A Sequential Bayesian Partitioning Approach for Online Steady-State Detection of Multivariate Systems

Jianguo Wu, Honglun Xu, Chen Zhang, and Yuan Yuan

***Abstract*— The steady-state detection is critically important in many engineering fields, such as fault detection and diagnosis and process monitoring and control. However, most of the existing methods were designed for univariate signals and, thus, are not effective in handling multivariate signals. In this paper, we propose an efficient online steady-state detection method for multivariate systems through a sequential Bayesian partitioning approach. The signal is modeled by a Bayesian piecewise constant mean and covariance model, and a recursive updating method is developed to calculate the posterior distributions analytically. The duration of the current segment is utilized for steady-state testing. Insightful guidance is also provided for hyperparameter selection. The effectiveness of the proposed method is demonstrated through thorough numerical and real case studies.**

***Note to Practitioners*—This paper addresses the problem of online steady-state detection of systems captured by multivariate signals. Existing approaches often monitor each signal independently, and the system is claimed steady when all signals reach steady state. These methods have many shortcomings, such as failing to consider the correlations among signals and suffering the multiple testing problems. In this paper, we propose a novel joint monitoring approach, where the multivariate signal is sequentially partitioned into segments of constant mean and covariance through an online Bayesian inference scheme, and once the current segment duration is sufficiently large, the signal is considered steady. We also provide several insightful guidelines to select appropriate hyperparameters under different scenarios. The proposed approach is much more accurate and robust than existing ones. However, this method may face prohibitive computational cost and ill-posed covariance inversion problem when there are hundreds or even thousands of variables in the system. In future research, we will develop efficient distributed monitoring and data fusion techniques to overcome these challenges.**

Manuscript received December 22, 2018; accepted February 15, 2019. Date of publication March 13, 2019; date of current version October 4, 2019. This paper was recommended for publication by Associate Editor K. Liu and Editor S. Reveliotis upon evaluation of the reviewers' comments. This work was supported by the Natural Science Foundation of China under Grant 51875003. *(Corresponding author: Jianguo Wu.)*

J. Wu is with the Department of Industrial Engineering and Management, College of Engineering, Peking University, Beijing 100871, China (e-mail: j.wu@pku.edu.cn).

H. Xu is with the Computational Science Program, The University of Texas at El Paso, El Paso, TX 79968 USA (e-mail: hxu3@miners.utep.edu).

C. Zhang is with the Department of Industrial Engineering, Tsinghua University, Beijing 100084, China (e-mail: c.zhang.thu@gmail.com).

Y. Yuan is with the IBM Research, Singapore 18983 (e-mail: polarisyy@gmail.com).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2019.2900482

***Index Terms*— Change-point detection, hyperparameter selection, multivariate system, sequential Bayesian partitioning, steady-state detection.**

I. Introduction

DETECTING whether a system is operating under steady-state condition is essential in process performance assessment and optimization [1], [2], fault detection and diagnosis [3], [4], and process automation and control [5]–[9]. It arises in many engineering fields, such as process industries, chemical engineering, and manufacturing process automation and control. In these applications, a steady-state condition is the basic requirement for process modeling, evaluation, monitoring, and control. In the discrete-event simulations [1], [10], for example, the steady state is not achieved until sometime after the system is started or initialized. The initial situation is often referred to as transient state, startup or warm-up period. Only the steady-state period data (e.g., throughput, work-in-process) represent the true performance of the system and, thus, needs to be identified for process assessment and optimization. In process or chemical industries, it is often mandatory to use steady-state data (e.g., flow rate, pH value, temperature, and pressure) for process modeling and design, and real-time optimization [6], [11]. In batch processes manufacturing [12], the operation is often unsteady during the startup period due to unstabilized material or machine conditions, which cannot guarantee a satisfied product quality. To avoid costly quality inspection and scrap costs, the steady-state operation needs to be identified. In process automation and control, the steady state can be used to trigger the next action. For example, in ultrasonic cavitation-based nanoparticle dispersion process, the particles are considered completely dispersed and the process can be stopped when the cavitation noise signal enters into the steady state [7]–[9].

In the past few decades, various types of univariate offline methods have been developed for initial bias truncation in discrete-event simulations [1], [10]. These methods often require a sufficiently large number of steady-state observations for warm-up truncation and, thus, are not applicable for online detection. In contrast, there are a relatively small number of online steady-state detection algorithms for univariate signals in the existing literature, most of which can be classified into the following categories.

1) Regression-based approach [13] where a simple linear regression is performed over a moving data window and the fitted slope is monitored. Once the magnitude of the slope is below a predefined threshold, the signal is considered steady.
2) Performing a t-test on the difference of the means of two adjacent moving windows. If the difference is significantly small, the signal is claimed steady [14].
3) Performing an F-test [variance ratio test (VRT)] on the ratio of two variances of a moving window calculated using different methods, namely, the mean squared deviation and the mean squared differences of successive data [15]. In the steady-state period, the ratio is expected to be near unity.
4) Monitoring the variance of a moving window [4].

When the variance is below a threshold, the signal is steady.

However, all the aforementioned online methods are developed for univariate signals. In practice, most of the systems or processes are inherently multivariate. With the rapid development of sensing technology, multiple sensor signals have become unprecedentedly available to better capture the system conditions. Therefore, multivariate steady-state detection algorithms are highly desirable. To the best of our knowledge, there are very limited methods for multivariate signals. Brown and Rhinehart [16] proposed to monitor each signal separately using an existing univariate steady-state detection algorithm. Once all signals are steady, the process is claimed to be steady. However, this strategy inevitably suffers the notorious multiple testing problem [17] with inflated type-II error or detection delay. In addition, it is incapable of detecting the change of correlation among different variables. Jiang *et al* [18] proposed to fuse the steady-state indices of all variables into one through the Dempster's rule of combination [19]. This approach is nevertheless a generalization of Brown and Rhinehart's method [16] and, thus, shares the same shortcomings. In addition, it requires sufficient historical steady-state data to determine the testing threshold, which is often unrealistic in practice due to the data unavailability and run-to-run trajectory variations. Note that if sufficient historical steady-state observations are available and follow the same statistical distribution for all runs, a large number of existing statistical process control (SPC) techniques [20] are readily available for the steady-state detection, e.g., detecting the first in-control sample. However, in the steady-state detection applications, the distribution of the steady-state data often varies from run to run, due to known or unknown process conditions. Therefore, the existing multivariate SPC techniques cannot be directly used. Most of the other multivariate methods are developed in the chemical batch processes [12], [21]. In these methods, a dimension reduction technique is applied first, such as the multiway principal component analysis [21], dynamic principal component analysis [12], and then either a univariate method is applied on a combined index or each extracted feature is monitored individually. These methods also have more or less the aforementioned disadvantages.

To overcome these disadvantages, this paper develops an efficient online multivariate steady-state detection method using a sequential Bayesian partitioning approach. In this method, the multivariate signal is sequentially segmented into phases of constant mean and covariance matrix under the Bayesian framework, and the posterior distribution of the phase duration is used to test the steady state. Once the duration is sufficiently large, the signal is claimed steady. The main challenges of this method are how to sequentially find the change-point regarding the mean and covariance, and how to select appropriate hyperparameters. To overcome these challenges, we develop an efficient recursive method to calculate the posterior distributions analytically and then provide several insightful guidelines on hyperparameter selection.

The rest of this paper is organized as follows. In section II, the steady-state detection problem is formulated into a piecewise constant modeling of multivariate signals. Section III presents the technical details of online change-point detection, computational cost reduction, and appropriate hyperparameter selection. The numerical illustration, performance comparison, and real case studies are provided in Section IV. Section V presents our conclusion and discussion.

II. Piecewise Constant Modeling of Multivariate Signals for Steady-State Detection

To detect whether a system is steady, it is necessary to first define what steady state is. In mathematics or statistics, an alternative term "stationary process," is often used, which is defined as a stochastic process x_t whose joint probability distribution $p(x_t, x_{t+1}, \ldots, x_{t+s})$ does not change over time t (strict or strong stationarity). Consequently, the mean and variance or covariance parameters do not change over time. In this paper, we define steady state as the condition where the mean and covariance of signals capturing the system dynamics are unchanging in a certain period. It is worth noting that in some practical applications, such as the discrete-event simulations and batch processes manufacturing, once steady state occurs, it is expected not to change anymore. Therefore, the online monitoring can be stopped once steady state is detected. However, in many other applications, e.g., monitoring temperature, pressure, and pH value in process or chemical industries [6], the transient state and steady state often occur alternatively, due to unexpected system faults, disturbances, or closed-loop control actions. Therefore, it requires the monitoring scheme to be able to detect the occurrence of multiple steady states and transient states.

Based on the definition, we propose to utilize a piecewise constant model to fit the multivariate signals, where each segment is modeled with a unique mean and covariance matrix. Take two univariate signals for example (Fig. 1). The segment duration would be very short in the transient period, whereas it is expected to be long in the steady-state period. In other words, the change-points between the successive segments occur more frequently and continuously in the transient period due to the rapid change of mean or covariance. When the system is in the steady-state period, there would be no change-points. In the online steady-state detection, the segment duration can be used as a monitoring statistic. Once it is sufficiently large, e.g., larger than a certain threshold, the system is considered steady.

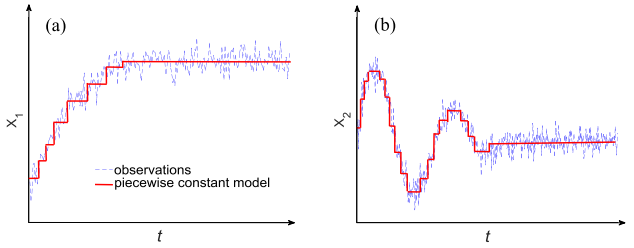


Fig. 1. Illustration of piecewise constant modeling of nonlinear signals with initial transient period. (a) Exponential function in the transient period. (b) Oscillating function in the transient period.

Let the multivariate signal be $\boldsymbol{X}_t, t = 1, 2, \ldots$, where $\boldsymbol{X}_t$ is a p-dimensional vector. Suppose $\boldsymbol{X}_t \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ and $\boldsymbol{X}_t, t = 1, 2, \ldots$, are independent. Define $\boldsymbol{\Phi}_t = (\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$. Suppose the change-points are at positions $\{c_1, c_2, \ldots, c_k, \ldots\}$, where $0 < c_1 < c_2 < \cdots < c_k < \cdots$, then, the piecewise constant model can be mathematically expressed as

$$\boldsymbol{\Phi}_t = \begin{cases} \boldsymbol{\Phi}^{(1)}, & \text{if } c_0 < t \le c_1 \\ \boldsymbol{\Phi}^{(2)}, & \text{if } c_1 < t \le c_2 \\ \quad \cdots & \\ \boldsymbol{\Phi}^{(k)}, & \text{if } c_{k-1} < t \le c_k \\ \quad \cdots & \end{cases} \tag{1}$$

where $\boldsymbol{\Phi}^{(i)}$ is the distribution parameter of the ith segment. Based on the above discussion, any interval $[c_{k-1}, c_k]$ with length $c_k - c_{k-1}$ greater than a certain value, e.g., L_0, can be claimed as a steady-state period.

In the online steady-state detection, the testing is performed each time a new observation is obtained. Therefore, the multivariate signal needs to be fitted sequentially using a piecewise constant model, and the duration of the current segment has to be estimated to decide if the current segment is long enough to claim a steady state. Although the idea is simple, how to efficiently estimate the latest change-point (i.e., the starting time of the current segment) in a real-time manner is nevertheless very challenging. Online Bayesian updating is particularly effective in the dynamic analysis of a sequence of data with change-points [22]. In addition, it provides uncertainty estimates in the number and locations of change-points, which are more realistic in applications. Therefore, in this paper, we propose to use a Bayesian approach, where a Bayesian piecewise constant model is formulated, and then, the posterior distribution of the latest change-point is calculated for the steady-state detection. The technical details are provided in Section III.

III. Online Bayesian Piecewise Constant Model Fitting and Steady-State Detection

A. Bayesian Formation and Prior Specification

For a Bayesian piecewise constant model, appropriate priors for the change-points, the mean, and covariance matrix of each segment need to be assigned. For a time series of fixed length, a joint prior can be specified for both the change-point number k and positions [23], e.g., $\pi(k, \{\delta_i\}_{i=1}^{i=k+1}) = \pi(k)\pi(\{\delta_i\}_{i=1}^{i=k+1}|k)$ where δ_i is the duration of the ith segment. However, this approach is not appropriate or straightforward for dynamic sequences with an increasing length. Instead, the prior is often specified by modeling the occurrence of change-points through a Markov process, where the next change-point only depends on the duration of the current segment [22]. For example, a Poisson point process can be assumed for the occurrence of the change-points or equivalently an exponential distribution is assumed for the segment durations. Another simple prior is the geometric prior applied to the segment duration, which corresponds to a constant Markov transition probability for the latest change-point at each time step. In fact, the stochastic process approach indirectly specifies a joint prior distribution for the number of change-points and their positions [24]. The advantage of this approach is that the prior transition probability of the latest change-point can be easily calculated, which is convenient for online change-point detection. Suppose the latest change-point (the time index for the last observation of the previous segment) at time step t is τ_t, then

$$\begin{aligned} &P(\tau_t = j|\tau_{t-1} = j') \\ &\quad = \begin{cases} \dfrac{1 - G(t - j')}{1 - G(t - 1 - j')}, & \text{if } j = j' \\ \dfrac{G(t - j') - G(t - 1 - j')}{1 - G(t - 1 - j')}, & \text{if } j = t - 1 \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{2}$$

where $0 \le j \le t - 1$, $G(\cdot)$ is the cumulative distribution function for the segment duration. The geometric distribution is the most popular and simplest one for the segment duration in sequential change-point inferences [22], [25]. Other priors include Poisson distribution and gamma distribution, which are often used in speech segmentation [26]. As observed in our study, the detection is not very sensitive to the distribution type. Therefore, we select the geometric distribution for the segment duration for the purpose of simplicity. It is easy to show that when a geometric distribution with parameter p_0 is assumed for the segment duration δ, i.e., $P(\delta = l) = (1 - p_0)^{l-1} p_0$, the prior transition probability is simply

$$P(\tau_t = j|\tau_{t-1} = j') = \begin{cases} 1 - p_0, & \text{if } j = j' \\ p_0, & \text{if } j = t - 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

As we can see, it specifies a constant prior Markov transition probability for the latest change-point.

For the changing parameters $\boldsymbol{\Phi}_t = (\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$, a conjugate prior is specified as follows for all segments:

$$\begin{aligned} p(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) &= p(\boldsymbol{\Sigma}_t) p(\boldsymbol{\mu}_t|\boldsymbol{\Sigma}_t) \\ &= \text{InvWish}_p(\boldsymbol{\Psi}_0, \upsilon_0) N\left(\boldsymbol{\mu}_0, \frac{1}{\gamma_0}\boldsymbol{\Sigma}_t\right) \end{aligned} \tag{4}$$

where $\boldsymbol{\Sigma}_t \sim \text{InvWish}_p(\boldsymbol{\Psi}_0, \upsilon_0)$ is a p-dimensional Inverse-Wishart distribution with degrees of freedom υ_0 and scale matrix $\boldsymbol{\Psi}_0$, and $\boldsymbol{\mu}_t|\boldsymbol{\Sigma}_t \sim N(\boldsymbol{\mu}_0, (1/\gamma_0)\boldsymbol{\Sigma}_t)$ is a p-dimensional normal distribution. In addition, to facilitate online Bayesian updating, we assume that the changing parameters are independent across different segments. In Section II-B, we will show that such prior is a conjugate prior and the posterior

distribution of both the latest change-point and changing parameters, i.e., $p(\tau_t|\boldsymbol{X}_{1:t})$ and $p(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t|\boldsymbol{X}_{1:t}, \tau_t)$, where $\boldsymbol{X}_{1:t} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_t\}$, are analytically tractable.

B. Sequential Bayesian Change-Point Detection and Steady-State Detection

As mentioned in Section II, the duration of the current segment is a critical parameter to determine if the system is steady. Therefore, it is essential to calculate the posterior distribution of the segmental duration or equivalently the latest change-point sequentially. In this section, the exact posterior probability mass function (PMF) of the latest change-points and the posterior density functions of the changing parameters will be derived.

The posterior distribution can be expressed as

$$\begin{aligned} P(\tau_{t+1} = j|\boldsymbol{X}_{1:t+1}) &\propto p(\tau_{t+1} = j, \boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t}) \\ &= P(\tau_{t+1} = j|\boldsymbol{X}_{1:t})p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t}, \tau_{t+1} = j) \end{aligned} \tag{5}$$

where $j = 0, 1, \ldots, t$ is the observation index. The predictive PMF of the latest change-point in (5) can be calculated by

$$P(\tau_{t+1} = j|\boldsymbol{X}_{1:t}) = \sum_{i=0}^{\min(j,t-1)} P(\tau_{t+1} = j|\tau_i = i)P(\tau_t = i|\boldsymbol{X}_{1:t}). \tag{6}$$

Based on (3), (6) can be further simplified as

$$P(\tau_{t+1} = j|\boldsymbol{X}_{1:t}) = \begin{cases} p_0, & j = t \\ P(\tau_t = j|\boldsymbol{X}_{1:t})(1-p_0), & j \le t-1 \end{cases} \tag{7}$$

The predictive density function $p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t}, \tau_{t+1} = j)$ in (5) can be expressed as

$$p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t}, \tau_{t+1} = j) = \begin{cases} p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{j+1:t}, \tau_{t+1} = j), & j \le t-1 \\ p(\boldsymbol{X}_{t+1}), & j = t. \end{cases} \tag{8}$$

Therefore, (5), (7), and (8) can be summarized as

$$P(\tau_{t+1} = j|\boldsymbol{X}_{1:t+1}) \propto \begin{cases} p_0 p(\boldsymbol{X}_{t+1}), & j = t \\ (1-p_0)P(\tau_t = j|\boldsymbol{X}_{1:t}) \\ \quad \times p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{j+1:t}, \tau_{t+1} = j), & j \le t-1. \end{cases} \tag{9}$$

Let $P_j^{(t+1)} = P(\tau_{t+1} = j|\boldsymbol{X}_{1:t+1})$ and $p_{j+1,t} = p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{j+1:t}, \tau_{t+1} = j)$, and then, (5), (7), and (8) can be summarized using a state transition equation (10), as shown at the bottom of this page.

The matrix in (10) can be considered as a posterior transition matrix. It is the only term involving the newest observation X_{t+1} and, thus, is essential in updating the posterior PMF of the latest change-point. From (10), we can see that if (8), or equivalently the posterior transition matrix, is tractable, the posterior PMF can be recursively calculated based on the posterior PMF obtained at the previous time step. Therefore, the calculation of (8) is critical for sequential change-point detection.

To get the analytical form of $p(X_{t+1}|\boldsymbol{X}_{1:t}, \tau_{t+1} = j)$, we first derive the posterior $p(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t|\boldsymbol{X}_{1:t}, \tau_t)$, which is provided in Theorem 1.

Theorem 1: Suppose the joint prior for $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ is specified as $\boldsymbol{\Sigma}_t \sim InvWish_{\mathrm{p}}(\boldsymbol{\Psi}_0, \upsilon_0)$ and $\boldsymbol{\mu}_t|\boldsymbol{\Sigma}_t \sim N(\boldsymbol{\mu}_0, (1/\gamma_0)\boldsymbol{\Sigma}_t)$, then, the posterior $p(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t|\boldsymbol{X}_{1:t}, \tau_t)$ can be derived as

$$\begin{aligned} (\boldsymbol{\Sigma}_t|\boldsymbol{X}_{1:t}, \tau_t) &\sim \mathrm{InvWish}_{\mathrm{p}}\left(\boldsymbol{\Psi}^*_{\tau_t+1,t}, \upsilon^*_{\tau_t+1,t}\right) \\ (\boldsymbol{\mu}_t|\boldsymbol{\Sigma}_t, \boldsymbol{X}_{1:t}, \tau_t) &\sim N\left(\boldsymbol{\mu}^*_{\tau_t+1,t}, \boldsymbol{\Sigma}_t/\gamma^*_{\tau_t+1,t}\right) \end{aligned} \tag{11}$$

where

$$\begin{aligned} \upsilon^*_{\tau_t+1,t} &= (t-\tau_t) + \upsilon_0, \quad \gamma^*_{\tau_t+1,t} = \gamma_0 + (t-\tau_t) \\ \boldsymbol{\mu}^*_{\tau_t+1,t} &= \frac{(t-\tau_t)\bar{X}_{\tau_t+1,t} + \gamma_0\boldsymbol{\mu}_0}{\gamma_0+(t-\tau_t)} \\ \boldsymbol{\Psi}^*_{\tau_t+1,t} &= \boldsymbol{\Psi}_0 + (t-\tau_t)S_{\tau_t+1,t} \\ &\quad + \frac{(t-\tau_t)\gamma_0}{(t-\tau_t)+\gamma_0}(\bar{X}_{\tau_t+1,t} - \boldsymbol{\mu}_0)(\bar{X}_{\tau_t+1,t} - \boldsymbol{\mu}_0)' \end{aligned} \tag{12}$$

where $\bar{X}_{\tau_t+1,t}$ and $S_{\tau_t+1,t}$ are the mean and variance of the observations $\boldsymbol{X}_{\tau_t+1:t}$ calculated as

$$\bar{X}_{\tau_t+1,t} = \frac{1}{t-\tau_t}\sum_{i=\tau_t+1}^{t} \boldsymbol{X}_i,$$

$$S_{\tau_t+1,t} = \frac{1}{t-\tau_t}\sum_{i=\tau_t+1}^{t} (\boldsymbol{X}_i - \bar{X}_{\tau_t+1,t})(\boldsymbol{X}_i - \bar{X}_{\tau_t+1,t})'.$$

The proof is given in Appendix A. Based on Theorem 1, the predictive density $p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t}, \tau_{t+1} = j)$ can be derived as follows.

Theorem 2: For $\tau_{t+1} < t$

$$(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t}, \tau_{t+1}) \sim t\left(d^*_{\tau_{t+1}+1,t}, \boldsymbol{\mu}^*_{\tau_{t+1}+1,t}, \frac{\left(\gamma^*_{\tau_{t+1}+1,t}+1\right)\boldsymbol{\Psi}^*_{\tau_{t+1}+1,t}}{\gamma^*_{\tau_{t+1}+1,t}d^*_{\tau_{t+1}+1,t}}\right). \tag{13}$$

For $\tau_{t+1} = t$

$$(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t}, \tau_{t+1} = t) \sim t\left(\upsilon_0 - p + 1, \boldsymbol{\mu}_0, \frac{(\gamma_0+1)\boldsymbol{\Psi}_0}{\gamma_0(\upsilon_0 - p + 1)}\right) \tag{14}$$

$$\left[P_0^{(t+1)}, P_1^{(t+1)}, \ldots, P_t^{(t+1)}\right] \propto \left[P_0^{(t)}, P_1^{(t)}, \ldots, P_{t-1}^{(t)}\right] \times \begin{bmatrix} (1-p_0)p_{1,t} & 0 & \cdots & 0 & p_0 \cdot p(\boldsymbol{X}_{t+1}) \\ 0 & (1-p_0)p_{2,t} & \cdots & 0 & p_0 \cdot p(\boldsymbol{X}_{t+1}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & (1-p_0)p_{t,t} & p_0 \cdot p(\boldsymbol{X}_{t+1}) \end{bmatrix} \tag{10}$$

where $d^*_{\tau_{t+1}+1,t} = v^*_{\tau_{t+1}+1,t} - p + 1$ is the degree of freedom, and the other two arguments are the mean and shape matrix of the p-dimensional multivariate t distribution, respectively. The proof is provided in Appendix B.

After the posterior PMF of the latest change-point is updated, the distribution of the duration of the current segment can be easily obtained to test if the multivariate signal is steady. Specifically, we define a probability index P_t, which is the posterior probability of the current segment being longer than a threshold L_0

$$P_t = P(t - \tau_t \geq L_0 | \boldsymbol{X}_{1:t}) = P(\tau_t \leq t - L_0 | \boldsymbol{X}_{1:t}) = \sum_{i=0}^{t-L_0} P_i^{(t)}. \tag{15}$$

Once P_t is larger than a threshold α, the signal is claimed to be steady. It is intuitive that L_0 directly influences the detection timeliness and false alarm rate (FAR) or misclassification rate, i.e., the probability of signaling a steady-state alarm in transient state. It may be determined based on engineering knowledge or process requirement. However, in most of the practical applications, the steady state needs to be detected as early as possible, yet without causing the FAR to exceed a certain level. Therefore, L_0 can be treated as a tuning parameter to make tradeoff between the detection delay and FAR. For α, since the probability index often increases rapidly to a value close to 1 (see Section IV-A for details), we simply set it to 0.9 and do not treat it as a tuning parameter.

C. Controlling the Computational Cost

From Section II, we know that the posterior distribution of the latest change-point can be calculated analytically. However, the computational and memory cost of each time step increase almost linearly with time t, as can be seen from (10). At time t, we need to calculate the posterior PMF $P(\tau_t = j | \boldsymbol{X}_{1:t})$ at t positions, i.e., $j = 0, 1, \ldots, t-1$. For a long multivariate signal, the computational cost may become very prohibitive for online applications and, thus, needs to be controlled.

As observed in applications, the posterior PMF often concentrates around a small region and is almost zero at all other positions, especially those far before the latest change-point. Therefore, a natural way to control the computational cost is to approximate the posterior PMF using a fixed-support-size strategy, where a fixed number of positions with high probabilities are selected to calculate the posterior PMF and set the posterior PMF at other locations to 0. Specifically, suppose the support size is m, then, at time $t \geq m+1$, $P(\tau_t = j | \boldsymbol{X}_{1:t})$ is calculated at the m positions selected at the previous time step along with the position $j = t-1$. Therefore, there are in total $m+1$ positions to update the posterior PMF $P(\tau_t = j | \boldsymbol{X}_{1:t})$. After the $m+1$ probabilities are calculated, we randomly select m positions using weighted sampling without replacement to approximate $P(\tau_t = j | \boldsymbol{X}_{1:t})$. The weight for each location in the random sampling is simply its posterior PMF. Note that from (9), we can see that if $P(\tau_t = j | \boldsymbol{X}_{1:t}) = 0$, then $P(\tau_{t+1} = j | \boldsymbol{X}_{1:t+1}) = 0$, therefore, we only need to calculate the posterior PMF at the m positions selected at the previous time step and the new position $j = t-1$. Using this fixed-support-size strategy, the computational cost can be effectively controlled and balanced without influencing much of the detection accuracy.

D. Hyperparameter Selection

The choice of hyperparameters is often crucial in Bayesian data analysis when the sample size is limited. In our online application, the change-point needs to be detected in a timely manner, e.g., detecting the occurrence of a new change-point with only a few observations in the new segment, yet without resulting in overfitting or excessive change-points. Therefore, the hyperparameters need to be selected appropriately. In Bayesian inference, if a sufficient amount of historical data is available, informative priors are more preferable and could be estimated through these data. However, in many applications, historical data are very limited. In addition, in our case, to simplify the problem, we assume that all segments of different characteristics (in terms of duration, noise, and amplitude) are independent and share the same hyperparameters. As a result, it may be unrealistic to obtain a set of hyperparameters that is informative for all segments. In this section, we provide some guidelines and heuristics for hyperparameter selection.

Recall that the priors are $\boldsymbol{\Sigma}_t \sim \text{InvWish}_p(\boldsymbol{\Psi}_0, v_0)$, $\boldsymbol{\mu}_t | \boldsymbol{\Sigma}_t \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_t/\gamma_0)$, and the prior transition probability given in (3). Therefore, the hyperparameters include p_0, v_0, $\boldsymbol{\Psi}_0$, γ_0, and $\boldsymbol{\mu}_0$. Similar to the proof of Theorems 1 and 2, we can get the prior distribution of $\boldsymbol{\mu}_t$ by integrating out $\boldsymbol{\Sigma}_t$ as

$$\boldsymbol{\mu}_t \sim t(v_0 - p + 1, \boldsymbol{\mu_0}, \boldsymbol{\Psi}_0/[(v_0 - p + 1)\gamma_0]). \tag{16}$$

As observed in the numerical and real case studies, the detection results are not sensitive to the prior transition probability p_0. Any values in the interval [0.05, 0.2] works quite well. For the covariance prior $\text{InvWish}_p(\boldsymbol{\Psi}_0, v_0)$, the mean value is $E(\boldsymbol{\Sigma}_t) = \boldsymbol{\Psi}_0/(v_0 - p - 1)$ for $v_0 > p+1$. Based on the mean and covariance of $\boldsymbol{\Sigma}_t$ [27], we can see that v_0 directly controls the noise level. The larger the value is, the smaller the noise level of the prior, and, thus, the more sensitive the change-point detection will be or the more change-points it will result in. In other words, if the prior noise level is much larger than the actual one, the algorithm may not be able to detect the mean-shift timely, as the shift is masked by the large noise specified by the prior. On the other hand, if the prior noise level is too low, an overfitting issue may occur, i.e., too many change-points are produced. To select $\boldsymbol{\Psi}_0$ and v_0 appropriately, several scenarios are considered.

1) The noise covariance is constant in the whole process and some historical data are available. We could calculate the sample covariance matrix $\boldsymbol{S}$ using the steady-state data and then select a very large v_0 and set $\boldsymbol{\Psi}_0 = v_0 \boldsymbol{S}$. Based on Theorem 1, as $v_0 \to \infty$, $(\boldsymbol{\Sigma}_t | \boldsymbol{X}_{1:t}, \tau_t) \to \boldsymbol{S}$. Therefore, the problem degenerates to a piecewise constant model with fixed covariance $\boldsymbol{\Sigma}_t$ (only mean-shift), which could significantly reduce the uncertainty and thus improve the detection accuracy.

2) No prior information is available but the noise level is roughly known (e.g., within certain range). In this case, for simplicity, we can roughly set $\boldsymbol{\Psi}_0 = \boldsymbol{I}_p$ and select υ_0 accordingly to match the noise level.
3) The prior information is not available, yet we want our algorithm to be robust enough to handle signals with significantly different noise levels. In such case, selecting an appropriate set of hyperparameters is not easy. In addition, using a single set of hyperparameters may be too restrictive and cannot handle all signals. To solve this problem, we propose to use an adaptive prior approach, where the covariance prior is dynamically updated or learned from the data in the monitoring process. More specifically, based on (11) we can get

$$E(\boldsymbol{\Sigma}_t|\boldsymbol{X}_{1:t}) = \sum_{j=0}^{t-1} E(\boldsymbol{\Sigma}_t|\boldsymbol{X}_{1:t}, \tau_t = j)P(\tau_t = j|\boldsymbol{X}_{1:t}) = \sum_{j=0}^{t-1} \frac{\boldsymbol{\Psi}^*_{\tau_t+1,t}}{\upsilon^*_{\tau_t+1,t} - p + 1} P(\tau_t = j|\boldsymbol{X}_{1:t}). \quad (17)$$

At time step $t+1$, we set $\boldsymbol{\Sigma}_{t+1} \sim InvWish_p(\boldsymbol{\Psi}_0^{(t+1)}, \upsilon_0)$ where

$$\boldsymbol{\Psi}_0^{(t+1)} = (\upsilon_0 - p - 1)E(\boldsymbol{\Sigma}_t|\boldsymbol{X}_{1:t}). \quad (18)$$

It is easy to show that $E(\boldsymbol{\Sigma}_{t+1}) = E(\boldsymbol{\Sigma}_t|\boldsymbol{X}_{1:t})$, which is often more informative than arbitrarily specified priors. This strategy is particularly effective for signals with only mean shift.

For the mean prior $\boldsymbol{\mu}_t|\boldsymbol{\Sigma}_t \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_t/\gamma_0)$ or (16), since different segments may have different means along the multivariate trajectories, a noninformative prior or a "flat" prior is recommended to reduce the influence of priors and let the data "speak" for themselves. To make the prior noninformative, we could roughly select a $\boldsymbol{\mu}_0$ (e.g., 0) based on the order of the signal magnitude and then select a very small positive value for γ_0. From (11) we can see that as $\gamma_0 \to 0$, $\boldsymbol{\mu}^*_{\tau_t+1,t} \to \bar{X}_{\tau_t+1,t}$ and $(\boldsymbol{\mu}_t|\boldsymbol{\Sigma}_t, \boldsymbol{X}_{1:t}, \tau_t) \sim N(\bar{X}_{\tau_t+1,t}, \boldsymbol{\Sigma}_t/(t-\tau_t))$, which does not involve $\boldsymbol{\mu}_0$.

IV. Numerical Studies for Illustration and Comparison

In this section, we use numerical studies to illustrate the sequential change-point and steady-state detection process, and compare our method with several existing approaches. To simulate signals with initial bias in the comparison, we use four types of bias functions as signal means, namely, the linear, quadratic, exponential, and oscillating functions, which are commonly used in testing initial bias elimination heuristics [10]. The bias functions and their shapes are shown in Table I.

TABLE I
Bias Functions and Their Shapes

Bias Type	Function Form	Shape
Linear	$x(i) = \begin{cases} \frac{i-1}{T_0} H, & i = 1, \ldots, T_0 \\ H, & i = T_0+1, \ldots, T \end{cases}$	
Quad.	$x(i) = \begin{cases} H\left[1 - \frac{(i-T_0-1)^2}{T_0^2}\right], & i = 1, \ldots, T_0 \\ H, & i = T_0+1, \ldots, T \end{cases}$	
Exp.	$x(i) = \begin{cases} H\left[1 - \frac{(i-T_0-1)^2}{T_0^2}\right], & i = 1, \ldots, T_0 \\ H, & i = T_0+1, \ldots, T \end{cases}$	
Osc.	$x(i) = \begin{cases} H\frac{T_0-i+1}{T_0}\sin\left(\frac{\pi i}{f}\right), & i = 1, \ldots, T_0 \\ 0, & i = T_0+1, \ldots, T \end{cases}$	

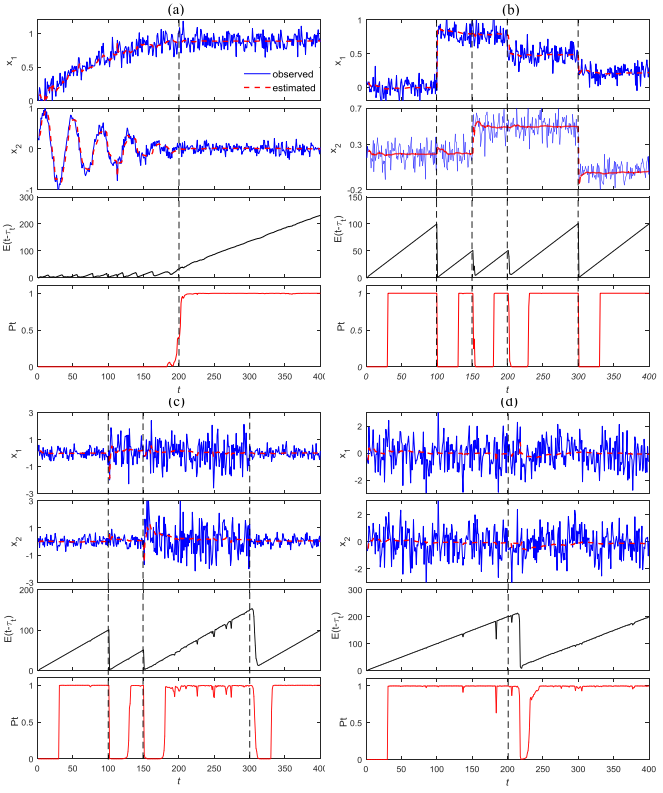


Fig. 2. Illustration of the sequential change-point detection and steady-state detection process. (a) Continuous mean and constant covariance. (b) Abrupt mean shift and constant covariance. (c) Constant mean and abrupt variance change. (d) Constant mean and abrupt correlation change.

A. Illustration

To illustrate the detection process and also show its robustness, we use four types of signals with different characteristics in terms of the change of mean and covariance matrix: 1) continuous mean and constant covariance; 2) abrupt mean shift and constant covariance; 3) constant mean and abrupt variance change; and 4) constant mean and abrupt correlation change, as shown in Fig. 2.

For display convenience, we only consider bivariate signals (i.e., $p = 2$) in the illustration. For the signal with continuous mean and constant covariance [Fig. 2(a)], the first dimension x_1 is an exponential signal, while the second dimension x_2 is an oscillating signal. The signal parameter is set as $H = 1, T_0 = 200, f = 30$, and $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_2$ where $\sigma = 0.1$. For the signal with abrupt mean shift and constant covariance [Fig. 2(b)], the covariance is $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_2$, the mean for

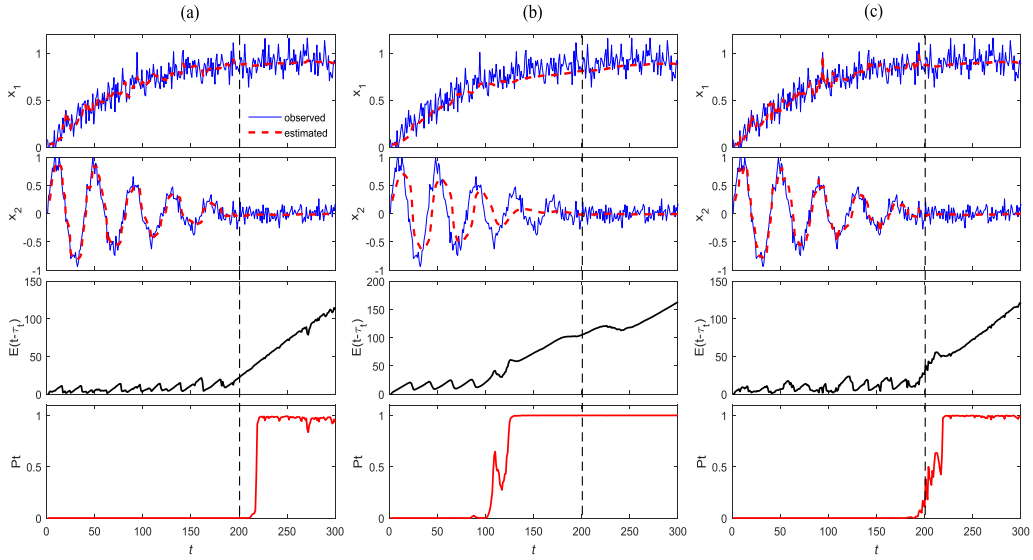


Fig. 3. Performance of adaptive prior in the steady-state detection process. (a) $\boldsymbol{\Psi}_0 = \boldsymbol{I}_2, \upsilon_0 = 100$, no dynamic updating. (b) $\boldsymbol{\Psi}_0 = \boldsymbol{I}_2, \upsilon_0 = 4$, no dynamic updating. (c) $\boldsymbol{\Psi}_0 = \boldsymbol{I}_2, \upsilon_0 = 4$, dynamic updating.

x_1 is set to 0, 0.8, 0.5, and 0.2 within the time intervals (0, 100), (100, 200), (200, 300), and (300, 400), respectively, and the mean for x_2 is set to 0.2, 0.5, and 0 within (0, 150), (150, 300), and (300, 400), respectively. For the signal with constant mean and abrupt variance change [Fig. 2(c)], $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2)$, where $\sigma_1 = 0.3, 1$, and 0.3 in the time intervals (0, 100), (100, 300), and (300, 400), respectively, and $\sigma_2 = 0.3, 1$, and 0.3 in the time intervals (0, 150), (150, 300), and (300, 400), respectively. For the case with constant mean and abrupt correlation change [Fig. 2(d)], $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, where $\sigma^2 = 1, \rho = 0.9$ for $t \in (0, 200)$ and $\rho = 0.6$ for $t \in (200, 400)$. The hyperparamters are set to $\boldsymbol{\mu}_0 = 0, \boldsymbol{\Psi}_0 = \boldsymbol{I}_2, p_0 = 0.1$, and $\gamma_0 = 0.01$ for all cases. For the hyperparameter υ_0, which is critical to control the prior noise level, we set it to 100 for Fig. 2(a) and (b), and 2 for Fig. 2(c) and (d). The duration threshold $L_0 = 30$ and probability threshold $\alpha = 0.9$.

In Fig. 2, the vertical dashed lines denote the true steady-state time. The dashed lines among the observations are the estimated means $E(\boldsymbol{\mu}_t|\boldsymbol{X}_{1:t})$ for the posterior distribution $\boldsymbol{\mu}_t|\boldsymbol{X}_{1:t}$. Similar to (16), we can prove that

$$(\boldsymbol{\mu}_t|\boldsymbol{X}_{1:t}, \tau_t) \sim t\left(d^*_{\tau_t+1,t}, \boldsymbol{\mu}^*_{\tau_t+1,t}, \frac{\boldsymbol{\Psi}^*_{\tau_t+1,t}}{\gamma^*_{\tau_t+1,t} d^*_{\tau_t+1,t}}\right).$$

Therefore

$$E(\boldsymbol{\mu}_t|\boldsymbol{X}_{1:t}) = \sum_{\tau_t=0}^{\tau_t=t-1} \boldsymbol{\mu}^*_{\tau_t+1,t} P(\tau_t|\boldsymbol{X}_{1:t}).$$

Clearly, the estimated means are very close to true values, indicating that the proposed method can effectively fit the signal sequentially through Bayesian inference. The third row of Fig. 2(a)–(d) shows the expected duration of the current segment, i.e., $E(t - \tau_t)$, which is used to demonstrate the change-point detection. The sharp decrease in the duration indicates a newly detected change-point. As we can see, the detection is very accurate and timely. The last row of Fig. 2(a)–(d) shows the probability index for the steady-state detection. Recall that the probability index is defined as the probability of the duration of current segment larger than that of the threshold L_0. We can see that the index often increases rapidly from a near-zero value to a value close to 1. Therefore, we simply select the probability threshold $\alpha = 0.9$ and do not treat it as a tuning parameter.

To show the effectiveness of the adaptive prior with dynamic updating strategy, we choose a signal with the same parameters as Fig. 2(a). As shown in Fig. 3, three cases are considered: 1) covariance prior with an appropriate noise level and without dynamic updating; 2) covariance prior with an excessively large noise level and without dynamic updating; and 3) covariance prior with an excessively large noise level but with dynamic updating. From Fig. 3(b), we can clearly see that if the prior noise level is too high, the algorithm is not able to timely detect the change-points, resulting in poor model fitting and steady-state detection. However, as shown in Fig. 3(c), if we use dynamic updating strategy to "correct" the prior, the model fitting and steady-state detection become much more accurate, even if the initial prior is specified inappropriately.

B. Performance Comparison With Other Methods

In this section, the performance of the proposed method is evaluated and compared with existing methods. In SPC area, two types of performance measures are often used, the α-error and the β-error (or detection delay). Usually, the α-error is specified at a desired level (e.g., 0.05) and the corresponding β-error is used as an evaluation metric to compare different control charts. However, this comparison scheme is not appropriate for the steady-state detection, as the α-error does not make any sense for non-i.i.d. (independent and identically distributed) samples in the transient period. Instead, another evaluation metric, namely, the FAR may be used, which is defined in our case as the probability of signaling a steady-state alarm in the transient period. Nevertheless, this metric still has shortcoming, in that, it does not capture the

TABLE II
FOUR NOISE TYPES

Type	Equation	Parameter
AR(0)	$\boldsymbol{\psi}_t = \boldsymbol{\epsilon}_t$	$\boldsymbol{\epsilon}_t \sim N(0, \sigma^2 \boldsymbol{I}_p)$
AR(1)	$\boldsymbol{\psi}_t = \phi_1 \boldsymbol{\psi}_{t-1} + \boldsymbol{\epsilon}_t$	$\boldsymbol{\epsilon}_t \sim N(0, \sigma^2 \boldsymbol{I}_p)$, $\phi_1 = 0.4$
AR(2)	$\boldsymbol{\psi}_t = \phi_2 \boldsymbol{\psi}_{t-1} + \phi_3 \boldsymbol{\psi}_{t-2} + \boldsymbol{\epsilon}_t$	$\boldsymbol{\epsilon}_t \sim N(0, \sigma^2 \boldsymbol{I}_p)$, $\phi_2 = -0.25, \phi_3 = 0.5$
CR	$\boldsymbol{\psi}_t = \boldsymbol{\epsilon}_t$	$\boldsymbol{\epsilon}_t \sim N(0, \sigma^2 \boldsymbol{C}_r)$

closeness of the false alarm time to the true steady-state time. In fact, the closeness of the alarm time in the transient period to the true steady state is very important as it directly reflects the amount of initial bias undetected or the remaining time needed to reach steady state. Naturally, we could use the closeness measure to evaluate the performance. Considering the fact that the detection delay is often better than false alarm with the same closeness, we use another metric, the weighted standard error (WSE) [8], defined as

$$\text{WSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} w(\hat{T}_i)(\hat{T}_i - T_{i0})^2} \tag{19}$$

where $\hat{T}_i$ is the detected time, T_{i0} is the true steady-state time, N is the total number of multivariate signals, and $w(\cdot)$ is the penalty weight ratio of detection delay over false alarm given as

$$w(\hat{T}_i) = \begin{cases} w \in (0, 1], & \text{if } \hat{T}_i \geq T_{i0} \\ 1, & \text{otherwise.} \end{cases} \tag{20}$$

Note that if $w = 1$, only the closeness is considered in performance assessment. In the comparison, $p = 4$ is selected and each dimension is simulated by a bias function, which is randomly selected from Table I to cover various initial bias. To further diversify the initial bias severity, different H, T_0, and noise levels are specified. Specifically, $H = 1$ and 2, $T_0 = 200, 300$. The length of the signal is set as $T = 500$. To test the robustness of the algorithm under different noise types, four scenarios are considered: 1) no auto-correlation and no correlation among variables, denoted by AR(0); 2) first-order auto-correlation and no correlation, denoted by AR(1); 3) second-order auto-correlation and no correlation, denoted by AR(2); and 4) no auto-correlation and with correlation among variables, denoted by CR. The noise types and their parameters are shown in Table II.

For each type of signal, three noise levels are considered. For AR(0) and CR, $\sigma = 0.06, 0.1, 0.14$. For CR, the correlation matrix $\boldsymbol{C}_r$ is randomly generated through vines method [28]. In the simulation, each signal is replicated 100 times, so that a total of 1200 signals ($2H \times 2T_0 \times 3\sigma \times 100$) are generated for each of the four noise types.

The proposed sequential Bayesian partitioning (SBP) method is compared with three existing methods. The first method is the exponentially weighted moving average-based variance ratio test (VRT) [16], where each dimension is monitored separately using the well-known method by Cao and Rhinehart [6], and the steady state is claimed once all dimensions reach steady state. The second method is the steady state detection (SSD) algorithm [29], which employs a moving window and tests if there is any nonstationary drift within that window. The third one is a wavelet transform (WT)-based method [18], where the status index for each dimension is combined using the Dempster's combination rule to form a global detection index.

The hyperparamters for SBP are set and fixed as $\boldsymbol{\mu}_0 = 0$, $\boldsymbol{\Psi}_0 = \boldsymbol{I}_2, \upsilon_0 = 100, p_0 = 0.1$, and $\gamma_0 = 0.01$ for all cases. The duration threshold L_0 is selected by optimizing the overall WSE under each noise type and weight w. For all the other three methods, the detection parameters are chosen by optimizing the overall WSE under each noise type and weight w. Note that in practical applications, the true steady-state times of the training data may be unknown or there may even not be sufficient training data. For the former case, some offline method could be used as a benchmark to estimate the steady-state times, and then, the estimated values can be used to evaluate WSE. For the latter case, we could use Monte Carlo simulation to generate a training database covering various initial bias conditions of different characteristics, e.g., noise level and changing rate to select an optimal L_0.

Fig. 4 shows the WSE and FAR of the four detection methods as functions of w under different noise types. It is worth noting that here the FAR is used only as an auxiliary metric to show the detection details. Clearly, the proposed SBP outperforms VRT, SSD, and WT methods significantly in terms of WSE. The FAR of SBP is also much lower than other three methods, indicating that if we reduce the FAR of all other methods to the same level as SBP, the WSE will become worse. For WT, the WSE does not change when w varies. The reason is that FAR is above 0.9 for all cases. Based on (20), the WSE will not change much when w varies. Note that the hyperparameters of the SBP are selected using only several trials under the guidelines of the hyperparameter selection in Section III-D. The performance could be further improved if these parameters are optimized.

Table III shows the detailed detection for each type of signal with noise AR(0) and penalty weight ratio $w \equiv 1$ (only consider closeness). Due to space limitation, the detailed results for other noise types and penalty weight ratios are not provided here. We can see that SBP is much more robust in handling signals of various noise levels and initial bias severity. Bear in mind that SSD and WT are moving window-based methods and, thus, are not robust. Too long a moving window may delay the detection, while too short a moving window may result in large FAR. As shown in Table III, all the three methods could not uniformly perform well across all types of signals. The proposed SBP method incorporates the sequential Bayesian inference scheme and, thus, could online "learn" the monitoring signal, which could significantly improve its robustness.

V. REAL CASE STUDIES

In this section, we apply the proposed method to two real cases to demonstrate its effectiveness: the Tennessee Eastman (TE) process [30] and a serial production line with perishable products [31].

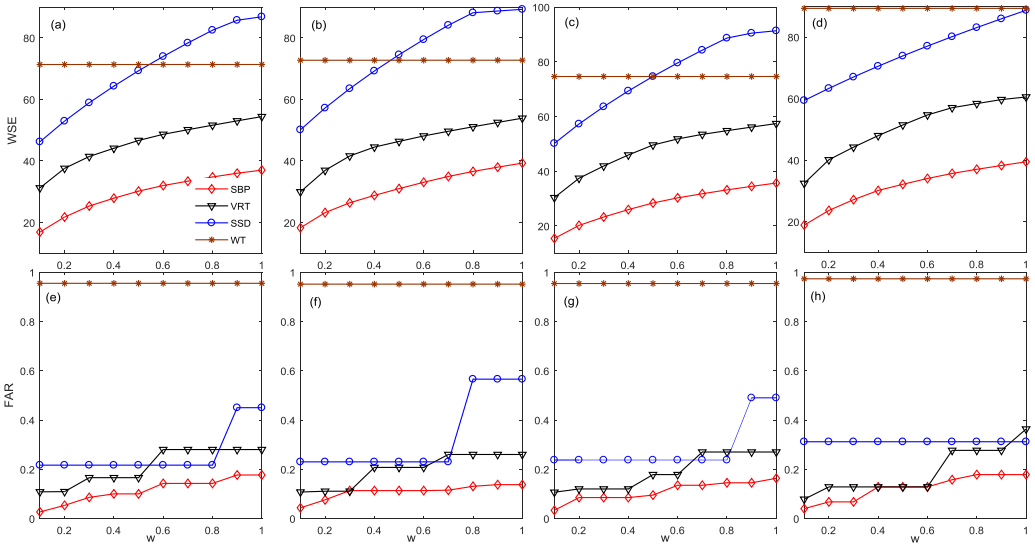


Fig. 4. WSE and FAR of SBP, VAR, SSD, and WT as functions of penalty weight ratio for different noise types. (a)–(e) AR(0). (b)–(f) AR(1). (e)–(g) AR(2). (d)–(h) CR.

TABLE III

COMPARISON OF SBP, VRT, SSD, AND WT FOR $w = 1$. THE DETECTION PARAMETERS ARE (1) SBP, $L_0 = 50$; (2) VRT, $\lambda_1 = 0.05$, $\lambda_2 = \lambda_3 = 0.2$, THRESHOLD = 0.8; (3) SSD, WINDOW SIZE = 24; AND (4) WT, $\Delta t = 20$

Signal			WSE($w = 1$)				FAR			
H	T_0	σ	SBP	VRT	SSD	WT	SBP	VRT	SSD	WT
1	200	0.06	27.3	47.8	66.3	38.2	0.03	0.01	0.24	0.91
		0.10	28.1	30.0	65.8	49.3	0.14	0.35	0.55	0.91
		0.14	26.6	35.1	70.1	59.8	0.15	0.69	0.68	0.94
	300	0.06	39.2	36.0	83.0	73.8	0.42	0.17	0.57	0.97
		0.10	35.9	74.8	127.5	105.9	0.37	0.75	0.78	1.00
		0.14	69.7	97.9	155.4	135.3	0.64	0.78	0.84	1.00
2	200	0.06	36.3	66.8	60.4	31.6	0	0.00	0.00	0.97
		0.10	36.9	52.4	63.3	36.6	0	0.00	0.11	0.92
		0.14	32.6	39.6	57.3	39.0	0.02	0.02	0.29	0.94
	300	0.06	28.3	51.8	70.2	61.1	0.1	0.00	0.13	0.99
		0.10	32.9	40.0	73.2	72.0	0.05	0.12	0.51	0.94
		0.14	29.8	39.8	91.2	76.6	0.2	0.47	0.70	0.97
Overall			37.0	54.4	86.8	71.4	0.18	0.28	0.45	0.96

A. Tennessee Eastman Process

The TE process is based on a simulation of a realistic chemical plant. It has been widely used as a benchmark process in the process monitoring community to test various fault detection, identification, diagnosis, and closed-loop control methodologies [32]. As shown in Fig. 5, the process consists of five major units: a reactor, condenser, compressor, separator, and stripper, and it contains eight chemical components: A, B, C, D, E, F, G, and H, where A, C, D, and E are the reactants, B is the inert gas, G and H are the products, and F is the byproduct. For the detailed process description, please refer to [30] and [32].

The process contains in total 53 measurement variables (see [30], [32] for details), out of which 41 are the process variables, i.e., XMEAS(1)–XMEAS(41), and 12 manipulated variables, i.e., XMV(1)–XMV(12). A total of 21 process faults are preprogrammed, i.e., IDV(1)–IDV(21), including A/C feed ratio step change, B composition step change, and D feed temperature step change. Here, we only consider Fault 1. When Fault 1 occurs at time step 160 (8 h), a step changed is induced for the A/C feed ratio, which results in a decrease in A feed in Stream 5 and control loop reacts to increase the A feed in Stream 1. After a certain amount of time, the A feed becomes steady in Stream 6. Fig. 6 shows the dynamic change of A feed in Stream 1 and Composition of A in Stream 6 once Fault 1 occurs.

To monitor the steady state of the whole system, we utilize all the 41 measurement variables. Since the historical data of all these variables under the normal operating condition are available or can be easily generated, we use these data to roughly estimate the hyperparameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\Psi}_0$. Suppose the mean and sample covariance of these normal data are $\bar{X}$ and $\boldsymbol{S}$, respectively. Then, we select $\boldsymbol{\mu}_0 = \bar{X}$, $\upsilon_0 = 1 \times 10^6$, and $\boldsymbol{\Psi}_0 = \upsilon_0 \boldsymbol{S}$. Other parameters are set as $p = 0.1$, $\gamma_0 = 1 \times 10^{-4}$, and $L_0 = 60$.

Fig. 7 shows the signal segmentation and steady-state detection results. Note that for space limitation, here, we only show the first six process variables, from XMEAS(1)–XMEAS(6). Fig. 7(g) shows the histogram of the simulated change-points. They are simulated in this way: 1) randomly draw a sample τ_T from $P(\tau_T | X_{1:T})$ and then randomly draw τ_{τ_T} from $P(\tau_{\tau_T} | X_{1:\tau_T})$, continue this process until we reach the beginning of the signal and 2) repeat the whole process 1000 times.

As we can see, the proposed method can accurately detect the onset of Fault 1 and can effectively partition the whole multivariate system into transient period and steady-state period. The detected onset time of Fault 1 is 166, which is very

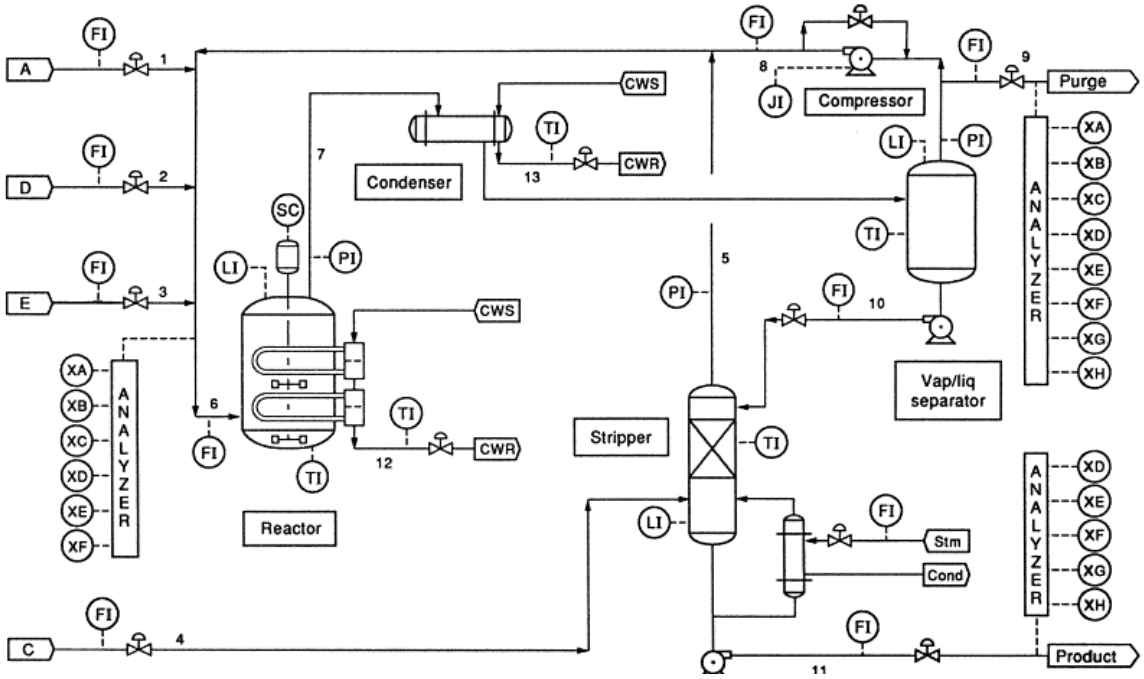


Fig. 5. TE process.

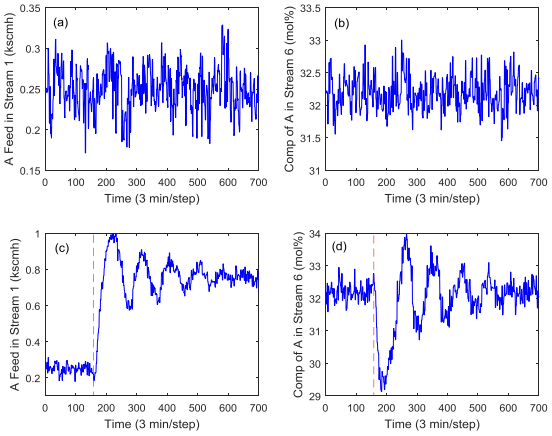


Fig. 6. Feed in Stream 1 and Composition of A in Stream 6 under normal operating condition and Fault 1. (a) Feed under normal operating condition [XMEAS(1)]. (b) Composition under normal operating condition [XMEAS(23)]. (c) Feed under Fault 1 [XMEAS(1)]. (d) Composition under Fault 1 [XMEAS(23)]. The vertical dashed lines denote the onset of Fault 1.

close to the true onset time 160. Since we know the true onset time, we use it to compare the detection accuracy between the proposed method and SSD, VRT, and WT methods. The optimal detection parameters that minimize the detection error are: 1) VRT: $\lambda_1 = 0.1, \lambda_2 = \lambda_3 = 0.2$, and threshold $= 1.7$; 2) SSD: window size $= 15$; and 3) WT: $\Delta t = 7$. The detection results are 139, 135, and 237 for VRT, SSD, and WT, respectively. We can see that the proposed SBP method is much more accurate. It is worth noting that in the other three methods, we need to build up to 41 monitoring charts, which are very time consuming, let alone the detection accuracy after fusing all detection results.

B. Serial Production Lines With Perishable Products

The perishable products refer to those having maximum allowable waiting time, exceeding which the item will be scrapped due to quality deterioration. Typical examples include yogurt and battery production [31]. Due to dynamic changes and frequent disruptions in the manufacturing process, and product perishability, the production system often operates partially or even entirely in the transient regime. After the production system is initiated, it needs some time (warm-up period) for the production to reach steady state, e.g., production rate and scrap rate. To facilitate process monitoring and real time control policy optimization, it is essential to detect when the production system reaches steady state.

Suppose there is a serial production with two Bernoulli reliability machines, m_1 and m_2, a finite buffer B_1 and perishable products, as shown in Fig. 8.

The performance measures that are of interest and are used to describe the system state include the following.

1) The production rate PR(t), which is the average number of parts producted by machine m_2 in the tth cycle.
2) The consumption rate CR(t), which is the average number of parts consumed by machine m_1 in the tth cycle.
3) Scrap rate SR(t), i.e., the expected number of scrapped parts in the tth cycle.
4) The work-in-process WIP(t), which is the average number of parts in buffer B_1 at the end of the tth cycle.

For the detailed description of the data set, please refer to [31] and [33].

The 4-D signal and detection results are shown in Fig. 9. All the other detection parameters are set using the same way as Fig. 7, except that $\boldsymbol{\mu}_0 = 0$ and $L_0 = 20$ are set here. To evaluate the detection accuracy, we use an offline method, namely, the adaptive minimal confidence region rule (AMCR), as a benchmark. AMCR determines the steady-state starting time by minimizing the confidence region of the mean estimate using all the observations since that time [33]. The detected

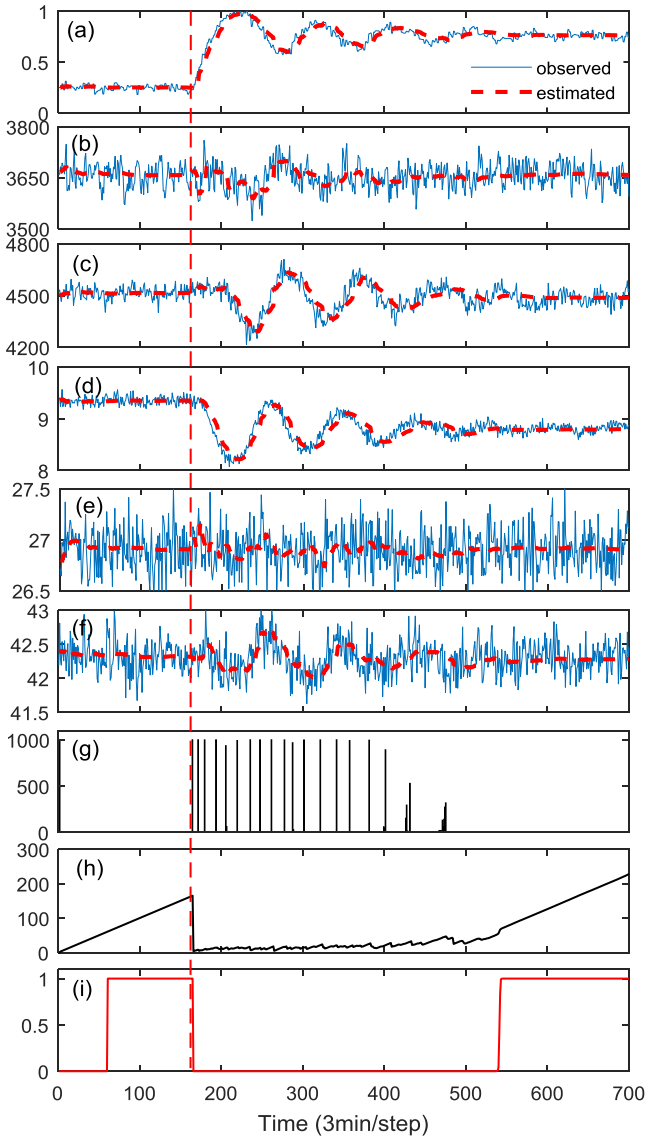


Fig. 7. Steady-state detection of the TE process under Fault 1 with 41 measurements (only the first 6 measures are shown here). (a)–(f) XMEAS(1)–XMEAS(6). (g) Histogram of change-points. (h) Mean duration of the current segment $E(t-\tau_t)$. (i) Probability index P_t. The vertical dashed lines denote the onset of Fault 1.

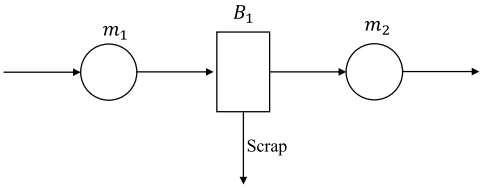


Fig. 8. Bernoulli line with perishable products.

steady-state starting time using the proposed online method is 37, which is very close to the AMCR detected time 35. In comparison, the VRT, SSD, and WT detected times are 38, 40, and 30, respectively. The corresponding optimal detection parameters are $\lambda_1 = \lambda_2 = \lambda_3 = 0.5$ and Thereshold $= 1$ for VRT, window size $= 5$ for SSD, and $\Delta t = 12$ for WT. We can see that the proposed SBP approach is still better than the other three methods. Note that the advantage of the proposed method

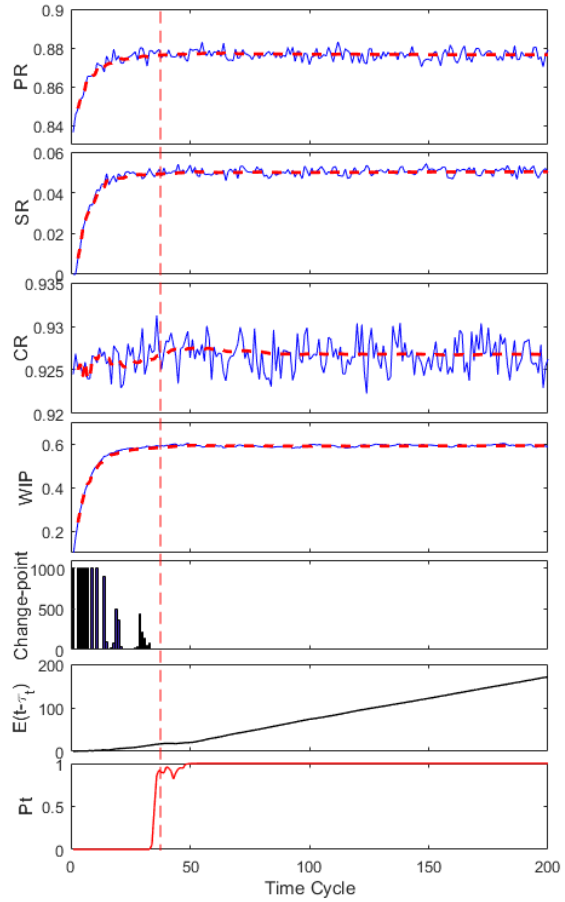


Fig. 9. Steady-state detection of the Bernoulli line of two machines. The histogram of the change-points, the mean duration of the current segment, and the probability index P_t, respectively (bottom). The vertical dashed line denotes the detected steady-state time.

is that it is much more robust in handling various signals of different characteristics using only one set of detection parameters. For one multivariate signal, the advantage may not be obvious, as other methods can always find a set of detection parameters that work well on that specific signal.

VI. Conclusion and Discussion

In this paper, an efficient online steady-state detection method has been developed for multivariate systems through a sequential Bayesian partitioning approach. In this approach, multivariate signals are modeled by piecewise constant models, where the mean and covariance are constant in each segment, and then, the duration of each segment is utilized to determine if the signal is steady. To facilitate online inference, a Bayesian formulation of the piecewise constant model is proposed. By using conjugate priors, it is found that the posterior distributions of the latest change-points can be calculated analytically through a recursive updating approach. Once the posterior probability of the duration is larger than a predefined threshold, the signal is considered steady. To control and balance the computational cost, a fixed-support-size strategy is proposed to approximate the posterior PMF of the latest change-points. The role and sensitivity of hyperparameters are discussed and several guidelines are provided to help select

hyperparameters appropriately. Thorough simulation and real case studies have demonstrated that the proposed method can timely detect the change-points and effectively partition the multivariate signal sequentially. The comparison results show that the proposed method is much more accurate and robust than existing methods in tackling signals of various characteristics.

On the other hand, we need to point out that although our approach can handle signals with mild autoregressive noise, due to its Gaussian noise assumption, it may not be able to detect the steady state accurately when the autocorrelation is very severe. In addition, in this paper, we assume that all the dimensions of the multivariate signal reach steady state at the same time. In practice, however, some dimensions may reach steady state earlier than the other ones, which will influence the detection performance. We will leave these problems in our future investigation. Last but not least, with the recent advances in sensing and information technology, we are more and more faced with the problem of monitoring up to hundreds or even thousands of process variables. In such cases, the proposed method may face severe challenges, e.g., prohibitive computational cost for online application and ill-posed covariance inversion. To overcome these challenges, data fusion and distributed monitoring techniques may be needed, which will be investigated in near future.

APPENDIX A
PROOF OF THEOREM 1

Based on the definition of change-point, $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} = \ldots = \boldsymbol{\mu}_{\tau_t+1}$, $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t-1} = \ldots = \boldsymbol{\Sigma}_{\tau_t+1}$

$$\begin{aligned} f(\boldsymbol{\mu}_t|\boldsymbol{\Sigma}_t, \boldsymbol{X}_{1:t}, \tau_t) &\propto f(\boldsymbol{\mu}_t|\boldsymbol{\Sigma}_t) f(\boldsymbol{X}_{\tau_t+1:t}|\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_{\mathrm{t}}, \tau_t) \\ &\propto \exp\left[\frac{-\gamma_0(\boldsymbol{\mu}_t - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\mu}_0)}{2}\right] \\ &\quad \times \exp\left[-\frac{\sum_{i=\tau_t+1}^{t}(\boldsymbol{X}_i - \boldsymbol{\mu}_t)'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_t)}{2}\right] \\ &\propto \exp\left[\frac{-\gamma^*_{\tau_t+1,t}(\boldsymbol{\mu}_t - \boldsymbol{\mu}^*_{\tau_t+1,t})'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\mu}^*_{\tau_t+1,t})}{2}\right] \end{aligned}$$

where

$$\begin{aligned} \gamma^*_{\tau_t+1,t} &= \gamma_0 + (t - \tau_t) \\ \boldsymbol{\mu}^*_{\tau_t+1,t} &= \frac{(t-\tau_t)\bar{X}_{\tau_t+1,t} + \gamma_0\boldsymbol{\mu}_0}{\gamma_0 + (t-\tau_t)}. \end{aligned}$$

Therefore

$$(\boldsymbol{\mu}_t|\boldsymbol{\Sigma}_t, \boldsymbol{X}_{1:t}, \tau_t) \sim N\left(\boldsymbol{\mu}^*_{\tau_t+1,t}, \frac{\boldsymbol{\Sigma}_t}{\gamma^*_{\tau_t+1,t}}\right).$$

For $(\boldsymbol{\Sigma}_t|\boldsymbol{X}_{1:t}, \tau_t)$

$$\begin{aligned} f(\boldsymbol{\Sigma}_{\mathrm{t}}|\boldsymbol{X}_{1:t}, \tau_t) &= \int f(\boldsymbol{\Sigma}_{\mathrm{t}}, \boldsymbol{\mu}_t|\boldsymbol{X}_{\tau_t+1:,t}, \tau_t) d\boldsymbol{\mu}_t \\ &\propto \int f(\boldsymbol{\Sigma}_{\mathrm{t}}, \boldsymbol{\mu}_t) f(\boldsymbol{X}_{\tau_t+1:,t}|\boldsymbol{\Sigma}_{\mathrm{t}}, \boldsymbol{\mu}_t, \tau_t) d\boldsymbol{\mu}_t \end{aligned}$$

where

$$\begin{aligned} f(\boldsymbol{\Sigma}_{\mathrm{t}}, \boldsymbol{\mu}_t) &\propto |\boldsymbol{\Sigma}_{\mathrm{t}}|^{-\frac{v_0+p+1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}_0\boldsymbol{\Sigma}_t^{-1}\right)\right] \left|\frac{1}{\gamma_0}\boldsymbol{\Sigma}_{\mathrm{t}}\right|^{-\frac{1}{2}} \\ &\quad \times \exp\left[\frac{-\gamma_0(\boldsymbol{\mu}_t - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\mu}_0)}{2}\right] \end{aligned}$$

and

$$\begin{aligned} &f(\boldsymbol{X}_{\tau_t+1:,t}|\boldsymbol{\Sigma}_{\mathrm{t}}, \boldsymbol{\mu}_t, \tau_t) \\ &\quad \propto |\boldsymbol{\Sigma}_t|^{-\frac{t-\tau_t}{2}} \exp\left[-\frac{\sum_{i=\tau_t+1}^{t}(\boldsymbol{X}_i - \boldsymbol{\mu}_t)'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_t)}{2}\right]. \end{aligned}$$

Therefore

$$\begin{aligned} &f(\boldsymbol{\Sigma}_{\mathrm{t}}|\boldsymbol{X}_{1:t}, \tau_t) \\ &\propto \int\int |\boldsymbol{\Sigma}_t|^{-\frac{v_0+p+1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}_0\boldsymbol{\Sigma}_t^{-1}\right)\right] |\boldsymbol{\Sigma}_t|^{-\frac{t-\tau_t+1}{2}} \\ &\quad \times \exp\left[-\frac{1}{2}\left((\boldsymbol{\mu}_t - \boldsymbol{\mu}^*_{\tau_t+1,t})'\left(\frac{\boldsymbol{\Sigma}_t}{\gamma_0 + t - \tau_t}\right)^{-1} \right.\right. \\ &\qquad \times (\boldsymbol{\mu}_t - \boldsymbol{\mu}^*_{\tau_t+1,t}) - (\gamma_0 + t - \tau_t)(\boldsymbol{\mu}^*_{\tau_t+1,t})'\boldsymbol{\Sigma}_t^{-1} \\ &\qquad \left.\left. \times \boldsymbol{\mu}^*_{\tau_t+1,t} + \gamma_0\mu_0'\boldsymbol{\Sigma}_t\mu_0 + \sum_{i=\tau_t+1}^{t}\boldsymbol{X}_i'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{X}_i\right)\right] d\boldsymbol{\mu}_t \\ &\propto |\boldsymbol{\Sigma}_t|^{-\frac{v_0+p+t-\tau_t+1}{2}} \\ &\quad \times \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}_0 + (t-\tau_t)S_{\tau_t+1,t} + \frac{(t-\tau_t)\gamma_0}{t-\tau_t+\gamma_0} \right.\right. \\ &\qquad \left.\left. \times (\bar{X}_{\tau_t+1,t} - \boldsymbol{\mu}_0)(\bar{X}_{\tau_t+1,t} - \boldsymbol{\mu}_0)'\right)\boldsymbol{\Sigma}_t^{-1}\right] \end{aligned}$$

where

$$\begin{aligned} \bar{X}_{\tau_t+1,t} &= \frac{1}{t-\tau_t}\sum_{i=\tau_t+1}^{t}\boldsymbol{X}_i, S_{\tau_t+1,t} \\ &= \frac{1}{t-\tau_t}\sum_{i=\tau_t+1}^{t}(\boldsymbol{X}_i - \bar{X}_{\tau_t+1,t})(\boldsymbol{X}_i - \bar{X}_{\tau_t+1,t})'. \end{aligned}$$

Therefore

$$(\boldsymbol{\Sigma}_t|\boldsymbol{X}_{1:t}, \tau_t) \sim \mathrm{InvWish_p}\left(\boldsymbol{\Psi}^*_{\tau_t+1,t}, v^*_{\tau_t+1,t}\right)$$

where

$$\begin{aligned} \boldsymbol{\Psi}^*_{\tau_t+1,t} &= \boldsymbol{\Psi}_0 + (t-\tau_t)S_{\tau_t+1,t} \\ &\quad + \frac{(t-\tau_t)\gamma_0}{t-\tau_t+\gamma_0}(\bar{X}_{\tau_t+1,t} - \boldsymbol{\mu}_0)(\bar{X}_{\tau_t+1,t} - \boldsymbol{\mu}_0)' \\ v^*_{\tau_t+1,t} &= (t-\tau_t) + v_0. \end{aligned}$$

APPENDIX B
PROOF OF THEOREM 2

If $\tau_{t+1} < t$, then, $\tau_{t+1} = \tau_t, \boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t = \ldots = \boldsymbol{\mu}_{\tau_{t+1}+1}, \boldsymbol{\Sigma}_{t+1} = \boldsymbol{\Sigma}_t = \ldots = \boldsymbol{\Sigma}_{\tau_{t+1}+1}$. For notational convenience, let $\boldsymbol{\mu} = \boldsymbol{\mu}_{t+1} = \ldots = \boldsymbol{\mu}_{\tau_{t+1}+1}, \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{t+1} = \boldsymbol{\Sigma}_t = \ldots = \boldsymbol{\Sigma}_{\tau_{t+1}+1}$.

$$
\begin{aligned}
&f(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t},\tau_{t+1})\\
&\quad=\int f(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t},\tau_{t+1},\boldsymbol{\Sigma})f(\boldsymbol{\Sigma}|\boldsymbol{X}_{1:t},\tau_{t+1})d\boldsymbol{\Sigma}\\
&\quad\propto\int\left|\left(\frac{1}{\gamma^*_{\tau_{t+1}+1,t}}+1\right)\boldsymbol{\Sigma}\right|^{-\frac{1}{2}}\times\exp\left[-\frac{\left(\boldsymbol{X}_{t+1}-\boldsymbol{\mu}^*_{\tau_{t+1}+1,t}\right)'\left[\left(\frac{1}{\gamma^*_{\tau_{t+1}+1,t}}+1\right)\boldsymbol{\Sigma}\right]^{-1}\left(\boldsymbol{X}_{t+1}-\boldsymbol{\mu}^*_{\tau_{t+1}+1,t}\right)}{2}\right]\\
&\qquad\times\left|\boldsymbol{\Psi}^*_{\tau_{t+1}+1,t}\right|^{\frac{v^*_{\tau_{t+1}+1,t}}{2}}|\boldsymbol{\Sigma}|^{-\frac{v^*_{\tau_{t+1}+1,t}+p+1}{2}}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}^*_{\tau_{t+1}+1,t}\boldsymbol{\Sigma}^{-1}\right)\right]d\boldsymbol{\Sigma}\\
&\quad\propto\int|\boldsymbol{\Sigma}|^{-\frac{v^*_{\tau_{t+1}+1,t}+p+2}{2}}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\left(\boldsymbol{\Psi}^*_{\tau_{t+1}+1,t}+\frac{\gamma^*_{\tau_t+1,t}}{\gamma^*_{\tau_{t+1}+1,t}+1}\left(\boldsymbol{X}_{t+1}-\boldsymbol{\mu}^*_{\tau_{t+1}+1,t}\right)\left(\boldsymbol{X}_{t+1}-\boldsymbol{\mu}^*_{\tau_{t+1}+1,t}\right)'\right)\boldsymbol{\Sigma}^{-1}\right)\right]d\boldsymbol{\Sigma}\\
&\quad\propto\left|\boldsymbol{\Psi}^*_{\tau_{t+1}+1,t}+\frac{\gamma^*_{\tau_t+1,t}}{\gamma^*_{\tau_{t+1}+1,t}+1}\left(\boldsymbol{X}_{t+1}-\boldsymbol{\mu}^*_{\tau_{t+1}+1,t}\right)\left(\boldsymbol{X}_{t+1}-\boldsymbol{\mu}^*_{\tau_{t+1}+1,t}\right)'\right|^{-\frac{v^*_{\tau_{t+1}+1,t}+1}{2}}.
\end{aligned}
$$

Based on Theorem 1

$$
\begin{aligned}
(\boldsymbol{\mu}|\boldsymbol{\Sigma},\boldsymbol{X}_{1:t},\tau_{t+1})&\sim N\left(\boldsymbol{\mu}^*_{\tau_t+1,t},\frac{\boldsymbol{\Sigma}}{\gamma^*_{\tau_{t+1}+1,t}}\right)\\
(\boldsymbol{\Sigma}|\boldsymbol{X}_{1:t},\tau_{t+1})&\sim\mathrm{InvWish}_{\mathrm{p}}\left(\boldsymbol{\Psi}^*_{\tau_{t+1}+1,t},v^*_{\tau_{t+1}+1,t}\right).
\end{aligned}
$$

Since

$$
(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t},\tau_{t+1},\boldsymbol{\Sigma})=(\boldsymbol{\mu}|\boldsymbol{\Sigma},\boldsymbol{X}_{1:t},\tau_{t+1})+N(0,\boldsymbol{\Sigma})
$$

it is easy to show that

$$
(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t},\tau_{t+1},\boldsymbol{\Sigma})\sim N\left(\boldsymbol{\mu}^*_{\tau_{t+1}+1,t},\left(\frac{1}{\gamma^*_{\tau_{t+1}+1,t}}+1\right)\boldsymbol{\Sigma}\right).
$$

Therefore, the equation shown at the top of this page is obtained.

According to the generalized matrix determinant lemma

$$
\begin{aligned}
\propto\Bigg|1+\frac{\left(\boldsymbol{X}_{t+1}-\boldsymbol{\mu}^*_{\tau_{t+1}+1,t}\right)'}{\left(v^*_{\tau_{t+1}+1,t}-p+1\right)}&\left(\frac{\left(\gamma^*_{\tau_{t+1}+1,t}+1\right)\boldsymbol{\Psi}^*_{\tau_{t+1}+1,t}}{\gamma^*_{\tau_{t+1}+1,t}\left(v^*_{\tau_{t+1}+1,t}-p+1\right)}\right)^{-1}\\
&\times\left(\boldsymbol{X}_{t+1}-\boldsymbol{\mu}^*_{\tau_{t+1}+1,t}\right)\Bigg|^{-\frac{\left(v^*_{\tau_{t+1}+1,t}-p+1\right)+p}{2}}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t},\tau_{t+1})\\
&\qquad\sim t\left(d^*_{\tau_{t+1}+1,t},\boldsymbol{\mu}^*_{\tau_{t+1}+1,t},\frac{\left(\gamma^*_{\tau_{t+1}+1,t}+1\right)\boldsymbol{\Psi}^*_{\tau_{t+1}+1,t}}{\gamma^*_{\tau_{t+1}+1,t}d^*_{\tau_{t+1}+1,t}}\right).
\end{aligned}
$$

Similarly, for $\tau_{t+1}=t$, we can get

$$
(\boldsymbol{X}_{t+1}|\boldsymbol{X}_{1:t},\tau_{t+1}=t)\sim t\left(v_0-p+1,\boldsymbol{\mu}_0,\frac{(\gamma_0+1)\boldsymbol{\Psi}_0}{\gamma_0(v_0-p+1)}\right)
$$

which is just a special case of (13) by setting $\tau_{t+1}=t$.

REFERENCES

[1] K. Hoad, S. Robinson, and R. Davies, "Automating warm-up length estimation," *J. Oper. Res. Soc.*, vol. 61, no. 9, pp. 1389–1403, 2009.

[2] A. Marchetti, A. Ferramosca, and A. H. González, "Steady-state target optimization designs for integrating real-time optimization and model predictive control," *J. Process Control*, vol. 24, no. 1, pp. 129–145, Jan. 2014.

[3] J. Chen and J. Howell, "A self-validating control system based approach to plant fault detection and diagnosis," *Comput. Chem. Eng.*, vol. 25, no. 2, pp. 337–358, Mar. 2001.

[4] M. Kim, S. H. Yoon, P. A. Domanski, and W. V. Payne, "Design of a steady-state detector for fault detection and diagnosis of a residential air conditioner," *Int. J. Refrig.*, vol. 31, no. 5, pp. 790–799, Aug. 2008.

[5] S. K. Mahuli, R. R. Rhinehart, and J. B. Riggs, "Experimental demonstration of non-linear model-based in-line control of pH," *J. Process Control*, vol. 2, no. 3, pp. 145–153, Jan. 1992.

[6] S. Cao and R. R. Rhinehart, "An efficient method for on-line identification of steady state," *J. Process Control*, vol. 5, no. 6, pp. 363–374, Dec. 1995.

[7] J. Wu, S. Zhou, and X. Li, "Acoustic emission monitoring for ultrasonic cavitation based dispersion process," *J. Manuf. Sci. Eng.*, vol. 135, no. 3, Jun. 2013, Art no. 031015.

[8] J. Wu, Y. Chen, S. Zhou, and X. Li, "Online steady-state detection for process control using multiple change-point models and particle filters," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 688–700, Apr. 2016.

[9] J. Wu, Y. Chen, and S. Zhou, "Online detection of steady-state operation using a multiple-change-point model and exact Bayesian inference," *IIE Trans.*, vol. 48, no. 7, pp. 599–613, Mar. 2016.

[10] K. P. White, M. J. Cobb, and S. C. Spratt, "A comparison of five steady-state truncation heuristics for simulation," in *Proc. Winter Simulation Conf.*, Dec. 2000, pp. 755–760.

[11] A. Mhamdi, W. Geffers, F. Flehmig, and W. Marquardt, *On-Line Optimization of MSF Desalination Plants*. Malvern, PA, USA: LPT, RWTH, 1999.

[12] Y. Yao, C. Zhao, and F. Gao, "Batch-to-batch steady state identification based on variable correlation and Mahalanobis distance," *Ind. Eng. Chem. Res.*, vol. 48, no. 24, pp. 11060–11070, Oct. 2009.

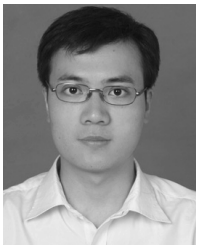
[13] W. Jianguo, S. Zhou, and X. Li, "Acoustic emission monitoring for ultrasonic cavitation based dispersion process," *J. Manuf. Sci. Eng.*, vol. 135, no. 3, Jun. 2013, Art. no. 031015.

[14] S. Narasimhan, C. S. Kao, and R. S. H. Mah, "Detecting changes of steady states using the mathematical theory of evidence," *AIChE J.*, vol. 33, no. 11, pp. 1930–1932, Nov. 1987.

[15] E. L. Crow, F. A. Davis, and M. W. Maxfield, *Statistics Manual: With Examples Taken From Ordnance Development*. New York, NY, USA: Dover, 1960.

[16] P. R. Brown and R. R. Rhinehart, "Demonstration of a method for automated steady-state identification in multivariable systems," *Hydrocarbon Process.*, vol. 79, no. 9, pp. 79–83, Sep. 2000.

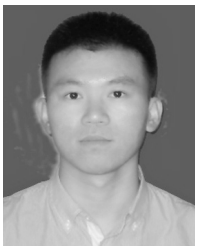
[17] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc. Ser. Methodol.*, vol. 57, no. 1, pp. 289–300, Jan. 1995.
[18] T. Jiang, B. Chen, X. He, and P. Stuart, "Application of steady-state detection method based on wavelet transform," *Comput. Chem. Eng.*, vol. 27, no. 4, pp. 569–578, Apr. 2003.
[19] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.
[20] D. C. Montgomery, *Introduction to Statistical Quality Control*. New York, NY, USA: Wiley, 2007.
[21] D. Aguado, A. Ferrer, A. Seco, and J. Ferrer, "Using unfold-PCA for batch-to-batch start-up process understanding and steady-state identification in a sequencing batch reactor," *J. Chemometrics*, vol. 22, no. 1, pp. 81–90, Jan. 2008.
[22] N. Chopin, "Dynamic detection of change points in long time series," *Ann. Inst. Statist. Math.*, vol. 59, no. 2, pp. 349–366, Jun. 2007.
[23] A. Hannart and P. Naveau, "An improved Bayesian information criterion for multiple change-point models," *Technometrics*, vol. 54, no. 3, pp. 256–268, May 2012.
[24] Y. Wen, J. Wu, and Y. Yuan, "Multiple-phase modeling of degradation signal for condition monitoring and remaining useful life prediction," *IEEE Trans. Rel.*, vol. 66, no. 3, pp. 924–938, Sep. 2017.
[25] P. Giordani and R. Kohn, "Efficient Bayesian inference for multiple change-point and mixture innovation models," *J. Bus. Econ. Statist.*, vol. 26, no. 1, pp. 66–77, Jan. 2008.
[26] X. Wang, "Incorporating knowledge on segmental duration in HMM-based continuous speech recognition," Ph.D. dissertation, Amsterdam Univ., Amsterdam, The Netherlands, 1997.
[27] S. J. Press, "Applied multivariate analysis: Using Bayesian and frequentist methods of inference," *J. Amer. Statist. Assoc.*, vol. 79, p. 386, May 1982.
[28] D. Lewandowski, D. Kurowicka, and H. Joe, "Generating random correlation matrices based on vines and extended onion method," *J. Multivariate Anal.*, vol. 100, no. 9, pp. 1989–2001, Oct. 2009.
[29] J. D. Kelly and J. D. Hedengren, "A steady-state detection (SSD) algorithm to detect non-stationary drifts in processes," *J. Process Control*, vol. 23, no. 3, pp. 326–331, Mar. 2013.
[30] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, Mar. 1993.
[31] F. Ju, J. Li, and J. A. Horst, "Transient analysis of serial production lines with perishable products: Bernoulli reliability model," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 694–707, Feb. 2017.
[32] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. London, U.K.: Springer, 2000.
[33] J. Wu, H. Xu, F. Ju, and B. Tseng, "Adaptive minimal confidence region rule for multivariate initialization bias truncation in discrete-event simulations," *Technometrics*, submitted for publication.



Jianguo Wu received the B.S. degree in mechanical engineering from Tsinghua University, Beijing, China, in 2009, the M.S. degree in mechanical engineering from Purdue University, West Lafayette, IN, USA, in 2011, and the M.S. degree in statistics and the Ph.D. degree in industrial and systems engineering from the University of Wisconsin–Madison, Madison, WI, USA, in 2014 and 2015, respectively.

From 2015 to 2017, he was an Assistant Professor with the Department of Industrial, Manufacturing and Systems Engineering (IMSE), The University of Texas at El Paso, El Paso, TX, USA. He is currently an Assistant Professor with the Department of Industrial Engineering and Management, Peking University, Beijing. 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.

Dr. Wu is a member of INFORMS, IISE, and SME.

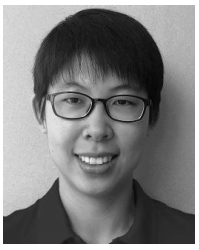


Honglun Xu received the B.S. degree in petroleum engineering from Northeast Petroleum University, Daqing, China, in 2013, and the M.S. degree in petroleum engineering from the China University of Petroleum, Beijing, China, in 2016. He is currently pursuing the Ph.D. degree in the Computational Science Program at The University of Texas at El Paso (UTEP), El Paso, TX, USA.

His research interests are focused on statistical modeling, process monitoring, and quality control.

Chen Zhang received the B.Eng. degree in electronic science and technology (optics) from Tianjin University, Tianjin, China, in 2012, and the Ph.D. degree in industrial systems engineering and management from the National University of Singapore, Singapore, in 2017.

She is currently an Assistant Professor of industrial engineering, Tsinghua University, Beijing, China. Her current research interests include developing approaches for modeling and monitoring of engineering systems with complex data.



Yuan Yuan received the B.E. degree from Tsinghua University, Beijing, China, in 2006 and the M.S. degree in industrial and systems engineering, the M.S. degree in statistics, and the Ph.D. degree in industrial and systems engineering from the University of Wisconsin–Madison, Madison, WI, USA, in 2010, 2011, and 2014, respectively.

She is currently a Research Scientist with the IBM Research, Singapore. Her research mainly focuses on data analytics, in particular, developing innovative and generic data-driven modeling and analysis methodologies for complex systems with massive data.

Dr. Yuan has received a number of awards including the QSR Best Student Paper Award from the Institute for Operations Research and the Management Sciences (INFORMS) in 2014 and the featured article award of *IE Magazine* in 2010.