Article:

Hinsheranan, S. and Stillman, E.C. (2021) The robustness of sufficient reduction methods for detecting shifts of various types in multivariate processes. Quality and Reliability Engineering International, 37 (5). pp. 2276-2287. ISSN 0748-8017

https://doi.org/10.1002/qre.2857

Title page

Article title: The Robustness of Sufficient Reduction Methods for Detecting Shifts of Various Types in Multivariate Processes

Short title: The Robustness of Sufficient Reduction Methods

Corresponding author:
Sawaporn Hinsheranan
Faculty of Science, Naresuan University, Phitsanulok, Thailand

Co-author:
Eleanor C. Stillman
School of Mathematics and Statistics, The University of Sheffield, Sheffield, UK

Email of corresponding author: sawaporns@nu.ac.th

Keywords: sufficient reduction, robustness, process shift, shift detection

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

The Robustness of Sufficient Reduction Methods for Detecting Shifts of Various Types in Multivariate Processes†

Sawaporn Hinsheranan*[1] | Eleanor C. Stillman[2]

[1]Faculty of Science, Naresuan University, Phitsanulok, Thailand
[2]School of Mathematics and Statistics, The University of Sheffield, Sheffield, UK

Correspondence
*Sawaporn Hinsheranan, Faculty of Science, Naresuan University, Phitsanulok, 65000, Thailand. Email: sawaporns@nu.ac.th

Summary

Applications of shift detection are of interest in several disciplines. Sufficient reduction (SR) methods have been developed for detecting a shift in a multivariate process under different conditions in several studies. However, all methods were proposed to detect only a constant, but persistent, mean shift. In practice, there might be other types of (mean) shift to be considered. Our purpose here is to investigate the robustness of SR methods for detecting different types of shift in a multivariate process. Four shift types are considered. The performances of the SR methods are compared against other statistical techniques used in multivariate process control. The evaluation was conducted via simulation by estimating four measures. The results show that in a process of independent observations the Wessman method performs well for detecting all sizes of single spike shift and small constant, linear and exponential shifts. In an autocorrelated process the Parallel, Frisén and Wessman methods produce a high number of false alarms. The Siripanthana and Stillman method gives shorter delays for detecting small shifts of all types, while the vector autoregressive chart gives a shorter delay for a large constant shift. The applications of SR methods to real health surveillance data are illustrated, with examples from food poisoning and pneumonia monitoring in Thailand.

KEYWORDS:
sufficient reduction, robustness, process shift, shift detection

1 | INTRODUCTION

Shift detection is of interest in several disciplines, such as industrial production, public health surveillance, business, finance and social network analysis.[1,2,3] Several collections of statistical techniques have been developed for detecting change in univariate and multivariate processes, such as statistical process control, time series techniques and statistical modelling techniques. However, these techniques were developed for specific purposes, and under different assumptions and evaluations[4], so they might perform well under the proposed conditions, but might not be applicable in other circumstances.[5,6] For example, parallel surveillance, where the multiple processes are monitored in parallel, is easy to implement, but it ignores the correlation between series and the multiplicity problem.[7,8,9] Statistical process control is widely used, however, it has the difficulty of signal interpretation, especially when using multivariate control charts.[10] Time series and statistical modelling techniques are

†The Robustness of Sufficient Reduction Methods.
[0]**Abbreviations:** SR, sufficient reduction.

more suitable for describing variation in the data and for retrospective assessment rather than for prospective surveillance.

Sufficient reduction (SR) is one statistical method proposed for detecting a shift in a multivariate process. The principle of SR is to reduce a p-dimensional multivariate series to a univariate sequence of statistics, which, having regard to the sufficiency property,[11] is proved to be sufficient to detect a shift in the process. Various SR methods have been proposed under different conditions. Wessman[12] and Frisén et al.[13] proposed methods to detect a mean shift in a multivariate process of independent observations. The former aims to detect a simultaneous change in all p constituent series and accounts for the correlation between series. The latter was proposed for detecting changes with time lag between series, but correlation between series is not allowed. Later, Siripanthana and Stillman[14] proposed an SR method for detecting a mean shift in a multivariate process of dependent observations, where both types of correlation (between and within series) are taken into account.

Frisén and de Maré[15] and Frisén[16] compared the performance of statistical surveillance methods for detecting a change in the mean level in a univariate process. A change divides the process into two parts: before change (or in control stage) and after change (or out of control stage). The comparison was conducted under the assumption that the data before and after change are independently and identically distributed. Frisén and de Maré[15] showed that a method based on the likelihood ratio of the probability density functions between in control and out of control stages is optimal for detecting a step change in the process in terms of minimal expected delay. This optimality is inherited by the SR method since this is also based on the likelihood ratio of the joint probability density functions of two stages.

With a single step change in mean, the process can be easily defined as being in one of two stages. In real life, there might be several kinds of shifts to consider, such as a single spike or gradual, linear and exponential shifts. However, these shifts result in a more complex situation than two distinct phases. So, the optimality of the SR methods might not hold and should be investigated. Whether or not strict optimality holds, since detecting various types of shift is of interest in several disciplines, the performance of SR relative to other methods might usefully be assessed. Thus, in this study, our purpose was to investigate the robustness of SR methods for detecting different types of shift in a multivariate process and to assess their detection performance against that of other statistical surveillance techniques. To evaluate performance, the change (or start of change) point of a process must be known in order to measure the delay in detection. However, in a real dataset, it is difficult to identify such a point unambiguously. Therefore, in this study detection performance was evaluated via simulation, where a change point and shift type can be pre-specified. More details of the SR methods and the system evaluation are given in the next sections following by the results from the simulation study, applications of SR methods to real data and a closing discussion.

2 | METHODS

2.1 | Sufficient reduction methods

Sufficient reduction methods use the principle of sufficiency[11] to reduce the dimension of a multivariate series to a univariate sequence of statistics summarizing all relevant information from the original series. The sequence, which can easily be monitored by a standard univariate control chart, is derived by factorizing the ratio of the joint probability density functions (pdfs) between stages according to the shift in parameter vector which we are attempting to detect. The derivation of the SR method can be found in[12,13,14].

The use of the SR method for monitoring shifts in multivariate processes was extended in several studies. Wessman[12] initially proposed SR to detect a shift in the parameter vector of a multivariate process under particular assumptions. The method aimed to detect only a simultaneous, sudden, but persistent, shift in the mean level of the multivariate series. Even though correlation between the series is allowed, observations in each series are assumed to be independent, and the variance of the process is assumed unchanged over time. The derived statistic at time t of Wessman is

$$T_t(\boldsymbol{x}_t) = \mathbf{c}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_t,\ t = 1, 2, ..., s \quad (1)$$

where $\boldsymbol{x}_t$ is the p-dimensional vector representing the observations made on a dimensional multivariate series at time t and s is a decision time point ($s = 1, 2, ...$). $\mathbf{c} = (c_1, ..., c_p)'$ is the vector of shift sizes, with c_i the shift we aim to detect in series i

($i,\ i = 1, ..., p$) and $\mathbf{\Sigma}$ is the covariance matrix of the process assumed unchanged over time.[12]

Later, Frisén et al.[13] extended the Wessman method to detect changes occurring with a time lag between series. Similar assumptions were made, except that correlation between series was not allowed. For the bivariate process, the derived statistic at time t of the Frisén method is

$$T_t(\boldsymbol{x}_t) = \begin{cases} c_1\ x_{1,t} + c_2\ x_{2,t+l} & t = 1, 2, ..., s - l \\ c_1\ x_{1,t} & t = s - l + 1, ..., s \end{cases} \tag{2}$$

where $x_{i,t}$ is an observation of a random variable X_i observed at time t, c_i is shift size to be detected in series i ($i = 1, 2$), and l is the time lag between series.[13]

In practice, many data are time dependent, with the current observation conditional on the previous one. Thus, to detect a shift in a process of dependent observations, correlation within series should be incorporated. Siripanthana and Stillman[14] proposed SR methods for detecting a mean shift in a multivariate process of autocorrelated data. Their method incorporated both correlation between series and correlation within series in the reduction. The correlation within series is investigated by considering the Gaussian autoregressive model of order 1 (AR(1) model). Simultaneous changes and changes with time lags were both considered. For the simultaneous changes, the derived statistic at time t is

$$T_t(\boldsymbol{x}_t) = \mathbf{c}'\mathbf{\Sigma}_\epsilon^{-1}(\boldsymbol{x}_t - \boldsymbol{\phi}\boldsymbol{x}_{t-1}),\ t = 1, 2, ..., s \tag{3}$$

where $\boldsymbol{\phi} = diag(\phi_i)$ is the diagonal matrix of autoregressive coefficients, with ϕ_i the coefficient for series i, $i = 1, 2, ..., p$. $\mathbf{\Sigma}_\epsilon = \mathbf{\Sigma} - \boldsymbol{\phi}\mathbf{\Sigma}\boldsymbol{\phi}'$, $\mathbf{\Sigma}$ is the covariance matrix of the data (now taken Gaussian AR(1)).[14]

For shift detection, the sequence of the derived statistics will be monitored by a standard univariate control chart. The detection performances of the SR methods were evaluated against those of other multivariate statistical methods such as principal component analysis,[12] parallel univariate control charts and multivariate control charts (Hotelling T^2 and multivariate exponentially weighted moving average control charts),[13] and VarR and Z charts.[14] The SR methods developed and tested in these simulations performed better by giving shorter delays and lower false alarms than other methods. This is because the reduction takes into account all relevant information on a shift in the mean vector of the original series, while the others have some limitations. For example, the principal component analysis loses some information since only the first few components are used. The parallel univariate control charts face the problem of multiplicity from multiple hypothesis testing and ignoring the correlation between the series, while the multivariate control chart has the difficulty of signal interpretation and threshold calculation.

2.2 | Types of shift

The detection performance depends on several aspects of the shift, such as size, duration and 'shape'. Obviously, larger shift sizes are more likely to be detected than smaller shift sizes; also, a long period of shift in the process increases the chance of detection. In this study, we considered four shapes of shift: single spike, constant, linear, and exponential shifts. Each was considered at three shift sizes, i.e. one, two and three times the standard deviation of the process, which was initially defined to be 1. To define precisely what we mean by these shapes and to make our different shifts comparable, we illustrate our approach in concrete terms. Suppose we are simulating daily data on cases of a disease, which, when in control, typically runs at a fixed mean value, and that the longest acceptable delay in detection would be a week, i.e. 7 days. We consider here the additional signal we will add to our basic process (which is generated separately, respecting a specified variance and correlation structure; see below) after the specified change point.

The single spike shift is a special case, which we do not seek to match in total effect, with an addition to the mean of 1, 2 or 3 occurring on day 1 only. A single-day spike is aimed at evaluating the detection of shift that exceeds the average level of the process within the 7-day observation period. It has been implemented in several simulation studies as it reflects the knowledge of a prior disease outbreak.[17,18,19] For the other shifts, which persist in the process after their introduction, we wish to match their total modification of the process over the 7-day observation period, T. From the constant shift case, it is easily seen that T is 7, 14 and 21, respectively for shift sizes 1, 2 and 3. To achieve the same total modification in a linearly increasing manner, the addition to the mean on day i is $a_i = T / \sum_{j=1}^{i} j$. For the exponential shift, the additional signal on day i is 2^{bi}, where b is chosen to satisfy $T = \sum_{j=1}^{7} 2^{bj}$.[20]

TABLE 1 The additional data for four shift types.

Type	Size	Day							Total
		1	2	3	4	5	6	7	
	1	1	-	-	-	-	-	-	1
Single spike	2	2	-	-	-	-	-	-	2
	3	3	-	-	-	-	-	-	3
	1	1	1	1	1	1	1	1	7
Constant	2	2	2	2	2	2	2	2	14
	3	3	3	3	3	3	3	3	21
	1	0.25	0.50	0.75	1.00	1.25	1.50	1.75	7
Linear	2	0.50	1.00	1.50	2.00	2.50	3.00	3.50	14
	3	0.75	1.50	2.25	3.00	3.75	4.50	5.25	21
	1	1.001	1.001	1.002	1.003	1.003	1.004	1.005	7
Exponential	2	1.174	1.379	1.620	1.903	2.235	2.624	3.082	14
	3	1.278	1.634	2.088	2.668	3.411	4.359	5.571	21

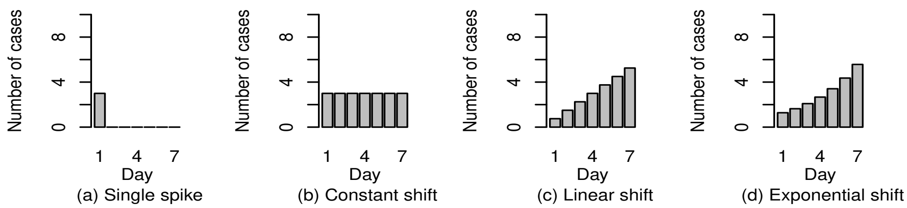

FIGURE 1 Four shapes of shift size 3.

Table 1 shows the number of additional data for each shift type of sizes 1-3. Figure 1 illustrates the four shapes of shift size 3 ($T = 21$) .

2.3 | Detection tool

By use of sufficient reduction, the dimension of the p multivariate series is reduced to just one. A univariate control chart is therefore a suitable means of detecting a process shift. Since we investigate several shift types including gradual shifts (e.g. linear and exponential shifts), we use an Exponentially Weighted Moving Average (EWMA) chart as a detection tool as it outperforms the Shewhart chart for detecting small shifts.[21] A further non-standard feature is that only positive shifts are of interest. Therefore, a one-sided EWMA chart for detecting increasing shifts was used.

The one-sided EWMA chart is designed to achieve an $\text{ARL}_0 = 370$ with a smoothing parameter $\lambda = 0.3$. Let z_t be the one-sided EWMA statistic at time t which can be calculated from

$$z_t = \max(\mu^I, \lambda T_t + (1 - \lambda) z_{t-1}),\ t > 0 \tag{4}$$

where μ^I is the mean of the in control stage and T_t is the statistic obtained from the sufficient reduction at time t. The one-sided EWMA statistics will be monitored with an upper control limit (UCL) calculated from the target mean and asymptotic variance of Z_t, defined as below.

$$UCL = \mu^I + L\sqrt{\sigma^2 \frac{\lambda}{2 - \lambda}} \tag{5}$$

where L is the chosen width of the band to achieve an $\text{ARL}_0 = 370$ and σ is the standard deviation of the in control process.[22,23,24] If the EWMA statistic exceeds the UCL, the alarm will be flagged and the process deemed out of control.

2.4 | Evaluation

The robustness of the SR methods was tested with 10,000 simulations. For simplicity, we investigated the detection performance in a bivariate process, where the data are generated from the bivariate normal distribution according to pre-specified process parameters (mean, variance, correlation between series (CBS) and correlation within series (CWS)) defined below.

$$\mathbf{X}_t \sim N_2\left(\begin{pmatrix}10\\5\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$$

Since the variance in each series equals one, the covariance between series is equal to correlation between series, ρ. Correlation within series is specified by a diagonal matrix of autoregressive coefficients from a bivariate AR(1) model. In this study we investigated the robustness of the SR methods in three different scenarios: Scenario 1 (CBS = 0.6, CWS = 0), Scenario 2 (CWS = 0.6, CBS = 0) and Scenario 3 (CWS = 0.6, CBS = 0.6). In each simulation run we generated 1,000 observations with the specified correlation structure. A change point of the process, denoted by τ, was randomly selected in the time interval (500,800). The four shift types of Section 2.2 were investigated, with an additional signal added to the original series from time τ.

To compare the detection performance of the SR methods against other statistical techniques, the original data with the added signals was also monitored with other univariate and multivariate statistical techniques including parallel univiariate control charts where the data in each series are monitored in parallel with a EWMA chart adjusted for multiplicity. The VarR (Vector autoregressive Residual) chart (i.e. monitoring the residuals from VarR(1) model with a Hotelling's T^2 chart),[25] and the Z chart (i.e. interpreting the signal resulting from a Hotelling's T^2 chart)[26] were also used. For sensible comparison, all charts are adjusted for detecting a positive shift and designed to meet the $\text{ARL}_0 = 370$ requirement.

Let t_A be a time of an alarm, then the detection performance of each method is evaluated using four measures. The conditional expected delay (CED) is the expected delay given that there is no false alarm before a change point τ. The false alarm rate (FAR) is the proportion of false alarms. As our objective was to detect a shift as soon as possible, the true alarm rate (TAR) is defined as the proportion of correct identification where a system gives an alarm during the testing period (i.e. within 7 time points after a true change occurs). On the other hand, the non-detection rate (NDR) is the proportion of cases in which the system fails to detect a shift within 7 time points since it has started.[13,2,14] These measures are defined as follows.

$$\text{CED} = E(t_A - \tau | t_A \geq \tau) \quad (6)$$

$$\text{FAR} = P(t_A < \tau) \quad (7)$$

$$\text{TAR} = P(t_A - \tau \leq 6 | t_A \geq \tau) \quad (8)$$

$$\text{NDR} = P(t_A - \tau \geq 7 | t_A \geq \tau) = 1 - \text{FAR} - \text{TAR} \quad (9)$$

Estimates of CED, FAR, TAR and NDR, denoted by $\widehat{ced}$, $\widehat{tar}$, $\widehat{far}$ and $\widehat{ndr}$, are taken from the 10,000 simulation runs. The results from the simulation study are summarized into three scenarios. The comparisons of detection performances between the methods are illustrated in bar charts, where the P, F, W, S, V and Z stand for Parallel, Frisén, Wessman, Siripanthana and Stillman methods, VarR and Z charts, respectively. Short $\widehat{ced}$ with high $\widehat{tar}$ and low $\widehat{far}$ and $\widehat{ndr}$ are desired.

3 | RESULTS

3.1 | Scenario 1: (CBS and no CWS)

In this scenario we investigated how the parallel method and SR methods, including the F and W methods, performed for detecting a shift in the bivariate process of independent observations (CBS = 0.6, CWS = 0); without CWS the other methods (S, V and Z) are not appropriate. The result is illustrated in Figure 2. As expected, the larger shift size is more likely to be detected than the smaller shift size. Overall, the W method performs better than other methods by giving shorter delays for

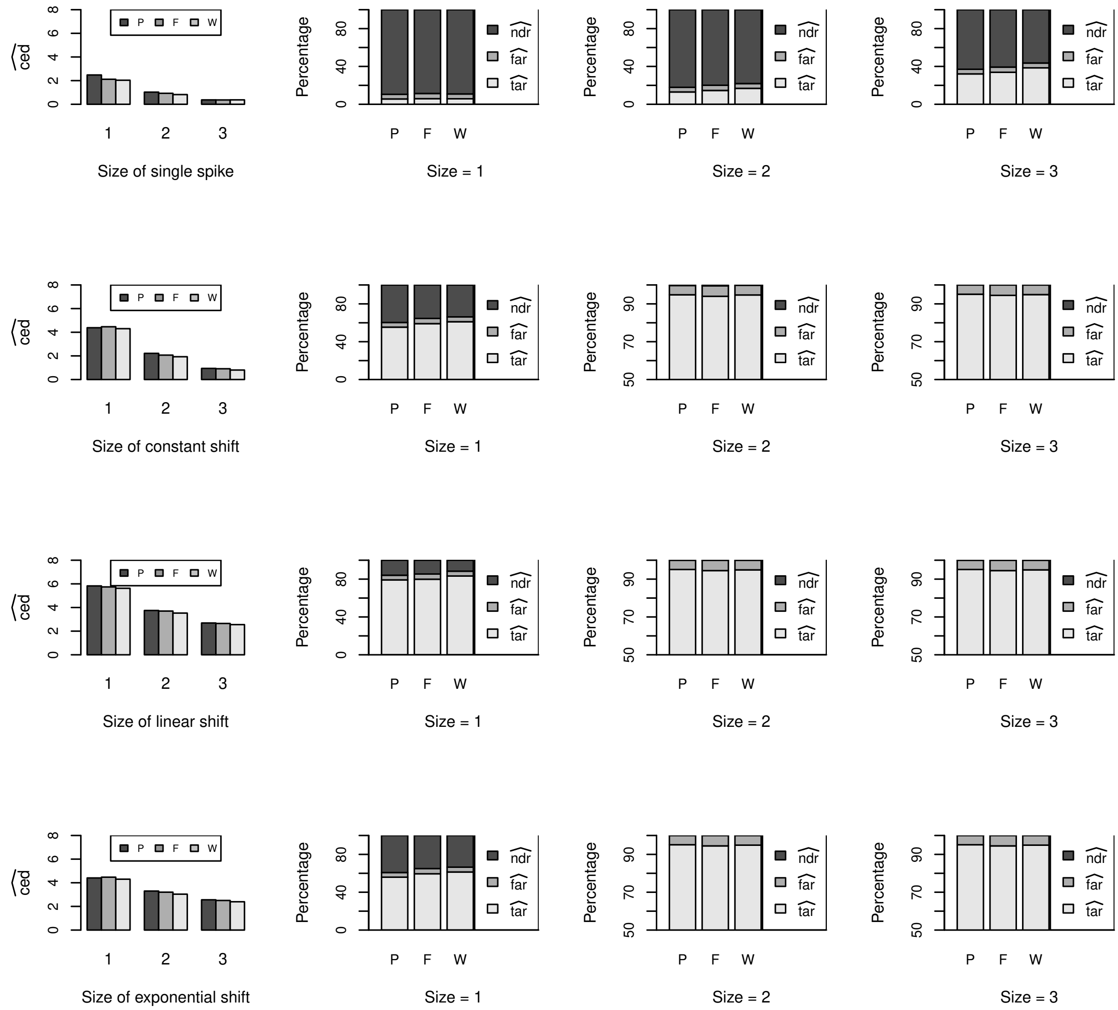

FIGURE 2 Bar charts comparing the detection performance in Scenario 1.

detecting all shift types and sizes.

Even though the single spike shift was detected rapidly due to the magnitude of the spike on the first day, the chance of being detected is rare resulting in high $\widehat{ndr}$s compared to the other shift types. However, the detection performance improved when shift sizes are large. The W method performed better than other methods by giving higher $\widehat{tar}$s and lower $\widehat{ndr}$s for detecting single spike shifts of size 2 or 3. The constant shift can be more quickly detected than linear and exponential shifts, by giving a lower $\widehat{ced}$s, while the linear shift is more likely to be detected correctly than other shift types by giving higher $\widehat{tar}$s and lower $\widehat{ndr}$s when shift size is small. However, the detection performances of the three methods are only slightly different when shift sizes are large.

3.2 | Scenario 2: (CWS and no CBS)

In the process of dependent observations (CWS = 0.6 and CBS = 0), all six methods were investigated. The result is shown in Figure 3. Unsurprisingly, the P, F and W methods produce high $\widehat{far}$s compared to the S, V and Z methods, since, while the S, V and Z methods take the CWS into account in the sufficient reduction, the P, F and W methods do not. Therefore, the effect of the CWS still remains in the derived sequence, and so monitoring such a sequence with a standard EWMA chart produces high

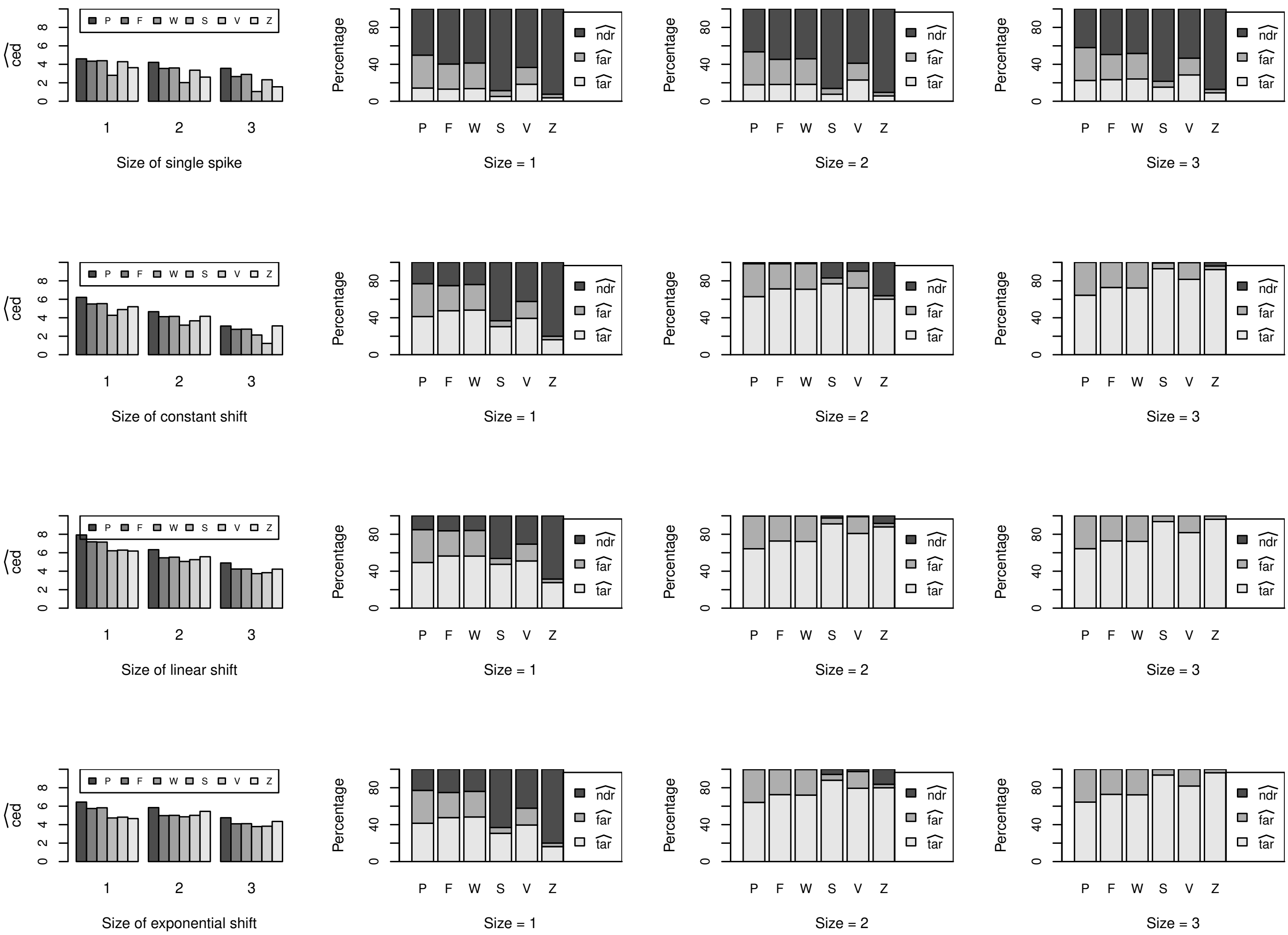

FIGURE 3 Bar charts comparing the detection performance in Scenario 2.

false alarms as the independence assumption of the chart is violated.

When CWS is present in the process, it is quite difficult to detect a single spike shift resulting in high $\widehat{ndr}$s. Even though the S and Z methods give lower $\widehat{ced}$s, their $\widehat{tar}$s are quite low compared to other methods. For the constant shift, the S method gives shorter delays for detecting shift sizes 1 and 2, while the V method detects the large shift size faster due to the use of the Hotelling T^2 chart, which performs well in detecting a large shift. All methods give lower $\widehat{ced}$s for detecting small exponential shift than small linear shifts. However, the small linear shift tends to be detected correctly more often than the small exponential shift resulting in high $\widehat{tar}$s. Again, S and Z methods give high $\widehat{tar}$s when shift sizes are large, especially the S method which typically has the lowest $\widehat{ced}$.

3.3 | Scenario 3: (CBS and CWS)

In this scenario we observed the detection performance when both CWS and CBS are present in the process (CWS = 0.6 and CBS = 0.6). The result is shown in Figure 4. Overall, the detection performances of six methods were quite similar to those in Scenario 2 due to the effect of the CWS. The P, F and W methods still produced high false alarms as the effect of the CWS remains. Also, all methods lose their ability to detect a single spike shift. For other shift types, the S and Z methods performed

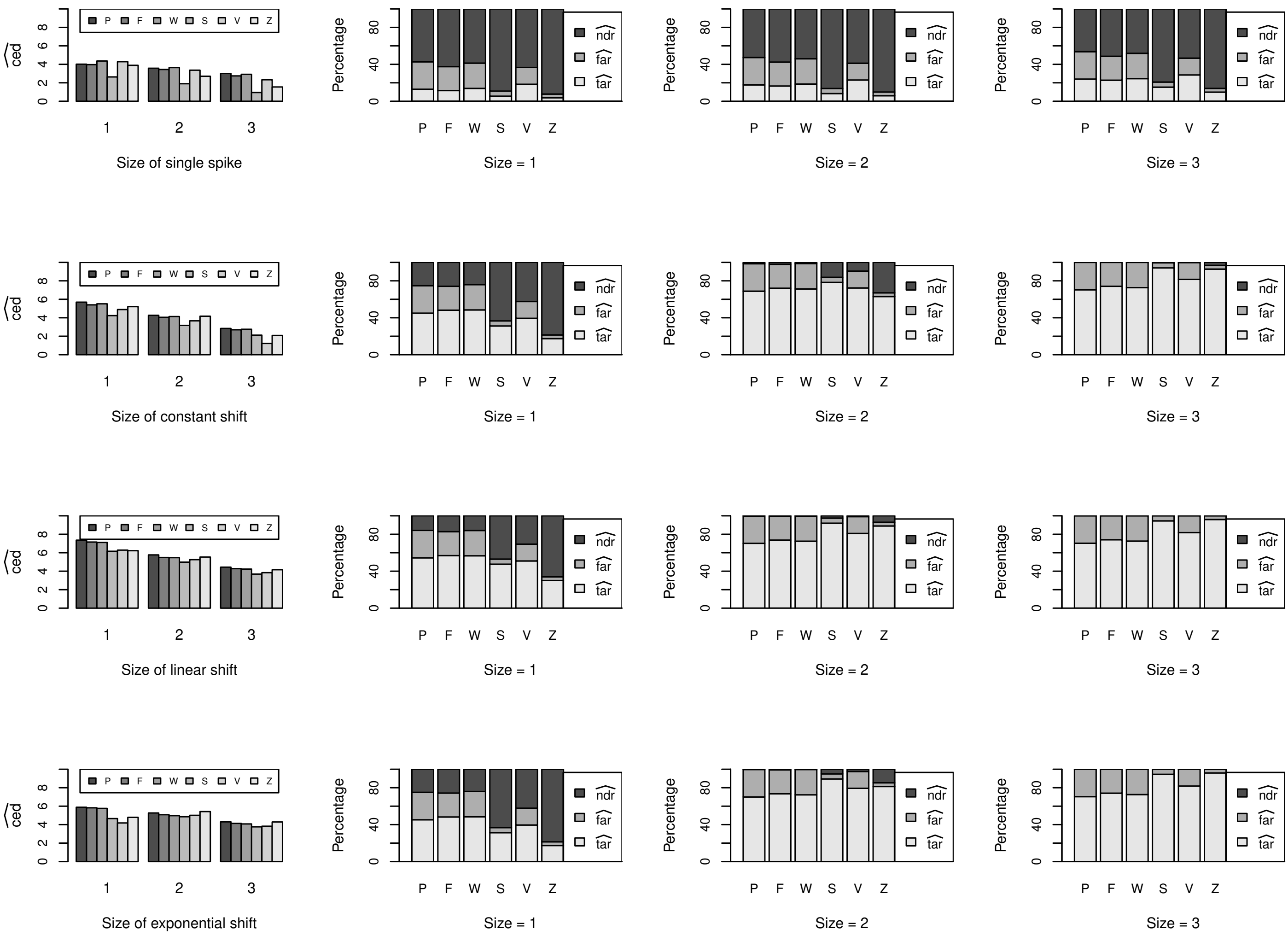

FIGURE 4 Bar charts comparing the detection performance in Scenario 3.

well for detecting large shift sizes by giving higher $\widehat{tar}$s than the other methods. However, the V method gave a shorter delay than the others for detecting a large constant shift size.

4 | APPLICATIONS

This section illustrates the use of SR methods for detecting a shift in real data. Since a change point is difficult to observe in real data, in our first example we illustrate the detection of simulated shifts in a real dataset. However, Example 2 shows the detection of a real shift in a real dataset.

Since we now have real data, we must use an initial ‘training period’ to estimate the parameters of the in control process, which is assumed bivariate AR(1), as in the simulation, though now with different variances and autoregressive coefficients. As previously, a change point divides the data into two periods. However, due to the lack of clarity in the change point, we define 12 time points prior to the declared change point as an arbitrary ‘guard region’. This region is also considered as a part of the ‘testing period’ in order to observe the possibility of false alarm; the ‘training period’ is that prior to the guard region.

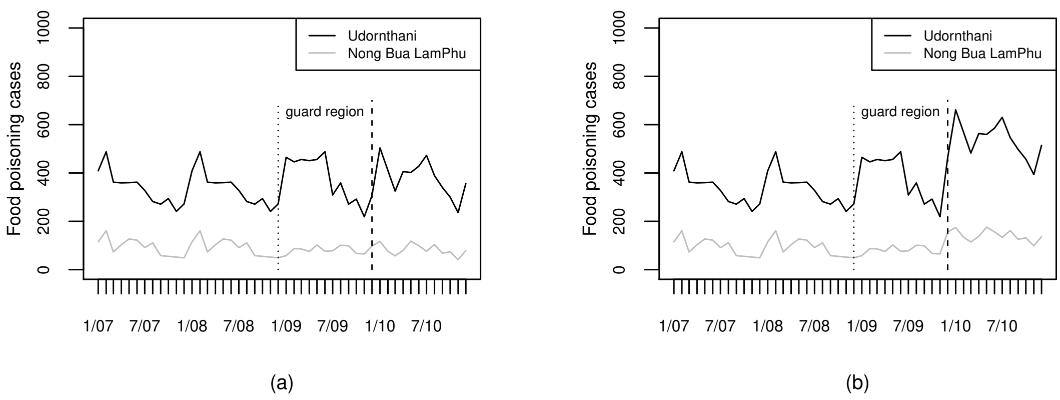

FIGURE 5 (a) Plot of food poisoning cases in Udornthani and Nong Bua LamPhu provinces in 2007 - 2010 (original data) and (b) plot of original data with added constant shift of size 3 standard deviations.

TABLE 2 Parameter estimation from the training period (Example 1).

Series	Mean	Standard deviation	Autoregressive coefficient	Correlation coefficient
Udornthani	335.67	67.87	0.61	0.83
Nong Bua LamPhu	93.08	35.16	0.35	

4.1 | Example 1

A plot of the number of food poisoning cases, reported monthly by the Bureau of Epidemiology, Thailand, in Nong Bua LamPhu and Udornthani provinces in 2007 - 2010 is shown in Figure 5. The vertical dashed line represents the change point. The vertical dotted line represents the start of the guard region. Since there is no evidence of a shift in the data, the system evaulation is investigated by injecting artificial extra cases into the original data. A change point is subjectively chosen in December 2009. The data are divided as below.

- the change point: December 2009 (vertical dashed line)
- an arbitrary guard region: December 2008 (vertical dotted line) - November 2009
- training period: January 2007 - November 2008
- testing period: December 2008 - December 2010

The parameter estimation from the training period is summarized in Table 2.

Four different shift types, each of three sizes (1, 2 and 3 standard deviations), are investigated. The shift is injected to the original data in December 2009 (the change point) and afterwards. The detection performance is evaluated by measuring the delay in detection (in months). Zero delay means that the shift can be detected in December 2009, while a negative delay is considered a false alarm. NA stands for failing to detect the shift within 7 time points (December 2009 - June 2010). The results in Table 3 show that the P and F methods give a false alarm midway through the guard region due to the CWS clearly present in the data. Even though V method takes the CWS into account, it still gives an even earlier false alarm. This corresponds to the results from the simulations in scenarios 2 and 3 showing that the V method is prone to false alarms. Surprisingly, the W method gives no false alarms. This may be due to the rather different CWS seen in the two real series (AR coefficients 0.61, 0.35), which

TABLE 3 The delay in detection for Example 1.

Type	Size	Method					
		P	F	W	S	V	Z
Single spike	1	-6	-6	NA	NA	-11	NA
	2	-6	-6	NA	NA	-11	NA
	3	-6	-6	NA	0	-11	0
Constant	1	-6	-6	NA	NA	-11	NA
	2	-6	-6	1	1	-11	1
	3	-6	-6	1	0	-11	0
Linear	1	-6	-6	NA	NA	-11	NA
	2	-6	-6	4	4	-11	5
	3	-6	-6	2	1	-11	1
Exponential	1	-6	-6	NA	NA	-11	NA
	2	-6	-6	2	1	-11	5
	3	-6	-6	1	1	-11	1

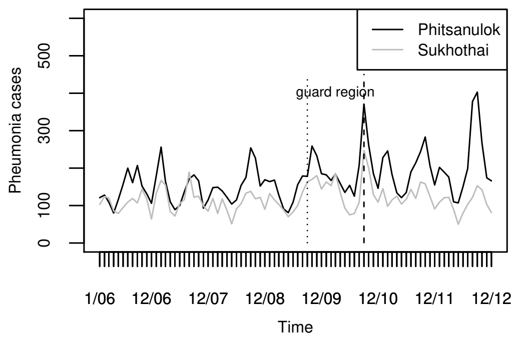

FIGURE 6 Plot of pneumonia cases in Phitsanulok and Sukhothai provinces in 2006 - 2012.

is not a situation we have explored in our simulation (AR coefficients both 0.6). Of course, the F method's allowance for a time lag offers no benefit here, as the injected shifts were simultaneous in both series. Also, the two series are strongly correlated ($\hat{\rho} = 0.83$), this might increase the possibility of detecting large shifts. Overall, S gives equal or shorter delays than W and Z methods.

4.2 | Example 2

Figure 6 shows the number of pneumonia cases, reported monthly by the Bureau of Epidemiology, Thailand, in Phitsanulok and Sukhothai provinces in 2006 - 2012. Apparently, due to the large spike in 2010, a change point of the process occurred in September 2010. Due to its short-lived nature, we consider this as a spike shift. The data are divided as below.

- change point: September 2010 (vertical dashed line)
- arbitrary guard region: September 2009 (vertical dotted line) - August 2010
- training period: January 2006 - August 2009

TABLE 4 Parameter estimation from the training period (Example 2).

Series	Mean	Standard deviation	Autoregressive coefficient	Correlation coefficient
Phitsanulok	145.66	44.38	0.59	0.71
Sukhothai	109.22	32.33	0.50	

TABLE 5 Delays in detection in Example 2.

Method	Delay in months
P	-6
F	-10
W	-10
S	0
V	0
Z	0

- testing period: September 2009 - December 2012

The parameter estimation from the training period is summarized in Table 4. The delays in detection (in months) are summarized in Table 5. Since CWS is clearly present in the data, the P, F and W methods produce false alarms, as expected, while other methods give zero delay as the large spike shift is quite easy to detect.

5 | DISCUSSION AND CONCLUSIONS

In this study we investigated how robust SR methods are for detecting various shift types and sizes in a bivariate process. Four shift types, each of three sizes, were considered. The detection performances of SR methods were compared against other statistical methods by estimating four measures, including the delay, true alarm, false alarm and non-detection rates, by simulation. Though the evaluation was conducted in a bivariate process, this can be extended to the multivariate case in the obvious manner.

Presence of a spike on only the first day in the testing period, means a single spike shift is not easy to detect, resulting in a high non-detection rate as compared to other shift types. Linear shifts can often be detected correctly (i.e. the true alarm rates are higher than other shift types), although their detection is typically delayed when compared to the constant and exponential shifts. This might be expected from Figure 1, which illustrates how the signal builds up over the application period, with sizes of the linear shift smaller than those of the constant and exponential shifts for the first half of the period and vice versa in the second half.

Even though the sufficient reduction methods (Wessman, Frisén and Siripanthana and Stillman), were proposed for detecting constant shifts in the multivariate process, they still perform robustly in some other circumstances. In a process of independent observations (Scenario 1), having incorporated the CBS in the sufficient reduction, the Wessman method detects faster and gives higher true alarm rates than the parallel and Frisén methods for all sizes of a single spike shift and small sizes of constant, linear and exponential shifts. In a process of dependent observations, whether or not CBS is present (Scenarios 2 and 3), the parallel, Frisén and Wessman methods produce high false alarm rates due to the effect of the CWS. The Siripanthana and Stillman method gives shorter delays for detecting small shifts of all types, while the VarR method gives a shorter delay for a large constant shift. Considering the ability to identify shifts correctly, Pan and Jarrett's VarR method gives higher true alarm rates than other methods for detecting all sizes of single spike shift. The Siripanthana and Stillman and Kalgonda and Kulkarni's Z methods both perform well for detecting large constant, linear and exponential shifts. However, they perform less well when those shift sizes are small.

Although the application of sufficient reduction methods in health surveillance was illustrated, it was quite difficult to define the exact time when the process shifted. The dramatic spike shift was easy to observe and was thus chosen as a change point of the process, but other shift types might be difficult to define in practice. The data in both examples show the presence of correlation within series. Methods not incorporating CWS produce the false alarms as expected, while other methods are able to detect the large spike shift without delay.

Here we have considered a negative delay (alarm in the guard region) as a false alarm. However, it might also be considered as a warning signal.[27] In health surveillance, disease prevention plays an important role; preventing the spread of disease can reduce the number of patients and cost of treatment. So the false alarm might be used as a warning sign for the practitioner to intervene or investigate the situation before an outbreak starts. On the other hand, in some situations, giving a false alarm causes more trouble such as panic, unnecessary intervention or investigation and the consequent cost of using limited resources. The relative 'costs' must be carefully considered as part of the design of a practical system, choosing the detection threshold to give the desired balance.

Sufficient reduction methods can be applied in several disciplines where the likely data might be positive counts and possibly contain lots of zero (e.g. number of nonconforming items in industrial processes or cases of a rare disease). Further study is focusing on developing sufficient reduction methods for detecting shifts in Poisson, or zero-inflated-Poisson, processes, along with a proper control chart for monitoring shifts in Poisson processes.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from Naresuan University in the form of research grant-number R2558C239 during the period in which the research was conducted.

Financial disclosure

None reported.

Conflict of interest

The authors declare no potential conflict of interests.

References

1. Strat Y. *Overview of temporal surveillance*; Chichester: John Wiley & Sons . 2005.
2. Frisén M. Methods and Evaluations for Surveillance in Industry, Business, Finance, and Public Health. *Qual Reliab Eng Int* 2011; 27: 611–621.
3. Noorossana R, Hosseini SS, Heydarzade A. An overview of dynamic anomaly detection in social networks via control charts. *Qual Reliab Eng Int* 2018; 34: 641–648.
4. Bédubourg G, Le Strat Y. Evaluation and comparison of statistical methods for early temporal detection of outbreaks: A simulation-based study. *PLoS One* 2017; 12(7): 1–18.
5. Shmueli G, Fienberg S. *Current and Potential Statistical Methods for Monitoring Multiple Data Streams for Biosurveillance*; New York: Springer . 2006.
6. Unkel S, Farrington C, Garthwaite P, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *J R Stat Ser A Stat Soc* 2012; 31(1): 49–82.

7. Marshall C, Best N, Bottle A, Aylin P. Statistical Issues in the Prospective Monitoring of Health Outcomes Across Multiple Units. *J R Stat Ser A Stat Soc* 2004; 167(3): 541–559.

8. Frisén M. *Principles for Multivariate Surveillance*: 133–144; Physica-Verlag HD . 2010.

9. Yahav I, Shmueli G. Directionally Sensitive Multivariate Control Charts in Practice: Application to Biosurveillance. *Qual Reliab Eng Int* 2014; 30: 159–179.

10. Bersimis S, Psarakis S, Panaretos J. Multivariate Statistical Process Control Charts: An Overview. *Qual Reliab Eng Int* 2007; 23: 517–543.

11. Cox DR, Hinkley DV. *Theoretical Statistics*. London: Chapman & Hall . 1974.

12. Wessman P. Some principles for surveillance adopted for multivariate processes with a common change point. *Commun Stat Theory Methods* 1998; 27: 1143–1161.

13. Frisén M, Andersson E, Schioler L. Sufficient Reduction in Multivariate Surveillance. *Commun Stat Theory Methods* 2011; 40(10): 1821-1838.

14. Siripanthana S, Stillman EC. Sufficient reduction methods for multivariate time-dependent health surveillance data. *Commun Stat Theory Methods* 2016; 45(4): 950–966.

15. Frisén M, De Maré J. Optimal Surveillance. *Biometrika* 1991; 78: 271–280.

16. Frisén M. Evaluation of Methods for Statistical Surveillance. *Stat Med* 1992; 11: 1489–1502.

17. Hutwagner L, Browne T, Seeman G, Fleischauer A. Comparing Aberration Detection Methods with Simulated Data. *Emerg Infect Dis* 2005; 11(2): 314-316.

18. Buckeridge DL, Okhmatovskaia A, Tu S, O'Connor M, Nyulas C, Musen MA. Predicting Outbreak Detection in Public Health Surveillance: Quantitative Analysis to Enable Evidence-Based Method Selection. *AMIA Annu Symp Proc* 2008: 76–80.

19. Mathes RW, Lall R, Levin-Rector A, et al. Evaluating and implementing temporal, spatial, and spatio-temporal methods for outbreak detection in a local syndromic surveillance system. *PLoS One* 2017; 12: 1-19.

20. Wang L, Ramoni M, Mandl K, Sebastiani P. Factors affecting automated syndromic surveillance. *Artif Intell Med* 2005; 34: 269–278.

21. Montgomery D. *Statistical Quality Control: A Modern Introduction*. New York: Wiley. 6th ed. 2009.

22. Robinson PB, Ho TY. Average Run Lengths of Geometric Moving Average Charts by Numerical Methods. *Technometrics* 1978; 20(1): 85–93.

23. Crowder SV. Design of Exponentially Weighted Moving Average Schemes. *J Qual Technol* 1989; 21: 155–162.

24. Shu L, Jiang W, Wu S. A One-Sided EWMA Control Chart for Monitoring Process Means. *Commun Stat Simul Comput* 2007; 36: 901–920.

25. Pan X, Jarrett JE. Why and how to use vector autoregressive models for quality control: the guideline and procedures. *Qual Quant* 2012; 46: 935–948.

26. Kalgonda AA, Kulkarni SR. Multivariate Quality Control Chart for Autocorrelated Processes. *J Appl Stat* 2004; 31: 317–327.

27. Bastos LS, Economou T, Gomes MFC, et al. A modelling approach for correcting reporting delays in disease surveillance. *Stat Med* 2019; 38: 4363–4377.

AUTHOR BIOGRAPHY

Sawaporn Hinsheranan is a member of academic staff of the Department of Mathematics, Faculty of Science at Naresuan University. She earned her BSc degree in Statistics from Chiang Mai University and received her MSc and PhD degrees in Statistics from the University of Sheffield. Her research interests include the dimensionality reduction and statistical process control.

Eleanor C. Stillman is a part time Senior Lecturer in the School of Mathematics and Statistics at the University of Sheffield. She took BSc and PhD degrees at Sheffield before taking up a lectureship there in 1987. Her research interests are varied but include the formulation and use of particle size and shape models, process control and optimization, and optimal experimental design.