

Fast Computing For Dynamic Screening Systems When Analyzing Correlated Data

Lu You and Peihua Qiu

Department of Biostatistics, University of Florida

2004 Mowry Road, Gainesville, FL 32610

Abstract

In practice, we often need to identify individuals whose longitudinal behavior is different from the behavior of those well-functioning individuals, so that some unpleasant consequences (e.g., stroke) can be avoided or early detected. To properly handle such applications, a new statistical method, called dynamic screening system, has been developed in the literature. A recent version of this method can analyze correlated data, using a sequential data de-correlation procedure. However, the computation involved is intensive, which may restrict the applicability of the method in certain cases. In this paper, we suggest a fast computing algorithm for the dynamic screening system. The algorithm only de-correlates a small portion of the history data that are collected after the previous time that the related CUSUM chart re-starts its charting statistic, and avoids unnecessary data de-correlation for the majority part of the history data. Thus, its computation is much faster than the conventional dynamic screening system. Furthermore, the new algorithm improves the effectiveness of the conventional dynamic screening system in certain cases, because it partially eliminates attenuation of a process mean shift that occurs with the conventional dynamic screening system due to data de-correlation. Numerical results show that the new algorithm works well in different cases.

Key Words: Computing time; Correlation; Data de-correlation; Sequential monitoring; Sprint length; Statistical process control.

1 Introduction

In the SHARe Framingham heart study of the National Heart, Lung and Blood Institute, one major task is to monitor health variables (e.g., blood pressure, cholesterol level) so that their irregular longitudinal pattern can be detected as soon as possible for timely medical treatments to avoid deadly cardiovascular diseases (e.g., stroke). This is the so-called dynamic screening (DS) problem in the literature (cf., Qiu and Xiang 2014). The DS problem has broad applications. Besides the

disease screening applications mentioned above, it has broad applications in industries when we are interested in monitoring the quality/performance of durable products or processes (e.g., airplanes, cars, health care systems) to make sure that they work stably and satisfactorily. This paper aims to make a contribution to solve the DS problem effectively.

The DS problem is related to statistical process control (SPC) because it mainly concerns about sequential process monitoring. However, conventional SPC charts cannot solve the DS problem properly because they usually require the process distribution to be longitudinally unchanged when the related process is in-control (IC) (cf., Hawkins and Olwell 1998, Page 1954, Qiu 2014, Roberts 1959, Shewhart 1931), which is often invalid in the DS problem. In the literature, there are several recent methods developed for solving the DS problem. Qiu and Xiang (2014) proposed the so-called dynamic screening system (DySS) in univariate cases for identifying individuals with irregular longitudinal patterns. The DySS method first estimates the regular longitudinal pattern of the performance variable using nonparametric longitudinal data modeling from an IC dataset in which all individuals are well-functioning. Then, for a given individual to monitor, her/his observations are first standardized using the estimated regular longitudinal pattern, and then the standardized observations are monitored by the regular cumulative sum (CUSUM) control chart well discussed in the SPC literature (Hawkins and Olwell 1998, Qiu 2014). A semiparametric version of the DySS method in Qiu and Xiang (2014) was suggested in Li et al. (2018), an improved version that accommodates unequally spaced observation times was discussed in Qiu et al. (2018), and a multivariate version was proposed in Qiu and Xiang (2015). The DySS methods discussed in these papers focus mainly on cases when the sequential process observations are independent of each other. In practice, the process observations are often temporally correlated. To handle correlated data effectively, Li and Qiu (2016) suggested a sequential data de-correlation procedure in univariate cases. This procedure de-correlates all history data based on recursive covariance matrix decompositions. In cases when the monitored process is stable for a quite long time period, the computation involved in this data de-correlation procedure could be extensive because it needs to compute an inverse matrix of high-dimensions at each time point and the dimensionality of the inverse matrix gets increased when time goes by. In order to save all necessary information about the history data, a large data storage space is also required. Furthermore, the data de-correlation procedure would attenuate a process mean shift as a price to pay for obtaining a sequence of uncorrelated observations. A multivariate version of this data de-correlation procedure was discussed

in Li and Qiu (2017), and it shared all limitations mentioned above.

In this paper, we suggest a novel algorithm for the DySS method to handle temporally correlated data. The main idea is that instead of de-correlating all history data, we only de-correlate a small portion of the history data observed after the previous time that the related CUSUM chart re-starts its charting statistic. Therefore, the unnecessary de-correlation for the majority part of the history data is avoided by this algorithm. Because the time period from the last time point that the related CUSUM chart re-starts its charting statistic to the current time point is usually short, the computation involved in the new method is fast and the required data storage space is small. Also, the new method improves the effectiveness of the conventional DySS method by partially eliminating the attenuation of a process mean shift due to data de-correlation. Numerical results show that the new method works well in practice.

The remaining part of the article is organized as follows. Our proposed new data de-correlation method is described in detail in Section 2. Its numerical performance is evaluated in Section 3, in comparison with the conventional DySS method for handling correlated data. Its application to a dataset from the SHARe Framingham heart study is presented in Section 4. Several remarks conclude the article in Section 5.

2 Proposed Methodology

We describe the proposed method in detail in this section. Because our proposed method is for fast computing in data de-correlation when we apply the DySS method to correlated data, the DySS method and the conventional data de-correlation algorithm that were first discussed in Qiu and Xiang (2014) and Li and Qiu (2016) will be briefly introduced first. The description of the proposed method is for univariate cases only in this paper. Multivariate cases can be discussed similarly.

2.1 DySS method for analyzing correlated data

Assume that y is a major performance variable of a longitudinal process, and we want to early detect any unpleasant event (e.g., stroke, airplane crash) in the process by sequentially monitoring the longitudinal pattern of y . To achieve this goal using the DySS method, we implement the following three steps:

- (i) Estimate the regular longitudinal pattern of y from an observed longitudinal dataset of a group of m well-functioning individuals. This dataset is called *IC dataset* hereafter.
- (ii) For a new individual to monitoring, standardize his/her observations using the estimated regular longitudinal pattern obtained in step (i).
- (iii) Monitor the standardized observations of the new individual, and give a signal as soon as all available data suggest a significant shift in his/her longitudinal pattern from the estimated regular pattern.

In step (i) of the DySS method, we assume that the IC data follow the model

$$y(t_{ij}) = \mu(t_{ij}) + \varepsilon(t_{ij}), \text{ for } j = 1, \dots, n_i \text{ and } i = 1, \dots, m, \quad (1)$$

where n_i is the total number of observations for the i th individual, $t_{ij} \in [0, T]$ is the j th observation time of the i th individual, $\mu(\cdot)$ is the mean function, and $\varepsilon(\cdot)$ is the error term with the covariance function $V(s, t) = \text{Cov}(\varepsilon(s), \varepsilon(t))$, for $s, t \in [0, T]$. Based on the local polynomial kernel smoothing procedures originally suggested in Chen et al. (2005) and Li (2011), Qiu and Xiang (2014) proposed a four-step procedure for estimating $\mu(t)$ and $V(s, t)$, which is briefly described below.

First, we obtain an initial estimate of $\mu(t)$ by the regular local p th-order polynomial kernel smoothing procedure (cf., Qiu 2005, Chapter 2):

$$\tilde{\mu}(t) = \mathbf{e}'_1 \left(\sum_{i=1}^m X'_i K_{ih}(t) X_i \right)^{-1} \left(\sum_{i=1}^m X'_i K_{ih}(t) \mathbf{y}_i \right),$$

where \mathbf{e}_1 is a $(p+1)$ -dimensional vector with the first element being 1 and the other elements being 0,

$$X_i = \begin{pmatrix} 1 & (t_{i1} - t) & \cdots & (t_{i1} - t)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (t_{in_i} - t) & \cdots & (t_{in_i} - t)^p \end{pmatrix},$$

$$K_{ih}(t) = \text{diag} \{ K((t_{i1} - t)/h)/h, \dots, K((t_{in_i} - t)/h)/h \},$$

$K(\cdot)$ is a kernel function, and $h > 0$ is a bandwidth. Second, we define residuals

$$\tilde{\varepsilon}(t_{ij}) = y(t_{ij}) - \tilde{\mu}(t_{ij}), \text{ for } j = 1, \dots, n_i \text{ and } i = 1, \dots, m,$$

and when $s \neq t$, estimate $V(s, t)$ by

$$\widehat{V}(s, t) = [A_1(s, t)V_{00}(s, t) - A_2(s, t)V_{10}(s, t) - A_3(s, t)V_{01}(s, t)] / B(s, t),$$

where $A_1(s, t) = S_{20}(s, t)S_{02}(s, t) - S_{11}^2(s, t)$, $A_2(s, t) = S_{10}(s, t)S_{02}(s, t) - S_{01}(s, t)S_{11}(s, t)$, $A_3(s, t) = S_{01}(s, t)S_{20}(s, t) - S_{10}(s, t)S_{11}(s, t)$, $B(s, t) = A_1(s, t)S_{00}(s, t) - A_2(s, t)S_{10}(s, t) - A_3(s, t)S_{01}(s, t)$,

$$S_{l_1 l_2}(s, t) = \frac{1}{Nh^2} \sum_{i=1}^m \sum_{j_1=1}^{n_i} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{n_i} \left(\frac{t_{ij_1} - s}{h}\right)^{l_1} \left(\frac{t_{ij_2} - t}{h}\right)^{l_2} K\left(\frac{t_{ij_1} - s}{h}\right) K\left(\frac{t_{ij_2} - t}{h}\right),$$

$$V_{l_1 l_2}(s, t) = \frac{1}{Nh^2} \sum_{i=1}^m \sum_{j_1=1}^{n_i} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{n_i} \tilde{\epsilon}(t_{ij_1}) \tilde{\epsilon}(t_{ij_2}) \left(\frac{t_{ij_1} - s}{h}\right)^{l_1} \left(\frac{t_{ij_2} - t}{h}\right)^{l_2} K\left(\frac{t_{ij_1} - s}{h}\right) K\left(\frac{t_{ij_2} - t}{h}\right),$$

and $N = \sum_{i=1}^m n_i(n_i - 1)$. Third, estimate $V(t, t)$ by

$$\widehat{V}(t, t) = \mathbf{e}'_1 \left(\sum_{i=1}^m X_i' W_i X_i \right)^{-1} \left(\sum_{i=1}^m X_i' W_i \tilde{\epsilon}_i^2 \right), \quad (2)$$

where $\tilde{\epsilon}_i^2 = (\tilde{\epsilon}^2(t_{i1}), \dots, \tilde{\epsilon}^2(t_{in_i}))'$, $W_i = K_{ih}^{1/2}(t) [J_i(t) \Sigma_i J_i(t)]^{-1} K_{ih}^{1/2}(t)$, $J_i(t) = \text{diag}\{I_{\{|t_{i1} - t| \leq h\}}, \dots, I_{\{|t_{in_i} - t| \leq h\}}\}$, and $I_{\{\cdot\}}$ is an indicator function that equals 1 when the event is “true” and 0 otherwise. Fourth, $\mu(t)$ is estimated by

$$\widehat{\mu}(t) = \mathbf{e}'_1 \left(\sum_{i=1}^m X_i' \widehat{W}_i X_i \right)^{-1} \left(\sum_{i=1}^m X_i' \widehat{W}_i \mathbf{y}_i \right), \quad (3)$$

where $\mathbf{y}_i = (y(t_{i1}), \dots, y(t_{in_i}))'$, $\widehat{W}_i = K_{ih}^{1/2}(t) [J_i(t) \widehat{\Sigma}_i J_i(t)]^{-1} K_{ih}^{1/2}(t)$, and

$$\widehat{\Sigma}_i = \begin{pmatrix} \widehat{V}(t_{i1}, t_{i1}) & \cdots & \widehat{V}(t_{i1}, t_{in_i}) \\ \vdots & \ddots & \vdots \\ \widehat{V}(t_{in_i}, t_{i1}) & \cdots & \widehat{V}(t_{in_i}, t_{in_i}) \end{pmatrix}.$$

In step (ii) of the DySS method, assume that observations of a new individual to monitor are sequentially collected at time points $t_1^*, t_2^*, \dots \in [0, T]$. To monitor his/her performance, we first standardize the observations $\{y(t_1^*), y(t_2^*), \dots\}$ by

$$\tilde{\epsilon}(t_j^*) = \frac{y(t_j^*) - \widehat{\mu}(t_j^*)}{\widehat{V}(t_j^*, t_j^*)}, \quad \text{for } j = 1, 2, \dots, \quad (4)$$

where $\widehat{\mu}(t)$ and $\widehat{V}(t)$ are obtained in (3) and (2) in step (i). By (4), we compare the new individual with the well-functioning individuals in the IC data cross-sectionally at the time points $\{t_j^*, j = 1, 2, \dots\}$.

In step (iii) of the DySS method, we try to monitor the standardized data $\{\tilde{\epsilon}(t_1^*), \tilde{\epsilon}(t_2^*), \dots\}$ using a conventional control chart. But, the standardized data are usually temporally correlated. To de-correlate them, Li and Qiu (2016) suggested using the following procedure based on the

Cholesky decomposition of the covariance matrix. First, we define $\widehat{\epsilon}(t_j^*) = y(t_j^*) - \widehat{\mu}(t_j^*)$, for each $j \geq 1$. Then, at t_1^* , define the standardized value $e^*(t_1^*) = \frac{\widehat{\epsilon}(t_1^*)}{\widehat{V}(t_1^*, t_1^*)}$. Now, at the current time point $j > 1$, let $\epsilon_j = (\epsilon'_{j-1}, \widehat{\epsilon}(t_j^*))'$, and the covariance matrix of ϵ_j is written as

$$\Sigma_{j,j} = \begin{pmatrix} \Sigma_{j-1,j-1} & \sigma_{j-1,j} \\ \sigma'_{j-1,j} & \sigma_{j,j} \end{pmatrix}.$$

Let the Cholesky decomposition of $\Sigma_{j,j}$ be

$$\Phi_j \Sigma_{j,j} \Phi_j' = D_j^2,$$

where Φ_j is a lower-triangle matrix with all diagonal elements being 1, and $D_j^2 = \text{diag}\{d_1^2, \dots, d_j^2\}$.

Then, we define

$$e^*(t_j^*) = \frac{\widehat{\epsilon}(t_j^*) - \sigma'_{j-1,j} \Sigma_{j-1,j-1}^{-1} \epsilon_{j-1}}{d_j}, \text{ for } j > 1. \quad (5)$$

In practice, $\Sigma_{j,j}$ can be estimated by the matrix $(\widehat{V}(t_{j_1}^*, t_{j_2}^*), 1 \leq j_1, j_2 \leq j)$. So, $e^*(t_j^*)$ can be calculated accordingly after $\Sigma_{j,j}$ is replaced by its estimator. In (5), Li and Qiu (2016) showed that $\Sigma_{j-1,j-1}^{-1}$ can be computed recursively as follows.

$$\Sigma_{j-1,j-1}^{-1} = \begin{pmatrix} \Sigma_{j-2,j-2}^{-1} + \Sigma_{j-2,j-2}^{-1} \sigma_{j-2,j-1} d_{j-1}^{-2} \sigma'_{j-2,j-1} \Sigma_{j-2,j-2}^{-1}, & -\Sigma_{j-2,j-2}^{-1} \sigma_{j-2,j-1} d_{j-1}^{-2} \\ -d_{j-1}^{-2} \sigma'_{j-2,j-1} \Sigma_{j-2,j-2}^{-1}, & d_{j-1}^{-2} \end{pmatrix}. \quad (6)$$

Then, the de-correlated observations are $\{e^*(t_1^*), e^*(t_2^*), \dots\}$, and they are uncorrelated with mean 0 and variance 1. To detect an upward mean shift in y , we can use the following CUSUM:

$$\begin{cases} C_0^+ = 0 \\ C_j^+ = \max(0, C_{j-1}^+ + e^*(t_j^*) - k), \text{ for } j \geq 1, \end{cases} \quad (7)$$

where $k > 0$ is an allowance constant, and the chart gives a signal when

$$C_j^+ > h, \quad (8)$$

where $h > 0$ is a control limit.

In many DS problems, the observation times $\{t_1^*, t_2^*, \dots\}$ are not equally spaced. See a real-data example about the SHARe Framingham heart study in Section 4. In such cases, to evaluate the performance of the CUSUM chart (7)-(8), we can use the IC average time to signal, denoted as ATS_0 , which is the average time from the beginning of process monitoring to the first signal when the process is IC, and the out-of-control (OC) average time to signal, denoted as ATS_1 , which is

the average time from a process shift to a signal. Usually, ATS_0 is pre-determined at a given level, and the chart is better for detecting a specific shift if its ATS_1 is smaller. In the CUSUM chart (7)-(8), the allowance constant k is often pre-determined, and the control limit h is chosen such that the pre-determined ATS_0 value is reached. See Hawkins and Olwell (1998) and Qiu (2014) for more detailed discussions about the design of a control chart.

2.2 Proposed methods

This paper tries to make two contributions, both of which aim to speed up the computation of the DySS method described in the previous subsection. First, we notice that computation of the data de-correlation procedure (5) through the Cholesky decomposition of $\Sigma_{j,j}$ and the recursive computation of $\Sigma_{j-1,j-1}^{-1}$ by (6) can be computationally simplified by the following idea. Instead of recursively computing $\Sigma_{j-1,j-1}^{-1}$, we can derive a recursive formula to compute $U_{j-1} = D_{j-1}^{-1}\Phi_{j-1}$. Then, $U_{j-1}\Sigma_{j-1,j-1}U_{j-1}' = I_{(j-1)\times(j-1)}$, and the de-correlated observations can be defined as

$$e^*(t_j^*) = \frac{\widehat{\epsilon}(t_j^*) - \boldsymbol{\sigma}'_{j-1,j}U_{j-1}'\mathbf{e}_{j-1}^*}{d_j}, \quad (9)$$

where $\mathbf{e}_{j-1}^* = (e^*(t_1^*), \dots, e^*(t_{j-1}^*))'$. We notice that computation in (9) only involves the multiplication of a matrix and a vector twice, plus some vector operations. So, its computational complexity is $2j^2 + O(j)$. If we take advantage of the fact that U_{j-1} is a lower triangular matrix, then, its computational complexity is actually $j^2 + O(j)$. As a comparison, the computational complexity of the original data de-correlation using (5) and (6) could be $4j^2 + O(j)$. So, much computation is saved by using the new data de-correlation algorithm (9).

In the original data de-correlation procedure (5)-(6), at each time point t_j^* , we try to transform the observation at that time point to be uncorrelated with observations obtained at *all* previous time points. However, a conventional CUSUM chart (e.g., (7)-(8)) has the re-starting mechanism. That is, its charting statistic is reset to be 0 at each time when the chart does not find any evidence of process distributional shift from the history data. So, the CUSUM chart tries to use all history data in process monitoring to improve its efficiency, and tries to get rid of all history data when it finds no evidence of shift from the history data to improve its sensitivity to a future shift. This re-starting mechanism makes the CUSUM chart have some optimal properties (e.g., Moustakides 1986, Ritov 1990). Based on this observation, Chatterjee and Qiu (2009) suggested the concept of

sprint length T_j as follows:

$$T_j = \begin{cases} 0 & \text{if } C_j^+ = 0, \\ k & \text{if } C_j^+ \neq 0, \dots, C_{j-k+1}^+ \neq 0, C_{j-k}^+ = 0, \end{cases}$$

Namely, T_j is the number of observations from the last time when the CUSUM is reset to the current time j , and the CUSUM makes a decision whether it should give a signal of process distributional shift based on the data within the sprint length of the current time point only. *This connection suggests that for data de-correlation, we only need to de-correlate the current observation with observations within T_j of the current time point j .* This idea will greatly simplify the computation. Also, the data storage requirement can be reduced dramatically.

Based on the two ideas described above, we propose the following DySS procedure for handling correlated data. For $j \geq 1$,

- If $T_j = 0$, then define $C_j^+ = 0$.
- If $T_j = 1$, then define $e^*(t_j^*) = \widehat{\epsilon}(t_j^*)/\sqrt{\widehat{V}(t_j^*, t_j^*)}$ and $C_j^+ = \max(0, e^*(t_j^*) - k)$.
- If $T_j > 1$, then define

$$\begin{aligned} \Sigma_{j,j}^* &= \begin{pmatrix} \Sigma_{j-1,j-1}^* & \boldsymbol{\sigma}_{j-1,j}^* \\ (\boldsymbol{\sigma}_{j-1,j}^*)' & \widehat{V}(t_j^*, t_j^*) \end{pmatrix} = \begin{pmatrix} \widehat{V}(t_{j-T_j+1}^*, t_{j-T_j+1}^*), & \cdots, & \widehat{V}(t_{j-T_j+1}^*, t_j^*) \\ \vdots & \ddots & \vdots \\ \widehat{V}(t_j^*, t_{j-T_j+1}^*), & \cdots, & \widehat{V}(t_j^*, t_j^*) \end{pmatrix}, \\ \boldsymbol{l}_j^* &= U_{j-1}^* \boldsymbol{\sigma}_{j-1,j}^*, \\ d_j^* &= \sqrt{\widehat{V}(t_j^*, t_j^*) - (\boldsymbol{l}_j^*)' \boldsymbol{l}_j^*}, \\ U_j^* &= \begin{pmatrix} U_{j-1}^* & 0 \\ -(1/d_j^*)(\boldsymbol{l}_j^*)' U_{j-1}^* & 1/d_j^* \end{pmatrix} \\ e^*(t_j^*) &= \frac{\widehat{\epsilon}(t_j^*) - (\boldsymbol{l}_j^*)' \boldsymbol{e}_{j-1}^*}{d_j^*} \\ C_j^+ &= \max(0, C_{j-1}^+ + e^*(t_j^*) - k), \end{aligned}$$

where $U_{j-T_j}^* = 1/d_{j-T_j}^*$, and $\boldsymbol{e}_{j-1}^* = (e^*(t_{j-T_j+1}^*), e^*(t_{j-T_j+2}^*), \dots, e^*(t_{j-1}^*))'$.

- The CUSUM chart gives a signal when $C_j^+ > h$.

It should be pointed out that the standardized observations $\{e^*(t_1^*), e^*(t_2^*), \dots, e^*(t_j^*)\}$ are uncorrelated only within the sprint length T_j of the current time point j , and the whole sequence

might be correlated. So, the control limit estimated using independent and standard normally distributed random numbers cannot guarantee the pre-specified ATS_0 value. For this reason, we suggest choosing the control limit h by a block bootstrap procedure, as discussed in Lahiri (2003) and Qiu and Xiang (2014). By this procedure, we partition the IC dataset into two parts. The first part of m_1 individuals is used for estimating IC functions $\mu(t)$ and $\Sigma(s, t)$, and the second part of $m - m_1$ individuals is used for estimating the control limit h . Then, we repeatedly select B individuals with replacement from the second part of the data, and apply the proposed DySS method to the sampled data for selecting h . See Qiu and Xiang (2014) for details.

Next, we explain through examples that the conventional data de-correlation procedure can sometimes sacrifice some power of the DySS method to detect a process distributional shift, in cases when process observations are positively correlated. By using the proposed method, this limitation is partially eliminated.

Example 1. In this example, we assume that the covariance matrix of $(\epsilon(t_1^*), \dots, \epsilon(t_j^*))'$ is

$$\sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{j-1} \\ \rho & 1 & \rho & \dots & \rho^{j-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{j-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{j-1} & \rho^{j-2} & \rho^{j-3} & \dots & 1 \end{pmatrix}.$$

In such cases, the correlation between two observations exponentially decays when they move apart. By the conventional decomposition procedure (5)-(6), we get

$$e^*(t_1^*) = \frac{\epsilon(t_1^*)}{\sigma}$$

$$e^*(t_j^*) = \frac{\epsilon(t_j^*) - \rho\epsilon(t_{j-1}^*)}{\sigma\sqrt{1-\rho^2}}, \text{ for } j \geq 2.$$

So, if there is a constant mean shift of size $\sigma\delta$ at the beginning of the process, then it is easy to check that

$$e^*(t_1^*) = \frac{\epsilon(t_1^*)}{\sigma} + \delta$$

$$e^*(t_j^*) = \frac{\epsilon(t_j^*) - \rho\epsilon(t_{j-1}^*)}{\sigma\sqrt{1-\rho^2}} + \delta\sqrt{1 - \frac{2\rho}{1+\rho}}, \text{ for } j \geq 2.$$

So, we can see that when $j \geq 2$, the shift size is offset in the standardized data by the factor of $\sqrt{1 - \frac{2\rho}{1+\rho}}$ in cases when $\rho > 0$.

Example 2. In this example, we assume the covariance matrix of $(\epsilon(t_1^*), \epsilon(t_2^*), \dots, \epsilon(t_j^*))'$ is

$$\sigma^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix}.$$

In such cases, the correlation between any two observations is a constant. By the conventional decomposition procedure (5)-(6), we get

$$e^*(t_1^*) = \frac{\epsilon(t_1^*)}{\sigma},$$

$$e^*(t_j^*) = \frac{\epsilon(t_j^*) - k_j \sum_{p=1}^j \epsilon(t_p^*)}{\sqrt{(1-\rho)(1+k_j)}}, \text{ for } j \geq 2,$$

where $k_j = \frac{\rho}{\rho(j-1)+(1-\rho)}$. If there is a constant mean shift of size $\sigma\delta$ at the beginning of the process, then we have

$$e^*(t_1^*) = \frac{\epsilon(t_1^*)}{\sigma} + \delta$$

$$e^*(t_j^*) = \frac{\epsilon(t_j^*) - k_j \sum_{p=1}^j \epsilon(t_p^*)}{\sigma\sqrt{(1-\rho)(1+k_j)}} + \frac{1 - jk_j}{(1-\rho)(1+k_j)}\delta, \text{ for } j \geq 2.$$

It can be checked that

$$\frac{1 - jk_j}{(1-\rho)(1+k_j)} = O(j^{-1}). \quad (10)$$

So, in this case, only the first few standardized observations carry meaningful information about the shift.

Both covariance structures assumed in the above two examples are common in the literature. From the above analysis (cf., (10)), we can see that the information about a process shift in the standardized observations is greatly compromised when j is large (i.e., in cases when we de-correlate a long sequence of observations). By using the sprint length idea, we decorrelate the data within T_j time points of the current time point j only and T_j is usually small. Therefore, the above problem is greatly alleviated.

3 Simulation Study

In this section, we present some simulation results to evaluate the performance of the proposed method described in the previous section. To this end, the following three versions of the DySS method will be considered: (i) the conventional DySS method discussed in Section 2.1, labeled as Method 1, (ii) the conventional DySS method with our proposed new data de-correlation algorithm (9), labeled as Method 2, and (iii) the proposed DySS method using both (9) and the sprint length T_j that is discussed in Section 2.2, labeled as Method 3. The IC data are assumed to follow model (1) with $\mu(t)$ being $\sin(2\pi t)$, for $t \in [0, 1]$, and $\varepsilon(t_{ij})$ generated from the model

$$\varepsilon(t_{ij}) = \xi_{0,ij} + \xi_{1i}\phi_1(t_{ij}) + \xi_{2i}\phi_2(t_{ij}) + \xi_{3i}\phi_3(t_{ij}),$$

where $\xi_{0,ij}, \xi_{1i}, \xi_{2i}, \xi_{3i}$ are independent random numbers generated from $N(0, 0.3)$, and $\phi_1(t) = t^2 + 0.5$, $\phi_2(t) = \sin(3\pi t)$ and $\phi_3(t) = \cos(3\pi t)$. This same mixed-effect model was also considered in Li (2011) and Li and Qiu (2016).

In all simulation examples, the observation times are generated as follows. We assume that the design interval $[0, 1]$ is divided into $1/\omega$ basic time units with $\omega = 0.01$. Within each 10 basic time units, the sampling rate d determines the number of observations and d equals 2, 5, or 10. When $d = 2$, for example, we randomly choose 2 basic time units as observation times for a specific block of 10 basic time units. The nominal ATS_0 value is chosen to be 25 or 50, and the allowance constant k is chosen to be 0.1, 0.2, or 0.5. Table 1 gives the control limit values h for different d , k and nominal ATS_0 values for Methods 1 and 2. This table is copied from Li and Qiu (2016).

Table 1: Control limit values h for different d , k and ATS_0 values.

k	$ATS_0 = 25$			$ATS_0 = 50$		
	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	0.991	2.039	3.149	1.938	3.375	4.937
0.2	0.828	1.750	2.633	1.688	2.875	3.984
0.5	0.431	1.109	1.648	1.092	1.820	2.406

We assume that there is an IC dataset with 1,000 individuals. As discussed in Section 2.2, half of this data is used for estimating IC quantities $\mu(t)$ and $V(s, t)$, and the remaining half is for determining the control limits of Method 3. To be consistent with the original setup in Li and

Qiu (2016), the bandwidth used for estimating $\mu(t)$ and $V(s, t)$ were chosen $h = 0.1$ when $d = 2$, $h = 0.05$ when $d = 5$, and $h = 0.02$ when $d = 10$. First, we investigate the IC performance of the related methods. During each simulation, each method monitors 10,000 IC individuals, and the times to signal and the computing times are recorded. Then, the ATS_0 values and the total computing times of the three methods are computed. This entire process, from estimation of $\mu(t)$ and $V(s, t)$ to computation of ATS_0 values and the total computing times, is repeated 100 times, and the averaged ATS_0 values and the averaged total computing times are computed accordingly. The averaged ATS_0 values are presented in Table 2, and the averaged total computing times (in seconds) are presented in Figure 1. From Table 2, we can see that all three methods can reach the nominal ATS_0 values well in different scenarios. From Figure 1, we can see that (i) the proposed new data de-correlation method (9) can indeed reduce the computing time of the DySS method original discussed in Li and Qiu (2016), and (ii) the proposed DySS method based on both (9) and the sprint length can further reduce the computing time in large margins.

Table 2: Computed actual ATS_0 values of the three methods with their standard errors (in parentheses) for different d , k and nominal ATS_0 values.

d	k	Nominal $ATS_0=25$			Nominal $ATS_0=50$		
		Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
2	0.1	25.775(0.052)	25.775(0.052)	24.895(0.136)	50.092(0.081)	50.092(0.081)	50.021(0.167)
2	0.2	25.756(0.054)	25.756(0.054)	24.890(0.145)	51.062(0.080)	51.062(0.080)	49.879(0.192)
2	0.5	25.954(0.055)	25.954(0.055)	25.104(0.118)	52.173(0.081)	52.173(0.081)	49.776(0.190)
5	0.1	25.183(0.044)	25.183(0.044)	24.958(0.128)	49.187(0.076)	49.187(0.076)	49.938(0.185)
5	0.2	25.510(0.044)	25.510(0.044)	24.883(0.122)	50.620(0.077)	50.620(0.077)	50.019(0.185)
5	0.5	25.543(0.047)	25.543(0.047)	24.900(0.126)	51.302(0.075)	51.302(0.075)	49.983(0.194)
10	0.1	24.900(0.049)	24.900(0.049)	24.921(0.124)	49.420(0.082)	49.420(0.082)	49.811(0.176)
10	0.2	25.061(0.049)	25.061(0.049)	24.922(0.129)	49.436(0.080)	49.436(0.080)	49.860(0.182)
10	0.5	24.708(0.050)	24.708(0.050)	24.945(0.128)	49.390(0.074)	49.390(0.074)	49.789(0.173)

Next, we study the OC performance of the three methods. The setup of the simulation is the same as that in the IC study above, and we consider constant shifts only here for simplicity. Namely, the OC model is $y(t_j^*) = \mu(t_j^*) + \delta + \epsilon(t_j^*)$, where δ is chosen to be 0.25, 0.50, 0.75 or 1.00. The calculated ATS_1 values of the three methods are shown in Figure 2. From the plots of the figure, Methods 1 and 2 are identical all the times, as expected, because their difference is on the computation in data de-correlation only. However, Method 3 has a better performance in all

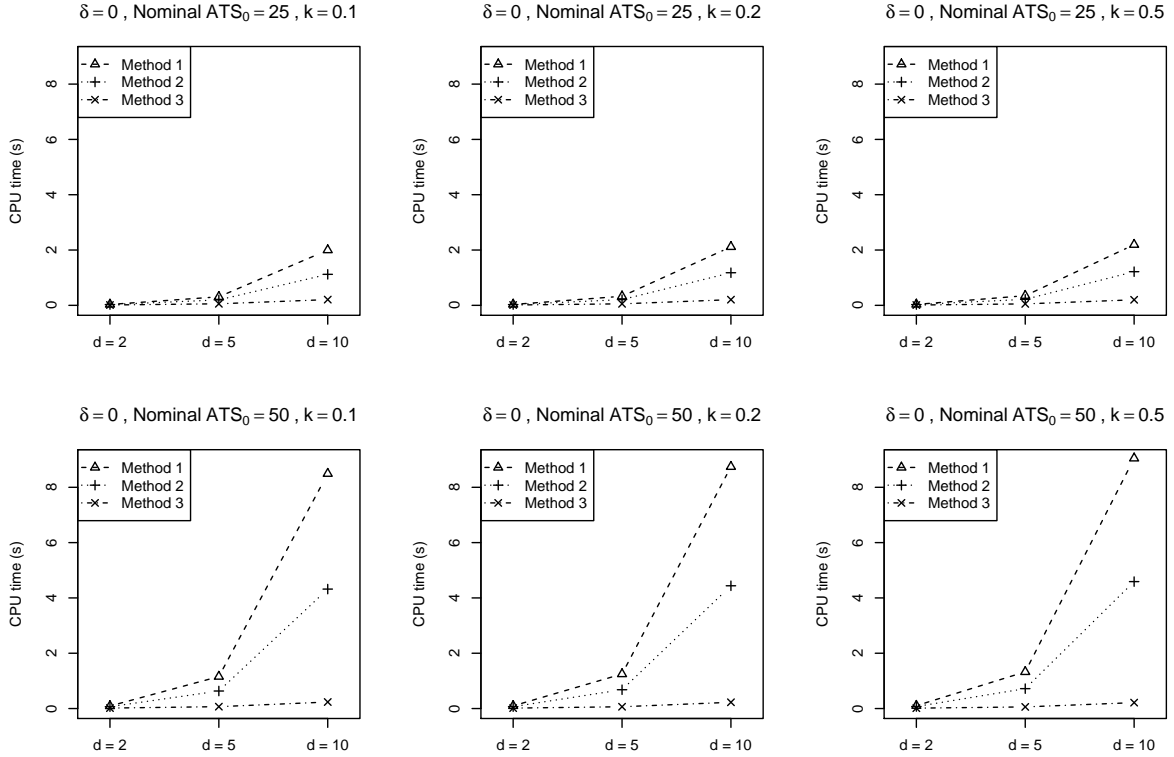


Figure 1: Computing times (in seconds) of three different methods for different d , k and nominal ATS_0 values when the process is IC.

cases considered. The reason behind this result has been explained at the end of Section 2 (cf., expression (10)). The averaged total computing times, calculated in the same way as that in the IC study, are shown in Figure 3. We can see that the proposed method decreases the computation dramatically in this case as well.

4 Application to the Stroke Data

In this section, we demonstrate the implementation of the three methods discussed in Section 3 on a dataset from the SHARe Framingham study mentioned at the beginning of the article. The dataset includes observations of the total cholesterol levels of 1028 non-stroke patients and 27 stroke patients. For each individual, 7 observations of total cholesterol levels were recorded at different ages, and the observation times are different for different individuals. We use the data of the 1028 non-stroke patients as the IC data. As in Li and Qiu (2016), we use the first 800 non-stroke patients

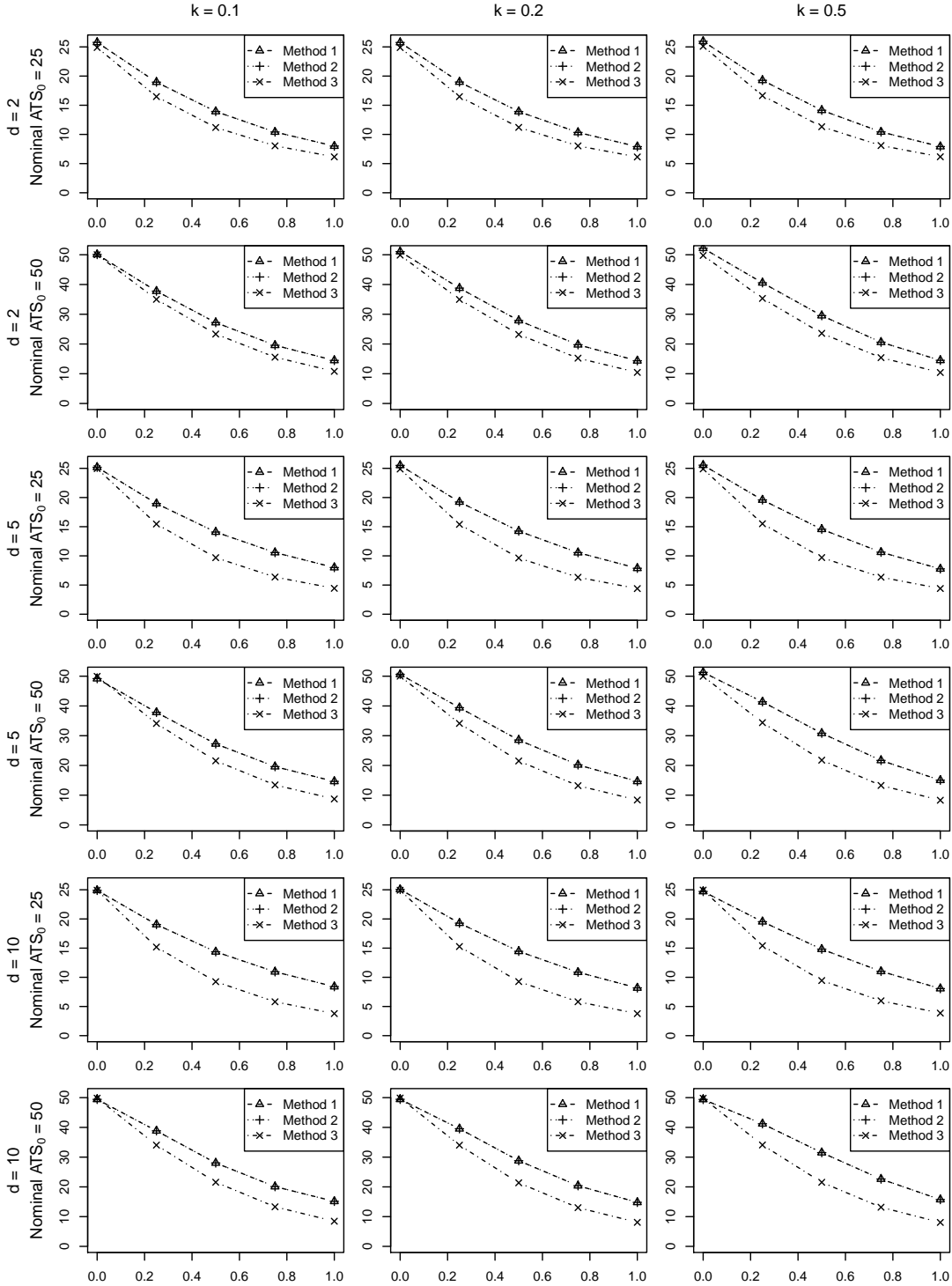


Figure 2: Computed ATS_1 values of three different methods for different d , k and nominal ATS_0 values. In each plot, x axis denotes the value of the shift size δ , and y axis denotes the computed ATS_1 value.

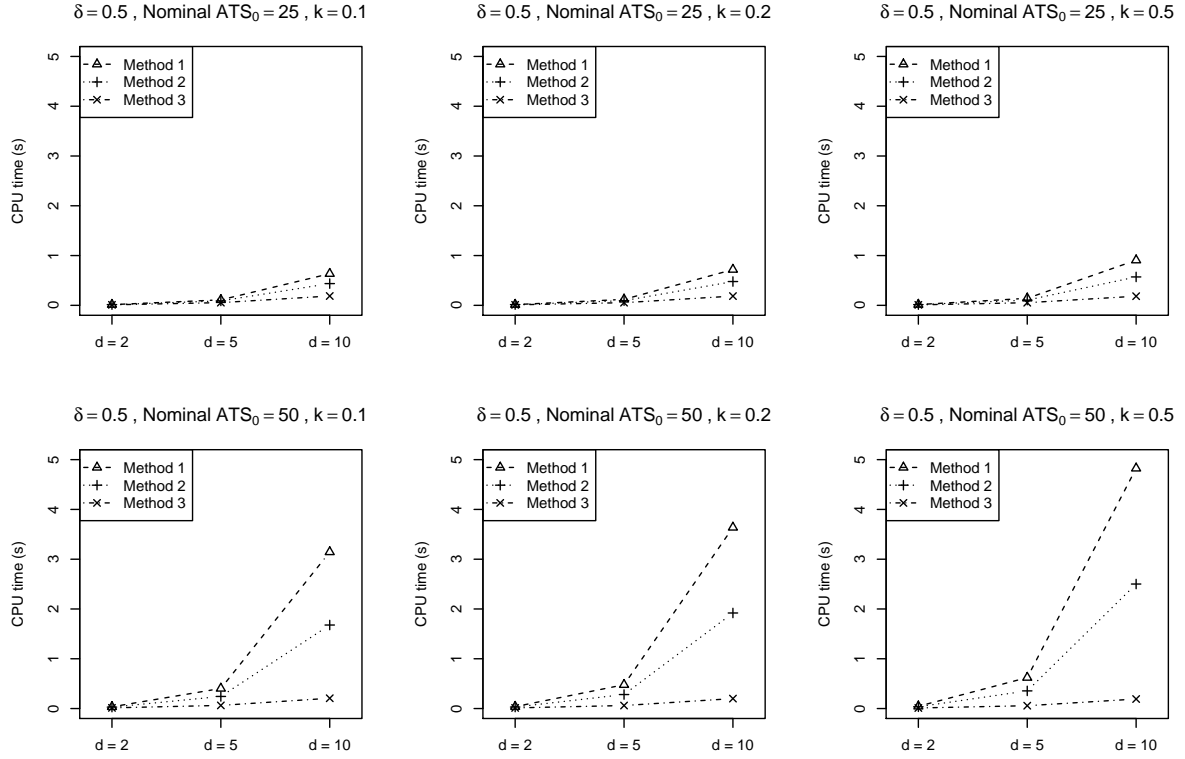


Figure 3: Computing times (in seconds) of three different methods for different d , k and nominal ATS_0 values when the process mean has a shift of size 0.5.

to estimate the IC mean function $\mu(t)$ and the IC covariance function $V(s, t)$ with a bandwidth of 0.15, the remaining 228 non-stroke patients for finding control limits by the bootstrap procedure. There are about 2.3 observations every 10 years in this dataset. So, we use $d = 2$, $k = 0.1$ and $ATS_0 = 25$ in this example.

Next, we use the three DySS methods to monitor the 27 stroke patients. Figure 4 presents their charting statistics C_j^+ for each stroke patients. Note that Methods 1 and 2 share the same charting statistics, which are shown in the upper panel of Figure 4, while the charting statistics of Method 3 are shown in the lower panel of Figure 4. From the figure, we can see that Methods 1 and 2 give signals for 22 stroke patients, while Method 3 gives signals for only 17 patients. For the five patients who receive signals from Methods 1 and 2 but not from Method 3, Figure 5 presents their original observations along with the estimated $\mu(t)$ in column 1, the original residuals in column 2, the de-correlated residuals and the corresponding charting statistics of Methods 1 and 2 in columns 3 and 4, and the de-correlated residuals and the corresponding charting statistic of

Method 3 in columns 5 and 6. Let us use patient #11 (2nd row in Figure 5) as an example to explain the results. Methods 1 and 2 both give signals to this patient whereas Method 3 does not. This patient’s observations are all below the estimated mean level (i.e., $\hat{\mu}(t)$). So, the algorithm of Method 3 restarts every observation time, and consequently no data de-correlation is implemented and no signal is given. However, for Methods 1 and 2, data de-correlation is implemented at each time point, and the de-correlated observation at a given time point is a linear combination of the residuals that are calculated at previous time points. When observations are positively correlated, previous residuals are negatively weighted in the de-correlation procedure. Thus, when the first few observations are smaller than the estimated mean level, the de-correlated residuals by Methods 1 and 2 at later times would tend to be large, resulting in signals by these two methods. Results about the other 4 patients can be explained similarly. From Figure 5, it seems that it is more reasonable not to give signals to the related 5 patients because their original observations are mostly below the estimated mean $\hat{\mu}(t)$.

The results about the stroke patients are related to the sensitivity of the three methods. To study their specificity, we apply the three methods to the 1028 non-stroke patients as well. The numbers of detected patients and the related computing times are presented in Table 3. From the table, we can notice that because there are only 7 observations for each patient in this data set, the computing efficiency of Method 3 is not fully exhibited. Also, despite the fact that Methods 1 and 2 detected more stroke patients, they also gave more signals to non-stroke patients in this example.

Table 3: The numbers of detected patients and the total computing times (in seconds) for three different methods in the stroke data example.

	Method 1		Method 2		Method 3	
	# Detected	CPU time	# Detected	CPU time	# Detected	CPU time
OC	22	0.03	22	0.03	17	0.02
IC	742	0.74	742	0.53	630	0.33

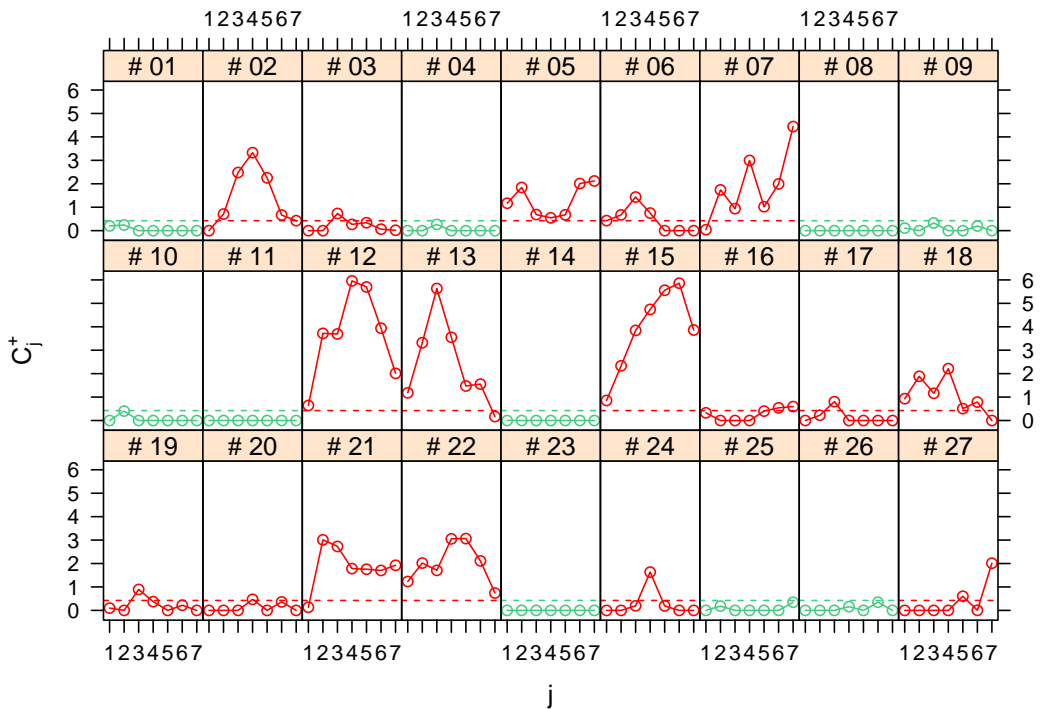
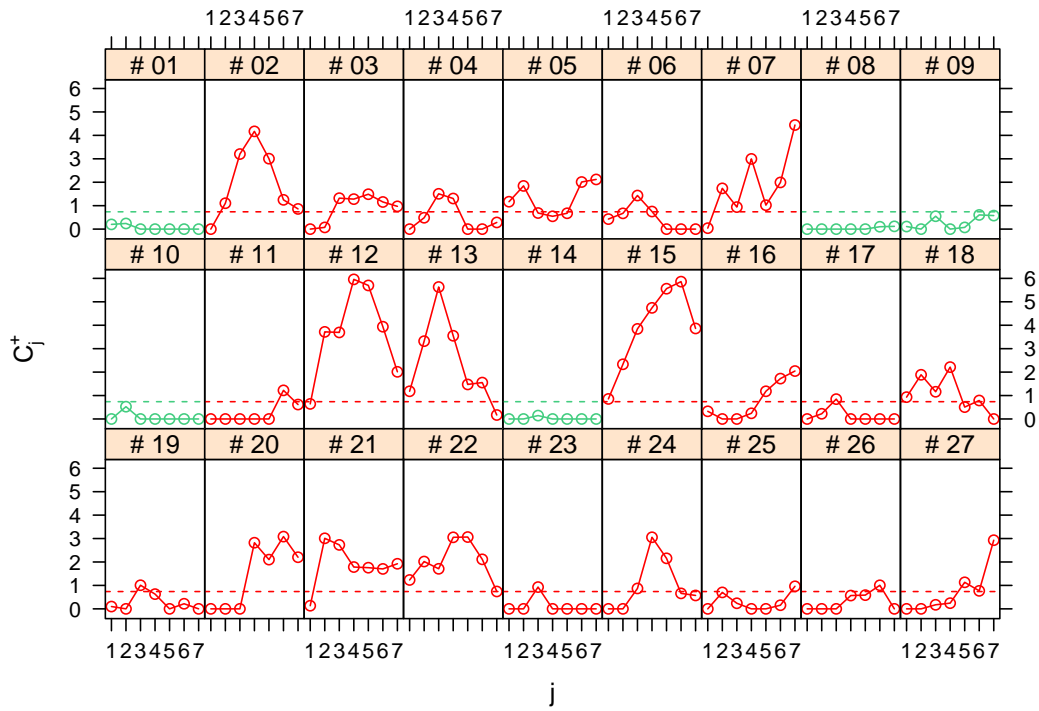


Figure 4: The upper panel displays the charting statistics of Methods 1 and 2 for all 27 stroke patients, while the lower panel displays the charting statistics of Method 3.

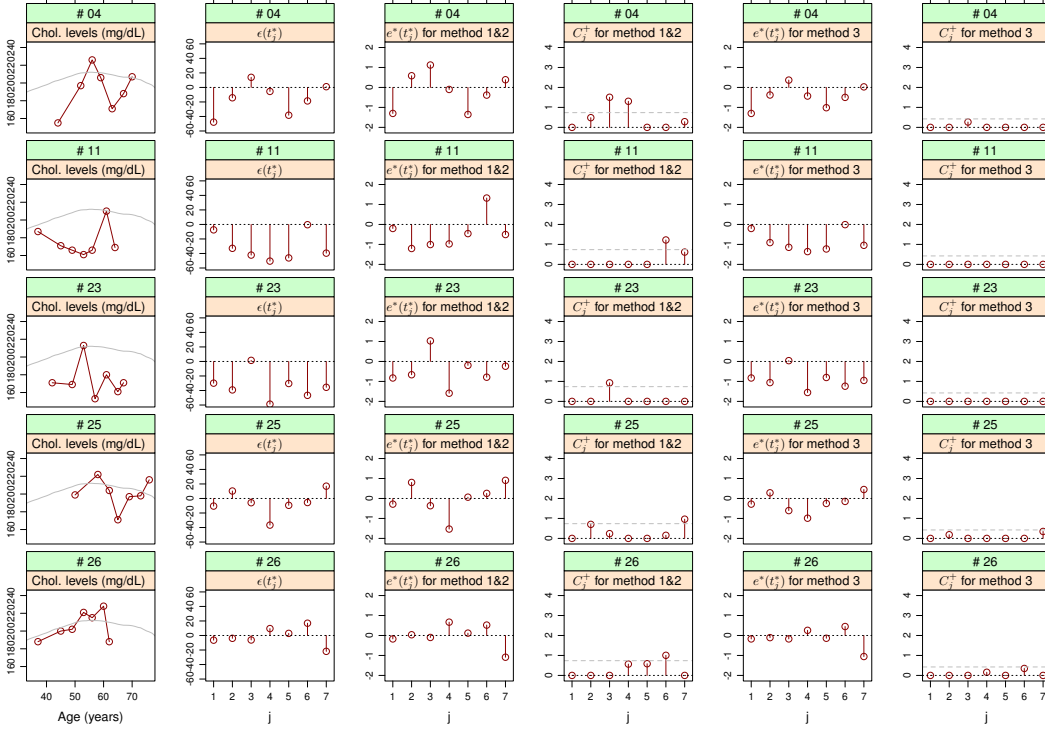


Figure 5: For the five patients who receive signals from Methods 1 and 2 but not from Method 3, column 1 presents their original observations (little circles) along with the estimated $\mu(t)$ (gray lines), column 2 shows the original residuals, columns 3 and 4 show the de-correlated residuals and the corresponding charting statistics of Methods 1 and 2, and columns 5 and 6 show the de-correlated residuals and the corresponding charting statistic of Method 3.

5 Concluding Remarks

As discussed in Section 1, the DS problem has broad applications in medicine, public health, engineering, and many other areas. The conventional DySS method for solving the DS problem has some limitations, including demanding computing and data storage requirements. In this paper, we have proposed two ideas to alleviate these shortcomings and make the DySS method more practical. One idea is related to the new recursive computing algorithm (9) for data de-correlation, and the other is related to the idea to de-correlate data within the sprint length of the current time point only. Numerical results show that the proposed methods do improve the computing time dramatically, and that they often improve the sensitivity of the DySS method to process distributional shifts as well. The suggested method here is discussed mainly for monitoring continuous numerical process data. It can also be used for monitoring correlated categorical data

(e.g., Ding et al. 2017, Xiang et al. 2018), which should be confirmed in the future research.

Acknowledgments: The authors thank a referee for a number of helpful comments and suggestions. This research is supported in part by an NSF grant.

References

- Chatterjee, S., and Qiu, P. (2009), “Distribution-free cumulative sum control charts using bootstrap-based control limits,” *Annals of Applied Statistics*, **3**, 349–369.
- Chen, K., and Jin, Z. (2005), “Local polynomial regression analysis of clustered data,” *Biometrika*, **92**, 59–74.
- Ding, D., Xiang, D., and Li, J. (2017), “Directional monitoring of categorical processes with serial dependence,” *Computers and Industrial Engineering*, **113**, 502–511.
- Hawkins, D.M., and Olwell, D.H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer-Verlag.
- Lahiri, S.N. (2003), *Resampling Methods for Dependent Data*, New York: Springer.
- Li, Y. (2011), “Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation,” *Biometrika*, **98**, 355–370.
- Li, J., and Qiu, P. (2016), “Nonparametric dynamic screening system for monitoring correlated longitudinal data,” *IIE Transactions*, **48**, 772–786.
- Li, J., and Qiu, P. (2017), “Construction of an efficient multivariate dynamic screening system,” *Quality and Reliability Engineering International*, **33**, 1969–1981.
- Li, W., Dou, W., Pu, X., and Xiang, D. (2018), “Monitoring individuals with irregular semiparametric longitudinal behaviour,” *Quality Technology & Quantitative Management*, **15**, 37–52.
- Moustakides, G.V. (1986), “Optimal stopping times for detecting changes in distributions,” *The Annals of Statistics*, **14**, 1379–1387.
- Page, E.S. (1954), “Continuous inspection scheme,” *Biometrika*, **41**, 100–115.
- Qiu, P. (2005), *Image Processing and Jump Regression Analysis*, New York: John Wiley & Sons.
- Qiu, P. (2014), *Introduction to Statistical Process Control*, Boca Raton, FL: Chapman Hall/CRC.
- Qiu, P., and Xiang, D. (2014), “Univariate dynamic screening system: an approach for identifying individuals with irregular longitudinal behavior,” *Technometrics*, **56**, 248–260.

- Qiu, P., and Xiang, D. (2015), “Surveillance of cardiovascular diseases using a multivariate dynamic screening system,” *Statistics in Medicine*, **34**, 2204–2221.
- Qiu, P., Zi, X., and Zou, C. (2018), “Nonparametric dynamic curve monitoring,” *Technometrics*, in press.
- Ritov, Y. (1990), “Decision theoretic optimality of the CUSUM procedure,” *The Annals of Statistics*, **18**, 1464–1469.
- Roberts, S.V. (1959), “Control chart tests based on geometric moving averages,” *Technometrics*, **1**, 239–250.
- Shewhart, W.A. (1931), *Economic Control of Quality of Manufactured Product*, New York: D. Van Nostrand Company.
- Xiang, D., Tsung, F., and Pu, X. (2017), “Statistical process control for latent quality characteristics using the up-and-down test,” *Technometrics*, **59**, 496–507.