Received: 5 February 2018 | Revised: 12 August 2018 | Accepted: 18 September 2018

DOI: 10.1002/qre.2409



RESEARCH ARTICLE

WILEY

# A Bayesian approach to sequential monitoring of nonlinear profiles using wavelets

Roumen Varbanov[1] | Eric Chicken[1] | Antonio Linero[1] | Yun Yang[2]

[1]Department of Statistics, Florida State University, Tallahassee, Florida, USA

[2]Department of Statistics, University of Illinois Urbana-Champaign, Champaign, Illinois

**Correspondence**
Roumen Varbanov, Department of Statistics, Florida State University, 214 Rogers Building (OSB), 117 N. Woodward Ave., PO Box 3064330, Tallahassee, Florida.
Email: r.varbanov@stat.fsu.edu

**Abstract**

We consider change-point detection and estimation in sequences of functional observations. This setting often arises when the quality of a process is characterized by such observations, called profiles, and monitoring profiles for changes in structure can be used to ensure the stability of the process over time. While interest in phase II profile monitoring has grown, few methods approach the problem from a Bayesian perspective. We propose a wavelet-based Bayesian methodology that bases inference on the posterior distribution of the change point without placing restrictive assumptions on the form of profiles. By obtaining an analytic form of this posterior distribution, we allow the proposed method to run online without using Markov chain Monte Carlo (MCMC) approximation. Wavelets, an effective tool for estimating nonlinear signals from noise-contaminated observations, enable us to flexibly distinguish between sustained changes in profiles and the inherent variability of the process. We analyze observed profiles in the wavelet domain and consider two possible prior distributions for coefficients corresponding to the unknown change in the sequence. These priors, previously applied in the nonparametric regression setting, yield tuning-free choices of hyperparameters. We present additional considerations for controlling computational complexity over time and their effects on performance. The proposed method significantly outperforms a relevant frequentist competitor on simulated data.

**KEYWORDS**

Bayesian, phase II, profile monitoring, statistical process control, wavelets

## 1 | INTRODUCTION

Technological advancements in manufacturing have increased the amount and complexity of data in process monitoring, prompting the development of statistical process control (SPC) methods to analyze these data. SPC techniques are used to ensure quality when analyzing sequences of observations output by the process. These observations are realizations of a quality characteristic, and changes in the sequence can indicate a deterioration in quality. While the most common SPC charts are designed to monitor univariate or low-dimensional multivariate quality characteristics, recent research in this area has predominately focused on high-dimensional or functional observations.

The sequentially observed units of study in this case are noise-contaminated, discretely-sampled functions called profiles. At each time $t$, a functional profile

$$y_i^t = f^t(x_i) + \epsilon_i^t, \quad i = 1, \ldots, n, \tag{1}$$

  

is observed, where $y_i^t$ are the values of the quality characteristic, $x_i$ are the corresponding values of an explanatory variable, $n$ is the number of observations in each profile, and the error terms $\epsilon_i^t$ represent the noise inherently associated with the process. The observations $y_i^t$ can be measurements made across positions $x_i$ of a manufactured part. While subject to some variability, reflected by $\epsilon_i^t$, the measurements should meet specified quality standards.

Examples of profile monitoring applications in literature include semiconductor manufacturing, where gate oxide thickness is a function of surface location;[1] a stamping operation, where stamping force is a function of crank angle;[2] artificial sweetener development, where amount dissolved is a function of temperature;[3] roundness evaluation of mechanical components, where radial measurement is a function of turning;[4] and automobile engine testing where torque produced is a function of engine speed.[5]

SPC is broken into two steps—phase I and phase II. In phase I, historical data are retrospectively analyzed to identify and model the desired, or in-control, performance of the process. In phase II, online procedures sequentially process new observations to detect deviations from the targets identified in phase I. Profile monitoring methods can be used to detect changes in the underlying structure or variability of profiles, as either change may be indicative of a deterioration in quality. If a process is operating in the presence of such conditions, which are termed faults, it is said to be out of control. Faults can be increasing, as a trend, or sustained at a constant magnitude. In some cases, the presence of a particular fault may be only temporary. In others, the process needs to be stopped, and the cause of the fault addressed before the process can resume functioning properly. A successful monitoring method will allow an in-control process to run without disruption but will stop an out-of-control process quickly. These criteria are usually evaluated in terms of average run length (ARL), the mean time until a false alarm is signaled for in an in-control process, and the mean time until a change is detected in an out-of-control process.

The transition from traditional SPC charts to profile monitoring has been followed by the evolution of profile monitoring itself. Approaches to profile monitoring can be classified by phase and the assumptions they make of profiles. In early work, monitoring methods focused on linear profiles. This assumption was relaxed in methods that assumed a nonlinear but parametric form of profile. The most robust methods consider nonparametric models for nonlinear profiles. For good reviews on profile monitoring, see Woodall et al,[6-8] Noorossana et al,[9] and Qiu.[10]

Parametric approaches generally summarize an in-control profile with a small set of parameters in phase I; estimates of the those parameters from new profiles are then monitored using univariate or multivariate control charts in phase II. Although there exist changes that may not be reflected in the parameters of the chosen model, these methods can be useful when profiles have the assumed parametric form. In practice, however, there are often cases with irregular profiles that are not parametrizable.

Many profile monitoring methods are designed with a specific application in mind. We approach the problem generally and require a nonparametric method to avoid restrictions on the form of the profile. Nonparametric methods generally rely on smoothing and dimension reduction tools to separate signal from noise in order to compute metrics that accurately quantify the difference between observed profiles and an in-control template.

Chang and Yadama[11] used wavelets to smooth profiles before fitting an estimate with B-splines and computing a statistic based on the control points. Williams et al[12] also used B-splines, but instead of using the control points, they proposed a variety of metrics based on the smoothed estimate. Gardner[1] used bivariate splines in the case of surface profiles with multiple explanatory variables. It should be noted that the use of splines implies that a profile has no explicitly unsmooth features such as jumps or points of nondifferentiability. Qiu et al[13,14] used local linear kernel smoothing in exponentially weighted moving average (EWMA) control schemes. Shiau et al[15] smoothed data with splines and used principal component analysis (PCA) to project profiles into a lower dimensional space. The principal components of projected profiles are then monitored using independent or combined control charts. In addition to PCA, Ding et al[16] explored the possibility utilizing a normality assumption in dimension reduction through independent component analysis (ICA).

Wavelets, a tool for both smoothing and dimension reduction, have proven to be a useful aid in many profile monitoring methods. Jin and Shi,[2,17] Lada et al,[18] and Zhou et al[19] used different thresholding schemes to select a subset of wavelet coefficients to monitor using multivariate control charts. By monitoring only a subset of coefficients, these nonparametric approaches suffer from some of the same pitfalls as parametric methods—changes not reflected in the chosen subset will not be detected. Chicken et al[20] incorporated thresholding when computing a statistic using all observed coefficients in a change-point procedure based on a likelihood ratio test (LRT). Wavelets have also been used in several procedures that relax common assumptions such as normality and independence on the noise structure of profiles.[21-23]

In statistical inference, a Bayesian framework can improve the interpretability of results and allow for the natural input of prior knowledge—two possibilities that are particularly useful in any type of process monitoring.

Despite demonstrated success in a variety of applications, Bayesian analysis has received little attention in profile monitoring. This can likely be attributed to the high computational cost typically associated with Bayesian inference. Abbas et al[24,25] considered the linear profile monitoring problem and monitored the posterior estimates of parameters for changes. Park and Shrivastava[26] considered a unique profile monitoring scenario involving nanoparticle growth during self-assembly. In a moderate dimensional setting, they used B-splines to parametrize geometric data in a time-series model. They used phase I analysis to facilitate computation of a Bayes factor approximation for inference.

In this paper, we propose a wavelet-based Bayesian (WBB) approach to phase II nonlinear profile monitoring. The proposed method monitors profiles for changes in underlying functional structure in the presence of Gaussian white noise. We establish a change-point framework and base inference on the posterior distribution of the change point given the observed profiles. The control chart statistic is the posterior probability that a change has occurred. Once the process is stopped, the maximum a posteriori probability (MAP) estimate of the change point can be used to identify the time when the process went out of control.

For computing the posterior of the change point, observed profiles are projected into the wavelet domain. We consider each observation in its entirety without direct dimension reduction. Instead, we assign the wavelet coefficients of the unknown functional change a prior that has been previously used for wavelet thresholding and reflects the sparsity of the wavelet representation. The posterior can be then be computed using all of the coefficients from a profile while maintaining the expectation that many will be zero, regardless of whether a change has occurred. The control chart statistic can then allow evidence of a subtle, but sustained, fault condition to build while taking into account information about the frequency and location of potential faults. Due to the orthogonality of the wavelet basis, we obtain an analytic form of the posterior and do not rely on computationally intensive Markov chain Monte Carlo (MCMC) for approximation. The proposed method shows substantial improvement over initial work in Varbanov et al.[27]

The remainder of the paper is organized as follows. Section 2 provides a brief overview of wavelet analysis in statistics. Section 3 describes the proposed method in detail. Section 4 demonstrates the proposed method's improvement over a relevant frequentist competitor through simulation. Section 5 concludes with a discussion of results, other considerations, and future work.

## 2 | BACKGROUND

### 2.1 | Wavelet representation

Wavelets are localized wave-like functions that can be used to generate a basis for a large class of functions. Often introduced as a refinement of Fourier analysis, wavelet analysis of functions allows for localization in both the frequency and $x$-axis domains. Wavelet bases can have several properties particularly useful in statistical applications: orthogonality, which can decorrelate data and maintain independence, and compactly supported basis elements, which can be scaled to reflect the finite nature of data. For good introductions to wavelets and their properties, see Ogden[28] and Vidakovic.[29]

Let $\phi$ and $\psi$ denote appropriately chosen father and mother wavelet functions, respectively. The many choices for this pair of functions carry different properties.[30] Here, we choose compactly supported $\phi$ and $\psi$ to generate an orthonormal basis. The translations and dilations of $\phi$ and $\psi$ given by

$$\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k), \quad \psi_{jk}(x) = 2^{j/2}\psi(2^j x - k) \tag{2}$$

generate the following orthonormal basis for the space of square-integrable functions $L_2(\mathbb{R})$:

$$\{\phi_{j_0 k}, \psi_{jk} | j_0, j, k \in \mathbb{Z}; j \geq j_0\}, \tag{3}$$

for any fixed $j_0$. A function $f \in L_2(\mathbb{R})$ can then be expressed as the infinite series

$$f(x) = \sum_k \xi_{j_0 k}\phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty}\sum_k \theta_{jk}\psi_{jk}(x), \tag{4}$$

where the wavelet coefficients $\xi_{j_0 k} = \langle f, \phi_{jk}\rangle$ and $\theta_{jk} = \langle f, \psi_{jk}\rangle$ are given by the usual inner product of the function $f$ and the corresponding basis function.

The smoothest structure of a function is represented by the first series in Equation 4, consisting of translations of the dilated father wavelet $\phi_{j_0 k}$. Higher frequency parts of the function are represented by separate series of translations and increasing dilations of the mother wavelet $\psi_{jk}$. As such, we refer to $\xi_{j_0 k}$ as *smooth*, or *coarse*, wavelet coefficients and $\theta_{jk}$ as *detail* wavelet coefficients. The projection of a function across this sequence of subspaces allows for the multiresolution property of wavelets. Varying the parameter $j$ changes the resolution level of the analysis, allowing one to zoom in or out on the smooth or detailed structure of $f$. Varying the parameter $k$ allows one to analyze a function in the $x$-axis domain. This combination of spatial and frequency adaptivity allows wavelets to model both irregular and smooth functions. The important features of functions can often be retained in parsimonious representations involving relatively few coefficients and basis functions.

## 2.2 | Discrete wavelet transform

In practice, functional data are not collected continuously but discretized into high-dimensional vectors. For these cases, there exist efficient algorithms for estimating the coefficients of the discrete analog of Equation 4. The pyramid algorithm of Mallat[31] for the discrete wavelet transform (DWT) estimates the coefficients of the function $f$ sampled at $n = 2^J$ equispaced points with computational cost $O(n)$. The DWT can be expressed in terms of an $n \times n$ orthogonal transformation matrix $W$. A vector of $n$ empirical coefficients for $\boldsymbol{f} = (f_1, f_2, \ldots, f_n)'$ is given by

$$\boldsymbol{d} = (c_{j_0 1}, c_{j_0 2}, \ldots, c_{j_0 2^{j_0}}, d_{j_0 1}, d_{j_0 2}, \ldots, d_{J-1,2^{J-1}})' = W\boldsymbol{f}. \tag{5}$$

The lowest resolution level of the decomposition, $j_0$, is specified by the user and may vary from 0 to $J-1$. Applying the inverse DWT, $W^{-1} = W'$, to the estimated coefficients will recover the original data $f = W'\boldsymbol{d} = W'Wf$. The DWT can be executed without numerical integration by treating the observed values $\boldsymbol{f}$ as the coefficients at resolution level $J$. With this approach, wavelet bases become wavelet filters. Hereafter, the term "wavelet coefficient" refers to the wavelet coefficient from the DWT of a discretely sampled function.

## 2.3 | Wavelet thresholding

In observed data, wavelet coefficients estimated from a noisy function are themselves noisy versions of the true wavelet coefficients corresponding to the underlying function. Wavelet thresholding can be used to obtain a parsimonious estimate of the underlying function in nonparameteric regression. A thresholding procedure assumes sparsity of the wavelet representation and modifies observed coefficients, setting a portion of them equal to zero, before using the coefficients to reconstruct the function.

A thresholding procedure requires the selection of a threshold value, $\lambda$, and a rule by which to modify the coefficients. Thresholding can be applied to coefficients individually (term-by-term) or in groups (blocks). For a threshold value $\lambda$, hard thresholding sets coefficients with magnitudes less than $\lambda$ to zero. Soft thresholding combines hard thresholding with the additional shrinkage of coefficients with magnitudes larger than $\lambda$. Global thresholding methods apply the same $\lambda$ to coefficients across all levels while level-dependent thresholds use different values of $\lambda$ for each level $j$. Thresholding is usually only applied to the detail coefficients, $d_{jk}$; smooth coefficients, $c_{j_0 k}$, are considered representative of a function's basic structure.

Bayesian approaches to wavelet estimation generally use a mixture prior to reflect the sparsity of coefficients. For the true detail coefficients of a function, several methods[29,32,33] considered the prior

$$\theta_{jk} \sim (1-\omega_j)\delta(0) + \omega_j N(0, s_j^2), \tag{6}$$

with hyperparameters that vary with resolution level. In this model, coefficients are independent and can be estimated with the posterior mean or median. The hyperparameters $\omega_j$ and $s_j$ can be chosen to reflect prior knowledge, or they can be estimated from the data. To improve theoretical properties and to reflect empirically observed marginal distributions of real data, Johnstone and Silverman[34] proposed modifying the nonzero portion of Equation 6 to a density with heavier tails than those of the normal distribution. They also proposed an empirical Bayes approach to choosing the hyperparameters by maximizing the marginal likelihood.

# 3 | METHODOLOGY

## 3.1 | Change-point framework

The process monitoring problem of interest can be framed as

$$\boldsymbol{y^t} = f^0(\boldsymbol{x}) + g(\boldsymbol{x})I(t \geq \tau) + \boldsymbol{\varepsilon^t}, \tag{7}$$

where $f^0$ is the functional relationship between the explanatory variable and the quality characteristic when the process is in control, and $g$ is an unknown functional change introduced when the process goes out of control. Here, we express a profile as the vector $\boldsymbol{y^t}$ and use the notation $f(\boldsymbol{x})$ to denote the values of the function $f$ at each component of the vector $\boldsymbol{x}$. Our inferential objectives are to (1) determine whether the process is in control given the profiles observed at times $t = 1, 2, \ldots, T$ and (2) update this classification with each newly observed $\boldsymbol{y^t}$. This is equivalent to considering the hypothesis test

$$\begin{aligned} H_0 &: f^0 = f^1 = f^2 = \ldots = f^T \\ H_a &: f^0 = f^1 = f^2 = \ldots = f^{\tau-1} \neq f^\tau = \ldots = f^T, \\ &\qquad \tau \in \{1, 2, \ldots, T\} \end{aligned} \tag{8}$$

at each time $T$. If the null hypothesis is rejected and the process is classified as out of control, we want to estimate $\tau$, the time of the first out-of-control profile.

As indicated by the framework in Equation 7, we consider a single change point after which the functional change $g$ remains constant. The proposed method is designed to detect changes in the underlying function of profiles and does not directly monitor profiles for changes in variability. We assume that the errors $\varepsilon_i^t$ are iid normal random variables with mean zero and variance $\sigma_\varepsilon^2$. We consider $\sigma_\varepsilon^2$ and the in-control function $f^0$ to be known a priori.

If unknown, $f^0$ and $\sigma_\epsilon^2$ must be estimated from historical data during phase I monitoring. Without loss of generality, we set $f^0(x) = 0$ and $\sigma_\epsilon^2 = 1$ for notational convenience. In practice, this is equivalent to analyzing the standardized profiles

$$\boldsymbol{y}^{t'} = \frac{\boldsymbol{y}^t - f^0(\boldsymbol{x})}{\sigma_\epsilon}. \tag{9}$$

This essentially reframes the problem of change detection as one of signal detection, because we no longer have the presence of a signal in observations prior to a change in the sequence.

When considering the observed data, we project each profile $\boldsymbol{y}^t$ into a vector of coefficients in the wavelet domain using the DWT. Let $\boldsymbol{d}^t$ denote the vector of all empirical wavelet coefficients of profile $\boldsymbol{y}^t$ and let $d_i^t$ denote the $i$th entry in this vector, as the vector is defined in Equation 5. As an orthogonal transformation, the DWT will maintain the independence of the initial noise structure. So when the process is in control and $t < \tau$,

$$\boldsymbol{d}^t = W\boldsymbol{\epsilon}^t = \boldsymbol{\epsilon}^{t\prime}, \tag{10}$$

and $\boldsymbol{d}^t$ will contain only Gaussian white noise $\boldsymbol{\epsilon}^{t\prime}$. Once the process is out of control and $t \geq \tau$,

$$\boldsymbol{d}^t = Wg(\boldsymbol{x}) + W\boldsymbol{\epsilon}^t = \boldsymbol{\theta} + \boldsymbol{\epsilon}^{t\prime}, \tag{11}$$

and $\boldsymbol{d}^t$ contains noise-contaminated observations of $\boldsymbol{\theta}$, the wavelet coefficients of $g(\boldsymbol{x})$. For the application of the DWT to $\boldsymbol{y}^t$, the constant vector $\boldsymbol{x}$ is assumed to contain equispaced values of the explanatory variable, and profiles are assumed to be of dyadic length $n = 2^J$, $J \in \mathbb{N}$.

## 3.2 | Obtaining the posterior distribution of $\tau$

We approach process monitoring from a Bayesian perspective, basing inference on the posterior distribution of the change point given the data. This framework was introduced by Shiryaev,[35] who first developed the optimal change detection procedure in the Bayesian formulation of the change-point problem. For our goal of developing an online monitoring method, we require a tractable posterior. To obtain an expression for this posterior, prior distributions need to be assigned to the unknown parameters in our model: $\tau$, the time of the first out-of-control profile, and $\boldsymbol{\theta}$, the wavelet coefficients of $g(\boldsymbol{x})$. Let $D = \{\boldsymbol{d^1}, \boldsymbol{d^2}, \ldots, \boldsymbol{d^T}\}$ denote the empirical wavelet coefficients, or data, collected until the current time. The marginal posterior of $\tau$ can be written as

$$\begin{aligned}\pi(\tau|D) &= \frac{\int_\theta f(D|\tau,\boldsymbol{\theta})\pi_\tau(\tau)\pi_\theta(\boldsymbol{\theta})d\boldsymbol{\theta}}{\sum_\tau \int_\theta f(D|\tau,\boldsymbol{\theta})\pi_\tau(\tau)\pi_\theta(\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &= \frac{f(D|\tau)\pi_\tau(\tau)}{\sum_{t=1}^{T} f(D|\tau=t)\pi_\tau(\tau=t) + f(D|\tau>T)\pi_\tau(\tau>T)},\end{aligned} \tag{12}$$

where $\pi_\tau(\tau)$ is the prior distribution of $\tau$, $\pi_\theta(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$, and $f(D|\tau)$ is the marginal density of the empirical wavelet coefficients conditioned only on $\tau$. The probability mass function of the posterior will allow us to compute $\pi(\tau \leq T|D)$, the probability that a change has occurred given our data. We can then classify the process as out of control when this value is large enough. The selection of the threshold for making this classification is discussed in Section 3.3. This threshold will define the amount of evidence required to stop the process by acting as an upper control limit (UCL) for $\pi(\tau \leq T|D)$. After classifying a process as out of control, we can estimate the exact change point by choosing

$$\hat{\tau} = \underset{1\leq t\leq T}{\arg\max}\ \pi(\tau = t|D), \tag{13}$$

the most probable change point according to the posterior distribution.

For the prior on $\tau$, we analogize each observed profile to a Bernoulli trial and consider $\tau$ to be the time of the first "success." This is reflected by the geometric prior

$$\pi_\tau(\tau = t) = (1-p)^{t-1}p, \quad t = 1, 2, \ldots. \tag{14}$$

In this analogy, a memoryless in-control process has a probability $p$ of being out of control at the next time step and will, on average, go out of control at time $t = 1/p$.

When choosing a prior for $\boldsymbol{\theta}$ without any knowledge of $g(\boldsymbol{x})$, we simply look to utilize the beneficial properties of wavelets—namely, orthogonality and sparsity. The components of the vector $\boldsymbol{\theta}$ correspond to the true wavelet coefficients resulting from the projection in Equation 5 and can accordingly be written as

$$\boldsymbol{\theta} = (\xi_{j_0 1}, \xi_{j_0 2}, \ldots, \xi_{j_0 2^{j_0}}, \theta_{j_0 1}, \theta_{j_0 2}, \ldots, \theta_{J-1,2^{J-1}})', \tag{15}$$

where $\xi_{j_0 k}$ are smooth coefficients, and $\theta_{jk}$ are detail coefficients. To model the sparsity of the detail coefficients, we modify the mixture prior from Equation 16, which has been previously used for wavelet thresholding in the nonparametric regression setting.[32,34] In these works, the authors increased the precision of estimates by varying the hyperparameters $\omega$ and $a$ by resolution level $j$ and estimated these values from the data to adapt to a given function. As we look to avoid unnecessary computational

cost, we propose using prespecified fixed values for the hyperparameters. To avoid overspecifying the structure of the unknown coefficients, we propose using global hyperparameters instead of the previously implemented level-dependent versions. The resulting prior is given by

$$\pi_{\theta_{jk}}(\theta_{jk}) = (1-\omega)\delta_0(\theta_{jk}) + \omega\gamma_s(\theta_{jk}), \quad j = j_0, \ldots, J-1, \quad k = 1, \ldots, 2^j, \tag{16}$$

where $\delta_0$ is the Dirac delta function at zero, and $\gamma_s$ is a unimodal symmetric density, centered at zero with scale parameter $s > 0$. The mixture represents the idea that with probability $\omega$, a coefficient will be drawn from $\gamma_s$ and with probability $1-\omega$, a coefficient will be exactly zero. For $\gamma_s$, we consider the normal density

$$\gamma_s(\theta_i) = \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{\theta_i^2}{2s^2}\right), \tag{17}$$

as in Abramovich et al[32] and the Laplace density

$$\gamma_s(\theta_i) = \frac{1}{2} s \exp(-s|\theta_i|), \tag{18}$$

as in Johnstone and Silverman.[34] A unique benefit of the normal density for this purpose is that it is the conjugate prior for the normal likelihood associated with the observed coefficients. Alternatively, the heavy tails of the Laplace distribution may be more suitable for modeling the extreme coefficients associated with certain functions, particularly when we model all nonzero coefficients with a single $\gamma_s$. We compare the performance of both choices of $\gamma_s$ in Section 4, referring to them as the normal and Laplace mixtures.

We complete the prior specification for smooth coefficients of $\boldsymbol{\theta}$ by setting $\omega = 1$ in Equation 16, resulting in

$$\pi_{\xi_{j_0k}}(\xi_{j_0k}) = \gamma_a(\xi_{j_0k}), \quad k = 1, \ldots, 2^{j_0}. \tag{19}$$

With this choice, we reflect the expectation that all smooth coefficients in our model will be nonzero and that these coefficients will be on the same scale as nonzero detail coefficients. Smooth coefficients are subject to the scale of the function being estimated. Because any effective monitoring method will detect large global changes reflected by large smooth coefficients, we prioritize the detection of subtle changes whose magnitude will be relatively small. In these cases, there is no reason to expect a significant difference in the magnitude of smooth and detail coefficients. Due to the orthogonality of the wavelet representation, the complete prior models each coefficient independently, a common assumption in Bayesian wavelet estimation.

To compute the posterior distribution of the parameter of interest $\tau$, we require $f(D|\tau)$, the density of the observed coefficients conditioned only the change point. Obtaining this density requires marginalizing out the unknown coefficients through integration over the prior distribution of $\boldsymbol{\theta}$. Both proposed choices of prior for $\boldsymbol{\theta}$ will result in closed-form expressions of $f(D|\tau)$ and, in turn, $\pi(\tau|D)$. Note that for $t < \tau$ (in-control profiles), $d_i^t \overset{\text{iid}}{\sim} N(0,1)$, and for $t \geq \tau$ (out-of-control profiles), $d_i^t \sim N(\theta_i, 1)$. When evaluating $f(D|\tau)$, we factor the density into two parts: the density of coefficients of in-control profiles and the density of coefficients of out-of-control profiles. The density of in-control coefficients can be written as a product of the individual densities of the independent coefficients. Empirical coefficients in position $i$ from out-of-control profiles, $\boldsymbol{d}_i^{t\geq\tau} = (d_i^\tau, d_i^{\tau+1}, \ldots, d_i^T)$, are dependent in that they correspond to the same parameter $\theta_i$. We account for the dependence of equivalently positioned coefficients by rewriting their density in terms of their respective sample means before integrating out the parameter $\theta_i$. The sample mean of each set of corresponding out-of-control coefficients is given by

$$\bar{d}_i^{t\geq\tau} = \frac{1}{T-\tau+1}\sum_{t=\tau}^{T} d_i^t \sim N\left(\theta_i, \frac{1}{T-\tau+1}\right). \tag{20}$$

We can now write

$$\begin{aligned} f(D|\tau) &= \int f(D \mid \boldsymbol{\theta}, \tau)\pi_\theta(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= \left[\prod_{t=1}^{\tau-1}\prod_{i=1}^{n}\phi(d_i^t \mid 0,1)\right]\int\left[\prod_{t=\tau}^{T}\prod_{i=1}^{n}\phi(d_i^t|\theta_i,1)\right]\pi_{\theta_i}(\theta_i)d\theta_i \\ &= \left[\prod_{t=1}^{\tau-1}\prod_{i=1}^{n}\phi(d_i^t \mid 0,1)\right]\left[\prod_{i=1}^{n}u(\boldsymbol{d}_i^{t\geq\tau})\int\phi\left(\bar{d}_i^{t\geq\tau}|\theta_i, \frac{1}{n_\tau}\right)\pi_{\theta_i}(\theta_i)d\theta_i\right] \\ &= \left[\prod_{t=1}^{\tau-1}\prod_{i=1}^{n}\phi(d_i^t \mid 0,1)\right]\left[\prod_{i=1}^{n}u(\boldsymbol{d}_i^{t\geq\tau})v(\bar{d}_i^{t\geq\tau})\right], \end{aligned} \tag{21}$$

where $n_\tau = T - \tau + 1$ is the number of out-of-control profiles, and $\phi(\cdot|\mu,\sigma^2)$ is the normal density function with mean $\mu$ and variance $\sigma^2$. The functions $u$ and $v$ result from the factorization of the density of $\boldsymbol{d}_i^{t\geq\tau}$ and are given below.

$$u(\boldsymbol{d}_i^{t\geq\tau}) = \frac{1}{\sqrt{n_\tau(2\pi)^{n_\tau-1}}}\exp\left(-\frac{\sum_{t=\tau}^{T}(d_i^t - \bar{d}_i^{t\geq\tau})^2}{2}\right). \tag{22}$$

If $\gamma_s$ is the normal density, then

$$v(\bar{d}_i^{t\geq\tau}) = (1-\omega)\phi\left(\bar{d}_i^{t\geq\tau}|0, \frac{1}{n_\tau}\right) + \omega\phi\left(\bar{d}_i^{t\geq\tau}|0, s^2 + \frac{1}{n_\tau}\right). \tag{23}$$

If $\gamma_s$ is the Laplace density, then

$$v(\bar{\theta}_i^g) = (1-\omega)\phi\left(\bar{d}_i^{t\geq\tau}|0, \frac{1}{n_\tau}\right) + \omega z_s\left(\bar{d}_i^{t\geq\tau}\right), \tag{24}$$

where

$$z_s(\bar{d}_i^{t\geq\tau}) = \frac{s}{2}\exp\left(\frac{s^2}{2n_\tau}\right)\left[\exp\left(-s\bar{d}_i^{t\geq\tau}\right)\Phi_\tau\left(\bar{d}_i^{t\geq\tau} - \frac{s}{n_\tau}\right) + \exp\left(s\bar{d}_i^{t\geq\tau}\right)\tilde{\Phi}_\tau\left(\bar{d}_i^{t\geq\tau} + \frac{s}{n_\tau}\right)\right]. \tag{25}$$

Here, $\Phi_\tau$ refers to the normal cumulative distribution function (CDF) with $\mu = 0$ and $\sigma^2 = \frac{1}{n_\tau}$ and $\tilde{\Phi}_\tau = 1 - \Phi_\tau$.

### 3.3 | Calibration and hyperparameters

We have now introduced each of the components needed to obtain the posterior of $\tau$ in Equation 12. To implement the process monitoring method, we need to select the hyperparameters for the prior distributions of $\tau$ and $\boldsymbol{\theta}$ and determine an appropriate UCL for the posterior probability that a change has occurred.

The prior distribution of $\boldsymbol{\theta}$ contains the hyperparameters $\omega$ and $s$, which reflect the sparsity and scale of the nonzero detail coefficients, respectively. When these priors were used to model detail coefficients in the nonparametric regression setting,[32,34] the authors proposed using the median of the posterior distribution as the estimate of a coefficient. This results in a true thresholding procedure that sets coefficients less than a certain magnitude to zero. The relationship between the posterior median $\text{med}(\theta_i|d_i)$, an observed wavelet coefficient $d_i$, and the hyperparameters $\omega$ and $s$ can be expressed in closed form.[32,34,36] Without loss of generality in the thresholding procedure, consider $d_i > 0$. The posterior median is given by

$$\text{med}(\theta_i|d_i) = \max(0, h(d_i, \omega, s)). \tag{26}$$

When $\gamma_s$ is the normal density,

$$h(d_i, \omega, s) = \frac{s^2}{1+s^2}d_i - \frac{s}{\sqrt{1+s^2}}\Phi^{-1}\left(\frac{1+\min(\omega_{\text{post}}, 1)}{2}\right), \tag{27}$$

where $\Phi$ is the standard normal CDF, and the posterior odds ratio for the coefficient at 0 is given by

$$\omega_{\text{post}} = \frac{1-\omega}{\omega}\sqrt{s^2+1}\exp\left(-\frac{s^2 d_i^2}{2(s^2+1)}\right). \tag{28}$$

When $\gamma_s$ is the Laplace density,

$$h(d_i, \omega, s) = d_i - s - \Phi^{-1}(\min(c, 1)), \tag{29}$$

where

$$c = s^{-1}\phi(d_i - s)\left(\omega^{-1} + \frac{1}{2}s\left(\frac{\Phi(d_i - s}{\phi(d_i - s} + \frac{\tilde{\Phi}(d_i + s)}{\phi(d_i + s)}\right) - 1\right). \tag{30}$$

Equations 26 to 30 also define the relationship between $\omega, s$, and the posterior median threshold, which is given by

$$\lambda_{\text{med}} = \min\{d : d \geq 0, h(d, \omega, s) > 0\}. \tag{31}$$

Because $\omega$ has the same meaning in either choice of $\gamma_s$, we treat $\omega$ as a user-specified quantity reflective of some sparsity level. In subsequent sections, we show results for the somewhat arbitrary choice $\omega = 0.05$. Given $\omega$, the user can choose $s$ that results in a desired median threshold. Without any prior knowledge of the function $g$, we propose mapping $\lambda_{\text{med}}$ to the universal threshold $\lambda_{\text{univ}} = \sqrt{2\log(n)}$ of Donoho and Johnstone,[37] where $n$ corresponds to the dimensionality of profiles. For $\omega = 0.05$, this mapping results in $s = 1.07, 1.31$ for the normal and Laplace mixtures, respectively.

The hyperparameter $p$ for the prior on $\tau$ and the UCL for $\pi(\tau \leq T|D)$ work in unison to control the sensitivity of the proposed method and are adjusted to control for false alarms. When calibrating the method to result in a desired in-control ARL, the user can either (1) adjust $p$ for a fixed UCL or (2) adjust the UCL for a fixed $p$. The former eases interpretation when comparing $\pi(\tau \leq T|D)$ to the UCL while the latter allows for more direct input of knowledge of the in-control properties of the process. For example, by setting UCL $= 0.5$, we utilize the intuitive interpretability of the Bayesian framework and classify the process as out-of-control when it is more likely than not that a change has occurred. Alternatively, if an in-control process will on average run $t$ time steps before going out of control, then setting $p = 1/t$ will match the mean of the prior distribution to that of the actual distribution. The performance of the proposed method under both calibration techniques is similar in terms of in-control stability and out-of-control detection time. Because the distribution of in-control run length is intractable, calibrating either $p$ or the UCL to result in a particular in-control ARL can be carried out through Monte Carlo simulation.

### 3.4 | Computational considerations

A fundamental problem with an online monitoring method based on the posterior distribution in Equation 12 is that the posterior requires $O(T)$ computations at time $T$. The linear growth of computation time needed to process a new observation is attributable to the growth in the number of possible change points. We circumvent this issue through the commonly used approach of using a moving

window or windowing (for a general overview of sample size considerations in SPC, see Montgomery[38]). With this approach, we only consider the most recent window, or subset, of data when computing the posterior. Let $W$ denote the window size in number of time steps and let $T_1^W, T_2^W, \ldots, T_W^W$ denote the respective time indices of profiles in the window at time $T$ (note that $T_1^W = T-W+1$ and $T_W^W = T$). The windowed posterior, or the posterior given only the data $D_W^T = \{\boldsymbol{d}^t : t = T_1^W, T_2^W, \ldots, T_W^W\}$, is then

$$\begin{aligned}\pi(\tau|D_W^T) &= \frac{f(D_W^T|\tau)\pi_\tau(\tau)}{\sum_{t=1}^{T} f(D_W^T|\tau=t)\pi_\tau(\tau=t) + f(D_W^T|\tau>T)\pi_\tau(\tau>T)} \\ &= \frac{f(D_W^T|\tau)\pi_\tau(\tau)}{f(D_W^T|\tau\le T_1^W)\pi_\tau(\tau\le T_1^W) + \sum_{t=T_2^W}^{T_W^W} f(D_W^T|\tau=t)\pi_\tau(\tau=t) + f(D_W^T|\tau>T)\pi_\tau(\tau>T)}.\end{aligned} \tag{32}$$

Unlike the full posterior, the windowed posterior treats data outside of the window as missing, resulting in the consideration of at most $W + 1$ specific change points at any time. This allows the method to maintain a fixed computational cost at each time of the process. Change points in the distant past, outside the current window, are weighted according to the prior in the same way as future change points. So while only the posterior utilizes only a portion of the data, its support covers the entire space of possible change points. Discarding data will still hinder the estimation of an exact change point outside of the current window and can prevent evidence of minor changes from accumulating. In practice, however, windowing would only be utilized for online process monitoring, and estimation of a change point can always be carried out offline using the full posterior after the process is stopped. In Section 4, we show the performance of the proposed method for $W = 10$.

## 4 | SIMULATION

We compared the performance of the proposed WBB method to the LRT method of Chicken et al[20] through Monte Carlo simulation. The comparison is of particular interest as both methods use wavelets to address the phase II profile monitoring problem in a change-point framework—one from a Bayesian perspective, the other from a frequentist one. The methods operate under the same assumptions: iid Gaussian errors and a single and sustained change in the functional structure of profiles. Both methods require that $f^0$ and $\sigma_\varepsilon$ be known or estimated prior to online monitoring. In order to compare WBB and LRT, we set $f^0(x) = 0$ and $\sigma_\varepsilon = 1$, without loss of generality. Using these values did not influence the comparison as both methods only monitor the differences between observed profiles and an in-control template—not the profiles themselves.

The parameters of the simulation emulate those used in Chicken et al,[20] where LRT outperformed the wavelet-based methods of Fan,[39] Jin and Shi,[17] and Jeong et al.[40] We considered profiles with $n = 512$ observations. When applying the DWT required by both methods, we used the Haar wavelet basis to obtain a complete decomposition (ie, $j_0 = 0$). For the implementation of WBB, we

**TABLE 1** A calibration summary from 100 replications of an in-control monitoring scenario[a]

| Method | Prior | $W$ | ARL | SDRL | UCL |
|---|---|---|---|---|---|
| WBB | Normal | Full | 211.26 | 183.61 | 0.170 |
| WBB | Normal | 10 | 216.30 | 163.30 | 0.200 |
| WBB | Laplace | Full | 202.95 | 173.73 | 0.170 |
| WBB | Laplace | 10 | 205.85 | 129.43 | 0.250 |
| LRT | N/A | N/A | 205.96 | 244.40 | 0.033 |

Abbreviations: ARL, average run length; LRT, likelihood ratio test; SDRL, standard deviation of run length; UCL, upper control limit; WBB, wavelet-based Bayesian.

[a]For each method, the summary includes the average and standard deviation of the 100 simulated run lengths as well as the upper control limit that was used.

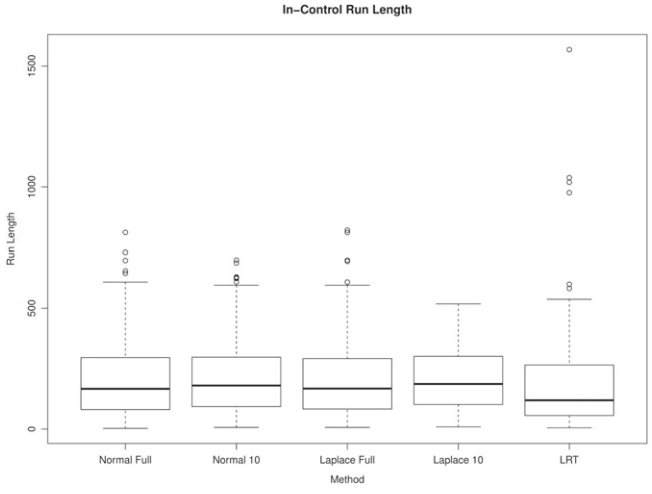


**FIGURE 1** The distribution of run length for each method from 100 simulated replications of an in-control monitoring scenario

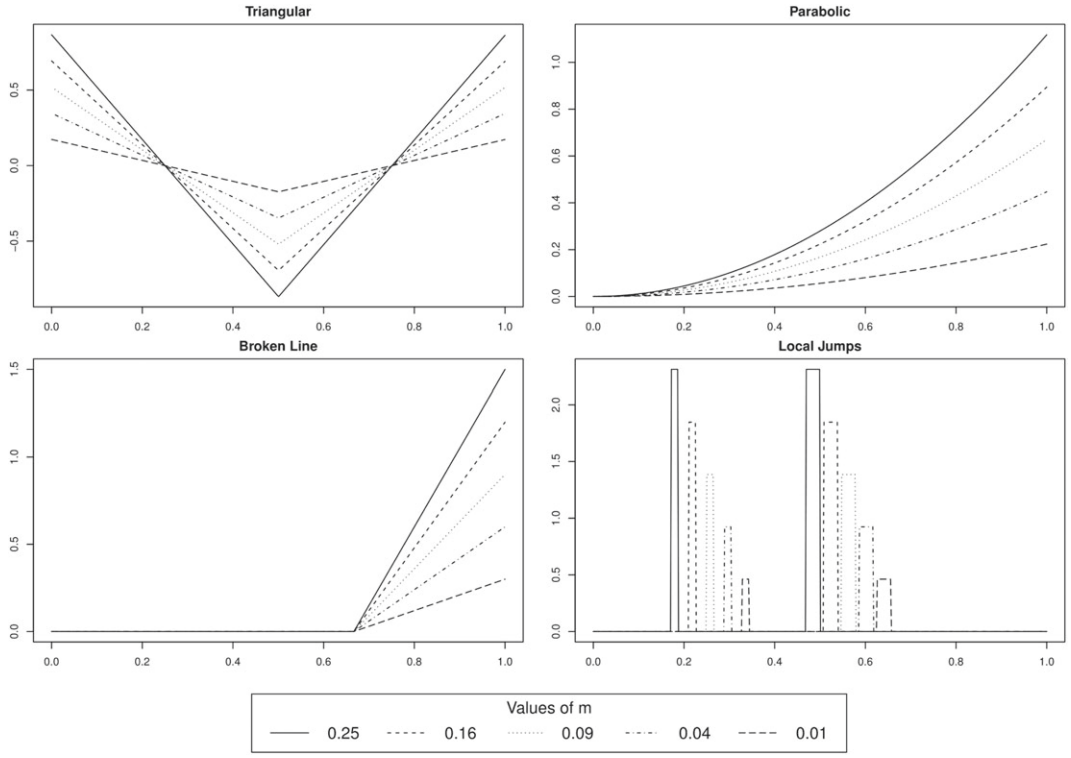


**FIGURE 2** The various forms and magnitudes of the introduced functional change $g$

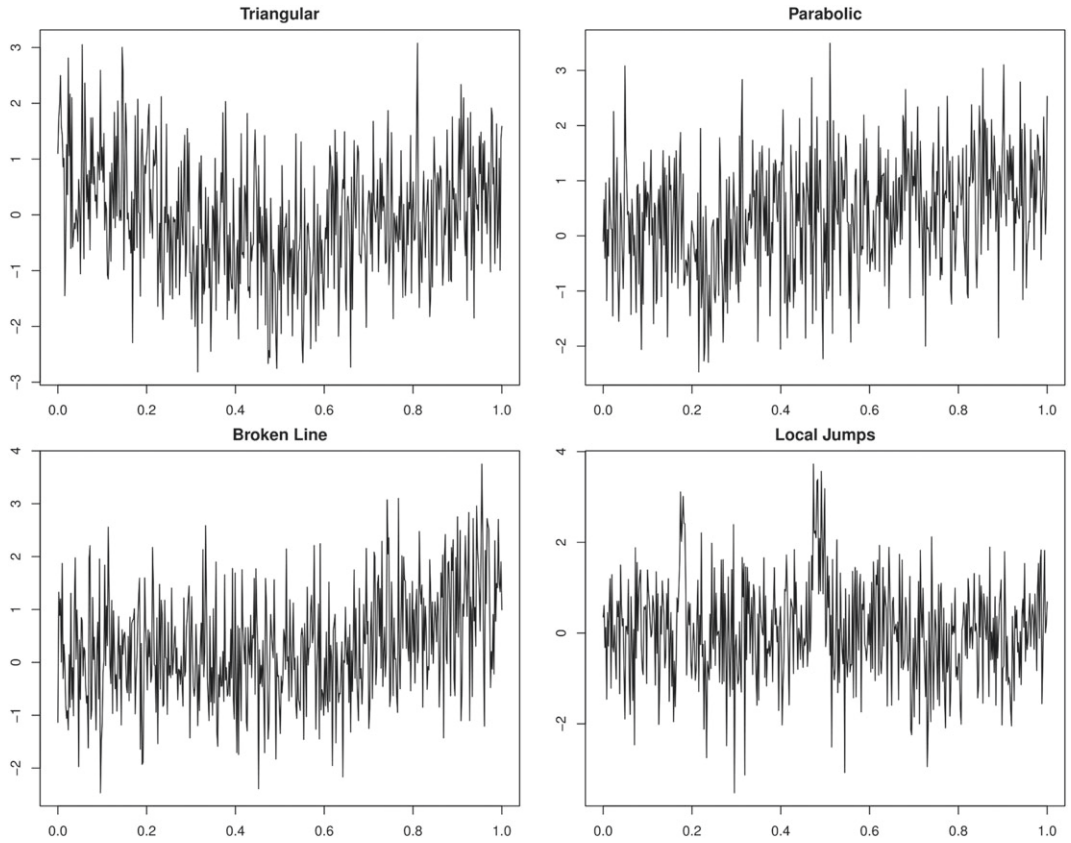


**FIGURE 3** The various forms of the functional change $g$ with magnitude $m = 0.25$ after the addition of noise

set $\omega = 0.05$ and mapped the posterior median threshold to the universal threshold, as described in Section 3.3, resulting in $s = 1.07$ for the normal mixture and $s = 1.31$ for the Laplace mixture. We considered both the full WBB

**TABLE 2** The ARL and (SDRL) for each method from 100 replications of each monitoring scenario when $\tau = 1$

| Method | $m$ | Horizontal Line | Triangular | Parabolic | Broken Line | Local Jumps |
|---|---|---|---|---|---|---|
| Normal Full | 0.01 | 4.16 (2.31) | 10.71 (4.67) | 6.37 (3.32) | 8.36 (3.92) | 11.44 (5.21) |
| | 0.04 | 1.32 (0.47) | 2.63 (0.77) | 1.82 (0.69) | 2.19 (0.73) | 2.96 (1.01) |
| | 0.09 | 1.02 (0.14) | 1.49 (0.52) | 1.06 (0.24) | 1.27 (0.45) | 1.65 (0.52) |
| | 0.16 | 1.00 (0.00) | 1.05 (0.22) | 1.00 (0.00) | 1.00 (0.00) | 1.06 (0.24) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| Normal 10 | 0.01 | 4.43 (2.37) | 12.58 (7.45) | 6.78 (3.57) | 9.03 (4.65) | 13.63 (7.91) |
| | 0.04 | 1.35 (0.50) | 2.71 (0.78) | 1.89 (0.69) | 2.21 (0.78) | 2.99 (1.00) |
| | 0.09 | 1.02 (0.14) | 1.51 (0.52) | 1.07 (0.26) | 1.30 (0.46) | 1.69 (0.51) |
| | 0.16 | 1.00 (0.00) | 1.05 (0.22) | 1.00 (0.00) | 1.00 (0.00) | 1.08 (0.27) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| Laplace Full | 0.01 | 4.11 (2.13) | 9.84 (4.02) | 6.21 (2.93) | 8.07 (3.41) | 10.79 (4.50) |
| | 0.04 | 1.25 (0.44) | 2.61 (0.79) | 1.79 (0.69) | 2.19 (0.76) | 2.96 (0.95) |
| | 0.09 | 1.01 (0.10) | 1.44 (0.52) | 1.06 (0.24) | 1.23 (0.42) | 1.64 (0.54) |
| | 0.16 | 1.00 (0.00) | 1.04 (0.20) | 1.00 (0.00) | 1.00 (0.00) | 1.04 (0.20) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| Laplace 10 | 0.01 | 4.60 (2.38) | 11.74 (6.54) | 6.69 (3.11) | 9.30 (4.65) | 13.11 (7.33) |
| | 0.04 | 1.33 (0.49) | 2.70 (0.78) | 1.86 (0.70) | 2.31 (0.79) | 3.08 (1.06) |
| | 0.09 | 1.02 (0.14) | 1.47 (0.52) | 1.06 (0.24) | 1.27 (0.45) | 1.69 (0.53) |
| | 0.16 | 1.00 (0.00) | 1.05 (0.22) | 1.00 (0.00) | 1.00 (0.00) | 1.07 (0.26) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| LRT | 0.01 | 53.40 (49.41) | 122.93 (109.31) | 107.13 (102.49) | 123.56 (101.73) | 114.20 (106.98) |
| | 0.04 | 2.42 (1.61) | 7.36 (6.00) | 5.48 (4.27) | 8.46 (6.08) | 11.41 (8.97) |
| | 0.09 | 1.13 (0.39) | 1.56 (0.87) | 1.35 (0.63) | 1.62 (0.92) | 2.29 (1.51) |
| | 0.16 | 1.00 (0.00) | 1.02 (0.14) | 1.02 (0.14) | 1.05 (0.22) | 1.05 (0.22) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |

Abbreviations: ARL, average run length; LRT, likelihood ratio test; SDRL, standard deviation of run length.

model and the windowed WBB model with $W = 10$. Hereafter, we may identify WBB implementations simply by the combination of prior and window size (eg, Normal 10, Laplace Full).

## 4.1 | Calibration and in-control performance

Before comparing the ability to detect changes, we considered each method's stability when monitoring an in-control process. The in-control statistical properties of a monitoring method allow us to control false alarms and make a fair comparison of each method's power.

We estimated the in-control ARL of each method using 100 replications of the monitoring scenario. Each method was calibrated by choosing a UCL to result in an in-control ARL of approximately 200. Given this target in-control ARL, we set $p = 1/200$ to match the mean of the prior on $\tau$. The chosen UCL was the smallest value for which the ARL of the simulated run lengths was at least 200. A calibration summary for all methods is given in Table 1. The table includes the standard deviation of run length (SDRL). A graphical comparison of the distributions of in-control run length for each method is shown in Figure 1.

From Table 1 and Figure 1, we see an important benefit of WBB. While WBB and LRT are both calibrated to result in the same in-control ARL, WBB attains that average with much less variability. LRT has a much higher SDRL and extreme run length values that far exceed those of any WBB setting. Given the presence of these extreme values in the distribution of run length, the LRT method will result in a higher proportion of early false alarms in order to match desired in-control ARL.

This discrepancy highlights a shortcoming of using ARL for in-control calibration. With increasingly complex process monitoring methods, error rates change in time and are often difficult to obtain. This prompts the use of ARL to control false alarms. However, without consideration of the stability with which that ARL is attained, a monitoring method may have undesirable in-control properties. In general, the in-control run length of a method is a random variable with a heavy-tailed, right-skewed distribution to which the mean is sensitive. A direction for future work would be to replace ARL with a robust characterization of run length.

**TABLE 3** The ARL and (SDRL) for each method from 100 replications of each monitoring scenario when $\tau = 20$

| Method | $m$ | Horizontal Line | Triangular | Parabolic | Broken Line | Local Jumps |
|---|---|---|---|---|---|---|
| Normal Full | 0.01 | 3.88 (1.82) | 10.47 (5.20) | 5.86 (2.74) | 8.01 (3.75) | 11.50 (5.08) |
| $p_{FA} = 0.05$ | 0.04 | 1.34 (0.52) | 2.66 (0.88) | 1.77 (0.66) | 2.20 (0.74) | 2.85 (0.97) |
| | 0.09 | 1.01 (0.10) | 1.56 (0.50) | 1.08 (0.27) | 1.26 (0.46) | 1.68 (0.53) |
| | 0.16 | 1.00 (0.00) | 1.01 (0.10) | 1.00 (0.00) | 1.02 (0.14) | 1.11 (0.31) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| Normal 10 | 0.01 | 3.93 (1.83) | 11.14 (6.52) | 5.92 (2.61) | 8.46 (4.32) | 12.27 (7.35) |
| $p_{FA} = 0.04$ | 0.04 | 1.38 (0.55) | 2.74 (0.89) | 1.79 (0.67) | 2.22 (0.72) | 2.89 (0.98) |
| | 0.09 | 1.02 (0.14) | 1.58 (0.50) | 1.09 (0.29) | 1.28 (0.47) | 1.71 (0.52) |
| | 0.16 | 1.00 (0.00) | 1.02 (0.14) | 1.00 (0.00) | 1.02 (0.14) | 1.11 (0.31) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| Laplace Full | 0.01 | 3.74 (1.72) | 9.67 (4.57) | 5.63 (2.50) | 7.76 (3.28) | 10.88 (4.49) |
| $p_{FA} = 0.05$ | 0.04 | 1.32 (0.51) | 2.63 (0.85) | 1.77 (0.66) | 2.22 (0.70) | 2.89 (0.95) |
| | 0.09 | 1.01 (0.10) | 1.47 (0.50) | 1.06 (0.24) | 1.24 (0.45) | 1.66 (0.54) |
| | 0.16 | 1.00 (0.00) | 1.01 (0.10) | 1.00 (0.00) | 1.02 (0.14) | 1.09 (0.29) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| Laplace 10 | 0.01 | 3.97 (1.75) | 10.57 (5.68) | 6.01 (2.49) | 7.95 (3.06) | 11.54 (6.20) |
| $p_{FA} = 0.02$ | 0.04 | 1.38 (0.55) | 2.72 (0.87) | 1.83 (0.68) | 2.36 (0.67) | 2.99 (0.97) |
| | 0.09 | 1.01 (0.10) | 1.52 (0.50) | 1.10 (0.30) | 1.29 (0.48) | 1.73 (0.53) |
| | 0.16 | 1.00 (0.00) | 1.01 (0.10) | 1.00 (0.00) | 1.02 (0.14) | 1.11 (0.31) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| LRT | 0.01 | 59.30 (50.47) | 121.99 (93.74) | 101.48 (86.20) | 129.53 (94.93) | 132.82 (97.58) |
| $p_{FA} = 0.08$ | 0.04 | 2.62 (1.80) | 8.76 (6.45) | 4.97 (3.40) | 9.31 (7.03) | 13.22 (10.04) |
| | 0.09 | 1.21 (0.48) | 1.73 (0.98) | 1.44 (0.72) | 1.72 (0.96) | 2.39 (1.45) |
| | 0.16 | 1.00 (0.00) | 1.00 (0.00) | 1.03 (0.22) | 1.06 (0.28) | 1.08 (0.27) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.01 (0.10) |

Abbreviations: ARL, average run length; LRT, likelihood ratio test; SDRL, standard deviation of run length.

The difference in the variability of a full WBB method and its windowed counterpart may be explained by the influence of the prior on the change point. The windowed models utilize less data and are therefore subject to greater influence by the prior. This removal of randomness results in more stable monitoring. It also shows the possibility of stably calibrating to any desired in-control ARL.

## 4.2 | Out-of-control performance and estimation

We used the UCL values from Section 4.1 to compare each method's ability to detect varying functional changes $g$ of different magnitudes $m$. The test functions are those used in Chicken et al.[20] We define $m$ to be the integrated squared error introduced by $g$ and considered magnitudes ranging from 0.01 to 0.25. We considered five different forms of $g$: a uniformly additive shift and the four different functions shown in Figure 2. The test functions represent changes of not only different forms but with different levels of localization. Figure 3 shows noise-contaminated versions of the test functions when $m = 0.25$, the largest magnitude considered. We see that even for the largest magnitude, the functional changes are small relative to the level of noise and may be difficult to detect visually.

Tables 2 to 3 summarize the detection performance of each method from 100 replications of each setting. Table 2 sets $\tau = 1$ (ie, the process is never in control), and Table 3 sets $\tau = 20$. The ARL and corresponding SDRL values refer to the number of out-of-control profiles observed before a change was signaled. For $\tau > 1$, if a method signaled a false alarm prior to time $\tau$, the monitoring method was reset and allowed to resume monitoring at the next time point, now closer to the change point than when the replicate originally began. The proportion of run lengths that ended in a false alarm is reported as $P_{FA}$. All methods excel at detecting changes of the largest magnitudes, often stopping the process after a single out-of-control observation. The starkest differences are seen for lower magnitude changes where all WBB methods significantly outperform LRT. This can be explained by the ability of the WBB method to incorporate frequency and location information across profiles. While LRT uses all coefficients in its control statistic, it only uses the sum of the squared

**TABLE 4** The mean and (RMSE) of the estimated $\hat{\tau}$ in each method from 100 replications of each monitoring scenario when $\tau = 1$

| Method | $m$ | Horizontal Line | Triangular | Parabolic | Broken Line | Local Jumps |
|---|---|---|---|---|---|---|
| Normal Full | 0.01 | 1.82 (1.80) | 2.63 (3.35) | 2.61 (2.98) | 2.31 (2.52) | 3.39 (3.93) |
| | 0.04 | 1.02 (0.14) | 1.04 (0.20) | 1.02 (0.14) | 1.02 (0.14) | 1.06 (0.24) |
| | 0.09 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| | 0.16 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| Normal 10 | 0.01 | 1.82 (1.80) | 6.39 (8.39) | 2.99 (3.38) | 3.53 (4.20) | 7.82 (10.00) |
| | 0.04 | 1.02 (0.14) | 1.04 (0.20) | 1.02 (0.14) | 1.02 (0.14) | 1.05 (0.22) |
| | 0.09 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| | 0.16 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| Laplace Full | 0.01 | 1.62 (1.55) | 2.59 (3.14) | 2.18 (2.34) | 2.08 (2.21) | 3.03 (3.53) |
| | 0.04 | 1.03 (0.17) | 1.04 (0.20) | 1.04 (0.20) | 1.04 (0.20) | 1.08 (0.32) |
| | 0.09 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| | 0.16 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| Laplace 10 | 0.01 | 1.64 (1.54) | 5.08 (6.73) | 2.43 (2.83) | 3.48 (4.25) | 6.68 (8.58) |
| | 0.04 | 1.03 (0.17) | 1.05 (0.22) | 1.04 (0.20) | 1.04 (0.20) | 1.06 (0.28) |
| | 0.09 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| | 0.16 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| LRT | 0.01 | 37.57 (55.20) | 79.68 (115.35) | 82.30 (129.51) | 87.40 (132.81) | 82.94 (120.86) |
| | 0.04 | 1.74 (1.40) | 4.52 (5.66) | 3.53 (4.26) | 5.28 (6.65) | 6.43 (8.30) |
| | 0.09 | 1.04 (0.20) | 1.29 (0.77) | 1.15 (0.46) | 1.34 (0.76) | 1.56 (1.26) |
| | 0.16 | 1.00 (0.00) | 1.02 (0.14) | 1.01 (0.10) | 1.02 (0.14) | 1.03 (0.17) |
| | 0.25 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |

coefficients, which discards the frequency and location information that those coefficients carry. Without carrying that information forward in some way, a method will struggle to quickly detect changes whose magnitude is small relative to the magnitude of the noise.

Tables 4-5 summarize the performance of change-point estimation from the same simulation replications used in the detection comparison. For each method, we report the mean estimate $\hat{\tau}$ after the process was stopped and its root mean squared error (RMSE). The results are in line with what we might expect from the detection simulation. LRT, which struggles to detect the smallest changes, also struggles to estimate their location. For WBB, the process shows more promising results for pinpointing the location of the change. Even in cases when it took WBB longer than 10 observations to detect the change, the estimate of the change point was much closer to the true change point than the stopping time.

As previously mentioned, the WBB results in this section show a significant improvement over initial variations of the framework.[27] These improvements are largely reflected in the ARL values of small magnitude changes for which the previous WBB approach performed similarly to LRT. The primary differences between the current implementation and the previous iteration include the structure of the prior on the unknown wavelet coefficients across resolution levels and how the parameters of that prior are determined. The stability of prespecified global values results in better performance than previous attempts to improve flexibility by estimating level-dependent hyperparameters from the data. Given that the simulation settings used in this paper match those used in the comparison of wavelet-based methods in Chicken et al,[20] the proposed WBB method would also be an improvement over the wavelet-based methods of Fan,[39] Jin and Shi,[17] and Jeong et al. [40]

## 5 | DISCUSSION

In this paper, we proposed a general WBB model for the phase II monitoring of nonlinear profiles. Under a variety of simulated settings, the proposed method demonstrated success in quickly detecting different functional changes and estimating the position of the change in the sequence. The proposed method is particularly suited for detecting changes that are minor relative to the noise associated with the process.

**TABLE 5** The mean and (RMSE) of the estimated $\hat{\tau}$ in each method from 100 replications of each monitoring scenario when $\tau = 1$

| Method | $m$ | Horizontal Line | Triangular | Parabolic | Broken Line | Local Jumps |
|---|---|---|---|---|---|---|
| Normal Full | 0.01 | 20.70 (1.85) | 20.94 (2.27) | 21.39 (2.70) | 21.27 (2.65) | 22.30 (4.58) |
| | 0.04 | 19.92 (0.51) | 19.97 (0.48) | 19.87 (0.54) | 19.89 (0.71) | 19.96 (0.86) |
| | 0.09 | 19.98 (0.14) | 19.96 (0.24) | 19.95 (0.22) | 19.94 (0.24) | 19.91 (0.39) |
| | 0.16 | 20.00 (0.00) | 19.99 (0.10) | 20.00 (0.00) | 20.00 (0.00) | 19.96 (0.20) |
| | 0.25 | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 19.98 (0.14) |
| Normal 10 | 0.01 | 20.48 (1.98) | 23.51 (6.44) | 20.92 (2.49) | 21.78 (3.92) | 25.30 (8.86) |
| | 0.04 | 19.92 (0.51) | 19.98 (0.47) | 19.87 (0.54) | 19.89 (0.71) | 19.96 (0.86) |
| | 0.09 | 19.98 (0.14) | 19.96 (0.24) | 19.95 (0.22) | 19.94 (0.24) | 19.92 (0.37) |
| | 0.16 | 20.00 (0.00) | 19.99 (0.10) | 20.00 (0.00) | 20.00 (0.00) | 19.96 (0.20) |
| | 0.25 | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 19.98 (0.14) |
| Laplace Full | 0.01 | 20.33 (1.57) | 20.69 (2.16) | 20.61 (2.17) | 20.82 (2.12) | 21.52 (3.42) |
| | 0.04 | 19.95 (0.48) | 19.92 (0.76) | 19.91 (0.52) | 19.90 (0.75) | 19.93 (0.88) |
| | 0.09 | 20.00 (0.00) | 19.96 (0.40) | 20.00 (0.00) | 19.98 (0.20) | 19.93 (0.36) |
| | 0.16 | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 19.98 (0.14) |
| | 0.25 | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) |
| Laplace 10 | 0.01 | 20.14 (1.62) | 22.68 (5.56) | 20.30 (1.87) | 20.88 (2.23) | 23.63 (6.96) |
| | 0.04 | 19.95 (0.48) | 19.92 (0.76) | 19.93 (0.48) | 19.92 (0.73) | 19.93 (0.88) |
| | 0.09 | 20.00 (0.00) | 19.96 (0.40) | 20.00 (0.00) | 19.99 (0.17) | 19.93 (0.36) |
| | 0.16 | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 19.98 (0.14) |
| | 0.25 | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) |
| LRT | 0.01 | 64.84 (64.66) | 107.61 (124.80) | 91.77 (105.08) | 116.18 (134.02) | 114.63 (134.94) |
| | 0.04 | 21.12 (2.01) | 25.15 (7.79) | 22.84 (4.20) | 24.32 (6.60) | 26.74 (10.72) |
| | 0.09 | 20.20 (0.51) | 20.43 (0.98) | 20.32 (0.75) | 20.43 (0.95) | 20.68 (1.22) |
| | 0.16 | 20.00 (0.00) | 20.00 (0.00) | 20.03 (0.22) | 20.03 (0.22) | 20.06 (0.24) |
| | 0.25 | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 20.00 (0.00) | 20.01 (0.10) |

Abbreviation: LRT, likelihood ratio test.

In future work, we are interested in exploring alternatives to windowing. Given our goal of detecting subtle changes, discarding data to maintain a certain level of computational complexity is counterproductive. This problem could possibly be solved through the construction of a posterior approximation that allows evidence of a change to build.

Other considerations for the method primarily focus on increasing its practical applicability. While common in the literature, the assumption of additive iid Gaussian noise will rarely be met. Considering a more flexible noise structure would result in a method more suited for real scenarios. Monitoring the process for changes in noise structure would address another common fault condition in SPC. Currently, the method requires knowledge of the in-control function and noise magnitude but does not incorporate the uncertainty that may be associated with those estimates.

The proposed method works well for the suggested hyperparameter values. The method's sensitivity to these settings (including hyperparameter values, calibration approach, and window size) is fully explored and reported in Varbanov.[41] We note that changing the proposed settings can have effects on performance, though generally not to the extent that would result in performance inferior to the other considered competitors.

The proposed method is unique in that it approaches the profile monitoring problem in a Bayesian framework without imposing restrictive assumptions on the form of the profiles or detectable functional changes. A Bayesian approach allows for the natural input of process knowledge and for an intuitive interpretation of the monitoring statistic. Wavelets allow the proposed method to flexibly monitor for a variety of profile changes, including the often irregular forms encountered in practice. By avoiding MCMC approximation of the posterior, the method incorporates these benefits of Bayesian analysis in an online monitoring procedure.

## ACKNOWLEDGEMENTS

This research was supported by the Office of the Secretary of Defense, Directorate of Operational Test and Evaluation, and the Test Resource Management Center under the Science of Test research program. Special thanks to both Dr Catherine Warner, OSD DOT&E, and Mr George Rumford, TRMC, for the sponsorship of the research program.

## DISCLAIMER

The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

## REFERENCES


1. Gardner MM, Lu JC, Gyurcsik RS, et al. Equipment fault detection using spatial signatures. *IEEE Trans Compon Packag Manuf Technol Sec C*. 1997;20:294-303.
2. Jin J, Shi J. Feature-preserving data compression of stamping tonnage information using wavelets. *Technometrics*. 1999;41:327-339.
3. Kang L, Albin SL. On-line monitoring when the process yields a linear profile. *J Qual Technol*. 2000;32:418-426.
4. Colosimo B, Semerado Q, Pacella M. Statistical process control for geometric specifications: on the monitoring of roundness profiles. *J Qual Technol*. 2008;40:1-18.
5. Amiri A, Jensen WA, Kazemzadeh RB. A case study on monitoring polynomial profiles in the automotive industry. *Qual Reliab Eng Int*. 2010;26:509-520.
6. Woodall W, Spitzner D, Montgomery D, Gupta S. Using control charts to monitor process and product quality profiles. *J Qual Technol*. 2004;36:309-320.
7. Woodall WH. Current research on profile monitoring. *Revista Producao*. 2007;17:420-425.
8. Woodall WH, Montgomery DC. Some current directions in the theory and application of statistical process monitoring. *J Qual Technol*. 2014;46:78-94.
9. Noorossana R, Saghaei A, Amiri A. *Statistical Analysis of Profile Monitoring*. Hoboken, NJ: Wiley; 2012.
10. Qiu P. *Introduction to Statistical Process Control*. Boca Raton, FL: CRC Press; 2014.
11. Chang SI, Yadama S. Statistical process control for monitoring nonlinear profiles using wavelet filtering and B-spline approximation. *Int J Protein Res*. 2010;48:1049-1068.
12. Williams JD, Woodall WH, Birch JB. Statistical monitoring of nonlinear product and process quality profiles. *Qual Reliab Int*. 2007;23:925-941.
13. Qiu P, Zou C, Wang Z. Nonparametric profile monitoring by mixed effects modeling. *Technometrics*. 2010;52(3):265-277.
14. Qiu P, Zou C. Control chart for monitoring nonparametric profiles with arbitrary design. *Technometrics*. 2010;52(3):265-277.
15. Shiau J, Huang H, Tsai M. Monitoring nonlinear profiles with random effects by nonparametric regression. *Commun Stat - Theory Methods*. 2009;38:1664-1679.
16. Ding Y, Zeng L, Zhou S. Phase I analysis for monitoring nonlinear profiles in manufacturing processes. *J Qual Technol*. 2006;38:199-216.
17. Jin J, Shi J. Automatic feature extraction of waveform signals for in-process diagnostic performance improvement. *J Intell Manuf*. 2001;12:140-145.
18. Lada E, Lu J-C, Wilson J. A wavelet-based procedure for process fault detection. *IEEE Trans Semicond Manuf*. 2002;15:79-90.
19. Zhou S, Sun B, Shi J. An SPC monitoring system for cycle-based waveform signal using haar transform. *IEEE Trans Autom Sci Eng*. 2006;3:60-72.
20. Chicken E, Pignatiello JJ, Simpson JR. Statistical process monitoring of nonlinear profiles using wavelets. *J Qual Technol*. 2009;41:198-212.
21. Paynabar K, Jin J. Characterization of non-linear profile variations using mixed-effect models and wavelets. *IIE Trans*. 2011;43:275-290.
22. Lee JJ, Hur Y, Kim SH, Wilson JR. Monitoring nonlinear profiles using a wavelet-based distribution-free CUSUM chart. *Int J Prod Res*. 2012;50(22):6574-6594.
23. McGinnity K, Chicken E, Pignatiello JJ. Nonparametric change-point estimation for sequential nonlinear profile monitoring. *Qual Reliab Eng Int*. 2015;31:57-73.
24. Abbas T, Qian Z, Ahmad S, Riaz M. On monitoring of linear profiles using Bayesian methods. *Comput Ind Eng*. 2016;94:245-268.
25. Abbas T, Qian Z, Ahmad S, Riaz M. Bayesian monitoring of linear profile monitoring using DEWMA charts. *Qual Reliab Eng Int*. 2017;33:1783-1812.
26. Park C, Shrivastava AK. Multimode geometric-profile monitoring with correlated image data and its application to nanoparticle self-assembly processes. *J Qual Technol*. 2014;46:216-233.
27. Varbanov R, Chicken E, Linero A, Yang Y. Wavelet-based Bayesian profile monitoring. In: IIE Annual Conference Proceedings; 2017; Miami, FL:621-626.
28. Ogden RT. *Essential Wavelets for Statistical Applications and Data Analysis*. Boston: Birkhauser; 1997.
29. Vidakovic B. *Statistical Modeling by Wavelets*. New York: Wiley; 1999.
30. Daubechies I. *Ten Lectures on Wavelets*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics; 1992.
31. Mallat SG. Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Trans Am Math Soc*. 1989;315:69-89.
32. Abramovich F, Sapatinas T, Silverman BW. Wavelet thresholding via a Bayesian approach. *J R Stat Soc Ser B (Stat Methodol)*. 1998;60(4):725-749.
33. Silverman BW. Wavelets in statistics: beyond the standard assumptions. *Philos Trans R Soc A Math Phys Eng Sci*. 1999;357:2459-2473.
34. Johnstone IM, Silverman BW. Empirical Bayes selection of wavelet thresholds. *Ann Stat*. 2005;33(4):1700-1752.
35. Shiryaev AN. On optimum methods in quickest change detection problems. *Theory Probability Appl*. 1963;8:22-46.
36. Johnstone IM, Silverman BW. EbayesThresh: R programs for empirical Bayes thresholding. *J Stat Softw*. 2005;12(8):1-38.
37. Donoho DL, Johnstone IM. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*. 1994;81(3):425-455.
38. Montgomery DC. *Statistical Quality Control*, 7th Edition. Hoboken, NJ: Wiley Global Education; 2012.
39. Fan J. Tests of significance based on wavelet thresholding and Neyman's truncation. *J Am Stat Assoc*. 1996;91:674-688.
40. Jeong MK, Lu JC, Wang N. Wavelet-based SPC procedure for complicated functional data. *Int J Prod Res*. 2006;44:729-744.
41. Varbanov R. *Wavelet-Based Bayesian Approaches to Sequential Profile Monitoring*. Florida: Florida State University, Dissertation; 2018.


## AUTHOR BIOGRAPHIES

**Roumen Varbanov** is a data scientist at Liberty Mutual Insurance in Boston, MA. He currently works on the Marketing, Direct, and Digital Analytics team within Global Retail Markets–United States.

**Eric Chicken** is a Professor in the Department of Statistics at Florida State University. His research interests include statistical process control, nonparametric regression and density estimation, statistical estimation via wavelets, and Bayesian estimation.

**Antonio Linero** is an Assistant Professor in the Department of Statistics at Florida State University. His research broadly focuses on developing flexible Bayesian methods, including appropriate methods for complex longitudinal data and model selection tools within the Bayesian nonparametric framework for high dimensional problems.

**Yun Yang** is an Assistant Professor in the Department of Statistics at University of Illinois Urbana-Champaign. His research interests lie broadly in machine learning, scalable Bayes inference, and theoretical foundations of high-dimensional problems.

**How to cite this article:** Varbanov R, Chicken E, Linero A, Yang Y. A Bayesian approach to sequential monitoring of nonlinear profiles using wavelets. *Qual Reliab Engng Int*. 2018;1–15. https://doi.org/10.1002/qre.2409