding and tuning the proposed method.

The performance of the proposed method was evaluated based on the WSDE. The simulation results demonstrated that the proposed method is much more robust in detecting bias signals under various noise levels/types and bias severity. It functions like an adaptive SDM, using adaptive window sizes based on the bias shape and noise amplitude and mainly using the observations, as the LCP for steady-state testing. In this article, only the non-informative priors are used for the parameters of the change-point model. In practice, however, we may have prior knowledge about the signals (e.g., bias shapes, steady-state transition point interval, etc.), and thus informative priors could be applied for further improve the detection accuracy. Note that for the purpose of exact calculation, hierarchal prior specifications are not implemented in the current algorithm, which may be one disadvantage of the proposed method. We will leave this to future work to improve its robustness.



**Figure 9.** Steady-state detection of CNP signals in the dispersion of 30 g  $Al_2O_3$  for (a) ultrasonic power 30 W and (b) ultrasonic power 40 W. The dash-dotted line denotes the detected time using the proposed method and the dashed line denotes the EWMA-MSER (off-line) (Wu et al., 2013) detected time.

## Notes on contributors

**Jianguo Wu** is an Assistant Professor in the Department of Industrial, Manufacturing and Systems Engineering at the University of Texas–El Paso, Texas. He received a B.S. degree in Mechanical Engineering from Tsinghua University, Beijing, China, in 2009; an M.S. degree in Mechanical Engineering from Purdue University, West Lafayette, Indiana, in 2011 an M.S. degree in Statistics in 2014 and a Ph.D. degree in Industrial and Systems Engineering in 2015, both from the University of Wisconsin–Madison, Madison, Wisconsin. His research interests are focused on statistical modeling, monitoring and analysis of complex processes/systems for quality control, and productivity improvement through integrated application of metrology, engineering domain knowledge, and data analytics. He is a member of INFORMS, IIE, and SME.

**Yong Chen** received a B.E. degree in Computer Science from Tsinghua University, Beijing, China, in 1998; an M.S. degree in Statistics and a Ph.D. degree in Industrial and Operations Engineering from the University of Michigan, Ann Arbor, Michigan, in 2003. He is currently an Associate Professor with the Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City. His research interests include reliability modeling, robust sensor data processing, and maintenance decision making. He received best paper awards from *IIE Transactions* in 2010 and 2004.

**Shiyu Zhou** is a Professor in the Department of Industrial and Systems Engineering at the University of Wisconsin–Madison. He received his B.S. and M.S. in Mechanical Engineering from the University of Science and Technology of China in 1993 and 1996, respectively, and his master's in Industrial Engineering and Ph.D. in Mechanical Engineering from the University of Michigan in 2000. His research interests include in-process quality and productivity improvement methodologies by integrating statistics, system and control theory, and engineering knowledge. He is a recipient of a CAREER Award from the National Science Foundation and the Best Application Paper Award from *IIE Transactions*. He is a member of IIE, INFORMS, ASME, and SME.

## Acknowledgments

The authors gratefully appreciate the editor and the referees for their valuable comments and suggestions. This research is supported by the National Science Foundation grants #1335129 and #1335454 and the Air Force Office of Scientific Research grant #FA9550-14-1-0384.

## References

- Adams, R.P. and MacKay, D.J. (2007) Bayesian online changepoint detection. Technical report, University of Cambridge, Cambridge, UK. arXiv preprint arXiv:0710.3742.
- Aguado, D., Ferrer, A., Seco, A. and Ferrer, J. (2008) Using Unfold-PCA for batch-to-batch start-up process understanding and steady-state identification in a sequencing batch reactor. *Journal of Chemometrics*, **22**, 81–90.
- Bethea, R.M. and Rhinehart, R.R. (1991) *Applied Engineering Statistics*, vol. 121, Marcel Dekker, New York.
- Bhat, S.A. and Saraf, D.N. (2004) Steady-state identification, gross error detection, and data reconciliation for industrial process units. *Industrial & Engineering Chemistry Research*, **43**, 4323–4336.
- Cao, S. and Rhinehart, R.R. (1995) An efficient method for on-line identification of steady-state. *Journal of Process Control*, **5**, 363–374.
- Cash, C.R., Nelson, B.L., Dippold, D.G., Long, J.M. and Pollard, W.P. (1992) Evaluation of tests for initial-condition bias. In Proceedings of the 24th Conference on Winter simulation, ACM, pp. 577–585.
- Chen, J. and Howell, J. (2001) A self-validating control system based approach to plant fault detection and diagnosis. *Computers & Chemical Engineering*, **25**(2), 337–358.
- Chopin, N. (2007) Dynamic detection of change points in long time series. *Annals of the Institute of Statistical Mathematics*, **59**, 349–366.
- Crow, E.L., Davis, F.A. and Maxfield, M.W. (1960) *Statistics Manual: With Examples Taken from Ordnance Development*, Dover Publications, New York.

- Fearnhead, P. and Liu, Z. (2007) On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**, 589–605.
- Franklin, W.W. (2009) *The Theoretical Foundation of the MSER Algorithm*, Ph.D. thesis, Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA.
- Gordon, G. (1977) *System Simulation*, Prentice Hall PTR, Upper Saddle River, NJ.
- Hoad, K., Robinson, S. and Davies, R. (2009) Automating warm-up length estimation. *Journal of the Operational Research Society*, **61**, 1389–1403.
- Holly, W., Cook, R. and Crowe, C. (1989) Reconciliation of mass flow rate measurements in a chemical extraction plant. *The Canadian Journal of Chemical Engineering*, **67**, 595–601.
- Kelton, W.D. and Law, A.M. (1983) A new approach for dealing with the startup problem in discrete event simulation. *Naval Research Logistics Quarterly*, **30**, 641–658.
- Kim, M., Yoon, S.H., Domanski, P.A. and Vance Payne, W. (2008) Design of a steady-state detector for fault detection and diagnosis of a residential air conditioner. *International Journal of Refrigeration*, **31**, 790–799.
- Kotz, S. and Nadarajah, S. (2004) *Multivariate t-Distributions and Their Applications*, Cambridge University Press, Cambridge, UK.
- Li, B. and Moor, B.D. (1998) A corrected normal approximation for the Student's *t* distribution. *Computational Statistics & Data Analysis*, **29**, 213–216.
- Li, H. (2004) A decoupling-based unified fault detection and diagnosis approach for packaged air conditioners. Ph.D. thesis, Purdue University, West Lafayette, IN.
- Mahuli, S.K., Rhinehart, R.R. and Riggs, J.B. (1992) Experimental demonstration of non-linear model-based in-line control of pH. *Journal of Process Control*, **2**, 145–153.
- Marchetti, A., Ferramosca, A. and González, A. (2014) Steady-state target optimization designs for integrating real-time optimization and model predictive control. *Journal of Process Control*, **24**, 129–145.
- Mhamdi, A., Geffers, W., Flehmig, F. and Marquardt, W. (1999) On-line optimization of MSF desalination plants, in *Desalination and Water Resources: Thermal Desalination Processes*, vol. 1, EOLSS, Eolss Publishers, Paris, France, pp. 136–163.
- Narasimhan, S., Kao, C.S. and Mah, R. (1987) Detecting changes of steady-states using the mathematical theory of evidence. *AIChE Journal*, **33**, 1930–1932.
- Natarajan, S. and Rhinehart, R.R. (1997) Automated stopping criteria for neural network training, in *Proceedings of the 1997 American Control Conference*, IEEE Press, Piscataway, NJ, pp. 2409–2413.
- Nelson, B.L. (1992) Statistical analysis of simulation results, in *Handbook of Industrial Engineering: Technology and Operations Management*, third edition, John Wiley, New York, pp. 2469–2495.
- Ricker, N. and Lee, J. (1995) Nonlinear modeling and state estimation for the Tennessee Eastman challenge process. *Computers & Chemical Engineering*, **19**, 983–1005.
- Robinson, S. (2002) New simulation output analysis techniques: a statistical process control approach for estimating the warm-up period, in *Proceedings of the 34th Conference on Winter Simulation: Exploring New Frontiers*, Society for Modeling and Simulation International, Vista, CA, pp. 439–446.
- Schladt, M. and Hu, B. (2007) Soft sensors based on nonlinear steady-state data reconciliation in the process industry. *Chemical Engineering and Processing: Process Intensification*, **46**, 1107–1115.
- Spratt, S. (1998) *Heuristics for the Startup Problem*. M.S. Thesis, Department of Systems Engineering, University of Virginia, Charlottesville, VA.
- Tricomi, F.G. and Erdélyi, A. (1951) The asymptotic expansion of a ratio of gamma functions. *Pacific Journal of Mathematics*, **1**, 133–142.
- Wang, X. (1997) Incorporating knowledge on segmental duration in HMM-based continuous speech recognition, Ph.D. Thesis, Institute of Phonetic Sciences, University of Amsterdam.
- White, K.P. Jr. (1997) An effective truncation heuristic for bias reduction in simulation output. *Simulation*, **69**, 323–334.
- White, K.P., Jr., Cobb, M.J. and Spratt, S.C. (2000) A comparison of five steady-state truncation heuristics for simulation, In *Proceedings of the 32nd Conference on Winter Simulation*, Society for Modeling and Simulation International, Vista, CA, pp. 755–760.

Wong, C.-K. and Easton, M.C. (1980) An efficient method for weighted sampling without replacement. *SIAM Journal on Computing*, **9**, 111–113.

Wu, J., Chen, Y., Li, X. and Zhou, S. (2015) Online steady-state detection for process control using multiple change-point models and particle filters. *IEEE Transactions on Automation Science and Engineering*. DOI: 10.1109/TASE.2014.2378150

Wu, J., Zhou, S. and Li, X. (2013) Acoustic emission monitoring for ultrasonic cavitation based dispersion process. *Journal of Manufacturing Science and Engineering*, **135**, 031015.

Yao, Y., Zhao, C. and Gao, F. (2009) Batch-to-batch steady-state identification based on variable correlation and Mahalanobis distance. *Industrial & Engineering Chemistry Research*, **48**, 11060–11070.

## Appendices

### Appendix A: Calculation of $P(s, t)$

$$\begin{aligned}
P(s, t) &= \int P(y_{st} | \sigma^2, \boldsymbol{\beta}) p(\sigma^2) p(\boldsymbol{\beta} | \sigma^2) d\sigma^2 d\boldsymbol{\beta} \\
&= \int \left[ \frac{\left(\frac{\gamma}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} (\sigma^2)^{-\frac{v}{2}-1} e^{-\frac{\gamma}{2\sigma^2}} \right] \left[ (2\pi)^{-1} |\sigma^2 \boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{2\sigma^2}} \right] \left[ (2\pi)^{-\frac{t-s+1}{2}} (\sigma^2)^{-\frac{t-s+1}{2}} e^{-\frac{y_{st}-\mathbf{X}_{st}^T \boldsymbol{\beta}}{2\sigma^2}} \right] d\sigma^2 d\boldsymbol{\beta} \\
&= \int \frac{(\gamma/2)^{\frac{v}{2}}}{\Gamma(v/2)} (2\pi)^{-\frac{t-s+3}{2}} (\sigma^2)^{-\frac{t-s+v+5}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{H_{st}}{2\sigma^2}\right] \exp\left[-\frac{(\boldsymbol{\beta}-\boldsymbol{\mu}_{st})^T (\mathbf{X}_{st}^T \mathbf{X}_{st} + \boldsymbol{\Sigma}^{-1}) (\boldsymbol{\beta}-\boldsymbol{\mu}_{st})}{2\sigma^2}\right] d\sigma^2 d\boldsymbol{\beta} \\
&= \int \frac{(\gamma/2)^{\frac{v}{2}}}{\Gamma(v/2)} (2\pi)^{-\frac{t-s+1}{2}} (\sigma^2)^{-\frac{t-s+v+3}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{H_{st}}{2\sigma^2}\right] |\mathbf{X}_{st}^T \mathbf{X}_{st} + \boldsymbol{\Sigma}^{-1}|^{-\frac{1}{2}} d\sigma^2 \\
&= \pi^{-(t-s+1)/2} \left(\frac{|\mathbf{M}_{st}|}{|\boldsymbol{\Sigma}|}\right)^{\frac{1}{2}} \frac{\gamma^{\frac{v}{2}}}{(H_{st})^{(t-s+1+v)/2}} \frac{\Gamma((t-s+1+v)/2)}{\Gamma(v/2)}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{M}_{st} &= (\mathbf{X}_{st}^T \mathbf{X}_{st} + \boldsymbol{\Sigma}^{-1})^{-1}, \\
\mathbf{N}_{st} &= (\boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}_0 + \mathbf{X}_{st}^T y_{st}), \\
\boldsymbol{\mu}_{st} &= \mathbf{M}_{st} \mathbf{N}_{st}, \\
H_{st} &= y_{st}^T y_{st} + \gamma + \boldsymbol{\beta}_0^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}_0 - \mathbf{N}_{st}^T \mathbf{M}_{st} \mathbf{N}_{st}.
\end{aligned}$$

### Appendix B: Proof of Lemma 1

1. For simplicity, we use  $\boldsymbol{\beta}$  and  $\sigma^2$  instead of  $\boldsymbol{\beta}_t$  and  $\sigma_t^2$  in the following derivation:

$$\begin{aligned}
P(\boldsymbol{\beta} | \tau_t = s, y_{st}) &= \int P(\boldsymbol{\beta}, \sigma^2 | \tau_t = s, y_{st}) d\sigma^2 = \int \frac{P(\boldsymbol{\beta}, \sigma^2) P(y_{st} | \boldsymbol{\beta}, \sigma^2, \tau_t = s)}{P(s, t)} d\sigma^2 \\
&= \frac{1}{P(s, t)} \int \left[ \frac{\left(\frac{\gamma}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} (\sigma^2)^{-\frac{v}{2}-1} e^{-\frac{\gamma}{2\sigma^2}} \right] \left[ (2\pi)^{-1} |\sigma^2 \boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{2\sigma^2}} \right] \left[ (2\pi)^{-\frac{t-s+1}{2}} (\sigma^2)^{-\frac{t-s+1}{2}} e^{-\frac{\|y_{st}-\mathbf{X}_{st}^T \boldsymbol{\beta}\|^2}{2\sigma^2}} \right] d\sigma^2 \\
&= \frac{\left(\frac{\gamma}{2}\right)^{\frac{v}{2}}}{P(s, t) \Gamma\left(\frac{v}{2}\right)} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} (2\pi)^{-\frac{t-s+3}{2}} \Gamma\left(\frac{t-s+3+v}{2}\right) \left[ \frac{\left(\|y_{st}-\mathbf{X}_{st}^T \boldsymbol{\beta}\|^2 + (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta}-\boldsymbol{\beta}_0) + \gamma\right)}{2} \right]^{-\frac{t-s+3+v}{2}} \\
&\propto \left[ \frac{\left(\|y_{st}-\mathbf{X}_{st}^T \boldsymbol{\beta}\|^2 + (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta}-\boldsymbol{\beta}_0) + \gamma\right)}{2} \right]^{-\frac{t-s+3+v}{2}} \\
&\propto [(\boldsymbol{\beta}-\boldsymbol{\mu}_{st})^T (\mathbf{X}_{st}^T \mathbf{X}_{st} + \boldsymbol{\Sigma}^{-1}) (\boldsymbol{\beta}-\boldsymbol{\mu}_{st}) + y_{st}^T y_{st} + \boldsymbol{\beta}_0^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}_0 + \gamma \\
&\quad - (y_{st}^T \mathbf{X}_{st} + \boldsymbol{\beta}_0^T \boldsymbol{\Sigma}^{-1}) (\mathbf{X}_{st}^T \mathbf{X}_{st} + \boldsymbol{\Sigma}^{-1})^{-1} (\mathbf{X}_{st}^T y_{st} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}_0)]^{-\frac{t-s+3+v}{2}} \\
&\propto \left[ 1 + \frac{1}{t-s+v+1} (\boldsymbol{\beta}-\boldsymbol{\mu}_{st})^T \boldsymbol{\Sigma}_{st}^{-1} (\boldsymbol{\beta}-\boldsymbol{\mu}_{st}) \right]^{-\frac{t-s+3+v}{2}},
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\mu}_{st} &= \mathbf{M}_{st} \mathbf{N}_{st}, \\
\boldsymbol{\Sigma}_{st} &= \frac{H_{st} \mathbf{M}_{st}}{(t-s+v+1)}.
\end{aligned}$$

Therefore,

$$(\boldsymbol{\beta}_t | \tau_t = s, y_{s:t}) \sim t_2(t - s + v + 1, \boldsymbol{\mu}_{st}, \boldsymbol{\Sigma}_{st}).$$

2. Suppose  $\boldsymbol{\Sigma}_{st} = \mathbf{K}_{st} \mathbf{R}_{st} \mathbf{K}_{st}^T$ , where

$$\mathbf{K}_{st} = \begin{bmatrix} k_{st}^{(1,1)} & 0 \\ 0 & k_{st}^{(2,2)} \end{bmatrix}$$

and  $\mathbf{R}_{st}$  is the correlation matrix. Let  $\boldsymbol{\beta}^* = \mathbf{K}_{st}^{-1} \boldsymbol{\beta}_t$ , then

$$P(\boldsymbol{\beta}^* | \tau_t = s, y_{s:t}) \propto \left[ 1 + \frac{1}{t - s + v + 1} (\boldsymbol{\beta}^* - \mathbf{K}_{st}^{-1} \boldsymbol{\mu}_{st})^T \mathbf{R}_{st}^{-1} (\boldsymbol{\beta}^* - \mathbf{K}_{st}^{-1} \boldsymbol{\mu}_{st}) \right]^{-\frac{t-s+3+v}{2}}.$$

Therefore,

$$(\boldsymbol{\beta}^* | \tau_t = s, y_{s:t}) \sim t_2(t - s + v + 1, \mathbf{K}_{st}^{-1} \boldsymbol{\mu}_{st}, \mathbf{R}_{st}).$$

According to Kotz and Nadarajah (2004), the marginal distribution

$$\left( \frac{a_t}{k_{st}^{(1,1)}} | \tau_t = s, y_{s:t} \right) \sim t_1(t - s + v + 1, (\mathbf{K}_{st}^{-1} \boldsymbol{\mu}_{st})^{(1)}, 1),$$

or  $((a_t - \boldsymbol{\mu}_{st}^{(1)})/k_{st}^{(1,1)} | \tau_t = s, y_{s:t})$  follows standard univariate  $t$  distribution with degrees of freedom  $t - s + v + 1$ .

This completes the proof. ■