Impact of Autocorrelation on Principal Components and Their Use in Statistical Process Control

Vanhatalo, Erik; Kulahci, Murat

Published in:
Quality and Reliability Engineering International

Link to article, DOI:
10.1002/qre.1858

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Research Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1858

Published online in Wiley Online Library

Impact of Autocorrelation on Principal Components and Their Use in Statistical Process Control

Erik Vanhatalo[a*†] and Murat Kulahci[a,b]

A basic assumption when using principal component analysis (PCA) for inferential purposes, such as in statistical process control (SPC), is that the data are independent in time. In many industrial processes, frequent sampling and process dynamics make this assumption unrealistic rendering sampled data autocorrelated (serially dependent). PCA can be used to reduce data dimensionality and to simplify multivariate SPC. Although there have been some attempts in the literature to deal with autocorrelated data in PCA, we argue that the impact of autocorrelation on PCA and PCA-based SPC is neither well understood nor properly documented.

This article illustrates through simulations the impact of autocorrelation on the descriptive ability of PCA and on the monitoring performance using PCA-based SPC when autocorrelation is ignored. In the simulations, cross-correlated and autocorrelated data are generated using a stationary first-order vector autoregressive model.

The results show that the descriptive ability of PCA may be seriously affected by autocorrelation causing a need to incorporate additional principal components to maintain the model's explanatory ability. When all variables have equal coefficients in a diagonal autoregressive coefficient matrix, the descriptive ability is intact, while a significant impact occurs when the variables have different degrees of autocorrelation. We also illustrate that autocorrelation may impact PCA-based SPC and cause lower false alarm rates and delayed shift detection, especially for negative autocorrelation. However, for larger shifts, the impact of autocorrelation seems rather small. © 2015 The Authors. *Quality and Reliability Engineering International* published by John Wiley & Sons Ltd.

Keywords: principal component analysis; multivariate data; serial dependence; vector autoregressive model

1. Introduction

The popularity of latent variable methods such as principal component analysis (PCA) keeps growing alongside the development of automatic data collection schemes with increasing availability of multivariate data. PCA is one of the most commonly used techniques in multivariate analysis, and according to Jolliffe[1] (p. 1), the central idea of principal component analysis (PCA) is "*to reduce dimensionality of a data set consisting of a large number of interrelated variables*".

In many applications of PCA, the purpose is descriptive. That is, we want to reduce the dimensions of the data and describe and interpret the data through a fewer number of latent variables. Reducing the dimensions is also important for inferential purposes, such as in statistical process control (SPC). In many SPC applications, the large number of quality characteristics makes univariate monitoring ineffective and inefficient; see MacGregor.[2] The difficulty for the engineer to simultaneously keep track of many univariate control charts is generally a strong argument in favor of a multivariate approach to SPC. Process monitoring by use of multivariate statistical methods has been an active research area during the last few decades. Excellent and comprehensive overviews of developments of methods within this area are given by, for example, Bersimis *et al*.,[3] Kourti,[4] and Qin.[5]

Traditional SPC techniques assume independent data in time. However, this assumption is becoming increasingly unrealistic in today's applications. Because of system dynamics combined with frequent sampling, successive observations will often be serially correlated; see Montgomery *et al*.[6] and Bisgaard and Kulahci.[7] The issue of autocorrelation in univariate SPC charts has been discussed by many authors.[8–13] Clearly, less research has been reported on the effects of and remedies for autocorrelation in multivariate SPC charts. Vanhatalo and Kulahci[14] show how the Hotelling T^2 chart is affected by autocorrelation. There are also articles proposing potential solutions to this problem in multivariate SPC.[14–20] In this article, we simply explore the impact of autocorrelation on PCA and PCA-based SPC and leave a more in-depth study on solutions to remedy this impact for our future research.

[a]*Luleå University of Technology, Luleå, Sweden*
[b]*Technical University of Denmark, Kongens Lyngby, Denmark*
**Correspondence to: Erik Vanhatalo, Luleå University of Technology, Luleå, Sweden.*
[†]*E-mail: erik.vanhatalo@ltu.se*

2. Motivation

Principal components are linear combinations of cross-correlated variables that are assumed to be independent in time. Although it is not required for calculating the principal components, useful interpretation of the principal components and inferences can be made if data are assumed to be independent and come from a multivariate normal distribution; see Johnson and Wichern[21] and Jolliffe.[1] In summary, Jollife[1] (p. 299) states that "*when the main objective of PCA is descriptive, not inferential, complications such as non-independence [in time] does not seriously affect this objective*". In this article, we aim to put this statement to test and investigate how the descriptive ability of PCA is affected by autocorrelation.

If the purpose of PCA is inferential, such as when PCA is used within SPC and the principal components are monitored for a process with multiple quality characteristics, serial dependence in the data is expected to affect monitoring performance. This is easily explained by the fact that the principal components are linear combinations of autocorrelated variables and therefore, the principal components will also be autocorrelated.

Often, empirical studies on PCA-based SPC do not discuss remedies for autocorrelated data but rather proceed under the (unspoken) assumption that the data are independent even though it can be more plausible from the specific cases that the data are autocorrelated; see, for example, Vanhatalo.[22] In general, we argue that there is no clear-cut recommendation on how to deal with autocorrelated data for PCA-based SPC. However, within chemometrics literature, a method of augmenting the input and/or output matrix with time-lagged values of the variables, as in the case of the so-called dynamic PCA (DPCA), has been put forward as a solution; see Ku *et al.*[23] and Kourti and MacGregor.[24] However, monitoring based on DPCA still suffers from autocorrelated principal components, and recently, Rato and Reis[25] suggested an improvement using decorrelated residuals in DPCA.

3. Principal component analysis

This section provides a short technical background on PCA. For a more complete explanation, see, for example, Johnson and Wichern,[21] Jackson,[26] and Jolliffe.[1]

Let the p variable random vector $\mathbf{X}' = [x_1, x_2, \ldots, x_p]$ have the associated covariance matrix $\boldsymbol{\Sigma}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$. In particular, $\boldsymbol{\Sigma}$ is the symmetric $p \times p$ matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}, \tag{1}$$

where the diagonal elements are the variances of each variable and the off-diagonal elements are the covariances among the variables. Let $\mathbf{e}_i$, $i = 1, 2, \ldots, p$, be the eigenvectors of $\boldsymbol{\Sigma}$, and given that the eigenvectors form columns of matrix $\mathbf{C}$, we have

$$\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C} = \boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, , \lambda_2, \ldots, , \lambda_p\} \tag{2}$$

The p principal components (scores) are then given as follows:

$$\begin{aligned} z_1 &= \mathbf{e}_1\mathbf{X} = e_{11}x_1 + \ e_{12}x_2 + \ldots + \ e_{1p}x_p \\ z_2 &= \mathbf{e}_2\mathbf{X} = e_{21}x_1 + \ e_{22}x_2 + \ldots + \ e_{2p}x_p \\ &\vdots \\ z_p &= \mathbf{e}_p\mathbf{X} = e_{p1}x_1 + \ e_{p2}x_2 + \ldots + \ e_{pp}x_p \end{aligned} \tag{3}$$

where e_{ij} is the jth entry of the ith eigenvector corresponding to the ith largest eigenvalue. In many cases, less than p principal components are sufficient to explain a reasonable amount of the total variance. The proportion of the total population variance that can be explained by the ith principal component is given by

$$\frac{\lambda_i}{(\lambda_1 + \lambda_2 + \ldots + \lambda_p)} \tag{4}$$

The components (loadings) of each eigenvector $\mathbf{e}_i' = [e_{i1}, , e_{i2}, \ldots, , e_{ip}]$ are important to interpret the underlying meaning of the principal component. For example, the sign and magnitude of e_{ij} determine the importance of the jth variable on the ith principal component and are proportional to the correlation between z_i and x_j.

In cases where the scale and the variance of the variables are substantially different, it is a common practice to obtain principal components from standardized variables, which is equivalent to obtaining the eigenvalue–eigenvector pairs from the correlation matrix $\boldsymbol{\rho}$ of $\mathbf{X}$.

Jolliffe[1] describes a number of ways to numerically calculate the principal components, such as through singular value decomposition. Another method popularized within chemometrics literature is the nonlinear iterative partial least squares algorithm,[27] which often is faster when only the first few principal components need to be calculated and also has the advantage of working for matrices with a moderate amount of normally distributed missing observations in the matrix.

4. Simulated data

Although it is not required by the PCA technique, we will assume that our quality characteristics (variables) follow a multivariate normal distribution. The multivariate normal distribution is an extension of the univariate normal distribution to a situation with multiple (p) variables; see, for example, Montgomery.[28] The multivariate normal density function is

$$f(\mathbf{X}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})} \tag{5}$$

where $\mathbf{X}' = [x_1, x_2, \ldots, x_p]$, $-\infty < x_j < \infty$, and $j = 1, 2, \ldots, p$, and $\boldsymbol{\mu}$ is a $p \times 1$ vector with the means of the p variables.

We simulate independent errors for the variables $\mathbf{X}' = [x_1, x_2, \ldots, x_p]$, from a multivariate normal distribution with mean zero and with covariance matrix $\mathbf{\Sigma}$. Autocorrelation in the variables is introduced by 'filtering' the errors through a first-order vector autoregressive model, VAR(1). Now, let

$$\mathbf{X}_t = \mathbf{c} + \mathbf{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$$

or in matrix form

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{p,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1p} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{p1} & \phi_{p2} & \cdots & \phi_{pp} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \\ \vdots \\ x_{p,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{p,t} \end{bmatrix} \tag{6}$$

or

$$\begin{aligned} x_{1,t} &= c_1 + \phi_{11}x_{1,t-1} + \phi_{12}x_{2,t-1} + \ldots + \phi_{1p}x_{p,t-1} + \varepsilon_{1,t} \\ x_{2,t} &= c_2 + \phi_{21}x_{1,t-1} + \phi_{22}x_{2,t-1} + \ldots + \phi_{2p}x_{p,t-1} + \varepsilon_{2,t} \\ &\vdots \\ x_{p,t} &= c_p + \phi_{p1}x_{1,t-1} + \phi_{p2}x_{2,t-1} + \ldots + \phi_{pp}x_{p,t-1} + \varepsilon_{p,t} \end{aligned}$$

where $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{\Sigma})$. This allows us to manipulate the autocorrelation structure in the variables through the $\mathbf{\Phi}$ matrix. For the process to be stationary, all absolute values of the eigenvalues of the autoregressive coefficient matrix $\mathbf{\Phi}$ should be less than one. For a stationary VAR(1) process, the expected value is

$$E(\mathbf{X}_t) = \boldsymbol{\mu} = (\mathbf{I} - \mathbf{\Phi})^{-1}\mathbf{c}\,, \tag{7}$$

where $\mathbf{I}$ is the identity matrix. The covariance matrix of the VAR(1) process is then computed using the following equations; see Reinsel[29]:

$$\begin{aligned} \mathrm{cov}(\mathbf{X}_t) &= \mathrm{cov}(\mathbf{c}) + \mathrm{cov}(\mathbf{\Phi}\mathbf{X}_{t-1}) + \mathrm{cov}(\boldsymbol{\varepsilon}_t) \\ \mathbf{\Gamma}(0) &= \mathbf{\Phi}\mathbf{\Gamma}(0)\mathbf{\Phi}' + \mathbf{\Sigma} \end{aligned} \tag{8}$$

where $\mathbf{\Gamma}(0)$ is the covariance matrix of the data and $\mathbf{\Sigma}$ is the covariance matrix for the errors. The covariance structure of the first-order autoregressive process is hence dependent on both the autocorrelation matrix $\mathbf{\Phi}$ and the covariance matrix $\mathbf{\Sigma}$ of the errors, which we need to consider in what follows. Note that by choosing $\mathbf{c}$ and $\mathbf{\Phi}$ as a zero vector/matrix, the model in (6) reduces back to generating data from a multivariate normal distribution with zero mean.

All simulations in this article were performed using R statistics software, and the R code for the simulations is available upon request.

5. Impact on the descriptive ability: a look at the bivariate case

To illustrate and visualize the impact of autocorrelation on PCA and PCA-based SPC, we start with a simple bivariate case:

$$\begin{aligned} x_{1,t} &= c_1 + \phi_{11}x_{1,t-1} + \phi_{12}x_{2,t-1} + \varepsilon_{1,t} \\ x_{2,t} &= c_2 + \phi_{21}x_{1,t-1} + \phi_{22}x_{2,t-1} + \varepsilon_{2,t} \end{aligned} \tag{9}$$

where we let

$$\mathbf{\Phi} = \begin{bmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{bmatrix} \text{ and } \mathbf{\Sigma} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix} \tag{10}$$

Thus, for time-independent data (i.e., $\mathbf{\Phi} = 0$), the eigenvalues $\mathbf{\Sigma}$ are 1.9 and 0.1, and we should thus expect one dominant latent variable. In other words, the first principal component would explain most of the variance in the data. Here, $\mathbf{\Sigma}$ can be viewed as giving the 'static relations' among the variables, while $\mathbf{\Phi}$ determines the 'dynamic relations' in the form of autocorrelation.

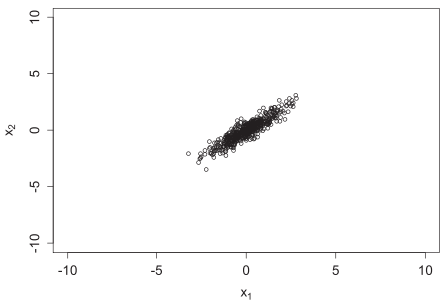
Figure 1a–d visualizes 500 observations of simulated data with different autocorrelation structures, the two estimated eigenvalues of the correlation matrix of the variables, and the proportion of explained variance per principal component. All four cases in Figure 1a–d are based on the same innovations (errors).

As illustrated in Figure 1a–d, as autocorrelation increases in a variable, so does the variance of that variable, at least for the tested cases with a diagonal $\mathbf{\Phi}$ matrix. It is worth noting that when the degree of autocorrelation in the two variables is unequal, the correlation between the variables is 'distorted', which also affects the results from PCA. For example, in Figure 1d, the first principal component explains around 54% of the variance in the data compared with around 95% for independent data in Figure 1a. It is thus evident from these limited simulations that autocorrelation affects the descriptive ability of PCA.

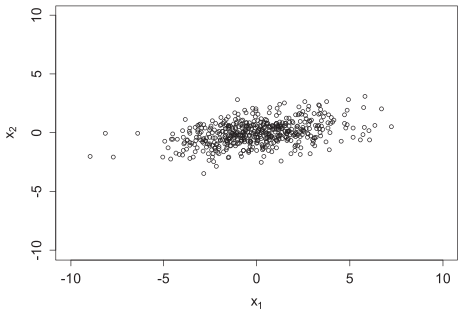
It turns out that for a diagonal $\mathbf{\Phi}$ matrix when the both variables have equal autocorrelation coefficients, that is, $\phi_{11} = \phi_{22} = \phi$, the true correlation between the variables is the same as for independent data. We can show that

$$\begin{aligned}\mathbf{\Gamma}(0) &= \mathbf{\Phi}\mathbf{\Gamma}(0)\mathbf{\Phi}' + \mathbf{\Sigma} \\ &= \begin{bmatrix}\phi & 0\\ 0 & \phi\end{bmatrix}\mathbf{\Gamma}(0)\begin{bmatrix}\phi & 0\\ 0 & \phi\end{bmatrix}' + \mathbf{\Sigma} \\ &= \phi\mathbf{I}\mathbf{\Gamma}(0)\phi\mathbf{I} + \mathbf{\Sigma} \\ &= \phi^2\mathbf{\Gamma}(0) + \mathbf{\Sigma} \\ &\Rightarrow \mathbf{\Gamma}(0)\frac{1}{1-\phi^2}\mathbf{\Sigma}\end{aligned}$$

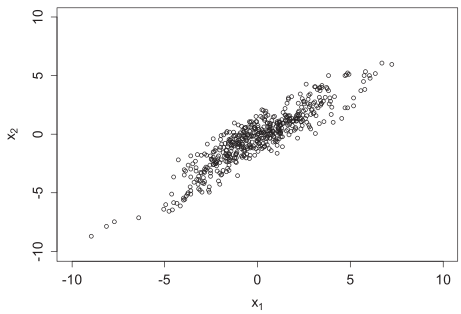
Because in this case, the covariance of the data is simply a scaled version of the covariance of the errors, the correlation between the two variables is the same as the correlation between two errors, and therefore, eigenvalues of the correlation matrix of the variables are the same as the eigenvalues of the correlation matrix of the errors. Furthermore, because we consider the stationary



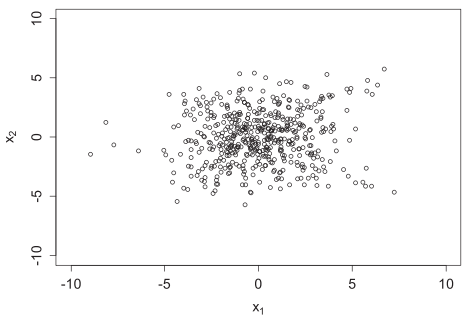
1a	PC_1	PC_2
Eigenvalue	1.908	0.092
Percent of variance	95.4 %	4.6 %



1b	PC_1	PC_2
Eigenvalue	1.400	0.600
Percent of variance	70.0 %	30.0 %



1c	PC_1	PC_2
Eigenvalue	1.910	0.090
Percent of variance	95.5 %	4.5 %



1d	PC_1	PC_2
Eigenvalue	1.088	0.912
Percent of variance	54.4 %	45.6 %

Figure 1. (a) No autocorrelation, $\phi_{11} = \phi_{22} = 0$; (b) autocorrelation in x_1, $\phi_{11} = 0.9$, $\phi_{22} = 0$; (c) autocorrelation in x_1 and x_2, $\phi_{11} = \phi_{22} = 0.9$; and (d) autocorrelation in x_1 and x_2, $\phi_{11} = 0.9$, $\phi_{22} = -0.9$

process, that is, $|\phi| < 1$, the variances of the variables are bigger than the variances of the errors as can be seen by comparing Figure 1a and 1c.

However, when $\phi_{11} \neq \phi_{22}$, the correlation between the variables is distorted with the biggest change occurring when the autocorrelation coefficients are large but with opposite signs; see Figure 1d. This phenomenon is further illustrated in Figure 2, which shows the 'true' correlation among the variables as a function of ϕ_{11} and ϕ_{22} for the $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ matrices given in (10).

6. Impact on PCA-based statistical process control in the bivariate case

In this section, we discuss the false alarm rate and the shift detection ability in phase II of PCA-based control charts for two variables only for illustration purposes. In our simulations, the PCA model is built based on observations from a process that initially is in control (phase I). This model is then used in the online monitoring stage (phase II). Because different autocorrelation structures will be tested in the simulations, the variances of the variables will not always be equal. Therefore, standardized variables will be used. In phase II, the variables are standardized based on their mean vector and standard deviations from the phase I sample. The simulations are based on a phase I sample of $m = 500$ observations of an in-control process. The scores of the two principal components in phase II are here monitored in separate Shewhart charts as well as in a Hotelling T^2 chart for individual observations, and the average run length is measured. Three-sigma limits are used in the Shewhart charts. That is,

$$\begin{aligned} UCL_{\text{Shewhart},PC_i} &= 3 \cdot \sqrt{\lambda_i} \\ LCL_{\text{Shewhart},PC_i} &= -3 \cdot \sqrt{\lambda_i} \end{aligned} \tag{11}$$

where PC_i is the ith principal component corresponding to the ith eigenvalue, λ_i. In the Hotelling T^2 chart, we use the traditional estimator $\mathbf{S}$ for the sample covariance matrix as Vanhatalo and Kulahci[14] showed it to be less sensitive to autocorrelation:

$$\mathbf{S} = \frac{1}{m-1} = \sum_{i=1}^{m} (x_i - \bar{x}) \times (x_i - \bar{x})' \tag{12}$$

The phase II upper control limit for the Hotelling T^2 chart using the traditional estimator $\mathbf{S}$ is based on the F distribution and given as follows:

$$UCL_{T^2} = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha,p,m-p} \tag{13}$$

where p is the number of variables, m is the number of samples (i.e., observations) in phase I, α is the acceptable false alarm rate, and $F_{\alpha,p,m-p}$ is the upper αth percentile of the F distribution with p and $m-p$ degrees of freedom.

The shift detection ability is measured by generating shifts in the process mean, δ_1 and δ_2, in terms of the 'true' standard deviation of the variables in the VAR(1) model, which can be calculated using (8). It should be noted that the control limits in (11) and (13) are used as we simply follow the naïve approach where the autocorrelation is ignored or overlooked.

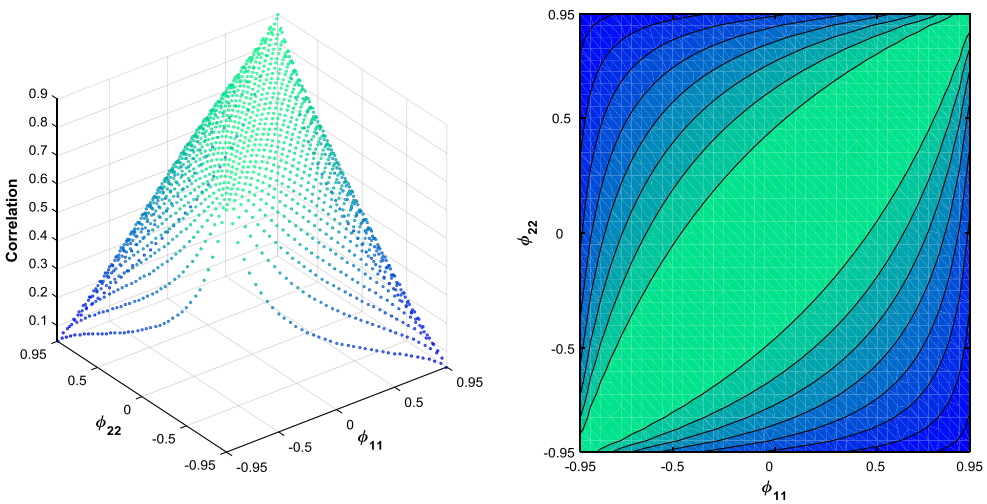


Figure 2. Visualization of the true correlation between the two variables depending on the autocorrelation coefficients ϕ_{11} and ϕ_{22} for the $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ matrices given in (10)

6.1. False alarm rate in the bivariate case

The average in-control run length (ARL_0), in phase II, for individual Shewhart charts for the principal components and the Hotelling T^2 chart based on both principal components, is given in Table I, which covers the four different cases of autocorrelation structures shown in Figure 1. From Table I, it can be seen that the false alarm rate decreases (ARL_0 values increase) with increasing magnitudes of autocorrelation in the variables. For the Shewhart charts, the most dramatic increase in ARL_0 comes for the case with high and positive autocorrelation in both variables, while for the Hotelling T^2 chart, the largest ARL_0 value is for the case with high autocorrelation but with opposite signs in both variables.

6.2. Shift detection capability in the bivariate case

The average out-of-control run length (ARL_1), in phase II, for individual Shewhart charts for the principal components and the Hotelling T^2 chart based on both principal components, is given in Table II. The behavior of the ARL_1 values depends on the different shift scenarios. For example, the shift scenario that is 'easiest' to detect in Table II is when both variables shift but in different directions, which goes against the correlation structure among the errors in the phase I sample. However, it is more difficult to detect the case with equal shifts in both variables. In general, the shift detection capability is reduced with increasing magnitude of autocorrelation in the variables.

The low ARL_1 values for the Shewhart chart for PC_2 for cases (a) and (c) when there is only a shift in x_1 are a consequence of the true correlation structure of the variables in phase I; see also Figure 2. In cases (a) and (b), the variables are highly correlated, which causes the shift in only x_1 to be unusual as the shift is in another direction than the main latent structure (PC_1). The shift is therefore easily detected in the direction of PC_2.

7. Impact on PCA-based statistical process control for a five-variable example

Finally, we explore the impact of autocorrelation in a somewhat more complex case. Here, we choose to explore a five-variable case, which will provide the possibility to study the performance of PCA-based SPC using the combination of the two complementing charts: [1] a Hotelling T^2 chart for the A first 'important' principal components (T^2_A chart) and [2] the squared prediction error (SPE); see, for example, Ferrer.[30]

The SPE for observation i is given as follows:

$$SPE = \mathbf{e}_i^T \mathbf{e}_i = \left(\mathbf{x}_i - \mathbf{x}_i^*\right)^T \left(\mathbf{x}_i - \mathbf{x}_i^*\right) \tag{14}$$

where $\mathbf{e}_i$ is the residual vector for the ith observation and $\mathbf{x}_i^*$ is the predicted observation vector from the PCA model based on the A first principal components.

An abnormal value in T^2_A chart can be viewed as an outlier in the dimensions of the retained principal components and indicates that the observation includes some extreme values in some (or all) of the original variables, while the correlation structure among the

Table I. Average in-control run length (ARL_0) for different autocorrelation structures based on 10,000 simulations in each case

			Shewhart		T^2
Case	$\phi_{11}=$	$\phi_{22}=$	PC_1	PC_2	PC_{1-2}
a	0	0	386.4	380.9	401.6
b	0.9	0	413.8	398.4	540.8
c	0.9	0.9	922.3	784.8	573.4
d	0.9	−0.9	565.8	552.6	901.6

Table II. Average out-of-control run length (ARL_1) for different autocorrelation structures based on 10,000 simulations in each case

			$\delta_1=1,\ \delta_2=0$			$\delta_1=1,\ \delta_2=1$			$\delta_1=1,\ \delta_2=-1$		
			Shewhart		T^2	Shewhart		T^2	Shewhart		T^2
Case	$\phi_{11}=$	$\phi_{22}=$	PC_1	PC_2	PC_{1-2}	PC_1	PC_2	PC_{1-2}	PC_1	PC_2	PC_{1-2}
a	0	0	156.1	4.6	6.3	42.8	373.0	69.7	386.8	1.1	1.2
b	0.9	0	184.9	100.7	166.1	58.0	369.3	73.8	400.2	24.4	29.2
c	0.9	0.9	514.4	34.0	39.8	184.6	777.5	192.4	958.4	10.4	12.0
d	0.9	−0.9	206.3	176.4	253.6	54.5	501.6	79.4	518.5	43.0	62.7

variables remains intact. An outlier in the SPE chart can be viewed as an outlier in the dimensions of the remaining principal components not included in the model and indicates that the correlation structure among the variables captured in phase I is different from what is being observed in phase II. In other words, the outlier does not behave in the same way as what is expected from the reference data set.

There are several different approximations for the upper control limit for the SPE chart; see, for example, Ferrer.[30] Here, we choose the approximation given by Jackson and Mudholkar[31] because, for our limited initial simulations, it seemed to provide in-control run lengths closest to the nominal value but still somewhat too high. The upper control limit for the SPE chart is given by

$$UCL_{SPE,\alpha} = \theta_1 \left[\frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 - \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0} \tag{15}$$

where z_α is the $100(1-\alpha)$ percentile of the standardized normal distribution, $\theta_i = \sum_{j=A+1}^{p} \lambda_j^i$, and $h_0 = 1 - (2\theta_1\theta_3)/3\theta_2^2$.

Even with as few as five variables, the number of possible combinations of covariance structures for the errors and autocorrelation basically becomes unfeasibly large. Therefore, we only consider a model for a given covariance structure of the errors and vary the autocorrelation through different diagonal $\mathbf{\Phi}$ matrices.

We assume that we have a five-variable VAR(1) model with the following error covariance matrix (static relations):

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0.9 & 0.8 & 0 & 0 \\ 0.9 & 1 & 0.7 & 0 & 0 \\ 0.8 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0 & 0.9 & 1 \end{bmatrix}$$

which essentially means that the resulting variables are correlated through two blocks of correlated errors. The first block contains correlated errors for x_1, x_2, x_3 and the second block, correlated errors for x_4 and x_5. In the simulations, we will change the parameters of the diagonal $\mathbf{\Phi}$ matrix (dynamic relations) as follows:

$$\mathbf{\Phi} = \begin{bmatrix} \phi_{11} & 0 & 0 & 0 & 0 \\ 0 & \phi_{22} & 0 & 0 & 0 \\ 0 & 0 & \phi_{33} & 0 & 0 \\ 0 & 0 & 0 & \phi_{44} & 0 \\ 0 & 0 & 0 & 0 & \phi_{55} \end{bmatrix}$$

applying different combinations of autocorrelation parameters using the following values:

$$\phi_{ii} = \begin{cases} +0.9 & \text{high positive autocorrelation} \\ 0 & \text{no autocorrelation (independent in time)} \\ -0.9 & \text{high negative autocorrelation} \end{cases}$$

We consider both the in-control and the out-of-control run lengths for different types of shifts in the variables.

In the simulations, we need to decide on A: how many principal components to retain? There are several different rules and recommendations that can be used, such as the SCREE plot, minimum eigenvalue, or cross-validation. However, because we aim to study the impact of autocorrelation on PCA-based SPC, we choose to treat the time-independent case as the baseline. For time-independent data, the two first principal components would here explain roughly 90.1% of the total variance. Therefore, in the simulations with autocorrelated data, we retain as many principal components as required to explain at least 90.1% of the variance. All the following simulations are based on a phase I sample of $m = 500$ observations from an in-control process.

7.1. False alarm rate for the five-variable example

The average in-control run length (ARL_0), in phase II, for the T_A^2 and SPE charts as well as the average number of retained principal components in the simulations, are provided in Tables III–V for a variety of autocorrelation coefficient combinations.

In Tables III–V, we observe that the average number of retained principal components varies for different magnitudes and signs of autocorrelation. This is because the autocorrelation distorts the correlation among the variables compared to the time-independent case where two principal components are enough to explain just over 90% of the variance. Instead, two to five principal components are needed to be retained depending on the dynamic relations in the $\mathbf{\Phi}$ matrix. This is also clearly visible from the 'true' correlation matrices for the variables and their eigenvalues. In other words, the impact of autocorrelation may be interpreted as a distortion of the relative importance among the latent structures in the data. Thus, the descriptive

Table III. Average in-control run length (ARL_0) for the T^2_A and SPE charts and the average number of retained principal components ($\bar{A}$) for **time-independent data** and different positive autocorrelation coefficients

	a	b	c	d	e	f
Autocorrelation matrix	$\mathbf{\Phi} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0.8 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 0.5 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0.9 \end{bmatrix}$
ARL_{0,T^2_A}	**417**	445	426	428	415	603
$ARL_{0,SPE}$	**489**	503	362	525	514	648
$\bar{A}$	**2.473**	3.000	4.000	3.000	2.996	2.451
True correlation matrix of the variables	$\begin{bmatrix} 1 & 0.9 & 0.8 & 0 & 0 \\ 0.9 & 1 & 0.7 & 0 & 0 \\ 0.8 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0 & 0.9 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.39 & 0.35 & 0 & 0 \\ 0.39 & 1 & 0.7 & 0 & 0 \\ 0.35 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0 & 0.9 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.39 & 0.35 & 0 & 0 \\ 0.39 & 1 & 0.7 & 0 & 0 \\ 0.35 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.39 \\ 0 & 0 & 0 & 0.39 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.84 & 0.35 & 0 & 0 \\ 0.84 & 1 & 0.42 & 0 & 0 \\ 0.35 & 0.42 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.84 \\ 0 & 0 & 0 & 0.84 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.84 & 0.67 & 0 & 0 \\ 0.84 & 1 & 0.68 & 0 & 0 \\ 0.67 & 0.68 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.89 \\ 0 & 0 & 0 & 0.89 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & .9 & 0.8 & 0 & 0 \\ 0.9 & 1 & 0.7 & 0 & 0 \\ 0.8 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0 & 0.9 & 1 \end{bmatrix}$
Eigenvalues of the true correlation matrix	**2.60, 1.90, 0.31, 0.10, 0.08**	1.98, 1.90, 0.72, 0.30, 0.10	1.98, 1.39, 0.72, 0.61, 0.30	2.11, 1.84, 0.74, 0.16, 0.16	2.47, 1.89, 0.37, 0.16, 0.11	2.60, 1.90, 0.31, 0.10, 0.08

The true correlation matrix for the variables and its eigenvalues are also reported. The ARL_0 values are based on 10,000 simulations in each case.

Table IV. Average in-control run length (ARL_0) for the T^2_A and SPE charts and the average number of retained principal components ($\bar{A}$) for different negative autocorrelation coefficients

	a	b	c	d	e
Autocorrelation matrix	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.9 & 0 \\ 0 & 0 & 0 & 0 & -0.8 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.8 & 0 & 0 & 0 \\ 0 & 0 & -0.7 & 0 & 0 \\ 0 & 0 & 0 & -0.6 & 0 \\ 0 & 0 & 0 & 0 & -0.5 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.9 & 0 & 0 & 0 \\ 0 & 0 & -0.9 & 0 & 0 \\ 0 & 0 & 0 & -0.9 & 0 \\ 0 & 0 & 0 & 0 & -0.9 \end{bmatrix}$
ARL_{0,T^2_A}	486	508	511	489	832
$ARL_{0,SPE}$	525	368	679	613	923
$\bar{A}$	3.000	4.000	3.00	2.997	2.450
True correlation matrix of the variables	$\begin{bmatrix} 1 & 0.39 & 0.35 & 0 & 0 \\ 0.39 & 1 & 0.7 & 0 & 0 \\ 0.35 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0 & 0.9 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.39 & 0.35 & 0 & 0 \\ 0.39 & 1 & 0.7 & 0 & 0 \\ 0.35 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.39 \\ 0 & 0 & 0 & 0.39 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.84 & 0.35 & 0 & 0 \\ 0.84 & 1 & 0.42 & 0 & 0 \\ 0.35 & 0.42 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.84 \\ 0 & 0 & 0 & 0.84 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.84 & 0.67 & 0 & 0 \\ 0.84 & 1 & 0.68 & 0 & 0 \\ 0.67 & 0.68 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.89 \\ 0 & 0 & 0 & 0.89 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.9 & 0.8 & 0 & 0 \\ 0.9 & 1 & 0.7 & 0 & 0 \\ 0.8 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0 & 0.9 & 1 \end{bmatrix}$
Eigenvalues of the true correlation matrix	1.98, 1.90, 0.72, 0.30, 0.10	1.98, 1.39, 0.72, 0.61, 0.30	2.11, 1.84, 0.74, 0.16, 0.16	2.47, 1.89, 0.37, 0.16, 0.11	2.60, 1.90, 0.31, 0.10, 0.08

The ARL_0 values are based on 10,000 simulations in each case.

Table V. Average in-control run length (ARL_0) for the T_A^2 and SPE charts and the average number of retained principal components ($\bar{A}$) for different mixed positive and negative autocorrelation coefficients

	a	b	c	d	e
Autocorrelation matrix	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & -0.9 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.8 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0 & -0.6 & 0 \\ 0 & 0 & 0 & 0 & 0.5 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0 & -0.9 & 0 \\ 0 & 0 & 0 & 0 & -0.8 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.8 & 0 & 0 & 0 \\ 0 & 0 & -0.7 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0.8 \end{bmatrix}$
ARL_{0,T_A^2}	483	482	477	529	537
$ARL_{0,SPE}$	374	NA	487	598	589
$\bar{A}$	4.000	4.984	4.000	3.000	3.000
True correlation matrix of the variables	$\begin{bmatrix} 1 & 0.09 & 0.35 & 0 & 0 \\ 0.09 & 1 & 0.31 & 0 & 0 \\ 0.35 & 0.31 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0 & 0.9 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.09 & 0.35 & 0 & 0 \\ 0.09 & 1 & 0.31 & 0 & 0 \\ 0.35 & 0.31 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.09 \\ 0 & 0 & 0 & 0.09 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.14 & 0.67 & 0 & 0 \\ 0.14 & 1 & 0.19 & 0 & 0 \\ 0.67 & 0.19 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.48 \\ 0 & 0 & 0 & 0.48 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.84 & 0.67 & 0 & 0 \\ 0.84 & 1 & 0.68 & 0 & 0 \\ 0.67 & 0.68 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.84 \\ 0 & 0 & 0 & 0.84 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.84 & 0.67 & 0 & 0 \\ 0.84 & 1 & 0.68 & 0 & 0 \\ 0.67 & 0.68 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.84 \\ 0 & 0 & 0 & 0.84 & 1 \end{bmatrix}$
Eigenvalues of the true correlation matrix	1.90, 1.51, 0.91, 0.58, 0.10	1.51, 1.09, 0.91, 0.91, 0.58	1.75, 1.48, 0.93, 0.52, 0.32	2.47, 1.84, 0.37, 0.16, 0.16	2.47, 1.84, 0.37, 0.16, 0.16

The true correlation matrix for the variables and its eigenvalues are also reported. The ARL_0 values are based on 10,000 simulations in each case. The ARL_0 value for the squared prediction error (SPE) chart in case (b) is not available because the SPE cannot be calculated in the majority of the cases.

capability of PCA and the possible simplification that PCA provides are affected by autocorrelation in the data. In fact, for some of the more extreme cases in Tables III–V, four or five principal components are needed to be retained to maintain the explanatory ability of the PCA model based on time-independent data.

An important conclusion is that when the autocorrelation coefficients are the same or nearly the same within the two blocks of correlated variables, the eigenvalues of the correlation matrix are the same or nearly the same as in the time-independent case. This is similar to the conclusion that we discussed in Section 5. We can therefore conclude that for a diagonal $\mathbf{\Phi}$ matrix when all variables have the same (or nearly the same) autocorrelation coefficients, the descriptive ability of PCA remains intact (or nearly intact).

We can also conclude that the ARL_0 values for the T_A^2 chart increase for many of the autocorrelation cases in Tables III–V; for some cases, the increase is moderate, whereas for others, it is more dramatic. The largest increase in the ARL_0 values for the T_A^2 chart comes with negative autocorrelation in the variables. The behavior of the ARL_0 values for the SPE chart is more varying where the run lengths decrease for some cases and increase for others. However, one needs to keep in mind that the SPE chart basically monitors the remaining unimportant principal components (the residuals) and that the number of remaining principal components varies among the cases. Also, for the SPE chart, the ARL_0 values increase the most for negative autocorrelation in the variables.

Table VI provides some examples when the autocorrelation coefficients are the same or roughly the same within the two blocks of variables. Although the ARL_0 values for the T_A^2 and SPE chart do not dramatically increase because of autocorrelation, there is a larger increase when there is a mix of positive and negative autocorrelation in the two blocks of variables and when all variables have negative autocorrelation.

It should be noted that the large values of ARL_0 may not seem as problematic but they imply an adverse effect on shift detection ability of the control chart.

7.2. *Shift detection capability for the five-variable example*

In this section, we study the average out-of-control run length (ARL_1) in phase II, for the T_A^2 and SPE charts as well as the average number of retained principal components for simulated shift scenarios. The shifts are expressed as multiples of the true standard deviation units of the variables. The impact of autocorrelation on the shift detection capability is studied for a limited number of examples, and the results are organized in two subsections: shift scenarios that for time-independent data are likely to be detected first in the Hotelling T^2 chart and other shift scenarios that are expected to generate faster shift detection in the SPE chart.

7.2.1. *Shifts likely to be detected first in the Hotelling T^2 chart*

Table VII provides results from a one-standard-deviation shift in the same direction for all variables within the first block; that is, $\delta_{x_1} = \delta_{x_2} = \delta_{x_3} = 1$. This scenario indicates an unusual event *in agreement with the PCA model* as the observations after the shift include unusual values in the first three variables although the correlation structure among the variables remains more or less intact. Another unusual event agreeing with the model is investigated in Table VIII, which provides the results from two-standard-deviation shifts in the same direction for the two variables within the second block; that is, $\delta_{x_4} = \delta_{x_5} = 2$.

As expected, the T_A^2 chart is faster to react to the shifts studied in Tables VII and VIII. The shift detection capability is however affected by autocorrelation, and for smaller shifts, the autocorrelation causes an increase in ARL_1 values for the T_A^2 chart. For larger shifts, the autocorrelation to a large extent does not seem to affect the shift detection capability.

7.2.2. *Shifts likely to be detected first in the squared prediction error chart*

Table IX provides results from one-standard-deviation shifts in different directions for the three variables within the first block; here, $\delta_{x_1} = -1$ and $\delta_{x_2} = \delta_{x_3} = 1$. This scenario is an unusual event *in disagreement with the PCA model* because the correlation structure among the variables captured in phase I data is different from that of the observations after the shift. Table X shows results from another shift of this kind, namely, a two-standard-deviation shift in different directions for the two variables within the second block; here, $\delta_{x_4} = 2$ and $\delta_{x_5} = -2$.

Note that for most simulations in case (e) in Tables VII–X, all five principal components were retained to explain at least 90.1% of the overall variation leaving no errors for SPE calculations.

As expected, the SPE chart is the fastest to react to the shifts studied in Tables IXa–c and Xa–c that goes against the correlation structure among the variables in the time-independent case. Although the SPE chart clearly has the lowest ARL_1 values for cases (a–c), there is a peculiar increase in ARL_1 values for the cases in Table IXd and Xd. This delay in shift detection in the SPE chart is caused by the fact that the SPE chart in these cases is based only on the fifth principal component. For these specific cases, it turns out that the signs and magnitudes of the average loadings of the fifth principal component more or less cancel out the simulated shifts in the variables. This cancelation produces a weak signal in the SPE chart yielding much slower shift detection.

Table VI. Average in-control run length (ARL_0) for the T_A^2 and SPE charts and the average number of retained principal components ($\bar{A}$) for different more realistic scenarios of autocorrelation coefficient combinations

	a	b	c	d
Autocorrelation matrix	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.5 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0.85 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 0.65 & 0 \\ 0 & 0 & 0 & 0 & 0.55 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0.85 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & -0.9 & 0 \\ 0 & 0 & 0 & 0 & -0.85 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.85 & 0 & 0 & 0 \\ 0 & 0 & -0.8 & 0 & 0 \\ 0 & 0 & 0 & -0.65 & 0 \\ 0 & 0 & 0 & 0 & -0.55 \end{bmatrix}$
ARL_{0,T_A^2}	482	456	657	550
$ARL_{0,SPE}$	563	451	694	538
$\bar{A}$	2.480	2.833	2.864	2.825
True correlation matrix of the variables	$\begin{bmatrix} 1 & 0.9 & 0.8 & 0 & 0 \\ 0.9 & 1 & 0.7 & 0 & 0 \\ 0.8 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0 & 0.9 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.88 & 0.75 & 0 & 0 \\ 0.88 & 1 & 0.69 & 0 & 0 \\ 0.75 & 0.69 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.89 \\ 0 & 0 & 0 & 0.89 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.88 & 0.75 & 0 & 0 \\ 0.88 & 1 & 0.69 & 0 & 0 \\ 0.75 & 0.69 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.88 \\ 0 & 0 & 0 & 0.88 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.88 & 0.75 & 0 & 0 \\ 0.88 & 1 & 0.69 & 0 & 0 \\ 0.75 & 0.69 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.89 \\ 0 & 0 & 0 & 0.89 & 1 \end{bmatrix}$
Eigenvalues of the true correlation matrix	2.60, 1.90, 0.31, 0.10, 0.08	2.55, 1.89, 0.34, 0.12, 0.11	2.55, 1.88, 0.34, 0.12, 0.12	2.55, 1.89, 0.34, 0.12, 0.11

The ARL_0 values are based on 10,000 simulations in each case.

Table VII. Average out-of-control run length (ARL_1) for the T_A^2 and SPE charts and the average number of retained principal components ($\bar{A}$) for **time-independent data**, different autocorrelation coefficient combinations, and for a one-standard-deviation shift in the first three variables

	a	b	c	d	e
Autocorrelation matrix	$\mathbf{\Phi} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0.85 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0.65 & 0 \\ 0 & 0 & 0 & 0 & 0.55 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.85 & 0 & 0 & 0 \\ 0 & 0 & -0.8 & -0.65 & 0 \\ 0 & 0 & 0 & 0 & -0.55 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & -0.9 \end{bmatrix}$
ARL_{0,T_A^2}	**70**	141	94	91	83
$ARL_{0,SPE}$	**456**	435	501	353	NA
$\bar{A}$	**2.489**	2.831	2.826	4.000	4.981
True correlation matrix of the variables	$\begin{bmatrix} 1 & 0.9 & 0.8 & 0 & 0 \\ 0.9 & 1 & 0.7 & 0 & 0 \\ 0.8 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0 & 0.9 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.88 & 0.75 & 0 & 0 \\ 0.88 & 1 & 0.69 & 0 & 0 \\ 0.75 & 0.69 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.89 \\ 0 & 0 & 0 & 0.89 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.88 & 0.75 & 0 & 0 \\ 0.88 & 1 & 0.69 & 0 & 0 \\ 0.75 & 0.69 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.89 \\ 0 & 0 & 0 & 0.89 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.39 & 0.35 & 0 & 0 \\ 0.39 & 1 & 0.7 & 0 & 0 \\ 0.35 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.39 \\ 0 & 0 & 0 & 0.39 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.09 & 0.35 & 0 & 0 \\ 0.09 & 1 & 0.31 & 0 & 0 \\ 0.35 & 0.31 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.09 \\ 0 & 0 & 0 & 0.09 & 1 \end{bmatrix}$
Eigenvalues of the true correlation matrix	**2.60, 1.90, 0.31, 0.10, 0.08**	2.55, 1.89, 0.34, 0.12, 0.11	2.55, 1.89, 0.34, 0.12, 0.11	1.98, 1.39, 0.72, 0.61, 0.30	1.51, 1.09, 0.91, 0.91, 0.58

The ARL_1 values are based on 10,000 simulations in each case. The ARL_1 value for the squared prediction error (SPE) chart in case (e) is not available because the SPE cannot be calculated for the vast majority of the simulations.

Table VIII. Average out-of-control run length (ARL_1) for the $T^2_{\bar{A}}$ and SPE charts and the average number of retained principal components ($\bar{A}$) for **time-independent data**, different autocorrelation coefficient combinations, and for a two-standard-deviation shift in the last two variables

	a	b	c	d	e
Autocorrelation matrix	$\mathbf{\Phi} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0.85 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 0.65 & 0 \\ 0 & 0 & 0 & 0 & 0.55 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.85 & 0 & 0 & 0 \\ 0 & 0 & -0.8 & 0 & 0 \\ 0 & 0 & 0 & -0.65 & 0 \\ 0 & 0 & 0 & 0 & -0.55 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.85 & 0 & 0 & 0 \\ 0 & 0 & -0.8 & 0 & 0 \\ 0 & 0 & 0 & -0.65 & 0 \\ 0 & 0 & 0 & 0 & -0.55 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & -0.9 \end{bmatrix}$
$ARL_{1,T^2_{\bar{A}}}$	**11**	21	7	17	11
$ARL_{1,SPE}$	**454**	409	468	332	NA
$\bar{A}$	**2.483**	2.833	2.830	4.000	4.981

The ARL_1 values are based on 10,000 simulations in each case. The ARL_1 value for the squared prediction error (SPE) chart in case (e) is not available because the SPE cannot be calculated for the vast majority of the simulations.

Table IX. Average out-of-control run length (ARL_1) for the $T^2_{\bar{A}}$ and SPE charts and the average number of retained principal components ($\bar{A}$) for **time-independent data**, different autocorrelation coefficient combinations, and for a one-standard-deviation shift in different directions in the first three variables

	a	b	c	d	e
Autocorrelation matrix	$\mathbf{\Phi} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0.85 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 0.65 & 0 \\ 0 & 0 & 0 & 0 & 0.55 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.85 & 0 & 0 & 0 \\ 0 & 0 & -0.8 & 0 & 0 \\ 0 & 0 & 0 & -0.65 & 0 \\ 0 & 0 & 0 & 0 & -0.55 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & -0.9 \end{bmatrix}$
$ARL_{1,T^2_{\bar{A}}}$	**233**	214	188	37	42
$ARL_{1,SPE}$	**2**	11	1	303	NA
$\bar{A}$	**2.497**	2.825	2.829	4.000	4.981
True correlation matrix of the variables	$\begin{bmatrix} 1 & 0.9 & 0.8 & 0 & 0 \\ 0.9 & 1 & 0.7 & 0 & 0 \\ 0.8 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0 & 0.9 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.88 & 0.75 & 0 & 0 \\ 0.88 & 1 & 0.69 & 0 & 0 \\ 0.75 & 0.69 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.89 \\ 0 & 0 & 0 & 0.89 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.88 & 0.75 & 0 & 0 \\ 0.88 & 1 & 0.69 & 0 & 0 \\ 0.75 & 0.69 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.89 \\ 0 & 0 & 0 & 0.89 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.39 & 0.35 & 0 & 0 \\ 0.39 & 1 & 0.7 & 0 & 0 \\ 0.35 & 0.7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.39 \\ 0 & 0 & 0 & 0.39 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.09 & 0.35 & 0 & 0 \\ 0.09 & 1 & 0.31 & 0 & 0 \\ 0.35 & 0.31 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.09 \\ 0 & 0 & 0 & 0.09 & 1 \end{bmatrix}$
Eigenvalues of the true correlation matrix	**2.60, 1.90, 0.31, 0.10, 0.08**	2.55, 1.89, 0.34, 0.12, 0.11	2.55, 1.89, 0.34, 0.12, 0.11	1.98, 1.39, 0.72, 0.61, 0.30	1.51, 1.09, 0.91, 0.91, 0.58

The ARL_1 values are based on 10,000 simulations in each case. The ARL_1 value for the squared prediction error (SPE) chart in case (e) is not available because the SPE cannot be calculated in the majority of the cases.

Table X. Average out-of-control run length (ARL_1) for the $T^2_{\bar{A}}$ and SPE charts and the average number of retained principal components ($\bar{A}$) for **time-independent data**, different autocorrelation structures, and for a two-standard-deviation shift in different directions in the last two variables

	a	b	c	d	e
Autocorrelation matrix	$\mathbf{\Phi} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0.85 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 0.65 & 0 \\ 0 & 0 & 0 & 0 & 0.55 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} -0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.85 & 0 & 0 & 0 \\ 0 & 0 & -0.8 & 0 & 0 \\ 0 & 0 & 0 & -0.65 & 0 \\ 0 & 0 & 0 & 0 & -0.55 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 \\ 0 & -0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & -0.9 \end{bmatrix}$
$ARL_{1,T^2_{\bar{A}}}$	**405**	413	488	7	8
$ARL_{1,SPE}$	**1**	2	1	273	NA
$\bar{A}$	**2.487**	2.834	2.828	4.000	4.982

The ARL_1 values are based on 10,000 simulations in each case. The ARL_1 value for the squared prediction error (SPE) chart in case (e) is not available because the SPE cannot be calculated in the majority of the cases.

8. Conclusions and discussion

In this article, we set out to investigate and illustrate the impact of autocorrelation on the descriptive ability of PCA and on the monitoring performance using PCA-based SPC. We intentionally show the impact of naïvely proceeding with PCA by ignoring the autocorrelation in the data. In the simulations, we introduce autocorrelation in the variables through a VAR(1) model. Conceptually, this means that we let the static relations among the variables be represented by the covariance matrix of the innovations (errors) in the VAR(1) model and the dynamic relations be represented by the autocorrelation coefficient matrix Φ of the VAR(1) model. In reality, the analyst will typically not be aware of the difference between static relations and dynamic relations but will simply model the sampled data – including both the static and dynamic relations' effect on the correlation structure of the variables. Our argumentation from an SPC point of view builds upon the assumption that the 'true' latent structure is given by the static relations among the variables. We believe that in that regard, there is a need to truly define what is meant by a latent structure/variable when data are autocorrelated. But this is beyond the scope of this paper.

The consequence of the naïve approach is that descriptive ability of PCA is affected by autocorrelation and the analyst may find additional important latent variables. The number of principal components that need to be retained to explain a given fraction of the variability in the data may increase as a result of autocorrelation. We show that the impact on the descriptive ability can vary from non-existent when Φ is diagonal and all variables have the same autocorrelation coefficients to dramatic when the variables have large autocorrelation coefficients with opposite signs. This means that the descriptive ability of PCA may be seriously affected by a situation where the variables have different degrees of autocorrelation, especially with different signs.

In this article, we also confirm and illustrate how autocorrelation can affect the naïve use of PCA for the inferential purpose of SPC. The false alarm rate and the shift detection capability of PCA-based SPC are affected by autocorrelation and can cause delayed shift detection. The impact on the false alarm rate and the shift detection capability of PCA-based SPC is largest for negative autocorrelation, but for shifts of larger magnitudes, the impact of autocorrelation seems to be rather small.

This article is not focused on solving the problem with autocorrelated data, but as mentioned earlier, a potential solution to reduce the impact of autocorrelation on PCA-based SPC is to adjust the control limits of the Hotelling T^2 chart for the A first principal components and the SPE chart. However, this would require an adjustment to each specific case, which is a time-consuming solution. Another time-consuming solution is to use a residuals approach where, for example, a multivariate time series model is fitted to the data and PCA is performed on the time-independent residuals. DPCA with decorrelated residuals[25] seems to be one of the most recently proposed solutions. However, we argue that expanding the data matrix with additional time-lagged variables increases model complexity, may have a negative impact on model interpretation, and makes it somewhat more difficult to identify causes of the fault. In our future research, we aim to compare different approaches in multivariate SPC to remedy the problems created by autocorrelation in the process data.

Acknowledgements

We gratefully acknowledge the financial support from the Swedish Research Council under grant number 340-2013-5108.

References

1. Jolliffe IT. *Principal Component Analysis* (2nd edn). Springer-Verlag: New York, NY, 2002.
2. MacGregor JF. Using on-line process data to improve quality: challenges for statisticians. *International Statistical Review* 1997; **65**(3):309–323. doi:10.1111/j.1751-5823.1997.tb00311.x.
3. Bersimis S, Psarakis S, Panaretos J. Multivariate statistical process control charts: an overview. *Quality and Reliability Engineering International* 2007; **23**(5):517–543. doi:10.1002/qre.829.
4. Kourti T. Application of latent variable methods to process control and multivariate statistical process control in industry. *International Journal of Adaptive Control and Signal Processing* 2005; **19**:213–246. doi:10.1002/acs.859.
5. Qin SJ. Statistical process monitoring: basics and beyond. *Journal of Chemometrics* 2003; **17**:480–502. doi:10.1002/cem.800.
6. Montgomery DC, Jennings CL, Kulahci M. *Introduction to Time Series Analysis and Forecasting* (2nd edn). Wiley: Hoboken, NJ, 2015.
7. Bisgaard S, Kulahci M. *Time Series Analysis and Forecasting by Example*. Wiley: Hoboken, NJ, 2011.
8. Johnson RA, Bagshaw M. The effect of serial correlation on the performance of CUSUM tests. *Technometrics* 1974; **16**(1):103–112.
9. Vasilopoulos AV, Stamboulis AP. Modification of control chart limits in the presence of data correlation. *Journal of Quality Technology* 1978; **10**(1):20–30.
10. Alwan LC, Roberts HV. Time-series modeling for statistical process control. *Journal of Business & Economic Statistics* 1988; **6**(1):87–95.
11. Montgomery DC, Mastrangelo CM. Some statistical process control methods for autocorrelated data. *Journal of Quality Technology* 1991; **23**(3):179–193.
12. Wardell DG, Moskowitz H, Plante RD. Run-length distributions of special-cause control charts for correlated processes. *Technometrics* 1994; **36**(1):3–17.
13. Zhang NF. Detection capability of residual control chart for stationary process data. *Journal of Applied Statistics* 1997; **24**(4):475–492.
14. Vanhatalo E, Kulahci M. The effect of autocorrelation on the Hotelling T^2 control chart. *Quality and Reliability Engineering International* 2014. Published online ahead of print. DOI: 10.1002/qre.1717
15. Kalgonda AA, Kulkarni SR. Multivariate quality control chart for autocorrelated processes. *Journal of Applied Statistics* 2004; **31**(3):317–327. doi:10.1080/0266476042000184000.
16. Pan X, Jarrett JE. Applying state space to SPC: monitoring multivariate time series. *Journal of Applied Statistics* 2004; **31**(4):397–418. doi:10.1080/02664860410001681701.

17. Pan X, Jarrett JE. Monitoring variability and analyzing multivariate autocorrelated processes. *Journal of Applied Statistics* 2007; **34**(4):459–469. doi:10.1080/02664760701231849.
18. Pan X, Jarrett JE. Using vector autoregressive residuals to monitor multivariate processes in the presence of serial correlation. *International Journal of Production Economics* 2007; **106**(1):204–216. doi:10.1016/j.ijpe.2006.07.002.
19. Pan X, Jarrett JE. Why and how to use vector autoregressive models for quality control: the guideline and procedures. *Quality & Quantity* 2012; **46**:935–948. doi:10.1007/s11135-011-9437-x.
20. Snoussi A. SPC for short-run multivariate autocorrelated processes. *Journal of Applied Statistics* 2011; **38**(10):2303–2312. doi:10.1080/02664763.2010.547566.
21. Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis* (5th edn). Prentice Hall: Upper Saddle River, NJ, 2002.
22. Vanhatalo E. Multivariate process monitoring of an experimental blast furnace. *Quality and Reliability Engineering International* 2010; **26**(5):495–508. doi:10.1002/qre.1070.
23. Ku W, Storer RH, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 1995; **30**:179–196.
24. Kourti T, MacGregor JF. Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics and Intelligent Laboratory Systems* 1995; **28**:3–21. doi:10.1016/0169-7439(95)80036-9.
25. Rato TJ, Reis MS. Advantage of using decorrelated residuals in dynamic principal component analysis for monitoring large-scale systems. *Industrial & Engineering Chemistry Research* 2013; **52**:13685–13698. doi:10.1021/ie3035306.
26. Jackson JE. *A User's Guide to Principal Components*. Wiley: Hoboken, NJ, 2003.
27. Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 1987; **2**:37–52.
28. Montgomery DC. *Statistical Quality Control: A Modern Introduction* (7th edn). Wiley: Hoboken, NJ, 2013.
29. Reinsel GC. *Elements of Multivariate Time Series Analysis* (2nd edn). Springer-Verlag: New York, NY, 1993.
30. Ferrer A. Latent structures-based multivariate statistical process control: a paradigm shift. (With discussions by William H. Woodall and Dennis K. J. Lin). *Quality Engineering* 2014; **26**(1):72–91. doi:10.1080/08982112.2013.846093.
31. Jackson JE, Mudholkar GS. Control procedures for residuals associated with principal component analysis. *Technometrics* 1979; **21**(3):341–349.

Authors' biographies

Erik Vanhatalo is an assistant professor and senior lecturer of quality technology and management at Luleå University of Technology (LTU). He holds an MSc degree in industrial and management engineering from LTU and a PhD in quality technology and management from LTU. His current research is focused on the use of statistical process control, experimental design, time series analysis, and multivariate statistical methods, especially in process industry. He is a member of the European Network for Business and Industrial Statistics (ENBIS).

Murat Kulahci is an associate professor in the Department of Applied Mathematics and Computer Science at the Technical University of Denmark and in the Department of Business Administration, Technology and Social Sciences at Luleå University of Technology in Sweden. His research focuses on design of physical and computer experiments, statistical process control, time series analysis and forecasting, and financial engineering. He has presented his work in international conferences and published over 50 articles in archival journals. He is the co-author of two books on time series analysis and forecasting.