



Sequential Analysis

Design Methods and Applications

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/lsqa20>

S3T: A score statistic for spatiotemporal change point detection

Junzhuo Chen, Seong-Hee Kim & Yao Xie

To cite this article: Junzhuo Chen, Seong-Hee Kim & Yao Xie (2020) S3T: A score statistic for spatiotemporal change point detection, Sequential Analysis, 39:4, 563-592, DOI: [10.1080/07474946.2020.1826796](https://doi.org/10.1080/07474946.2020.1826796)

To link to this article: <https://doi.org/10.1080/07474946.2020.1826796>



Published online: 27 Jan 2021.



Submit your article to this journal 



Article views: 22



View related articles 



View Crossmark data 



S³T: A score statistic for spatiotemporal change point detection

Junzhuo Chen, Seong-Hee Kim, and Yao Xie

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

ABSTRACT

We present an efficient score statistic, called the S³T statistic, to detect the emergence of a spatially and temporally correlated signal from either fixed-sample or sequential data. The signal may cause a mean shift and/or a change in the covariance structure. The score statistic can capture both the spatial and temporal structures of the change and hence is particularly powerful in detecting weak signals. The score statistic is computationally efficient and statistically powerful. Our main theoretical contribution is accurate analytical approximations to the false alarm rate of the detection procedures, which can be used to calibrate the threshold analytically. Numerical experiments on simulated and real data, as well as a case study of water quality monitoring using sensor networks, demonstrate the good performance of our procedure.

ARTICLE HISTORY

Received 11 January 2019
Revised 10 November 2019,
2 February 2020
Accepted 17 April 2020

KEYWORDS

Change point detection;
false alarm control; score
statistics; spatiotemporal
data analysis

SUBJECT CLASSIFICATIONS

Primary 62L10; Secondary
62G10; 62G32

1. INTRODUCTION

Detection of the emergence of a signal in a noisy background arises in many multisensor spatiotemporal surveillance applications. When the monitored process is in control, sensors observe noise. When the monitored process is out of control, a signal emerges in the noise. A variety of applications possess particular spatial and temporal correlation structures. One application is an environmental sensor network, which is used to monitor river systems to detect a contaminant hazard (Kim et al. 2017). When the signal emerges, observations from sensors may have a time-varying mean and spatiotemporal correlation structures due to water flow.

Exploiting spatiotemporal structures of the change is crucial for detecting weaker signals. However, most existing methods only capture either spatial correlations (Healy 1987; Crosier 1988; Jiang et al. 2011; Lee et al. 2014, Lee, Goldsman, and Kim 2015) or temporal correlations (Xie and Siegmund 2012). It is still unclear how to jointly capture the spatial and temporal correlations in detection statistics. Moreover, computational complexity is often a concern when we try to jointly model spatial and temporal correlation, especially when there are a large number of sensors, which leads to high-dimensional observations. In particular, one issue with the likelihood ratio statistic is that one

CONTACT Yao Xie yao.c.xie@gmail.com H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Dr NW, Atlanta, GA, 30332-0205, USA.
Recommended By S. Lee.

© 2020 Taylor & Francis Group, LLC

has to invert the sample covariance matrix, which can be computationally expensive and numerically unstable. An alternative to the likelihood ratio statistic is the score statistic, which can sometimes lead to a simpler test statistic. When the hypothesis involves a univariate parameter, the score test is the locally most powerful test (Rao and Poti 1946).

In this article, we propose a new efficient score statistic for spatial–temporal change point detection, which we call the S^3T statistic. The S^3T statistic can capture both spatial and temporal correlations of the signal. Hence, it can react quickly to a change in the mean and/or the spatiotemporal covariance. The score statistic is computationally efficient. By avoiding the inversion of the sample covariance matrix, the S^3T statistic has computation complexity $O(p^3)$, where p is the dimensionality of the observations, whereas the likelihood ratio statistic has $O(p^3N^3)$ complexity, which grows with the time horizon N . Our main theoretical contributions are accurate analytic approximations for the false alarm rate in the offline case and the average run length in the online case. Using the theoretical approximations, calibrating thresholds to control the false alarm rate of our procedure can be done efficiently without resorting to onerous numerical simulations. This is quite useful in practice, because the simulation-based approach to calibrate thresholds can be quite time consuming, due to the search over the unknown parameters when evaluating the detection statistics. For scalar observations, our statistic S^3T reduces to the score detector considered in Xie and Siegmund (2012). Our work provides a novel extension of Xie and Siegmund (2012) for multidimensional observations when there are both spatial and temporal correlations.

The rest of the article is organized as follows. Section 2 formulates the problem. Section 3 presents detection statistics for both offline and online change point detection. Section 4 presents theoretical approximations for the false alarm rate in the offline case and average run length in the online case. Section 5 contains numerical examples for simulated data and real data, as well as a case study of water quality monitoring. Finally, Section 6 concludes the article followed by Appendices.

2. PROBLEM FORMULATION

Consider a sequence of samples $\mathbf{y}_\ell \in \mathbb{R}^p, \ell = 1, 2, \dots, N$, where p is the dimension, and N is the sample size, which is fixed in the offline setting and grows in the online setting. We assume that under the null hypothesis, $\{\mathbf{y}_\ell\}$ forms a series of independent and identically distributed normal random vectors with spatial correlation caused by, for instance, sensor measurement errors or background noises from the environment. At an unknown time k , $1 \leq k \leq N - 1$, which corresponds to the unknown change point, a signal emerges on top of the noise. The change may alter not only the mean of $\{\mathbf{y}_\ell\}$ but also the spatiotemporal correlation structure, which we will explain in more details.

First, consider an offline setting, where the goal is to detect a change in retrospect from the samples. Formally, this can be formulated as the following hypothesis test:

$$\begin{aligned} H_0 : & \quad \mathbf{y}_\ell = \mathbf{w}_\ell, & \ell = 1, 2, \dots, N, \\ H_1 : & \quad \begin{cases} \mathbf{y}_\ell = \mathbf{w}_\ell, & \ell = 1, 2, \dots, k, \\ \mathbf{y}_\ell = \mathbf{x}_\ell + \mathbf{w}_\ell, & \ell = k + 1, \dots, N, \end{cases} \end{aligned} \quad (2.1)$$

where $\mathbf{w}_\ell \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \Sigma)$ and Σ is the spatial covariance matrix of the noise. We assume that before the change, the samples have no temporal correlation. This is reasonable because we often have enough reference data before the change to estimate and then remove the temporal correlation.

Below we describe a model for the signal $\{\mathbf{x}_\ell\}$ after the change has occurred. The signal can be spatially and temporally correlated. We capture the temporal correlation using multivariate time series models. Two examples are the first-order vector autoregressive VAR(1) model (Brockwell and Davis 1991),

$$\mathbf{x}_\ell = (1 - \theta)\boldsymbol{\mu} + \theta\mathbf{x}_{\ell-1} + \boldsymbol{\epsilon}_\ell, \quad \ell = 1, 2, \dots, \quad (2.2)$$

where $\theta \in \mathbb{R}$, $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}_\ell]$ and $\boldsymbol{\epsilon}_\ell$ is the process noise, and the VARMA(1, 1) model, given by

$$\mathbf{x}_{\ell+1} + \phi\mathbf{x}_\ell = (1 + \phi - \eta)\boldsymbol{\mu} + \eta\boldsymbol{\epsilon}_\ell + \boldsymbol{\epsilon}_{\ell+1}, \quad \ell = 1, 2, \dots,$$

where the parameters are $\eta \in \mathbb{R}$ and $\phi \in \mathbb{R}$. Models with higher orders can also be used if necessary.

We capture the spatial correlation of the signal using standard spatial correlation models (Gaetan and Guyon 2010). Denote $\text{Var}[\mathbf{x}_\ell] = \gamma\mathbf{\Lambda} \in \mathbb{R}^{p \times p}$, where $\mathbf{\Lambda}$ is the spatial correlation matrix of the signal $\mathbf{x}_\ell$, and $\gamma \geq 0$ is the magnitude (assuming that the model is stationary and $\text{Var}[\mathbf{x}_\ell]$ does not change over ℓ). Note that the variance of the signal $\text{Var}[\mathbf{x}_\ell]$ depends on the variance of the process noise, $\text{Var}[\boldsymbol{\epsilon}_\ell]$. Here we assume that the *structure* of $\mathbf{\Lambda}$ is known but the parameter γ may be unknown. This is a common practice, because once a spatial correlation model is assumed, $\mathbf{\Lambda}$ is usually specified by the location of the samples and some unknown parameters. In particular, each entry of the spatial covariance $\mathbf{\Lambda}$ is determined by a correlation function, $C(d|\rho)$, of the distance d between two samples (sensors) and is parameterized by ρ . Let $\mathbb{1}\{A\}$ denote an indicator function, which takes a value of 1 when the event A is true and 0 otherwise. Several commonly used correlation functions are the following:

- (i) Spherical model (Lee et al. 2014):

$$C(d|\rho) = 1 - \mathbb{1}\{d = 0\} + \rho \mathbb{1}\{d = 1\} + \frac{\rho}{2} \mathbb{1}\{d = \sqrt{2}\}, \quad \rho \in [0, 1], \quad (2.3)$$

- (ii) Exponential model (Gaetan and Guyon 2010):

$$C(d|\rho) = 1 - \mathbb{1}\{d = 0\} + e^{-d/\rho} \mathbb{1}\{d > 0\}, \quad \rho > 0,$$

- (iii) Matérn model (Gaetan and Guyon 2010):

$$C(d|\rho) = 1 - \mathbb{1}\{d = 0\} + \frac{1}{2^{v-1}\Gamma(v)} \left(\sqrt{2}v^{1/2}d/\rho \right)^v K_v \left(\sqrt{2}v^{1/2}d/\rho \right) \mathbb{1}\{d > 0\},$$

$$\rho > 0,$$

where $\rho > 0$, v is the order of the Matérn model that determines the degree of smoothness of the correlation function, $\Gamma(\cdot)$ is the gamma function, and $K_v(\cdot)$ is the modified Bessel function of the second kind (Abramowitz and Stegun 1970). Note that when $v = p + 0.5$, $p \in \mathbb{R}^+$, the Matérn model is a product of an exponential and a polynomial of

order p . When $\nu=0.5$, the Matérn model is equivalent to the exponential model. When $\nu \rightarrow \infty$, it converges to the squared exponential covariance function.

Now we derive our detection statistic. For an assumed change location k , let

$$\tau = N - k$$

denote the number of postchange samples. Define a vector by concatenating all samples after the assumed change point location k ,

$$\mathbf{y}_{(k+1:N)} = \left[\mathbf{y}_{k+1}^\top, \dots, \mathbf{y}_N^\top \right]^\top \in \mathbb{R}^{p\tau}, \quad (2.4)$$

where $a^\top$ denotes the transpose of a vector a . Define $\mathbf{x}_{(k+1:N)}$ and $\mathbf{w}_{(k+1:N)}$ similarly. Then after the change, we have

$$\mathbf{y}_{(k+1:N)} = \mathbf{x}_{(k+1:N)} + \mathbf{w}_{(k+1:N)}.$$

The covariance matrix of the concatenating observation vector consists of two terms that are due to the signal and the noise, respectively:

$$\text{Var}[\mathbf{y}_{(k+1:N)}] = \gamma \mathbf{V}_\tau(\theta) + \mathbf{\Sigma}_\tau,$$

where $\gamma \mathbf{V}_\tau(\theta) = \text{Var}[\mathbf{x}_{(k+1:N)}]$, $\mathbf{\Sigma}_\tau = \text{Var}[\mathbf{w}_{(k+1:N)}]$, and θ is the parameter related to temporal correlation, which we will specify next. The second term in the covariance matrix is given by

$$\mathbf{\Sigma}_\tau = \mathbf{I}_\tau \otimes \mathbf{\Sigma} \in \mathbb{R}^{p\tau \times p\tau}, \quad (2.5)$$

where $\mathbf{I}_\tau$ is a τ by τ identity matrix and $\otimes$ denotes the Kronecker product.

By concatenating the observation vectors as in (2.4), we can jointly model spatial and temporal correlation of the signal by one matrix $\mathbf{V}_\tau(\theta)$. For instance, for the VAR(1) model,

$$\mathbf{V}_\tau(\theta) = \mathbf{R}_\tau(\theta) \otimes \mathbf{\Lambda}, \quad (2.6)$$

where $\mathbf{R}_\tau(\theta) \in \mathbb{R}^{\tau \times \tau}$ and $[\mathbf{R}_\tau(\theta)]_{i,j} = \theta^{|i-j|}, \forall i, j \in \{1, \dots, \tau\}$ is due to the temporal correlation in (2.2). Similarly, if the signal follows the VARMA(1,1) model, the matrix $\mathbf{V}$ can be parameterized by $\theta \triangleq (\phi, \eta)$ with the following form:

$$\mathbf{V}_\tau(\theta) = \mathbf{R}_\tau(\phi, \eta) \otimes \mathbf{\Lambda}, \quad (2.7)$$

where $\mathbf{R}_\tau(\phi, \eta) \in \mathbb{R}^{\tau \times \tau}$; $[\mathbf{R}_\tau(\phi, \eta)]_{i,j} = 1 + \eta^2 - 2\phi\eta$ if $i=j$ and $[\mathbf{R}_\tau(\phi, \eta)]_{i,j} = \phi^{|i-j|-1}(\phi - \eta)(1 - \phi\eta)$ otherwise. For other models, similar forms of $\mathbf{V}_\tau$ can be derived: the temporal dependence of the signal is captured by $\mathbf{R}_\tau$, the spatial dependence by $\mathbf{\Lambda}$, and the spatial-temporal covariance is a Kronecker product of the two (Genton 2007).

Using the representation above, the detection problem can be reformulated as the following hypothesis test:

$$\begin{aligned} H_0: & \mathbf{y}_{(1:k)} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_k), \quad \mathbf{y}_{(k+1:N)} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\tau), \\ H_1: & \mathbf{y}_{(1:k)} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_k), \quad \mathbf{y}_{(k+1:N)} \sim \mathcal{N}(\boldsymbol{\mu}_{(k+1:N)}, \gamma \mathbf{V}_\tau(\theta) + \mathbf{\Sigma}_\tau), \end{aligned} \quad (2.8)$$

for $k = 1, \dots, N-1$, where $\mathbf{0}$ is a vector of zeros, $\boldsymbol{\mu}_{(k+1:N)} = \mathbb{E}[\mathbf{y}_{(k+1:N)}] \in \mathbb{R}^{p\tau}$ and $\gamma \in \mathbb{R} > 0$. Note that we assume that $\boldsymbol{\mu}_{(k+1:N)}$ is unknown. Equivalently, under the null

Table 1. Notation.

p	Dimension of samples
N	Sample size in offline change point detection
k	Change point location
τ	Number of postchange samples, $\tau = N - k$
Σ	Spatial covariance matrix of the noise, $\Sigma = \text{Var}[\mathbf{w}_\ell]$
Λ	Structure of the spatial covariance matrix of the signal $\text{Var}[\mathbf{x}_\ell] = \gamma\Lambda$
γ	Magnitude of the spatial covariance matrix of the signal $\text{Var}[\mathbf{x}_\ell] = \gamma\Lambda$
Σ_τ	Covariance of noise in concatenated observations $\Sigma_\tau = \text{Var}[\mathbf{w}_{(k+1:N)}] = \mathbf{I}_\tau \otimes \Sigma$
$\gamma\mathbf{V}_\tau(\theta)$	Covariance of signal in concatenated observations $\text{Var}[\mathbf{x}_{(k+1:N)}] = \gamma\mathbf{V}_\tau(\theta) = \gamma\mathbf{R}_\tau(\theta) \otimes \Lambda$
$\mathbf{R}_\tau(\theta)$	Matrix that captures temporal dependence of the signal

hypothesis, $\gamma = 0$ and $\boldsymbol{\mu}_{(k+1:N)} = \mathbf{0}$, and under the alternative hypothesis, $\gamma > 0$ or $\boldsymbol{\mu}_{(k+1:N)} \neq \mathbf{0}$. Using this form of the hypothesis, we can derive our score statistic.

Table 1 provides a list of notation used throughout the article.

3. STATISTIC FOR OFFLINE AND ONLINE DETECTION

We now derive the $\mathbf{S}^3\mathbf{T}$ statistic for offline change point detection. The log-likelihood function for the hypothesis test in (2.8) is given by

$$\begin{aligned} \ell(\gamma, \boldsymbol{\mu}, \tau, \theta) = & -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log |\gamma\mathbf{V}_\tau(\theta) + \Sigma_\tau| \\ & - \frac{1}{2} (\mathbf{y}_{(k+1:N)} - \boldsymbol{\mu}_{(k+1:N)})^\top (\gamma\mathbf{V}_\tau(\theta) + \Sigma_\tau)^{-1} (\mathbf{y}_{(k+1:N)} - \boldsymbol{\mu}_{(k+1:N)}). \end{aligned} \quad (3.1)$$

To cope with unknown parameters, we may use the generalized likelihood ratio (GLR) statistic based on (3.1). However, (3.1) involves the inversion of a $p\tau$ by $p\tau$ dimensional matrix $\gamma\mathbf{V}_\tau(\theta) + \Sigma_\tau$, which incurs a complexity of $O(p^3\tau^3)$ for a given τ . Recall that $\tau = N - k$, so $\tau = 1, 2, \dots, N$. Hence, the complexity of computing the GLR statistic is $O(p^3N^3)$, which grows polynomially with N (the time horizon), and the computation of the likelihood statistic becomes prohibitive.

3.1. Quadratic score statistic

Define the following notation. Let $\mathbf{A}_\tau(\theta) = \Sigma_\tau^{-1} \mathbf{V}_\tau(\theta)$, $\mathbf{B}_\tau(\theta) = \Sigma_\tau^{-1/2} \mathbf{V}_\tau(\theta) \Sigma_\tau^{-1/2}$, $c(\tau, \theta) = \text{tr}(\mathbf{A}_\tau(\theta))$, and $d(\tau, \theta) = 2\text{tr}[\Sigma_\tau^{-1} \mathbf{V}_\tau(\theta) \Sigma_\tau^{-1} \mathbf{V}_\tau(\theta)]$, where $\text{tr}(\cdot)$ denotes the trace of a matrix. We now derive the score statistic for detection. Take the derivative of $\ell(\gamma, \boldsymbol{\mu}, \tau, \theta)$ in (3.1) with respect to γ and $\boldsymbol{\mu}$ and evaluate at $\gamma = 0$ and $\boldsymbol{\mu} = \mathbf{0}$. Then we obtain

$$\boldsymbol{\varsigma}(\tau, \theta) = \begin{bmatrix} \left. \frac{\partial \ell}{\partial \gamma} \right|_{\boldsymbol{\mu}=\mathbf{0}, \gamma=0} \\ \left. \frac{\partial \ell}{\partial \boldsymbol{\mu}} \right|_{\boldsymbol{\mu}=\mathbf{0}, \gamma=0} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} c(\tau, \theta) + \frac{1}{2} \mathbf{y}_{(k+1:N)}^\top \Sigma_\tau^{-1} \mathbf{V}_\tau(\theta) \Sigma_\tau^{-1} \mathbf{y}_{(k+1:N)} \\ \Sigma_\tau^{-1} \mathbf{y}_{(k+1:N)} \end{bmatrix}. \quad (3.2)$$

The derivation of (3.2) is given in Appendix A. It can be verified that $\mathbb{E}[\boldsymbol{\varsigma}(k, \theta)] = \mathbf{0}$ under the null hypothesis, where $\mathbf{0}$ represents the zero vector. It can also be shown that the covariance of the score vector $\boldsymbol{\varsigma}(\tau, \theta)$ is given by

$$\text{Cov}[\varsigma(\tau, \theta)] = \begin{bmatrix} \frac{1}{4}d(\tau, \theta) & 0 \\ \mathbf{0} & \mathbf{\Sigma}_\tau^{-1} \end{bmatrix}.$$

As suggested by Rao (1948), when the likelihood function involves multiple parameters, the score statistic is a quadratic function of the efficient score vector. In our case, this becomes

$$\begin{aligned} S(\tau, \theta) &= \varsigma(\tau, \theta)^\top \text{Cov}[\varsigma(\tau, \theta)]^{-1} \varsigma(\tau, \theta) \\ &= \frac{\left[\mathbf{y}_{(k+1:N)}^\top \mathbf{\Sigma}_\tau^{-1} \mathbf{V}_\tau(\theta) \mathbf{\Sigma}_\tau^{-1} \mathbf{y}_{(k+1:N)} - c(\tau, \theta) \right]^2}{d(\tau, \theta)} + \mathbf{y}_{(k+1:N)}^\top \mathbf{\Sigma}_\tau^{-1} \mathbf{y}_{(k+1:N)}. \end{aligned} \quad (3.3)$$

The most expensive part in evaluating (3.3) is computing $\mathbf{\Sigma}_\tau^{-1}$. According to (2.5), we have $\mathbf{\Sigma}_\tau^{-1} = \mathbf{I}_\tau \otimes \mathbf{\Sigma}^{-1}$, which means that we only need to compute $\mathbf{\Sigma}^{-1}$, which has a complexity $O(p^3)$. Hence, the computational complexity of evaluating $S(\tau, \theta)$ is much lower than that of the GLR statistic. Moreover, because $\mathbf{\Sigma}$ is assumed known and fixed, its inversion can be precomputed. However, in (3.1), the likelihood function involves $(\gamma \mathbf{V}_\tau(\theta) + \mathbf{\Sigma}_\tau)^{-1}$, which has to be computed for each τ value.

Because the expected value of $S(\tau, \theta)$ increases as τ increases, it needs to be normalized to have mean 0 and variance 1 under the null hypothesis. This leads to the following *quadratic score statistic*,

$$\tilde{S}(\tau, \theta) = \frac{S(\tau, \theta) - \mathbb{E}[S(\tau, \theta)]}{\sqrt{\text{Var}[S(\tau, \theta)]}}, \quad (3.4)$$

where $\mathbb{E}[S(\tau, \theta)] = p\tau + 1$, and the variance is given by (the derivation can be found in Appendix B)

$$\begin{aligned} \text{Var}[S(\tau, \theta)] &= 2p\tau + 10 - 24 \frac{c(\tau, \theta)}{d(\tau, \theta)^2} \text{tr}(\mathbf{\Sigma}_\tau^{-1} \mathbf{V}_\tau(\theta) \mathbf{\Sigma}_\tau^{-1} \mathbf{V}_\tau(\theta) \mathbf{\Sigma}_\tau^{-1} \mathbf{V}_\tau(\theta)) \\ &\quad + \frac{48}{d(\tau, \theta)^2} \text{tr}(\mathbf{\Sigma}_\tau^{-1} \mathbf{V}_\tau(\theta) \mathbf{\Sigma}_\tau^{-1} \mathbf{V}_\tau(\theta) \mathbf{\Sigma}_\tau^{-1} \mathbf{V}_\tau(\theta) \mathbf{\Sigma}_\tau^{-1} \mathbf{V}_\tau(\theta)). \end{aligned}$$

Then we may construct the *quadratic detector* using $\tilde{S}(\tau, \theta)$, which detects a signal when the maximum standardized score statistic over all possible parameter values of $\theta \in \Theta$ and τ exceeds a prespecified threshold $b > 0$,

$$\max_{\theta \in \Theta, 1 \leq \tau \leq N} \tilde{S}(\tau, \theta) \geq b.$$

3.2. $\mathbf{S}^3\mathbf{T}$ statistic for offline change point detection

Although the quadratic score statistic achieves the maximum discrimination between the null and the alternative distribution (Rao 1948), theoretical analysis of the detection statistic is intractable; thus, it is difficult to calibrate the threshold b to control the false alarm rate. In this section, we propose a simpler statistic, namely, the $\mathbf{S}^3\mathbf{T}$ statistic, which is the score statistic with respect to γ only:

$$W(\tau, \theta) = \frac{\frac{\partial \ell}{\partial \gamma} \big|_{\mu=0, \gamma=0}}{\sqrt{\text{Var} \left[\frac{\partial \ell}{\partial \gamma} \big|_{\mu=0, \gamma=0} \right]}} = \frac{\mathbf{y}_{(N-\tau+1:N)}^\top \boldsymbol{\Sigma}_\tau^{-1} \mathbf{V}_\tau(\theta) \boldsymbol{\Sigma}_\tau^{-1} \mathbf{y}_{(N-\tau+1:N)} - c(\tau, \theta)}{\sqrt{d(\tau, \theta)}}. \quad (3.5)$$

Note that both the spatial and temporal correlations are still captured in the statistic by $\boldsymbol{\Sigma}_\tau^{-1}$ in (2.5) and $\mathbf{V}_\tau(\theta)$ in (2.6), respectively. Under the null hypothesis, the detection statistic $W(\tau, \theta)$ has mean 0 and unit variance. The detection procedure claims a change when the maximum of the score statistic exceeds a prespecified threshold $b > 0$,

$$\max_{\theta \in \Theta, 1 \leq \tau \leq N} W(\tau, \theta) \geq b. \quad (3.6)$$

3.3. $\mathbf{S}^3\mathbf{T}$ statistic for online change point detection

We now present an online change point detection procedure based on the $\mathbf{S}^3\mathbf{T}$ statistic. In the online setting, the sample size N is not fixed and samples are sequentially collected. A signal may occur at an unknown time k . Our goal is to detect the emergence of the signal as soon as possible.

Consider a sequential version of the hypothesis test in (2.1), where the number of samples N increases. We adopt a sliding window approach for online detection and construct the detection statistic using the most recent ω samples at each time, where ω is a prespecified window length (demonstrated in Appendix C). Given a current time t , the detection statistic constructed using the most recent ω samples is given by

$$W_t(\omega, \theta) = \frac{\mathbf{y}_{(t-\omega+1:t)}^\top \boldsymbol{\Sigma}_\omega^{-1} \mathbf{V}_\omega(\theta) \boldsymbol{\Sigma}_\omega^{-1} \mathbf{y}_{(t-\omega+1:t)} - c(\omega, \theta)}{\sqrt{d(\omega, \theta)}}. \quad (3.7)$$

The detection procedure for online change point detection is a stopping time, which raises an alarm when the detection statistic exceeds a threshold $b > 0$ for the first time:

$$\mathcal{T} = \inf \left\{ t : \max_{\theta \in \Theta} W_t(\omega, \theta) \geq b \right\}, \quad (3.8)$$

where b is a prespecified threshold. Note that this corresponds to a type of Shewhart chart (Shewhart 1931).

4. THEORETICAL APPROXIMATIONS

4.1. Significance level for offline $\mathbf{S}^3\mathbf{T}$ statistic

We present a theoretical approximation for the significance level of the detection procedure defined in (3.6). The approximation is quite accurate and can be used to avoid the time-consuming simulation when choosing an appropriate b . Denote the standard normal density function by $\phi(x)$ and its distribution function by $\Phi(x)$, and define a special function (Siegmund and Yakir 2007):

$$\nu(x) = \frac{\frac{x}{2} \left[\Phi\left(\frac{x}{2}\right) - \frac{1}{2} \right]}{\frac{x}{2} \Phi\left(\frac{x}{2}\right) + \phi\left(\frac{x}{2}\right)}. \quad (4.1)$$

Define the following quantities, which are useful to state our theoretical approximation results:

$$\mu(\tau, \theta) = \tau \left[\frac{\text{tr}(\mathbf{A}_{\tau+1}(\theta) \mathbf{A}_{\tau+1}(\theta))}{\text{tr}(\mathbf{A}_{\tau}(\theta) \mathbf{A}_{\tau}(\theta))} - 1 \right], \quad (4.2)$$

$$H(\tau, \theta) = - \frac{\partial^2 E[W(\tau, \theta) W(\tau, s)]}{\partial^2 s} \Big|_{s=\theta}, \quad (4.3)$$

$$g(\tau, \theta) = \left(\frac{\exp - \xi_0(\tau, \theta) b + \psi(\xi_0(\tau, \theta))}{\sigma_{\xi_0} \sqrt{2\pi}} \right), \quad (4.4)$$

$$\psi(\xi) = -\xi \frac{c(\tau, \theta)}{\sqrt{d(\tau, \theta)}} - \frac{1}{2} \log \left| \mathbf{I}_{p\tau} - \frac{2\xi \mathbf{B}_{\tau}(\theta)}{\sqrt{d(\tau, \theta)}} \right|. \quad (4.5)$$

Note that $\psi(\xi)$ is the cumulant generating function (a.k.a. the log-moment generating function) of the detection statistic $W(\tau, \theta)$. The following theorem is our main theoretical result, which provides an analytical approximation for the significance level of the detection procedure defined in (3.6).

Theorem 4.1 (Approximation for significance level). *When the threshold $b \rightarrow \infty$ and $\theta \in \Theta \subset \mathbb{R}^d$, under the null hypothesis, the probability of false alarm for the procedure defined in (3.6) is given by*

$$\begin{aligned} & \mathbb{P}_{H_0} \left(\max_{\substack{\theta \in \Theta \\ 1 \leq \tau \leq N}} W(\tau, \theta) \geq b \right) \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \sum_{\tau=1}^N \int_{\theta \in \Theta} \frac{[b \xi_0(\tau, \theta)]^{\frac{d}{2}}}{\xi_0(\tau, \theta)} g(\tau, \theta) |H(\tau, \theta)|^{\frac{1}{2}} \frac{b^2 \mu(\tau, \theta)}{2\tau} \nu \left(\sqrt{\frac{b^2 \mu(\tau, \theta)}{\tau}} \right) d\theta + o(1), \end{aligned} \quad (4.6)$$

where

$$\sigma_{\xi_0}^2 = d(\tau, \theta)^{-1} \text{tr} \left(\left[\mathbf{I}_{p\tau} - \frac{2\xi_0 \mathbf{B}_{\tau}(\theta)}{\sqrt{d(\tau, \theta)}} \right]^{-1} \mathbf{B}_{\tau}(\theta) \left[\mathbf{I}_{p\tau} - \frac{2\xi_0 \mathbf{B}_{\tau}(\theta)}{\sqrt{d(\tau, \theta)}} \right]^{-1} \mathbf{B}_{\tau}(\theta) \right),$$

and $\xi_0(\tau, \theta)$ is the solution to

$$\frac{1}{\sqrt{d(\tau, \theta)}} \text{tr} \left(\left[\mathbf{I}_{p\tau} - \frac{2\xi_0 \mathbf{B}_{\tau}(\theta)}{\sqrt{d(\tau, \theta)}} \right]^{-1} \mathbf{B}_{\tau}(\theta) - \mathbf{A}_{\tau}(\theta) \right) = b. \quad (4.7)$$

Note that the solution of (4.7) can be obtained by a simple grid search when the dimension of θ is not too large.

The main proof technique for Theorem 4.1 is *change of measure*, which evaluates the boundary hitting probability of Gaussian processes (Siegmund 1985; Yakir 2013). See Appendix D for the derivation of (4.5) and Appendix E for the proof of Theorem 4.1, when the dimension of parameter θ is 1 (i.e., $d=1$). The proof can be generalized to multidimensional cases.

Table 2. Simulated and approximated significance level when the signal $\{\mathbf{x}_\ell\}$ follows a VAR(1) model ($\theta \in [0.1, 0.9]$, $N = 50$ and $\rho = 0.3$).

b	$p = 2$		$p = 9$		$p = 36$	
	Simulation	Approximation	Simulated	Approximation	Simulated	Approximation
3.5	0.097	0.097	0.065	0.057	0.036	0.042
4	0.063	0.068	0.036	0.030	0.013	0.019
4.5	0.038	0.047	0.018	0.019	0.006	0.008
5	0.033	0.032	0.011	0.012	0.003	0.003
5.5	0.022	0.021	0.005	0.007	0.002	0.001
6	0.015	0.014	0.003	0.004	0.0004	0.0005
6.5	0.006	0.009	0.002	0.002	0.0002	0.0002

Theorem 4.1 is useful for calibrating the threshold b analytically, avoiding the onerous numerical simulations. Because the detection statistic requires evaluating the maximum over the set of θ , if we determine the threshold by running numerical simulations, we will have to run a large number of Monte Carlo trials for each discretized value of θ to obtain an estimation with good accuracy. On the other hand, applying the theorem to obtain b is relatively easy, because we only need to evaluate the integral numerically.

Although the theorem is an asymptotic result for large b , we find that this holds even for not very large b values in numerical studies. We verify the accuracy of **Theorem 4.1** by comparing the approximated significance levels with simulated ones. In the experiment, we assume that the signal $\{\mathbf{x}_\ell\}$ follows a VAR(1) model, $\mathbf{x}_\ell = (1 - \theta)\boldsymbol{\mu} + \theta\mathbf{x}_{\ell-1} + \boldsymbol{\epsilon}_\ell$, where $\theta \in \mathbb{R}$. Hence, $\mathbf{V}_\tau(\theta)$ has the form in (2.6). We further assume that the spatial correlation of the signal follows a spherical model, as defined in (2.3), with parameter $\rho = 0.3$. Set $N = 50$. The search space of θ is $\{0.1, 0.2, \dots, 0.9\}$. In addition, the covariance matrix of the noise process $\boldsymbol{\Sigma}$ is assumed to be a p by p identity matrix. Simulation results are based on 5,000 independent replications. Both simulated and approximated false alarm rates are reported in **Table 2**. As one can observe, the approximation is quite accurate.

In the proof of **Theorem 4.1**, we approximate the detection statistic $W(\tau, \theta)$ as a two-dimensional Gaussian random field. In the following, we verify by simulation that such an approximation is reasonable. We generate data under the null hypothesis and verify the distribution of the detection statistic W for a set of fixed values of θ and τ . For the signal, we use a VAR(1) model, $\mathbf{x}_\ell = (1 - \theta)\boldsymbol{\mu} + \theta\mathbf{x}_{\ell-1} + \boldsymbol{\epsilon}_\ell$, as the temporal correlation model and a spherical model for spatial correlation model. We assume that the noise has the same spatial correlation structure as the signal. We set $N = 50$ and $p = 9$. **Figure 1** shows the histograms and q-q plots of W for fixed values of θ and τ based on 1,000 repetitions, which indicate that the Gaussian random field approximation is reasonable.

4.2. Average Run Length for online S^3T statistic

In the online setting, the false alarm rate is characterized by the average run length (ARL), which is the expected stopping time of the procedure when there is no signal, denoted as $E_{H_0}(\mathcal{T})$. The following theorem provides an approximation for $E_{H_0}(\mathcal{T})$.

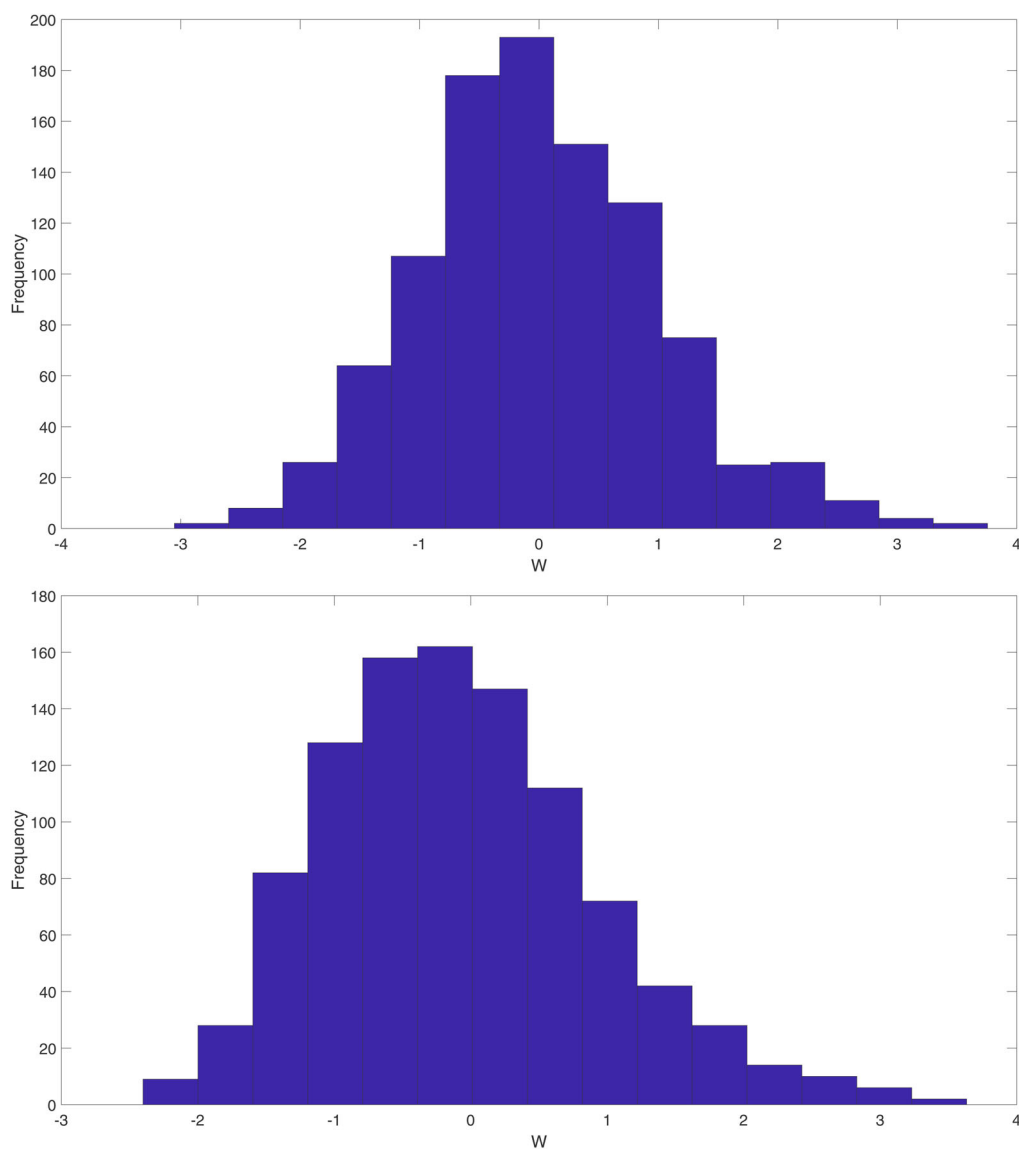


Figure 1. Histograms and q-q plots of $W(\theta, \tau)$ for fixed values of θ and τ : (a), (c) $\tau = 30$, $\theta = 0.3$ and (b), (d) $\tau = 40$, $\theta = 0.2$.

Theorem 4.2 (Approximation of average run length). Assume that $b \rightarrow \infty$. For the stopping time defined in (3.8),

$$E_{H_0}(\mathcal{T}) = (2\pi)^{\frac{d}{2}} \left(\int_{\theta \in \Theta} \frac{[b\xi_0(\omega, \theta)]^{\frac{d}{2}}}{\xi_0(\omega, \theta)} g(\omega, \theta) |H(\omega, \theta)|^{\frac{1}{2}} \frac{b^2 \mu(\omega, \theta)}{2\omega} \nu \left(\sqrt{\frac{b^2 \mu(\omega, \theta)}{\omega}} \right) d\theta \right)^{-1} (1 + o(1)). \quad (4.8)$$

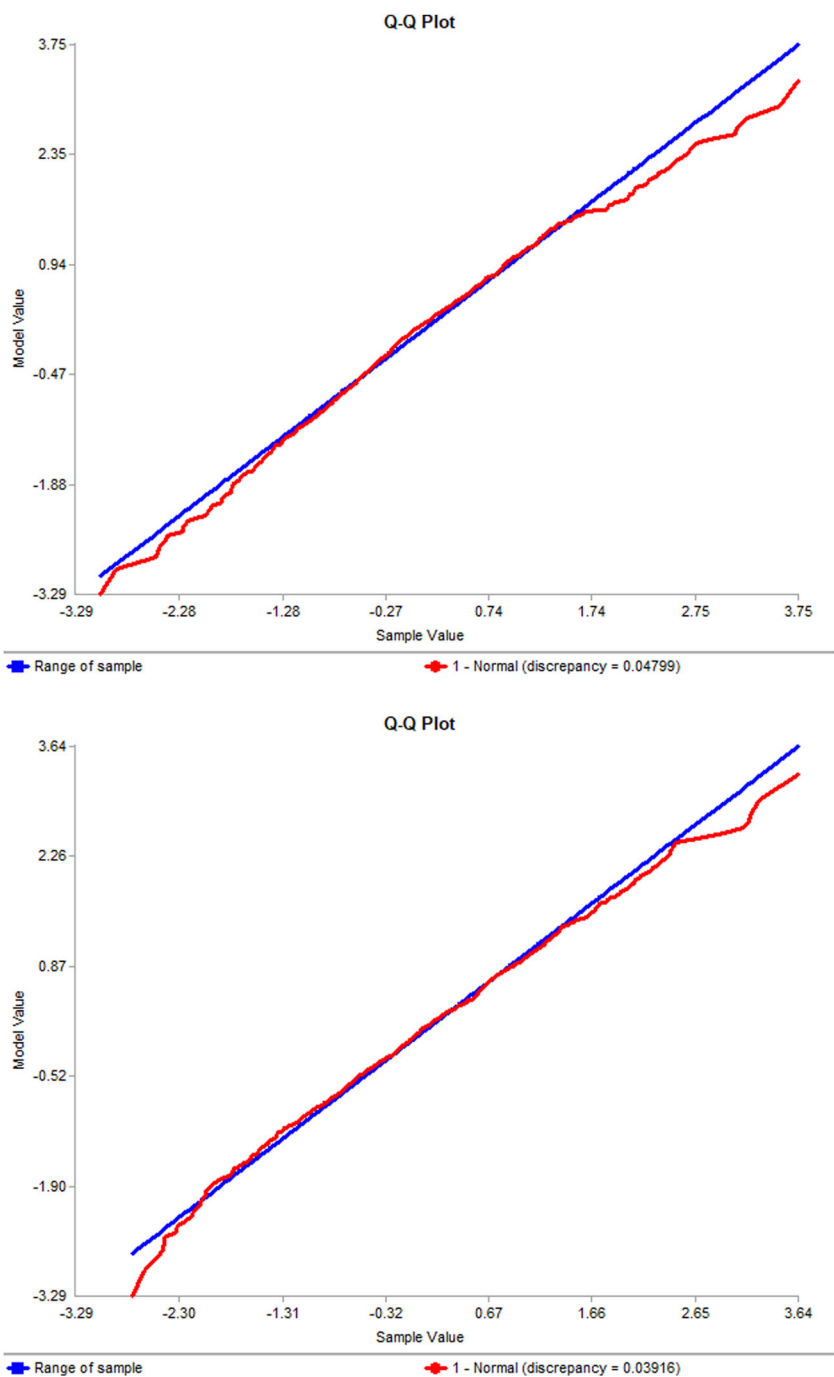


Figure 1. Continued.

The derivation of Theorem 4.2 uses a similar technique based on the change of measure as in the derivation of Theorem 4.1. By Theorem 4.1, we can first obtain an approximation to the probability $\mathbb{P}_{H_0}(T \leq m)$, where $m > 0$ is fixed and sufficiently large:

$$\begin{aligned}
\mathbb{P}_{H_0}(\mathcal{T} \leq m) &= \mathbb{P}_{H_0} \left(\max_{\substack{\theta \in \Theta \\ 1 \leq t \leq m}} W_t(\omega, \theta) \geq b \right) \\
&= (2\pi)^{-\frac{d}{2}} \left(\sum_{t=1}^m \int_{\theta \in \Theta} \frac{[b\xi_0(\omega, \theta)]^{\frac{d}{2}}}{\xi_0(\omega, \theta)} g(\omega, \theta) |H(\omega, \theta)|^{\frac{1}{2}} \frac{b^2 \mu(\omega, \theta)}{2\omega} \nu \left(\sqrt{\frac{b^2 \mu(\omega, \theta)}{\omega}} \right) d\theta \right) + o(1).
\end{aligned} \tag{4.9}$$

As argued in Siegmund and Venkatraman (1995) and Siegmund and Yakir (2008), the stopping time $\mathcal{T}$ is asymptotically exponentially distributed and is uniformly integrable. Hence, for large m , $\mathbb{P}_{H_0}(\mathcal{T} \leq m) - [1 - \exp(-\lambda m)] \rightarrow 0$, where λ is approximately equal to the right-hand side of (4.9) divided by m . Then by the first-order Taylor expansion of an exponential term, we can obtain $E_{H_0}(\mathcal{T}) \approx \lambda^{-1}$, which leads to (4.8).

The accuracy of Theorem 4.2 is verified by comparing the simulated and the approximated $E_{H_0}(\mathcal{T})$. In the experiments, the signal $\{\mathbf{x}_\ell\}$ is generated by a VAR(1) model, $\mathbf{x}_\ell = (1 - \theta)\boldsymbol{\mu} + \theta\mathbf{x}_{\ell-1} + \boldsymbol{\epsilon}_\ell$, where $\theta \in \mathbb{R}$. Hence, $V_\tau(\theta)$ has the form in (2.6). Meanwhile, we assume that the spatial correlation of the signal follows a spherical model, as defined in (2.3), with parameter $\rho = 0.3$. The search space of parameter θ is $\{0.1, 0.2, \dots, 0.9\}$. In addition, the covariance matrix of the noise process $\boldsymbol{\Sigma}$ is assumed to be a p by p identity matrix. The results based on 5,000 replications are presented in Figure 2. The comparison between the simulated and approximated ARLs shows that the approximation in Theorem 4.2 is quite accurate.

5. NUMERICAL EXAMPLES

In this section, we demonstrate the performance of the proposed detection procedures. Online change point detection is the focus here because it is the most relevant setting for our targeted applications of water quality monitoring. The performance comparison for offline change point detection will be similar. We adopt the commonly used performance metric for sequential change detection, the expected detection delay (EDD), after a change has occurred. There is a trade-off between the in-control ARL and the EDD. Typically, we choose the threshold for each procedure so that its ARL meets a prespecified large value (e.g., 5,000 or 10,000), so that there is rarely a false alarm. We also compare with other methods on simulated and real data.

5.1. Simulation

The detection procedure defined in (3.8) is compared with two other procedures: (i) an online detection procedure defined similarly to (3.8) using the quadratic score statistic $\tilde{S}(\tau, \theta)$ and (ii) a multivariate cumulative sum (MCUSUM) procedure (Healy 1987). In the MCUSUM procedure, at each time step, a T^2 statistic (Hotelling 1947) is calculated, which is combined with a CUSUM procedure.

In the experiment, the signal is generated from a VAR(1) model, $\mathbf{x}_\ell = (1 - \theta)\boldsymbol{\mu} + \theta\mathbf{x}_{\ell-1} + \boldsymbol{\epsilon}_\ell$, with $p = 2$ and parameter $\theta = 0.5$. The spatial model of the signal follows the spherical model defined in (2.3) with $\rho = 0.3$. For both procedures, based on $\mathbf{S}^3\mathbf{T}$ and the quadratic score statistic, respectively, we use a window length $\omega = 50$ and the

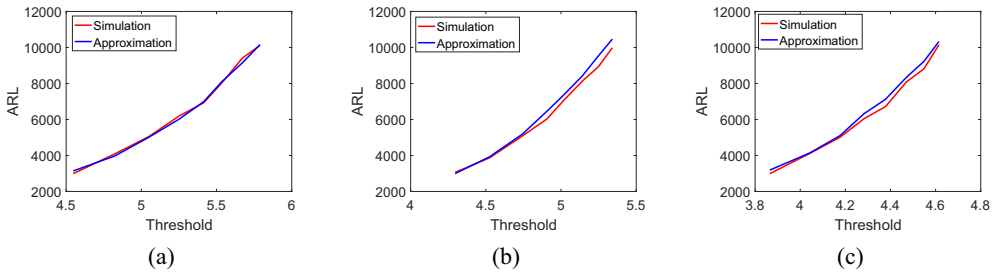


Figure 2. Comparison of approximated and simulated ARL for (a) $p = 1$, (b) $p = 2$, and (c) $p = 9$.

search space for the parameter θ , $\{0.1, 0.2, \dots, 0.9\}$. Thresholds for all three procedures are calibrated so that they have the same false alarm rate $E_{H_0}(\mathcal{T}) = 100$. To evaluate the EDD, we assume that the change occurs at $t = 1$. The mean of the signal $\boldsymbol{\mu} = E[\mathbf{x}_\ell] = \mu \mathbb{1}_p$, $\mu \geq 0$. We explore different values of μ for the mean shift and γ for the magnitude of covariance matrix of the signal. If $\mu = 0$ and $\gamma > 0$, then there is only a change in covariance; if both μ and γ are positive, then there are both mean shift and covariance change. Hence, the experiments demonstrate that the proposed detection procedure is suitable for both cases where there is either mean and/or covariance change.

Table 3 reports the simulated EDD of three procedures based on 5,000 repetitions. The smallest EDD values for each setting are marked in bold. The comparison shows that the two score statistic procedures, which capture both spatial and temporal correlation, outperform the MCUSUM procedure (which only captures the spatial correlation information). Such an advantage is more significant when the signal is weak; that is, when γ or μ are both small. This demonstrates that incorporating temporal correlation information indeed improves detection performance. We also find that $\mathbf{S}^3\mathbf{T}$ outperforms the quadratic score statistic in many settings. This can be explained by the fact that the quadratic score statistic needs to search more unknown parameters (the unknown μ); thus, the statistic is noisier than $\mathbf{S}^3\mathbf{T}$ when there is no change. Therefore, to achieve the same ARL, the threshold for the quadratic score statistic tends to be higher, which may cause a larger detection delay. Given that $\mathbf{S}^3\mathbf{T}$ enjoys tractable theoretical analysis and an accurate approximation for its false alarm rate, it is a good option for practitioners.

5.2. Real data example: Solar flare detection

We apply our detection procedure to a data set, which is acquired by the Solar Data Observatory (National Aeronautics and Space Administration 2013). The data are a video sequence that contains an abrupt emergence of a solar flare that occurs around time $t = 227$. In this video, the normal state is a sequence of slowly drifting images of the solar surface, and the changes are much brighter transient solar flares. Figure 3 shows a snapshot when a solar flare occurs at $t = 227$.

The size of the images is 232×292 pixels. After vectorizing the images, this leads to 67,744 dimensional vectors. Due to the high dimensionality, it is computationally expensive to apply our detection procedure on the original images directly. Hence, we

Table 3. Simulated expected detection delay.

$\gamma \backslash \mu$	S^3T						Quadratic score statistic						MCUSUM					
	0	0.1	0.5	1	2		0	0.1	0.5	1	2		0	0.1	0.5	1	2	
0.01	97.27	59.08	6.37	2.80	1.49		98.05	65.82	6.45	2.77	1.51		98.37	77.67	9.43	3.56	1.79	
0.05	96.28	57.96	5.95	2.72	1.49		95.32	63.19	6.74	2.81	1.52		96.79	71.97	9.28	3.54	1.79	
0.1	72.93	53.16	6.04	2.78	1.50		82.49	56.78	6.74	2.86	1.49		80.70	65.16	9.21	3.54	1.78	
0.2	65.32	46.16	5.96	2.77	1.50		74.87	48.83	6.28	2.78	1.47		67.33	55.17	9.02	3.52	1.79	
0.5	39.40	30.32	5.81	2.78	1.56		37.07	33.42	6.07	2.80	1.50		41.52	35.87	8.36	3.47	1.78	
1	20.91	19.42	5.65	2.75	1.51		22.75	20.51	5.64	2.76	1.55		23.71	21.31	7.45	3.45	1.77	

Smallest EDD values for each setting shown in bold.

apply a spatial scanning scheme by breaking the original image into overlapping patches of dimension 20×20 , as demonstrated in the right panel of Figure 3. The detection statistic is calculated for each image patch (of dimension $p = 400$). Then, we take the maximum of the detection statistic over all of the patches.

We assume that before the solar flare, the data form a white noise process with no spatial and temporal correlation. The mean and variance of the noise process are estimated by the first 50 samples in the sequence. For the signal, we use a VAR(1) model, $\mathbf{x}_\ell = (1 - \theta)\boldsymbol{\mu} + \theta\mathbf{x}_{\ell-1} + \boldsymbol{\epsilon}_\ell$ to capture the temporal correlation. The spatial model of the signal is captured by a spherical model defined in (2.3). Online procedures are implemented with a window length of $\omega = 10$. Figures 4(a)–4(c) show the values of S^3T statistic, the quadratic score statistic, and the MCUSUM statistic on a logarithmic scale, respectively. Because in this case we do not have the ground truth, we cannot evaluate the true EDD. However, as we can observe, both S^3T and the quadratic score statistics obtain peak detection statistics at around $t = 227$, and another solar flare at around $t = 173$, indicating that both statistics can successfully detect the emergence of solar flares. However, the MCUSUM statistic misses both solar flares.

5.3. Case study: Water quality monitoring

In this section, we consider a case study of real-time water quality monitoring using a sensor network deployed along the Altamaha River in Georgia, United States. The goal is to detect contaminant spills that pollute the river as quickly as possible.

We study the Altamaha River in Georgia, United States. The shape of the river is shown in Figure 5(a). The nodes in the river network represent monitoring locations, where concentration data are collected. The contaminant concentration data for such a river network are simulated by the Storm Water Management Model (SWMM; see Rossman 2010) developed by the United States Environmental Protection Agency. SWMM is widely used in environmental engineering for water-related studies. SWMM requires geologic, geometric, and fundamental hydrodynamics data to construct a river network. Given rainfall information, as well as the location, intensity, and duration of a contaminant spill, SWMM simulates the contaminant transport process through the river over a period. Data generated by SWMM are highly close to real data and hence are widely used for water-related study when real data are not available.

In river dynamic simulation systems, rain events and spill events bring randomness to the contaminant transport. We use the same data as those in Telci and Aral (2011) to generate rain events. The Altamaha River watershed is divided into 10

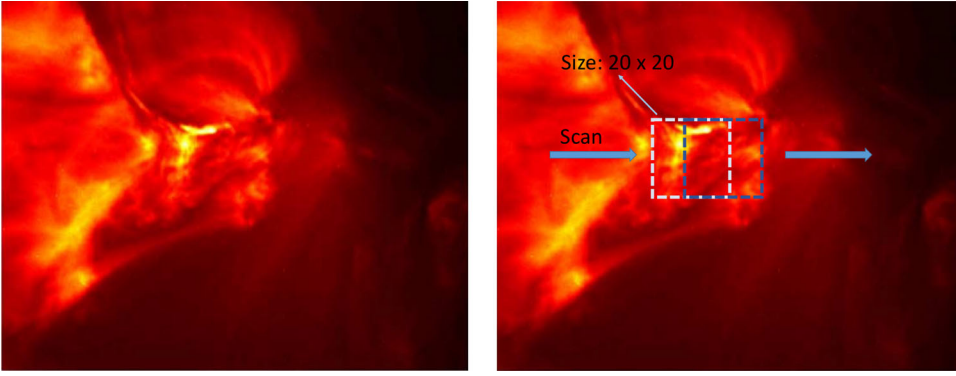


Figure 3. Detection of solar flare at $t = 227$: (left) snapshot of the original SDO data at $t = 227$; (right) overlapping image patches for dimensionality reduction.

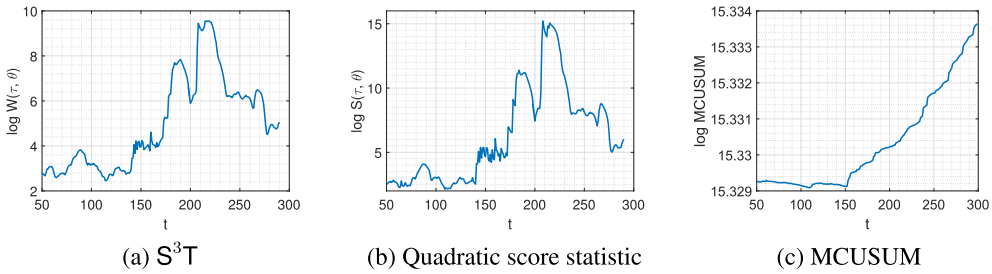


Figure 4. Detection statistics on logarithmic scale.

subcatchments, as shown in Figure 5(b). The rainfall measurements are obtained from different United States Geological Survey stations close to these 10 subcatchments in 2006. Based on the statistical analysis of these measurements, five rain patterns are generated for each subcatchment. Each rain pattern describes time-dependent rainfall events and keeps changing hydrologic conditions in each catchment during the simulation. Note that the rain patterns for each subcatchment are different, and thus there are 5^{10} possible combinations for the entire watershed.

Due to the nature of hydrodynamics, there is a strong spatial correlation among the concentration data collected at different locations in the river network. However, the shape of the network and the direction of the stream impose constraints on spatial modeling. For example, there should not be a correlation for data collected at two locations that do not share a common flow. A reasonable spatial correlation model is critical here.

We adopt the so-called tail-up spatial model for stream networks, which is proposed based on the moving average constructions in Ver Hoef and Peterson (2010). The tail-up models have the following desired properties: (i) they use stream distance rather than the Euclidean distance, which is defined as the shortest distance along the stream network between two locations; (ii) statistical independence is imposed on the samples located on stream segments that do not share a common flow; and (iii) proper

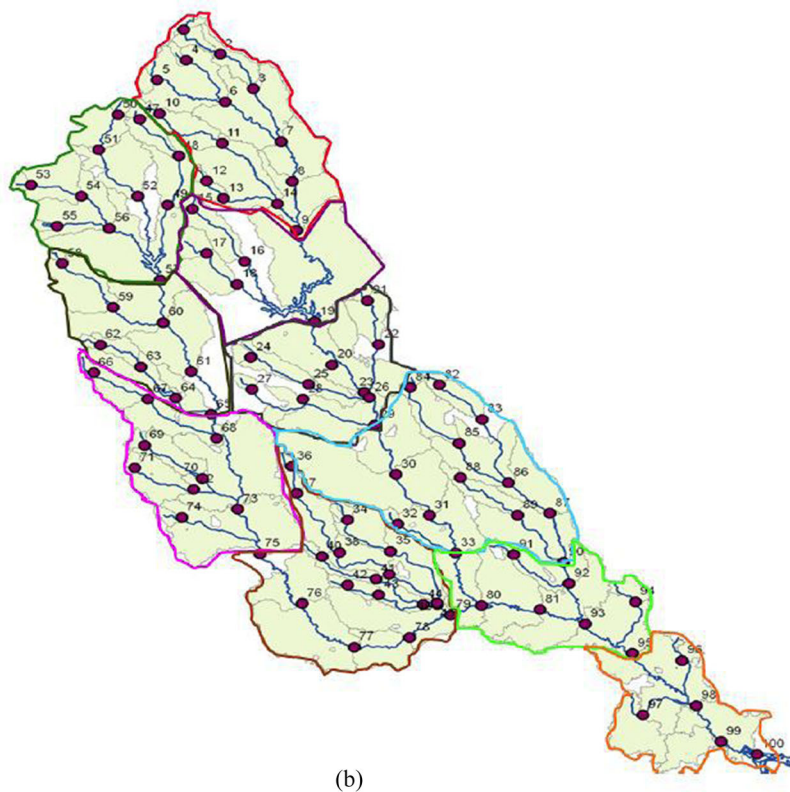
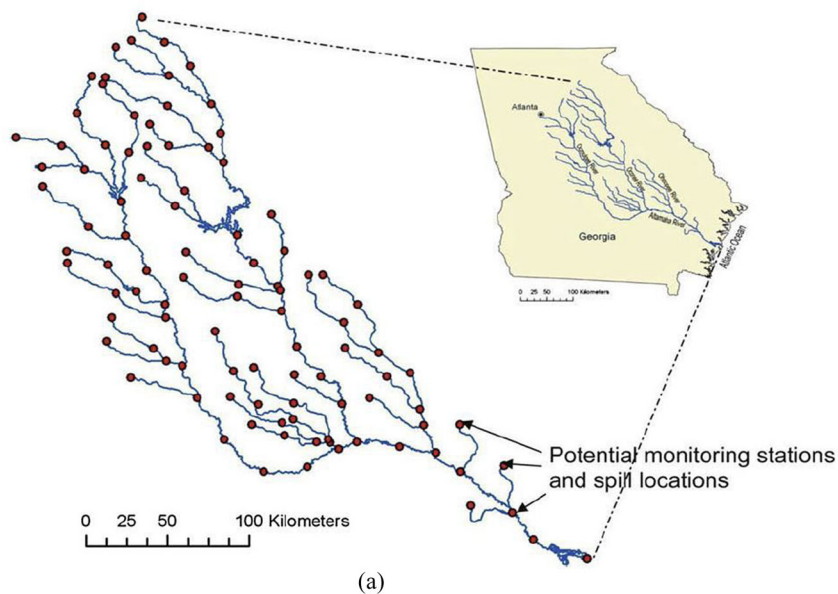


Figure 5. Water quality monitoring using sensor network: (a) shape and monitoring locations, (b) 10 subcatchments of the Altamaha River (Telci and Aral 2011), and (c) an example of stream network with nine stream segments ($i = 1, \dots, 9$) and three locations s_1, s_2, s_3 .

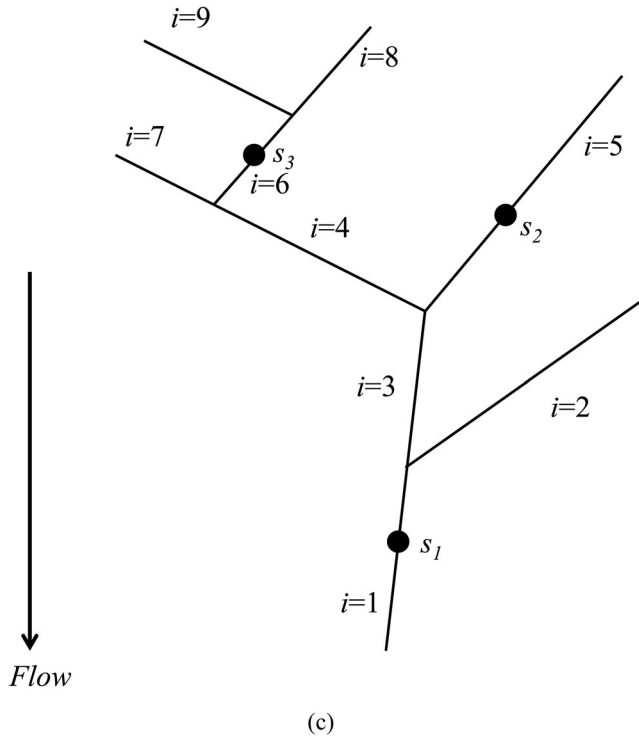


Figure 5. Continued.

weighting is incorporated on the entries of the covariance matrix when the line segments in the network are splitting into multiple segments to ensure that the resulting covariance is stationary.

To explain the tail-up models, we first introduce some notation. A stream network consists of a finite number of stream segments. We index them with $i = 1, 2, \dots$. Denote the index set of stream segment as I and the locations on the network as s_j , $j = 1, 2, \dots$. Let $D_{s_j} \subseteq I$ be the index set of all stream segments that are downstream of location s_j (which means that water from s_j flows into these segments) including the segment containing s_j . Figure 5(c) illustrates a simple stream network with $I = \{1, 2, \dots, 9\}$, $D_{s_1} = \{1\}$, $D_{s_2} = \{1, 3, 5\}$, and $D_{s_3} = \{1, 3, 4, 6\}$. Two locations, s_j and s_k , are said to be “flow connected” if $D_{s_j} \cap D_{s_k} = D_{s_j}$ or D_{s_k} . Finally, define

$$B_{s_j, s_k} = \begin{cases} \overline{(D_{s_j} \cap D_{s_k})} \cap (D_{s_j} \cup D_{s_k}), & \text{if } s_j \text{ and } s_k \text{ are flow-connected;} \\ \emptyset, & \text{otherwise.} \end{cases}$$

Here B_{s_j, s_k} is the set of stream segments between two locations, including the segment for the upstream location but excluding the ones for the downstream location. For example, $B_{s_1, s_3} = \{3, 4, 6\}$ and $B_{s_2, s_3} = \emptyset$. To ensure the stationarity of the variances, Ver Hoef and Peterson (2010) suggested assigning weights to each stream segment in the network. In a stream network, one segment splits into two segments when it goes upstream. For example, segment 1 splits into segments 2 and 3 in Figure 5(c). One way to weight the segments is based on the flow volume of each segment. For example, we

weight segments 2 and 3 by w_2 and w_3 , where $w_2 + w_3 = 1$ and w_2/w_3 equals the ratio of the flow volume between segments 2 and 3. Using tail-up models, the covariance between two locations, s_j and s_k on the stream network is given by

$$C(s_j, s_k | \zeta) = \begin{cases} 0, & \text{if } s_j \text{ and } s_k \text{ are not flow-connected;} \\ \zeta_1, & \text{if } s_j = s_k; \\ \prod_{i \in B_{s_j, s_k}} \sqrt{w_i} \zeta_1 \rho(d(s_j, s_k)/\zeta_2), & \text{otherwise,} \end{cases} \quad (5.1)$$

where $d(s_j, s_k)$ is the stream distance between s_j and s_k , ζ_1 is the variance parameter, $\rho(\cdot|\zeta_2)$ is the correlation function with a parameter ζ_2 , and w_i is the weights on the segment i . The correlation function $\rho(\cdot|\zeta_2)$ can be derived from many commonly used spatial models that we discussed in Section 2.

For illustration, consider the example in Figure 5(c). If an exponential model is used for spatial correlation, the covariance matrix of s_1 , s_2 , and s_3 can be constructed based on (5.1) as follows:

$$\begin{pmatrix} 1 & \sqrt{w_3 w_5} & \sqrt{w_3 w_4 w_6} \\ \sqrt{w_3 w_5} & 1 & 0 \\ \sqrt{w_3 w_4 w_6} & 0 & 1 \end{pmatrix} \odot \begin{pmatrix} \zeta_0 + \zeta_1 & \zeta_1 e^{-d(s_1, s_2)/\zeta_2} & \zeta_1 e^{-d(s_1, s_3)/\zeta_2} \\ \zeta_1 e^{-d(s_1, s_2)/\zeta_2} & \zeta_0 + \zeta_1 & \zeta_1 e^{-d(s_2, s_3)/\zeta_2} \\ \zeta_1 e^{-d(s_1, s_3)/\zeta_2} & \zeta_1 e^{-d(s_2, s_3)/\zeta_2} & \zeta_0 + \zeta_1 \end{pmatrix},$$

where $\odot$ denotes the Hadamard (element-wise) product operation between two matrices.

In our case study, we use the tail-up model with an exponential correlation function to model the data collected at different locations on the Altamaha River network. Both the signal and the noise share the same spatial correlation structure. The spatial covariance matrix for $p = 100$ nodes on the river network is constructed based on the stream distance and flow volume information. We use SWMM to generate data when there is no change and obtain the maximum likelihood estimates for the parameters in the model, $\hat{\zeta}_1 = 0.027$ and $\hat{\zeta}_2 = 0.68$. The covariance matrix is illustrated in Figure 6. For temporal correlation, we use a VAR(1) model $\mathbf{x}_\ell = (1 - \theta)\boldsymbol{\mu} + \theta\mathbf{x}_{\ell-1} + \boldsymbol{\epsilon}_\ell$ to capture the temporal correlation of a contaminant spill as suggested in Clement and Thas (2007) and Clement et al. (2006).

We apply the online change point detection procedure based on $\mathbf{S}^3\mathbf{T}$ to detect contaminant spills in the Altamaha River network. We also compare it with two other methods: (i) online detection based on the quadratic score statistic and (ii) the Hotelling's T^2 chart. Among the 100 nodes on the river network, 10 of them (nodes 1, 15, 19, 33, 36, 50, 58, 67, 84, 95, marked by red stars in Figure 6(c)) are used as possible contaminant spill locations, and the remaining 90 nodes are used for collecting measurements every 15 min. In each replication, we run SWMM to simulate the river network during a 10-day period. A single instantaneous spill is generated, with a spill location randomly selected from the 10 possible locations. The spill starting time is uniformly distributed between the first 15 to 20 h. The intensity of the contaminant spills

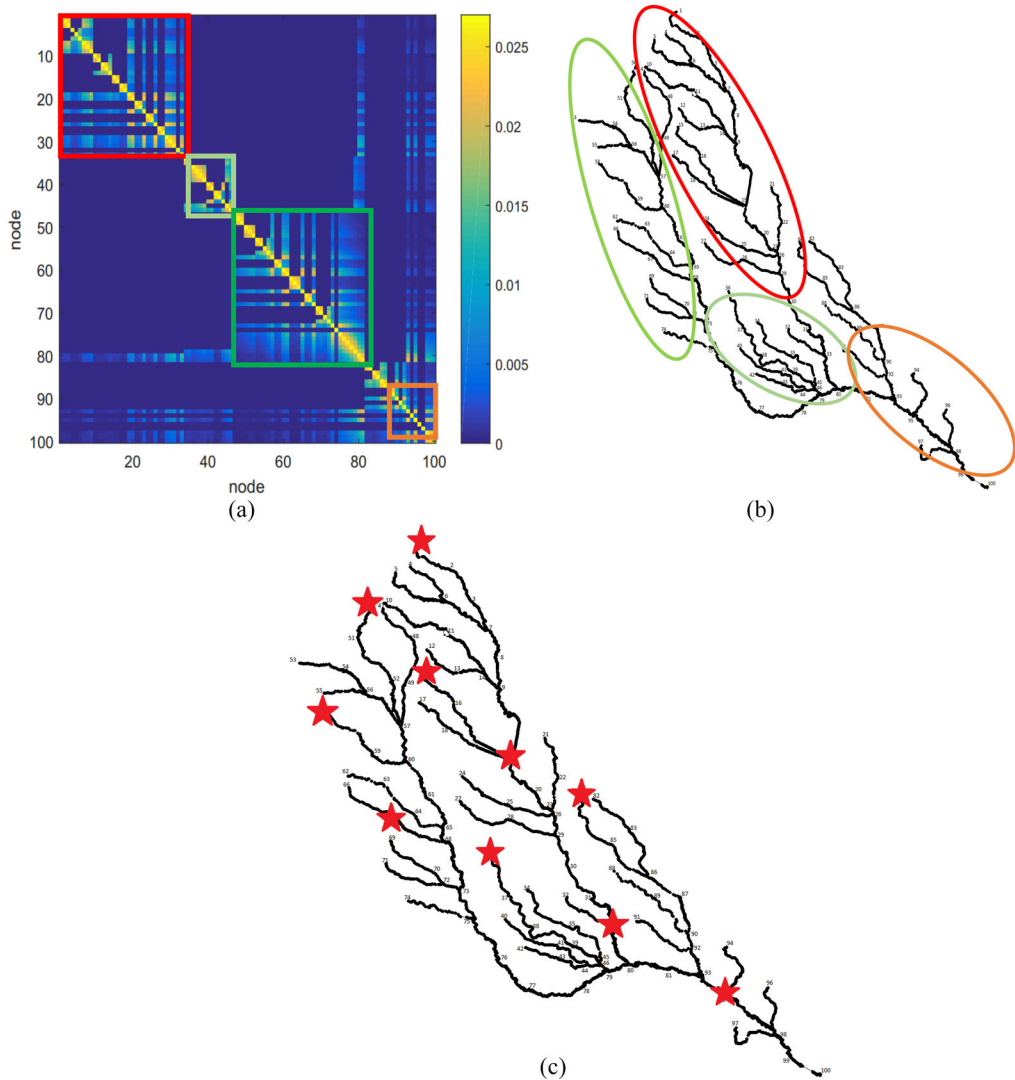


Figure 6. (a) Visualization of the spatial covariance matrix using the tail-up model for 100 sensor networks over the river system; the spatial covariance matrix has a block structure, with blocks in the matrix corresponding to the branches of the river with matching colors in (b); (c) node indexes of the Altamaha River network and potential spill locations marked by red stars.

follows a uniform distribution, and we consider three different levels: $U(10, 100)$ (low), $U(100, 250)$ (medium), and $U(250, 500)$ (high) in units of grams per liter.

The thresholds for the three detection procedures are adjusted so that the in-control ARLs are set to be 10 days (960 samples). For the two procedures based on S^3T and the quadratic score statistic, the length of the sliding window is chosen to be 12.5 h (50 samples). Table 4 reports the average and standard error of detection delays obtained from 100 simulated spills. For spills with high intensity, all three methods achieve similar performance regarding detection delay, because strong signals are easier to detect. However, when the signal is relatively weak (low and medium spill intensity), the proposed detection statistic S^3T significantly outperforms the other two methods.

Table 4. Simulated expected detection delay in hours (numbers in parentheses are standard errors).

Spill intensity	S^3T	Quadratic score statistic	$\bar{T}^2$
Low	38.285 (3.655)	45.822 (4.675)	52.959 (5.035)
Medium	26.301 (1.679)	28.522 (1.873)	30.753 (2.192)
High	25.519 (1.697)	25.489 (1.667)	25.563 (1.860)

Smallest EDD values for each spill intensity shown in bold.

6. CONCLUSIONS

In this article, we propose a novel efficient score statistic S^3T to detect the emergence of a spatial-temporal signal from a noisy background in both the offline and online settings. The statistic captures the spatial and temporal correlation simultaneously and enjoys a relatively low computational cost. An accurate approximation for its false alarm rate is presented. Numerical results based on simulated data, real solar flare data, and a real case study of water quality monitoring show that the proposed S^3T statistic has a clear advantage over existing methods.

APPENDIX A: DERIVATION OF $\frac{\partial \ell}{\partial \mu}|_{\mu=0, \gamma=0}$.

The following propositions are used in the derivation of $\frac{\partial \ell}{\partial \mu}|_{\mu=0, \gamma=0}$.

Proposition A.1. Let $M(t)$ be a nonsingular square matrix whose elements are functions of a scalar parameter α . Then

$$\frac{\partial M(t)^{-1}}{\partial \alpha} = -M(t)^{-1} \frac{\partial M(t)}{\partial \alpha} M(t)^{-1}.$$

Proposition A.2. Let $M(t)$ be a nonsingular square matrix whose elements are functions of a scalar parameter α . Then

$$\frac{\partial |M(t)|}{\partial \alpha} = |M(t)| \text{tr} \left(M(t)^{-1} \frac{\partial M(t)}{\partial \alpha} \right).$$

By Proposition A.1, we can calculate

$$\begin{aligned} \left. \frac{\log |\gamma V_\tau(\theta) + \Sigma_\tau|}{\partial \gamma} \right|_{\mu=0, \gamma=0} &= \frac{1}{|\gamma V_\tau(\theta) + \Sigma_\tau|} \left. |\gamma V_\tau(\theta) + \Sigma_\tau| \text{tr} \left((\gamma V_\tau(\theta) + \Sigma_\tau)^{-1} V_\tau(\theta) \right) \right|_{\gamma=0} \\ &= \text{tr}(\Sigma_\tau^{-1} V_\tau(\theta)). \end{aligned}$$

For convenience, here we use y and μ to denote $y_{(k+1:N)}$ and $\mu_{(k+1:N)}$. By Proposition A.2, we have

$$\begin{aligned} \left. \frac{\partial (y - \mu)^\top (\gamma V_\tau(\theta) + \Sigma_\tau)^{-1} (y - \mu)}{\partial \gamma} \right|_{\mu=0, \gamma=0} &= (y - \mu)^\top \left. \frac{\partial (\gamma V_\tau(\theta) + \Sigma_\tau)^{-1}}{\partial \gamma} (y - \mu) \right|_{\mu=0, \gamma=0} \\ &= -y^\top (\gamma V_\tau(\theta) + \Sigma_\tau)^{-1} V_\tau(\theta) (\gamma V_\tau(\theta) + \Sigma_\tau)^{-1} y|_{\gamma=0} \\ &= -y^\top \Sigma_\tau^{-1} V_\tau(\theta) \Sigma_\tau^{-1} y. \end{aligned}$$

Hence, we have

$$\left. \frac{\partial \ell}{\partial \mu} \right|_{\mu=0, \gamma=0} = -\frac{1}{2} \text{tr}(\Sigma_\tau^{-1} V_\tau(\theta)) + \frac{1}{2} y^\top \Sigma_\tau^{-1} V_\tau(\theta) \Sigma_\tau^{-1} y,$$

as appeared in equation (3.2).

APPENDIX B: DERIVATION OF $\text{Var}[S(\tau, \theta)]$

Here we calculate the variance of the statistic $S(\tau, \theta)$ defined in (3.3). For convenience, we use $\mathbf{y}$ to denote $\mathbf{y}_{(k+1:N)}$, use $\mathbf{V}$ to denote the matrix $\Sigma_\tau^{-1} \mathbf{V}_\tau(\theta) \Sigma_\tau^{-1}$, use Σ to denote Σ_τ , and use c and d to denote $c(\tau, \theta)$ and $d(\tau, \theta)$, respectively. Then we can write

$$S(\tau, \theta) = \frac{(\mathbf{y}^\top \mathbf{V} \mathbf{y} - c)^2}{d} + \mathbf{y}^\top \Sigma^{-1} \mathbf{y}.$$

We have that $\mathbb{E}[S] = p\tau + 1$, and

$$\text{Var}[S(\tau, \theta)] = \mathbb{E}[S^2] - \mathbb{E}[S]^2.$$

In the following, we calculate $\mathbb{E}[S^2]$.

$$\begin{aligned} \mathbb{E}[S^2] &= \mathbb{E} \left[\left(\frac{(\mathbf{y}^\top \mathbf{V} \mathbf{y} - c)^2}{d} + \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \right)^2 \right] \\ &= \mathbb{E} \left[(\mathbf{y}^\top \Sigma^{-1} \mathbf{y})^2 \right] + 2\mathbb{E} \left[\frac{(\mathbf{y}^\top \mathbf{V} \mathbf{y} - c)^2}{d} \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \right] + \mathbb{E} \left[\frac{(\mathbf{y}^\top \mathbf{V} \mathbf{y} - c)^4}{d^2} \right]. \end{aligned} \quad (\text{B.1})$$

The first term can be calculated as

$$\mathbb{E} \left[(\mathbf{y}^\top \Sigma^{-1} \mathbf{y})^2 \right] = p^2 \tau^2 + 2p\tau. \quad (\text{B.2})$$

We then expand the second term,

$$\begin{aligned} \mathbb{E} \left[\frac{(\mathbf{y}^\top \mathbf{V} \mathbf{y} - c)^2}{d} \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \right] &= \frac{1}{d} \mathbb{E} \left[(\mathbf{y}^\top \mathbf{V} \mathbf{y})^2 \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \right] \\ &\quad - \frac{2c}{d} \mathbb{E} [\mathbf{y}^\top \mathbf{V} \mathbf{y} \mathbf{y}^\top \Sigma^{-1} \mathbf{y}] + \frac{c^2}{d} \mathbb{E} [\mathbf{y}^\top \Sigma^{-1} \mathbf{y}]. \end{aligned}$$

We calculate the three expectations separately:

$$\begin{aligned} \mathbb{E} [\mathbf{y}^\top \Sigma^{-1} \mathbf{y}] &= p\tau. \\ \mathbb{E} [\mathbf{y}^\top \mathbf{V} \mathbf{y} \mathbf{y}^\top \Sigma^{-1} \mathbf{y}] &= (p\tau + 2)c. \\ \mathbb{E} [(\mathbf{y}^\top \mathbf{V} \mathbf{y})^2 \mathbf{y}^\top \Sigma^{-1} \mathbf{y}] &= (p\tau + 4)(2d + c^2). \end{aligned}$$

Combining we get

$$\mathbb{E} \left[\frac{(\mathbf{y}^\top \mathbf{V} \mathbf{y} - c)^2}{d} \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \right] = 2p\tau + 4. \quad (\text{B.3})$$

Next we calculate the last term in (B.1),

$$\mathbb{E} \left[\frac{(\mathbf{y}^\top \mathbf{V} \mathbf{y} - c)^4}{d^2} \right] = 3 - 2p\tau - 24 \frac{c}{d^2} \text{tr}(\Sigma^{-1} \mathbf{V} \Sigma^{-1} \mathbf{V} \Sigma^{-1} \mathbf{V}) + \frac{48}{d^2} \text{tr}(\Sigma_\tau^{-1} \mathbf{V} \Sigma^{-1} \mathbf{V} \Sigma^{-1} \mathbf{V} \Sigma^{-1} \mathbf{V}). \quad (\text{B.4})$$

Note that the tedious calculation steps for (B.4) are omitted here.

Combining (B.2), (B.3), and (B.4), we can obtain

$$\text{Var}[S(\tau, \theta)] = 2p\tau + 10 - 24 \frac{c}{d^2} \text{tr}(\Sigma^{-1} \mathbf{V} \Sigma^{-1} \mathbf{V} \Sigma^{-1} \mathbf{V}) + \frac{48}{d^2} \text{tr}(\Sigma^{-1} \mathbf{V} \Sigma^{-1} \mathbf{V} \Sigma^{-1} \mathbf{V} \Sigma^{-1} \mathbf{V}).$$

APPENDIX C: DIAGRAM TO ILLUSTRATE OFFLINE AND ONLINE S^3T STATISTIC

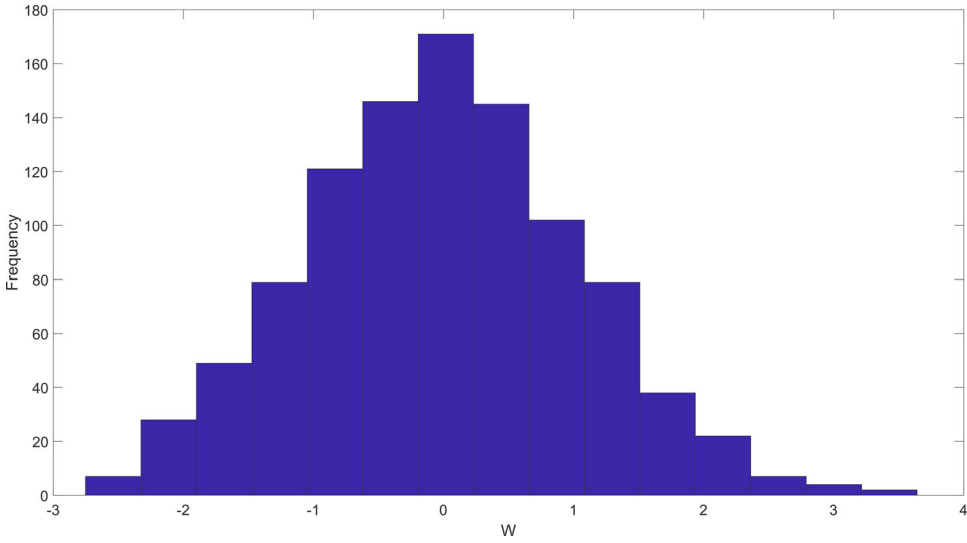


Figure C.1. (a) Diagram showing the construction of offline change point detection statistic and (b) sliding window of length w for online detection.

APPENDIX D: DERIVATION OF THE CUMULANT GENERATING FUNCTION OF W

Here we present the derivation of the cumulant generating function of $W(\tau, \theta)$ under the null hypothesis; that is, [equation \(4.5\)](#).

Let $\mathbf{z} = \Sigma_\tau^{-\frac{1}{2}} \mathbf{y}_{(k+1:N)}$. Under the null hypothesis, $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}_{p\tau})$. For convenience, here we use $\mathbf{B}$ to denote the $p\tau$ by $p\tau$ matrix $\Sigma_\tau^{-\frac{1}{2}} \mathbf{V}_\tau(\theta) \Sigma_\tau^{-\frac{1}{2}}$ and use c and d to denote $c(\tau, \theta)$ and $d(\tau, \theta)$, respectively. Then, we have

$$W(\tau, \theta) = \frac{\mathbf{z}^\top \mathbf{B} \mathbf{z} - c}{\sqrt{d}}.$$

Under the null hypothesis, the cumulant generating function of $W(\tau, \theta)$ can be calculated as

$$\begin{aligned} \psi(\xi) &= \log E[\exp(\xi W(\tau, \theta))] = \log E\left[\exp\left(\xi \left(\frac{\mathbf{z}^\top \mathbf{B} \mathbf{z} - c}{\sqrt{d}}\right)\right)\right] \\ &= -\xi \frac{c}{\sqrt{d}} + \log E\left[\exp\left(\frac{\xi \mathbf{z}^\top \mathbf{B} \mathbf{z}}{\sqrt{d}}\right)\right] \\ &= -\xi \frac{c}{\sqrt{d}} + \log \int_{\mathbf{z}} \exp\left(\frac{\xi \mathbf{z}^\top \mathbf{B} \mathbf{z}}{\sqrt{d}}\right) \frac{1}{(2\pi)^{\frac{p\tau}{2}}} \exp\left(-\frac{1}{2} \mathbf{z}^\top \mathbf{z}\right) d\mathbf{z} \\ &= -\xi \frac{c}{\sqrt{d}} + \log \int_{\mathbf{z}} \frac{1}{(2\pi)^{\frac{p\tau}{2}}} \exp\left(-\frac{1}{2} \mathbf{z}^\top \left(\mathbf{I}_{p\tau} - \frac{2\xi \mathbf{B}}{\sqrt{d}}\right) \mathbf{z}\right) d\mathbf{z} \\ &= -\xi \frac{c}{\sqrt{d}} + \log \left| \mathbf{I}_{p\tau} - \frac{2\xi \mathbf{B}}{\sqrt{d}} \right|^{-\frac{1}{2}}, \end{aligned}$$

which is equivalent to [equation \(4.5\)](#). Note that the last equation uses the fact that

$$\int_{\mathbf{z}} \frac{1}{(2\pi)^{\frac{p\tau}{2}}} \exp\left(-\frac{1}{2}\mathbf{z}^\top\left(\mathbf{I}_{p\tau} - \frac{2\xi\mathbf{B}}{\sqrt{d}}\right)\mathbf{z}\right) d\mathbf{z} = \left|\mathbf{I}_{p\tau} - \frac{2\xi\mathbf{B}}{\sqrt{d}}\right|^{-\frac{1}{2}}.$$

APPENDIX E: PROOF OF THEOREM 4.1

After discretizing the parameter space, $W(\tau, \theta)$ can be treated as a two-dimensional Gaussian random field, which can be completely characterized by its covariance function. The following lemma computes the covariance function of $W(\tau, \theta)$.

Lemma E.1. *Under the null hypothesis, the covariance function of $W(\tau, \theta)$ is*

$$\text{Cov}[W(n, \theta_1), W(m, \theta_2)] = \frac{\text{tr}(\mathbf{A}_n(\theta_1)\mathbf{A}_n(\theta_2))}{[\text{tr}(\mathbf{A}_n(\theta_1)\mathbf{A}_n(\theta_1))\text{tr}(\mathbf{A}_m(\theta_2)\mathbf{A}_m(\theta_2))]^{1/2}}, \quad (\text{E.1})$$

where $n \leq m$.

The following lemma shows that the first-order approximation of the covariance function in (E.1) does not have any cross-product term. Thus, the two-dimensional random field is further decomposed as a sum of two independent one-dimensional random processes.

Lemma E.2. *Assuming that δ and $i \in \mathbb{Z}$ are small relative to θ and τ , respectively, the first-order approximation of the covariance function in (E.1) is given as*

$$\text{Cov}[W(\tau, \theta), W(\tau + i, \theta + \delta)] \approx 1 - \gamma^2(\tau, \theta)\delta^2 - \frac{\mu(\tau, \theta)}{2\tau}i + o(\delta^2) + o(i), \quad (\text{E.2})$$

where

$$\gamma(\tau, \theta) = \frac{\text{tr}(\dot{\mathbf{A}}_\tau(\theta)\mathbf{A}_\tau(\theta))}{\text{tr}(\mathbf{A}_\tau(\theta)\mathbf{A}_\tau(\theta))}, \quad (\text{E.3})$$

$\mu(\tau, \theta)$ is defined in (4.2), and $\dot{\mathbf{A}}_\tau(\theta) = \partial\mathbf{A}_\tau(\theta)/\partial\theta$.

The following two lemmas are needed in the proof. Both lemmas are proved in Xie and Siegmund (2012).

Lemma E.3. *Assume $\xi \rightarrow \infty$, $b \rightarrow \infty$, $N \rightarrow \infty$, with $\frac{\xi}{b} \approx 1$ and $\frac{b}{N} \approx c$, where $c > 0$ is some constant. The discretized process $b\left[W\left(\tau + i, \theta + \frac{\Delta}{\sqrt{N}j}\right) - \xi\right]$, where i is an integer and $j \geq 0$, conditioned on $W(\tau, \theta) = \xi$, can be written as a sum of two independent processes:*

$$\left\{b\left[W\left(\tau + i, \theta + \frac{\Delta}{\sqrt{N}j}\right) - \xi\right] \middle| W(\tau, \theta) = \xi\right\} = S_i + V_j,$$

where $S_i = \sum_{\ell=1}^i a_\ell$, with

$$a_\ell \sim N\left(-\frac{\mu(\tau, \theta)}{2\tau}b^2, \frac{\mu(\tau, \theta)}{\tau}b^2\right),$$

and

$$V_j = \sqrt{2}\gamma(\tau, \theta)\frac{b}{\sqrt{N}}\Delta jV - \gamma^2(\tau, \theta)\frac{b^2}{N}\Delta^2j^2,$$

with $V \sim N(0, 1)$. $\mu(\tau, \theta)$ and $\gamma(\tau, \theta)$ are defined in (4.2) and (E.3), respectively.

Lemma E.4. Assume that $x_1, x_2, \dots$ are independent and identically distributed $N(-\mu_1, \sigma_1^2)$ random variables ($\mu_1 > 0$). Define the random walk $S_0 = 0$, $S_i = \sum_{\ell=1}^i x_\ell$, $i = 1, 2, \dots$, and the smooth varying random process $V_j = \beta \Delta j V - \frac{\beta^2}{2} \Delta^2 j^2$, for some constants $\Delta > 0$, $\beta > 0$. As $\Delta \rightarrow 0$, for some constant α , we have

$$\frac{1}{\Delta} \int_0^\infty e^{-\alpha x} \mathbb{P}\left(\max_{i \geq 1} S_i \leq -x\right) \mathbb{P}\left(\max_{i \leq 0} S_i + \max_{j \geq 1} V_j \leq -x\right) dx \xrightarrow{\Delta \rightarrow 0} \frac{|\beta|}{\sqrt{2\pi}} \left(\frac{2\mu_1^2}{\sigma_1^2}\right) \nu\left(\frac{2\mu_1}{\sigma_1}\right),$$

where $\nu(x)$ is defined in (4.1).

In the following, we go through the main steps that lead to the approximation of the false alarm rate in Theorem 4.1 for the case of $d = 1$.

Step 1: We first discretize the parameter $\theta \in [\theta_1, \theta_2]$ by a rectangular mesh grid of size $\frac{\Delta}{\sqrt{N}}$, where $\Delta > 0$ is a small number. Note that the discretization mentioned here is used for asymptotic analysis only. The probability of false alarm can be approximated as

$$\mathbb{P}\left(\max_{(i,j) \in D} W\left(i, j \frac{\Delta}{\sqrt{N}}\right) \geq b\right), \quad (\text{E.4})$$

where D is the index set

$$D = \left\{(i, j) : 1 \leq i \leq N, \theta_1 \leq j \frac{\Delta}{\sqrt{N}} \leq \theta_2\right\},$$

which covers the entire parameter space. Let $J(i_0, j_0)$ denote everything to the “future” of the current index (i_0, j_0) in the parameter space; that is,

$$J(i_0, j_0) = \{(i, j) \in D : j \geq j_0, \text{ or } i \geq i_0 \text{ and } j = j_0\}.$$

Using a similar approach as in Siegmund (1988), the event

$$\left\{\max_{(i,j) \in D} W\left(i, j \frac{\Delta}{\sqrt{N}}\right) \geq b\right\}$$

can be decomposed into a series of “last hitting events” for which (i_0, j_0) is the “last” location where $W\left(i, j \frac{\Delta}{\sqrt{N}}\right)$ hits the threshold b . Then, the probability in (E.4) can be written as the sum of probabilities of $W\left(i, j \frac{\Delta}{\sqrt{N}}\right)$ last hits b at (i_0, j_0) over all possible (i_0, j_0) :

$$\begin{aligned} \mathbb{P}\left(\max_{(i,j) \in D} W\left(i, j \frac{\Delta}{\sqrt{N}}\right) \geq b\right) &\approx \sum_{(i_0, j_0) \in D} \mathbb{P}\left(W\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right) \geq b, \max_{(i,j) \in J(i_0, j_0)} W\left(i, j \frac{\Delta}{\sqrt{N}}\right) < b\right) \\ &= \sum_{(i_0, j_0) \in D} \int_0^\infty \mathbb{P}\left(W\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right) \\ &\quad \times \mathbb{P}\left(\max_{(i,j) \in J(i_0, j_0)} W\left(i, j \frac{\Delta}{\sqrt{N}}\right) < b \mid W\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right) \frac{dx}{b}. \end{aligned} \quad (\text{E.5})$$

Step 2: In the following, we obtain an approximation on the probability

$$\mathbb{P}\left(W\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right) \frac{dx}{b}.$$

To simplify the notation, we denote $W\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right)$ as W here. The key idea is to approximate W as a Gaussian random field. The Gaussian approximation performs well when the probability of

interest is close to the mean of the true distribution but suffers from deviation if the probability is in the tail of the true distribution. Hence, we apply the change of measure technique to shift the mean of the random field W to the threshold b .

Denote the cumulant generating function of W as $\psi(\xi) = \log E[\exp(\xi W)]$. To construct the new probability measure, we first choose a $\xi_0 > 0$ such that $\psi'(\xi) = b$. The new probability measure dF_{ξ_0} is constructed using exponential embedding, as follows:

$$dF_{\xi_0} = \exp(\xi_0 W - \psi(\xi_0)) dF,$$

where dF is the original distribution of W . Let E_{ξ_0} and $\mathbb{P}_{\xi_0}$ denote the expectation and probability under the new measure dF_{ξ_0} , respectively. It can be verified that under the new measure

$$E_{\xi_0}[W] = E[W \exp(\xi_0 W - \psi(\xi_0))] = e^{-\psi(\xi_0)} \frac{\partial e^{\psi(\xi)}}{\partial \xi} \Big|_{\xi=\xi_0} = \psi'(\xi) = b,$$

namely, the mean of W is close to the threshold b under the new probability measure.

The threshold crossing probability can be rewritten as

$$\begin{aligned} \mathbb{P}\left(W = b + \frac{x}{b}\right) &= E_{\xi_0} \left[\frac{1}{\exp[\xi_0 W - \psi(\xi_0)]} 1\left\{W = b + \frac{x}{b}\right\} \right] \\ &= \exp\left[\psi(\xi_0) - \xi_0 \left(b + \frac{x}{b}\right)\right] \mathbb{P}_{\xi_0}\left(W = b + \frac{x}{b}\right). \end{aligned} \quad (\text{E.6})$$

Now we can apply the Gaussian approximation to obtain $\mathbb{P}_{\xi_0}(W = b + \frac{x}{b})$ and use (E.6) to get the original probability. By treating W as a normal random variable with mean b and variance $\sigma_{\xi_0}^2$, we have

$$\mathbb{P}_{\xi_0}\left(W = b + \frac{x}{b}\right) = \frac{1}{\sqrt{2\pi}\sigma_{\xi_0}} \exp\left(\frac{-x^2}{2b^2\sigma_{\xi_0}^2}\right) \approx \frac{1}{\sqrt{2\pi}\sigma_{\xi_0}}.$$

Note that in (E.5), the integrands with smaller x values contribute more to the integration, because the integrand decays exponentially fast with x . Now, when $b \rightarrow \infty$, $\frac{x}{b} \rightarrow 0$ for small x , and hence $\exp\left(\frac{-x^2}{2b^2\sigma_{\xi_0}^2}\right) \rightarrow 1$. The above argument is similar to those used for Laplace's method.

The cumulant generating function of W can be calculated as

$$\psi(\xi) = -\xi \frac{\text{tr}(\Sigma_{\tau}^{-1} \mathbf{V}_{\tau}(\theta))}{[2\text{tr}(\Sigma_{\tau}^{-1} \mathbf{V}_{\tau}(\theta) \Sigma_{\tau}^{-1} \mathbf{V}_{\tau}(\theta))]^{1/2}} - \frac{1}{2} \log \left| \mathbf{I}_{p_{\tau}} - \frac{2\xi \Sigma_{\tau}^{1/2} \mathbf{V}_{\tau}(\theta) \Sigma_{\tau}^{1/2}}{[2\text{tr}(\Sigma_{\tau}^{-1} \mathbf{V}_{\tau}(\theta) \Sigma_{\tau}^{-1} \mathbf{V}_{\tau}(\theta))]^{1/2}} \right|.$$

Hence, ξ_0 can be obtained by solving the following equation numerically:

$$\frac{1}{\sqrt{d(\tau, \theta)}} \text{tr} \left(\left[\mathbf{I}_{p_{\tau}} - \frac{2\xi_0 \mathbf{B}_{\tau}(\theta)}{\sqrt{d(\tau, \theta)}} \right]^{-1} \mathbf{B}_{\tau}(\theta) - \mathbf{A}_{\tau}(\theta) \right) = b.$$

Eventually, we have

$$\mathbb{P}\left(W\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right) \approx g(i_0, j_0) \exp\left(-\frac{\xi_0}{b} x\right), \quad (\text{E.7})$$

where $g()$ follows the definition in (4.4).

Step 3: Next we tackle with the conditional probability

$$\mathbb{P}\left(\max_{(i,j) \in J(i_0, j_0)} W\left(i, j \frac{\Delta}{\sqrt{N}}\right) < b \mid W\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right).$$

The first-order expansion of the covariance function given by Lemma E.2 does not have any cross-product term, which implies that if we approximate $W(\tau, \theta)$ as a Gaussian random field, it can be decomposed as a sum of two independent one-dimensional random processes.

By Lemma E.3, the conditional probability can be written in terms of the decomposed random processes using the techniques in Siegmund (1988) and H.-J. Kim and Siegmund (1989) as follows:

$$\begin{aligned}
 & \mathbb{P}\left(\max_{(i,j) \in J(i_0, j_0)} W\left(i, j \frac{\Delta}{\sqrt{N}}\right) < b \mid W\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right) \\
 &= \mathbb{P}\left(\max_{(i,j) \in J(i_0, j_0)} b \left[W\left(i, j \frac{\Delta}{\sqrt{N}}\right) - W\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right)\right] \leq -x \mid W\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right) \\
 &\approx \mathbb{P}\left(\max_{i \geq 1} S_i \leq -x\right) \mathbb{P}\left(\max_{i \leq 0} S_i + \max_{j \geq 1} V_j \leq -x\right).
 \end{aligned} \tag{E.8}$$

Step 4: Combining the approximations in (E.7) and (E.8), the approximated false alarm rate becomes

$$\begin{aligned}
 & \mathbb{P}\left(\max_{(i,j) \in D} W\left(i, j \frac{\Delta}{\sqrt{N}}\right) \geq b\right) \\
 &\approx \sum_{(i_0, j_0) \in D} g\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right) \frac{\Delta}{\sqrt{N}} \frac{\sqrt{N}}{\Delta b} \int_0^\infty \exp\left(-\frac{\xi_0}{b} x\right) \\
 &\quad \times \mathbb{P}\left(\max_{i \geq 1} S_i \leq -x\right) \mathbb{P}\left(\max_{i \leq 0} S_i + \max_{j \geq 1} V_j \leq -x\right) dx.
 \end{aligned} \tag{E.9}$$

Lemma E.4 enables us to find an expression for the integration in (E.9).

Finally, by Lemma E.4 with $\alpha = \frac{\xi_0}{b}$, $\beta = \sqrt{2}\gamma(\tau, \theta) \frac{b}{\sqrt{N}}$, $\mu_1 = \frac{\mu(\tau, \theta)}{2\tau} b^2$ and $\sigma_1^2 = \frac{\mu(\tau, \theta)}{\tau} b^2$, we have the approximated significance level

$$\frac{1}{2\sqrt{\pi}} \sum_{(i_0, j_0) \in D} g\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right) \frac{b^2 \mu\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right)}{N - i_0} \cdot \nu\left(\sqrt{\frac{b^2 \mu\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right)}{N - i_0}}\right) \gamma\left(i_0, j_0 \frac{\Delta}{\sqrt{N}}\right) \frac{\Delta}{\sqrt{N}}. \tag{E.10}$$

As $\Delta \rightarrow 0$, the Riemann sum (E.10) converges to the approximation in Theorem 4.1.

APPENDIX F: PROOF OF LEMMA E.1: COVARIANCE FUNCTION OF W

Proof. Let $\mathbf{C}_\tau(\theta) = \Sigma_\tau^{-1} \mathbf{V}_\tau(\theta) \Sigma_\tau^{-1}$, and rewrite $\mathbf{C}_m(\theta_2)$ as

$$\mathbf{C}_m(\theta_2) = \begin{bmatrix} \mathbf{C}_{11}(\theta_2) & \mathbf{C}_{12}(\theta_2) \\ \mathbf{C}_{21}(\theta_2) & \mathbf{C}_n(\theta_2) \end{bmatrix}.$$

Denote $\mathbf{y}_{(T-\tau+1:T)}$ as $\mathbf{Y}_\tau$, and let

$$\mathbf{Y}_m = \begin{bmatrix} \mathbf{Y}_\Delta \\ \mathbf{Y}_n \end{bmatrix}.$$

We have

$$\begin{aligned}
 \text{Cov}[W(n, \theta_1), W(m, \theta_2)] &= \frac{\mathbb{E}[\mathbf{Y}_n^\top \mathbf{C}_n(\theta_1) \mathbf{Y}_n \mathbf{Y}_m^\top \mathbf{C}_m(\theta_2) \mathbf{Y}_m] - \mathbb{E}[\mathbf{Y}_n^\top \mathbf{C}_n(\theta_1) \mathbf{Y}_n] \mathbb{E}[\mathbf{Y}_m^\top \mathbf{C}_m(\theta_2) \mathbf{Y}_m]}{2(\text{tr}\{\mathbf{A}_n(\theta_1) \mathbf{A}_n(\theta_1)\} \text{tr}\{\mathbf{A}_m(\theta_2) \mathbf{A}_m(\theta_2)\})^{1/2}}.
 \end{aligned} \tag{F.1}$$

The first term in the numerator is

$$\begin{aligned}
 & \mathbb{E}[Y_n^\top \mathbf{C}_n(\theta_1) Y_n Y_m^\top \mathbf{C}_m(\theta_2) Y_m] \\
 &= \mathbb{E}\left[(Y_n^\top \mathbf{C}_n(\theta_1) Y_n)(Y_\Delta^\top \mathbf{C}_{11}(\theta_2) Y_\Delta + Y_n^\top \mathbf{C}_n(\theta_2) Y_n + Y_n^\top \mathbf{C}_{21}(\theta_2) Y_\Delta + Y_\Delta^\top \mathbf{C}_{12}(\theta_2) Y_n)\right] \\
 &= \mathbb{E}[Y_n^\top \mathbf{C}_n(\theta_1) Y_n Y_n^\top \mathbf{C}_n(\theta_2) Y_n] + \mathbb{E}[Y_n^\top \mathbf{C}_n(\theta_1) Y_n] \mathbb{E}[Y_\Delta^\top \mathbf{C}_{11}(\theta_2) Y_\Delta] \\
 &= 2\text{tr}\{\mathbf{A}_n(\theta_1) \mathbf{A}_n(\theta_2)\} + \text{tr}\{\mathbf{A}_n(\theta_1)\} \text{tr}\{\mathbf{A}_n(\theta_2)\} + \mathbb{E}[Y_n^\top \mathbf{C}_n(\theta_1) Y_n] \mathbb{E}[Y_\Delta^\top \mathbf{C}_{11}(\theta_2) Y_\Delta].
 \end{aligned} \tag{F.2}$$

Note that we have utilized the fact that under null hypothesis, Y_Δ and Y_n are independent and $\mathbb{E}[Y_\Delta] = 0$.

The second term in the numerator is

$$\begin{aligned}
 & \mathbb{E}[Y_n^\top \mathbf{C}_n(\theta_1) Y_n] \mathbb{E}[Y_m^\top \mathbf{C}_m(\theta_2) Y_m] \\
 &= \mathbb{E}[Y_n^\top \mathbf{C}_n(\theta_1) Y_n] \mathbb{E}[Y_\Delta^\top \mathbf{C}_{11}(\theta_2) Y_\Delta + Y_n^\top \mathbf{C}_n(\theta_2) Y_n + Y_n^\top \mathbf{C}_{21}(\theta_2) Y_\Delta + Y_\Delta^\top \mathbf{C}_{12}(\theta_2) Y_n] \\
 &= \mathbb{E}[Y_n^\top \mathbf{C}_n(\theta_1) Y_n] \mathbb{E}[Y_n^\top \mathbf{C}_n(\theta_2) Y_n] + \mathbb{E}[Y_n^\top \mathbf{C}_n(\theta_1) Y_n] \mathbb{E}[Y_\Delta^\top \mathbf{C}_{11}(\theta_2) Y_\Delta] \\
 &= \text{tr}\{\mathbf{A}_n(\theta_1)\} \text{tr}\{\mathbf{A}_n(\theta_2)\} + \mathbb{E}[Y_n^\top \mathbf{C}_n(\theta_1) Y_n] \mathbb{E}[Y_\Delta^\top \mathbf{C}_{11}(\theta_2) Y_\Delta].
 \end{aligned} \tag{F.3}$$

By combining (F.1), (F.2), (F.3), we obtain the covariance function in Lemma E.1.

APPENDIX G: PROOF OF LEMMA E.2: FIRST-ORDER EXPANSION OF THE COVARIANCE FUNCTION OF W

Proof. We approximate the covariance function by expanding each term in (E.1) at θ and keeping only the first-order terms.

The numerator in (E.1) is approximated as

$$\begin{aligned}
 & \text{tr}(\mathbf{A}_\tau(\theta + \delta) \mathbf{A}_\tau(\theta)) \approx \text{tr}(\mathbf{A}_\tau(\theta) \mathbf{A}_\tau(\theta)) + \delta \text{tr}(\dot{\mathbf{A}}_\tau(\theta) \mathbf{A}_\tau(\theta)) \\
 &= \text{tr}(\mathbf{A}_\tau(\theta) \mathbf{A}_\tau(\theta)) (1 + \delta \gamma(\tau, \theta)).
 \end{aligned} \tag{G.1}$$

Partition the matrix $\mathbf{A}_{\tau+i}(\theta + \delta)$ as follows:

$$\mathbf{A}_{\tau+i}(\theta + \delta) = \begin{bmatrix} \mathbf{A}_{11}(\theta + \delta) & \mathbf{A}_{12}(\theta + \delta) \\ \mathbf{A}_{21}(\theta + \delta) & \mathbf{A}_\tau(\theta + \delta) \end{bmatrix}.$$

Then rewrite the second term in the denominator in (E.1) as

$$\begin{aligned}
 & \text{tr}(\mathbf{A}_{\tau+i}(\theta + \delta) \mathbf{A}_{\tau+i}(\theta + \delta)) \\
 &= \text{tr}(\mathbf{A}_{11}(\theta + \delta) \mathbf{A}_{11}(\theta + \delta)) + \text{tr}(\mathbf{A}_{12}(\theta + \delta) \mathbf{A}_{21}(\theta + \delta)) \\
 &+ \text{tr}(\mathbf{A}_{21}(\theta + \delta) \mathbf{A}_{12}(\theta + \delta)) + \text{tr}(\mathbf{A}_\tau(\theta + \delta) \mathbf{A}_\tau(\theta + \delta)).
 \end{aligned}$$

After expanding each term at θ , the denominator in (E.1) can be approximated as

$$\left[\text{tr}(\mathbf{A}_\tau(\theta) \mathbf{A}_\tau(\theta)) \text{tr}(\mathbf{A}_{\tau+i}(\theta + \delta) \mathbf{A}_{\tau+i}(\theta + \delta)) \right]^{1/2} \approx \text{tr}(\mathbf{A}_\tau(\theta) \mathbf{A}_\tau(\theta)) \sqrt{1 + 2\delta a \sqrt{1 + b}}, \tag{G.2}$$

where

$$a = \frac{\text{tr}(\dot{\mathbf{A}}_{11}(\theta) \mathbf{A}_{11}(\theta)) + \text{tr}(\dot{\mathbf{A}}_{12}(\theta) \mathbf{A}_{21}(\theta)) + \text{tr}(\dot{\mathbf{A}}_{21}(\theta) \mathbf{A}_{12}(\theta)) + \text{tr}(\dot{\mathbf{A}}_\tau(\theta) \mathbf{A}_\tau(\theta))}{\text{tr}(\mathbf{A}_{11}(\theta) \mathbf{A}_{11}(\theta)) + \text{tr}(\mathbf{A}_{12}(\theta) \mathbf{A}_{21}(\theta)) + \text{tr}(\mathbf{A}_{21}(\theta) \mathbf{A}_{12}(\theta)) + \text{tr}(\mathbf{A}_\tau(\theta) \mathbf{A}_\tau(\theta))}, \tag{G.3}$$

and

$$b = \frac{2i \frac{1}{2i\tau} \left[\text{tr}(\mathbf{A}_{\tau+i}(\theta) \mathbf{A}_{\tau+i}(\theta)) - \text{tr}(\mathbf{A}_\tau(\theta) \mathbf{A}_\tau(\theta)) \right]}{\frac{1}{\tau^2} \text{tr}(\mathbf{A}_\tau(\theta) \mathbf{A}_\tau(\theta))}. \tag{G.4}$$

Because i and δ are small compared to τ and θ , the terms $\text{tr}(\mathbf{A}_{\cdot\tau}(\theta)\mathbf{A}_\tau(\theta))$ and $\text{tr}(\mathbf{A}_\tau(\theta)\mathbf{A}_\tau(\theta))$ are relatively larger than the subdiagonal elements in (G.3) and, hence, a can be further approximated as

$$a \approx \frac{\text{tr}(\mathbf{A}_{\cdot\tau}(\theta)\mathbf{A}_\tau(\theta))}{\text{tr}(\mathbf{A}_\tau(\theta)\mathbf{A}_\tau(\theta))}. \quad (\text{G.5})$$

Meanwhile, we approximate the term $\frac{1}{i\tau} \left[\text{tr}(\mathbf{A}_{\tau+i}(\theta)\mathbf{A}_{\tau+i}(\theta)) - \text{tr}(\mathbf{A}_\tau(\theta)\mathbf{A}_\tau(\theta)) \right]$ in (G.4) using $\frac{1}{\tau} \left[\text{tr}(\mathbf{A}_{\tau+1}(\theta)\mathbf{A}_{\tau+1}(\theta)) - \text{tr}(\mathbf{A}_\tau(\theta)\mathbf{A}_\tau(\theta)) \right]$, and then we have

$$b \approx \frac{i}{\tau} \mu(\tau, \theta). \quad (\text{G.6})$$

The argument for the above approximation is as follows. First, note that

$$\begin{aligned} \mathbf{A}_{\tau+i}(\theta) &= \boldsymbol{\Sigma}_{\tau+i}^{-1} \mathbf{V}_{\tau+i}(\theta) = (\mathbf{I}_{\tau+i} \otimes \boldsymbol{\Sigma})^{-1} (\mathbf{R}_{\tau+i}(\theta) \otimes \boldsymbol{\Lambda}) \\ &= \mathbf{R}_{\tau+i}(\theta) \otimes (\boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda}). \end{aligned}$$

Then we have

$$\begin{aligned} \text{tr}(\mathbf{A}_{\tau+i}(\theta)\mathbf{A}_{\tau+i}(\theta)) &= \text{tr}((\mathbf{R}_{\tau+i}(\theta) \otimes (\boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda}))(\mathbf{R}_{\tau+i}(\theta) \otimes (\boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda}))) \\ &= \text{tr}((\mathbf{R}_{\tau+i}(\theta)\mathbf{R}_{\tau+i}(\theta)) \otimes (\boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda})) \\ &= \text{tr}(\mathbf{R}_{\tau+i}(\theta)\mathbf{R}_{\tau+i}(\theta)) \text{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda}) \\ &= \text{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda}) \sum_j \sum_k [\mathbf{R}_{\tau+i}(\theta)]_{jk}^2 \\ &= \text{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda}) \left(i \sum_j \sum_k [\mathbf{R}_{\tau+1}(\theta)]_{jk}^2 + \sum_{|j-k|>\tau} [\mathbf{R}_{\tau+1}(\theta)]_{jk}^2 \right) \\ &\approx \text{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda}) \left(i \sum_j \sum_k [\mathbf{R}_{\tau+1}(\theta)]_{jk}^2 \right). \end{aligned}$$

The last approximation is due to the fact that the (j, k) th element of $\mathbf{R}_{\tau+1}(\theta)$ such that $|j - k| > \tau$ is small.

Combining (G.1), (G.2), (G.5), (G.6) and the Taylor expansion $\frac{1}{\sqrt{1+x}} \approx 1 - \frac{1}{2}x + o(x)$, we obtain the approximation in (E.2).

DISCLOSURE

The authors have no conflicts of interest to report.

FUNDING

This work is sponsored by NSF CMMI-1538746.

ORCID

Yao Xie  <http://orcid.org/0000-0001-6777-2951>

REFERENCES

- Abramowitz, M., and I. A. Stegun. 1970. "Bessel Functions of Integer Order." In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun, chapter 9, 355–434. New York: Dover Publications.

- Brockwell, P. J., and R. A. Davis. 1991. "Stationary ARMA Processes." In *Time Series: Theory and Methods*, chapter 3, 77–110. New York: Springer.
- Clement, L., and O. Thas. 2007. "Estimating and Modeling Spatio-Temporal Correlation Structures for River Monitoring Networks." *Journal of Agricultural, Biological, and Environmental Statistics* 12 (2):161–76. doi:[10.1198/108571107X197977](https://doi.org/10.1198/108571107X197977)
- Clement, L., O. Thas, P. Vanrolleghem, and J. Otttoy. 2006. "Spatio-Temporal Statistical Models for River Monitoring Networks." *Water Science and Technology: A Journal of the International Association on Water Pollution Research* 53 (1):9–15. doi:[10.2166/wst.2006.002](https://doi.org/10.2166/wst.2006.002)
- Crosier, R. B. 1988. "Multivariate Generalizations of Cumulative Sum Quality-Control Schemes." *Technometrics* 30 (3):291–303. doi:[10.1080/00401706.1988.10488402](https://doi.org/10.1080/00401706.1988.10488402)
- Gaetan, C., and X. Guyon. 2010. "Second-Order Spatial Models and Geostatistics." In *Spatial Statistics and Modeling*, chapter 1, 1–52. New York: Springer.
- Genton, M. G. 2007. "Separable Approximations of Space-Time Covariance Matrices." *Environmetrics* 18 (7):681–95. doi:[10.1002/env.854](https://doi.org/10.1002/env.854)
- Healy, J. D. 1987. "A Note on Multivariate CUSUM Procedures." *Technometrics* 29 (4):409–12. doi:[10.1080/00401706.1987.10488268](https://doi.org/10.1080/00401706.1987.10488268)
- Hotelling, H. 1947. "Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights." In *Techniques of Statistical Analysis*, edited by Eisenhart, C., Hastay, M.W. and Wallis, W.A., 111–84. New York: McGraw Hill.
- Jiang, W., S. W. Han, K.-L. Tsui, and W. H. Woodall. 2011. "Spatiotemporal Surveillance Methods in the Presence of Spatial Correlation." *Statistics in Medicine* 30 (5):569–83. doi:[10.1002/sim.3877](https://doi.org/10.1002/sim.3877)
- Kim, H.-J., and D. Siegmund. 1989. "The Likelihood Ratio Test for a Change-Point in Simple Linear Regression." *Biometrika* 76 (3):409–23. doi:[10.1093/biomet/76.3.409](https://doi.org/10.1093/biomet/76.3.409)
- Kim, S.-H., M. M. Aral, Y. Eun, J. J. Park, and C. Park. 2017. "Impact of Sensor Measurement Error on Sensor Positioning in Water Quality Monitoring Networks." *Stochastic Environmental Research and Risk Assessment* 31 (3):743–56. doi:[10.1007/s00477-016-1210-1](https://doi.org/10.1007/s00477-016-1210-1)
- Lee, M. L., D. Goldsman, and S.-H. Kim. 2015. "Robust Distribution-Free Multivariate CUSUM Charts for Spatiotemporal Biosurveillance in the Presence of Spatial Correlation." *IIE Transactions on Healthcare Systems Engineering* 5 (2):74–88. doi:[10.1080/19488300.2015.1017674](https://doi.org/10.1080/19488300.2015.1017674)
- Lee, M. L., D. Goldsman, S.-H. Kim, and K.-L. Tsui. 2014. "Spatiotemporal Biosurveillance with Spatial Clusters: control Limit Approximation and Impact of Spatial Correlation." *IIE Transactions* 46 (8):813–27. doi:[10.1080/0740817X.2013.785296](https://doi.org/10.1080/0740817X.2013.785296)
- National Aeronautics and Space Administration. 2013. (Retrieved 7- 2012. 30 SDO Instruments, <http://sdo.gsfc.nasa.gov/mission/instruments.php>.
- Rao, C. R. 1948. "Large Sample Tests of Statistical Hypotheses concerning Several Parameters with Applications to Problems of Estimation." *Mathematical Proceedings of the Cambridge Philosophical Society* 44:50–7.
- Rao, C. R., and S. J. Poti. 1946. "On Locally Most Powerful Tests When Alternatives Are One Sided, *Sankhyā*." *The Indian Journal of Statistics* 7:439.
- Rossmann, L. A. 2010. *Storm Water Management Model User's Manual, Version 5.0*, National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency, Washington, DC.
- Shewhart, W. A. 1931. *Economic Control of Quality of Manufactured Product*, London: Macmillan.
- Siegmund, D. 1985. *Sequential Analysis: Tests and Confidence Intervals*, Springer: New York.
- Siegmund, D. 1988. "Approximate Tail Probabilities for the Maxima of Some Random Fields." *The Annals of Probability* 16 (2):487–501. doi:[10.1214/aop/1176991769](https://doi.org/10.1214/aop/1176991769)
- Siegmund, D., and E. Venkatraman. 1995. "Using the Generalized Likelihood Ratio Statistic for Sequential Detection of a Change-Point." *The Annals of Statistics* 23 (1):255–271. doi:[10.1214/aos/1176324466](https://doi.org/10.1214/aos/1176324466)
- Siegmund, D., and B. Yakir. 2007. *The Statistics of Gene Mapping*, New York: Springer.

- Siegmund, D., and B. Yakir. 2008. "Detecting the Emergence of a Signal in a Noisy Image." *Statistics and Its Inference* 1:3–12.
- Telci, I. T., and M. M. Aral. 2011. "Contaminant Source Location Identification in River Networks Using Water Quality Monitoring Systems for Exposure Analysis." *Water Quality, Exposure and Health* 2 (3-4):205–218. doi:[10.1007/s12403-011-0039-6](https://doi.org/10.1007/s12403-011-0039-6)
- Ver Hoef, J. M., and E. E. Peterson. 2010. "A Moving Average Approach for Spatial Statistical Models of Stream Networks." *Journal of the American Statistical Association* 105 (489):6–18. doi:[10.1198/jasa.2009.ap08248](https://doi.org/10.1198/jasa.2009.ap08248)
- Xie, Y., and D. Siegmund. 2012. "Spectrum Opportunity Detection with Weak and Correlated Signals." in 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), pp. 128–132, IEEE.
- Yakir, B. 2013. *Extremes in Random Fields: A Theory and Its Applications*, New Jersey: Wiley.