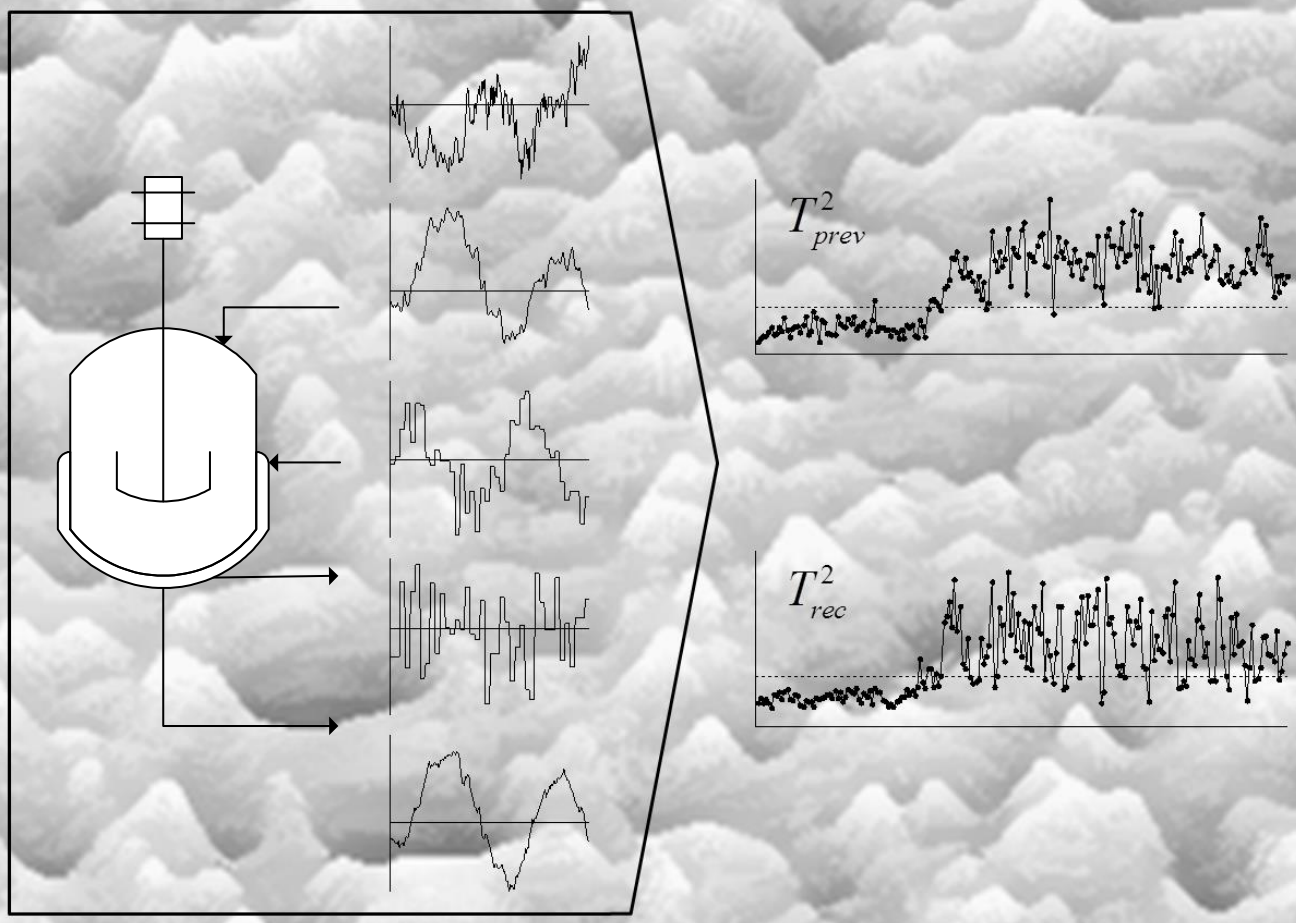


Tiago Miguel Janeiro Rato

Development of integrated data-driven frameworks for the monitoring, analysis and control of complex systems

Doctoral Thesis in Chemical Engineering with specialization in Chemical Processes, under the supervision of Professor Doctor Marco Paulo Seabra dos Reis, presented to the Department of Chemical Engineering, Faculty of Sciences and Technology of the University of Coimbra

Coimbra, January 2014

· U C ·

UNIVERSIDADE DE COIMBRA

• U C •

FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Development of integrated data-driven frameworks for the monitoring, analysis and control of complex systems

Tiago Miguel Janeiro Rato

(Master in Chemical Engineering)

Doctoral Thesis in Chemical Engineering with specialization in Chemical Processes, under the supervision of Professor Doctor Marco Paulo Seabra dos Reis, presented to the Department of Chemical Engineering, Faculty of Sciences and Technology of the University of Coimbra

Coimbra, January 2014

Acknowledgements

The work presented in this Ph.D. thesis would not be possible without the contribution of some people and entities to which I want to express my gratitude:

In the first place, I want to express my deepest gratitude to my supervisor and mentor Professor Doctor Marco Paulo Seabra dos Reis for his constant support, guidance and constructive suggestions on how to improve the work developed during all these years.

I am also thankful to Professor Bart De Ketelaere, and Eric Schmitt from the Katholieke Universiteit Leuven to whom I had the pleasure to collaborate in a joint work conducted during the last phase of this thesis.

I thank to all my colleagues and friends who accompanied me on this long journey. Especially to Pedro Almeida, who always supported me with his friendship.

I am also very thankful to my family, in particular my parents, Rogério and Glória, who always gave me their full support and the strength I needed in all the phases of my life.

At last, I would like to acknowledge the financial support of the Portuguese Foundation for Science and Technology (FCT) under PhD grant SFRH/BD/65794/2009. I also acknowledge the financial support in Chapter 13 through project PTDC/EQU-ESI/108374/2008 co-financed by the Portuguese FCT and European Union's FEDER through "Eixo I do Programa Operacional Factores de Competitividade (POFC)" of QREN (with ref. FCOMP-01-0124-FEDER-010397).

To my parents

Abstract

Statistical process control (SPC) is a critical activity in all industrial process and since the introduction of the Shewhart control chart many methodologies have been developed to promptly detect process upsets as soon as possible, while avoiding unnecessary corrective actions that only introduce more variation into the system. Furthermore, current industrial data present a set of distinctive and challenging features that raise important difficulties in the implementation of efficient SPC approaches. In particular, they present megavariate, multiscale and dynamical features that need proper treatment. These three data characteristics are the main focus of this thesis.

The base approach to handle both megavariate and dynamic features is dynamic principal component analysis (DPCA), which was the ability to simultaneously handle the cross- and auto-correlative behavior of processes. The concurrent modeling of such dependencies, allows for a more rigorous description of the normal behavior of processes, setting the ground for the development of improved SPC methodologies that are able to robustly detect finer deviations from normal operation conditions. A key point in the application of DPCA is the definition of its structure, namely the selection of the number of time-shifted replicates for each variable to include and the number of components to retain in the final model. To address the former aspect, two new lag selection methods are proposed in this thesis. The first method estimates a single lagged structure for all variables, while the second one refines this procedure by selecting the number of lags to be used for each individual variable. The application of these methods lead to a more rigorous estimation of the process lagged structure, and thus when implemented with SPC schemes that rely on a DPCA framework, significant improvements are observed. Still, the traditional DPCA approach leads to monitoring statistics with considerable levels of autocorrelation, a feature that seriously hinders its dissemination in practice. To handle this issue, a new set of multivariate statistics based in DPCA and missing data imputation methods were developed. The obtained DPCA with decorrelated residuals (DPCA-DR) methodology presents low autocorrelation levels, and therefore is better positioned to implement SPC over complex systems, in a more consistent and stable way. The performance of the proposed DPCA-DR was compared with a variety of current monitoring schemes for large scale systems, under different dynamic scenarios and for different types of process upsets and fault magnitudes. The results obtained clearly

indicate that the statistics based on DPCA-DR consistently present superior performances regarding detection ability and decorrelation power. For instance, in the Tennessee Eastman process, the proposed DPCA-DR statistics had the highest detection rates on 19 of the 21 faults, are statistically superior to their PCA and DPCA counterparts and present low levels of autocorrelation, which simplifies their implementation and improves their reliability.

Another subject treated in this thesis is the monitoring of the process correlation structure. On this regard, the analysis of the literature shows that most of the multivariate SPC methods developed so far are essentially focused on detecting changes in the process mean. Yet, the monitoring of the process multivariate dispersion is also a relevant issue in SPC, since a process failure may not manifest itself so notoriously as a deviation from the nominal mean values, especially due to the action of control systems struggling to maintain key process variables close to their target values.

The monitoring of multivariate process dispersion is usually carried out through monitoring statistics based on the generalized variance or likelihood ratio tests, which are strictly based on marginal correlations. This global measure of association is not sensitive to the inner association between variables and therefore is unable, by design, to efficiently discern changes in the local correlation structure. To access and use information on the local association structure, in this thesis several monitoring statistics and sensitivity enhancing transformations (SET) are proposed that are based on partial correlations. The SET was found to be a key aspect in the development of such monitoring schemes which, for maximum effectiveness, makes use of information regarding the causal network underlying the process in order to construct a set of uncorrelated transformed variables around which the detection of changes in correlation is maximized. The results obtained in the comparison study involving the current methodologies to monitor the correlation structure, in both off-line and on-line cases, show that the proposed methods based on partial correlations are able to efficiently detect changes in the process structure and presented higher sensitivity than their marginal counterparts.

Keywords: lag selection; principal component analysis; multivariate statistical process control; mathematical modeling; multivariate dynamical processes; partial correlation; marginal correlation; variable transformation; causal network.

Resumo

O controlo estatístico de processos (SPC) é uma atividade crítica em qualquer unidade industrial. Desde a introdução da primeira carta de controlo, muitas outras metodologias têm vindo a ser desenvolvidas com o intuito de detetar rapidamente diferentes tipos de falhas processuais ao mesmo tempo que evitam o recurso a intervenções desnecessárias que apenas introduzem mais variação nos processos. Além disso, os processos industriais atuais apresentam características específicas que dificultam a análise dos dados recolhidos, nomeadamente a sua escala elevada, natureza multiescala e presença de dinâmica em todas as variáveis. Estas características requerem a implementação de novas metodologias de SPC, as quais serão o principal foco desta tese.

A abordagem base para lidar com estes fenómenos apoia-se no conceito da análise dos componentes principais, nomeadamente a sua versão dinâmica (DPCA), uma vez que esta é capaz de descrever a correlação cruzada e autocorrelação dos dados em simultâneo. Deste modo, DPCA permite uma descrição mais rigorosa dos processos e como tal, o desenvolvimento de metodologias SPC aptas para a deteção de desvios subtis às condições normais de operação. Um dos passos mais relevantes na aplicação de DPCA recai na correta definição da sua estrutura dinâmica, mais especificamente a seleção do número de variáveis desfasadas necessárias para descrever o processo corretamente. Neste sentido, dois métodos para a seleção do número de desfasamentos são propostos nesta tese. O primeiro método estima um número de desfasamentos igual para todas as variáveis, enquanto o segundo, refina este resultado ao estimar o número de desfasamentos mais apropriado para cada variável. A aplicação direta destes métodos em DPCA, permite por si só melhorar a capacidade de deteção das técnicas existentes devido a uma melhor descrição da variabilidade processual. No entanto, como DPCA tradicional não consegue lidar corretamente com a dinâmica dos dados, uma vez que as estatísticas de monitorização baseadas em DPCA tendem a apresentar autocorrelação, uma nova metodologia foi desenvolvida para mitigar este efeito. O novo método, designado por DPCA com resíduos descorrelacionados (DPCA-DR), baseia-se na combinação de DPCA com técnicas de estimação de dados em falta. DPCA-DR apresenta baixos níveis de autocorrelação nas suas estatísticas sem por em causa a sua capacidade de deteção de falhas. Como tal, DPCA-DR é uma mais-valia para a monitorização de sistemas complexos, apresentando um melhor desempenho quando comparado com as técnicas

atuais para lidar com sistemas multivariados. Por exemplo, no conhecido caso de estudo do processo Tennessee Eastman, as estatísticas baseadas em DPCA-DR apresentam as melhores taxas de deteção em 19 das 21 falhas consideradas, sendo que estas são também estatisticamente superiores às taxas de deteção obtidas com PCA e DPCA tradicionais. Além disso, as estatísticas baseadas em DPCA-DR apresentam baixos níveis de autocorrelação, o que simplifica a sua implementação e aumenta a sua fiabilidade.

Outro tema abordados nesta tese consiste na monitorização da estrutura correlacionada dos processos. Neste aspeto, da análise da literatura verificou-se que a maioria dos métodos SPC desenvolvidos até então focam-se essencialmente na deteção de desvios na média processual. No entanto, a monitorização da dispersão multivariada é igualmente importante pois uma falha processual pode não se manifestar tão notoriamente em desvios no valor nominal das variáveis, já que estas estão normalmente sob o efeito de sistemas de controlo que as mantêm próximas do seu valor pretendido.

Em processos multivariados, a monitorização da dispersão é usualmente feita por recurso a estatísticas baseadas na variância generalizada ou em testes de verosimilhança que apenas têm em conta a correlação marginal das variáveis. Esta medida global da associação entre as variáveis não é sensível à relação direta entre as mesmas e, como tal, a correlação marginal não é uma medida adequada para a deteção de desvios na associação local entre as variáveis. A fim de explorar este tipo de informação, abordagens baseadas na correlação parcial entre as variáveis foram consideradas nesta tese. Dos métodos desenvolvidos, destaca-se o pré-processamento das variáveis por via de uma transformação linear com vista a descorrelacionar as variáveis. Esta transformação baseia-se na rede causal inerente ao processo, incorporando a sua estrutura de modo a obter um novo conjunto de variáveis descorrelacionadas e em torno das quais a deteção de desvios na correlação é maximizada. Os resultados obtidos durante a comparação destes novos métodos com outros alternativos baseados na correlação marginal, demostram que o uso de correlações parciais, aliadas a uma melhor descrição do processo, levam a uma maior capacidade de deteção de desvios na estrutura dos processos, tendo também um elevado potencial para diagnosticar a origem da falha.

Palavras-chave: seleção de desfasamentos; análise dos componentes principais; controlo estatístico multivariado; modelação matemática; processos dinâmicos multivariados; correlação parcial; correlação marginal; transformação de variáveis; redes causais.

Symbols and Abbreviations

Symbols

$\mathbf{B}$	Matrix of regression coefficient
J_{max}	Decomposition depth in the wavelet decomposition
l	Number of lags
m	Number of variables
n	Number of observations
$\mathbf{P}$	Matrix of loadings
p	Number of latent variables
r	Sample correlation
$\mathbf{r}_0$	Vector of marginal correlation coefficients
$\mathbf{r}_1$	Vector of 1st order partial correlation coefficients.
$\mathbf{S}$	Sample covariance matrix
s	Sample standard deviation
$\mathbf{T}$	Matrix of latent variables
$\mathbf{v}$	Vector of variables' variance
$\mathbf{x}$	Vector of observation
$\mathbf{X}$	Matrix of observations
$\mathbf{\Lambda}$	Matrix of eigenvalues
$\boldsymbol{\mu}$	Population mean vector
v	Degrees of freedom
ρ	Population correlation
σ	Population standard deviation
$\mathbf{\Sigma}$	Population covariance matrix
$\tilde{\mathbf{X}}$	Extended data matrix
$\odot$	Hadamard product
$\otimes$	Kronecker product

Abbreviations

AR	Autoregressive
ARIMA	Autoregressive integrated moving average
ARL	Average run length
ARMA	Autoregressive moving average
CL	Center line
CSTR	Continuous stirred-tank reactor
CUSUM	Cumulative sum
CVA	Canonical variate analysis
DPCA	Dynamic principal component analysis
DPCA-DR	Dynamic principal components analysis with decorrelated residuals
DPLS	Dynamic partial least squares
EPC	Engineering process control
EWMA	Exponentially weighted moving average
i.i.d.	Independent and identically distributed
KSV	Key singular value
KSVR	Key singular value ratio
LCL	Lower control limit
LRT	Likelihood ratio test
MS-DPCA-DR	Multiscale dynamic principal components analysis with decorrelated residuals
MSE	Mean squared error
MSPC	Multivariate statistical process control
MS-PCA	Multiscale principal component analysis
NOC	Normal operation conditions
OLS	Ordinary least squares
PC	Principal component
PCA	Principal component analysis
PLS	Partial least squares
SET	Sensitivity enhancing transformation
SI	System identification
SNR	Signal-to-noise ratio
SPC	Statistical process control

SPE	Squared prediction error
UCL	Upper control limit
VAR	Vector autoregressive
VARIMA	Vector autoregressive integrated moving average
VARMA	Vector autoregressive moving average
VMA	Vector moving average

Table of Contents

List of Figures

List of Tables

Part I – Introduction and Goals

1 Introduction

The main topics addressed in this thesis are introduced in this chapter. It is divided in four sections where the contextualization of the research conducted in the scope of this thesis is provided as well as the underlying motivation and basic concepts. In the first section, the general research setting is described. Then, in the next section, the main goals are defined and a brief description of the approaches followed in order to attain such goals is provided. Afterwards, the main contributions are summarized as well as the critical concepts employed in their development and associated advantages regarding current approaches. Finally, an overview of the thesis' structure is provided.

1.1 Scope and Motivation

The analysis and operation of industrial and biological systems faces nowadays significant challenges due to the massive quantities of information they generate. Moreover, the data collected from these systems is usually composed by highly correlated variables with different time/length resolutions, as a result of the complexity of all the underlying phenomena going on. Thus, there is a need for integrated data processing and analysis methodologies capable to extract and use the information that is in fact relevant for the proper implementation of subsequent tasks, such as process monitoring, optimization and control. Various methodologies for dealing with particular aspects of industrial data were already developed and presented in the literature. For instance, the highly interactive and correlated nature of data can be explored with multivariate and multi-way methods, such as principal component analysis (PCA) [1], partial least squares (PLS) [2], independent component analysis (ICA) [3], Tucker3 and PARAFAC [4]. Certain multiresolution and multiscale dynamic features can also be properly described, namely using data-driven multiscale frameworks [5-9], as well as issues such as measurement noise/uncertainty can also be taken into account [10, 11]. However, there are some key elements that still remain to be properly integrated in data-driven systems analysis frameworks, even considering the state-of-the-art analysis methodologies, and especially in what concerns to industrial process systems: the explicit incorporation of

systems network topology along with information about the causal directionality between connected elements. With such descriptions available, approaches that are more consistent with the true nature of systems can be developed.

In this context, the general scope of this thesis regards the development of new frameworks for handling the complex nature of systems and of the data they generate, in tasks such as process monitoring, diagnosis and analysis. These frameworks will combine recent developments on multiresolution theory, network analysis and multivariate statistics/data-mining, in order to flexibly handle the several intricate features of data generated in complex industrial processes and biosystems.

1.2 Goals

The present research aimed at developing new methodologies that exploit the information contained in the data collected from complex industrial systems and biosystems as well as from some *a priori* knowledge, in order to:

(i) improve the quality of the analysis and the effectiveness of the extraction of useful knowledge regarding the underlying physical-chemical phenomena, speeding up the learning process and thus providing the basis for the efficient development of innovative solutions;

(ii) improve the tasks of process monitoring, diagnosis and control, given the extended access to more information and more precise descriptions of the reality trough models inferred from data.

In the context of these general goals, new integrated frameworks that enable the prompt detection of process upsets will be developed, as well as analysis and diagnosis tools able to suggest lines of action to plant personnel, towards their effective elimination or mitigation. To do so, the developed frameworks will combine dynamical modeling methods, multiresolution and multivariate methodologies along with several other approaches such as network analysis and inference methodologies based on partial correlation, in order to accommodate the main complexity features present in data from industrial systems and biosystems.

1.3 Contributions

In accordance with the general goals described above, the following developments can be considered as the main contributions from this thesis:

(i) Two new methodologies for defining the structure of dynamic PCA models (DPCA). They consist of lag selection methods that determinate the number of time-shifted replicates to be included in DPCA. The proposed methods are based on a succession of singular value decomposition problems and lead to better estimates of the true lagged structure of the underlying processes. Moreover, contrary to current approaches, one of the proposed methods selects a different number of lags for each variable, which ultimately leads to better descriptions of the process under analysis;

(ii) A new methodology called dynamic principal component analysis with decorrelated residuals (DPCA-DR), that finally successfully addresses one of the main limitations of DPCA, namely its inability to remove the autocorrelation from the monitoring statistics. This methodology is based on the combined use of DPCA, to explain the process dynamics and correlation structure, and missing data techniques, to intrinsically perform one-step-ahead predictions without the need of fitting a multivariate time series model. When DPCA-DR is applied to process monitoring of complex dynamic systems a set of monitoring statistics with low autocorrelation levels and high fault detection performances are obtained;

(iii) New sensitivity enhancing transformations (SET) for process monitoring are proposed and discussed. These transformations are based on the process causal structure and are capable of increasing the sensitivity to abnormal structural changes. These SET incorporate the inner relationship between the variables in order to construct a new set of uncorrelated variables around which the deviations in the correlation coefficients are maximized, so that even small changes can be easily detected;

(iv) Off-line and on-line monitoring statistics based on local measures of association, namely partial correlations, are proposed for the first time. They can exploit the benefits from the use of the SET and lead to higher fault detection performances than current approaches (which are mostly invariant to

SET). The task of fault diagnosis also benefits from the use of partial correlations and SET's, where the new methods show greater capability to identify the correct root causes of process upsets when compared with the current marginal-based counterparts;

(v) An equivalence relationship is proposed for the off-line and on-line methodologies for monitoring the process multivariate dispersion using partial correlations. Its impact goes beyond the methods proposed in this thesis;

(vi) New multiscale monitoring procedures for detecting changes in the process location and multivariate dispersion. In the case of location monitoring, the use of DPCA-DR in combination with the wavelet transform showed to be capable to isolate the scales related with the fault and consequently leading to higher detection rates than the current single-scale and multiscale PCA approaches. As for the monitoring of the multivariate dispersion, the major contribution is given by the SET, which proved to allow for a suitable monitoring of multiscale systems even when single-scale monitoring schemes are used. In some sense, the SET transformations already incorporate multiscale modelling features.

Most of the subjects covered in the above topics were already published or are in the final stages of preparation at the time of submission of this thesis, as summarized in Table 1.1. Furthermore, the algorithms related with more relevant contributions are available in the form of Matlab functions in the digital supplementary material that complements this thesis. A summary of such functions, as well as their relation with the thesis contributions, is provided in Table 1.2.

Table 1.1 Summary of the published works related with the topics addressed in this thesis.

Chapter	Contribution	Reference
5	(i)	Rato, T.J. and Reis, M.S., *Defining the structure of DPCA models and its impact on process monitoring and prediction activities.* Chemometrics and Intelligent Laboratory Systems, 2013. **125**(15): p. 74-86.
6	(ii)	Rato, T.J. and Reis, M.S., *Statistical Process Control of Multivariate Systems with Autocorrelation*, in *Computer Aided Chemical Engineering*, M.C.G. E.N. Pistikopoulos and A.C. Kokossis, Editors. 2011, Elsevier. p. 497-501.
6	(ii)	Rato, T.J. and Reis, M.S., *Advantage of Using Decorrelated Residuals in Dynamic Principal Component Analysis for Monitoring Large-Scale Systems.* Industrial & Engineering Chemistry Research, 2013. **52**(38): p. 13685-13698.
6	(ii)	Rato, T.J. and Reis, M.S., *Fault detection in the Tennessee Eastman benchmark process using dynamic principal components analysis based on decorrelated residuals (DPCA-DR).* Chemometrics and Intelligent Laboratory Systems, 2013. **125**(15): p. 101-108.
9, 10	(iii), (iv)	Rato, T.J. and Reis, M.S., *Non-causal data-driven monitoring of the process correlation structure: a comparison study with new methods.* **Submitted**.
9, 10	(iii), (iv)	Rato, T.J. and Reis, M.S., *Sensitivity Enhancing Transformations for Monitoring the Process Structure.* **Submitted**.

Table 1.2 Summary of the Matlab functions developed in this thesis.

Chapter	Contribution	Function	Description
5	(i)	*detlag_method1*	Method 1 to define the lagged structure of processes: maximum number of lags in DPCA
5	(i)	*detlag_method2*	Method 2 to define the lagged structure of processes: number of lags for each variable
6	(ii)	*stDPCADR_model*	Constructs a DPCA-DR model based on a NOC data set
6	(ii)	*stDPCADR*	Computes the proposed DPCA-DR monitoring statistics based on the model given by *stDPCADR_model*
9	(iii)	*network_edge*	Network reconstruction: identification of the undirected edges between the variables
9	(iii)	*network_direct1*	Network reconstruction: define the directionality of the causal network of stationary systems
9	(iii)	*network_direct2*	Network reconstruction: define the directionality of the causal network of dynamic systems
9	(iii)	*SET_Net*	Sensitivity enhancing transformation based on the process causal network

1.4 Thesis Overview

The present thesis is organized in five parts representing different research scopes, as presented in Figure 1.1.

The Part I comprises an introduction and overview of the main themes considered in this thesis as well as the description of the motivation, goals and contributions of the work carried out.

In Part II a review of the State of the Art regarding multivariate and megavariate statistics process control (SPC) is provided for the cases of monitoring the process location (Chapter 3) and multivariate dispersion (Chapter 4). In this part, the fundamental modeling methodologies underlying both the current and proposed methodologies, are also briefly described (Chapter 2).

Part III regards the monitoring of the process location by means of megavariate monitoring schemes based on latent variables models. The developed lag selection methodologies (Chapter 5) and DPCA-DR technique (Chapter 6) are described and assessed in this part. The DPCA-DR methodology was compared in a study encompassing twenty two monitoring schemes designed to deal with the megavariate and dynamic characteristics of the process. From this study it was found that the DPCA-DR based monitoring statistics were the most consistent and reliable with low levels of autocorrelation and higher fault detection performances. The multiscale version of DPCA-DR is also introduced in this part of the thesis (Chapter 7).

In Part IV the problem of monitoring the process structure is addressed. This study is first conducted for the case of off-line monitoring, based on non-overlapping moving windows (Chapter 10), from where it was observed that a set of proposed monitoring statistics based on the maximum deviation, in absolute value, of the partial correlation coefficients (*RMAX*) and variance (*VnMAX*), were the ones leading to the best monitoring performances. Afterwards, their application was extended to on-line process monitoring, based on individual observations (Chapter 11). An important factor in these proposed monitoring schemes for the process correlation structure is the use of sensitivity enhancing transformations (Chapter 9). These variable transformations proved to be crucial for the fault detection capability of the proposed monitoring statistics. The SET's

are able to provide good descriptions of the process structure, extracted with the aid of partial correlations.

Finally, in Part V, the thesis main conclusions are summarized (Chapter 14) and a set of application fields to be explored in future work are described (Chapter 15).

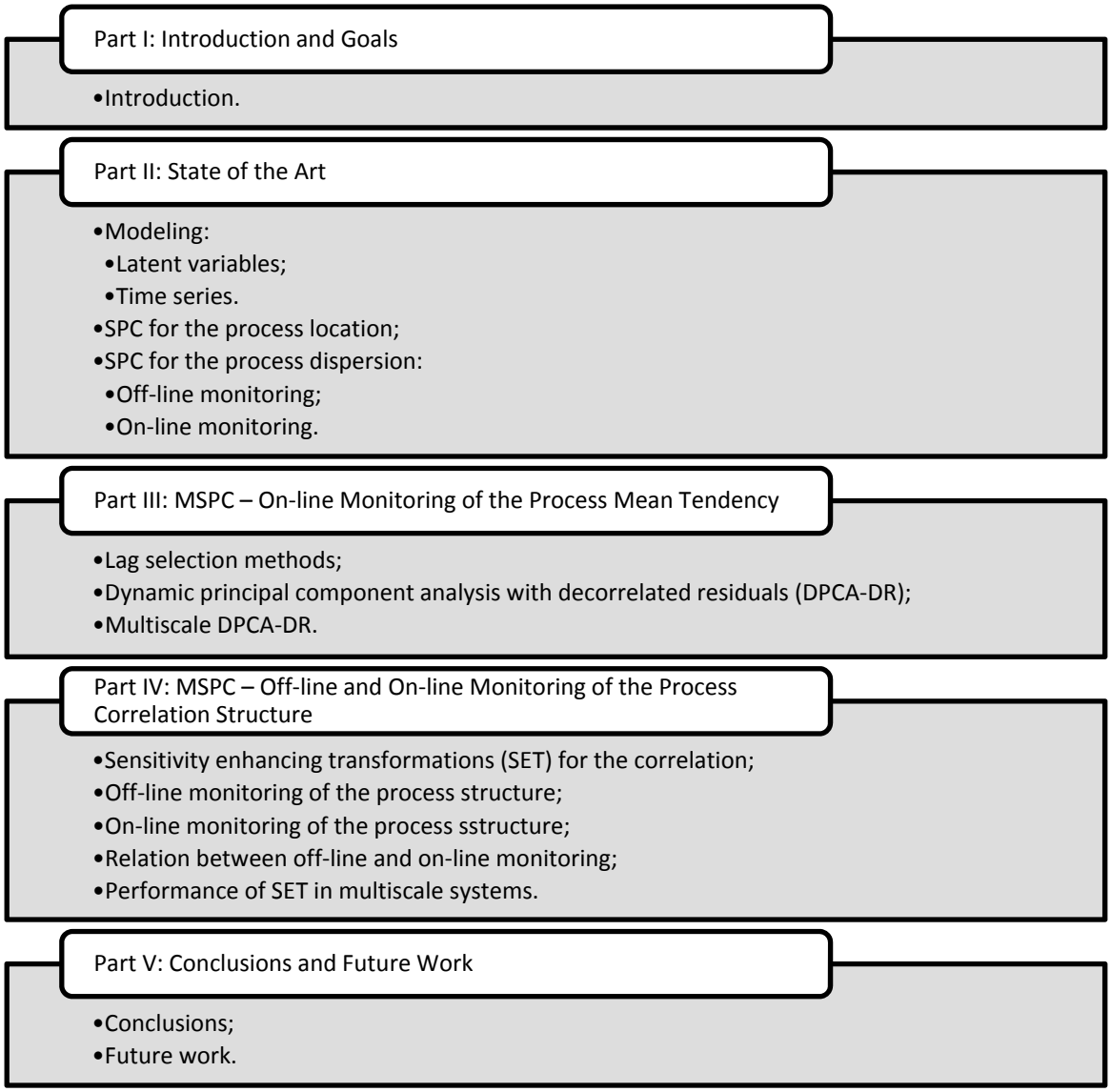


Figure 1.1 Organization of the five parts that compose this thesis.

Part II – State of the Art

2 Modeling

A fundamental task underlying every process systems engineering application, namely process monitoring, is the proper modeling of the process under analysis. In the scope of process monitoring, the complex relationship between process variables as well as the dynamic, non-linear and time-varying dependencies should be accommodated in the monitoring scheme, in an efficient and robust way. For instance, for handling process dynamics it is common to use time series to model the process autocorrelation and then apply control charts to the residuals [12]. Similarly, for time-varying processes a model can be employed to explain the data trend prior to the application of the traditional control charts [13, 14].

Another problem often found in industrial processes arises from the huge amounts of information collected from the process plant. This data is characterized by highly levels of cross- and auto-correlation, which cannot be monitored or modeled directly in an efficient way using more traditional approaches such as classical time series analysis. Instead, latent variables models are used to reduce the data dimension into a few latent variables that explain most of the process underlying phenomena. Principal component analysis (PCA) and partial least squares (PLS) are the main representatives from the class of models used to perform such task. The reduced set of uncorrelated latent variables can then be monitored through appropriated multivariate control charts [15-19]. Moreover, these methods are also useful to develop predictive approaches, variable selection tasks and outlier detection schemes [20]. PCA can also be used in combination with time series in order to handle the computation complexity associated with the modeling of multivariate time series. To do so, PCA is firstly applied to the original data and then its principal components are modeled with resource to time series [1].

In certain cases, data measurements may not be available, due to sensor failures, routine maintenance, gross measurement errors and the existence of different sampling acquisition rates [21, 22]. Such missing measurements need to be estimated in order to proceed with the monitoring schemes, since the suspension of their application until all measurements are available is usually unacceptable [21]. To address this issue, several methods to estimate missing measurements based on existing PCA or PLS models can be

employed as discussed by Nelson *et al.* (1996) [21], Walczak and Massart (2001) [23, 24] and Arteaga and Ferrer (2002) [22].

In the following sections, these three classes of methodologies (latent variable modeling, missing data estimation and multivariate time series modeling) are reviewed for the purpose of introducing such base concepts behind both current and the proposed monitoring schemes, and to set the nomenclature to be followed in the presentation of the proposed methods.

2.1 Latent Variables Modeling

Latent variables models are the backbone of many multivariate SPC techniques due to their ability to explain the main correlation features of data through a reduced set of latent variables. These latent variables usually correspond to a linear combination of the original set of variables, maximizing some optimization criteria, such as maximum variance or prediction power. Variables combinations can also involve dynamic and non-linear terms, in order to incorporate such features in the modeling formalism. Given their relevancy in the context of process monitoring tasks, the PCA, dynamic PCA and PLS models will be introduced in the following sections.

2.1.1 Principal Component Analysis

Principal component analysis (PCA) is a multivariate data analysis technique focused on finding a low dimensional subspace around which the majority of data variability is concentrated (the PCA subspace). The new variables of such a low dimensional subspace, are linear combinations of the original ones, and are called principal components (PCs), being uncorrelated quantities by design, with an implicit ordering: the first PC concentrates the greatest portion of data variability around him, followed by the second PC, which is orthogonal to the first one, and only explains the variability not accounted for by it, and so on, and so forth [1, 20].

PCA provides a decomposition of a data matrix, $\mathbf{X}$, with n observations and m original variables as:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{2.1}$$

where $\mathbf{T}_{n\times p}$ is the matrix of PCA scores (the i^{th} column contains the scores for the i^{th} PC), $\mathbf{P}_{m\times p}$ is the matrix with the PCA loadings (the i^{th} column contains the variable loadings for the i^{th} PC), and $\mathbf{E}_{n\times m}$ is the residual matrix. p stands for the number of PCs retained (the data pseudo-rank), i.e., the dimension of the PCA subspace. As PCA is scale-dependent, the data matrix, $\mathbf{X}$, must be properly pre-processed in some meaningful way, in order to guarantee the quality of data analysis. The most common method of pre-processing, is to center all variables to zero mean and scale them to unit variance, known as "autoscaling", but many other pre-processing methods are available for more specific situations [2, 25].

The computation of principal components can be made with resource to the covariance matrix $\mathbf{\Sigma}$ or, when this is unknown (as happens in practice), to its estimator, the sample covariance matrix, $\mathbf{S} = (n-1)^{-1}\mathbf{X}_c{}^{\mathrm{T}}\mathbf{X}_c$, where $\mathbf{X}_c$ is the matrix of mean centered measurements [18]. As a result of the spectral decomposition, the $\mathbf{\Sigma}$ matrix can be written as [26],

$$\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^{\mathrm{T}} \tag{2.2}$$

where

$$\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_1, \ldots, \lambda_m\right) \tag{2.3}$$

is the diagonal matrix of eigenvalues and,

$$\mathbf{\Gamma} = \begin{bmatrix} \boldsymbol{\gamma}_1 & \boldsymbol{\gamma}_2 & \cdots & \boldsymbol{\gamma}_m \end{bmatrix} \tag{2.4}$$

is an orthogonal matrix consisting of the corresponding eigenvectors of $\mathbf{\Sigma}$.

It is easily shown that the loading matrix, $\mathbf{P}$, is equal to the eigenvectors matrix $\mathbf{\Gamma}$. Furthermore, if $\boldsymbol{\gamma}_i$ is chosen to have unit length (i.e., $\boldsymbol{\gamma}_i^{\mathrm{T}}\boldsymbol{\gamma}_i = 1$), then $\mathrm{var}\left(t_i\right) = \lambda_i$ [18, 20].

2.1.2 Dynamic Principal Component Analysis

PCA only describes the correlation among variables, but does not incorporate any feature to address the correlation along the observations' mode, i.e., the variables autocorrelation. Ku *et al.* (1995) [27] presented an approach for incorporating a linear time series modeling framework into conventional PCA, through a "time lag shift" methodology. It consists in adding several time-shifted replicates of the variables under analysis, to the original set of variables, and then apply PCA to such an extended matrix, say $\tilde{\mathbf{X}}$, in order to also model the variables' dynamic structure, in addition to all the static relationships present. A possible notation for describing this computational scheme is provided in Equation (2.5):

$$\tilde{\mathbf{X}} = \left[\overbrace{\mathbf{x}_1(0)\cdots\mathbf{x}_m(0)}^{\mathbf{x}(0)} \; \overbrace{\mathbf{x}_1(1)\cdots\mathbf{x}_m(1)}^{\mathbf{x}(1)} \cdots \overbrace{\mathbf{x}_1(l)\cdots\mathbf{x}_m(l)}^{\mathbf{x}(l)} \right] \tag{2.5}$$

where $\mathbf{x}_i(j)$ represents the i^{th} variable (in column format) shifted j times into the past (i.e., with j lags). In Equation (2.5), $\mathbf{x}(j)$ is the submatrix containing all the original variables shifted j times; and $\tilde{\mathbf{X}}$ is the resulting extended matrix (with l lags), which, in this case, has the form of an Hankel matrix [28]. However, in general, different lags may be used for different variables, in which case the structure of the extended matrix is no longer that of an Hankel matrix.

Therefore, in simple terms, DPCA is essentially the same method as the original PCA approach, except that the data matrix, $\tilde{\mathbf{X}}$, is now composed by additional time-shifted replicates of the original variables. Thus, after parameter estimation, it corresponds to an implicit vector autoregressive model [29] (VAR or VARX, if process inputs are also included; in fact, in more precise terms, the actual model structure corresponds to a latent variable VAR or VARX). A key aspect when applying DPCA is the definition of its lagged structure, for which some guidelines were already proposed Ku *et al.* (1995) [27].

After the construction of the extended set of process variables, $\tilde{\mathbf{X}}$, a PCA analysis can be carried out as usual, leading to the final DPCA model.

2.1.3 Partial Least Squares

Partial least squares (or projection to latent structures, PLS) is a multivariate method that relates two data matrices, **X** and **Y**, through a linear latent variables model [2, 30-32]. More specifically, given a $(n \times m)$ data matrix of input variables, $\mathbf{X}_{n \times m}$, and a $(n \times r)$ data matrix with the corresponding output variables, $\mathbf{Y}_{n \times r}$, PLS successively finds those pairs of orthogonal linear combinations of the input and output variables, with maximal covariance. In case there is just one output variable, no linear combinations are considered in the **Y** block. By doing so, PLS describes the part of the X-variability with predictive potential for explaining the variability of the outputs. The PLS model structure defines a common latent variable space relating both data blocks, as follows [1, 2, 33]:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{2.6}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{B}\mathbf{Q}^{\mathrm{T}} + \mathbf{F} \tag{2.7}$$

where $\mathbf{T}_{n \times p}$ is the matrix of the X-scores, defining the common latent variable space relating $\mathbf{X}$ and $\mathbf{Y}$; $\mathbf{P}_{m \times p}$ is the X-loading matrix; $\mathbf{Q}_{r \times p}$ is the Y-loading matrix; $\mathbf{B}_{p \times p}$ is a regression coefficient matrix (for the inner latent variable relationships); $\mathbf{E}_{n \times m}$ and $\mathbf{F}_{n \times r}$ are residual matrices; and p is the number of PLS components considered (i.e. latent variables). For the special case where p is equal to the number of input variables in **X**, Equations (2.6) and (2.7) reduce to,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{F} \tag{2.8}$$

which corresponds to a multivariate linear regression model. Moreover, in these circumstances the PLS solution becomes identical to the ordinary least squares (OLS) solution, which is given by [1, 12],

$$\boldsymbol{\beta} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} \tag{2.9}$$

As in the PCA case, a dynamic PLS model can be obtained by including past values of input and/or output variables in the **X** block.

2.2 Missing Data Estimation in PCA

Nelson *et al.* (1996) [21] studied several approaches for estimating the PCA or PLS scores and responses when some observations are missing, but an estimated model is already available. Among the approaches studied, conditional mean replacement was found to be, in general, the most recommendable option, which will be reviewed here for the case of PCA.

Let us consider a data matrix $\mathbf{X}$, which can be decomposed by PCA, as $\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E}$, as referred above. When a new measurement vector is collected, say $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix}^{\mathrm{T}}$, it may contain some missing measurements. For notational convenience and without loss of generality, let us assume that such missing measurements are the first elements of this data vector, which then presents the following partitioned structure:

$$\mathbf{x}^{\mathrm{T}} = \begin{bmatrix} \mathbf{x}^{\#\mathrm{T}} & \mathbf{x}^{*\mathrm{T}} \end{bmatrix} \tag{2.10}$$

where $\mathbf{x}^{\#}$ denotes the missing measurements and $\mathbf{x}^{*}$ the observed variables. Correspondingly, the loading matrix, $\mathbf{P}$, can also be rearranged to conform to such a partition: $\mathbf{P}^{\mathrm{T}} = \begin{bmatrix} \mathbf{P}^{\#\mathrm{T}} & \mathbf{P}^{*\mathrm{T}} \end{bmatrix}$. The methodology for estimating the scores in this case, is a particular case of the application of the Expectation-Maximization (EM) algorithm [34]. This algorithm can, in general, be used to estimate a statistical model in the presence of missing data, $\mathbf{x}^{\#}$. In this approach, the successive estimates of missing data, $\hat{\mathbf{x}}^{\#}$, obtained with updated parameter estimates, during the Estimation stage, are employed to refine the model parameters, during the Maximization stage. The successive refinements obtained for $\hat{\mathbf{x}}^{\#}$, correspond to the expected values of $\mathbf{x}^{\#}$, given the knowledge of the observed variables, $\mathbf{x}^{*}$, and the current estimates of the model parameters, $\boldsymbol{\theta}$:

$$\hat{\mathbf{x}}^{\#} = E\left(\mathbf{x}^{\#} \middle| \mathbf{x}^{*}, \boldsymbol{\theta}\right) \tag{2.11}$$

In the present situation, it is assumed that a PCA model is already available (estimated from reference data), and therefore only the Expectation stage of the EM algorithm is necessary, in order to compute the expected values for missing measurements, conditionally to the knowledge of $\mathbf{x}^{*}$ and the model parameters (the PCA loadings and pre-processing parameters). Let us also consider the following notation for the spectral

decomposition of the covariance matrix, $\mathbf{\Sigma} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{\mathrm{T}}$, after introducing the above mentioned missing data/observed variables partition (where $\mathbf{\Lambda}$ is the diagonal matrix, with the PCA eigenvalues along the main diagonal; all the other quantities have the same meanings introduced before):

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{P}^{\#}\mathbf{\Lambda}\mathbf{P}^{\#\mathrm{T}} & \mathbf{P}^{\#}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}} \\ \mathbf{P}^{*}\mathbf{\Lambda}\mathbf{P}^{\#\mathrm{T}} & \mathbf{P}^{*}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}} \end{bmatrix} \tag{2.12}$$

Using this expression for $\mathbf{\Sigma}$, the conditional expectation of the missing measurements, is simply given by:

$$\hat{\mathbf{x}}^{\#} = \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{x}^{*} = \mathbf{P}^{\#}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\left(\mathbf{P}^{*}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\right)^{-1}\mathbf{x}^{*} \tag{2.13}$$

The estimated missing measurements can then be used in the score calculation along with the observed data, as if no measurements were missing, leading to the following expression for the first *p* scores (*p* is the pseudo-rank):

$$\begin{aligned} \hat{\mathbf{t}} &= \mathbf{P}_{1:p}^{\mathrm{T}}\begin{bmatrix} \hat{\mathbf{x}}^{\#} \\ \mathbf{x}^{*} \end{bmatrix} = \mathbf{P}_{1:p}^{\mathrm{T}}\begin{bmatrix} \mathbf{P}^{\#}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\left(\mathbf{P}^{*}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\right)^{-1}\mathbf{x}^{*} \\ \left(\mathbf{P}^{*}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\right)\left(\mathbf{P}^{*}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\right)^{-1}\mathbf{x}^{*} \end{bmatrix} \\ &= \mathbf{P}_{1:p}^{\mathrm{T}}\begin{bmatrix} \mathbf{P}^{\#} \\ \mathbf{P}^{*} \end{bmatrix}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\left(\mathbf{P}^{*}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\right)^{-1}\mathbf{x}^{*} \\ &= \mathbf{P}_{1:p}^{\mathrm{T}}\mathbf{P}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\left(\mathbf{P}^{*}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\right)^{-1}\mathbf{x}^{*} \\ &= \begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\left(\mathbf{P}^{*}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}\right)^{-1}\mathbf{x}^{*} \end{aligned} \tag{2.14}$$

where $\mathbf{P}_{1:p}$ is composed by the first *p* columns of the loading matrix, $\mathbf{P}$, $\mathbf{I}$ is an ($p \times p$) identity matrix, and $\mathbf{0}$ is an ($p \times (m - p)$) matrix of zeros.

A possible limitation of the conditional mean replacement solution is that $\mathbf{X}^{*\mathrm{T}}\mathbf{X}^{*}$ (i.e., $\mathbf{P}^{*}\mathbf{\Lambda}\mathbf{P}^{*\mathrm{T}}$) may become ill-conditioned. In such case, traditional regression solutions, such as ridge regression, principal components regression and PLS can be employed. Alternatively, one may also use the projection to model plane approach [21] or trimmed scores regression (TSR) [22], to estimate the scores corresponding to missing observations.

In the case of the projection to model plane method all scores are estimated at once by regressing $\mathbf{x}^*$ onto the plane defined by $\mathbf{P}^*$, which results in [21],

$$\hat{\mathbf{t}} = \left(\mathbf{P}_{1:p}^{*\,\mathrm{T}}\mathbf{P}_{1:p}^{*}\right)^{-1}\mathbf{P}_{1:p}^{*}\mathbf{x}^{*} \tag{2.15}$$

as the least squares estimator based on the measured variables.

As for the trimmed scores regression proposed by Arteaga and Ferrer (2002) [22], the scores are estimated by application of a linear regression model based on the known scores ($\mathbf{t}^* = \mathbf{P}_{1:p}^{*\,\mathrm{T}}\mathbf{x}^*$). That is,

$$\hat{\mathbf{t}} = \mathbf{B}^{\mathrm{T}}\mathbf{t}^{*} \tag{2.16}$$

where,

$$\mathbf{B} = \left(\mathbf{P}_{1:p}^{*\,\mathrm{T}}\mathbf{P}^{*}\boldsymbol{\Lambda}\mathbf{P}^{*\mathrm{T}}\mathbf{P}_{1:p}^{*}\right)^{-1}\mathbf{P}_{1:p}^{*\,\mathrm{T}}\mathbf{P}_{1:p}^{*}\boldsymbol{\Lambda}_{1:p} \tag{2.17}$$

is the least squares estimator of the regression model. Thus, the scores are estimated as [22],

$$\hat{\mathbf{t}} = \boldsymbol{\Lambda}_{1:p}\mathbf{P}_{1:p}^{*\,\mathrm{T}}\mathbf{P}_{1:p}^{*}\left(\mathbf{P}_{1:p}^{*\,\mathrm{T}}\mathbf{P}^{*}\boldsymbol{\Lambda}\mathbf{P}^{*\mathrm{T}}\mathbf{P}_{1:p}^{*}\right)^{-1}\mathbf{P}_{1:p}^{*\,\mathrm{T}}\mathbf{x}^{*} \tag{2.18}$$

2.3 Multivariate Time Series Analysis

In the context of statistical process control, multivariate time series models are usually applied in order to explain the autocorrelation present on the data. This is done by either a direct application of the model to obtain serial uncorrelated residuals, or implicitly, through a DPCA model as proposed by Ku *et al.* (1995) [27]. Time series models are also applied for forecast and structural analysis. One of these applications is the Granger causality, which resort to vector autoregressive models to infer the causal relation between pairs of variables. A subset of these models will be described below, along with a brief description of their main characteristics. Further information about these models can be found elsewhere [29, 35-37].

2.3.1 Vector Autoregressive Model

The vector autoregressive (VAR) model of order p, VAR(p), represents a linear dependency between $\mathbf{y}_t$ and its past realizations according to [36],

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{d}_t, \quad t = 0, \pm 1, \pm 2, \ldots \tag{2.19}$$

where $\mathbf{y}_t = \left[y_{1t}, \ldots, y_{kt}\right]^{\mathrm{T}}$ is a $(k \times 1)$ vector of measurements at instant t, $\mathbf{A}_i$ is a $(k \times k)$ matrix of coefficients, $\mathbf{v} = \left[v_1, \ldots, v_k\right]^{\mathrm{T}}$ is a $(k \times 1)$ vector of intercept terms and $\mathbf{d}_t = \left[d_{1t}, \ldots, d_{kt}\right]^{\mathrm{T}}$ is a $(k \times 1)$ vector of white noise, i.e., $E(\mathbf{d}_t) = \mathbf{0}$, $E\left(\mathbf{d}_t\mathbf{d}_t^{\mathrm{T}}\right) = \mathbf{\Sigma}_d$ and $E\left(\mathbf{d}_t\mathbf{d}_s^{\mathrm{T}}\right) = \mathbf{0}$ for $s \neq t$.

This linear dependency can be easily exemplified for a 1st order bivariate process (i.e., k = 2, $\mathbf{y}_t = \left[y_{1t}, y_{2t}\right]^{\mathrm{T}}$ and $\mathbf{d}_t = \left[d_{1t}, d_{2t}\right]^{\mathrm{T}}$). In this case, Equation (2.19) can be rewritten as,

$$y_{1t} = v_1 + \alpha_{11}y_{1,t-1} + \alpha_{12}y_{2,t-1} + d_{1,t} \tag{2.20}$$

$$y_{2t} = v_2 + \alpha_{21}y_{1,t-1} + \alpha_{22}y_{2,t-1} + d_{2,t} \tag{2.21}$$

where α_{ij} is the (i, j) element of $\mathbf{A}_1$ and v_i is the i^{th} element of $\mathbf{v}$.

In Equation (2.20), α_{12} represents the linear dependency of y_{1t} on $y_{2,t-1}$ in the presence of $y_{1,t-1}$. Therefore, if α_{12} is equal to zero, y_{1t} does not depend on $y_{2,t-1}$, and the model shows that y_{1t} only depends on its own past history. Likewise, if α_{21} (Equation (2.21)) is zero, y_{2t} does not depend on $y_{1,t-1}$ [35]. This situation corresponds to the particular case where no interaction exists between the two variables.

2.3.1.1 Stationarity Condition and Moments

To better illustrate the properties of the VAR model, the characteristics of the VAR(1) model are first introduced and latter extended to the general model. The model under consideration has the form,

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{d}_t \tag{2.22}$$

Assuming that the VAR(1) model in Equation (2.22) is *stable*, the model's expected value, for $E(\mathbf{d}_t) = \mathbf{0}$, becomes,

$$E(\mathbf{y}_t) = \mathbf{v} + \mathbf{A}_1 E(\mathbf{y}_{t-1}) \tag{2.23}$$

Since $E(\mathbf{y}_t)$ is time-invariant [35],

$$\boldsymbol{\mu} \equiv E(\mathbf{y}_t) = (\mathbf{I}_k - \mathbf{A}_1)^{-1} \mathbf{v} \tag{2.24}$$

Using $\mathbf{v} = (\mathbf{I}_k - \mathbf{A}_1)\boldsymbol{\mu}$, the VAR(1) model can be rewritten as,

$$\mathbf{y}_t - \boldsymbol{\mu} = \mathbf{A}_1 (\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \mathbf{d}_t \tag{2.25}$$

By defining $\tilde{\mathbf{y}}_t = \mathbf{y}_t - \boldsymbol{\mu}$, a further simplification can be obtained, resulting in,

$$\tilde{\mathbf{y}}_t = \mathbf{A}_1 \tilde{\mathbf{y}}_{t-1} + \mathbf{d}_t \tag{2.26}$$

The model represented by Equation (2.26) can then be used to derive the remaining properties of the VAR(1) model.

By repeated substitutions, the VAR(1) model of Equation (2.26) assumes the form of a vector moving average (VMA) model of infinite order:

$$\tilde{\mathbf{y}}_t = \mathbf{d}_t + \mathbf{A}_1 \mathbf{d}_{t-1} + \mathbf{A}_1^2 \mathbf{d}_{t-2} + \mathbf{A}_1^3 \mathbf{d}_{t-3} + \cdots \tag{2.27}$$

From this expression, several characteristics of the VAR(1) model can be observed. One of such characteristics is a result of the serial independency of $\mathbf{d}_t$, from where $\text{cov}(\mathbf{y}_{t-h}, \mathbf{d}_t) = 0$ for all $h > 0$ [35]. Other property can be obtained by post multiplying Equation (2.27) by $\mathbf{d}_t^{\mathrm{T}}$ and taking its expectation, from where it follows that $\text{cov}(\mathbf{y}_t, \mathbf{d}_t) = \boldsymbol{\Sigma}_d$ [35]. This relation also shows the dependency between $\mathbf{y}_t$ and the past values of the white noise sequence ($\mathbf{d}_{t-j}$) through the coefficients matrix, $\mathbf{A}_1^j$. Fur such dependency to be meaningful, $\mathbf{A}_1^j$ must converge to zero as $j \to \infty$ [35]. Therefore, the eigenvalues of $\mathbf{A}_1$ must be less than 1 in modulus [35, 36]. This is the necessary and sufficient condition for defining a VAR(1) model as *stable* [35, 36]. Due to the relation

presented in Equation (2.28), the stability of a VAR(1) process can also be determined by Equation (2.29), with $\upsilon = 1/\lambda$ [35, 36].

$$\det\left(\lambda \mathbf{I}_k - \mathbf{A}_1\right) = \lambda^k \det\left(\mathbf{I}_k - \mathbf{A}_1 \frac{1}{\lambda}\right) \tag{2.28}$$

$$\det\left(\mathbf{I}_k - \mathbf{A}_1 \upsilon\right) \neq 0 \quad \text{for} \quad |\upsilon| \leq 1 \tag{2.29}$$

Thus, a VAR(1) model is *stable*, if all the zeros (υ) in Equation (2.29) are greater than 1 in modulus [35, 36].

Recalling Equation (2.27), it can be verified that the covariance of $\mathbf{y}_t$ is given by [35],

$$\operatorname{cov}\left(\mathbf{y}_t\right) = \boldsymbol{\Gamma}_y\left(0\right) = \boldsymbol{\Sigma}_d + \mathbf{A}_1 \boldsymbol{\Sigma}_d \mathbf{A}_1^{\mathrm{T}} + \mathbf{A}_1^{\,2} \boldsymbol{\Sigma}_d \left(\mathbf{A}_1^{\,2}\right)^{\mathrm{T}} + \cdots = \sum_{i=0}^{\infty} \mathbf{A}_1^{\,i} \boldsymbol{\Sigma}_d \left(\mathbf{A}_1^{\,i}\right)^{\mathrm{T}} \tag{2.30}$$

where $\mathbf{A}_1^{\,0} = \mathbf{I}_k$.

Alternatively, the covariance of $\mathbf{y}_t$ can also be determined by post multiplying $\tilde{\mathbf{y}}_{t-h}^{\mathrm{T}}$ to Equation (2.27) and taking its expectation, resulting in,

$$\boldsymbol{\Gamma}_y\left(h\right) = E\left(\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_{t-h}^{\mathrm{T}}\right) = \mathbf{A}_1 E\left(\tilde{\mathbf{y}}_{t-1} \tilde{\mathbf{y}}_{t-h}^{\mathrm{T}}\right) + E\left(\mathbf{d}_t \tilde{\mathbf{y}}_{t-h}^{\mathrm{T}}\right) \tag{2.31}$$

Considering that [35],

$$\operatorname{cov}(\mathbf{y}_{t-h}, \mathbf{d}_t) = \begin{cases} \boldsymbol{\Sigma}_d & h = 0 \\ \mathbf{0} & h > 0 \end{cases} \tag{2.32}$$

for $h = 0$ one obtains [36],

$$\boldsymbol{\Gamma}_y\left(0\right) = \mathbf{A}_1 \boldsymbol{\Gamma}_y\left(-1\right) + \boldsymbol{\Sigma}_d = \mathbf{A}_1 \boldsymbol{\Gamma}_y\left(1\right) + \boldsymbol{\Sigma}_d \tag{2.33}$$

and for $h > 0$ [36],

$$\boldsymbol{\Gamma}_y\left(h\right) = \mathbf{A}_1 \boldsymbol{\Gamma}_y\left(h-1\right) \tag{2.34}$$

If $\mathbf{A}_1$ and $\mathbf{\Sigma}_d$ are known, $\mathbf{\Gamma}_y(0)$ can then be determined from Equation (2.34) with $h=1$, from where it results that $\mathbf{\Gamma}_y(1)=\mathbf{A}_1\mathbf{\Gamma}_y(0)$. By substitution in Equation (2.33), one finally gets [36],

$$\mathbf{\Gamma}_y(0)=\mathbf{A}_1\mathbf{\Gamma}_y(0)\mathbf{A}_1^{\mathrm{T}}+\mathbf{\Sigma}_d \tag{2.35}$$

or

$$\mathrm{vec}\left(\mathbf{\Gamma}_y(0)\right)=\mathrm{vec}\left(\mathbf{A}_1\mathbf{\Gamma}_y(0)\mathbf{A}_1^{\mathrm{T}}\right)+\mathrm{vec}\left(\mathbf{\Sigma}_d\right) \tag{2.36}$$

Hence [36],

$$\mathrm{vec}\left(\mathbf{\Gamma}_y(0)\right)=\left(\mathbf{I}_{k^2}-\mathbf{A}_1\otimes\mathbf{A}_1\right)^{-1}\mathrm{vec}\left(\mathbf{\Sigma}_d\right) \tag{2.37}$$

where $\otimes$ is the Kronecker product.

2.3.1.2 Properties of VAR(p) Models

The properties described previously for the VAR(1) model are easily extended for the VAR(p) models, since any VAR(p) process with $p > 1$ can be written in the form of a VAR(1) [36].

For a VAR(p) model described by Equation (2.19), the corresponding VAR(1) model is given by,

$$\mathbf{x}_t=\mathbf{v}^*+\mathbf{A}^*\mathbf{x}_{t-1}+\mathbf{b}_t \tag{2.38}$$

where

$$\mathbf{x}_t=\begin{bmatrix}\mathbf{y}_t\\ \mathbf{y}_{t-1}\\ \vdots\\ \mathbf{y}_{t-p+1}\end{bmatrix}_{(kp\times 1)} \qquad \mathbf{v}^*=\begin{bmatrix}\mathbf{v}\\ \mathbf{0}\\ \vdots\\ \mathbf{0}\end{bmatrix}_{(kp\times 1)}$$

$$\mathbf{A}^* = \underset{(kp\times kp)}{\begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_k & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k & & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_k & \mathbf{0} \end{bmatrix}} \qquad \mathbf{b}_t = \underset{(kp\times 1)}{\begin{bmatrix} \mathbf{d}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}}$$

From this representation, it results that $\mathbf{x}_t$ is *stable* if [35, 36],

$$\begin{gathered} \det\left(\mathbf{I}_{kp} - \mathbf{A}^* \upsilon\right) \neq 0 \Leftrightarrow \\ \Leftrightarrow \det\left(\mathbf{I}_k - \mathbf{A}_1\upsilon - \cdots - \mathbf{A}_p\upsilon^p\right) \neq 0 \quad \text{for} \quad |\upsilon| \leq 1 \end{gathered} \tag{2.39}$$

The mean value of this model is then given by Equation (2.40) and its covariance by Equation (2.41) [36],

$$\boldsymbol{\mu} \equiv E\left(\mathbf{x}_t\right) = \left(\mathbf{I}_{kp} - \mathbf{A}^*\right)^{-1} \mathbf{v}^* \tag{2.40}$$

$$\operatorname{vec}\left(\boldsymbol{\Gamma}_x\left(0\right)\right) = \left(\mathbf{I}_{(kp)^2} - \mathbf{A}^* \otimes \mathbf{A}^*\right)^{-1} \operatorname{vec}\left(\Sigma_b\right) \tag{2.41}$$

where,

$$\Sigma_b = E\left(\mathbf{b}_t \mathbf{b}_t^{\mathrm{T}}\right) = \underset{(kp\times kp)}{\begin{bmatrix} \boldsymbol{\Sigma}_d & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}} \qquad \boldsymbol{\Gamma}_x\left(0\right) = \underset{(kp\times kp)}{\begin{bmatrix} \boldsymbol{\Gamma}_y\left(0\right) & \boldsymbol{\Gamma}_y\left(1\right) & \cdots & \boldsymbol{\Gamma}_y\left(p-1\right) \\ \boldsymbol{\Gamma}_y\left(-1\right) & \boldsymbol{\Gamma}_y\left(0\right) & \cdots & \boldsymbol{\Gamma}_y\left(p-2\right) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_y\left(-p+1\right) & \boldsymbol{\Gamma}_y\left(-p+2\right) & \cdots & \boldsymbol{\Gamma}_y\left(0\right) \end{bmatrix}}$$

Finally, note that the conversion of a *stable* VAR(p) model (given by Equation (2.19)) into a VMA (given by Equation (2.42)), can be made through Equation (2.43), which provides the equivalent coefficient matrices $\boldsymbol{\Phi}_i$, where $\boldsymbol{\Phi}_0 = \mathbf{I}_k$ [36].

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \mathbf{d}_{t-i} \tag{2.42}$$

$$\boldsymbol{\Phi}_i = \sum_{j=1}^{i} \boldsymbol{\Phi}_{i-j} \mathbf{A}_j, \quad i = 1, 2, \ldots \tag{2.43}$$

2.3.2 Vector Moving Average Model

A vector moving average models of order q, VMA(q), relates $\mathbf{y}_t$ with the past values of the white noise, $\mathbf{d}_{t-j}$, as [35, 36],

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{d}_t + \mathbf{M}_1\mathbf{d}_{t-1} + \cdots + \mathbf{M}_q\mathbf{d}_{t-q}, \quad t = 0, \pm 1, \pm 2, \ldots \tag{2.44}$$

where $\mathbf{y}_t = \left[y_{1t}, \ldots, y_{kt}\right]^{\mathrm{T}}$ is a $(k \times 1)$ vector of measurements, $\mathbf{M}_i$ is a $(k \times k)$ matrix of coefficients, $\boldsymbol{\mu} = \left[\mu_1, \ldots, \mu_k\right]^{\mathrm{T}}$ is a $(k \times 1)$ vector of the mean values of $\mathbf{y}_t$ (i.e., $E(\mathbf{y}_t) = \boldsymbol{\mu}$) and $\mathbf{d}_t = \left[d_{1t}, \ldots, d_{kt}\right]^{\mathrm{T}}$ is a $(k \times 1)$ vector of white noise with covariance $\boldsymbol{\Sigma}_d$.

For determining the properties of the VMA model, the VMA(1) scenario is first analyzed. This VMA(1) model, with zero mean (for simplification purposes), is described as,

$$\mathbf{y}_t = \mathbf{d}_t + \mathbf{M}_1\mathbf{d}_{t-1} \tag{2.45}$$

This model, can be converted into the corresponding VAR model, by rewriting Equation (2.45) in its equivalent form through successive substitutions, as [36]:

$$\begin{aligned} \mathbf{d}_t &= \mathbf{y}_t - \mathbf{M}_1\mathbf{d}_{t-1} \\ &= \mathbf{y}_t + \sum_{i=1}^{\infty}\left(-\mathbf{M}_1\right)^i \mathbf{y}_{t-i} \end{aligned} \tag{2.46}$$

Then, if $\mathbf{M}_1^{\,i} \to 0$ as $i \to \infty$,

$$\mathbf{y}_t = -\sum_{i=1}^{\infty}\left(-\mathbf{M}_1\right)^i \mathbf{y}_{t-i} + \mathbf{d}_t \tag{2.47}$$

which is the corresponding VAR model of infinite order. Because $\left(-\mathbf{M}_1\right)^i$ may be equal to zero for i greater than some finite number p, the VMA model can in fact be represented by a finite order VAR(p) [36].

For the representation in Equation (2.47) to be meaningful, $\mathbf{M}_1^{\,i}$ must approach zero when i tends to infinity. This condition is obtained if the eigenvalues of $\mathbf{M}_1$ are all less than 1 in modulus, i.e.,

$$\det\left(\mathbf{I}_k + \mathbf{M}_1 \upsilon\right) \neq 0 \quad \text{for} \quad |\upsilon| \leq 1 \tag{2.48}$$

This condition is analogous to the stability condition of the VAR(1) model.

In a more general form, a VMA(q) model with zero mean (Equation (2.49)) can be represented by a VAR model (Equation (2.50)) if the condition in Equation (2.51) is verified [36].

$$\mathbf{y}_t = \mathbf{d}_t + \mathbf{M}_1 \mathbf{d}_{t-1} + \cdots + \mathbf{M}_q \mathbf{d}_{t-q}, \quad t = 0, \pm 1, \pm 2, \ldots \tag{2.49}$$

$$\mathbf{y}_t = \sum_{i=1}^{\infty} \mathbf{\Pi}_i \mathbf{y}_{t-i} + \mathbf{d}_t \tag{2.50}$$

$$\det\left(\mathbf{I}_k + \mathbf{M}_1 \upsilon + \cdots + \mathbf{M}_q \upsilon^q\right) \neq 0 \quad \text{for} \quad |\upsilon| \leq 1 \tag{2.51}$$

A VMA model with this property is defined as *invertible*. The coefficient matrices $\mathbf{\Pi}_i$ of this model can be recursively computed by Equation (2.52), using $\mathbf{\Pi}_1 = \mathbf{M}_1$ and $\mathbf{M}_j = \mathbf{0}$ for $j > q$ [36].

$$\mathbf{\Pi}_i = \mathbf{M}_i - \sum_{j=1}^{i-1} \mathbf{\Pi}_{i-j} \mathbf{M}_j, \quad i = 2, 3, \ldots \tag{2.52}$$

To determine the covariance matrix of the VMA model, Equation (2.53) can be used with $\mathbf{M}_0 = \mathbf{I}_k$ [35, 36].

$$\mathbf{\Gamma}_y(h) = E\left(\mathbf{y}_t \mathbf{y}_{t-h}^{\mathrm{T}}\right) = \begin{cases} \sum_{i=0}^{q-h} \mathbf{M}_{i+h} \mathbf{\Sigma}_d \mathbf{M}_i^{\mathrm{T}} & h = 0, 1, \ldots, q \\ \mathbf{0} & h > q \end{cases} \tag{2.53}$$

2.3.3 Vector Autoregressive Moving Average Model

A VAR model of finite order is generally represented by Equation (2.54).

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \tag{2.54}$$

In the case where the error term ($\boldsymbol{\varepsilon}_t$) has autocorrelation, described by a VMA model of finite order given by,

$$\boldsymbol{\varepsilon}_t = \mathbf{d}_t + \mathbf{M}_1\mathbf{d}_{t-1} + \cdots + \mathbf{M}_q\mathbf{d}_{t-q} \tag{2.55}$$

where $\mathbf{d}_t$ is white noise with zero mean and covariance $\boldsymbol{\Sigma}_d$, the broader class of vector autoregressive moving average (VARMA) models is obtained [36].

These models are composed by an autoregressive component of order p and another moving average component of order q and are defined as VARMA(p,q). Their mathematical representation is given as [36],

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{d}_t + \mathbf{M}_1\mathbf{d}_{t-1} + \cdots + \mathbf{M}_q\mathbf{d}_{t-q}, \quad t = 0, \pm 1, \pm 2, \ldots \tag{2.56}$$

As described in Section 2.3.1, the autoregressive component of the VARMA model can be converted into an equivalent moving average representation if the model is *stable*, that is, if

$$\det\left(\mathbf{I}_k - \mathbf{A}_1\upsilon - \cdots - \mathbf{A}_p\upsilon^p\right) \neq 0 \quad \text{for} \quad |\upsilon| \leq 1 \tag{2.57}$$

In this case, the VARMA model assumes the equivalent form of Equation (2.58) [36].

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \boldsymbol{\varepsilon}_{t-i} \tag{2.58}$$

By use of Equation (2.55) the pure VMA representation, Equation (2.59), is obtained [36].

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \left(\mathbf{d}_{t-1} + \mathbf{M}_1\mathbf{d}_{t-i-1} + \cdots + \mathbf{M}_q\mathbf{d}_{t-i-q}\right) \\ &= \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{N}_i\mathbf{d}_{t-i} \end{aligned} \tag{2.59}$$

The coefficients matrices $\mathbf{N}_i$ are computed by use of Equation (2.60) with $\mathbf{N}_0 = \mathbf{I}_k$ [36].

$$\mathbf{N}_i = \mathbf{M}_i - \sum_{j=1}^{i} \mathbf{A}_j\mathbf{N}_{i-j}, \quad i = 1, 2, \ldots \tag{2.60}$$

Likewise, if the moving average components is *invertible*, a pure VAR representation, Equation (2.61), can be obtained [36].

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=1}^{\infty} \boldsymbol{\Theta}_i \mathbf{y}_{t-i} + \mathbf{d}_t \tag{2.61}$$

The coefficient matrices $\boldsymbol{\Theta}_i$ of this model are determined through Equation (2.62) [36].

$$\boldsymbol{\Theta}_i = \mathbf{A}_i + \mathbf{M}_i - \sum_{j=1}^{i-1} \mathbf{M}_{i-j} \boldsymbol{\Theta}_j, \quad i = 1, 2, \ldots \tag{2.62}$$

3 Statistical Process Control: an Overview

Statistical process control (SPC) encompasses a collection of techniques that ultimately aim to improve the capability to meet consumers' and others stakeholders' requirements. Among these techniques, control charts are a powerful tool to verify if the process is only subjected to *common causes* of variation and detect any unusual source of variability [12]. These *common causes* of variation are inherent to the process and cannot be removed without changing the process itself. Therefore, they are unavoidable and cause the process to fluctuate around its desired specifications with a stable and recognizable variability pattern. As long as the process is only subjected to such natural and random fluctuations, it is considered to be under a state of statistical control. Otherwise, an *assignable cause* is declared, which should be promptly detected, diagnosed and corrected, in order to reduce the quality losses to a minimum. The detection of *assignable causes* is usually done through control charts.

The base concept of control charts amounts to the successive application of hypothesis tests to check process (statistical) stability, in particular to verify whether the location and/or dispersion of key process variables have not changed [38]. In practice, this is performed by a graphical representation of the assessed variables quality characteristics against control limits, which delimitate the region of normal operation conditions (NOC). Thus, a point plotted within the control limits is consistent with the maintenance of the state of statistical control [39]. The first control chart was the univariate chart introduced by Shewhart (1931) [40], consisting of a center line (CL), an upper control limit (UCL) and a lower control limit (LCL) as follows,

$$\begin{aligned} UCL &= \mu_0 + k\sigma_0 \\ CL &= \mu_0 \\ LCL &= \mu_0 - k\sigma_0 \end{aligned} \tag{3.1}$$

where μ_0 is the target value (usually the mean value of the assessed characteristic under NOC), σ_0 is the standard deviation of the assessed characteristic and k is the distance of the control limits from the center line in standard deviations units [39]. The value of k is usually set to 3, representing "3-sigma" limits, which contain 99.73% of the NOC process

variability, if the variable follows a normal distribution [38]. An example of the Shewhart control chart is given in Figure 3.1, for random draws of the standardized normal distribution with 3-sigma limits.

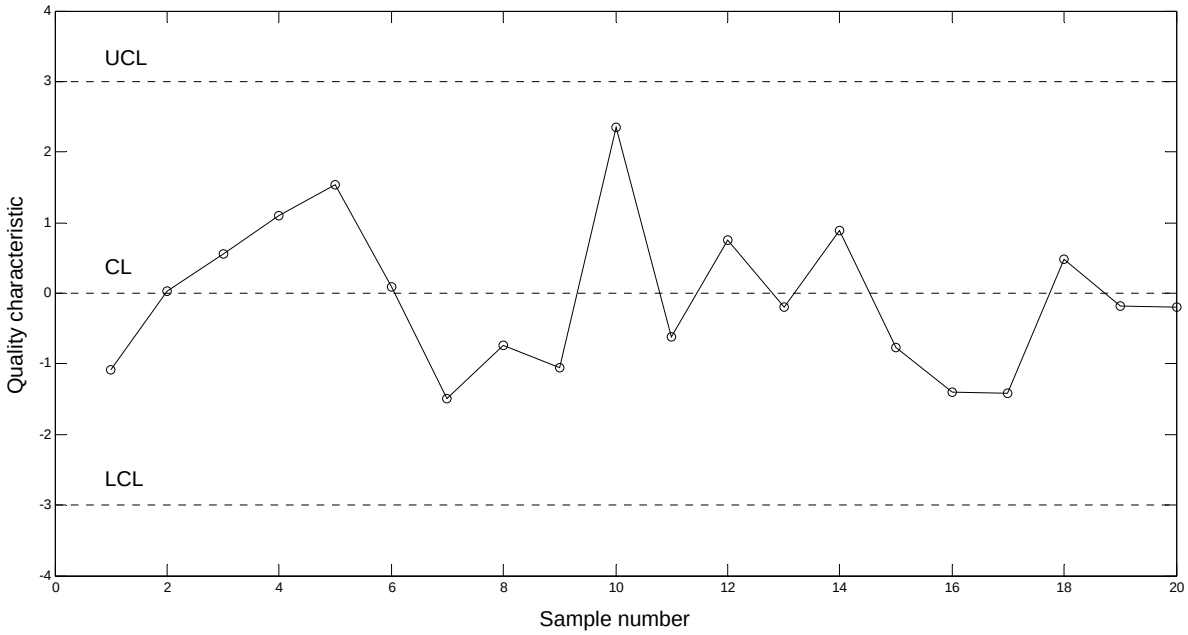


Figure 3.1 Typical representation of a Shewhart control chart.

The Shewhart control chart can be used in a variety of applications and is generally the base model for the more sophisticated approaches. In its construction it is assumed that the quality characteristic under assessment is approximately normally distributed and that the observations are independent [38]. Moreover, this control chart is relatively insensitive to small deviations (say, smaller than $1.5\sigma_0$) since only the current value is used to monitor the process [39]. To overcame this issue, other approaches were developed that bring "memory" to the monitoring procedure, such as the cumulative sum (CUSUM) control chart [41] and exponentially weighted moving average (EWMA) control chart [42].

The previously mentioned control charts are classified as univariate, since they only monitor a single isolated variable. This type of approach was initially justifiable due to the low number of measured variables in industrial plants in the early times. However, the development of data acquisition techniques, allied with increasingly inexpensive sensors

for measuring process variables such as temperature, flow rates and pressure, and the availability of affordable computational power and memory, allowed the collection of large amounts of industrial data which could no longer be properly monitored by univariate approaches. Not only would be impossible to efficiently monitor all possible univariate control charts, but it would be even undesirable to do so. The reasons are the increased false alarm rates associated with driving so many parallel testing schemes, and the intrinsic limitation of univariate monitoring schemes, which fail to take into consideration the inner correlation between process variables [38]. As the relationships between variables are not considered in univariate schemes, they are much less sensitive to detect changes in the way variables are associated. A simple example illustrating the limitations of the univariate Shewhart control charts is presented in Figure 3.2 for the case of two process variables. From the analysis of the individual univariate control charts, one would conclude that the process is in-control over all the monitoring period. However, analyzing the scatter plot, it is possible to observe that while most observations do lie inside the 99.73% confidence ellipse, observation 40 (represented by a circle) has a clear deviation from the normal behavior which is not signaled by any of the univariate control charts.

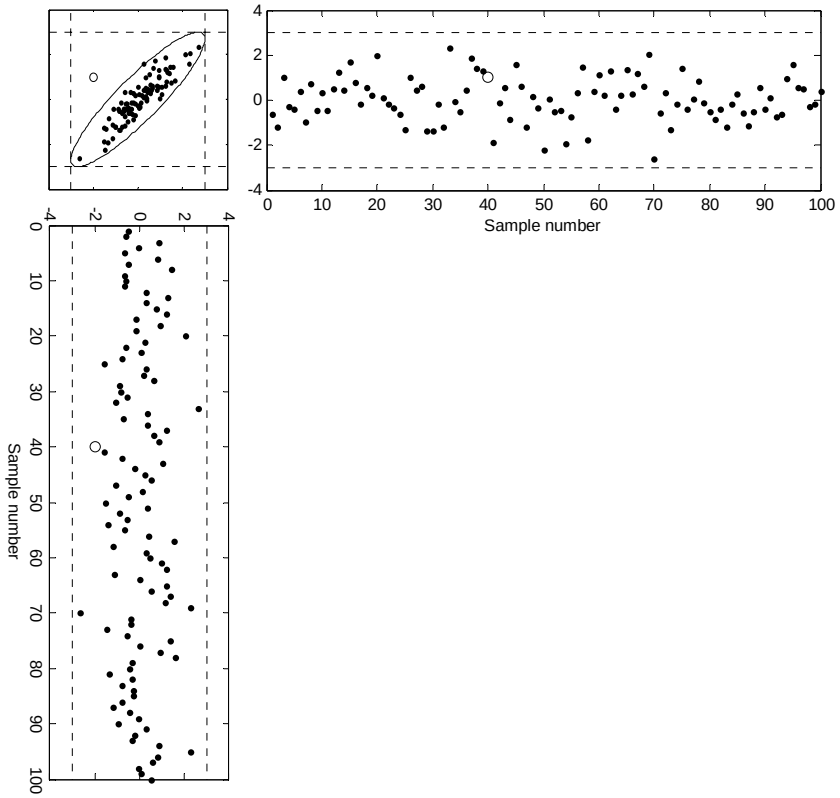


Figure 3.2 Limitations of monitoring a two variable process using two univariate Shewhart control charts.

To overcome the limitations mentioned earlier, multivariate control schemes, such as the Hotelling's T^2 [43], multivariate CUSUM [44] and multivariate EWMA [45], were developed. However, when the number of measured variables (m) becomes large (order of several dozens or even higher), even these schemes begin to experiment problems, mostly due to the issues raised by the variables collinearity and usually associated with the need to invert the covariance matrix of all the variables under simultaneous monitoring. A common solution for dealing with this issue consists of adopting a latent variable modeling framework, developed for these types of processes, whose parameters can be estimated with simple and stable methods. Examples of such latent variables models are principal component analysis (PCA) [15, 16] and partial least squares (PLS) [17-19], the former for problems involving a single block of process variables and the latter for those situations where two blocks of variables need to be explicitly and simultaneously handled. A more detailed revision of these procedures is provided in the following sections.

3.1 Univariate Cumulative Sum and Exponentially Weighted Moving Average Control Charts

In this section the univariate cumulative control chart and exponentially weighted moving average control chart are briefly introduced. These control charts aim to mitigate the relative insensitiveness of the Shewhart control chart to small deviations on the monitored variable by inclusion of past information in the monitoring procedure. They differ by the way through which they bring such additional information from the past to their respective control charts.

3.1.1 Cumulative Sum Control Chart

The base cumulative sum (CUSUM) control chart incorporates all measurements since the beginning of the monitoring procedure by application of cumulative sums of deviations from the target value (μ_0) by defining [12, 38],

$$C_i = \sum_{j=1}^{i} \left(x_j - \mu_0 \right) \tag{3.2}$$

and plotting it against the sample number *i*. To avoid storing all the data, C_i can also be recursively computed as,

$$C_i = \left(x_i - \mu_0 \right) + C_{i-1} \tag{3.3}$$

If the process remains in-control, C_i will present a random walk around zero. However, when a fault occurs, a trend is expected to appear in the plot. To detect this trend, a V-mask procedure was employed in the earlier applications of this scheme. The base idea behind this approach consists in placing a V-shaped mask over each new observation. If all the previous cumulative sums lie within the mask, then the process is considered to be in-control. As an alternative to the graphical V-mask, the tabular CUSUM was developed. In this case, an one-sided upper CUSUM, C_i^+, is used to accumulate positive deviations from the target value (Equation (3.4)), while an one-sided lower CUSUM, C_i^-, accumulates negative deviations (Equation (3.5)) [12]:

$$C_i^+ = \max\left\{0, x_i - \left(\mu_0 + k \right) + C_{i-1}^+ \right\} \tag{3.4}$$

$$C_i^- = \max\left\{0, \left(\mu_0 - k \right) - x_i + C_{i-1}^- \right\} \tag{3.5}$$

In Equations (3.4) and (3.5) $C_0^+ = C_0^- = 0$ and k is a slack value usually selected as halfway between the target value μ_0 and the out-of-control value μ_1 that is of interest to be detected, i.e., $k = \left| \mu_1 - \mu_0 \right| / 2$ [12]. Whenever C_i^+ or C_i^- exceed a control limit (referred as a decision interval) the process is considered to be out-of-control. From Equations (3.4) and (3.5) one can observe that, for the tabular case, the CUSUM control charts only accumulate values from a window in the past where the cumulative sums have magnitudes exceeding the value of the slack constant, for the positive or negative sides. Therefore, this methodology avoids the accumulation of deviations considered not to be insignificant in the process.

3.1.2 Exponentially Weighted Moving Average Control Chart

The exponentially weighted moving average (EWMA) control chart was introduced by Roberts (1959) [42] and uses a weighted window that gradually gives lower weights to past observations according to [12],

$$z_i = \lambda x_i + (1-\lambda) z_{i-1} \tag{3.6}$$

where $0 < \lambda \leq 1$ is a forgetting factor and $z_0 = \mu_0$. As a consequence of the definition of Equation (3.6), the weight given to each observation decreases exponentially, with a decay speed determined by the value of the forgetting factor λ. For $\lambda = 1$ the EWMA chart resumes to the Shewhart control chart and for $\lambda \to 0$ it approaches the CUSUM control chart [38]. Typical values for λ are in the range of 0.05 to 0.25 [12]. Furthermore, as the monitored quantity z_i is an weighted average, the EWMA control chart presents some robustness to deviation from the normality assumption [12, 38].

To determine the control limits for the EWMA control chart, one take into account that in the cases where observations x_i are independent random variables with variance σ^2, the variance of z_i is given by [12],

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2-\lambda} \right) \left[1 - (1-\lambda)^{2i} \right] \tag{3.7}$$

which asymptotically leads to,

$$\sigma_z^2 = \sigma^2 \left(\frac{\lambda}{2-\lambda} \right) \tag{3.8}$$

since $\left[1 - (1-\lambda)^{2i} \right]$ converges to 1 as i increases [12]. The UCL and LCL for the EWMA control chart can then be set as,

$$\begin{aligned} UCL &= \mu_0 + L\sigma^2 \sqrt{\left(\frac{\lambda}{2-\lambda} \right) \left[1 - (1-\lambda)^{2i} \right]} \\ CL &= \mu_0 \\ LCL &= \mu_0 - L\sigma^2 \sqrt{\left(\frac{\lambda}{2-\lambda} \right) \left[1 - (1-\lambda)^{2i} \right]} \end{aligned} \tag{3.9}$$

where L is the width of the control limits, chosen along with λ in order to obtain a desired ARL performance [12].

3.2 Multivariate SPC

As the application of multiple parallel univariate control charts is not a suitable solution for monitoring several related variables, since the joint relationships between them is not explicitly considered, it becomes necessary to use multivariate control schemes. This class of approaches is here presented for the multivariate extension of the univariate control charts, namely the Hotelling's T^2 as well as for the multivariate CUSUM and multivariate EWMA. These monitoring control charts work relatively well for fairly small process (with only a few dozen of variables).

3.2.1 Hotelling's T^2

The natural extension of the univariate Shewhart control chart for monitoring multivariate processes is the Hotelling's T^2 control chart [19]. This chart assumes the process to be *i.i.d.* multivariate normal distributed, and the monitoring statistic, for single observations samples, is just the squared Mahalanobis distance between each multivariate observation and the overall reference mean. Assuming the mean and covariance matrix to be known, the monitoring statistic has the form [19, 46, 47]:

$$\chi_0^2 = \left(\mathbf{x} - \boldsymbol{\mu}_0\right)^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_0\right) \tag{3.10}$$

where $\mathbf{x}$ is a $(m \times 1)$ measurement vector, $\boldsymbol{\mu}_0$ is the $(m \times 1)$ population mean vector and $\boldsymbol{\Sigma}_0$ is the $(m \times m)$ in-control population covariance matrix. Under multivariate normal conditions, this statistic follows a central χ^2 distribution with m degrees of freedom. Therefore, a multivariate χ_0^2 control chart can be constructed by plotting χ_0^2 *versus* time with an UCL given by $\chi_{\alpha,m}^2$ where α is an appropriate level of significance (e.g. $\alpha = 0.01$) [19, 48].

When the in-control mean vector $\boldsymbol{\mu}_0$ and the covariance matrix $\boldsymbol{\Sigma}_0$ are unknown, they can be estimated from a sample of *n* past multivariate observations, using the usual well-known unbiased estimators of these population parameters, namely the sample mean and the sample covariance matrix [48]:

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \tag{3.11}$$

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{x}_i - \overline{\mathbf{x}}\right)\left(\mathbf{x}_i - \overline{\mathbf{x}}\right)^{\mathrm{T}} \tag{3.12}$$

In this case, when new multivariate observations are obtained, the Hotelling's T^2 statistic is given by [19, 47, 48],

$$T^2 = \left(\mathbf{x} - \overline{\mathbf{x}}\right)^{\mathrm{T}} \mathbf{S}^{-1}\left(\mathbf{x} - \overline{\mathbf{x}}\right) \tag{3.13}$$

whose control chart has the following UCL [16, 17, 47, 48]:

$$UCL = \frac{m(n-1)(n+1)}{n^2 - nm} F_{\alpha,m,n-m} \tag{3.14}$$

where $F_{\alpha,m,n-m}$ is the upper α percentile of the *F* distribution with *m* and $n-m$ degrees of freedom.

3.2.2 Multivariate CUSUM

The simplest extension of the CUSUM procedure for the multivariate case was proposed by Woodall and Ncube (1985) [49] and relies on the simultaneous application of a tabular CUSUM for each variable. Therefore, this approach is subjected to the same constrains associated with univariate control charts when applied to multivariate processes. Moreover, its performance is shown to be dependent of the faults direction, especially when the monitored variables are correlated [50]. As the direction of the faults cannot always be anticipated, the use of directionally invariant charts is preferable. This class of control charts is discussed below.

The first multivariate CUSUM was proposed by Crosier (1988) [44]. In this monitoring scheme the cumulative sum is computed as,

$$\mathbf{C}_i = \begin{cases} \mathbf{0} & \text{if } d_i \le k \\ \left(\mathbf{C}_{i-1} + \mathbf{x}_i - \boldsymbol{\mu}_0\right)\left(1 - \dfrac{k}{d_i}\right) & \text{otherwise} \end{cases} \tag{3.15}$$

$$d_i = \sqrt{\left(\mathbf{C}_{i-1} + \mathbf{x}_i - \boldsymbol{\mu}_0\right)^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} \left(\mathbf{C}_{i-1} + \mathbf{x}_i - \boldsymbol{\mu}_0\right)} \tag{3.16}$$

where $k > 0$ is a reference value, $\mathbf{C}_i = \mathbf{0}$, $\boldsymbol{\mu}_0$ is the $(m \times 1)$ target vector and $\boldsymbol{\Sigma}_0$ is the $(m \times m)$ in-control covariance matrix. The monitoring statistic is then defined as $MCUCUM_i = \sqrt{\mathbf{C}_i^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} \mathbf{C}_i}$ and is compared against a suitable control limit.

Later on, Pignatiello and Runger (1990) [50] proposed two alternative multivariate CUSUM control charts. The control chart showing the best ARL performance considers the multivariate sum given by [50],

$$\mathbf{C}_i = \sum_{j=i-n_i+1}^{i} \left(\mathbf{x}_i - \boldsymbol{\mu}_0\right) = n_i \left(\frac{1}{n_i} \sum_{j=i-n_i+1}^{i} \left(\mathbf{x}_i\right) - \boldsymbol{\mu}_0 \right) \tag{3.17}$$

where n_i is given by Equation (3.19). This cumulative sum is a measure of the difference between the sample mean based on the last n_i observations and its target value. The norm of these deviations can then be compared against a reference value k related with the allowed deviation, resulting in a monitoring statistic given by [50],

$$MC1_i = \max\left\{0, \sqrt{\mathbf{C}_i^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} \mathbf{C}_i} - n_i k\right\} \tag{3.18}$$

$$n_i = \begin{cases} n_{i-1} + 1 & \text{if } MC1_{i-1} > 0 \\ 1 & \text{otherwise} \end{cases} \tag{3.19}$$

The second multivariate CUSUM control chart suggested by Pignatiello and Runger (1990) [50] basically applies the one-sided upper CUSUM to the square of the distance between each observation and the target value, i.e.,

$$d_i^2 = \left(\mathbf{x}_i - \boldsymbol{\mu}_0\right)^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} \left(\mathbf{x}_i - \boldsymbol{\mu}_0\right) \tag{3.20}$$

$$MC2_i = \max\left\{0, MC2_{i-1} + d_i^2 - k\right\} \tag{3.21}$$

where, k is a reference value and $MC2_0 = 0$.

3.2.3 Multivariate EWMA

In order to improve the detection capability of the Hotelling's T^2 control chart, a multivariate EWMA procedure can be employed. The reasoning behind this approach is the same as for the univariate case and aims to increase the detection of small and moderate deviations in the process mean using information from the past in the monitoring scheme. This is done by application of a weighted average according to the procedure proposed by Lowry *et al.* (1992) [45]. In this case, the observations are weighted as,

$$\mathbf{z}_i = \lambda \mathbf{x}_i + (1-\lambda)\mathbf{z}_{i-1} \tag{3.22}$$

where $0 \le \lambda \le 1$ is a forgetting factor and $\mathbf{z}_0 = \mathbf{0}$. Then, a Hotelling's T^2 like statistic can be computed as,

$$T_i^2 = \mathbf{z}_i^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{z}_i}^{-1} \mathbf{z}_i^{\mathrm{T}} \tag{3.23}$$

where the covariance matrix is given by [12],

$$\boldsymbol{\Sigma}_{\mathbf{z}_i} = \frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]\boldsymbol{\Sigma}_0 \tag{3.24}$$

A natural extension of Equation (3.22) was presented by Hawkins *et al.* (2007) [51], who replaced λ by a smoothing matrix, $\mathbf{R}$, with non-zero diagonal elements, resulting in,

$$\mathbf{z}_i = \mathbf{R}\mathbf{x}_i + (\mathbf{I}-\mathbf{R})\mathbf{z}_{i-1} \tag{3.25}$$

Other extensions have also been proposed. For instance, Kramer and Schmid (1997) [52] developed a generalization of the multivariate EWMA for time dependent observations and Yumin (1996) [53] suggested the use of latent variables. A review of these methods can be found in [54].

3.3 Megavariate SPC

Current approaches for implementing large scale process monitoring, are essentially based on latent variable frameworks, given their intrinsic ability to deal with a high number of correlated variables. In fact, the classical full-rank MSPC approach, based on the Hotelling's T^2 statistic, is only rarely applied to processes with more than a dozen of variables, leaving outside its scope the vast majority of current industrial applications. This class of approaches is called megavariate SPC and in this section two representatives will be presented, namely the monitoring schemes based on PCA and PLS models.

3.3.1 Megavariate SPC Based on PCA

A common procedure for handling highly collinear process data in megavariate SPC consists in using a PCA model to describe the NOC behavior of process variables. As the number of retained principal components (PC) is low, say *p*, and they are uncorrelated by design, the Hotelling's T^2 procedure can be applied without any limitation [15, 16]. In this case, the following monitoring statistic is applied, after preliminarily centering all variables to zero mean, which might be also properly scaled:

$$T_{PCA}^2 = \sum_{i=1}^{p} \frac{t_i^2}{\lambda_i} = \mathbf{x}^{\mathrm{T}} \mathbf{P} \boldsymbol{\Lambda}_p^{-1} \mathbf{P}^{\mathrm{T}} \mathbf{x} \tag{3.26}$$

where $\boldsymbol{\Lambda}_p$ is a diagonal matrix with the first *p* eigenvalues in the main diagonal (the eigenvalue associated to a given PC, also provides the value for its variance in the data set used to estimate the model). If the process under normal operation conditions (NOC) follows a multivariate normal distribution, then the UCL for T_{PCA}^2 is given by [18, 47]:

$$UCL = \frac{p(n-1)(n+1)}{n^2 - np} F_{\alpha,p,n-p} \tag{3.27}$$

Where $F_{\alpha,p,n-p}$ is the upper α percentile of the *F* distribution, with *p* and $n-p$ degrees of freedom.

However, process monitoring strictly based on the PCA subspace, via T^2_{PCA} applied to the first *p* PCs is not sufficient, as it lacks a very important piece of information arising from the variability around the PCA subspace. This variability is captured by a monitoring statistic based on the squared prediction error (SPE) of the observations residuals, $\mathbf{e}_{m\times 1}$ [15]. This statistic is also known as the Q statistic:

$$Q = \mathbf{e}^{\mathrm{T}}\mathbf{e} = \left(\mathbf{x}-\hat{\mathbf{x}}\right)^{\mathrm{T}}\left(\mathbf{x}-\hat{\mathbf{x}}\right) = \mathbf{x}^{\mathrm{T}}\left(\mathbf{I}-\mathbf{P}\mathbf{P}^{\mathrm{T}}\right)\mathbf{x} \tag{3.28}$$

where $\hat{\mathbf{x}}$ stands for the projection of $\mathbf{x}$ onto the PCA subspace, i.e., it corresponds to the reconstruction in the original variables space, of the score "observed" in the latent variables subspace. The process is considered to be under statistic control, if this statistic is below its UCL [15, 55],

$$UCL = \theta_1\left(\frac{z_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0\left(h_0-1\right)}{\theta_1^2}\right)^{1/h_0} \tag{3.29}$$

where,

$$\theta_i = \sum_{j=p+1}^{m}\lambda_j^i, \quad i = 1,2,3 \tag{3.30}$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \tag{3.31}$$

where *p* is the number of retained principal components and z_α is the standard normal variable, corresponding to the upper $1-\alpha$ percentile.

3.3.2 Megavariate SPC Based on PLS

A similar megavariate SPC procedure can also be derived for PLS, using the two orthogonal components of variability, namely the variability in the PLS predictive subspace (PLS X-scores) and the residuals around such subspace or, alternatively, those regarding the prediction of the response variable(s) [17-19]. In this case, the Hotelling's T^2 statistic for a new score vector, $\mathbf{t}_{p\times 1}$, is given by [38]:

$$T_{PLS}^2 = \mathbf{t}^{\mathrm{T}} \mathbf{S}_{\mathbf{t}}^{-1} \mathbf{t} \tag{3.32}$$

where $\mathbf{S}_{\mathbf{t}}$ represents the estimated covariance matrix of the PLS X-scores. The UCL for T_{PLS}^2 is given by:

$$UCL = \frac{p(n-1)(n+1)}{n^2 - np} F_{\alpha,p,n-p} \tag{3.33}$$

where p is the number of latent variables retained in the PLS model and $F_{\alpha,p,n-p}$ is the upper α percentile of the F distribution with p and $n-p$ degrees of freedom. As to the SPE (or Q statistic), it can be either computed for the $\mathbf{X}$ or the $\mathbf{Y}$ block, according to the following expressions [17, 56]:

$$Q_{\mathbf{X}} = \left(\mathbf{x} - \hat{\mathbf{x}}\right)^{\mathrm{T}} \left(\mathbf{x} - \hat{\mathbf{x}}\right) \tag{3.34}$$

$$Q_{\mathbf{Y}} = \left(\mathbf{y} - \hat{\mathbf{y}}\right)^{\mathrm{T}} \left(\mathbf{y} - \hat{\mathbf{y}}\right) \tag{3.35}$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the reconstructed variables in $\mathbf{X}$ and $\mathbf{Y}$ original domains, using the PLS model defined by the Equations (2.6) and (2.7).

The control limits for the $Q_{\mathbf{X}}$ and $Q_{\mathbf{Y}}$ statistics, can be determined assuming an approximation to a $g \cdot \chi^2$ distribution: $UCL = g \cdot \chi_{\alpha,h}^2$ [38, 57]. This equation is also well approximated by [38]:

$$UCL = gh\left[1 - \frac{2}{9h} + z_\alpha \left(\frac{2}{9h}\right)^{1/2}\right]^3 \tag{3.36}$$

where g is a weighting factor and h the effective number of degrees of freedom for the χ^2 distribution, which can easily be obtained by matching the moments of these distribution with those from the empirical distribution, leading to $g = v/(2m)$ and $h = 2m^2/v$, where v is the variance and m the mean of the SPE values ($Q_{\mathbf{X}}$ and $Q_{\mathbf{Y}}$); z_α is the standard normal variable, corresponding to the upper $1-\alpha$ percentile.

3.4 Multiscale Principal Component Analysis

The main advantages of multiscale principal component analysis (MS-PCA) arise from the combination of the PCA power to diagonalize the variables correlation with the wavelet transformation ability to diagnozalize their autocorrelation. Moreover, it allows for a better description of the monitored process, since data collected from most industrial process typically present multiscale features [9].

The first step in the MS-PCA methodology proposed by Bakshi (1998) [9] is the decomposition of each variable into multiple scales by application of the wavelet transform as,

$$\mathbf{WX} = \begin{bmatrix} \mathbf{H}_{J_{max}} & \mathbf{G}_{J_{max}} & \cdots & \mathbf{G}_j & \cdots & \mathbf{G}_1 \end{bmatrix}^{\mathrm{T}} \mathbf{X} \tag{3.37}$$

where $\mathbf{W}$ is an $(n \times n)$ orthogonal matrix representing the orthogonal wavelet transforms. $\mathbf{G}_j = \mathbf{G}\prod_{j-1}\mathbf{H}$ is the $\left(2^{\log_2(n-j)} \times n\right)$ matrix containing wavelet filters coefficients corresponding to scale $j = 1, 2, \ldots, J_{max}$ and $\mathbf{H}_{J_{max}} = \prod_{J_{max}} \mathbf{H}$ is the matrix of scaling-function filter coefficients at the coarsest scale. J_{max} is the wavelet decomposition depth. For the case of the Haar wavelet transform, the filters are given by [9, 58],

$$\mathbf{H} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \tag{3.38}$$

$$\mathbf{G} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \tag{3.39}$$

Other wavelet filters can be found in the literature [59, 60].

From this decomposition, it is possible to define, for each variable, the approximation coefficients as,

$$\mathbf{a}_{J_{max}} = \mathbf{H}_{J_{max}} \mathbf{x} \tag{3.40}$$

and the detail coefficients as,

$$\mathbf{d}_j = \mathbf{G}_j \mathbf{x} \quad j = 1, 2, \ldots, J_{max} \tag{3.41}$$

These wavelet coefficients are then independently monitored by a MSPC-PCA procedure. In order to maintain an on-line approach, a moving window of length $2^{J_{max}}$ is used to compute the wavelet coefficients and only the ones related to the current observation are inspected. To do so, a PCA model is determined for each scale and during the monitoring stage each scale is monitored by the corresponding $T^{2}_{PCA}{}^{(j)}$ and $Q^{(j)}$ statistics for $j = 1, 2, \ldots J_{max} + 1$. Given the increase in the number of control charts (2 for each scale, in a total of $2(J_{max}+1)$ control charts), their detection limits need to be adjusted in order to maintain the desired overall false alarm rate (α). These corrected limits can be obtained through the application of the Bonferroni inequality, which leads to a false alarm rate for each control chart (α_A) given by [9, 58],

$$\alpha_A = \frac{\alpha}{2\left(J_{max}+1\right)} \tag{3.42}$$

After selecting the relevant scales (i.e., the ones where the $T^{2}_{PCA}{}^{(j)}$ or $Q^{(j)}$ statistics are above their control limits), the data is reconstructed back to the original time domain using only such scales. Finally, the actual state of the process is assessed by analysis of the reconstructed signal, based only on the wavelet coefficients that violate their control limits. The general steps behind this monitoring scheme are represented in Figure 3.3.

By application of MS-PCA, a filtering procedure is performed before checking if the process remains under normal operation conditions [9]. As a result, only the scales related with the fault are analyzed and therefore a fault oriented SPC filter with multiple resolutions is applied.

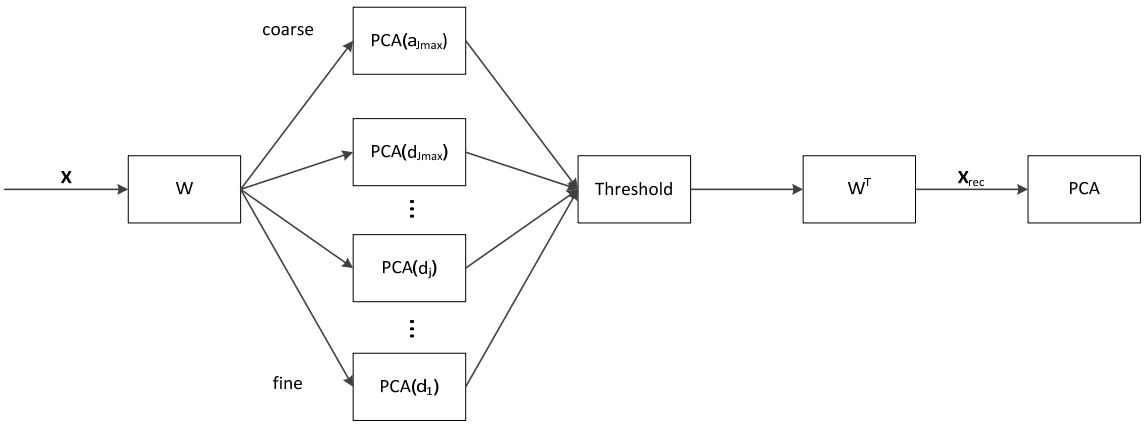


Figure 3.3 Schematic representation of the MS-PCA procedure. **X** is the original signal and $\mathbf{X}_{rec}$ is the reconstructed signal based on the relevant wavelet coefficients.

4 Statistical Process Control of the Correlation Structure

The optimized and safe operation of current industrial processes requires the simultaneous monitoring of multiple related variables. To do so, a variety of multivariate statistical process control (MSPC) methods, namely control charts, have been developed and applied in order to determine whether the process is only subjected to common causes of variability or if a special or assignable cause, related with some abnormality inside or outside the process, has occurred. Among these procedures, one can refer the Hotelling's T^2 chart [43], the multivariate CUSUM [44] and multivariate EWMA control charts [45], presented in the previous chapter.

Analyzing the literature, one can verify that most multivariate process monitoring methodologies developed so far, including the latent variables methodologies [15, 16, 27, 61-63] and state-space or time series approaches [64, 65], are essentially non-causal and focused on detecting changes in the process mean [66-69]. Regarding the procedures developed for monitoring the process covariance, two main approaches are employed: (i) monitoring subgroups of observations or, equivalently, non-overlapping windows (off-line case) and (ii) monitoring individual observations (on-line case).

As for the off-line approaches, the most widely adopted ones are based on the generalized variance (determinant of the covariance matrix), for which several approaches were proposed, namely by Alt (1984) [70], Aparisi *et al.* (2001) [71] and Djauhari (2005) [72]. However, the generalized variance is a rather ambiguous measure of multivariate variability, as quite different covariance matrices can lead to similar values for the determinant. As an alternative to the generalized variance, Guerrero-Cusumano (1995) [73] proposed the conditional entropy and Djauhari *et al.* (2008) [74] the vector variance, which is the sum of the squares of all eigenvalues of the sample covariance matrix. Other approaches are based on the likelihood ratio test (LRT), as for example those found in the works of Alt and Smith (1988) [75] and Levinson *et al.* (2002) [76]. More recently, Yen and Shiau (2010) [77] presented a control chart, based on LRT, specifically designed to detect an increase in process dispersion, which was later extended to variations in both

directions (increase and decrease) [69]. A different procedure was also proposed by Tang and Barnett (1996) [78], based on the decomposition of the sample covariance matrix into various independent components.

The task of monitoring individual multivariate observations is substantially more complex, since the sample covariance matrix is not defined in such cases. For that purpose Yeh *et al.* (2005) [79] developed a procedure based on the univariate EWMA that allows the recursive estimation of the sample covariance matrix. Then, they monitor the squared distances between the estimates obtained and the in-control covariance matrix. However, this EWMA recursion scheme assumes the process mean to be constant. To extend its application to cases where the process mean may change, Huwang *et al.* (2007) [80] employed a similar EWMA recursion that includes an estimation of the process mean and compared both procedures through the monitoring of the trace of the covariance matrix estimates. In a similar approach, Reynolds and Cho (2006) [81] proposed to monitor the Mahalanobis distance of the EWMA recursion applied to the squared deviations from target. Later on, Hawkins and Maboudou-Tchao (2008) [82] also applied the EWMA recursion of the covariance matrix to extend Alt's likelihood ratio statistic to individual observations.

These monitoring statistics will be described more thoroughly in the following sections for the off-line case (based on non-overlapping windows) as well as for on-line case (based on individual observations). From this revision, it can be verified that all the monitoring statistics mentioned above are based only on the marginal covariance matrix and do not take into consideration the inherent structure of the process.

4.1 Off-line Monitoring Statistics for the Process Dispersion

In this section the current methodologies for monitoring the process dispersion based on subgroups of non-overlapping windows are reviewed. These approaches will constitute the benchmarks against which the performance of the methodologies proposed in this thesis will be assessed and compared in Chapter 10.

4.1.1 *W* Statistic

Alt and Smith (1988) [75] presented three procedures for monitoring process variability by following its marginal covariance. One of the schemes is a direct extension of the univariate S^2 control chart [48] and is equivalent to successively performing statistical hypothesis tests of the form, H_0: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ *vs.* H_1: $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$. The monitoring statistic is based on the likelihood ratio test, and is defined as,

$$W = -m(n-1) - (n-1)\ln\left(|\mathbf{S}|/|\boldsymbol{\Sigma}_0|\right) + (n-1)\,\mathrm{tr}\left(\boldsymbol{\Sigma}_0^{-1}\mathbf{S}\right) \tag{4.1}$$

where m is the number of variables, n is the number of observations, $\boldsymbol{\Sigma}_0$ is the in-control covariance matrix and $\mathbf{S}$ is the sample covariance matrix. Anderson (2003) [46] showed that W is asymptotically distributed as $\chi^2_{m(m+1)/2}$, and therefore the process dispersion is considered to be out-of-control if W exceeds the $\mathrm{UCL} = \chi^2_{m(m+1)/2,\alpha}$ ($\chi^2_{\nu,\alpha}$ is the $100\times(1-\alpha)^{\text{th}}$ percentile of a chi-squared distribution with ν degrees of freedom).

4.1.2 |S| Control Charts

The other two approaches presented by Alt and Smith (1988) [75] are based on the sample generalized variance (i.e., the determinant of the sample covariance matrix, $|\mathbf{S}|$), which is a widely used measure of multivariate dispersion. One of these approaches makes use of the sample distribution of $|\mathbf{S}|$. Under Gaussian conditions, the distribution of $|\mathbf{S}|$ is equal to the distribution of $|\boldsymbol{\Sigma}_0|/(n-1)^m$ times the product of m independent factors, where the i^{th} factor is distributed as χ^2_{n-i}, i.e., [46]

$$|\mathbf{S}| = |\boldsymbol{\Sigma}_0|(n-1)^{-m}\prod_{i=1}^{m}\chi^2_{n-i} \tag{4.2}$$

For the case of two variables, it can be shown that $|\mathbf{S}|$ is distributed as $|\boldsymbol{\Sigma}_0|\left(\chi^2_{2n-4}\right)^2 \Big/ \left[4(n-1)^2\right]$. Consequently, the control limits for the $|\mathbf{S}|$-chart with two variables are [70, 75],

$$LCL = |\boldsymbol{\Sigma}_0|\left(\chi^2_{2n-4,1-\alpha/2}\right)^2 \Big/ \left[4(n-1)^2\right]$$
$$UCL = |\boldsymbol{\Sigma}_0|\left(\chi^2_{2n-4,\alpha/2}\right)^2 \Big/ \left[4(n-1)^2\right] \tag{4.3}$$

When more than two variables are available, one can use Equation (4.2) or its normal approximation presented by Anderson (2003) [46] to compute the control limits.

Finally, the third |**S**|-chart presented by Alt and Smith (1988) [75] uses only the first two moments of |**S**| and the property that most of the distribution of |**S**| is confined in the interval $E(|\mathbf{S}|) \pm 3\sqrt{\text{var}(|\mathbf{S}|)}$, where $E(|\mathbf{S}|) = b_1 |\boldsymbol{\Sigma}_0|$ and $\text{var}(|\mathbf{S}|) = b_2 |\boldsymbol{\Sigma}_0|^2$, with,

$$b_1 = (n-1)^{-m} \prod_{i=1}^{m} (n-i)$$
$$b_2 = (n-1)^{-2m} \prod_{i=1}^{m} (n-i) \left[\prod_{j=1}^{m} (n-j+2) - \prod_{j=1}^{m} (n-j) \right] \tag{4.4}$$

From this result, it follows that the control limits for this control chart are given by [70],

$$LCL = |\boldsymbol{\Sigma}_0|\left(b_1 - 3b_2^{1/2}\right)$$
$$UCL = |\boldsymbol{\Sigma}_0|\left(b_1 + 3b_2^{1/2}\right) \tag{4.5}$$

However, since |**S**| is positive definite, it is not meaningful to have a negative LCL, and therefore the LCL is usually set to zero.

The only difference between these two |**S**|-charts procedures lies in the control limits adopted: in the former case they are probability limits, Equation (4.3), while in the latter case, they are 3-sigma limits, Equation (4.5).

4.1.3 *G* Statistic

Kramer and Jensen (1969) [83] proposed an alternative method to test the equality of two population matrices, through the so called *G* statistic. This hypothesis test is based on the idea that, in a stable process, the covariance matrix estimated with the complete data set of collected data, should be approximately equal to the one obtained from the mean square of successive differences. This concept was later explored by Levinson *et al.*

(2002) [76] in order to verify if any change has occurred in the process covariance matrix over time. To obtain the monitoring statistics one has to calculate first the in-control sample covariance matrix, $\mathbf{S}_0$, from a reference data set with n_0 observations. After that, the i^{th} test sample covariance matrix, $\mathbf{S}_{1,i}$, determined from a subgroup with n_1 observations is combined with $\mathbf{S}_0$ in the pooled estimator of the covariance [76],

$$\mathbf{S}_{pool,i} = \frac{(n_0 - 1)\mathbf{S}_0 + (n_1 - 1)\mathbf{S}_{1,i}}{n_0 + n_1 - 2} \tag{4.6}$$

The monitoring statistic is then computed as $G_i = gM_i$, where,

$$M_i = (n_0 + n_1 - 2)\ln\left|\mathbf{S}_{pool,i}\right| - (n_0 - 1)\ln\left|\mathbf{S}_0\right| - (n_1 - 1)\ln\left|\mathbf{S}_{1,i}\right| \tag{4.7}$$

$$g = 1 - \left(\frac{1}{n_0 - 1} + \frac{1}{n_1 - 1} - \frac{1}{n_0 + n_1 - 2}\right)\left(\frac{2m^2 + 3m - 1}{6(m+1)}\right) \tag{4.8}$$

When the process is under statistical process control, the G statistic is distributed as $\chi^2_{m(m+1)/2}$. Therefore, its control limits can be established as [76],

$$\begin{aligned} LCL &= \chi^2_{m(m+1)/2,\alpha/2} \\ UCL &= \chi^2_{m(m+1)/2,1-\alpha/2} \end{aligned} \tag{4.9}$$

4.1.4 *E* Statistic

The previous monitoring statistics (Sections 4.1.1 to 4.1.3) are mainly based on some form of the generalized variance, which is somewhat insensitive to the correlation structure of the variables and therefore some changes may be masked or pass undetected. To address this issue, Guerrero-Cusumano (1995) [73] proposed the use of a conditional entropy measure. The entropy of a vector is a measure of the dispersion of its values and, for a continuous m-multivariate random variable $\mathbf{x}$, it is defined as [73],

$$H(\mathbf{x}) = -\int f(\mathbf{x})\ln f(\mathbf{x})\,d\mathbf{x} = E\left[-\ln f(\mathbf{x})\right] \tag{4.10}$$

where $f(\mathbf{x})$ is the probability density of $\mathbf{x}$.

If **x** follows a normal distribution with mean **μ** and covariance **Σ**, then the entropy is given by [73],

$$\begin{aligned} H(\mathbf{x}) &= \frac{1}{2} m \ln(2\pi e) + \frac{1}{2} \ln|\boldsymbol{\Sigma}| \\ &= \frac{1}{2} m \ln(2\pi e) + \frac{1}{2} \ln|\boldsymbol{\Sigma}_d^2| + \frac{1}{2} \ln|\mathbf{P}_0| \\ &= \frac{1}{2} m \ln(2\pi e) + \frac{1}{2} \sum_{i=1}^{m} \ln(\sigma_i^2) - T(\mathbf{x}) \end{aligned} \tag{4.11}$$

where $\mathbf{P}_0$ is the correlation matrix, $\boldsymbol{\Sigma}_d = \text{diag}(\sigma_i)$ is a diagonal matrix with σ_i in its main diagonal, σ_i is the standard deviation of the i^{th} variable and $T(\mathbf{x})$ is the mutual information. By estimation of σ_i through the sample standard deviation, and assuming that the mutual information is known, the sample entropy is given by,

$$\hat{H}(\mathbf{x}) = \frac{1}{2} m \ln(2\pi e) + \frac{1}{2} \sum_{i=1}^{m} \ln(s_i^2) - T(\mathbf{x}) \tag{4.12}$$

The difference between the sample and theoretical entropy, $\delta = \hat{H}(\mathbf{x}) - H(\mathbf{x}) = 1/2 \sum_{i=1}^{m} \ln(s_i^2/\sigma_i^2)$, is then considered a conditional entropy, since it is conditioned on $\mathbf{P}_0$. Based on this, the monitoring statistic, E, is defined as,

$$E = k\delta = \sqrt{\frac{n-1}{2m}} \sum_{i=1}^{m} \ln\left(\frac{s_i^2}{\sigma_i^2}\right) \tag{4.13}$$

where $k = [2(n-1)/m]^{1/2}$ is a normalization constant.

The control limits of E are calculated by [73],

$$\begin{aligned} LCL &= kp\left[\Phi'\left(\frac{n-1}{2}\right) - \ln\left(\frac{n-1}{2}\right)\right] - z_{1-\alpha/2} k \sqrt{m\Phi''\left(\frac{n-1}{2}\right) + \frac{2}{n-1}\text{tr}(\mathbf{P}_0 - \mathbf{I})^2} \\ UCL &= kp\left[\Phi'\left(\frac{n-1}{2}\right) - \ln\left(\frac{n-1}{2}\right)\right] + z_{1-\alpha/2} k \sqrt{m\Phi''\left(\frac{n-1}{2}\right) + \frac{2}{n-1}\text{tr}(\mathbf{P}_0 - \mathbf{I})^2} \end{aligned} \tag{4.14}$$

where $\Phi'(\cdot)$ and $\Phi''(\cdot)$ are the first and second derivative of the natural logarithm of the gamma function.

Although the E statistic is much simpler than the statistics based on the generalized variance, it requires the constancy of $\mathbf{P}_0$. Therefore one must verify if $\mathbf{P}_0$ indeed remains under statistical control, before testing E.

4.1.5 *VMAX* statistic

A simpler and more efficient control chart than the |**S**|-chart was proposed by Costa and Machado (2009) [84]. They proposed the use of the *VMAX* statistic, which is the maximum value of the sample variances of the data after normalization, i.e.,

$$VMAX = \max\left\{s_1^2, s_2^2, \ldots, s_m^2\right\} \tag{4.15}$$

where $s_i^2 = \mathbf{z}_i^{\mathrm{T}}\mathbf{z}_i / n - 1$ and $\mathbf{z}_i = \left(\mathbf{x}_i - \boldsymbol{\mu}_i\right)/\sigma_i$. The normalization of the data is an important step on this monitoring statistic since it guarantees that all the sample variances have the same probability of exceeding a certain UCL. The upper control limit for a two dimensional process is obtained through,

$$\alpha = 1 - \int_0^{n \cdot UCL} \Pr\left[\chi^2_{n,\left(t\rho^2/1-\rho^2\right)} < \frac{n \cdot UCL}{1-\rho^2}\right] \frac{1}{2^{n/2}\Gamma\left(n/2\right)} e^{-t/2} t^{(n/2)-1} dt \tag{4.16}$$

where $\chi^2_{n,\left(t\rho^2/1-\rho^2\right)}$ is the non-central chi-square distribution with n degrees of freedom and non-centrality parameter given by $\left(t\rho^2/1-\rho^2\right)$, and ρ is the correlation between the variables.

When compared to the generalized variance, the *VMAX* statistic presents a faster detection performance and better diagnostic features. The reasoning underlying this statistic was later on applied to the simultaneous monitoring of the mean and covariance matrix [85], and to sample ranges [86].

4.2 On-line Monitoring Statistics for the Process Dispersion

In this section a review of the methodologies for monitoring the process multivariate dispersion based on single observations is provided. The monitoring statistics will be grouped according to their base monitoring principle and proposing authors. In Chapter 11, the performance of these monitoring schemes will be compared against the methodologies proposed in this thesis for on-line monitoring.

4.2.1 Cumulative Deviations

In order to monitor the process multivariate dispersion through an on-line procedure, Yeh *et al.* (2005) [79] applied an EWMA recursive scheme to obtain updated estimates of the covariance matrix. They have considered data following a multivariate normal distribution, where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are assumed to be known and, without loss of generality, equal to $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}_m$. The covariance matrix estimation scheme proposed by these authors is given by,

$$\mathbf{S}_t = \lambda \mathbf{z}_t \mathbf{z}_t^{\mathrm{T}} + (1-\lambda)\mathbf{S}_{t-1} \tag{4.17}$$

where $0 < \lambda < 1$ is a forgetting factor. $\mathbf{S}_t$ is a positive definite matrix when $t \geq m$ and therefore it is a reasonable estimator of the covariance matrix. After estimation, the covariance matrix is monitored by defining $\mathbf{S}_{1,t}$ and $\mathbf{S}_{2,t}$ as the stacked diagonal $(m \times 1)$ and non-diagonal elements $(m(m-1)/2 \times 1)$ of $\mathbf{S}_t$ respectively, i.e.,

$$\mathbf{S}_{1,t} = \left[s_{(1,1),t}, s_{(2,2),t}, \ldots, s_{(m,m),t} \right]^{\mathrm{T}} \tag{4.18}$$

$$\mathbf{S}_{2,t} = \left[s_{(1,2),t}, s_{(1,3),t}, \ldots, s_{(i,j),t}, \ldots, s_{(m-1,m),t} \right]^{\mathrm{T}}, \quad (i < j) \tag{4.19}$$

The vector $\mathbf{S}_{1,t}$ is a vector with estimates of the m population variances, which is $\mathbf{1}_{m\times 1}$, and $\mathbf{S}_{2,t}$ is an estimator of the vector of $(m(m-1)/2 \times 1)$ population covariance's, which is $\mathbf{0}_{(m(m-1)/2)\times 1}$. The deviations from target values are squared and accumulated as follows,

$$D_{1,t} = \sum_{i=1}^{m} \left(s_{(i,i),t} - 1 \right)^2 \tag{4.20}$$

$$D_{2,t} = \sum_{i=1<j=m} \left(s_{(i,i),t} - 0 \right)^2 \tag{4.21}$$

The monitoring statistic is then given by the maximum of the standardized cumulated shift in the variances and covariances,

$$MaxD_t = \max \left\{ \frac{D_{1,t} - \mu_{D_1}}{\sigma_{D_1}}, \frac{D_{2,t} - \mu_{D_2}}{\sigma_{D_2}} \right\} \tag{4.22}$$

The respective asymptotic moments are given by [79],

$$\begin{aligned}
\mu_{D_1} &= 2m \frac{\lambda}{2-\lambda} \\
\sigma^2_{D_1} &= m \left[\frac{48\lambda^4}{1-(1-\lambda)^4} + \frac{8\lambda^2}{(2-\lambda)^2} \right] \\
\mu_{D_2} &= \frac{m(m-1)}{2} \frac{\lambda}{2-\lambda} \\
\sigma^2_{D_2} &= m(m-1)(2m-1) \frac{\lambda^4}{1-(1-\lambda)^4} + m(m-1) \frac{\lambda^2}{(2-\lambda)^2}
\end{aligned} \tag{4.23}$$

This control chart signals a fault whenever *MaxD* exceeds a predetermined UCL.

4.2.2 Trace of the Covariance Matrix

Huwang *et al.* (2007) [80] investigated the performance of monitoring statistics based on the trace of the covariance matrix estimated by an EWMA recursion. One of such monitoring statistics is given by,

$$MEWMS_t = \frac{\mathrm{tr}(\mathbf{S}_t) - \mu_{\mathrm{trS}}}{\sigma_{\mathrm{trS}}} \tag{4.24}$$

where,

$$\begin{aligned}
\mu_{\mathrm{trS}} &= m \\
\sigma^2_{\mathrm{trS}} &= 2m \left[(1-\lambda)^{2(t-1)} + \sum_{i=2}^{t} \lambda^2 (1-\lambda)^{2(t-i)} \right]
\end{aligned} \tag{4.25}$$

and is compared to a LCL and UCL in order to verify if a change on the covariance matrix occurred. In this case, the covariance matrix $\mathbf{S}_t$ is estimated by Equation (4.17) where the variables are pre-transformed as $\mathbf{z} = \boldsymbol{\Sigma}_0^{-1/2}\left(\mathbf{x} - \boldsymbol{\mu}_0\right)$, so that the in-control $\mathbf{z}$ is distributed as $N(\mathbf{0},\mathbf{I}_m)$.

The $MEWMS_t$ statistic assumes that the process mean does not shift during the monitoring period. To mitigate this limitation, the same authors included an estimation stage of the process mean on Equation (4.17), resulting on a new updating scheme of the covariance matrix given by [80],

$$\mathbf{V}_t = \lambda\left(\mathbf{z}_t - \mathbf{y}_t\right)\left(\mathbf{z}_t - \mathbf{y}_t\right)^{\mathrm{T}} + \left(1-\lambda\right)\mathbf{V}_{t-1} \tag{4.26}$$

$$\mathbf{y}_t = \omega\mathbf{z}_t + \left(1-\omega\right)\mathbf{y}_{t-1} \tag{4.27}$$

where $0 < \lambda < 1$, $\mathbf{V}_0 = \left(\mathbf{z}_1 - \mathbf{y}_1\right)\left(\mathbf{z}_1 - \mathbf{y}_1\right)^{\mathrm{T}}$, $0 < \omega < 1$ and $\mathbf{y}_0 = \mathbf{0}$. For $t \geq m$, $\mathbf{V}_t$ is a positive definite matrix and $E\left(\mathbf{V}_t\right) \to 2(1-\omega)^2/(2-\omega)\boldsymbol{\Sigma}$ as $t \to \infty$. Therefore $(2-\omega)/[2(1-\omega)^2]\mathbf{V}_t$ can be used to estimate $\boldsymbol{\Sigma}$ and hence monitored by,

$$MEWMV_t = \frac{\mathrm{tr}\left(\mathbf{V}_t\right) - \mu_{\mathrm{trV}}}{\sigma_{\mathrm{trV}}} \tag{4.28}$$

The first and second moments of $\mathrm{tr}(\mathbf{V}_t)$ can be found on the original work of Huwang *et al.* (2007) [80].

4.2.3 L_1-norm and L_2-norm of Variance Deviations

Memar and Niaki (2009) [87], proposed the use of a sum of deviations of variances from their target value based on the L_1-norm and L_2-norm. When applied to the covariance matrix estimated by Equation (4.17), $\mathbf{S}_t$, the monitoring statistics become [87],

$$D_{1,S} = \sum_{i=1}^{m}\left|s_{(i,i),t} - 1\right| \tag{4.29}$$

$$D_{2,S} = \sum_{i=1}^{m}\left(s_{(i,i),t} - 1\right)^2 \tag{4.30}$$

The same applies for the covariance matrix estimated by Equation (4.26), $\mathbf{V}_t$ [87],

$$D_{1,V} = \sum_{i=1}^{m} \left| \frac{2-\omega}{2(1-\omega)^2} v_{(i,i),t} - 1 \right| \tag{4.31}$$

$$D_{2,V} = \sum_{i=1}^{m} \left(\frac{2-\omega}{2(1-\omega)^2} v_{(i,i),t} - 1 \right)^2 \tag{4.32}$$

All four monitoring statistics signal an alarm when they fall outside their UCL.

4.2.4 Likelihood Ratio Test

The off-line generalized likelihood ratio statistic proposed by Alt (1984) [70],

$$W = -m(n-1) - (n-1)\ln\left(|\mathbf{S}|/|\boldsymbol{\Sigma}_0|\right) + (n-1)\operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}\mathbf{S}\right) \tag{4.33}$$

where $\mathbf{S}$ is the $(m \times m)$ sample covariance matrix of a rational subgroup with n observations, was extended by Hawkins and Maboudou-Tchao (2008) [82] to individual observations. On their work, the data is previously transformed, as $\mathbf{z} = \mathbf{B}(\mathbf{x} - \boldsymbol{\mu}_0)$, where $\mathbf{B}$ is any matrix that satisfies $\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^{-1} = \mathbf{I}_m$. Then, an on-line estimation of $\mathbf{S}_t$ is obtained through Equation (4.17), with $\mathbf{S}_0 = \mathbf{I}_m$. The covariance matrix is monitored by Alt's likelihood ratio defined as (unneeded constants are omitted):

$$c_t = \operatorname{tr}(\mathbf{S}_t) - \log\left(|\mathbf{S}_t|\right) - m \tag{4.34}$$

and compared to an UCL.

4.2.5 Mahalanobis Distance of Squared Deviations

Reynolds and Cho (2006) [81] suggested a set of control charts based on the Mahalanobis distance of the EWMA recursion applied to the squared deviations from target. On their methodology, for each variable, the EWMA statistic of standardized deviations from target is defined as [81],

$$EZ_{i,t}^2 = \lambda z_{i,t}^2 + (1-\lambda) EZ_{i,t-1}^2, \quad i = 1,\ldots,m \tag{4.35}$$

where $EZ_{i,0}^2 = 1$, $0 < \lambda < 1$ is a weighting or tuning parameter and $z_{i,t}$ is the standardized observation (i.e., $z_{i,t} = (x_{i,t} - \mu_i)/\sigma_i$). The in-control covariance of $\left[EZ_{1,t}^2 \quad \ldots \quad EZ_{m,t}^2 \right]^{\mathrm{T}}$ is given by [81],

$$\boldsymbol{\Sigma}_{EZ^2} = 2\left(\frac{\lambda}{2-\lambda}\right)\left(\boldsymbol{\Sigma}_{Z0} \odot \boldsymbol{\Sigma}_{Z0}\right) \tag{4.36}$$

where $\odot$ is the Hadamard product (element by element product) and $\boldsymbol{\Sigma}_{Z0}$ is the in-control covariance matrix of the standardized variables. The Mahalanobis distance of $EZ_{i,t}^2$ is given by:

$$M_1 Z_t^2 = \left[EZ_{1,t}^2 - 1 \quad \ldots \quad EZ_{m,t}^2 - 1 \right] \left(\boldsymbol{\Sigma}_{EZ^2} \right)^{-1} \left[EZ_{1,t}^2 - 1 \quad \ldots \quad EZ_{m,t}^2 - 1 \right]^{\mathrm{T}} \tag{4.37}$$

This statistic is used with a UCL and detects both increases and decreases in variance. However, the authors reported that, for large values of λ, the detection of decreases in variance may become difficult. To improve the ability to detect these decreases they suggested the use of,

$$M_2 Z_t^2 = \left[EZ_{1,t}^2 \quad \ldots \quad EZ_{m,t}^2 \right] \left(\boldsymbol{\Sigma}_{EZ^2} \right)^{-1} \left[EZ_{1,t}^2 \quad \ldots \quad EZ_{m,t}^2 \right]^{\mathrm{T}} \tag{4.38}$$

with an LCL and UCL.

The performance of the M_1Z^2 and M_2Z^2 statistics changes substantially when regressed-adjusted variables are used instead of the original variables. The vector of regressed-adjusted variables in this case is given by [81],

$$\mathbf{a} = \left(\mathrm{diag}\left(\boldsymbol{\Sigma}_{Z0}^{-1} \right) \right)^{-1/2} \boldsymbol{\Sigma}_{Z0}^{-1} \mathbf{z} \tag{4.39}$$

where $\mathrm{diag}\left(\boldsymbol{\Sigma}_{Z0}^{-1} \right)$ is a diagonal matrix with the same diagonal elements as $\boldsymbol{\Sigma}_{Z0}^{-1}$. From this transformation, the in-control covariance matrix of the transformed variables becomes $\Sigma_{A0} = \left(\mathrm{diag}\left(\boldsymbol{\Sigma}_{Z0}^{-1} \right) \right)^{-1/2} \boldsymbol{\Sigma}_{Z0}^{-1} \left(\mathrm{diag}\left(\boldsymbol{\Sigma}_{Z0}^{-1} \right) \right)^{-1/2}$. Thus, the monitoring statistics are given as,

$$M_1 A_t^2 = \left[EA_{1,t}^2 - 1 \quad \ldots \quad EA_{m,t}^2 - 1 \right] \left(\boldsymbol{\Sigma}_{EA^2} \right)^{-1} \left[EA_{1,t}^2 - 1 \quad \ldots \quad EA_{m,t}^2 - 1 \right]^{\mathrm{T}} \tag{4.40}$$

$$M_2 A_t^2 = \begin{bmatrix} EA_{1,t}^2 & \dots & EA_{m,t}^2 \end{bmatrix} \left(\boldsymbol{\Sigma}_{EA^2} \right)^{-1} \begin{bmatrix} EA_{1,t}^2 & \dots & EA_{m,t}^2 \end{bmatrix}^{\mathrm{T}} \tag{4.41}$$

by replacing $z_{i,t}$ by $a_{i,t}$ and $\boldsymbol{\Sigma}_{Z0}$ by $\boldsymbol{\Sigma}_{A0}$ in Equations (4.35) to (4.38).

4.2.6 Variable Transformation Based on the Singular Wishart Distribution

While the previous methods (Sections 4.2.1 to 4.2.4) resort to an EWMA recursion scheme to estimate the covariance matrix, a different approach based on a single observation and the properties of the singular Wishart distribution was proposed by Bodnar *et al.* (2009) [88]. In this methodology, the properties of the estimation $\mathbf{M}_t = \mathbf{x}_t \mathbf{x}_t^{\mathrm{T}}$ are used to transform $\mathbf{M}_t$ into a set of vectors following a multivariate Gaussian distribution and then, the conventional monitoring statistics for the mean are applied to monitor each of the obtained vectors.

The transformation is carried by defining $\sigma_{(i,i)}$ as the (i,i) element of $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_{1,i}$ as the i^{th} column of $\boldsymbol{\Sigma}_0$ without $\sigma_{(i,i)}$ and $\boldsymbol{\Sigma}_{2,i}$ as a $(m-1)\times(m-1)$ matrix obtained by deleting the i^{th} row and the i^{th} column of $\boldsymbol{\Sigma}_0$. The same decomposition is performed in $\mathbf{M}_t$, resulting in $m_{(i,i),t}$, $\mathbf{M}_{1,i,t}$ and $\mathbf{M}_{2,i,t}$, respectively. The new set of variables is given by [88],

$$\boldsymbol{\eta}_{i,t} = \boldsymbol{\Sigma}_{2,i}^{-1/2} \left(\mathbf{M}_{1,i,t} / m_{(i,i),t} - \boldsymbol{\Sigma}_{1,i} / \sigma_{(i,i)} \right) m_{(i,i),t}^{1/2} \sim N_{m-1} \left(\mathbf{0}_{m-1}, \mathbf{I}_{m-1} \right) \tag{4.42}$$

resulting in a total of m vectors $\boldsymbol{\eta}_{i,t}$ that need to be monitored individually. To do so, the authors suggested monitoring each $\boldsymbol{\eta}_{i,t}$ vector by standard multivariate EWMA and CUSUM control charts as follows.

On the CUSUM schemes, they considered the methodology of Crosier (1988) [44] defined by,

$$\begin{aligned} c_{i,t} &= \left\| \mathbf{Q}_{i,t-1} + \boldsymbol{\eta}_{i,t} \right\| \\ \mathbf{Q}_{i,t} &= \begin{cases} \mathbf{0} & \text{if } c_{i,t} \le k \\ \left(\mathbf{Q}_{i,t-1} + \boldsymbol{\eta}_{i,t} \right) \left(1 - \dfrac{k}{c_{i,t}} \right) & \text{otherwise} \end{cases} \end{aligned} \tag{4.43}$$

where $\mathbf{Q}_{i,0} = \mathbf{0}$ and $k > 0$ is a reference value. For each vector $\boldsymbol{\eta}_{i,t}$ the $MCUSUM_{i,t}$ is determined as the norm of the vector $\mathbf{Q}_{i,t}$,

$$MCUSUM_{i,t} = \left(\mathbf{Q}_{i,t}^{\mathrm{T}}\mathbf{Q}_{i,t}\right)^{1/2} = \max\left\{0, c_{i,t} - k\right\} \tag{4.44}$$

The individual control chart signals an alarm if any of the $MCUSUM_{i;t}$ statistics exceeds some preselected critical value, and therefore a unique control chart is obtained by considering,

$$MCUSUM_{t} = \max\left\{MCUSUM_{i,t}\right\} \tag{4.45}$$

Another multivariate CUSUM scheme applied was the one proposed by Pignatiello and Runger (1990) [50] based on,

$$MC_{t} = \max\left\{MC_{i,t}\right\} \tag{4.46}$$

where

$$MC_{i,t} = \max\left\{0, MC_{i,t-1} + D_{i,t}^{2} - m - k\right\} \tag{4.47}$$

with $D_{i,t}^{2} = \boldsymbol{\eta}_{i,t}^{\mathrm{T}}\boldsymbol{\eta}_{i,t}$ and $MC_{i,0} = 0$.

Regarding the EWMA approaches, Bodnar *et al.* (2009) [88] used the schemes proposed by Lowry *et al.* (1992) [45], by defining $\mathbf{z}_{i,t} = \lambda\boldsymbol{\eta}_{i,t} + (1-\lambda)\mathbf{z}_{i,t-1}$ with $\mathbf{z}_{i,0}$ = **0** and considering both exact and asymptotic variances of $\mathbf{z}_{i,t}$. Subsequently, the monitoring statistics for each of the $\boldsymbol{\eta}_{i,t}$ vector become,

$$Q_{i,t} = \frac{2-\lambda}{\lambda\left(1-(1-\lambda)^{2t}\right)}\mathbf{z}_{i,t}^{\mathrm{T}}\mathbf{z}_{i,t} \tag{4.48}$$

for the exact variance and

$$Qa_{i,t} = \frac{2-\lambda}{\lambda}\mathbf{z}_{i,t}^{\mathrm{T}}\mathbf{z}_{i,t} \tag{4.49}$$

for the asymptotic variance.

The join EWMA statistics are then defined as,

$$MEWMA_{t} = \max\left\{Q_{i,t}\right\} \tag{4.50}$$

$$MEWMAa_t = \max\left\{Qa_{i,t}\right\} \tag{4.51}$$

The same authors also proposed monitoring the Mahalanobis distance of $\boldsymbol{\eta}_{i,t}$ through an EWMA recursion scheme. For this monitoring scheme, the individual statistics are given by,

$$QM_{i,t} = \lambda D_{i,t}^2 + (1-\lambda) QM_{i,t-1} \tag{4.52}$$

with $D_{i,t}^2 = \boldsymbol{\eta}_{i,t}^{\mathrm{T}} \boldsymbol{\eta}_{i,t}$ and $QM_{i,0} = m - 1$. Finally, the individual statistics are combined as,

$$MEWMAM_t = \max\left\{QM_{i,t}\right\} \tag{4.53}$$

Part III – MSPC – On-line Monitoring of the Process Mean Tendency

5 Defining the Structure of DPCA models

Statistical process control (SPC) methodologies have been routinely applied in many different industrial contexts, from laboratories to discrete manufacturing industries and chemical processing industries. With the increasing availability of data through faster and more informative sensors and measurement systems, the dynamic or autocorrelated nature of systems became an aspect that must be incorporated into SPC methodologies. The usual *i.i.d.* assumption for the definition of the normal operation conditions (NOC) region is no longer valid under these circumstances, and several alternative methodologies were proposed to the classic univariate [40, 41, 89], multivariate [43-45] and megavariate [15-17, 19] approaches. These can be organized into three distinct classes of methods: (i) methods based on correcting/adjusting control limits for the existent SPC methods, using knowledge of the specific dynamic model underlying data generation [90]; (ii) methods based on time series modeling followed by the monitoring of one-step-ahead prediction residuals [91, 92]; (iii) methods based on time-domain variable transformations, that diagonalize, in an approximate way, the autocorrelation matrix of process data [5, 9].

The first class of approaches (i), is restricted to very particular situations (univariate processes with rather simple dynamic structures), for which correction formulas were derived and made available. As to the time series based approach (ii), an usually criticism concerns the difficulty of defining proper time series model structures (the specification problem), which requires a significant amount of expertise. Perhaps even more important than this, the fact that estimating classic multivariate time series models (e.g., VARMA, VARIMA) for small-medium sized systems (> 10 variables) is a complex or maybe unfeasible task, which limits their use in practice. Finally, the third class of approaches (iii) does provide effective solutions to the autocorrelation problem, but its implementation requires a high load of computational programming. The current lack of software packages through which such methods can be conveniently made available, has been hindering their diffusion into practical applications.

However, an alternative approach has quickly gained popularity, given its conceptual simplicity and relationship with a well-known and accepted technique: SPC using dynamic principal component analysis (DPCA) [27]. DPCA is a methodology proposed

by Ku *et al.* (1995) [27], which essentially attempts to model the autocorrelation structure present in data, through a "time lag shift" method. This method consists in including time-shifted replicates of the variables under analysis, in order to simultaneously capture the static relationships and the dynamic structure, through the application of standard PCA. DPCA has been applied in different application scenarios, that include not only multivariate process control and fault diagnostic [27, 93-95] but also maintenance activities planning [96] and sensitivity analysis [97]. On a different context, DPCA was also applied in economical forecasts after the initial work of Brillinger (1964) [98]; other related applications include the construction and analysis of economic indicators [99] and volatility modeling [100].

A key point in the implementation of the DPCA method is the selection of the number of lags to be used, i.e. the number of time-shifted versions for each variable to include in the DPCA model. This problem is similar to selecting the lag structure in time series models (ARMA, ARIMA, ARMAX, etc.) [28, 29]. The solution proposed by Ku *et al.* (1995) [27] consists in implementing parallel analysis, a technique that combines the scree plot obtained from a PCA analysis applied to the collected data, with the scree plot resulting from the analysis of a random data set of the same size. The interception of these two curves represents the cut-off for the selection of the number of components to retain. This is followed by the analysis of the correlations exhibited by the scores, in order to determine the number of linear relationships present in data. The underlying reasoning is that the scores corresponding to low magnitude eigenvalues correspond to the existence of linear relationships (static and/or dynamic), involving the variables under analysis, including their time-shifted versions. Such scores should also be approximately uncorrelated, as the authors illustrated with resource to several examples. Time-shifted variables are added until no additional linear relationships are detected. The existence of a new linear relationship is verified through the difference between the number of low magnitude eigenvalues (associated with uncorrelated scores) obtained with the addition of a new time-shifted variable, and the expected number of such coefficients assuming that the previous lag structure was correct.

Other approaches to the lag-selection problem were also proposed. Autoregressive (AR) models were employed to determine the number of lags to use in DPCA [101]. In this case, the authors suggested the application of an AR model only to the output variable, from which a single lag is proposed for all the input variables. This is a very simple

approach that does not explicitly incorporate the relationships between variables. Wachs and Lewin (1999) [102] proposed the delay-adjusted PCA, that determines the most appropriated time delays, between inputs and outputs variables, by shifting inputs until their correlation with the outputs is maximized (maximum of the cross-correlation function). This approach assumes a two block variable structure (**X** and **Y**), where the output variables are correlated among themselves with no delays present, and inputs are independent of each other. The authors point out that this may not always be true, especially when analyzing closed-loop data. Guerfel *et al.* (2009) [103] proposed an approach where the number of lags is selected as the minimum number needed for detecting a specific fault, therefore requiring *a priori* knowledge of possible systems faults. Other proposed methods result from identification techniques based on Akaike information criterion, such as those employed by Li and Qin (2001) [94] and by Russell *et al.* (2000) [95]. However, the first approach assumes a two block variable structure (**X** and **Y**) and both methodologies propose a unique delay structure for all variables, which may not be true in general.

To addresses this current major weakness of the DPCA methodology, which constitutes a central problem in the implementation of the method in real world application scenarios, it is proposed in this thesis a new method to determine, in a more rigorous way, not only the maximum number of shifts to adopt in DPCA models, but also the specific lag structure for each variable. Therefore, contrary to the works published so far, the number of time-shifts used for describing the dynamic behavior of each variable can be different. Furthermore, no explicit segmentation in input/output variables is strictly required. The advantages of adopting the proposed method in different process system activities, such as process monitoring and system identification are also illustrated in this chapter.

5.1 Defining the Lagged Structure in DPCA

In this section, the benchmark method for selecting the number of time-shifted replicates to include in the extended matrix is briefly reviewed, and the new alternative methodologies presented. The benchmark is the method most extensively used in the literature dealing directly with the lag structure definition problem, which is the one proposed by Ku *et al.* (1995) [27].

5.1.1 Defining the Lagged Structure in DPCA: Ku *et al.* Method

Ku *et al.* (1995) [27] pointed out that the eigenvectors corresponding to the last principal components with small eigenvalues, represent exact or near-exact linear relationships between the original variables [20]. This characteristic was explored by these authors who proposed an algorithm based on the identification of the number of linear relationships needed to describe the system, including those involving time-shifted variables with a given number of lags (l). The extended data matrix ($\tilde{\mathbf{X}}$) has, in this case, the simple form of a Hankel matrix. The presence of linear relationships, originated from static or dynamic relationships, manifests themselves through two types of effects: (i) small eigenvalues in the spectral decomposition of the covariance matrix of $\tilde{\mathbf{X}}$, (ii) and by the fact that the corresponding scores should be, in these conditions, approximately independent (a feature that can be checked through, for instance, auto- and cross-correlation plots). The pseudo-code for the algorithm proposed by the authors is presented in Table 5.1 [27].

The above procedure assumes the implementation of a methodology for selecting the number of principal components, for which the authors suggest the use of parallel analysis followed by the analysis of cross-correlation plots of the principal component scores. The number of linear relationships identified is given by the difference between the number of variables considered in the extended matrix and the number of principal components retained.

Table 5.1 Pseudo-code for the lag-selection methodology proposed by Ku *et al.* (1995) [27].

1. Set $l = 0$;
2. Form the extended data matrix $\tilde{\mathbf{X}} = \left[\mathbf{x}(0)\ \mathbf{x}(1) \cdots \mathbf{x}(l)\right]$;
3. Perform PCA and calculate all the principal scores;
4. Set $j = m \times (l+1)$ and $r(l) = 0$;
5. Determine if the j^{th} component represents a linear relation. If yes proceed, if no go to step 7;
6. Set $j = j - 1$ and $r(l) = r(l) + 1$, repeat step 5;
7. Calculate the number of new relationships:
$$r_{new}(l) = r(l) - \sum_{i=0}^{l-1} (l - i + 1) r_{new}(i)$$
8. If $r_{new}(l) \leq 0$, go to step 10, otherwise proceed;
9. Set $l = 1 + 1$, go to step 2;
10. Stop.

According to the authors, the number of lags obtained by this procedure is usually 1 or 2, depending on the order of the dynamic system, and is the same for all variables. However, they refer that, in the case of non-linear systems, *l* could be set to higher values in order to get a better linear approximation of the non-linear relationships [27].

This methodology will be adopted as the benchmark method against which the proposed approaches will be tested and compared. In this decision, it is taken into account the fact that it is eventually the most well-known and widely used approach for addressing this problem, which furthermore is theoretical driven and thoroughly tested. Other less tested methods available in the literature either present similar limitations or require the consideration of two blocks of variables (please refer to the review presented in the beginning of Chapter 5).

5.1.2 Defining the Coarse Lagged Structure in DPCA: Method 1 – Selection of the Maximum Number of Lags

In this section, an alternative way to select a single number of time-shifts for all variables is proposed. In this case, the extended matrix to be used in DPCA has also the form of a Hankel matrix. This method can be used separately, or in a first stage preceding the implementation of Method 2, to be described in the next section, which will refine the number of shifts to consider for each individual variable.

The proposed method has roots in the work of Ku *et al.* (1995) [27] and is focused on following the singular values obtained in each stage, where the extended matrix is augmented with the variables replicates for an additional time-shift. By analyzing the sequence of the singular values, one can estimate the point (i.e., the number of lags), after which the introduction of additional time-shifted variables become redundant, i.e., they are no longer necessary for explaining the existent stationary and dynamic relationships.

The proposed method assumes the existence of a total of *m* linear dynamic relationships to be identified (as many as the variables involved), whose order is not known *a priori*, but is at least 1 (the simplest dynamic model). This simple hypothesis allows for the derivation of an algorithm that consistently leads to better estimates of the dynamic lag-structure involved, as illustrated in Section 5.2. The algorithm consists in sequentially

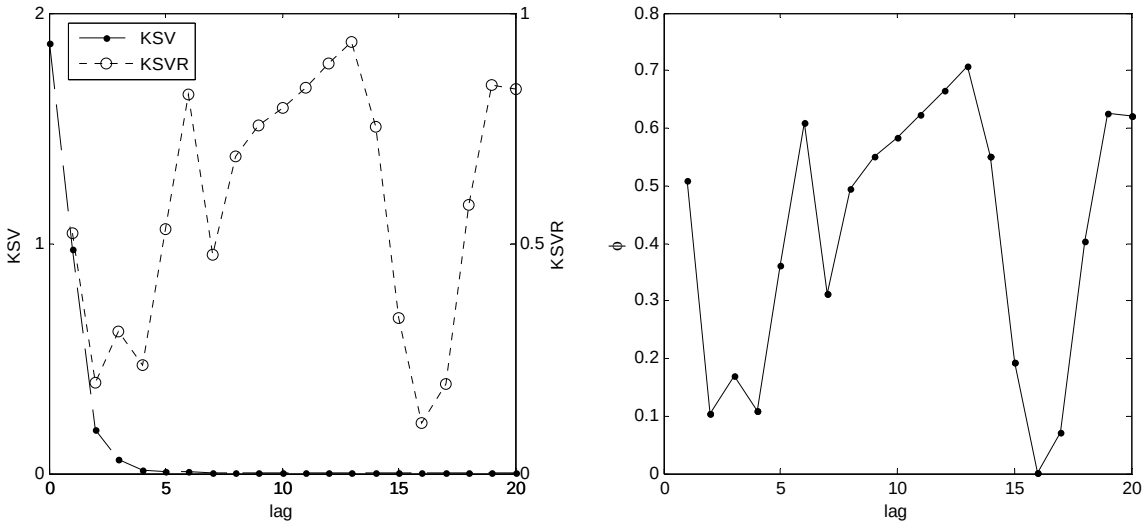
introducing an additional set of time-shifted replicates for all original variables (which corresponds to the consideration of one more lag in the extended matrix), after which the singular values are computed for the corresponding covariance matrix. This procedure is repeated until a pre-defined upper limit on the number of lags to use is achieved, $l_{\max}$, where it stops (this limit is usually a number high enough for allowing the description of all the dynamic features present in the data, but it can be adjusted if, during the analysis, it is concluded that it was underestimated initially). In each stage, l (which also coincides with the number of lags introduced in the extended matrix), the following quantities are computed from the singular values:

- ***Key Singular Value***. The key singular value in the l^{th} stage ($l \geq 1$), $KSV(l)$, is defined as the $(m \times l + 1)^{\text{th}}$ singular value, after sorting the set of singular values according to the decreasing order of their magnitude. As there are m linear dynamic relationships to be identified, by hypothesis, only after introducing the number of lags necessary to model the linear dependency with the highest delay (i.e., that goes further into the past, or having the maximum number of lags in the DPCA model), one is expected to get a small magnitude for the singular value in such a position. Therefore, the key singular value should signal the point where the linear relationship that requires the larger number of lags to be properly described was finally achieved. The other relationships, requiring a lower number of lags, give rise to multiple low singular values (one per additional lag, after the numbers of lags necessary to fully describe their linear relations are achieved). All of them appear at the end of the ordered vector of singular values. It can be shown that, only after the point where there is a sufficient number of lags to describe all linear descriptions present, a small value appears in the KSV, indicating that the addition of variable replicates with more lags is no longer relevant for the DPCA model.

- ***Key Singular Value Ratio***. Under the assumption of having m dynamic relationships present, a small value for the KSV at stage l, indicates that one has attained the point where no more lags are necessary to be added. However, due to the presence of noise and small-moderate non-linear effects, the identification of this condition is not always clear. Therefore, a second element in the algorithm is introduced, in order to increase the robustness of the detection of the maximum number of lags required. In fact, following the behavior of the successive values obtained for the $KSV(l)$,

$(l = 1, 2, \ldots)$, it can be observed that there is a point where it decays more or less sharply (corresponding to the stage where *KSV*(*l*) starts getting low values) and then becomes approximately constant (see the left plot in Figure 5.1, for an example). Therefore, by defining the Ratio of successive Key Singular Values at stage *l*, *KSVR*(*l*), by $KSVR(l) = KSV(l) / KSV(l-1)$, one can capture this behavior more efficiently. Moreover, from this definition, one can verify that the required number of lags should have a low value for *KSVR*, indicating that a significant decrease in the *KSV* has just occurred, i.e., the current singular value is significant lower than the previous one. After this point, the ratio tends to have values closer to 1 and to be approximately constant.

In resume, the maximum number of lags to be considered in the extended matrix for implementing DPCA should obey the following two criteria: (i) have a small *KSV* and (ii) have a low value for *KSVR*.



(a) (b)

Figure 5.1 *KSV* and *KSVR* obtained in the analysis of the Wood and Berry case study (a). The analysis of the parameter *KSVR*, leads, in this case, to an estimated maximum number of lags of 16. Also shown, is the objective function for the auxiliary optimization problem for selecting the number of lags (b).

In order to find the number of lags that match both of these conditions, a procedure is implemented that seeks for the number of lags introduced, *l*, for which *KSV* and *KSVR* are closer to their minimums – the minimums attained individually in the analysis, i.e., min(*KSV*) and min(*KSVR*). This task is performed by minimizing the objective function, *Distance To Optimum*, ϕ, given by Equation (5.1), where KSV_N and $KSVR_N$ are

normalized versions of *KSV* and *KSVR*, constructed in order to remove the effects of scale and provide equal weight to both criteria in the analysis (Equations(5.2) and (5.3)).

$$\phi = \sqrt{KSV_N(l)^2 + KSVR_N(l)^2} \tag{5.1}$$

$$KSV_N(l) = \frac{KSV(l) - \min(KSV)}{\max(KSV) - \min(KSV)} \tag{5.2}$$

$$KSVR_N(l) = \frac{KSVR(l) - \min(KSVR)}{\max(KSVR) - \min(KSVR)} \tag{5.3}$$

Thus, with this additional criteria available, one can select the number of lags simply by finding out the minimum of the objective function ϕ, Equation (5.1). The pseudo-code for the proposed algorithm for estimating the maximum number of lags to use in DPCA, is presented in Table 5.2. The plot in the right hand side of Figure 5.1 illustrates this objective function for the Wood and Berry case study.

Table 5.2 Pseudo-code for the Method 1: selection of the maximum number of lags to use in DPCA.

1. Set $l = 0$;
2. Form the extended data matrix $\tilde{\mathbf{X}} = \left[\mathbf{x}(0)\ \mathbf{x}(1) \cdots \mathbf{x}(l)\right]$;
3. Perform the singular value decomposition of the covariance of the extended matrix: $\mathbf{\Sigma}_{\tilde{\mathbf{X}}} = \mathbf{USV}^{\mathrm{T}}$;
4. Set $KSV(l) = s_{ml+1}$;**(1)**
5. If $l > 0$ set $KSVR(l) = KSV(l)/KSV(l-1)$;
6. If $l < l_{max}$, set $l = l + 1$ and go to step 2, otherwise proceed;
7. Normalize *KSV* and *KSVR*;
8. Determine: $\arg\min_{l \in [1, lmax]} \sqrt{KSV_N(l)^2 + KSVR_N(l)^2}$, s.t. $l \geq l^*$ (where l^* is the first l, such that $KSVR(l) < KSVR(l-1)$);**(2)**

Notes: **(1)** s_{ml+1} is the $(ml+1)^{\text{th}}$ singular value of $\mathbf{\Sigma}_{\tilde{\mathbf{X}}}$. **(2)** A justification for this condition will be provided in Section 5.2.2.1.

5.1.3 Defining the Fine Lagged Structure in DPCA: Method 2 – Selection of the Number of Lags for Each Variable

Method 1 provides an approach for selecting a single number of lags to be used for all variables (as in the benchmark method). Such lag is the one corresponding to the dynamic relationship requiring a longer tail into the past, in terms of the number of lags involved. However, analyzing different multivariate systems, one can verify that quite often the dynamic order is not the same for all variables. Therefore, the number of lags required to describe them will also be different. Under these circumstances, it is both opportune and important to devise a methodology for fine tuning the number of lags to be adopted for each variable, in order to increase de accuracy and stability of the DPCA approach. In order to obtain such a finer selection of the number of lags for each variable, a second algorithm (Method 2) is proposed. This algorithm presents some similarities with Method 1, but includes, in each stage, a variable-wise analysis.

In Method 2, at each stage k, m versions of the extended matrix are analyzed, $\tilde{\mathbf{X}}^1(k), \tilde{\mathbf{X}}^2(k), \ldots, \tilde{\mathbf{X}}^m(k)$. These matrices are obtained from that relative to the preceding stage, $\tilde{\mathbf{X}}(k-1)$, modified by the inclusion of a single time-shifted variable, with one more lag than the number of lags used in the preceding stage for the same variable. Let us define $l^{(k)}$, as the m-dimensional vector containing in its entries the number of lags considered for each variable in stage k: $\lambda_i^{(k)}$ ($i = 1, \ldots, m$). In other words, the first entry of $l^{(k)}$ contains the number of shifted versions for x_1, $\lambda_1^{(k)}$, the second entry contains the number of shifted version for x_2, $\lambda_2^{(k)}$, and so on and so forth until the m^{th} entry:

$$l^{(k)} = \begin{bmatrix} \lambda_1^{(k)} & \lambda_2^{(k)} & \cdots & \lambda_m^{(k)} \end{bmatrix} \tag{5.4}$$

Additionally, let us consider the m-dimensional indicator vector, $\delta^{(k)}$, as a vector composed by zeros, except for the k^{th} position, where it has a 1:

$$\delta^{(k)} = \begin{bmatrix} \overset{1}{\overset{\frown}{0}} & \overset{2}{\overset{\frown}{0}} & \cdots & \overset{k}{\overset{\frown}{1}} & \cdots & \overset{m}{\overset{\frown}{0}} \end{bmatrix} \tag{5.5}$$

In this circumstances, the lag structure corresponding to the i^{th} version of the extended matrix at stage k, $\tilde{\mathbf{X}}^i(k)$, summarized in the vector $l_i^{(k)}$, is given by (see also Figure 5.2 for more details about this process):

$$l_i^{(k)} = l^{(k-1)} + \delta^{(i)} \rightarrow \tilde{\mathbf{X}}^i(k) \tag{5.6}$$

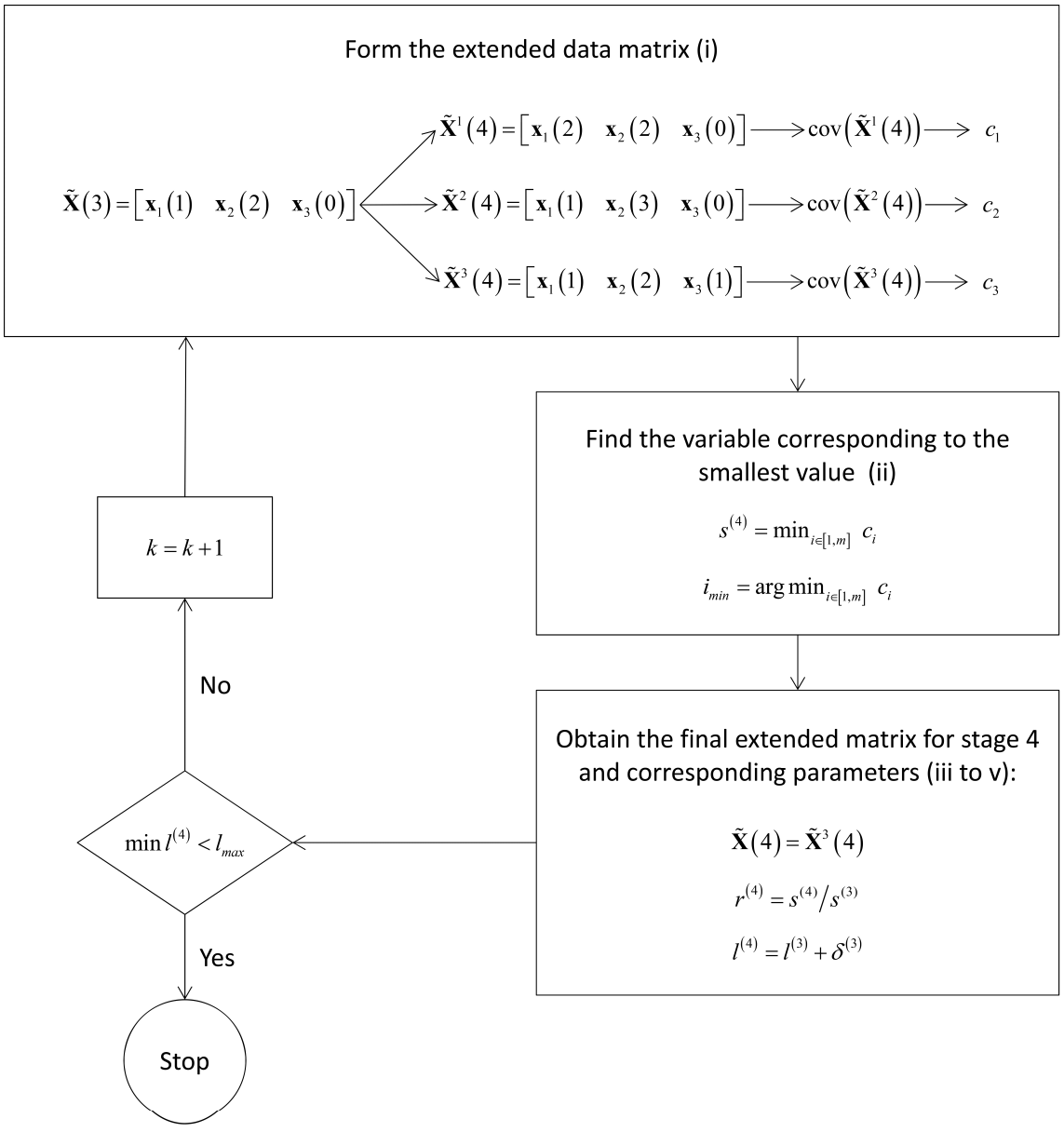


Figure 5.2 Illustration of the implementation of Method 2. In this example, the stage 4 of a 3 variables system is presented. On the previous stages 1 lag was selected for $\mathbf{x}_1$ and 2 lags to $\mathbf{x}_2$. In the current stage, the several versions of the extended data matrices are defined (three in this case), and their singular values determined. Based on the analysis of quantities computed from them, variable $\mathbf{x}_3$ is selected to be incorporated into the extended matrix with an additional time-shift.

Now, for each version of the extended matrix, $\tilde{\mathbf{X}}^i(k)$, the smallest singular value of the corresponding covariance matrix, $\text{cov}\left(\tilde{\mathbf{X}}^i(k)\right)$, is determined and saved as c_i. This set of values, c_i (i = 1, …, m), are then analyzed in order to find the minimum at stage k, say $(s^{(k)})$. The lagged variable to be incorporated in the extended matrix is the one for which such minimum was found. This will result in the extended matrix for stage k, $\tilde{\mathbf{X}}(k)$, and the procedure is repeated again for stage $k+1$, where another lagged version of some variable is added again (which can be any of the m variables under consideration). This process continuous until the maximum number of lags to analyze is attained. The total number of stages will always be equal to the sum of the lags for all variables, as in each stage a single lag is introduced for some variable under analysis. The maximum number of lags to analyze, $l_{\max}$, is a parameter that can be either provided after the implementation of Method 1, or selected in a more conservative way (i.e., slightly above of what it is expected to be a reasonable value for this parameter).

In order to remove potentially redundant lagged variables that might be included in this forward addition process, a final pruning stage is performed, where the results obtained at all stages are analyzed for their significance and improvement of the objective function, using a similar criteria to the one presented before in Equation (5.1) (the only difference lies in the redefinition of the normalization factors). The complete procedure for implementing Method 2 is summarized in the pseudo-code presented in Table 5.3.

Table 5.3 Pseudo-code for the algorithm that estimates the fine lag structure of the extended matrix, for implementing DPCA (Method 2).

1. Set stage $k = 0$, and initialize the lag vector (whose i^{th} entry, contains the number of lags for the corresponding i^{th} variable), $l^{(k)} = [0 \quad \cdots \quad 0]$;
2. Form $\tilde{\mathbf{X}}(0)$ and determine the smallest singular value ($s^{(0)}$);
3. Set $k = k + 1$;
   i. For i = 1 to m
      - Form $\tilde{\mathbf{X}}^i(k)$, whose lagged structure is given by the vector: $l_i^{(k)} = l^{(k-1)} + \delta^{(i)}$;
      - Compute the covariance matrix of $\tilde{\mathbf{X}}^i(k)$;
      - Determine the smallest singular value of the covariance matrix obtained in the previous step: c_i;

   ii. Find the variable corresponding to the smallest value of the set $\{c_i\}_{i=1,\ldots,m}$: $s^{(k)} = \min_{i\in[1,m]} c_i$; **(1)**

   iii. Obtain the final extended matrix for stage k: $\tilde{\mathbf{X}}(k)$; with $l^{(k)} = l^{(k-1)} + \delta^{(i_{min})}$ where $i_{min} = \arg\min_{i\in[1,m]} c_i$; **(1)**

   iv. Set $r^{(k)} = s^{(k)} / s^{(k-1)}$;

   v. If $\min l^{(k)} < l_{max}$, go to step 3, otherwise proceed;
4. Determine the stage that provides the best description of the linear dynamics involved: $k^* = \arg\min_{k\in[1,k_{\max}]} \phi(k) = \sqrt{\left(\tilde{s}^{(k)}\right)^2 + \left(\tilde{r}^{(k)}\right)^2}$ **(2)**, s.t. $l \geq l^*$ (where l^* is the first l, such that $r(l) < r(l-1)$ **(3)**).

Notes: **(1)** in ii), $s^{(k)}$ contains the minimum singular values, whereas in iii), i_{min} corresponds to the index for the minimum value.
(2) $\tilde{y}$ represents the normalized score corresponding to y.
(3) A justification for the condition used in stage 4 will be provided in Section 5.2.2.1.

5.2 Comparative Assessment Study

To demonstrate the increased lag estimation accuracy of the proposed methodologies, two testing scenarios were considered. In the first scenario, a large number of systems from the same class were randomly generated, and then the benchmark and Method 1 were employed to estimate the appropriate maximum number of lags necessary to describe the dynamic relationship for each realization of the model structure. In the second scenario, several multivariate systems found in the literature, with known dynamics, are employed, in order to test the estimation accuracy performance of Method 2 in selecting the specific number of lags for each variable.

5.2.1 Systems with Random Lagged Structure

To assess the accuracy performance of the methods presented in Section 5.1, they were employed in the estimation of the number of lags for a large number of systems with randomly generated lagged structures. The systems under study were based on the following continuous first order dynamic transfer function with time delay, defined by

$$g = \frac{Ke^{-\theta s}}{\tau s + 1} \tag{5.7}$$

where K is the system gain, τ is the time constant and θ is the time delay. A large number of different realizations of this set of parameters were generated (following independent uniform distributions), which will imply different time lags in the corresponding discrete models.

Following this procedure, 5000 SISO systems and another 5000 MIMO 2×2 systems were generated, all of them subjected to additive white noise (d) with different magnitudes of signal-to-noise ratio (SNR) and noise structures (with and without autocorrelation). The SNR is defined by:

$$SNR = 10\log_{10}\left(\frac{\mathrm{var}(x)}{\mathrm{var}(d)}\right) \tag{5.8}$$

The deviations obtained between the estimated and the true maximum delay of the system (equivalent to the maximum lag), are presented graphically in Figures 5.3 to 5.6.

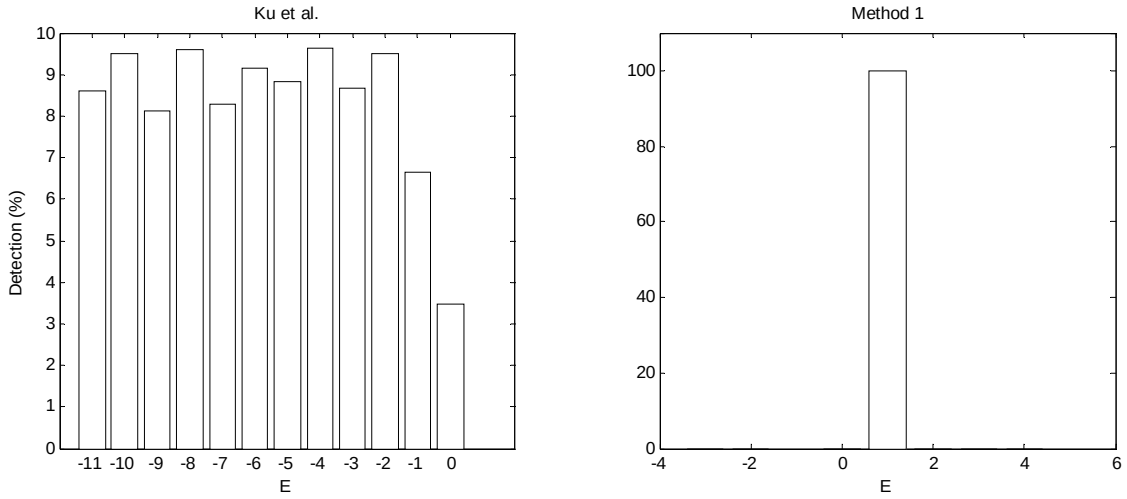


Figure 5.3 Graphical representation of the deviation between the estimated and the true number of lags (E = Estimated - True) for SISO systems, corrupted with additive white noise (20 dB), without autocorrelation.

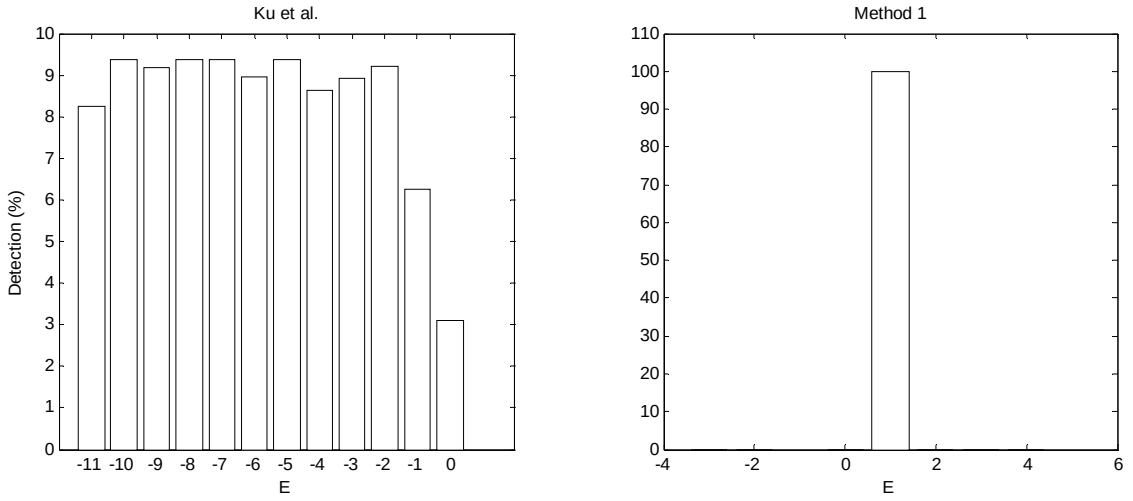


Figure 5.4 Graphical representation of the deviation between the estimated and the true number of lags (E = Estimated - True) for SISO systems, corrupted with additive autocorrelated noise (20 dB).

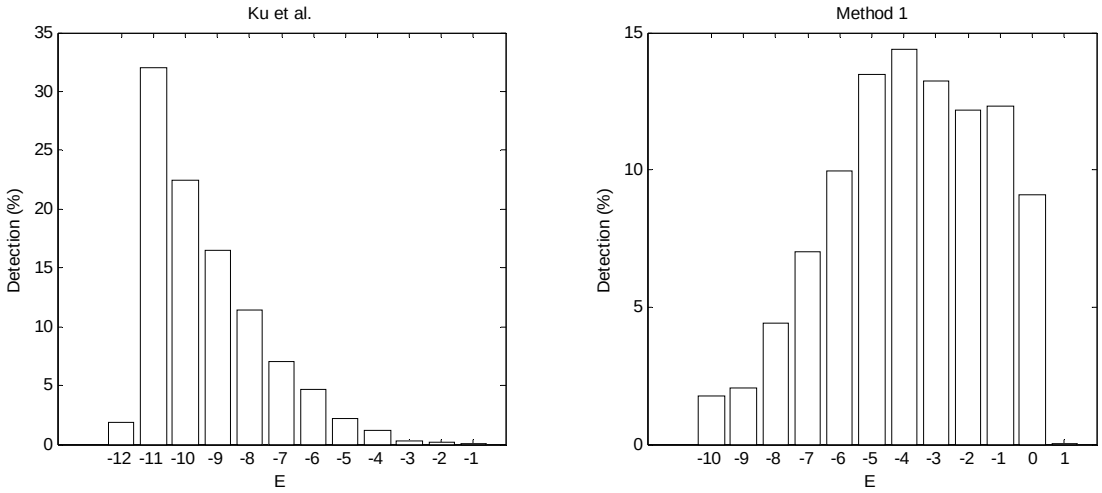


Figure 5.5 Graphical representation of the deviation between the estimated and true number of lags (E = Estimated - True) for MIMO 2×2 systems, corrupted with additive white noise (20 dB), without autocorrelation.

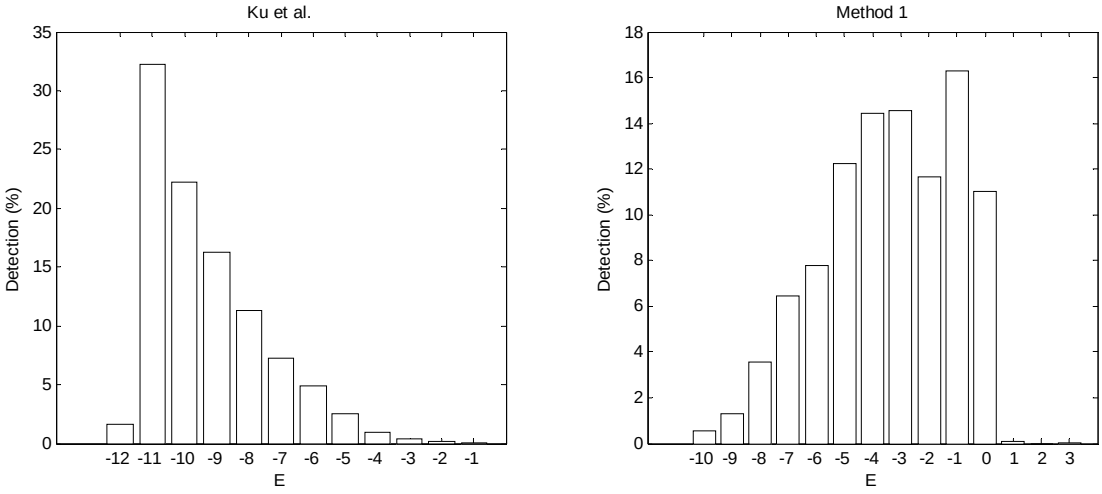


Figure 5.6 Graphical representation of the deviation between the estimated and true number of lags (E = Estimated - True) for MIMO 2×2 systems, corrupted with additive autocorrelated noise (20 dB).

Regarding the SISO systems, Figures 5.3 and 5.4 indicate that the deviations distribution arising from the application of Ku *et al.* method is almost uniform between –11 and –2. This may happen because the true delays were also generated by a uniform distribution and the estimates from this method typically vary little, indicating a 0-lag model in about 95% of the SISO systems, and a 1-lag model in approximately 95% of the MIMO 2×2 systems. These results are consistent with the authors' comments about their method, namely that it usually provides estimates of the systems order in the range 1-2. However, the method fails in the estimation of the true number of lags, which necessarily leads to

less adequate DPCA models, which are not correctly modeling the dynamic behavior of the systems.

On the other hand, the proposed method estimates 1 more lag than the correct one in all the generated SISO systems (Figures 5.3 and 5.4). In the case of the MIMO 2×2 systems (Figures 5.5 and 5.6), one can see that it also presents some estimation error, which nevertheless is much lower than that for the Ku *et al.* method. In fact, the absolute deviations obtained with the Ku *et al*. method (e_1) are greater than the ones obtained by the proposed method (e_2). This can be seen in Figure 5.7, where the difference $D = e_1 - e_2$ is represented. From this figure it can be concluded that the proposed approach indeed presents higher estimation accuracy of the true number of lags. The statistical significance of the difference between the methods, was also assessed with a permutation test [104] leading to highly significant *p*-values, much lower than 0.01, confirming the lower absolute deviations obtained with the proposed method.

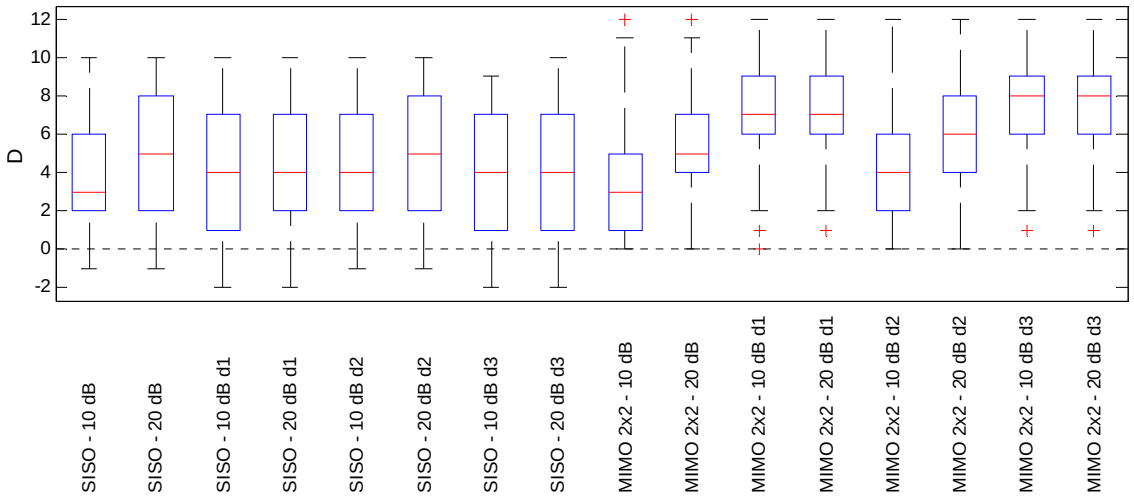


Figure 5.7 Graphical representation of the difference between the absolute deviation obtained on the number of lags estimated by the Ku *et al.* method and the proposed method. d1, d2 and d3 refer to 3 different dynamic transfer functions applied to the additive autocorrelated noise.

5.2.2 Multivariate Dynamic Systems Collected from the Literature

In this section Method 2 is applied to three MIMO systems found in the literature, namely those proposed by (i) Wood and Berry, (ii) Wardle and Wood, and (iii) Ogunnaike and Ray. The corresponding transfer functions are presented in Table 5.4 [105]. The proposed method was applied in two parts: in the first part, Method 2 is employed for selecting the number of lags for the output variables only. In the second part, the same was done for the input variables, using the previously selected number of lags for the outputs. This procedure turned out to be the most effective one for handling complex higher-order input-output dynamic systems. Nevertheless, it should be noticed that an explicit segmentation of the data as input/output variables is not strictly required to the implementation of Method 2.

Table 5.4 Transfer functions for the three MIMO systems used in the comparison study.

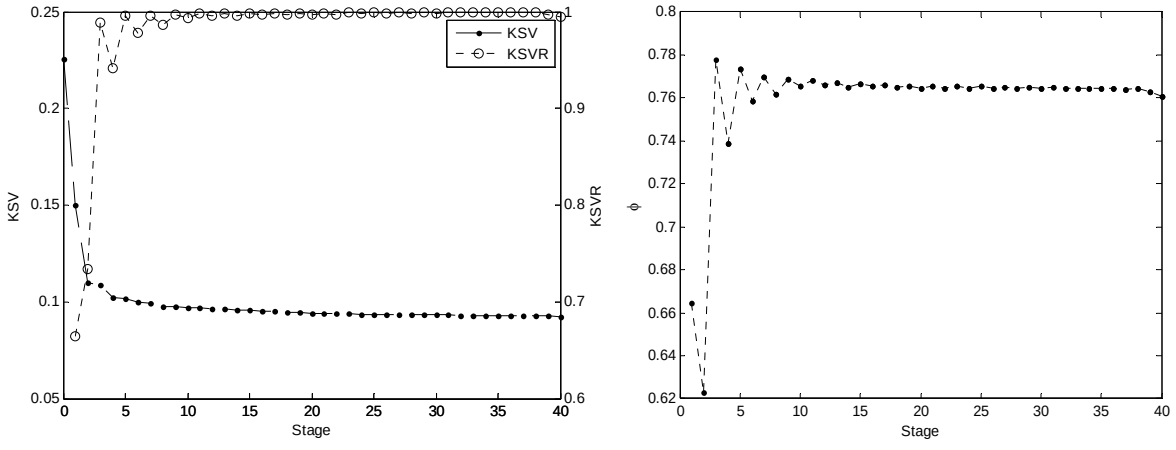
	Wood and Berry	Wardle and Wood	Ogunnaike and Ray
g_{11}	$\frac{12.8e^{-s}}{16.7s+1}$	$\frac{0.126e^{-6s}}{60s+1}$	$\frac{0.66e^{-2.6s}}{6.7s+1}$
g_{12}	$\frac{-18.9e^{-3s}}{21s+1}$	$\frac{-0.101e^{-12s}}{(48s+1)(45s+1)}$	$\frac{-0.61e^{-3.5s}}{8.64s+1}$
g_{13}			$\frac{-0.0049e^{-s}}{9.06s+1}$
g_{21}	$\frac{6.6e^{-7s}}{10.9s+1}$	$\frac{0.094e^{-8s}}{38s+1}$	$\frac{1.11e^{-6.5s}}{3.25s+1}$
g_{22}	$\frac{-19.4e^{-3s}}{14.4s+1}$	$\frac{-0.12^{-8s}}{35s+1}$	$\frac{-2.36e^{-3s}}{5s+1}$
g_{23}			$\frac{-0.01e^{-1.2s}}{7.09s+1}$
g_{31}			$\frac{-34.68e^{-9.2s}}{8.15s+1}$
g_{32}			$\frac{46.2e^{-9.4s}}{10.9s+1}$
g_{33}			$\frac{0.87(11.6s+1)e^{-s}}{(3.89s+1)(18.8s+1)}$
True number of lags (after discretization)	$[2 \quad 2 \quad 9 \quad 5]$	$[2 \quad 2 \quad 10 \quad 15]$	$[3 \quad 3 \quad 4 \quad 14 \quad 14 \quad 5]$

The true number of lags for each system is a function of the sampling rate adopted, and was determined here through the Matlab function *c2d* (that converts a continuous model into a discrete one), and by the transformation equations described by Roffel and Betlem (2006) [106]. All systems were simulated using the Matlab function *lsim*, and subjected to a SNR of 10 dB. Each data set analyzed was composed by 3000 observations.

5.2.2.1 Wood and Berry system

By application of the proposed method to the Wood and Berry system, on the first part of the procedure, the number of lags of the outputs was determined according to the algorithm presented in Table 5.3. This algorithm makes use of the singular values of a data matrix successively extended by additional lagged variables and the ratio between the singular values before and after the inclusion of each lagged variable. These values are introduced into an optimization function (ϕ), from which the lagged variable leading to the lower value is the one to be included in the extended matrix. This procedure is repeated until the maximum number of stages is achieved.

In the case of the Wood and Berry systems, the results obtained in each stage are presented in Figure 5.8.

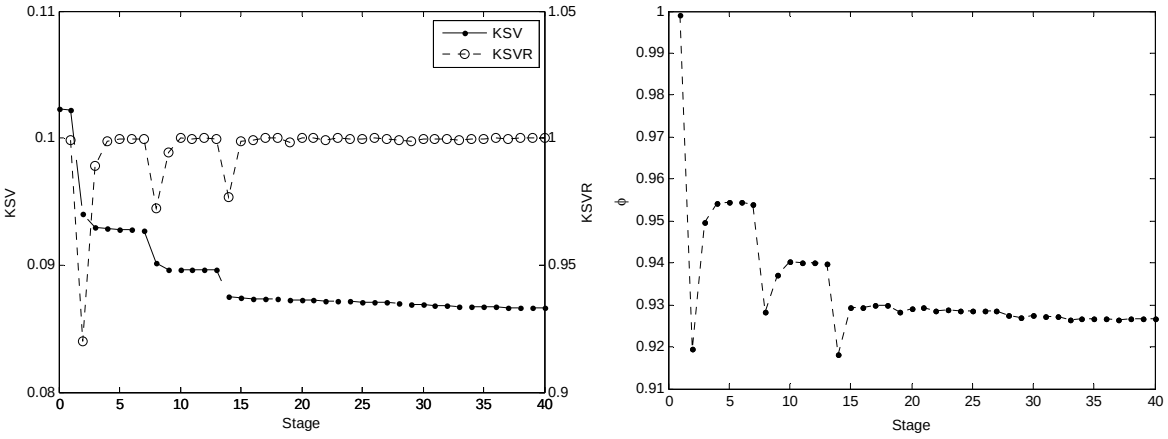


(a) (b)

Figure 5.8 Graphical representation of (a) the singular values, their ratio and (b) the output from the optimization algorithm, ϕ, during the first part of the proposed method, in the Wood and Berry system.

From Figure 5.8 (b) it can be verified that stage $k = 2$ led to the lowest value of ϕ. However, this value is greatly affected by the singular values ratio, $\tilde{r}^{(k)}$ (see Table 5.3, step 4), which is expected to be low in the first stages, because of the rapid decrease in the singular values (s) after the first inclusions of lags. Furthermore, the use of the ratio in the optimization functions, ϕ, is mainly for the purpose of identifying significant changes in s, once its value is already low, and not in the initial stages. This is the reason way the ratio condition is present in the proposed methods (Method 1 see Table 5.2, step 8; Method 2 see Table 5.3, step 4). It should be also noted, that the decreasing profile of s is related to the amount of variability explained. Thus, after the point where all significant lagged variables have been included, the decrease on the singular values will be lower. Consequently, the desirable combination of lags should have a low value for the singular values, s, followed by an almost constant profile (which is equivalent to a ratio near 1). From Figure 5.8 (a) it is observed that stage 4 (that has the second lowest ϕ) has a closer match to these specifications, and was therefore, chosen for providing the number of lags in the output variables.

On the second part of the method, the number of lags for the input variables is also determined, given the information about the number of lags for the output variables. Following the same procedures and considerations, stage 14 is selected for the outputs (see Figure 5.9). In this part, the singular value profile is not as ambiguous as in the outputs case, and a clearer, almost constant, profile appears after stage 14 (see Figure 5.9 (a)), which is also identified by the lowest value of ϕ. By selecting stage 14, a lag vector of $\begin{bmatrix} 2 & 2 & 9 & 5 \end{bmatrix}$ was obtained, which is equal to the theoretical one (see Table 5.4).



(a) (b)

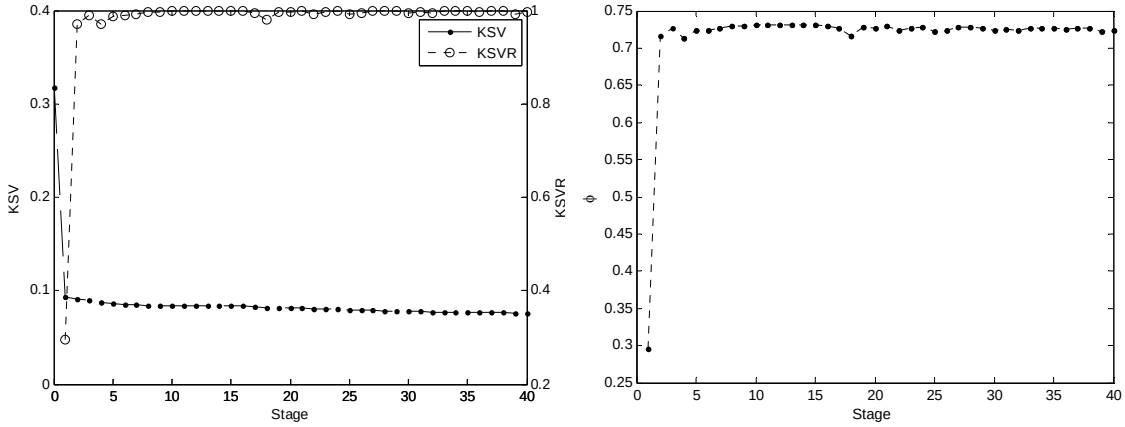
Figure 5.9 Graphical representation of (a) the singular values, their ratio and (b) ϕ, during the second part of the proposed method, in the Wood and Berry system.

5.2.2.2 Wardle and Wood system

The second case study was the Wardle and Wood system. The two part procedure was also implemented, as for the previous example. From the first part of the method the profiles presented in Figure 5.10 was obtained.

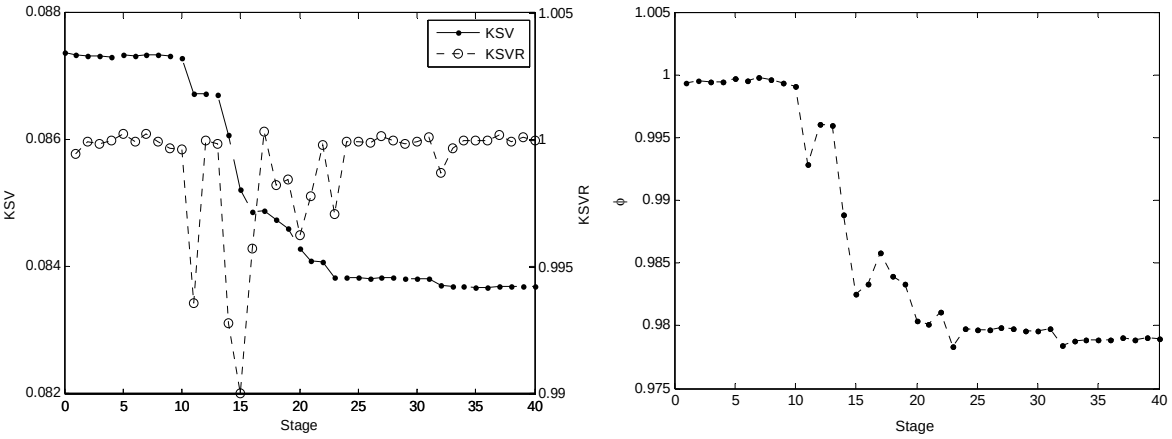
As can be seen in Figure 5.10 (b), the minimum of ϕ is attained on stage 1. However, as in the case of the Wood and Berry system, the ratio in stage 1 has a big effect on the value of ϕ and it also results from the high reduction in the singular value after the addition of the first lagged variable, and therefore should not be considered, in accordance with the ratio condition. Stage 4 respects this condition and has the 2nd lowest ϕ, along with an almost constant singular values profile after it. Given these considerations, stage 4 was selected for the first part of the method.

In the second part of the method, the minimum value of ϕ is obtained in stage 23 (see Figure 5.11(b)). Note that this stage is also related to the almost constant profile of the singular values (Figure 5.11(a)), even with the small decrease in stage 32, which is considered as less relevant due to its smaller ratio. Therefore, stage 23 is the one selected, giving a lag vector of $\begin{bmatrix}2 & 2 & 10 & 14\end{bmatrix}$ which is quite similar to the theoretical one, $\begin{bmatrix}3 & 2 & 10 & 15\end{bmatrix}$ (see Table 5.4).



(a) (b)

Figure 5.10 Graphical representation of (a) the singular values, their ratio and (b) ϕ, during the first part of the proposed method, in the Wardle and Wood system.



(a) (b)

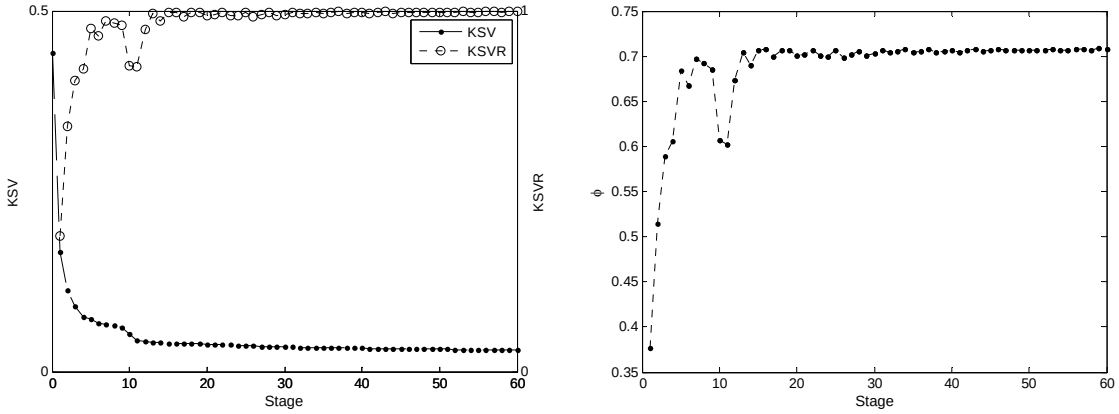
Figure 5.11 Graphical representation of (a) the singular values, their ratio and (b) ϕ, during the second part of the proposed method, in the Wardle and Wood system.

5.2.2.3 Ogunnaike and Ray system

The third case study regards the Ogunnaike and Ray system. From the application of the proposed method, the profiles presented in Figure 5.12 for the output variables (first part) were obtained. In this case, the progressive decrease in *s* is more evident, and therefore a careful analysis of the first stages should be conducted. As can be seen in Figure 5.12 (b), the first stages have the lowest values of ϕ. However, as explained before, the first stages correspond to a rapid, but not stable, reduction of *s*. Note that the required state of almost constant values of *s*, only starts between stage 10 and 20 (Figure 5.12 (a)). The algorithm is capable of dealing with this situation by the inclusion of the ratio condition explained in Section 5.2.2.1. With this condition, the algorithm selects stage 11 for the outputs. This

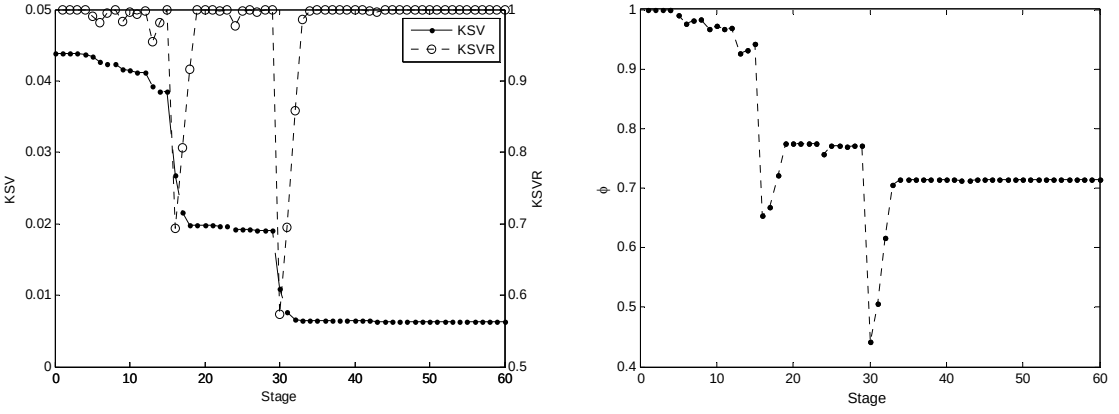
can be considered a good choice since s is already low and, in the subsequent stages, no significant decrease of s occurred.

In the second part, a characteristic profile that becomes almost constant after stage 30 is obtained. This stage also corresponds to the lowest ratio, i.e., the greatest decrease in s relatively to its previous value (see Figure 5.13 (a)). Therefore stage 30 presents all the desirable characteristics and is, in fact, the stage selected by the algorithm, which leads to a lag vector of $\begin{bmatrix}4 & 3 & 4 & 14 & 14 & 2\end{bmatrix}$. This lag vector is quite similar to the theoretical one, $\begin{bmatrix}3 & 3 & 4 & 14 & 14 & 5\end{bmatrix}$ (see Table 5.4).



(a) (b)

Figure 5.12 Graphical representation of (a) the singular values, their ratio and (b) ϕ, during the first part of the proposed method, in the Ogunnaike and Ray system.



(a) (b)

Figure 5.13 Graphical representation of (a) the singular values, their ratio and (b) ϕ, during the second part of the proposed method, in the Ogunnaike and Ray system.

5.3 Application of the Proposed Methodologies to Statistical Process Control (SPC) and System Identification (SI) Activities

The proposed lag selection methods aim to better estimate the real dynamic relationships involving all the system variables, leading to more precise and reliable models. This will have a natural impact in the activities built over DPCA models, such as statistical process control (SPC) and system identification (SI), as analyzed in the following two sections.

5.3.1 Application to Statistical Process Control

In this section, the effect of using the proposed lag selection methods for DPCA models in the task of multivariate statistical process control, is assessed and compared. For such, several monitoring methodologies were implemented, namely the well-known MSPC-PCA procedure [15-17, 19], and two other related procedures, based on DPCA models. One of the methods based on DPCA uses the current lag selection method proposed by Ku *et al.* (1995) [27], DPCA-LS1, and the other employs the proposed methodology (Method 2), DPCA-LS2.

The system studied was the Tennessee Eastman benchmark process, developed by Downs and Vogel (1993) [107], which has been widely used for comparing process monitoring and control approaches. The simulation model has 41 measurements (XMEAS), 12 manipulated (XMV) variables and allows for the analysis of 21 process upsets; more details are provided in Section 6.3.

In this study, the data provided by Russell *et al.* (2000) [108] was used. Each data set contains 960 observations with a sampling interval of 3 min. The faults were introduced 8 hours after the beginning of the simulation runs. All the manipulated and measurement variables, except the agitation speed of the reactor's stirrer (which is always constant) were collected, giving a total of 52 variables.

The data set without faults was used to estimate the PCA and DPCA models. The number of principal components was determined by parallel analysis and the number of lags for the DPCA model was first selected with the approach proposed by Ku *et al.* (1995) [27]. These methods led to a PCA model with 17 PCs and a DPCA model with 3 lags and 29 PCs (DPCA-LS1). These estimates are in line with those obtained before by Russell *et al.*

(2000) [95]. Implementing DPCA-LS2, one would obtain the fine lag structure presented in Table 5.5, along with a total of 69 PCs.

Table 5.5 Number of lags for each variable obtained with DPCA-LS2, in the Tennessee Eastman process.

XMEAS(1)	17	XMEAS(14)	4	XMEAS(27)	17	XMEAS(40)	12
XMEAS(2)	17	XMEAS(15)	17	XMEAS(28)	13	XMEAS(41)	17
XMEAS(3)	8	XMEAS(16)	12	XMEAS(29)	3	XMV(1)	17
XMEAS(4)	17	XMEAS(17)	17	XMEAS(30)	17	XMV(2)	17
XMEAS(5)	17	XMEAS(18)	17	XMEAS(31)	17	XMV(3)	17
XMEAS(6)	16	XMEAS(19)	17	XMEAS(32)	8	XMV(4)	17
XMEAS(7)	17	XMEAS(20)	17	XMEAS(33)	8	XMV(5)	15
XMEAS(8)	15	XMEAS(21)	17	XMEAS(34)	17	XMV(6)	16
XMEAS(9)	17	XMEAS(22)	17	XMEAS(35)	17	XMV(7)	17
XMEAS(10)	17	XMEAS(23)	17	XMEAS(36)	17	XMV(8)	17
XMEAS(11)	16	XMEAS(24)	17	XMEAS(37)	17	XMV(9)	16
XMEAS(12)	17	XMEAS(25)	17	XMEAS(38)	17	XMV(10)	17
XMEAS(13)	17	XMEAS(26)	17	XMEAS(39)	4	XMV(11)	17

The pair of monitoring statistics (T^2 and Q) for each model (PCA, DPCA-LS1 and DPCA-LS2) were then applied to a second data set representing normal operation conditions in order to determine their respective control limits. The control limits were set by trial and error, so that all the statistics present the same (observed) false alarm rate of 1%. The methods where then applied to the battery of 21 data sets with different types of faults. The results obtained, are presented in Table 5.6 in terms of detection rate (i.e., the number of detections in the faulty regions over the total number of observations in the faulty regions).

From the analysis of Table 5.6, it is possible to verify that the DPCA-LS2 statistics have the higher fault detection rates for 17 out of 21 faults, and comparable detection rates on the remaining ones. The superiority of the proposed lag selection method is also formally confirmed upon application of a paired t-test (5% significance level).

Table 5.6 Detection rates for each fault. The best performance achieved for each type of fault is signaled in bold.

Fault	PCA		DPCA-LS1		DPCA-LS2	
	T^2	Q	T^2	Q	T^2	Q
1	0.991	**0.995**	0.990	0.994	0.989	0.993
2	**0.985**	0.984	0.984	0.981	0.981	**0.985**
3	0.036	0.006	0.035	0.010	**0.059**	0.032
4	0.218	0.980	0.165	**0.999**	0.397	**0.999**
5	0.257	0.217	0.293	0.228	0.326	**0.486**
6	0.989	**0.999**	0.989	**0.999**	0.988	**0.999**
7	**0.999**	**0.999**	0.986	**0.999**	0.888	**0.999**
8	**0.974**	0.968	0.973	**0.974**	0.970	0.971
9	0.034	0.010	0.030	0.002	**0.070**	0.017
10	0.367	0.154	0.439	0.172	0.508	**0.531**
11	0.414	0.638	0.340	0.829	0.542	**0.991**
12	0.985	0.925	0.990	0.964	0.994	**0.996**
13	0.943	**0.950**	0.943	**0.950**	0.938	0.948
14	0.988	**0.999**	0.990	**0.999**	0.996	0.998
15	0.035	0.007	0.059	0.009	**0.072**	0.040
16	0.174	0.137	0.217	0.145	0.305	**0.474**
17	0.787	0.905	0.790	0.953	0.954	**0.969**
18	0.893	**0.901**	0.890	0.898	0.894	**0.901**
19	0.115	0.059	0.046	0.298	0.072	**0.956**
20	0.340	0.423	0.408	0.493	0.609	**0.777**
21	0.362	0.414	0.429	0.409	0.444	**0.456**

5.3.2 Application to System Identification

Even though it is not its natural application area, DPCA can also be used in the analysis of input/output systems, namely in SI contexts [27, 94, 102]. In this section, this application scenario will be addressed, mostly to consolidate the results presented above and to illustrate the added-value of properly estimating the dynamic structure of a DPCA model. This analysis will be based on the evaluation of the one-step-ahead prediction performance of the models derived from the application of the various lag selection methods under consideration. Input/output relationships are extracted from the singular vectors relative to the smallest singular values, as they represent the linear relations present in the extended data covariance matrix.

The process under analysis is the Wood and Berry system described before (see Table 5.4), from which 5000 observations were generated with a SNR of 10 dB. By application of the method proposed by Ku *et al.* (1995) [27] the number of lags estimated is 1. On the

other hand, with the proposed method, the estimated lag vector is $[2 \quad 2 \quad 9 \quad 5]$, as referred in Section 5.2.2.1. With such lagged structures, the extended data matrices for DPCA using both approaches were constructed and their corresponding covariance matrices determined. Then, the singular value decomposition was applied to each covariance matrix, and the singular vectors relative to the two smallest singular values were used to estimate the intrinsic systems models.

The models thus obtained from the application of the two lag selection approaches were then used to provide one-step-ahead predictions in independent data sets of 5000 observations, repeated 1000 times. The prediction quality was assessed by the mean squared error (MSE). The results are presented in Table 5.7. For illustration proposes, the MSE for models with 2 and 9 lags are also presented.

Table 5.7 Mean squared error of one-step-ahead predictions for the Wood and Berry system.

Number of lags	MSE
1	67,8 ± 2,91
2	311,22 ± 80,05
9	1,89 ± 0,89
$[2 \quad 2 \quad 9 \quad 5]$	1,06 ± 0,56

From the results on Table 5.7, it can be easily concluded that the proposed method led to the lowest MSE, not only comparing with the results obtained with 1 lag (Ku *et al*. method) but also with the 9 lags model (maximum lag number). This indicates that an individual number of lags for each variable is indeed preferable to a lag number common to all variables, since a more accurate model can be obtained in this circumstances.

5.4 Discussion

In the previous sections, the increased accuracy of the proposed methodologies relatively to the benchmark method proposed by Ku *et al.* (1995) [27] has been demonstrated in several testing scenarios. The added value associated with these methods was also highlighted in practical situations where DPCA models are employed, namely in SPC and

SI tasks. The proposed methodologies for defining the structure of DPCA models are theoretical-driven, but incorporate also results and improvements arising from an extensive analysis of possible alternatives to address the lag selection problem. This dose of empiricism is reflected in the adoption of solutions (such as the "ratio" restriction), which consistently led to better results.

Other alternatives, even when grounded on a well-established theoretical background, failed to provide better results, or presented implementation problems. For instance, one approach was tested to select the maximum number of lags, that consisted in finding the point after which the number of new relationships are the expected ones assuming that all the relevant linear relations were extracted until the previous stage. In fact, if one is able to extract all the relationships with a certain number of lags, by adding one more lag than necessary for all variables, one should theoretically obtain all the previous extracted relationships, replicated one more time. This process will go on, as more lags are added. Therefore, by finding the onset of such regular behavior, one could establish the maximum number of lags necessary to describe the dynamic behavior of all the variables. However, such a methodology led to implementation problems that manifest in imprecise estimates of the onset of the replication process, which translate in even worse results than those provided by Method 1. As an example, consider the Wood and Berry system presented earlier. For this specific system, the correct number of lags is 9, and therefore it was expected that the number of linear relations (NLR) increased in a proportional way after this lag. Nevertheless, as can be seen in Figure 5.14, despite the near linear relation between the NLR and the number of lags, there is no significant difference in the regions before (I) and after (II) lag 9, making it impossible to identify it as the appropriate transition point. In this sense, the proposed methods proved to be empirically accurate and stable in most of the circumstances studied, providing a usable solution to this non-trivial problem of model structure definition for DPCA.

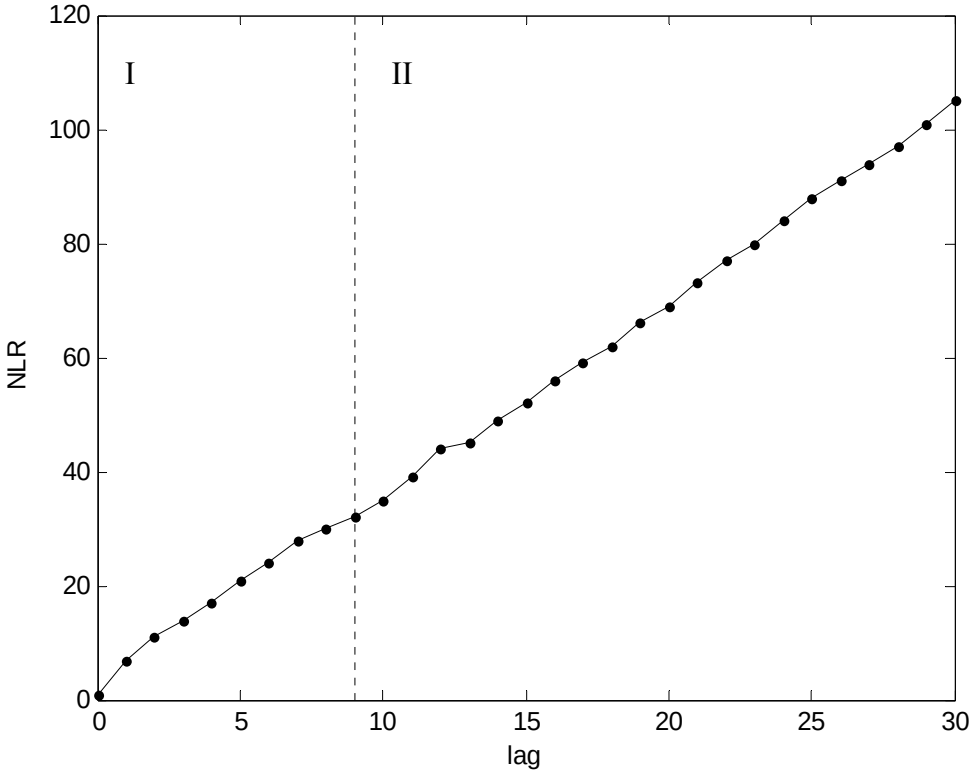


Figure 5.14 Number of linear relations (NLR) obtained by parallel analysis in each lag for the Wood and Berry system.

5.5 Conclusions

In this chapter, two new methods for selecting the number of lags in DPCA models were proposed. These methods are based on the correspondence between the smallest singular values and near linear relationships present on collected data.

The methods for selecting the maximum number of lags were compared with the procedure proposed by Ku *et al.* (1995) [27] and applied to a series of systems with randomly generated structures. From this analysis, it was concluded that the proposed method to select the number of lags led to a more accurate estimation of the true number of lags and was significantly better than the benchmark method, in a statistical sense (at a 5% significance level). The same conclusion was drawn from the implementation of these methods to several models collected from the literature. It was also noted that, although the proposed algorithm is capable of selecting the correct lagged structure most of the times, it is recommend to analyze the graphical representation of the singular values, their ratio and ϕ, and select a stage with simultaneous low values for the singular value and

ratio, with no subsequent lower ratio values. This recommendation should be considered in particular for the first stages of the procedure, as discussed in the text.

Finally, it was shown that the use of the proposed methods to select the number of lags, ultimately leads to superior performances in other activities based on DPCA models, such as statistical process control and system identification.

6 Multivariate Statistical Process Control for Large Scale Processes

Since its introduction in the early 1930's by W.A. Shewhart, statistical process control (SPC) was adopted by the industry as a useful and valuable tool for properly handling the natural or common cause variability in the daily operations and to quickly detect process anomalies requiring intervention [40]. With the continuous increasing demand for improved products and processes over the years, new SPC procedures were developed in order to meet the needs for higher detection sensitivities and to cope with the new data structures that become available (spectra, images, hyphenated instruments measurements, etc.) as well as with the more complex nature of industrial processes and systems. However, these SPC methodologies rely upon the conventional *i.i.d.* assumption for process measurements, and therefore are of limited application scope in most current industrial processes, where the intrinsic process dynamics (driven by mass, energy and momentum conservation, as well as by the action of control systems under external perturbations), associated with high sampling rates (very easy to achieve with the current state of instrumentation technology), lead, very frequently, to strong violations of such hypothesis.

To deal with this limitation, a common solution involves the development of a dynamic model derived for the process under analysis, using normal operation data. This model is then employed to predict the current value of the process (say, at time k), using data acquired until the previous sampling time (time $k-1$). The difference between the estimate and measured values is finally computed at time k, which, if the estimated model is appropriate, should be approximately normal (under Gaussian assumptions) and independently distributed, with zero mean and constant variance. Several explicit model structures can be used for such propose, with the aim of removing the autocorrelation trend, such as: time series models, namely ARIMA [91, 92]; state-space models [65, 109, 110]; and dynamic latent variable models, such as PCA [27] or PLS [111]. In spite of requiring an additional stage of modeling, for which the necessary skills must be present in practice in the development team, this approach tends to be preferentially used.

Another possible approach is based on transforming the time series of the original variables into other(s) sequence(s), presenting much less autocorrelation than the original one, and therefore easily to be monitored by the methodologies based on the *i.i.d.* assumption. A family of transforms with the ability to approximately diagonalize the autocorrelation function for a large class of signals, is the wavelet transform family. The monitoring procedures thus obtained, fall under the heading of multiscale statistical process control [5, 8, 9], and avoid extensive time series modeling, leading to good results for a large class of process upsets and perturbation magnitudes, given their adaptive nature. However, this approach usually requires a considerable programming overhead and a significant developing time.

When addressing the problem of monitoring large scale processes with autocorrelation, the usual adoption of a preliminary stage of classical time series modeling to generate monitoring residuals soon is discarded, as it is very hard, if possible at all, to estimate a VARIMA model even for low/moderate number of variables (such as 10). The solution usually proposed relies on adopting latent variable frameworks, which are well known for their ability to cope with a large number of correlated variables. These frameworks are then modified in order to incorporate the ability to also describe the autocorrelation structure of collected data. This is the basic reasoning underlying the dynamic principal component analysis methodology (DPCA), proposed by Ku *et al.* (1995) [27] in the scope of process monitoring, which has been adopted as the standard technique for handling large scale processes with autocorrelation.

However, even though DPCA constitutes an improved approach for conducting MSPC in large scale dynamic systems when compared to the PCA-based methodology [15-17], one can easily verify that the usual monitoring statistics, T^2 and Q (or SPE – squared prediction error), as well as the individual scores, still present significant amounts of autocorrelation. Thus, such statistics still cannot be properly monitored with control charts based on *i.i.d.* assumptions, in particular when the amount of residual autocorrelation is relatively high.

The failure to generate decorrelated statistics in the original DPCA approach has a direct impact in the detection ability of the technique, as will be demonstrated in this chapter. This originated the development of other approaches, more complex and involved, for handling the same monitoring problem. Examples of such methodologies, include the use

of state-space models [65, 95, 109] and the combination of state-space or multivariate time series modeling frameworks with latent variable models [110, 112, 113]. The relationship of these methodologies with the present work is addressed in the Discussion section.

In this chapter, several approaches related to DPCA are proposed and analyzed, some of them being new and with the ability to generate statistics with very low levels of autocorrelation. Their monitoring performances were compared with those from a variety of current statistics, under different testing scenarios. Among the approaches proposed, it has been found out that the statistics derived from DPCA based on Decorrelated Residuals (DPCA-DR) consistently lead to superior detection performances, having also the advantages of being robust, easy to compute, structurally simpler and more consistent, as the prediction task, necessary to remove the dynamic effects from data, instead of being carried out with resource to parallel modeling techniques, such as time series or state-space models, is entirely accomplished within the same estimated DPCA model.

6.1 Methods

In this section, the monitoring procedures and the corresponding statistics for handling situations where the dynamic behavior of the process variables can no longer be overlooked, are presented. These statistics are introduced, by the first time, according to a general presentation framework that incorporates new monitoring approaches for large scale systems with autocorrelation, together with the current ones. By using such a presentation scheme, the overall picture of the spectrum of approaches for dealing with this class of problems can be transmitted more efficiently. The monitoring statistics are based on a combination of approaches able to cope with the high dimensionality and collinearity of process data (correlation), with others that take into account their dynamic features (autocorrelation). In particular, they encompass a subset of the following methodologies: PCA, DPCA, PLS, ARMA models and missing data imputation methodologies.

Despite the variability of methodologies referred in this section, they inherit some common features from the static approach already in use. In particular, they share the same latent variable backbone to cope with the correlation structure of data, leading to a

decomposition of the whole data variability into two orthogonal parts, namely the part captured and described by the model (followed by the Hotelling's T^2 of the scores) and the one not captured by it (residual variability, monitored with the *SPE* or *Q* statistic). As the several methodologies under consideration encompass multiple combinations of latent variables models, lag selection methods, prediction approaches and monitoring subspaces, it becomes necessary to designate each monitoring statistics in an unambiguous, concise and systematic way. To do so, a compound code was developed for defining which alternative for these different dimensions is actually being considered in a given monitoring statistic. More specifically, the dimensions to specify are: (i) latent variable model adopted (PCA, DPCA, DPLS) and, when required, the method used to select time-shifted variables (Table 6.1); (ii) the prediction method used to cope with the autocorrelation structure of data; and the (iii) subspace the statistic will monitor (Table 6.2). The complete structure of this coding system will be presented in the end of this section. The several alternatives considered in each dimension are as follows.

Code dimension: latent variable model. The latent variable model structures considered in this study are PCA and DPCA as well as dynamic PLS (DPLS). Regarding the DPCA approach, two different lag selection (LS) methodologies were adopted for defining its lagged structure, i.e., the number of time-shifted replicates for each variable (Table 6.1). The LS1 method, is the one proposed by Ku *et al.* (1995) [27] and is based on the number of linear relationships required to properly describe the system (see Section 5.1.1 for more details). The LS2 method, estimates the number of lags for each variable based on a succession of singular value decomposition problems and subsequent analyses of the results following a set of rules, and is fully described in Section 5.1.3.

Table 6.1 Definitions of the codes used to identify the lag selection method adopted in DPCA.

Designation	Lag selection method
LS1	Proposed by Ku *et al.* (1995) [27]
LS2	New method proposed in Section 5.1.3

Code dimension: prediction method. In this chapter, "observed scores" are referred as the scores obtained directly from observed data (using a completely defined latent variable model) through a direct projection operation, and not estimated by some other mean, such

as with resource to a time series model (in the latent variable space) or using a missing data imputation technique. In case the scores are estimated by one of such approaches, they are called "estimated scores", and the estimation methodology is specified by this code dimension. The estimated scores ($\hat{\mathbf{t}}$) can either be obtained through the use of a time series model (autoregressive), or by application of the conditional estimation method presented in Section 2.2. For the cases where an autoregressive model (AR) of order *p* is employed, the predicted score for the i^{th} PC at current time *j*, ($\hat{t}_{i,j}$), is given by (see Section 2.3.1),

$$\hat{t}_{i,j} = \nu_i + a_{i,1} t_{i,j-1} + a_{i,2} t_{i,j-2} + \cdots + a_{i,p} t_{i,j-p} \tag{6.1}$$

The appropriate AR model was fitted with resource to the Matlab algorithm *ARfit* developed by Schneider and Neumaier (2001) [114]. The AR model order was optimally selected using the Schwarz's Bayesian Criterion as the default method. In case this is the prediction method adopted for estimating scores, the code "TS" (Time Series), will appear in the designation of the statistic. For the situations where the conditional estimation is used, the respective code is "MD" (from its origin, Missing Data Imputation theory). The corresponding statistics will also be called DPCA-DR, as the residuals involved, ($\mathbf{t} - \hat{\mathbf{t}}$) and ($\mathbf{x} - \mathbf{P}\hat{\mathbf{t}}$), will lead to statistics with very low levels of serial correlation (DR stands for Decorrelated Residuals). A code "0" indicates that no predicted scores are used in the statistic.

Code dimension: monitored subspace. Table 6.2, refers the codes used to refer to the different situations considered, regarding the complementary partition of data variability. In this table, "reconstructed data", means the reconstruction to the original variables domain, of the variable values corresponding to the scores in the latent variables space.

Table 6.2 Codes used to identify the part of data variability under monitoring with a given statistic (S-latent variable/R-residual), as well as the specific type of statistic used (1,2,…).

Latent variables statistics (S) (PCA subspace)		**Residual statistics (R) (Original variables' subspace)**	
Code	Statistic type	Code	Statistic type
S1	T^2 for observed scores.	R1	Q (or SPE) obtained from reconstructed data using observed scores.
S2	T^2 for observed and estimated scores.	R2	T^2 for the residuals of reconstructed data obtained with the estimated scores.
S3	T^2 for the difference between the observed and the estimated scores (residual scores).		

A more complete description of the statistics used in this study is now provided. The S1 and R1 statistics result from the direct application of the Hotelling's T^2 and Q statistics formulas from PCA, respectively. They are defined by:

$$S_1 = \mathbf{t}^{\mathrm{T}} \mathbf{S}_{\mathbf{t}}^{-1} \mathbf{t} \tag{6.2}$$

$$R_1 = \left(\mathbf{x} - \hat{\mathbf{x}}\right)^{\mathrm{T}} \left(\mathbf{x} - \hat{\mathbf{x}}\right) \tag{6.3}$$

where, following the usual definition of these statistics, $\mathbf{t}_{p\times 1}$ is the vector of scores, $\mathbf{S}_{\mathbf{t}}$ is the sample covariance matrix of the scores and $\hat{\mathbf{x}}$ is the reconstructed data using the observed scores. When the MSPC monitoring methodologies make use of a prediction framework, such as the one based on a time series model (TS) or an estimation methodology for current data (MD), one can also obtain the one-step-ahead estimates for the current scores, $\hat{\mathbf{t}}_{p\times 1}$ (i.e., the values of the scores at the current time, using information from the past, until the last sampling time). Once available, these estimates can be treated in different ways. One approach is to incorporate such predicted scores and the observed ones, in a Hotelling's T^2 statistic, as follows (S2):

$$S_2 = \begin{bmatrix} \mathbf{t} \\ \hat{\mathbf{t}} \end{bmatrix}^{\mathrm{T}} \mathbf{S}_{\mathbf{t},\hat{\mathbf{t}}}^{-1} \begin{bmatrix} \mathbf{t} \\ \hat{\mathbf{t}} \end{bmatrix} \tag{6.4}$$

where $\mathbf{S}_{\mathbf{t},\hat{\mathbf{t}}}$ is the sample covariance matrix of the combined vector of scores. This augmented vector of observed ($\mathbf{t}$) and estimated ($\hat{\mathbf{t}}$) scores, allows for accommodating some structured variability that the autoregressive models were not able to capture.

The R2 and S3 statistics are similar to S1 and R1 for the observed scores, but now involving their one-step-ahead prediction counterparts ($\hat{\mathbf{t}}$), and are given by:

$$S_3 = \left(\mathbf{t} - \hat{\mathbf{t}}\right)^{\mathrm{T}} \mathbf{S}_{\mathbf{t}-\hat{\mathbf{t}}}^{-1} \left(\mathbf{t} - \hat{\mathbf{t}}\right) \tag{6.5}$$

$$R_2 = \left(\mathbf{x} - \mathbf{P}\hat{\mathbf{t}}\right)^{\mathrm{T}} \mathbf{S}_{\mathbf{r}}^{-1} \left(\mathbf{x} - \mathbf{P}\hat{\mathbf{t}}\right) \tag{6.6}$$

where $\mathbf{S}_{\mathbf{t}-\hat{\mathbf{t}}}$ is the sample covariance matrix for the difference between the observed and estimated scores ($\mathbf{t} - \hat{\mathbf{t}}$) and $\mathbf{S}_{\mathbf{r}}$ is the sample covariance matrix of the residuals, in the original variables space, obtained from the reconstructed data using the estimated scores ($\hat{\mathbf{x}} = \mathbf{P}\hat{\mathbf{t}}$) and the actual measurements ($\mathbf{x}$).

Each one of these statistics, is completely specified by the following compound code: **[Latent variable method]-[Prediction Method]-[Type of Statistic]**, where the field: "Latent variable method" may contain PCA, DPCA-LS1, DPCA-LS2 or PLS, in it (i.e., for the case of DPCA, one also specifies the lag selection methodology used, LS1 or LS2); "Prediction Method" including TS, MD or 0; and "Type of Statistic", referring to the specific type of statistics used (see Table 6.2; in the case of PLS, the statistic R1 can be applied over the original domain of the X-variables or Y-variables, being designated by R1x and R1y, respectively). For instance, the statistic with designation DPCA-LS2-MD-S3, is computed according to the formula for S3 (it makes use of the scores estimated by missing data, MD) and is based on a DPCA model, where the number of lags was estimated by the LS2 method. Thus, the DPCA-DR statistics are all those sharing the following code backbone: DPCA-xx-MD-xx.

A resume of all the studied statistics is presented in Table 6.3. In this table it is also indicated the original reference for the current statistics, while the new contributions from this thesis' research, are marked with: "New". For the statistics that are not entirely new, as they are based on concepts already cited in literature, the respective reference field was left in blank.

Table 6.3 Compound coding scheme used to designate all the monitoring statistics studied in this work.

Class of Latent Variable Model	With / Without time-shifted variables	Lag selection method	Prediction model	Designation of the Statistic	Equation	Reference
PCA	Without	-	-	PCA-0-S1	(6.2)	[16]
				PCA-0-R1	(6.3)	[15]
			Time series	PCA-TS-S2	(6.4)*	New
				PCA-TS-S3	(6.5)*	-
				PCA-TS-R2	(6.6)*	New
	With	Ku *et al.* (LS1)	-	DPCA-LS1-0-S1	(6.2)	[27]
				DPCA-LS1-0-R1	(6.3)	[27]
			Time series	DPCA-LS1-TS-S2	(6.4)*	New
				DPCA-LS1-TS-S3	(6.5)*	-
				DPCA-LS1-TS-R2	(6.6)*	New
			Missing data	DPCA-LS1-MD-S3	(6.5)†	New
				DPCA-LS1-MD-R2	(6.6)†	New
		Proposed method (LS2)	-	DPCA-LS2-0-S1	(6.2)	-
				DPCA-LS2-0-R1	(6.3)	-
			Time series	DPCA-LS2-TS-S2	(6.4)*	New
				DPCA-LS2-TS-S3	(6.5)*	-
				DPCA-LS2-TS-R2	(6.6)*	New
			Missing data	DPCA-LS2-MD-S3	(6.5)†	New
				DPCA-LS2-MD-R2	(6.6)†	New
PLS	With	Genetic Algorithms (GA)‡	-	DPLS-GA-S1	(6.2)	-
				DPLS-GA-R1x	(6.3)	-
				DPLS-GA-R1y	(6.3)	-

* The estimated scores ($\hat{\mathbf{t}}$) are obtained through a time series model, Equation (6.1).

† The estimated scores ($\hat{\mathbf{t}}$) are obtained by conditional estimation, Equation (2.14).

‡ The time-sifted variables to be included on the X-block were determined by application of a genetic algorithm on an extended data matrix.

6.2 Performance Assessment of the Monitoring Statistics

In this section, the results of applying the monitoring statistics summarized in Table 6.3 are presented for several simulated scenarios, where faults of different types and magnitudes were introduced in order to accurately measure and compare their detection performance. The systems studied include the Wood and Berry column, the multivariate AR(l) process presented by Ku *et al.* (1995) [27], a large scale process with 100 variables and a continuous stirred-tank reactor (CSTR) system with a heating jacket, under feedback control.

The adoption of simulated examples with different levels of complexity is justified by the need to have, at this stage, precise measures for comparing all the approaches considered, for which is required complete control over the exact moments of occurrence of the faults, their duration, location and main characteristics (namely magnitude). Only under such conditions can they be comparatively assessed in a rigorous way.

6.2.1 Scenario 1: Wood and Berry Column

In this scenario, a model for the approximated dynamical behavior of a binary distillation column separating methanol from water, is adopted [115]. In this dynamic model, the methanol weight fraction in the distillate (x_D) and in the reboiler (x_B) (output variables), are expressed as a function of the reflux (F_R) and reboiler's steam flow rate (F_S) (input variables). It consists of a linear dynamic model, whose transfer function in the Laplace domain is given by Equation (6.7):

$$\begin{bmatrix} x_D(s) \\ x_B(s) \end{bmatrix} = \begin{bmatrix} \dfrac{12.8e^{-s}}{16.7s+1} & \dfrac{-18.9e^{-3s}}{21s+1} \\ \dfrac{6.6e^{-7s}}{10.9s+1} & \dfrac{-19.4e^{-3s}}{14.4s+1} \end{bmatrix} \begin{bmatrix} F_R(s) \\ F_S(s) \end{bmatrix} \tag{6.7}$$

During the simulations, F_R and F_S are assumedly normally distributed with zero mean and unit variance. x_D and x_B are computed according to Equation (6.7), with a random disturbance superimposed, given by the following transfer function, related to the feed flow rate and feed composition [113]:

$$\begin{bmatrix} x_{D,d}(s) \\ x_{B,d}(s) \end{bmatrix} = \begin{bmatrix} \dfrac{3.8e^{-8.1s}}{14.9s+1} & \dfrac{0.22e^{-7.7s}}{21s+1} \\ \dfrac{4.9e^{-3.4s}}{13.2s+1} & \dfrac{0.14e^{-9.2s}}{12.1s+1} \end{bmatrix} \begin{bmatrix} d_1(s) \\ d_2(s) \end{bmatrix} \tag{6.8}$$

where d_1 and d_2 follow a normal distribution with zero mean and their variance is set so that the variability added corresponds to a signal-to-noise ratio of approximately 10 dB (see Equation (5.8)).

The observation measurement vector, is defined as $\mathbf{x} = \begin{bmatrix} x_D & x_B & F_R & F_S \end{bmatrix}^{\mathrm{T}}$. In order to construct the latent variables models, 3000 observations were collected under normal operation conditions ($\mathbf{X}_{ref}$). The data matrix $\mathbf{X}_{ref}$ was then used to estimate the number of lags required for the construction of the DPCA model. From this analysis the number of lags obtained through the use of the Ku *et al.* (1995) [27] approach (LS1) was 2 for all variables. On the other hand, the LS2 method led to a lag structure of $l = \begin{bmatrix} 2 & 2 & 9 & 4 \end{bmatrix}$, that is, 2 additional time-shifted replicates for x_D and x_B, 9 for F_R and 4 lags for F_S. It should be noted that the continuous-time model presented in Equation (6.7), when sampled every minute, gives rise to the following discrete-time difference equations [102]:

$$\begin{cases} x_{D,i} = 1.985x_{D,i-1} - 0.898x_{D,i-2} + 0.744F_{R,i-2} - 0.709u_1F_{R,i-3} \\ \qquad\qquad -0.879F_{S,i-4} + 0.828F_{S,i-5} \\ x_{B,i} = 184x_{B,i-1} - 0.851x_{B,i-2} + 0.579F_{R,i-8} - 0.54F_{R,i-9} \\ \qquad\qquad -1.302F_{S,i-4} + 1.187F_{S,i-5} \end{cases} \tag{6.9}$$

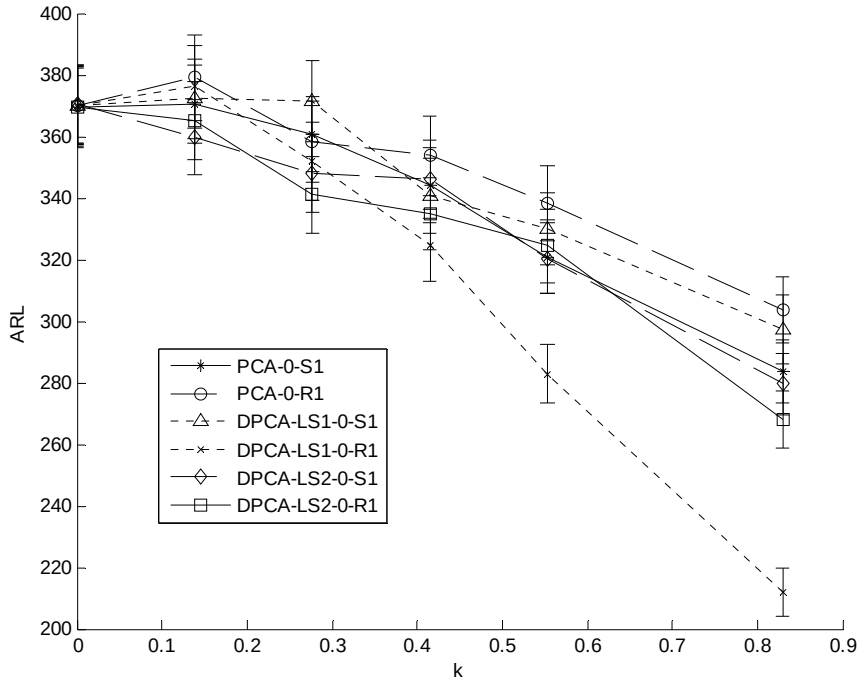
From these equations, one can verify that the actual lag structure is indeed, $l = \begin{bmatrix} 2 & 2 & 9 & 5 \end{bmatrix}$. These number of lags, being regarded as the "true" ones, are an evidence of the superior lag estimation accuracy of the LS2 method.

The system was then subjected to a set of step perturbations in the sensor measurements, and the corresponding average run length (ARL) was determined for each perturbation. In this process, the upper control limits (UCL) for all statistics were previously adjusted, by trial and error, in order to enforce an equal in-control average run length (ARL_0) of 370 for all of them. This is a necessary procedure for enabling a sound comparison of all statistics, as it assures that they all present the same in-control behavior, and therefore, the differences in fault detection performance only arise from their intrinsic characteristics and not from an arbitrary specification of the detection thresholds.

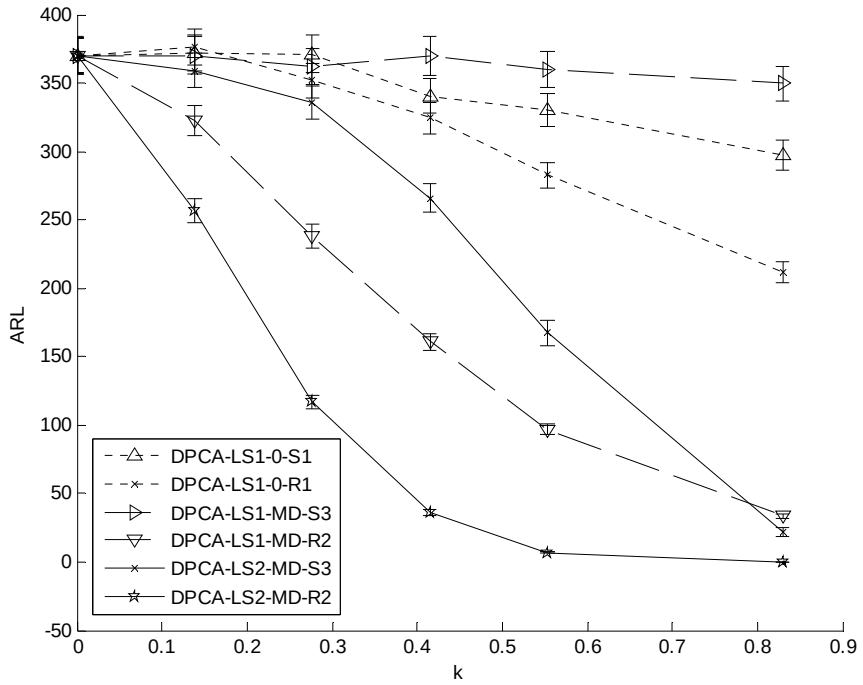
For each perturbation studied, 3000 data sets were generated, leading to the computation of 3000 run lengths, from which the ARLs were computed for all statistics. The ARL values (along with their associated 95% confidence intervals, obtain through bootstrap), for a step perturbation in the mean of the first sensor, with magnitude *k* times the variable's standard deviation, are presented in Figure 6.1. From analysis of Figure 6.1 it

can be seen that, in this case, there is no significant difference between the static and the dynamic versions of PCA (Figure 6.1 (a)), even with the use of the LS2 method to estimate the number of lags. In fact, the DPCA-LS1-0-R1 (that uses the Ku *et al.* (1995) [27] approach) gives better results than DPCA-LS2-0-R1. However, even with DPCA, the resulting statistics still present some autocorrelation (see Figure 6.2). Figure 6.1 (b) displays the results obtained with methodologies that incorporate an implicit prediction methodology, namely through the conditional estimation approach (MD) (for comparison purposes, the results for the best monitoring statistic found in Figure 6.1 (a) are also presented). The results obtained clearly show that the application of such an approach, not only reduces the statistics autocorrelation (Figure 6.3), but also improves the control chart performance (Figure 6.1 (b)).

In Figure 6.4, a performance comparison index (N) based on the computation of the area under the ARL curves (such as those shown in Figure 6.1), is presented. This index is normalized so that it falls in the range $[0,1]$, where 1 represents the best performance (smallest area under the ARL curve). In this analysis, the missing data based statistics, especially DPCA-LS2-MD-R2, presents the best performance, and lead to the weakest final autocorrelation.

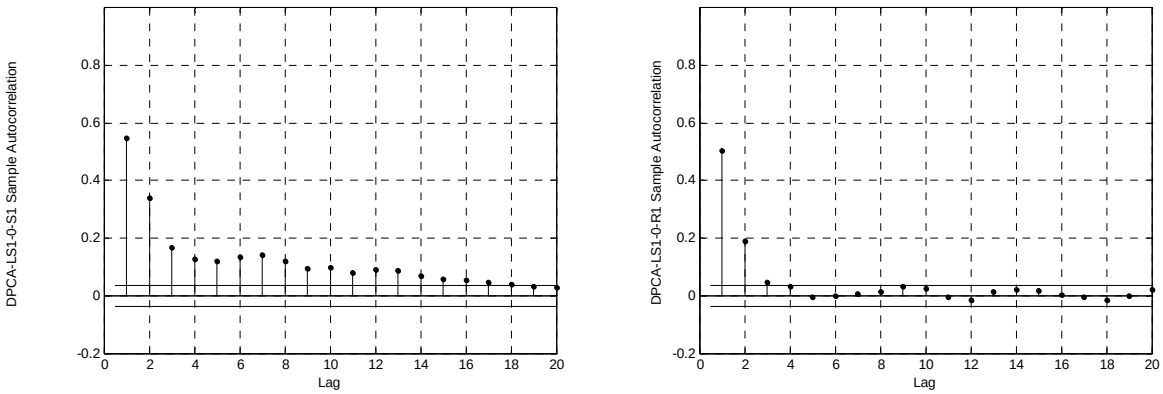


(a)



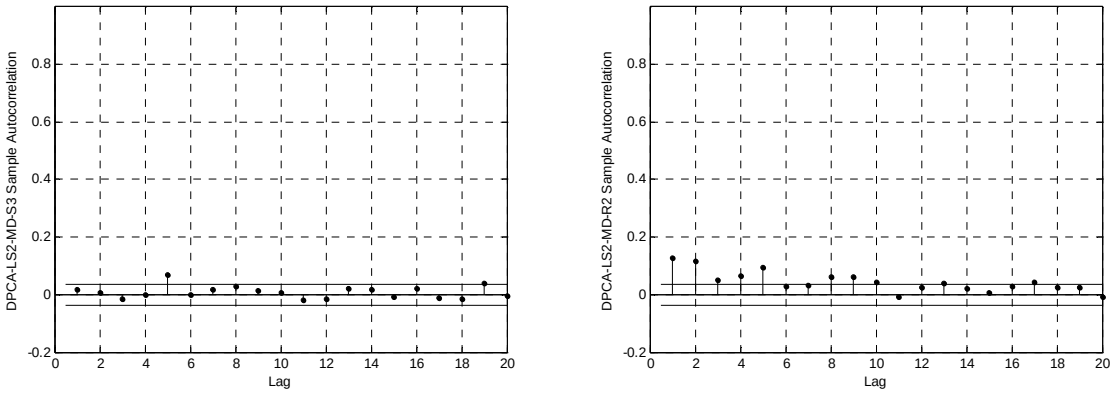
(b)

Figure 6.1 ARL for the tested methodologies used to monitor the Wood and Berry distillation column subject to deviations in the first sensor measurement. In (a) no predictive methodologies are employed, whereas in (b) it is presented the results for the case where the approaches based on MD are adopted.



(a) (b)

Figure 6.2 Sample autocorrelation function for the (a) DPCA-LS1-0-S1 and (b) DPCA-LS1-0-R1 statistics.



(a) (b)

Figure 6.3 Sample autocorrelation function for the (a) DPCA-LS2-MD-S3 and (b) DPCA-LS2-MD-R2 statistics.

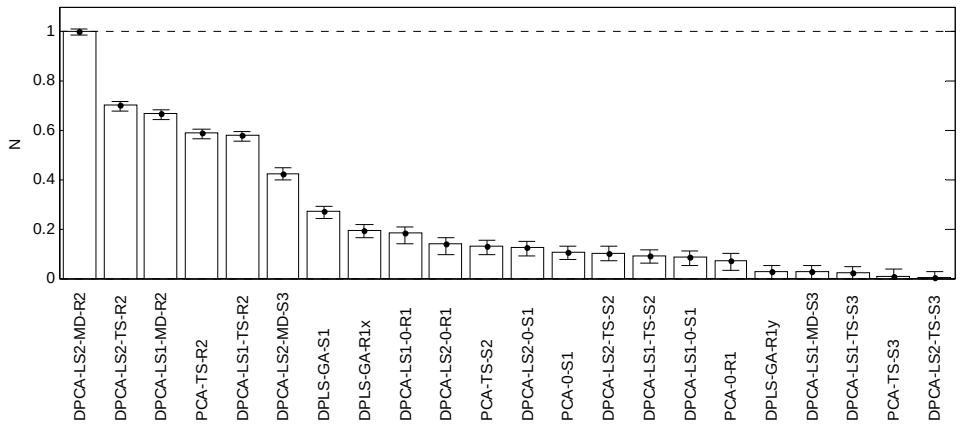


Figure 6.4 Comparison of the methods performance in the Wood and Berry column: values of the performance index (*N*) obtained when the system was subject to deviations in the first sensor measurement. Bars' heights correspond to the associated mean values.

6.2.2 Scenario 2: Multivariate AR(1) Process

The following multivariate AR(1) process was presented by Ku *et al.* (1995) [27] to demonstrate the application of DPCA on multivariate statistical process control.

$$\begin{aligned}\mathbf{z}(k) &= \begin{bmatrix} 0.118 & -0.191 \\ 0.847 & 0.264 \end{bmatrix}\mathbf{z}(k-1) + \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}\mathbf{u}(k-1), \\ \mathbf{y}(k) &= \mathbf{z}(k) + \mathbf{v}(k)\end{aligned} \tag{6.10}$$

where **u** is the correlated input:

$$\mathbf{u}(k) = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{bmatrix}\mathbf{u}(k-1) + \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix}\mathbf{w}(k-1) \tag{6.11}$$

The input **w** is a random noise sequence with zero mean and variance 1. The output **y** is equal to **z** plus another random noise component, **v**(*k*), with zero mean and variance 0.1. The observation measurement vector is defined as $\mathbf{x} = \begin{bmatrix} \mathbf{y}^{\mathrm{T}} & \mathbf{u}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$.

As in the previous example, the reference data was composed by 3000 observations and was used to estimate the reference models and to estimate the number of lags according to the two considered methodologies. In this case, both lag selection methods gave an estimation of 1 lag for all variables, and therefore there are no differences between monitoring statistics that only differ on LS1 and LS2.

To assess the monitoring statistics performance, the system was subjected to changes on the mean of **w**, from which 3000 observations were collected for different magnitudes of the change. Each one of these changes was repeated 3000 times. The corresponding ARL values for the more relevant monitoring statistics and their 95% confidence limits are presented in Figure 6.5.

From Figure 6.5 it is noticeable that the proposed missing data based monitoring statistics present a superior performance, namely DPCA-LS2-MD-R2, which also presents no residual autocorrelation, contrary to what happens with the other monitoring approaches.

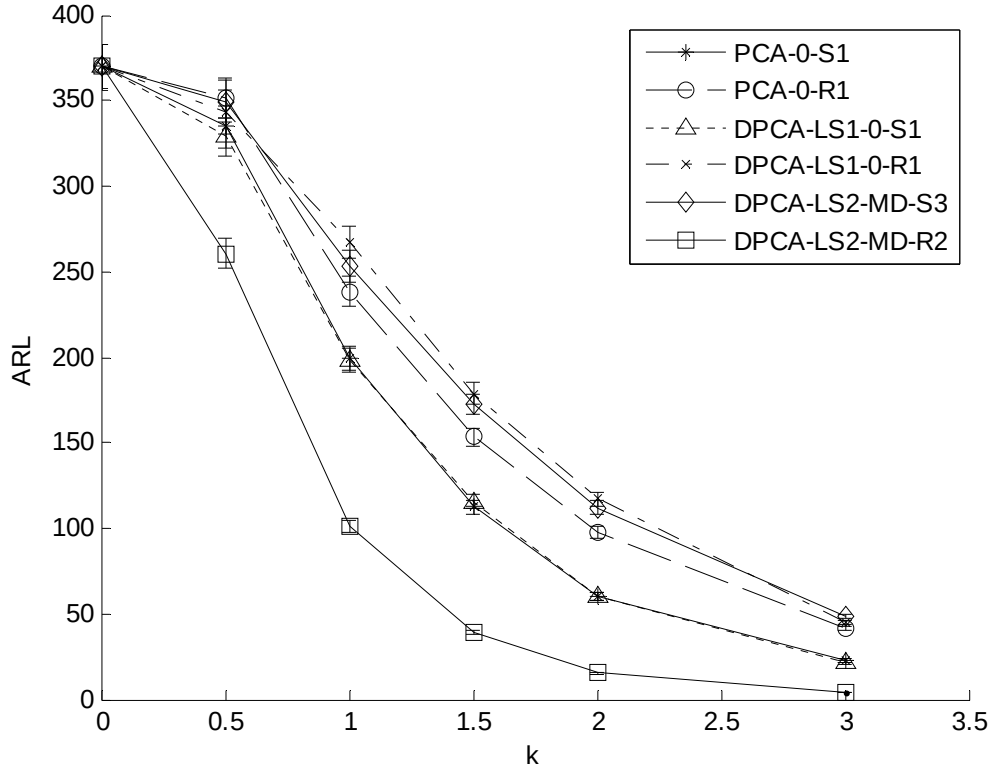


Figure 6.5 ARL for the tested methodologies used to monitor the multivariate AR(1) process subject to step perturbations in the mean of **w**, with magnitude *k* times the standard deviation of **w**.

6.2.3 Scenario 3: Large Scale Process

To assess the ability of the proposed methodologies to cope with a large number of variables, a megavariate process with the following latent variable model structure was simulated:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{6.12}$$

where $\mathbf{X}$ is an $(n \times m)$ matrix of measured variables, $\mathbf{T}$ is a $(n \times p)$ matrix of latent variables, $\mathbf{P}$ is an $(m \times p)$ matrix of orthogonal loadings and $\mathbf{E}$ is an $(n \times m)$ matrix of errors. In this study, the number of measured variables (m) was set to 100, with 5 latent variables (p) following independent AR(1) processes with autoregression coefficients of 0.90. The $\mathbf{P}$ matrix was randomly generated, but forced to have orthogonal columns.

As a result of the application of the LS1 lag selection methods, 0 lags were attributed to all variables, which is clearly an underestimation of the system's dynamics. Furthermore, this situation corresponds to the standard PCA procedure and therefore, in this case, it

was decided to use 1 lag for all variables in order to have a more diversified set of monitoring methodologies to compare. In contrast, when the LS2 lag selection method was applied, it selected 0 lags for 30 variables, 1 lag for 32 variables and 2 lags for the remaining 38 variables. After estimating the reference models from NOC data, the control limits for the monitoring statistics were adjusted to the same in-control ARL_0 of 370. The system was then subjected to step deviations of different magnitudes on one of the variables of the **X** matrix, emulating for instance a sensor failure. The corresponding ARL for each perturbation was computed based on 3000 replications (leading to 3000 run lengths) which are represented along with the associated 95% confidence limits obtained by bootstrapping in Figure 6.6 for a selected set of the most relevant monitoring statistics in this case study.

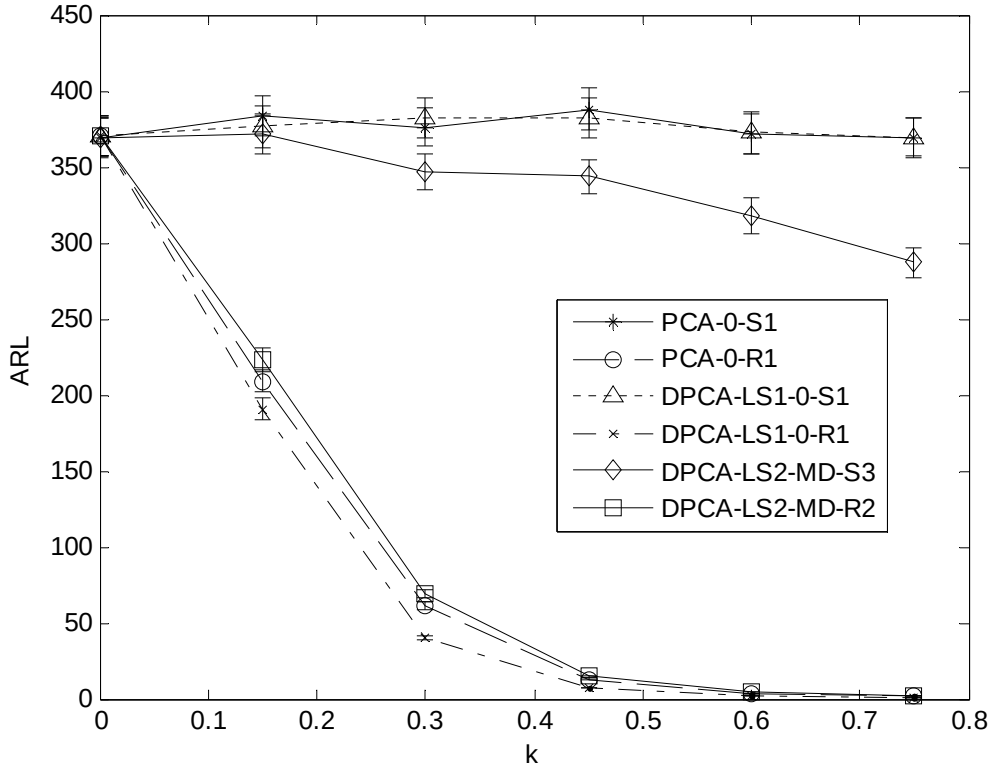
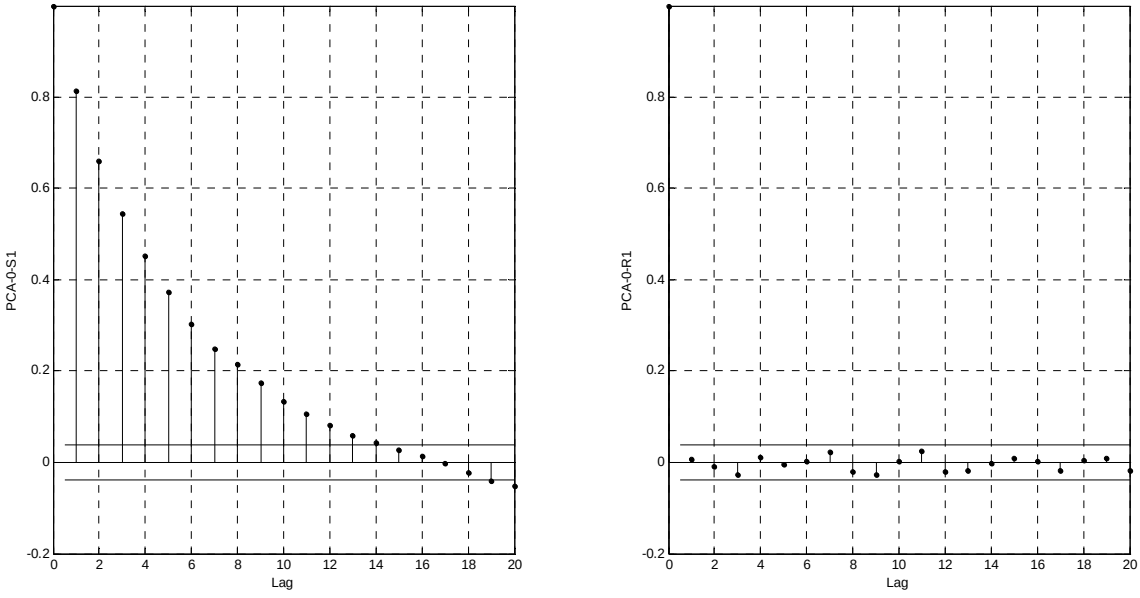


Figure 6.6 ARL for the tested methodologies used to monitor the large scale process subject to a step perturbation in one of the **X** variables, with magnitude *k* times the standard deviation of the corresponding variable.

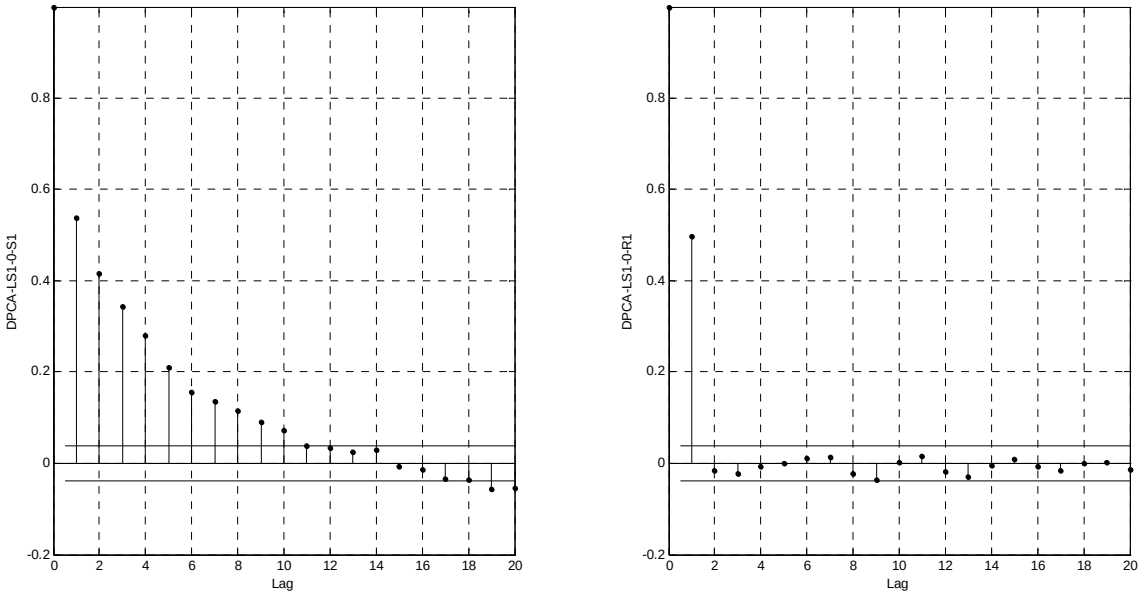
From Figure 6.6, it is clear that, in general, the monitoring methodologies are capable to detect the simulated faults through their residuals statistics (i.e., PCA-0-R1, DPCA-LS1-0-R1 and DPCA-LS2-MD-R2). However, under normal operation conditions, some of

these procedures present a considerably high autocorrelation as can be seen in Figures 6.7 to 6.9. The effects of autocorrelation are more evident on the monitoring statistics for the scores subspace, and in a less extent on the DPCA-LS1-0-R1 statistic, applied on the residuals subspace, which implies that its underlying monitoring latent variable models do not fully describe the system structure. It is also noticeable that the residual monitoring statistic of the static PCA model (Figure 6.7) has no autocorrelation, which explains its good detection capabilities in this case. Yet, the PCA monitoring statistic for the scores subspace is highly autocorrelated. On the other hand, when the decorrelated residuals approach is considered (DPCA-LS2-MD-S3 and DPCA-LS2-MD-R3) the monitoring statistics' autocorrelation remains low at all times (Figure 6.9) without compromising the detection ability.



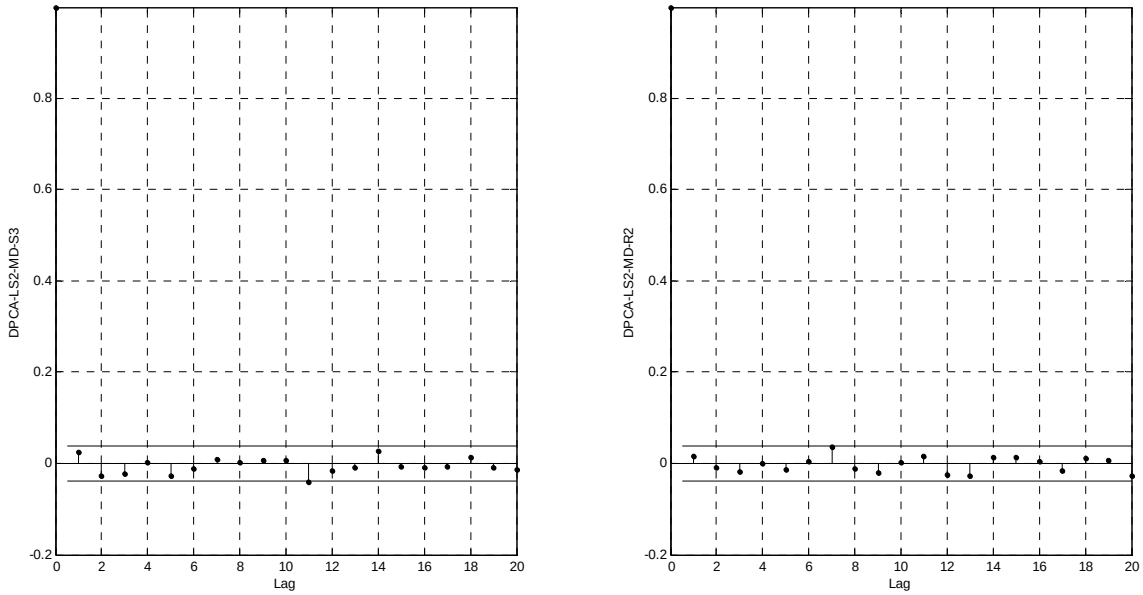
(a) (b)

Figure 6.7 Sample autocorrelation function for the (a) PCA-0-S1 and (b) PCA-0-R1 statistics.



(a) (b)

Figure 6.8 Sample autocorrelation function for the (a) DPCA-LS1-0-S1 and (b) DPCA-LS1-0-R1 statistics.



(a) (b)

Figure 6.9 Sample autocorrelation function for the (a) DPCA-LS2-MD-S3 and (b) DPCA-LS2-MD-R2 statistics.

From the performance comparison index (*N*) presented in Figure 6.10, it is also clear that the best monitoring statistics for this case study either use a time series or a missing data technique to estimate the scores. However, the latter approach is considered to be

preferable, since it does not require the use of a different model structure besides DPCA, namely a different time series model for each score. Finally, even though DPCA-LS2-MD-R3 is, in this particular case study, the third best monitoring statistic, its performance is very close to the best one, as can be seen in Figure 6.6, and has the advantage of using a more consistent procedure with low autocorrelation for both monitoring statistics (see Figure 6.9), which provides more robustness and simplicity in practice.

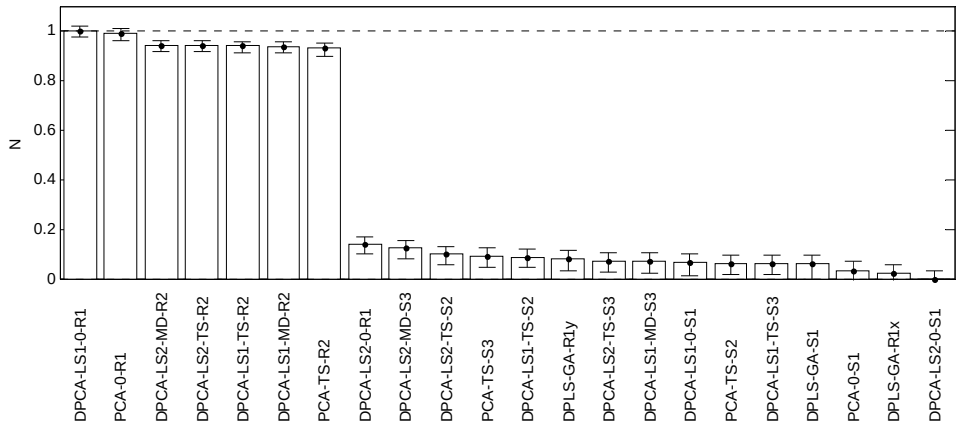


Figure 6.10 Comparison of the methods' performance in the large scale system: values of the performance index (*N*) obtained when the system was subjected to a step perturbations in one of the X variables. Bars' heights correspond to the associated mean values.

6.2.4 Scenario 4: Continuous Stirred-Tank Reactor

In order to assess the statistics performance in a more realistic system, a dynamic model of a continuous stirred-tank reactor (CSTR) with a heating jacket and under feedback control was employed. In this system, an endothermic reaction of the type $A \rightarrow B$ takes place in a CSTR with free discharge flow. This system is under a PI control loop in order to maintain the temperature and fluid level close to their set-points. The inputs of the system are the feed stream concentration and temperature, and the heating fluid inlet temperature. The system outputs are the CSTR level, concentrations and temperature, and the heating fluid outlet temperature. These 3 input and 4 output variables were considered to form the set of measured variables.

The method's parameters were determined from a reference data set composed of 3000 observations, leading to 2 lags for all the mentioned variables through the LS1 method.

On the other hand, the LS2 method only selected 3 lags for the reactor temperature and 4 lags for the heating fluid outlet temperature.

The system was then subjected to perturbations on the discharge coefficient and heat transfer coefficient, which inherently change the systems dynamics and therefore affect several variables simultaneously. Each perturbation was repeated 3000 times and in each run 3000 observations were recorded.

In this case study most of the monitoring statistics presented a similar performance, as can be seen in Figures 6.11 and 6.12 where the results regarding the performance comparison index (*N*) are presented. Nevertheless, the inclusion of MD techniques eliminated the autocorrelation when applied with LS1, which is still outperformed by DPCA-LS2-MD-R2.

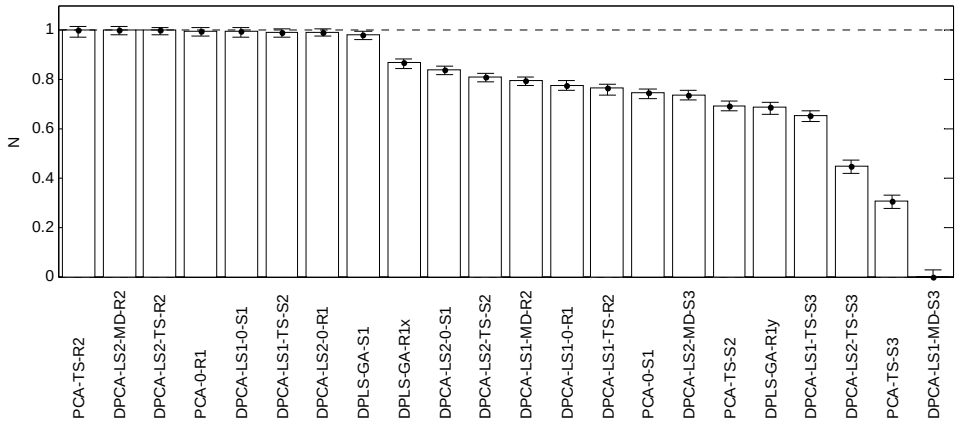


Figure 6.11 Comparison of the methods' performance in the CSTR system: values of the performance index (*N*) obtained when the system was subjected to changes on the discharge coefficient. Bars' heights correspond to the associated mean values.

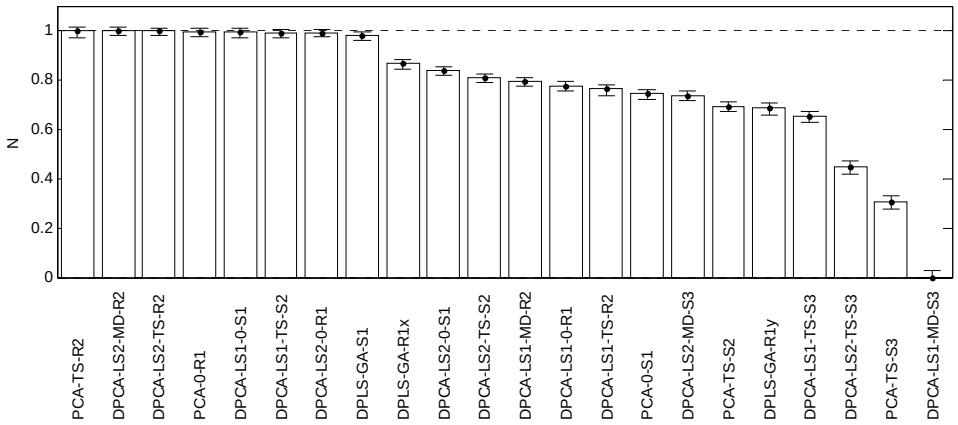


Figure 6.12 Comparison of the methods' performance in the CSTR system: values of the performance index (*N*) obtained when the system was subjected to changes on the heat transfer coefficient. Bars' heights correspond to the associated mean values.

6.2.5 Discussion

Analyzing the results presented in the previous section, one can verify that the DPCA-LS2-MD-R2 statistic tends to present a consistently superior performance over the remaining methodologies, as can be observed in Figure 6.13, where the global results are summarized.

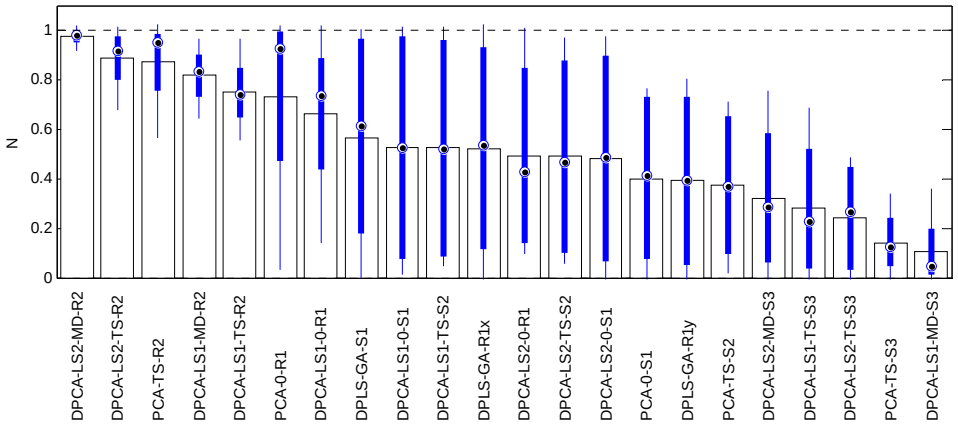


Figure 6.13 Comparison of the methods' performance in all systems considered in this study: box-plots of the performance index (*N*) obtained in all simulations performed. Bars' heights correspond to the associated mean values.

The DPCA-LS2-MD-R2 statistic was one of the new statistics introduced in this chapter and is a representative of a new class of monitoring schemes that make use of data imputation methods to predict the current scores of a DPCA model, called DPCA-DR. These results indicate that DPCA-DR provide a competitive alternative to the current MSPC methodologies, as its monitoring statistics usually present better detection performances and lower autocorrelation levels. However, they also depend on a suitable method to estimate the number of lags for the DPCA-DR model. This issue was also addressed in this thesis, where two methods were adopted to estimate the number of lags and it was found that the new LS2 method was clearly superior to LS1.

Given these results, it was found pertinent to highlight the two most promising monitoring statistics based on DPCA with decorrelated residuals obtained by missing data imputation (DPCA-LS2-MD-S3 and DPCA-LS2-MD-R2).

For monitoring the latent variables subspace, DPCA-LS2-MD-S3, is renamed as T^2_{PREV} and is computed as,

$$T^2_{PREV} = \left(\mathbf{t} - \hat{\mathbf{t}}\right)^{\mathrm{T}} \mathbf{S}^{-1}_{\mathbf{t}-\hat{\mathbf{t}}} \left(\mathbf{t} - \hat{\mathbf{t}}\right) \tag{6.13}$$

where $\mathbf{S}_{\mathbf{t}-\hat{\mathbf{t}}}$ is the sample covariance matrix of the difference between the observed and estimated scores, $(\mathbf{t} - \hat{\mathbf{t}})$. Likewise, the complementary monitoring statistic for the residuals subspace, DPCA-LS2-MD-R2, becomes T^2_{RES} and is given by,

$$T^2_{RES} = \mathbf{r}^{\mathrm{T}} \mathbf{S}^{-1}_{\mathbf{r}} \mathbf{r} = \left(\mathbf{x} - \mathbf{P}\hat{\mathbf{t}}\right)^{\mathrm{T}} \mathbf{S}^{-1}_{\mathbf{r}} \left(\mathbf{x} - \mathbf{P}\hat{\mathbf{t}}\right) \tag{6.14}$$

where $\mathbf{S}_{\mathbf{r}}$ is the sample covariance matrix of the residuals in the reconstructed data, obtained with the estimated scores ($\mathbf{r} = \mathbf{x} - \mathbf{P}\hat{\mathbf{t}}$). These two statistics are then considered as the core of the DPCA-DR procedure and their performance will be further assessed in the following section for the Tennessee Eastman case study.

Looking to its construction, the DPCA-DR approach implicitly conducts a multivariate time series modeling and prediction step in order to compute the residuals in the scores and variables spaces, but in a fully integrated way on a conventional DPCA framework, avoiding the usual two-stage modeling/projection sequence. Up until now, the applications of the DPCA model have not considered the computation of serial

decorrelated residuals [27, 95, 110, 116], which has been limiting its performance, and therefore the application of this approach in practice. For instance, in the work of Russell *et al.* (2000) [95] DPCA was implemented as suggested by Ku *et al.* (1995) [27] and therefore no decorrelated residuals were computed [95]. This justifies the similar detection performances obtained with the PCA and DPCA statistics in this work, as well as the inferior sensitivity of DPCA when compared with CVA, which is also a linear modeling approach for estimating a state-space model.

A possible limitation of the DPCA-DR approach is that it is based on the conditional mean replacement solution to the missing data imputation problem, which faces problems when the dispersion matrix of the past observations becomes ill-conditioned. In such case, other missing data techniques, such as projection to model plane [21] or trimmed scores regression (TSR) [22] can be employed. For more detail in this subject, please refer to Section 2.2.

6.3 Case Study: Tennessee Eastman Benchmark Process

To further test and compare the monitoring features and performance of the proposed DPCA-DR methodology, an application scenario which has been widely used in process monitoring and fault detection studies was selected: the Tennessee Eastman benchmark process. This case study not only provides an additional challenging testing environment for the specific comparison study carried out in this chapter, but also enables and simplifies the extension of the comparison scope to other methods tested in the same system, such as [27, 95, 117-119].

As referred before in Section 5.3.1, a model of this process was developed by Downs and Vogel (1993) [107], consisting of five major transformation units, which are a reactor, a condenser, a compressor, a separator, and a stripper, as shown in Figure 6.14. From this model, 41 measurements (XMEAS) are generated along with 12 manipulated (XMV) variables. A total of 21 different process upsets are simulated for testing the detection ability of the monitoring methods, as presented in Table 6.4 [95, 120]. In the current study the analysis was conducted with the data set used by Russell *et al.* (2000) [108], where the Tennessee Eastman process is controlled with the approach suggested by Lyman and Georgakis (1995) [120]. Each data set contains 960 observations collected at

a sample interval of 3 min and the faults were introduced 8 hours after the simulations start. All the manipulated and measurement variables, except the agitation speed of the reactor's stirrer (which is always constant), were collected, giving a total of 52 variables.

A data set with no faults, representing normal operation conditions was used to estimate the reference PCA, DPCA and DPCA-DR models. The number of principal components for PCA and DPCA was determined by parallel analysis and the number of lags was selected by the approach proposed by Ku *et al.* (1995) [27]. Using these methods, a PCA model with 17 PCs and a DPCA model with 3 lags and 29 PCs were constructed. These results are in good accordance with those obtained by Russell *et al.* (2000) [95]. For selecting the number of lags for the DPCA model, the algorithm proposed in Section 5.1.3 was used to obtain the number of lags for each variable lag structure as presented in Table 5.5, which led to a model with 69 PCs.

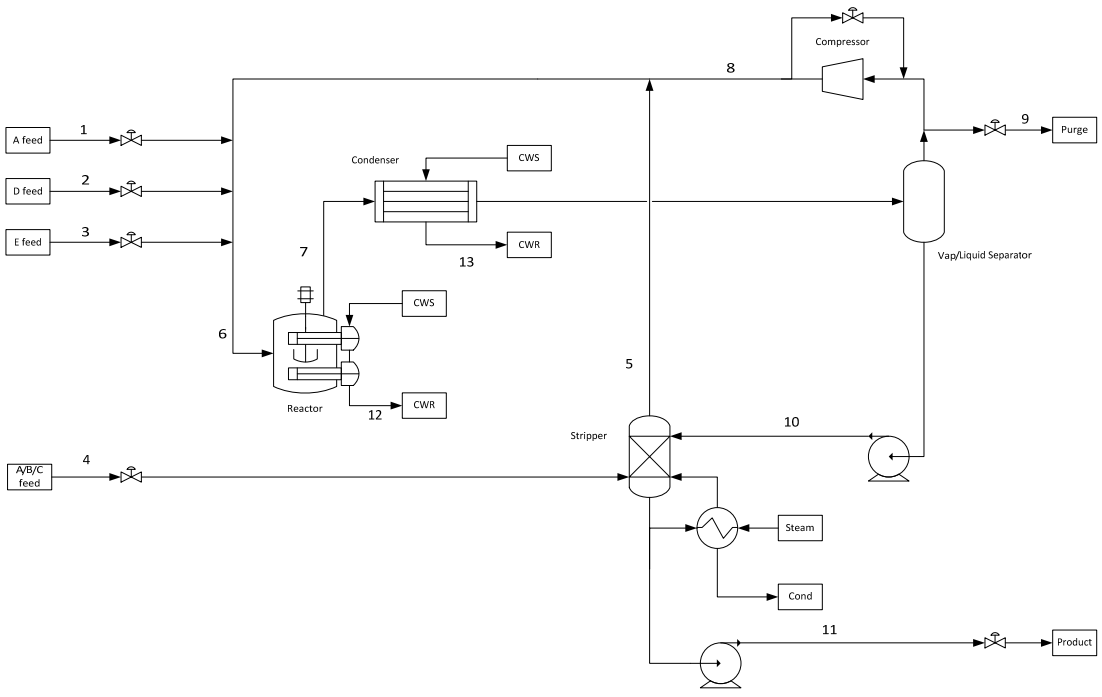


Figure 6.14 The Tennessee Eastman process flow sheet.

Table 6.4 Process faults for the Tennessee Eastman process simulator.

Variable	Description	Type
IDV(1)	A/C feed ratio, B composition constant(Stream 4)	Step
IDV(2)	B composition, A/C ratio constant (Stream 4)	Step
IDV(3)	D feed temperature (Stream 2)	Step
IDV(4)	Reactor cooling water inlet temperature	Step
IDV(5)	Condenser cooling water inlet temperature	Step
IDV(6)	A feed loss (Stream 1)	Step
IDV(7)	C header pressure loss - reduced availability (Stream 4)	Step
IDV(8)	A, B, C feed composition (Stream 4)	Random variation
IDV(9)	D feed temperature (Stream 2)	Random variation
IDV(10)	C feed temperature (Stream 4)	Random variation
IDV(11)	Reactor cooling water inlet temperature	Random variation
IDV(12)	Condenser cooling water inlet temperature	Random variation
IDV(13)	Reaction kinetics	Slow drift
IDV(14)	Reactor cooling water valve	Sticking
IDV(15)	Condenser cooling water valve	Sticking
IDV(16)	Unknown	
IDV(17)	Unknown	
IDV(18)	Unknown	
IDV(19)	Unknown	
IDV(20)	Unknown	
IDV(21)	The valve for Stream 4 was fixed at the steady state position	Constant position

An important point to consider before proceeding with the comparison study is that the direct use of the theoretical significance levels for establishing the statistical control limits for the various methods may lead to widely different observed false alarm rates, which distorts any comparison on the methods detection performances. This undesirable effect can be removed by manipulating the theoretical significance level of the control limits in such a way that the effectively observed performance for all methods under normal operations conditions (i.e., their false alarm rate), becomes equal. Therefore, the UCL for the various methods were set to a false alarm rate of 1% under normal operation conditions, by trial and error, on a second data set with no faults. With this preliminary but important procedure concluded, the fault detection rates for all the methods regarding each fault were finally determined. A summary of the results obtained is presented in Table 6.5.

Table 6.5 Fault detection rates for the various methods under study, regarding each faulty scenario (a description of each fault can be found in Table 6.4). The top scores are signaled in bold.

Fault	PCA		DPCA		DPCA-DR	
	T^2	Q	T^2	Q	T^2_{PREV}	T^2_{RES}
1	0.991	0.995	0.990	0.994	0.996	**0.998**
2	**0.985**	0.984	0.984	0.981	**0.985**	0.983
3	**0.036**	0.006	0.035	0.010	0.021	0.016
4	0.218	0.980	0.165	**0.999**	0.998	**0.999**
5	0.257	0.217	0.293	0.228	**0.999**	**0.999**
6	0.989	**0.999**	0.989	**0.999**	**0.999**	**0.999**
7	**0.999**	**0.999**	0.986	**0.999**	**0.999**	**0.999**
8	0.974	0.968	0.973	0.974	**0.985**	0.981
9	**0.034**	0.010	0.030	0.002	0.020	0.010
10	0.367	0.154	0.439	0.172	**0.956**	0.933
11	0.414	0.638	0.340	0.829	**0.965**	0.865
12	0.985	0.925	0.990	0.964	**0.998**	**0.998**
13	0.943	0.950	0.943	0.950	**0.958**	0.956
14	0.988	**0.999**	0.990	**0.999**	0.998	**0.999**
15	0.035	0.007	0.059	0.009	**0.385**	0.047
16	0.174	0.137	0.217	0.145	**0.976**	0.945
17	0.787	0.905	0.790	0.953	**0.976**	0.975
18	0.893	0.901	0.890	0.898	**0.905**	0.900
19	0.115	0.059	0.046	0.298	**0.971**	0.843
20	0.340	0.423	0.408	0.493	0.908	**0.916**
21	0.362	0.414	0.429	0.409	0.539	**0.577**

From the analysis of Table 6.5, it is possible to verify that the DPCA-DR monitoring statistics tend to present highest fault detection rates. In fact, T^2_{PREV} was the best statistic in 14 out of 21 faults, and T^2_{RES} in 9 of them. Globally, they were capable to detect 19 of 21 faults, failing only in the detection of faults number 3 and 9, where all methods present problems. Fault number 15 is another example of a fault difficult to detect, but where the statistic T^2_{PREV} achieved the best score. The lower capability for detecting these three specific faults was expected, as other methods reported in the literature (e.g. PCA, DPCA and CVA) also fail to detect them [95].

In order to better illustrate the monitoring behavior of the methods under analysis, the control charts for some of the process faults are presented in Figures 6.15 and 6.16. From these representations it is possible to clearly observe that only the DPCA-DR statistics present a consistent out-of-control state in both statistics, simultaneously (T^2_{PREV} and T^2_{RES}, see Figure 6.15). This is a relevant issue, since the PCA and DPCA statistics may lead to

the erroneous conclusion that the process has returned to their normal operation conditions and that it is no longer under the effect of a fault. In the case of Fault 10 (Figure 6.16) only the DPCA-DR statistics signals an out-of-control state during the total duration of the fault, while the PCA and DPCA statistics only became out-of-control when the data also exceeds their normal values.

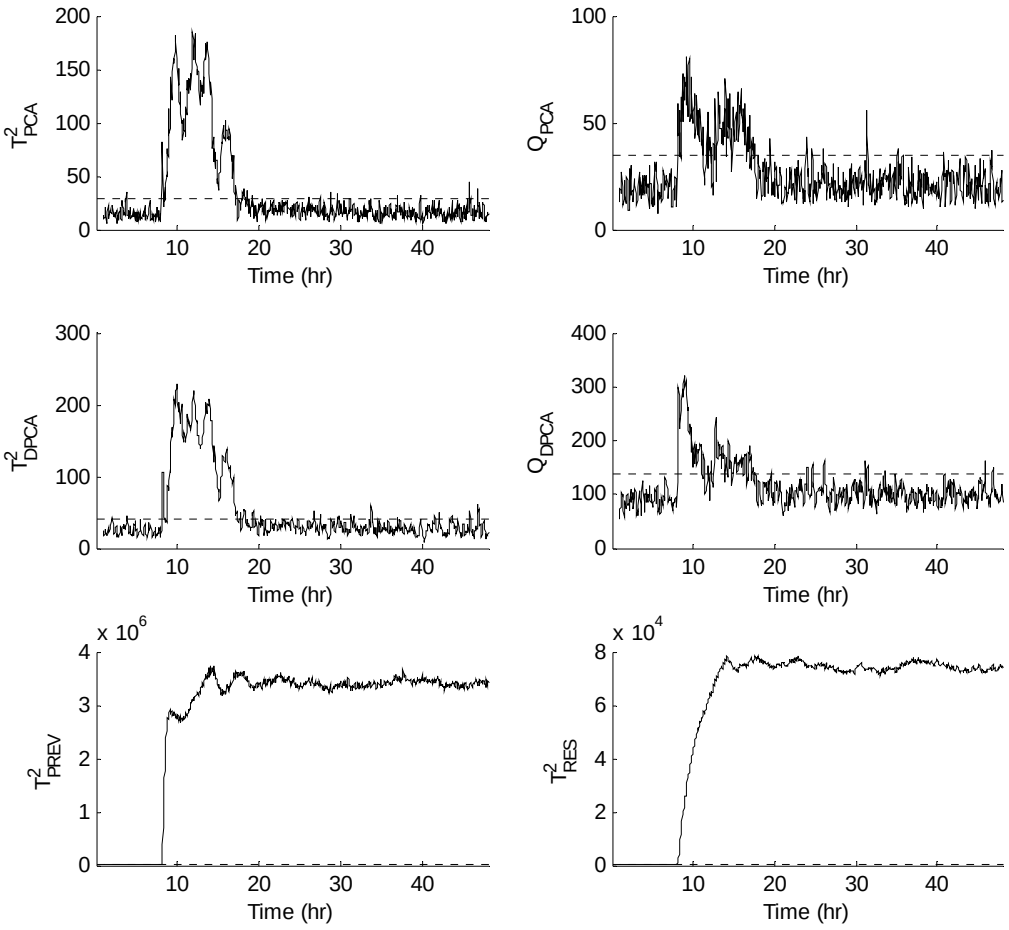


Figure 6.15 The multivariate statistics under test for Fault 5: PCA statistics (first or top row), DPCA statistics (second or middle row) and DPCA-DR statistics (third or bottom row).

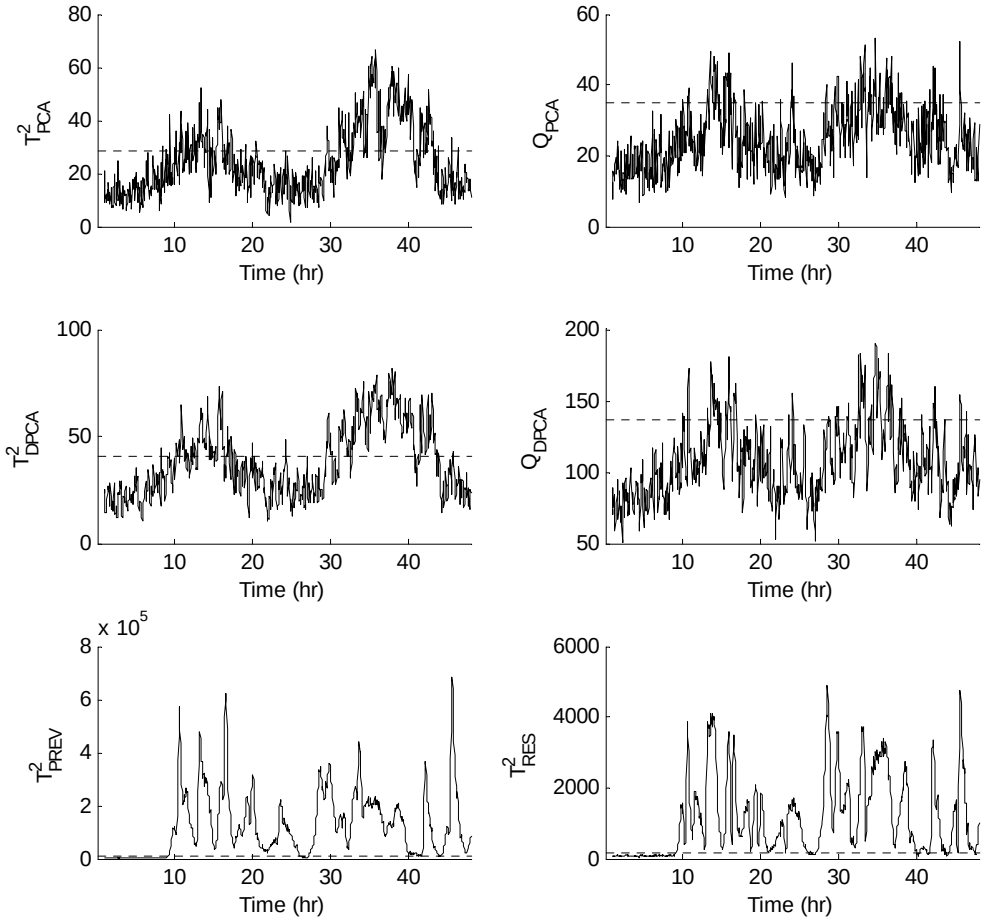


Figure 6.16 The multivariate statistics under test for Fault 10: PCA statistics (first or top row), DPCA statistics (second or middle row) and DPCA-DR statistics (third or bottom row).

To confirm the overall superiority of the DPCA-DR statistics in this case study, a paired t-test between all the statistics was conducted as presented in Table 6.6. Note that as the monitoring statistics were implemented over the same data sets, they are paired by design in this comparison study. The test statistic is given by $t_0 = \bar{D} \Big/ \left(s_D / \sqrt{n} \right)$, where $\bar{D}$ is the sample average of the differences between two methods under analysis in the n different testing conditions, D_1, D_2, … , D_n, and s_D is the sample standard deviation of these differences [39]. From this analysis it can be concluded that, with a 5% significance level, the DPCA-DR statistics are indeed significantly better than all the PCA and DPCA monitoring statistics.

Another advantage of the DPCA-DR method is the lower autocorrelation levels of its monitoring statistics, where much of its success may lie, as this characteristic makes the DPCA-DR statistics more reliable and consistent with the type of control charts used to monitor them (Figure 6.17).

Table 6.6 *p*-values for the paired t-test involving the detection rates obtained with method A (see first column) and method B (see first line), on all simulated faults, along with the signal of the test statistic, *i.e.* sign(t_0). For instance, a plus (+) signal, indicates that method A leads to higher detections rates, on average, when compared to method B. Values in bold indicate *p*-values lower than 0.05 (i.e., statistically significant differences at this level).

A \ B	PCA		DPCA		DPCA-DR	
	T^2	Q	T^2	Q	T^2_{PREV}	T^2_{RES}
PCA						
T^2		0.388 (-)	0.414 (-)	0.152 (-)	**0.002 (-)**	**0.003 (-)**
Q	0.388 (+)		0.540 (+)	**0.046 (-)**	**0.004 (-)**	**0.008 (-)**
DPCA						
T^2	0.414 (+)	0.540 (-)		0.257 (-)	**0.002 (-)**	**0.004 (-)**
Q	0.152 (+)	**0.046 (+)**	0.257 (+)		**0.007 (-)**	**0.013 (-)**
DPCA-DR						
T^2_{PREV}	**0.002 (+)**	**0.004 (+)**	**0.002 (+)**	**0.007 (+)**		0.115 (+)
T^2_{RES}	**0.003 (+)**	**0.008 (+)**	**0.004 (+)**	**0.013 (+)**	0.115 (-)	

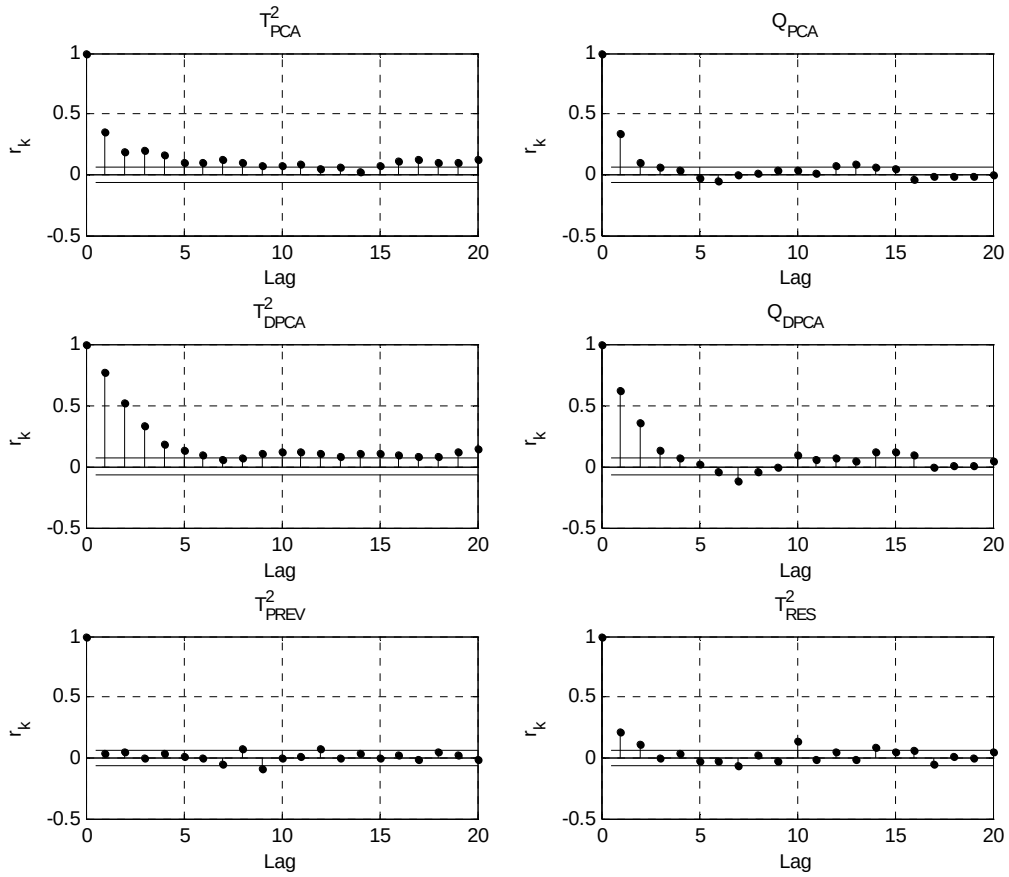


Figure 6.17 Autocorrelation plots for the monitoring statistics when the process is under normal operation conditions (data set with no faults). The proposed DPCA-DR statistics present the lowest levels of correlation among all the studied ones. PCA statistics - first or top row -, DPCA statistics - second or middle row -, DPCA-DR statistics - third or bottom row -.

6.4 Conclusions

In this chapter the problem of monitoring large processes with autocorrelated or dynamic data was addressed. Twenty two monitoring statistics were presented in a systematic way and studied, including ten statistics introduced for the first time. They encompass a variety of methods, including PCA, DPCA, PLS, time series and conditional estimation frameworks, and it was concluded that those derived from the class of DPCA-DR statistics tend to show, in general, better detection performances. Such monitoring statistics require a proper method to estimate the number of lags necessary to construct the DPCA model, for which two alternatives were considered, being observed that the LS2 methods led to the best results.

The DPCA-DR based monitoring statistics also presented better results when applied to the Tennessee Eastman benchmark process. From the analysis of these results, it was verified that DPCA-DR was capable to efficiently detect 19 out of the 21 faults. Moreover, the DPCA-DR statistics also presented the lowest autocorrelation levels and were able to sustain the out-of-control signals throughout the whole faults duration, while PCA and DPCA statistics often return to their in-control regions leading to a false sense of normality.

Consequently, given the consistency of the results obtained, the DPCA-DR statistics seem to be more effective, reliable and consistent regarding their counterparts tested in this study.

7 Multiscale Dynamic Principal Component Analysis

Current industrial processes contemplate a combination of phenomena occurring at different time-frequency scales. Thus, they are multiscale by nature, and the same applies to the data they generate, which require the development and application of suitable monitoring methodologies capable to cope with such distinctive features. However, most of the current monitoring schemes operate in single-scale frameworks and their detection abilities are dependent on the specific scale where the fault manifests itself. For instance, Shewhart control charts are suitable to detect large shifts at the finest scales, while CUSUM and EWMA control charts, depending on their tuning parameters, are better to detect small changes at coarser scales [5, 9, 58]. The same applies to most of the current multi- and megavariate monitoring statistics based on latent variables models, such as PCA, which also model data at a single-scale. Therefore, this characteristic also affects DPCA-DR, which inherits the single-scale nature from PCA and DPCA.

The performance of monitoring schemes based on PCA frameworks can be improved by proper integration with methods that allow for a multiscale representation of process data. One example is multiscale PCA (MS-PCA), proposed by Bakshi (1998) [9], which combines the ability of PCA for decorrelating the variables covariance, with that from wavelets transforms to decorrelate the serial dependencies of signals. The basic procedure underlying MS-PCA consists in first decomposing each variable into multiple scales through the use of a wavelet transform; then PCA models are developed at each scale, based on which process monitoring can be conducted in a simultaneous and independent way; when abnormal, the relevant scales that actually contribute to the fault are selected and used to reconstruct the fault signature at the original time scale; finally, the reconstructed signal is submitted to a confirmatory test, in order to verify if the process is indeed out-of-control, or not. As a result of these decomposition and reconstruction steps, the monitoring procedure becomes sensitive to a wide range of faulty patterns with different magnitudes and localizations in the time-frequency plane [5]. Several improvements and modifications to the base MS-PCA scheme, as well as alternative multiscale monitoring strategies, have been proposed over the last years, as reviewed by

Reis (2005) [121]. As a result of these contributions, the application scope of multiscale methodologies was extended to cases where the data also presents dynamic [122], non-stationary [123] and multiresolution [8] features.

Even though the wavelet transforms have the ability to reduce the signal autocorrelation, the coarser scales coefficients may still present some autocorrelation. This happens because of the moving window approach used to implement the wavelet transform on-line, or even due to the particular dynamic nature of the signals, which leads to serial correlations in the wavelet coefficients. This characteristic cannot be handled with PCA and therefore a DPCA-DR approach has the potential to increase the fault detection capability of the monitoring scheme based on MS-PCA. In this context, and given the superiority of the DPCA-DR scheme relatively to its PCA-based counterparts, as illustrated in Chapter 6, it becomes natural to develop a multiscale monitoring scheme based on DPCA-DR. The base procedure follows the main steps proposed by Bakshi (1998) [9] for MS-PCA, which will be described and compared in the following sections.

7.1 Multiscale DPCA-DR (MS-DPCA-DR)

In this section, a multiscale version of DPCA-DR is proposed in order to improve its ability to detect faults with a variety of time-frequency features. The proposed methodology is based on the work of Bakshi (1998) [9] where monitoring schemes based on PCA model are simultaneously applied to different scales in order to determine if the process is under a state of statistical control. The original procedure was described in more detail in Section 3.4, and will be here adapted as follows.

As for MS-PCA, the first step of the proposed multiscale DPCA-DR (MS-DPCA-DR) involves the decomposition of the data using a wavelet decomposition with depth J_{max}. Then, an MSPC procedure based on DPCA-DR is employed independently at each scale, in order to select the wavelet coefficients to be included in the reconstructed data and the corresponding $T^2_{PREV}{}^{(j)}$ and $T^2_{RES}{}^{(j)}$ statistics are determined for $j = 1,\ldots,J_{max}+1$, where j is the scale index, representing the J_{max} decomposition levels for detail coefficients (d_j) and the coarsest approximation level ($a_{J_{max}}$). For each scale, the monitoring statistics are

compared against their control limits and if either $T^{2}_{PREV}{}^{(j)}$ or $T^{2}_{RES}{}^{(j)}$ exceed such limits, the corresponding scale is used to reconstruct the data. The reconstructed data is finally subjected to a confirmatory monitoring stage, also using a DPCA-DR control scheme.

To implement this procedure it is required to construct a DPCA-DR model for each wavelet scale as well as for each of the possible reconstruction combinations. The modeling stage of the wavelet scales is straightforward, with only a minor remark on the determination of the UCL limits, as explain later. As for the reconstructed data, in order to avoid the combinatorial computation of all possible reconstructions, a database composed by two independent NOC data sets are used to determine the DPCA-DR parameters (if they were not previously determined) by request. The first data set is used to reconstruct train data, with the same scales as the monitoring data, and employed to construct the DPCA-DR model. After that, the second data set is applied to determine the UCLs by trial and error adjustment in order to ensure that a pre-established significance level (α) is obtain. The choice of this approach is justified by the fact that the monitoring statistics often violate their construction assumptions (such as Gaussian distribution of the data or *i.i.d*. assumptions) and therefore the theoretical limits do not reflect with the actual observed false alarm rate.

Regarding the UCL for each of the monitoring statistics, it is important to note that, overall, $2\times\left(J_{max}+1\right)$ control charts are under use (i.e., two control charts for each wavelet scale). Therefore, in order to maintain the desired overall significance level (α), the significance level for each control chart (α_A) is corrected by application of the Bonferroni inequality as $\alpha_A = \alpha/2\left(J_{max}+1\right)$. Nevertheless, this issue does not have a direct impact on the method's false alarm rate, since this is ultimately defined by the final confirmatory monitoring stage based on the DPCA-DR model applied to the reconstructed data.

The described MS-DPCA-DR procedure is summarized in Table 7.1 in the form of a pseudo-code and schematically represented in Figure 7.1.

Table 7.1 Pseudo-code for the proposed MS-DPCA-DR methodology.

1. Get window of data with dyadic length containing the current observation (length equal to $2^{J_{max}}$);
2. Obtain the current wavelet coefficients at all scales, $j = 1, \ldots, J_{max} + 1$;
3. Implement DPCA-DR-based MSPC at each scale, using $T^{2}_{PREV}{}^{(j)}$ and $T^{2}_{RES}{}^{(j)}$ statistics, and select the relevant scales, i.e. the scales where $T^{2}_{PREV}{}^{(j)} \geq ULC\left(T^{2}_{PREV}{}^{(j)}\right)$ or $T^{2}_{RES}{}^{(j)} \geq ULC\left(T^{2}_{RES}{}^{(j)}\right)$;
4. Reconstruct data using the relevant scales;
5. Obtain the corresponding DPCA-DR model from the database. If a DPCA-DR model is not available,
    a. Reconstruct train and validation data with the same relevant scales;
    b. Determine the DPCA-DR parameters from the reconstructed train data;
    c. Determine the UCL from the reconstructed validation data;
    d. Save DPCA-DR model and UCL on database.
6. Implement DPCA-DR-based MSPC on the reconstructed data, using $T^{2}_{PREV}{}^{(rec)}$ and $T^{2}_{RES}{}^{(rec)}$ statistics;
7. Compare $T^{2}_{PREV}{}^{(rec)}$ and $T^{2}_{RES}{}^{(rec)}$ against their UCL.

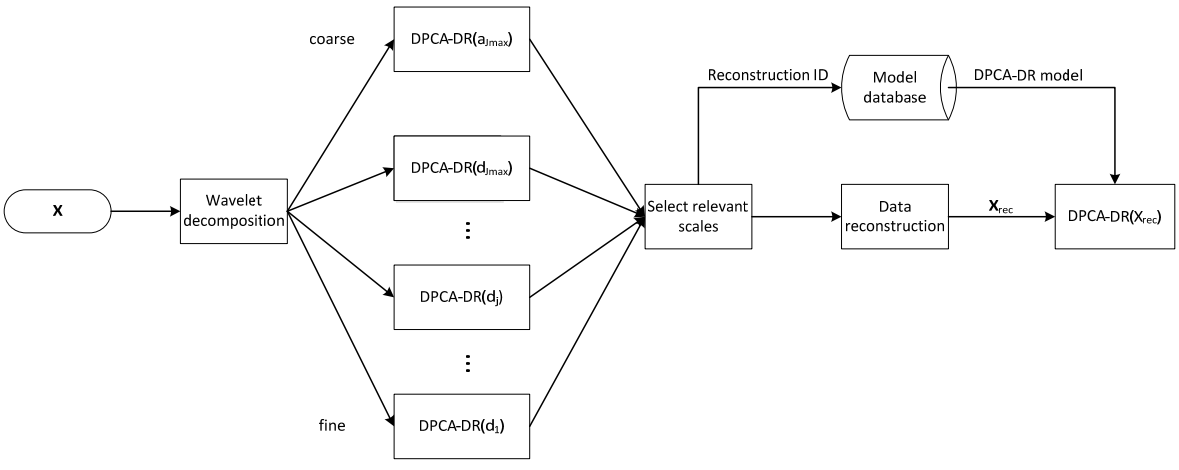


Figure 7.1 Schematic representation of the proposed MS-DPCA-DR methodology.

7.2 Comparative Study

In order to assess the characteristics of the proposed multiscale methodology, a comparison study was conducted including the monitoring approaches based on the following base models: PCA, DPCA-DR, MS-PCA and MS-DPCA-DR. This comparison is made in terms of their detection rates (i.e. the ratio between the number of alarms and simulation observations) when a continuous stirred-tank reactor (CSTR) is subjected to step perturbations on the heat transfer coefficient and discharge coefficient. The false detection rate of all methods was set to 1% and the detection rates were computed based on 2000 observations, repeated 100 times in order to determine their reliability.

In the studied CSTR model, an endothermic reaction of the type A → B takes place in a reactor with an heating jacket. The system inputs are the feed stream concentration (C_{A0}) and temperature (T_0) and the heating fluid inlet temperature (T_{j0}). The system outputs are the CSTR level (h), concentration of compound A (C_A), temperature (T) and the heating fluid outlet temperature (T_j). Moreover, the CSTR level (h) and temperature (T) are under a PI feedback control loop.

From the distribution of the variance and cumulative variance at different scales, depicted in Figures 7.2 and 7.3, it is observed that the data has multiscale characteristics. The input variables (C_{A0}, T_0 and T_{j0}) are generated from random draws of the normal distribution and, therefore, they are essentially white noise. Consequently, most of their variability is captured by the detail coefficients d_1 to d_3. As for the outputs, the CSTR level's energy (h) is mainly localized in d_4 to d_6, the concentration (C_A) dynamics in d_5 to a_8, the temperature (T) in d_3 to d_6 and the heating fluid outlet temperature (T_j) in d_3 to d_5. Even though most of their variability overlaps on several scales, there is a clear difference in their distribution. Therefore, it is expected that multiscale techniques present better detection capabilities as a result of their higher flexibility to describe more accurately such NOC multiscale dynamic features. This analysis also shows that most of the variability can be properly described by a decomposition level (J_{max}) equal to 8.

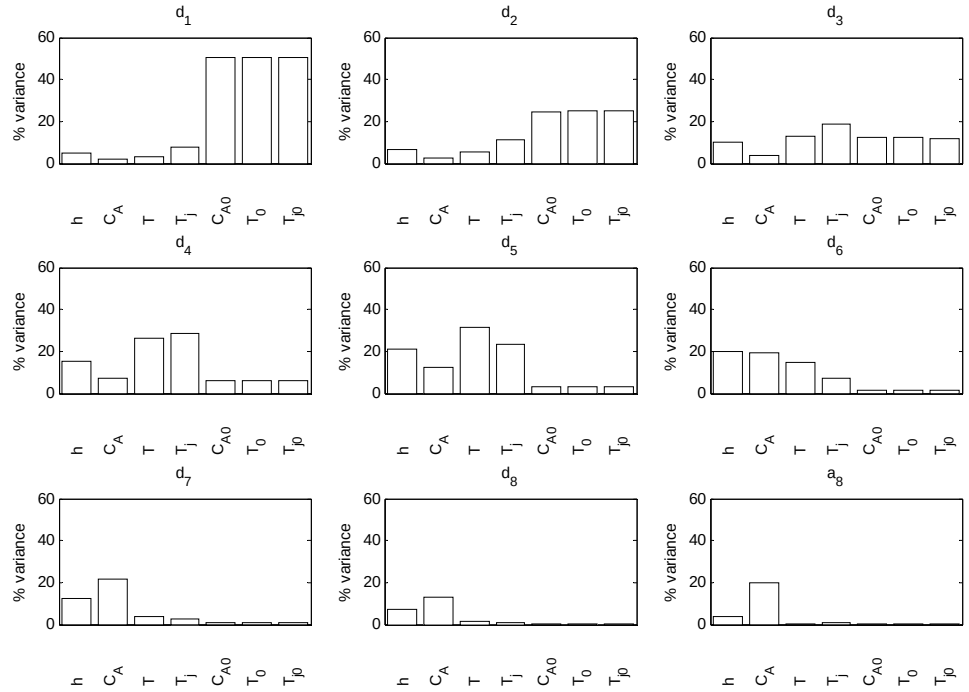


Figure 7.2 Distribution of the variance of the CSTR process variables for J_{max} = 8.

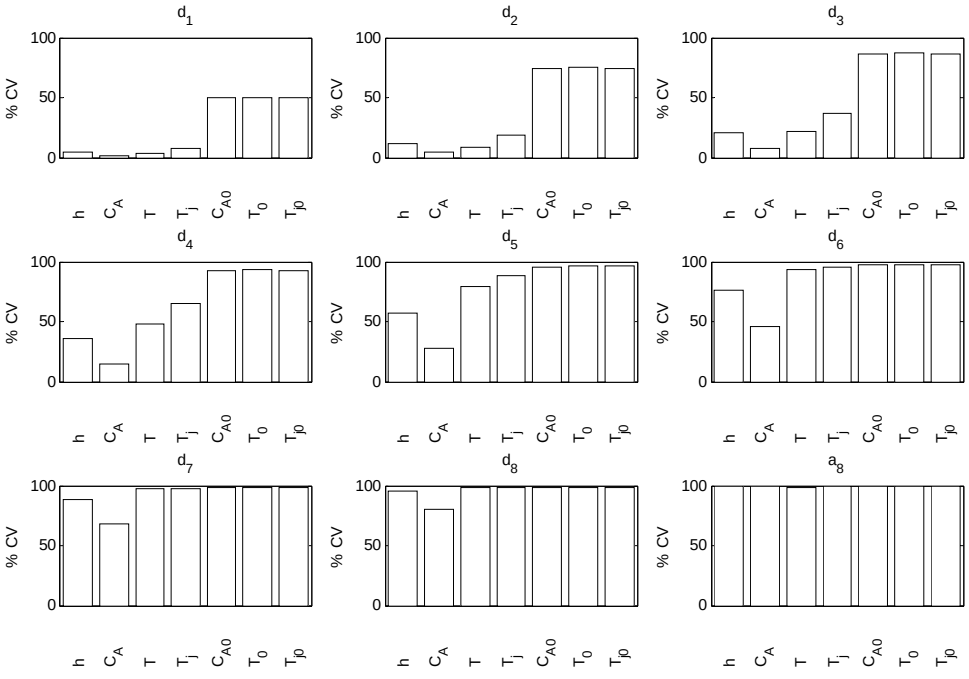


Figure 7.3 Distribution of the cumulative variance (CV) of the CSTR process variables for J_{max} = 8.

When the system is subjected to a perturbation on the heat transfer coefficient, it is observed a change on the heating fluid outlet temperature (T_j). Under normal operation conditions, this variable is mostly described by the detail coefficients d_3 to d_5, which are related to high frequency signals. However, when a fault occurs, a low frequency change is observed in T_j (see Figure 7.4 (h)). Consequently, the single-scale procedures (PCA and DPCA-DR) are not able to promptly detect this fault, since the signal is masked by the high frequency components and the fault contributions passes mostly unnoticed (see Figure 7.4 (a)). On the other hand, both MS-PCA and MS-DPCA-DR are capable to isolate the low frequency (that in this case, appear in a_8, Figure 7.4 (h)) and remove the in-control scales from the reconstructed signal. However, as the fault is located at a coarser scale, the signal presents high levels of autocorrelation, which causes the monitoring statistics to present some detection delay. A delay is also observed when the process returns to normal operation conditions. These situations are a result of the application of a moving window during the wavelet transformation.

Another interesting observation is the relatively low ability of MS-DPCA-DR to cope with wavelet coefficients dynamics. This result was not expected since the single-scale DPCA-DR is characterized by monitoring statistics with low levels of autocorrelation. Nevertheless, the increase in performance is significant, as can be verified in Figure 7.5, where the fault detection rates are represented for simulated faults of magnitude δ times the reference heat transfer coefficient.

Similar results were obtained when the system was subjected to perturbations on the discharge coefficient (see Figure 7.6). These results show that the MS-DPCA-DR is consistently better than its counterparts and capable to detected even finer faults than MS-PCA.

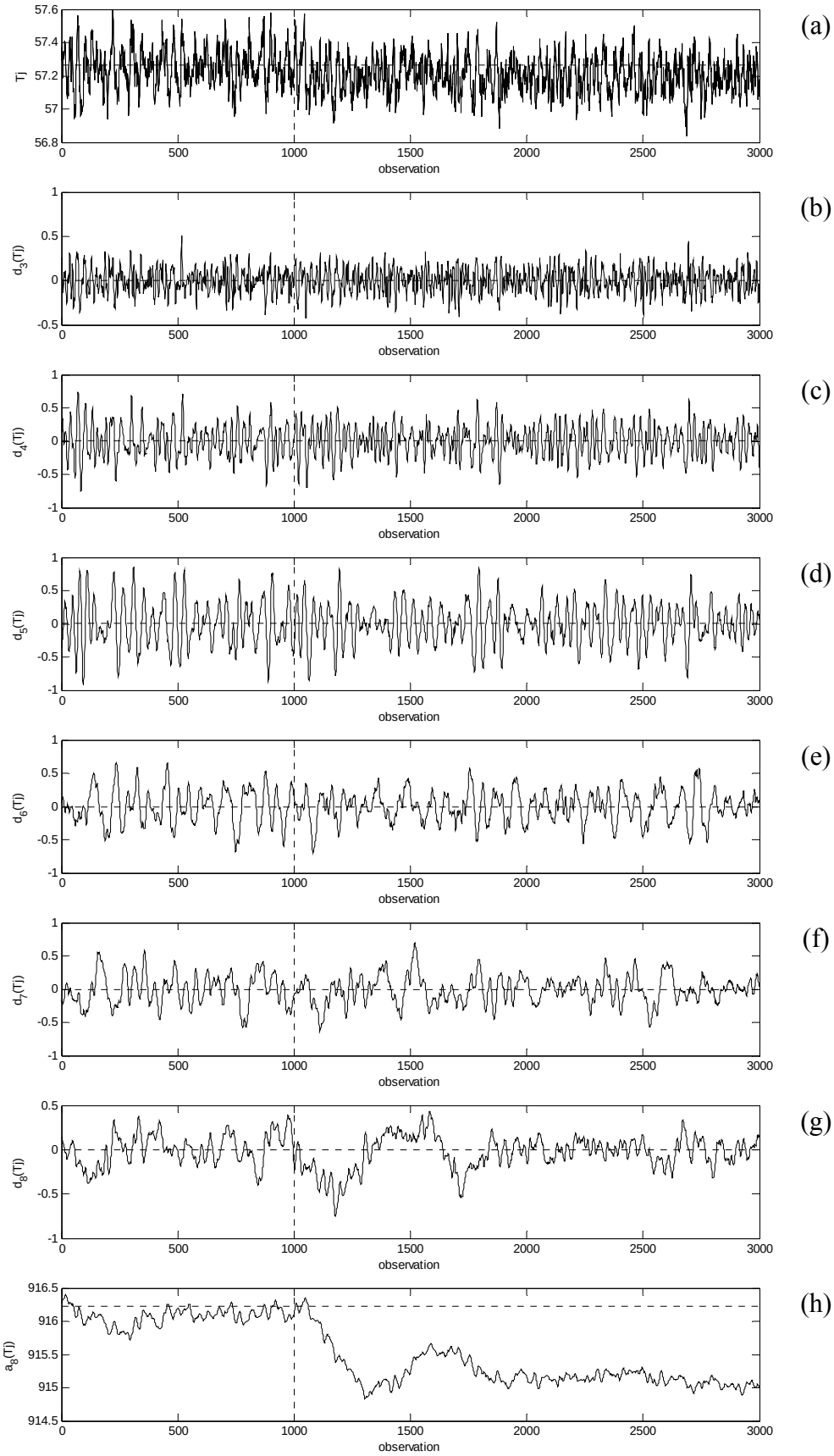


Figure 7.4 Graphical representation of T_j over time when the CSTR model is subjected to a step deviation in the heat transfer coefficient of magnitude $\delta = 1.005$ after observation 1000: (a) original data; (b) wavelet coefficient at scale d_3; (c) wavelet coefficient at scale d_4; (d) wavelet coefficient at scale d_5; (e) wavelet coefficient at scale d_6; (f) wavelet coefficient at scale d_7; (g) wavelet coefficient at scale d_8; (h) wavelet coefficient at scale a_8;

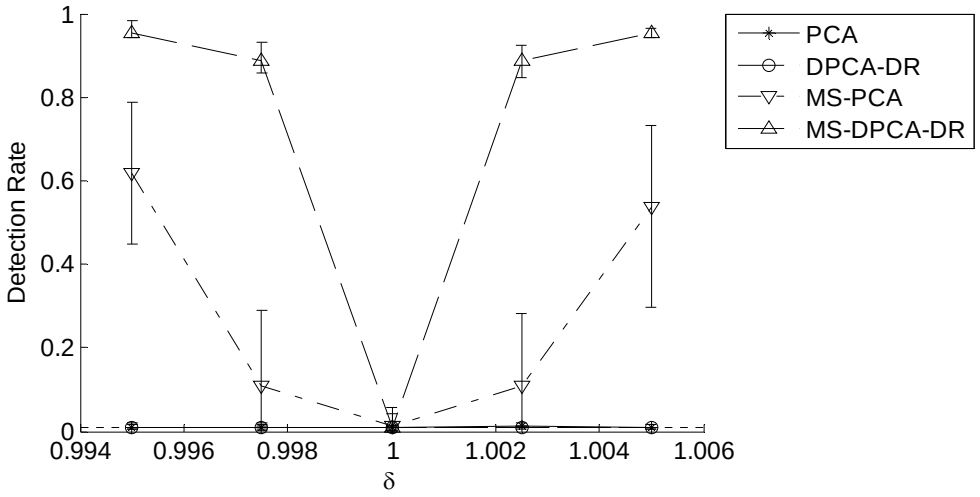


Figure 7.5 Curve of the detection rates for a step perturbation in the heat transfer coefficient of the CSTR system. δ is a multiplicative factor that introduces a deviation in the heat transfer coefficient (under NOC $\delta = 1$).

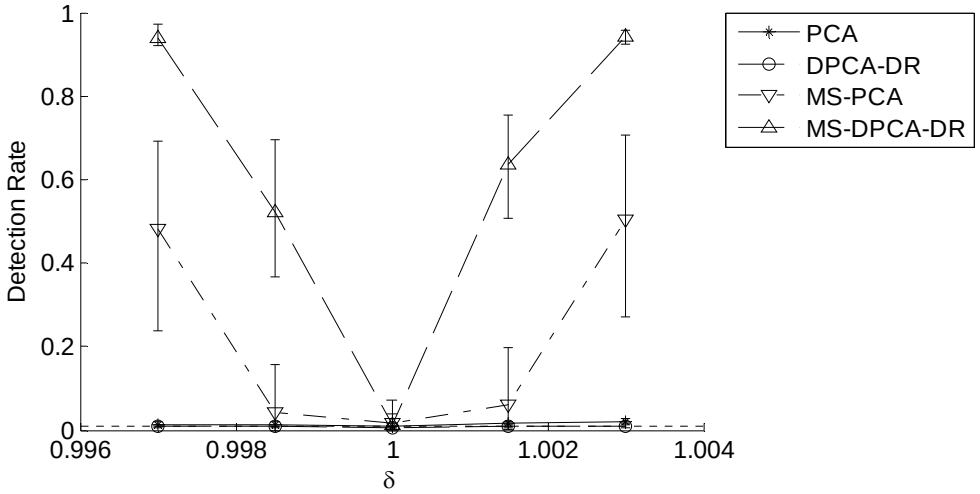


Figure 7.6 Curve of the detection rates for a step perturbation in the discharge coefficient of the CSTR system. δ is a multiplicative factor that introduces a deviation in the discharge coefficient (under NOC $\delta = 1$).

7.3 Conclusions

The combined used of latent variables models and wavelet transforms is a powerful tool to monitor systems with multiscale characteristics. These characteristics are common to most of the industrial process and therefore methodologies that take advantage of such features have the ability to better explain the system under monitoring.

DPCA-DR adds detection capabilities to the static PCA by both inclusion of time-shifted variables (which allows for the capture of dynamic features) and reduction of the monitoring statistics' autocorrelation. This is an important factor when it is combined with the wavelet transform, since even for stationary system, some autocorrelation can be found in the wavelet coefficients computed on-line by a moving window procedure. Interestingly, DPCA-DR is not capable to remove all the autocorrelation introduced by the wavelets computation scheme, especially for wavelet coefficients corresponding to lower frequencies. Nevertheless, the final performance is greatly increased and therefore this situation is taken as an acceptable side effect.

The results obtained in this study also showed that the proposed methodology is capable to focus on the time-frequency scales related with the fault. In the case study, the faults occurred at low frequency scales, which were masked by the higher frequency components when a single-scale approach was used. Therefore, the single-scale methodologies had low sensitivity to these faults. However, when the data is "filtered-out" by the multiscale monitoring methodology and only the relevant scales are used to monitor the process, a better isolation of the faulty signature is obtained and consequently the detection performance is improved.

Part IV – MSPC – Off-line and On-line Monitoring of the Process Correlation Structure

8 Introduction: the Networked Structure of Processes

Analyzing the contributions for monitoring the correlation structure that have been proposed in the literature (see Chapter 4), it is possible to verify that all of them are strictly based on the marginal covariance matrix of process data. Even multivariate statistical process control based on principal component analysis (MSPC-PCA) [15-17], which implicitly has the ability to detect changes in the correlation structure of data through the *Q* or *SPE* statistics, is based on the marginal covariance matrix. However, as process variables may present a significant marginal covariance even though they do not directly interact in a causal way (as long as they are affected by some common causes of variation), monitoring procedures based on this quantity are unable, by design, to efficiently detect and discern changes in the local causal correlation structure.

In order to access and use local information of the correlation structure of variables, alternative measures of association must be adopted in process monitoring procedures. Partial correlation is one such quantity [124]. It evaluates the correlation between pairs of variables, after controlling for the effect of others, i.e., after removing their indirect effect in inducing any association between the variables under analysis. As partial correlation coefficients are able to retain, to a larger extent, local information between the direct association of variables (even though in a non-causal sense, i.e., without the associated causal directionality), they can provide a finer map of the direct association structure connecting process variables. Thus, statistical process control (SPC) based on partial correlation should be able to detect changes in the local structure of variables (fault detection) and to identify the root causes of specific process upsets (fault diagnosis) in a more effective and efficient way. Therefore, the total time invested in fault detection and diagnosis activities may be improved using such an alternative measure of local association, as both the primary detection and especially the subsequent diagnosis process will be improved. Moreover, it is worth noticing that, even though partial correlations have been proposed a short time after PCA, their potential to improve process monitoring and fault detection activities have not yet been explored. This fact is quite surprising, as detection and, in particular, diagnosis tasks, can potentially benefit significantly from the

use of local measurements of association. This is in major contrast with the widely explored use of marginal correlation approaches such has PCA based MSPC and most of the current monitoring methodologies.

Some causal methodologies have also been proposed, such as those developed by Bauer *et al.* (2007) [125], that uses transfer entropy in order to identify the directionality of the fault's propagation path, and by Yuan and Qin (2012) [126], where a combination of Granger causality and PCA are employed to perform features selection for locating the origin of faults with oscillatory characteristics. However, these methodologies are strongly oriented to fault diagnosis rather than fault detection, which is an obvious pre-requirement before their application. Chiang and Braatz (2003) [117] also suggested the combined use of the Kullback–Leibler information distance and the process causal map to detect and diagnose faults. Yet, this approach requires the knowledge of the causal map in the form of a digraph. Therefore, as partial correlations convey information about the inner relationship between the variables, without the need of *a priori* information about the system structure, they will be the main focus of the proposed approaches for monitoring the process correlation structure.

In this context, the concepts behind partial correlations are introduced in the next section. Based on their properties, several methods to monitor changes on the partial correlation coefficients with the potential to detect finer local changes in the process structure, and to identify their source in a more effective way, are presented and studied in the following chapters. Another key contribution of this work to the process monitoring is the use of sensitivity enhancing transformations (SET), which will be addressed in Chapter 9. The SET's exploit the process networked structure in order to decorrelate the measured variables in a meaningful way. Interestingly, most of the current approaches to monitor the process correlation resort to some sort of decorrelation transform [51, 79, 80, 88], but none of them actually benefit from such variable transformations. In fact, they are mostly used only for simplification purposes. Reynolds and Cho (2006) [81] briefly studies the effects of employing regressed-adjusted variables, as proposed by Hawkins (1993) [127], instead of the original variables. Yet, as will be shown latter on Chapter 11, this particular transformation does not always lead to the best results.

The proposed procedures will be presented and compared with marginal-based approaches available in the literature in Chapters 10 for the off-line case (i.e. based on

subgroups of non-overlapping moving windows) and in Chapter 11 for the on-line case (i.e. based on individual observations). Furthermore, a conceptual relation between these two monitoring approaches (off-line and on-line) will be presented in Chapter 12, which allows for the analysis of their properties in a unified way.

8.1 Local Association Measures

The current approaches to monitor process correlation structure and multivariate dispersion are based on the application of a sequence of statistical hypothesis tests in order to determine if some significant change has occurred in the covariance matrix. However, the covariance matrix does not convey a detailed information about the local association structure of the system [128]. Therefore methodologies based on the marginal covariance present intrinsic limitations regarding the detection capability to changes in the local structure of the process variables' relationships, as well as to the subsequent diagnosis of the fault's origin once it is signaled. For instance, two variables, x and y, can be related in several different ways such as (i) direct relation $x \rightarrow y$, (ii) co-regulated by a third variable z ($z \rightarrow x$ and $z \rightarrow y$) or (iii) indirect relation $x \rightarrow z \rightarrow y$ [128]. In all these examples, the correlation between x and y may be similar, and a change on their underlying relationship may pass undetected by just monitoring the marginal correlation. An example of such situation is represented in Table 8.1 for two of the cases mentioned. In both cases, all marginal correlations are close to 0.99, which indicates that all variables may have a similar behavior. However, when 1st order partial correlation are employed, only the variables with a direct association have a significant correlation value (in these cases, close to 0.71), while the variables with an indirect relation (x and y) have a 1st order partial correlation close to zero (i.e., there is no direct relationship between them).

From this simple example it is obvious the limitation of using marginal-based measures of association and the pertinence of adopting local measures in order to capture the inner associations between variables. The basic idea of partial correlations is to remove the effect of third-party variables before checking for an association between the two designated variables. Therefore the correlation between two variables is quantified, after conditioning upon (i.e., controlling for, or holding constant) one or several other variables [129, 130]. In the above example, the partial correlation between x and y conditioned to z

would remove the common effect of z on x and y, providing a clearer picture of the local correlation structure between these variables. More specifically, this can be achieved by first regressing x on z, and y on z, after which both regression residuals are saved. These residuals are the parts of x and y that are uncorrelated with z. The correlation between the two residual vectors corresponds to the partial correlation between x and y conditioned on z ($r_{xy.z}$). A pseudo-code for this regression based approach is presented in Table 8.2.

Table 8.1 Marginal and partial correlations of two causal networks with different connective structures. Solid arrows represent causal relationships and dashed arcs represent indirect relationships. The values displayed are the sample correlations computed from 3000 observations.

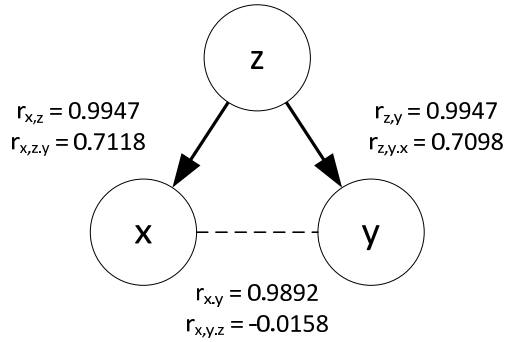
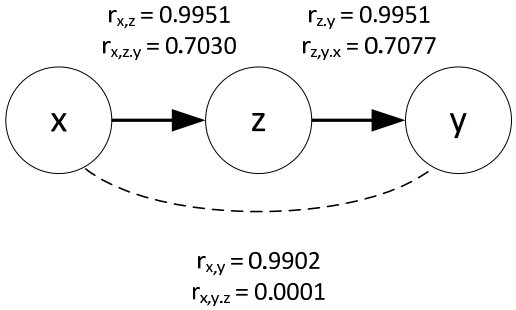
Network	
(a) co-regulation by a third variable	(b) indirect relation
	

Table 8.2 Pseudo-code for the computation of partial correlation.

1. For the par of variables (x,y), fit a regression model with $\mathbf{Z} = \left[\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_q\right]$ as the matrix of controlled variables:
   a. $\mathbf{x} = \mathbf{Z}\mathbf{B}_\mathbf{x}$, where $\mathbf{B}_\mathbf{x} = \left(\mathbf{Z}^\mathrm{T}\mathbf{Z}\right)^{-1}\mathbf{Z}\mathbf{x}$;
   b. $\mathbf{y} = \mathbf{Z}\mathbf{B}_\mathbf{y}$, where $\mathbf{B}_\mathbf{y} = \left(\mathbf{Z}^\mathrm{T}\mathbf{Z}\right)^{-1}\mathbf{Z}\mathbf{y}$.
2. Compute the regression residuals:
   a. $\mathbf{e} = \mathbf{x} - \mathbf{Z}\mathbf{B}_\mathbf{x}$;
   b. $\mathbf{f} = \mathbf{y} - \mathbf{Z}\mathbf{B}_\mathbf{y}$.
3. Compute the q^{th} order partial correlation as:

$$r_{\mathbf{xy}\cdot\mathbf{Z}} = \frac{\mathrm{cov}\left(\mathbf{e},\mathbf{f}\right)}{\sqrt{\mathrm{var}\left(\mathbf{e}\right)\mathrm{var}\left(\mathbf{f}\right)}}$$

The order of the partial correlation coefficient is determined by the number of variables it is conditioned on (which is q in Table 8.2). For instance, $r_{xy.z}$ is a 1st order partial correlation coefficient because it is conditioned solely on one variable (z). Partial correlations can be obtained either using the above referred regression based approach, or through analytical formulas, in a recursive way. Equations (8.1) to (8.3) illustrate the computation of the partial correlation coefficients for orders 0, 1 and 2. Similar equations exist for higher order representatives.

0th order partial correlation:

$$r_{xy} = \frac{\operatorname{cov}(x, y)}{\sqrt{\operatorname{var}(x)\operatorname{var}(y)}} \tag{8.1}$$

1st order partial correlation:

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{\left(1 - r_{xz}^2\right)\left(1 - r_{yz}^2\right)}} \tag{8.2}$$

2nd order partial correlation:

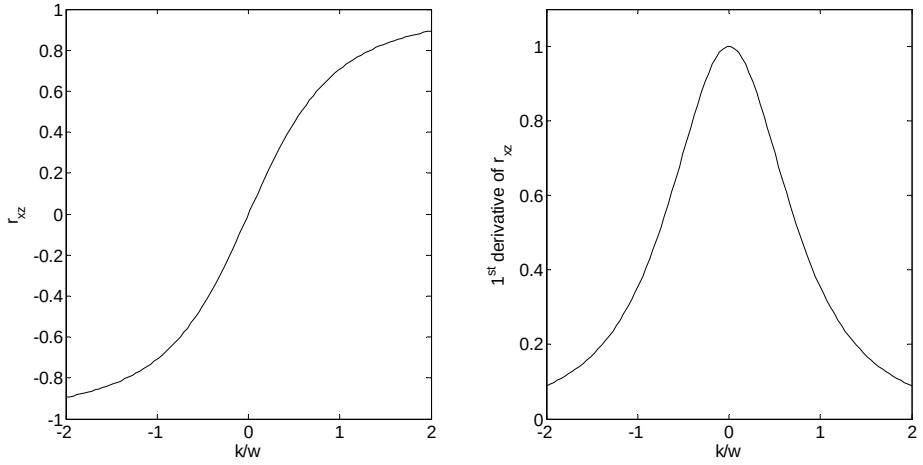
$$r_{xy.zq} = \frac{r_{xy.z} - r_{xq.z}r_{yq.z}}{\sqrt{\left(1 - r_{xq.z}^2\right)\left(1 - r_{yq.z}^2\right)}} \tag{8.3}$$

As partial correlation coefficients bring out differences between the direct and indirect relationships between variables, it is possible to apply a thresholding procedure in order to identify connected variables and variables that are not directly associated. The cut-off values depend on the statistical significance to observe in this process [130, 131]. This scheme has also been applied in variables selection and classification methods [128, 129, 132]. Partial correlations have also a great potential to improve the activities related with statistic process monitoring, namely fault detection and diagnosis. This will be the topic of the next chapters of this thesis.

9 Sensitivity Enhancing Transformations for the Monitoring of the Process Correlation Structure

Analyzing closely the sensitivity of correlation measures to changes in the population features, it is possible to observe that the detection potential to changes in the correlation coefficients is highly dependent on their nominal values (i.e., under normal operation conditions). More specifically, it is related to the strength of the relationship between the two variables under analysis: the higher the strength of the association, the more difficult it becomes to detect a change in its value. On the contrary, if two variables are initially uncorrelated, any change on their association status is detected in a more sensitive way.

In order to exemplify this finding, consider that $x = kz + w\varepsilon$, where both z and ε follow *i.i.d.* $N(0,1)$ distributions. In this case, the correlation of interest, say that between x and z (r_{xz}), can be represented as a function of the ratio k/w according to Figure 9.1 (a). It is apparent that there can be rather different sensitivities of r_{xz} to changes on the population constants (k and w), depending on the particular position in the curve. The 1st derivative of r_{xz} (Figure 9.1 (b)) is a proper measure of such sensitivity for detecting changes on r_{xz}, and one can verify that changes are more difficult to detect as $|r_{xz}|$ gets closer to 1 and easier when $|r_{xz}|$ is close to 0. The maximum sensitivity is reached at $r_{xz} = 0$, which corresponds to the case where variables are uncorrelated (either $k = 0$ or $w \to \infty$, i.e. $w \gg k$). Therefore, changes in the correlation coefficients will be easier to detect when a previously non-existent relationship gives rise to some association between the variables. Furthermore, it was also found out that the same behavior occurs with partial correlations. Thus, in order to exploit this feature in process monitoring and fault detection activities, the original variables should be preliminarily rotated to a regression subspace of uncorrelated variables, by application of what will be called here, a "sensitivity enhancing transformation".



(a) (b)

Figure 9.1 Graphical representation of (a) the effect of k/w on the correlation (r_{xz}) and (b) first derivative of r_{xz}.

One of the transformations usually applied in the treatment of multivariate Gaussian data relies on the Cholesky decomposition [60], which factorizes the covariance matrix, $\mathbf{\Sigma}$, into a lower triangular matrix $\mathbf{L}$, such that,

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^{\mathrm{T}} \tag{9.1}$$

This matrix, can then be used to obtain uncorrelated variables with unit variance through the transformation [60]:

$$\mathbf{u} = \mathbf{L}^{-1}\left(\mathbf{x} - \boldsymbol{\mu}\right) \tag{9.2}$$

This transformation was adopted by Hawkins and Maboudou-Tchao (2008) [82] on their extension of the W statistic to on-line monitoring and corresponds to a succession of regression operations, where u_i is the residual of the regression of x_i on $x_1,\ldots, x_{i-1}$, rescaled to unit variance. Yet, the W statistic is invariant to any linear transformation and therefore does not benefit from such transformation. On the rest of this work the use of the original, untransformed data will be referred by T_X, whereas the use of this particular data transformation will be denoted as T_{Ch}.

The transformation based on the Cholesky decomposition described on Equation (9.2) is only suitable for stationary linear systems. In order to extend its application to dynamic and non-linear systems, it is proposed the addition of time-shifted variables or polynomial terms (depending of the situation) to the data matrix. This type of approach is in

accordance with other methodologies, such as DPCA which incorporates time-shifted variables to implicitly model an AR process [27]. Likewise, the approximation of non-linear functions by polynomial terms is justifiable by the Taylor series expansion theorem since, under normal operation conditions, the processes tend to experiment only mild fluctuations around the nominal operation state. These additional variables should be placed at the beginning of the extended data matrix, $\tilde{\mathbf{X}}$, which for the case of the inclusion of time-shifted variables becomes:

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}(l) & \cdots & \mathbf{X}(1) & \mathbf{X}(0) \end{bmatrix} \tag{9.3}$$

where $\mathbf{X}(j)$ is an $(n \times m)$ matrix of variables shifted j times into the past (i.e., with j lags). When the inclusion of polynomial terms is required, $\mathbf{X}(j)$ is replaced by powers of the type $x^{(j+1)}$.

After this step, the regular Cholesky decomposition can be performed, Equation (9.4), from which a new set of uncorrelated variables are obtained by Equation (9.5).

$$\tilde{\boldsymbol{\Sigma}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\mathrm{T}} \tag{9.4}$$

$$\tilde{\mathbf{u}} = \tilde{\mathbf{L}}^{-1}\left(\tilde{\mathbf{x}} - \tilde{\boldsymbol{\mu}}\right) \tag{9.5}$$

In the case where $\tilde{\mathbf{X}}$ is obtained by the addition of time-shifted variables, only the regression variables related to the present state are of interest (i.e., the last m variables in $\tilde{\mathbf{u}}$), since they correspond to the residuals of the linear regressions of the variables in the present, onto those from the past. Using this procedure, both cross- and auto-correlations are handled simultaneously for whitening the process data at the current time. The same procedure can also be applied to non-linear systems through the use of polynomial terms instead of time-shifted variables. This type of transformation will be referred as $\mathrm{T}_{\mathrm{ChExt}}$.

As stated earlier, the Cholesky decomposition, is essentially a triangularization method that regresses the i^{th} variable onto the other $(i - 1)$ variables that preceded it. Therefore, the sequence by which variables are included in the model is defined by their particular order of appearance in the original data matrix. As this ordering is completely arbitrary, hardly it will provide the best description of the system structure from a cause-effect sense. In fact, the more meaningful variable ordering corresponds to placing the more important variables (i.e., variables that affect most of the others or that appear on the root

nodes of the true underlying variable interaction network) at the beginning of the matrix (first columns), for which some *a priori* knowledge of the process may be necessary. As the base Cholesky decomposition procedure regresses each variable onto all its predecessors, it may end up relating variables that are not originally associated in any causal way, leading to rather poor monitoring performances.

In order to deal with this problem, an alternative transformation is proposed. The goal of the new transformation is to break the relevant variables relationships upon application of a similar linear regression scheme but only on the variables that are indeed related with each other, as exemplified on Figure 9.2 ("breaking" means here the removal of an association after it is properly explained by a model). For such, the relevant edges between variables must be first identified, either with resort to *a priori* knowledge of the process or through the use of some data-driven network reconstruction technique, as the one described in the next paragraph (see also Figure 9.2 (a) and (b)).

A new procedure to identify the edges linking directly associated variables through the use of partial correlations up to the 2nd order is suggested here, which is has some communalities with others proposed in the literature [129, 130, 132]. The main difference factor relies on the inclusion of time-sifted variables or polynomial terms in the regression models, for addressing dynamic and non-linear process features, respectively. By conditioning on these additional variables, this algorithm becomes capable to detect both dynamic and non-linear relationships, resulting in a more accurate identification of the networks structure. Once the connections are established, one has available an undirected network. The causal directions of these undirected edges are determined by applying another algorithm based on the variables' cross-correlation or, alternatively, in the concepts of transfer entropy and Granger causality. A more detailed description of these algorithms will be provided in Section 9.2.

The second step of the transformation involves the regression of each variable onto its parents or causal predecessors (see Figure 9.2 (c)), resulting in a final regression model where only the directly connected variables are considered in order to obtain the new set of residual variables. On dynamic systems, time-shifted variables should also be included on the regression model and for non-linear systems, polynomial terms must also be added. As a possible additional step, a Cholesky decomposition can be applied to the residuals in order to ensure that uncorrelated variables are indeed obtained and also to

accommodate for the possibility that some relationships may not be accurately described by the linear regression models. On the rest of this work, these NETwork-oriented transformations will be referred as T_{Net} (without a Cholesky decomposition of the residuals) and T_{NetCh} (with a final Cholesky decomposition of the residuals).

These new proposed transformations are able to integrate information about the local correlation structure, leading to improvements in the monitoring statistics performance. Furthermore, in the absence of *a priori* knowledge about variables connectivity, the edges linking connected variables can also be estimated through the proposed network reconstruction method, and the nature of the transformations remain, in this case, fully data-driven. The complete set of sensitivity enhancing transformations covered in this study, and the associated nomenclature, is summarized in Table 9.1.

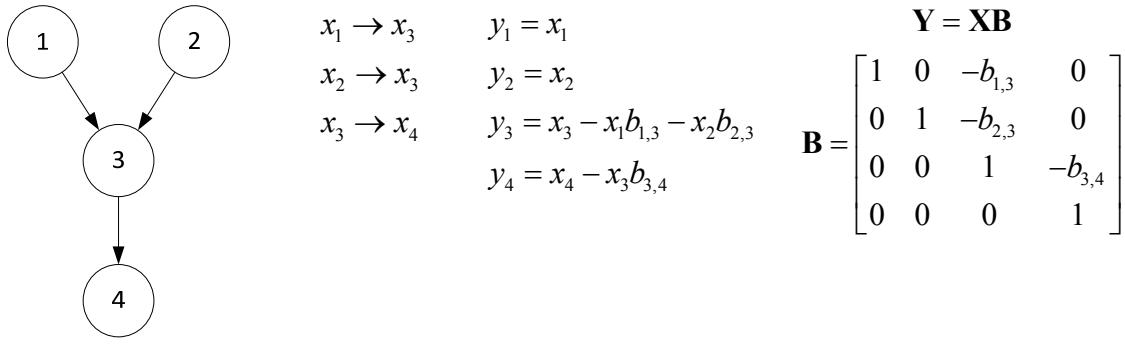


Figure 9.2 Implementation of the proposed transformation procedure: (a) original network, (b) identification of the relevant edges, (c) implementation of successive linear regressions involving the directly connected variables, (d) final model.

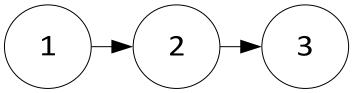
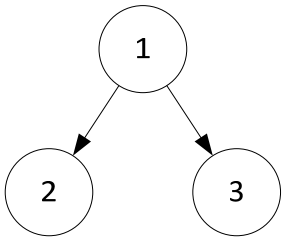
Table 9.1 Nomenclature associated with the sensitivity enhancing transformations covered in this study.

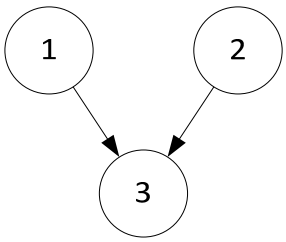
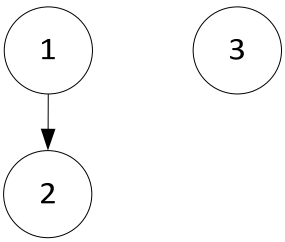
Transformation	Description
T_X	Original variables.
T_{Ch}	Variables transformed using the Cholesky decomposition, Equation (9.2).
T_{ChExt}	Variables transformed using the Cholesky decomposition after incorporation of time-shifted variables and/or polynomial terms, Equation (9.5).
T_{Net}	Variables transformed using a set of linear regression models constructed only from related variables (see Figure 9.2).
T_{NetCh}	Variables transformed using a set of linear regression models constructed only from related variables with an additional Cholesky decomposition of the resulting residuals.

9.1 Example of the Effect of SET in SPC

In order to better illustrate the advantage of employing the sensitivity enhancing transformations, let us consider Table 9.2, where four systems are presented. In each system the variables are connected in different ways, according to the networks represented (for instance, in system (a) variable 1 has a direct influence on variable 2, which in turn affects variable 3). Under normal operation conditions the connected variables are related by linear stochastic equations and, during a fault, the slop of one of these equations was increased by 5%, which consequently changes the systems correlation.

Table 9.2 Effect of the sensitivity enhancing transformation on the partial correlation coefficients. The results presented are the average value of 200 samples (standard deviations are all close to 1).

	Network							
	(a) 				(b) 			
	T_X		T_{Ch}		T_X		T_{Ch}	
	NOC	Fault	NOC	Fault	NOC	Fault	NOC	Fault
$r_{1,2.3}$	0.11	0.01	-0.03	24.11	-0.03	1.74	0.08	24.39
$r_{1,3.2}$	-0.20	0.08	0.002	-0.03	0.11	-2.06	-0.03	-0.09
$r_{2,3.1}$	0.17	-0.09	0.20	-0.07	-0.06	0.17	-0.05	0.17

	Network							
	(c) 				(d) 			
	T_X		T_{Ch}		T_X		T_{Ch}	
	NOC	Fault	NOC	Fault	NOC	Fault	NOC	Fault
$r_{1,2.3}$	-0.06	-1.35	-0.04	-0.23	-0.07	2.60	0.06	24.32
$r_{1,3.2}$	0.09	2.55	-0.01	24.79	0.03	0.02	0.02	0.13
$r_{2,3.1}$	0.01	-0.003	0.001	0.02	-0.02	-0.02	-0.02	-0.02

In this study, 200 sample covariance matrices were collected for NOC and faulty conditions, each one of them computed from 3000 observations. Note that the partial correlation coefficients computed from 3000 observations are approximately normally distributed [46]. Therefore, the partial correlation coefficients, normalized according to, $r_n = \left(r - \overline{r}\right)/s_r$, where $\overline{r}$ is the sample mean and s_r is the sample standard variance under NOC, are approximately distributed as $N(0,1)$. These values are depicted in Table 9.2 for both original (T_X) and transformed (T_{Ch}) variables.

The results presented in Table 9.2 clearly show the importance of using a transformation, since when a change occurs, a deviation of more than 24 standard deviations is observed when the transformation T_{Ch} is used, while for raw data, without any transformation (T_X), only a maximum change of 2.60 standard deviations is registered. This general result has a great impact on the monitoring statistics performance.

9.2 Data-driven Reconstruction of the Causal Network

The reconstruction of the causal network consists of two main stages: (i) the identification of the edges associated with the directly associated variables (or nodes) and (ii) the establishment of the causal directionality of each edge (i.e., the directionality of the dependency). In order to address these two problems, several algorithms were developed, which will be presented in this section. Note however that the goal of the proposed algorithms is not to retrieve the exact causal network, but to explain the local correlation structure of the process with enough accuracy in order to potentiate the effective application of the SET's.

As for the first stage, one way to identify the variables that are directly related is based on the information retrieved by partial correlation. This approach, has been explored in several works, such as the ones by Fuente *et al.* (2004) [130] and Pellet and Elisseeff (2007) [133], where partial correlations, usually up to the 2nd order, are employed in order to recover the significant edges between variables. However, these procedures are limited to detect stationary linear relationships. Therefore, in order to allow for the identification of more complex relationships, time-sifted variables and polynomial terms need to be included in the regression models. By doing so, the final algorithm for identifying the

undirected edges of the network, presented in Table 9.3, is obtained. Note that this procedure is the same as the one applied elsewhere [129, 130, 132], differing only in the inclusion of additional time-shifted and polynomial terms in order to capture the existing dynamic and non-linear relationships.

Table 9.3 Pseudo-code for the identification of the network undirected edges.

1. Set an edge for all possible pair of variables;
2. For all possible pairs of variables determinate the 0^{th} order partial correlation between each pair (x,y):
    a. Construct an extended matrix **X** composed of **x** and its time-shifted replicates;
    b. Fit a model of the type **y** = **Xb**;
    c. Determine the association between variables through the analysis of the statistical significance of a = corr(**y**,**Xb**) (If a is lower than a threshold, eliminate the edge for the pair (x,y)).
3. For all remaining edges determine the 1^{th} order partial correlation between the pair (x,y) controlled by z:
    a. Construct an extended matrix **Z,** composed by **z** and its time-shifted replicates;
    b. Fit a model of the type **x** = **Zb** and take the residuals **e** = **x** – **Zb**;
    c. Fit a model of the type **y** = **Zb** and take the residuals **f** = **y** – **Zb**;
    d. Repeat Step 2 for the pair of variables (**e**,**f**).
4. For all remaining edges compute the 2^{th} order partial correlations between the pair (x,y) controlled by z and q:
    a. Construct an extended matrix **Z** composed by **z** and its time-shifted replicates;
    b. Construct an extended matrix **Q** composed by **q** and its time-shifted replicates;
    c. Fit a model of the type **x** = [**Z Q**]**b** and take the residuals **e** = **x** – [**Z Q**]**b**;
    d. Fit a model of the type **y** = [**Z Q**]**b** and take the residuals **f** = **y** – [**Z Q**]**b**;
    e. Repeat Step 2 for the pair of variables (**e**,**f**).

The directions of the undirected edges identified in this way are determined in the second stage of the causal network reconstruction procedure. Two algorithms are proposed for addressing this task. The first algorithm was derived for stationary systems, where it was observed that, when one variable (a child node) has more than one predecessor (parent node), setting the wrong direction causes the appearance of previously non-identified edges involving other parents, as a result of the transfer of information between variables that are not directly associated, during the computation of the regression residuals (an example will be provided further ahead in the text, in Table 9.5). Thus, by testing both possible directions ("tentative parent" → "tentative child") it is possible to extract the underlying causal directionality, by simple employment of network reconstruction

techniques based on partial correlations. To do so, for each "tentative direction" a linear model is fitted between the "tentative parent" (regressor) and the "tentative child" (response). The "tentative child" is then replaced by its regression residuals and a new undirected network is estimated by the algorithm in Table 9.3. As a result of this variable replacement, if the "real child" has more than one parent, some of the information related with these other parents is transferred to the regression residuals when the direction is wrongly attributed (i.e. when the "tentative child" is in fact the "real parent"). Consequently, a larger number of new edges between the wrong "tentative child" and the other variables (namely the other parents of the "tentative parent", which is in fact the "real child") will arise. As a result of this observation, the following decision rule is derived:

Rule 1: in the cases where the child (y) has more than one parent (x), when x is considered as the "tentative child", the replacement of x by its regression residuals (e = x – f(y)) cause the appearance of edges between the residuals (e) and the other parents of y when a new network is estimated. Otherwise, if y is considered as the "tentative child", no new edges are expected to appear. Therefore, the direction that leads to the lowest number of new edges is related with the correct (or at least, predominant) direction.

After application of this decision rule, a second decision rule can be drawn:

Rule 2: in the cases where the child (y) has more than one parent (x), all its parents are expected to be determined by Rule 1. Therefore, if y has parents identified by Rule 1, the remaining edges involving y are related with the child of y.

Based on these decision rules, the algorithm presented in Table 9.4 can be used to establish the edges directions in stationary systems. This algorithm can also be applied to dynamic and non-linear systems, if time-shifted variables or polynomial terms are employed during the regression stages.

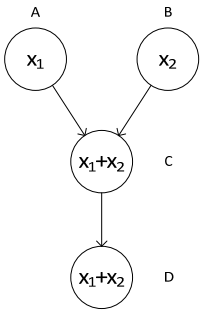
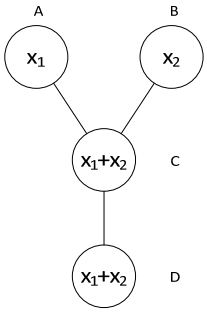
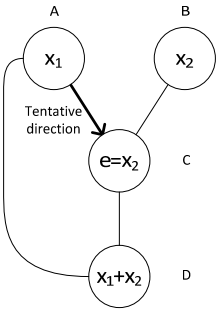
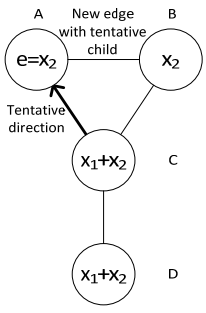
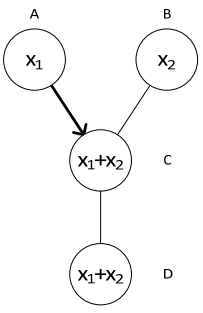
Table 9.4 Pseudo-code for determining the direction of the edges identified by the algorithm of Table 9.3 for stationary systems.

1. Determine the significant edges between the variables, though the algorithm in Table 9.3. Save the resulting network structure as $\mathbf{E}_0$;
2. For all the significant edges, analyze the (x,y) pairs of unidentified directions:
    a. Consider the hypotheses that x causes y ($x \rightarrow y$):
        i. Construct an extended matrix **X** composed by **x** and its time-shifted replicates;
        ii. Remove the effect of x on y by fitting a model of the type $\mathbf{y} = \mathbf{Xb}$ and take the residuals $\mathbf{e} = \mathbf{y} - \mathbf{Xb}$;
        iii. On the original data set, replace **y** by **e** and determine the new network. Save the new network structure as $\mathbf{E}_{new}$.
        iv. Determine the number of new edges (i.e., edges present on $\mathbf{E}_{new}$ that were not identified on $\mathbf{E}_0$). Save this value as $N_{x\rightarrow y}$;
        v. Determine the number of new edges that involve variable y. Save this value as $L_{x\rightarrow y}$.
    b. Repeat Step 2.a by considering the hypotheses that y causes x ($y \rightarrow x$), obtaining $N_{y\rightarrow x}$ and $L_{y\rightarrow x}$.
    c. Set the direction based on Rule 1:
        i. If $L_{x\rightarrow y} < L_{y\rightarrow x}$, then $x \rightarrow y$;
        ii. Else, if $L_{y\rightarrow x} < L_{x\rightarrow y}$, then $y \rightarrow x$.
3. For the remaining (x,y) pairs of unidentified directions, set the direction based on Rule 2;
4. For the remaining (x,y) pairs of unidentified directions, set the direction that leads to the number of new edges:
        i. If $N_{x\rightarrow y} < N_{y\rightarrow x}$, then $x \rightarrow y$;
        ii. Else, if $N_{y\rightarrow x} < N_{x\rightarrow y}$, then $y \rightarrow x$.
5. For the remaining (x,y) pairs of unidentified directions:
    a. Verify if the edge is still relevant by conditioning on their previously identified parents (the algorithm in Table 9.3 only considers partial correlations up to 2nd order and at this point there may be identified more than 2 parents for any of the variables).
        i. If the edge is no longer relevant, eliminate it;
        ii. Otherwise, set the direction as the one that leads to the best regression model.

The reasoning behind the decision rules used in the algorithm presented in Table 9.4 is illustrated for a four variable network, represented in Table 9.5. In this case, one is testing whether A causes C or C causes A. If it is assumed that A causes C, the residual of the regression model eliminates the component of x_1 in C, and consequently, when the new network is reconstructed, there are no new edges between the residual of C and the remaining nodes (note that the new edge between A and D is not related with the "tentative child" C). On the other hand, when it is considered that C causes A, the component of x_2 is added in the residual of A and a new edge between the "tentative child" A and B appears. This new edge is a result of the wrong attribution of the edge

direction and therefore the direction that leads to the lowest number of new edges with the "tentative child" is regarded as the correct one. However, the decision rule 1 requires that the "real child" has at least two parents, so that the information of one parent is transferred to the other when it is considered as a "tentative child". Therefore, this rule cannot be applied to determine the direction between nodes C and D, since D has only one parent and therefore no new edges with the "tentative child" appear in both situations. To select the directionally in this situation, the rule 2 has to be considered instead. This rule is a direct extension of the previous one and assumes that rule 1 is able to identify all the node's parents (which is in fact verified, as long as the node has more than one parent). Therefore, the remaining undirected edges are considered to be related with the node's children and their direction established accordingly.

Table 9.5 Example of the determination of the direction in one of the network edges, by application of the proposed decision rules.

Real causal network	Identified edges	Determine direction		Obtained direction
		Replace tentative child (C) by its residual ($e = x_2$)	Replace tentative child (A) by its residual ($e = x_2$)	
				

The second algorithm proposed for the second stage of network reconstruction is presented in Table 9.6, and makes uses of the variables' cross-correlation (Step 2) in order to identify the time dependency between them, based on the simple principle that the present can only be affected by the past and not the other way around. As an alternative, Granger causality [126] or transfer entropy [125] can also be applied in Step

2. The remaining steps are similar with the ones used in the previous algorithm of Table 9.4 (note that only rule 1 is employed in this case).

The main difference between the algorithms in Tables 9.4 and 9.6 is their application scope. The first algorithm (Table 9.4) is more general, encompassing non-linear and dynamic dependencies. On the other hand, the second algorithm (Table 9.6) includes measures of causality that are specifically related with time dependency and therefore can only be applied to dynamic system. These algorithms for network reconstruction will be applied in the analysis of the case studies presented in Chapter 10.

Table 9.6 Pseudo-code for determining the direction of the edges identified by the algorithm of Table 9.3, valid for dynamic systems.

1. Determine the significant edges between the variables, though the algorithm in Table 9.3. Save the resulting network structure as $\mathbf{E}_0$;
2. For all the significant edges, analyze the cross-correlation between the connected variables (x,y). The variable with significant greatest cross-correlation (y) is set as the child ($x \rightarrow y$);
3. For the remaining (x,y) pairs of unidentified directions:
    a. Consider the hypotheses that x causes y ($x \rightarrow y$):
        i. Construct an extended matrix $\mathbf{X}$ composed by $\mathbf{x}$ and its time-shifted replicates;
        ii. Remove the effect of x on y by fitting a model of the type $\mathbf{y} = \mathbf{Xb}$ and take the residuals $\mathbf{e} = \mathbf{y} - \mathbf{Xb}$;
        iii. On the original data set, replace $\mathbf{y}$ by $\mathbf{e}$ and determine the new network. Save the new network structure as $\mathbf{E}_{new}$.
        iv. Determine the number of new edges (i.e., edges present on $\mathbf{E}_{new}$ that were not identified on $\mathbf{E}_0$). Save this value as $N_{x\rightarrow y}$;
        v. Determine the number of new edges that involve variable y. Save this value as $L_{x\rightarrow y}$.
    b. Repeat Step 3.a by considering the hypotheses that y causes x ($y \rightarrow x$), obtaining $N_{y\rightarrow x}$ and $L_{y\rightarrow x}$.
    c. Set the direction based on Rule 1:
        i. If $L_{x\rightarrow y} < L_{y\rightarrow x}$, then $x \rightarrow y$;
        ii. Else, if $L_{y\rightarrow x} < L_{x\rightarrow y}$, then $y \rightarrow x$;
        iii. Else, if $N_{x\rightarrow y} < N_{y\rightarrow x}$, then $x \rightarrow y$;
        iv. Else, if $N_{y\rightarrow x} < N_{x\rightarrow y}$, then $y \rightarrow x$.
4. For the remaining (x,y) pairs of unidentified directions:
    a. Verify if the edge is still relevant by conditioning on their previously identified parents (the algorithm in Table 9.3 only considers partial correlations up to 2nd order and at this point there may be identified more than 2 parents for any of the variables).
        i. If the edge is no longer relevant, eliminate it;
        ii. Otherwise, set the direction as the one that leads to the best regression model.

In order to illustrate the application of the two-stage procedure for reconstructing the causal network, let us consider a dynamic version of the network proposed by Tamada *et al.* (2003) [134]. This system is composed by 16 nodes (variables), related as shown in Figure 9.3, with the dynamic equations given by Equations (9.6), where ε_i is a white noise sequence with a signal-to-noise ratio of 10 dB (see Equation (5.8)). From the edge identification algorithm (Table 9.3), the correct undirected network is retrieved, as can be verified in Table 9.7. This structure is then the input for the second stage of the reconstruction algorithm (Table 9.6), where most of the directions are correctly attributed in Step 2 (see Table 9.7). The only exception is the relationship between g_1 and g_4 which is reversed. The two remaining edges are addressed in Step 3 of the proposed algorithm. These results show that the proposed algorithms are able to produce a reasonable estimate of the underlying causal network, using only data collected from the process. However, in this case study, one edge is wrongly directed by the criteria based on the cross-correlation (the use of the Granger causality in Step 2 led to the same result). Nevertheless, for the purposes of the sensitivity enhancing transformation, if an additional Cholesky decomposition of the residuals is performed, the effects of the wrong direction become negligible.

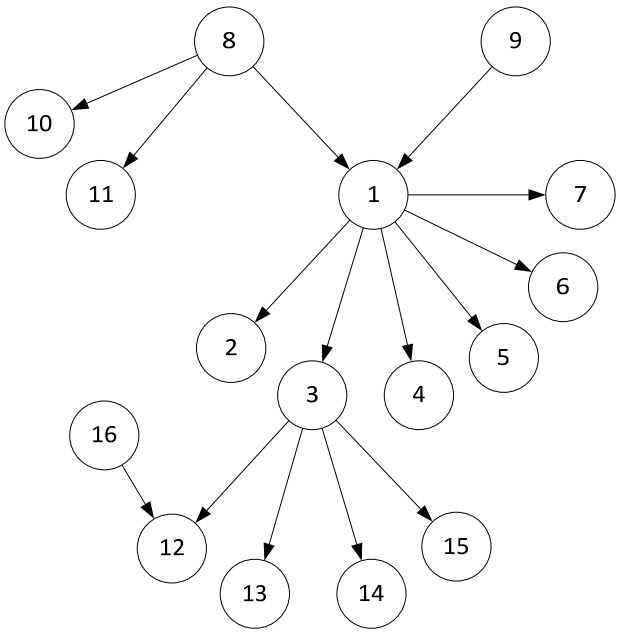
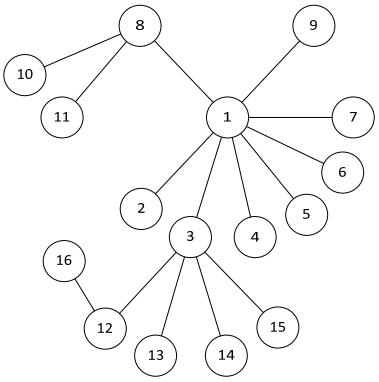
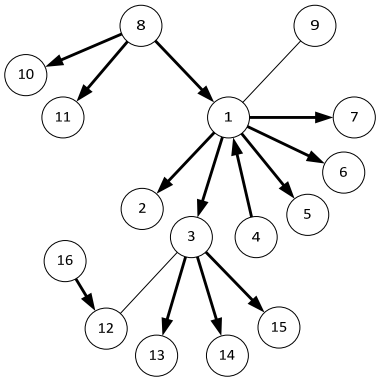
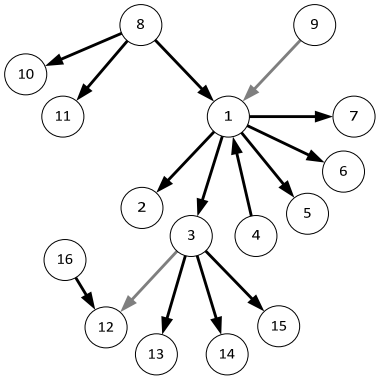


Figure 9.3 Graphical representation of the original causal structure for the causal network under analysis.

$$
\begin{aligned}
&g_{8,t} = \varepsilon_{8,t},\ g_{9,t} = \varepsilon_{9,t},\ g_{16,t} = \varepsilon_{16,t} \\
&g_{10,t} = 1 + 0.40\left(g_{8,t} + 0.60g_{8,t-1} - 0.30g_{8,t-2}\right) + \varepsilon_{10,t} \\
&g_{11,t} = 0.56 + 0.15\left(g_{8,t} + 0.40g_{8,t-1} + 0.60g_{8,t-2}\right) + \varepsilon_{11,t} \\
&g_{1,t} = 1.2\left(g_{8,t} + 0.60g_{8,t-1} + 0.30g_{8,t-2}\right) + 0.80g_{9,t} + \varepsilon_{1,t} \\
&g_{2,t} = 0.60\left(g_{1,t} + 0.50g_{1,t-1} + 0.20g_{1,t-2}\right) + \varepsilon_{2,t} \\
&g_{3,t} = 0.05 + 0.22\left(g_{1,t} - 0.40g_{1,t-1} - 0.20g_{1,t-2}\right) + \varepsilon_{3,t} \\
&g_{4,t} = 1 + 0.4\left(g_{1,t} - 0.20g_{1,t-1} - 0.10g_{1,t-2}\right) + \varepsilon_{4,t} \\
&g_{5,t} = 0.062 + 0.16\left(g_{1,t} + 0.40g_{1,t-1} + 0.60g_{1,t-2}\right) + \varepsilon_{5,t} \\
&g_{6,t} = 0.60\left(g_{1,t} + 0.80g_{1,t-1} + 0.10g_{1,t-2}\right) + \varepsilon_{6,t} \\
&g_{7,t} = 0.70\left(g_{1,t} + 0.40g_{1,t-1} + 0.20g_{1,t-2}\right) + \varepsilon_{7,t} \\
&g_{12,t} = 0.80\left(g_{16,t} + 0.60g_{16,t-1} + 0.30g_{16,t-2}\right) + 0.51g_{3,t} + \varepsilon_{12,t} \\
&g_{13,t} = 1.30\left(g_{3,t} + 0.50g_{3,t-1} + 0.50g_{3,t-2}\right) + \varepsilon_{13,t} \\
&g_{14,t} = 1 + 0.40\left(g_{3,t} + 0.40g_{3,t-1} + 0.60g_{3,t-2}\right) + \varepsilon_{14,t} \\
&g_{15,t} = 0.028 + 1.30\left(g_{3,t} + 0.60g_{3,t-1} - 0.30g_{3,t-2}\right) + \varepsilon_{15,t}
\end{aligned}
\tag{9.6}
$$

Table 9.7 The successive results obtained in the several phases of the proposed network reconstruction procedure.

Stage 1	Stage 2 – Step 2	Stage 2 – Step 3
		

9.3 Final Remarks on the Use of SET

The proposed sensitivity enhancing transformations essentially rotate the original variables into a new set of uncorrelated variables, where changes in the correlation and partial correlation coefficients are easier to detect due to the increased sensitivity around the zero correlation state.

The concept of using uncorrelated variables to monitor the marginal covariance was already proposed in the literature, mainly for simplification purposes and not to increase the methods performances (e.g. [80, 82]). Moreover, they are often based on the inverse of the covariance matrix, which may be ill-conditioned for multivariate systems, or on triangularization method (like the Cholesky decomposition [60]), which may not decorrelate the data in a meaningful way. Still, the latter approach remains quite useful, depending on the variables ordering, and capable to break the autocorrelation by including lagged variables in the data matrix.

Although the Cholesky decomposition produces uncorrelated variables, the resulting statistics performance is dependent on the variables ordering. To circumvent this problem, it is suggested a transformation that makes use of the process network structure in order to construct linear/polynomial regression models involving only the variables that are actually causally related (i.e. only the parents nodes/variables are used as regressors). These new transformations amount to fitting separate regression models for each variable which, for multivariate processes, can be seen as a disadvantage. However, it is worth noticing that other similar approaches have already been proposed, such as Hawking's regressed-adjusted variables [127] or Ottestad's regression components [135]. Other examples can also be found in [1, 20]. Principal component analysis, which is extensively used to explain the process variability, was also tested to decorrelate the variables. However, as it also ends up with a linear combination of variables that may not be directly related, it was found to be unsuitable for explaining the underlying process structure. In fact, none of the methods mentioned above are driven by the inner process structures.

In this study, ordinary lest squares regression (OLS) is applied to develop the modeling tasks, but principal component regression (PCR) or partial least squares regression (PLSR) can also be adopted, in case collinearity becomes a problem.

The new transformations may also require the estimation of the causal network, which can be a quite complex task. Still, when the regression step is complemented by an additional Cholseky decomposition over the regression residuals, the final model has some robustness to miss specifications of the true network, namely the identification of wrong edges and the incorrect attribution of causal directions. Regarding this point, it is important to underline that the goal of the proposed algorithms is to identify an approximate causal network that lead to a more effective use of the SET's, and not to retrieve the exact network.

10 Off-line Monitoring of the Process Correlation Structure

Statistical process control is a pervasive industrial task with the aim to monitor the evolution of the overall process operation state. It is implemented mostly by means of control charts designed to assess whether the process is only subjected to normal causes of variation, inherent to its operation, or if a special cause of variation has occurred, that urges to be detected, diagnosed and fixed or accommodated. Special causes can result from equipment failure, abnormal changes in raw materials properties or any internal or external events that affect process evolution and increase its variability, putting in risk not only product quality and process efficiency, but also the safety of people and the environment. Therefore, the rapid detection of faults and associated diagnosis of their origins is of outmost importance, and have attracted the interest of many researches and practitioners since the proposal of the first tool, the univariate control chart, by Shewhart (1931) [40]. In this context proposals were rapidly put forward for extending the methodology to the multivariate case through the Hotelling's T^2 statistic [43], CUSUM charts [44] and multivariate EWMA charts [45], and then to the megavariate case, by the PCA and PLS-based multivariate statistical process control scheme (MSPC-PCA, MSPC-PLS, respectively) [15, 16] and all the associated methodologies derived from the same latent variable formalism. Even without entering in a more exhaustive enumeration of the methodologies proposed through the decades, one can easily acknowledge that the proposed control charts were essentially designed to monitor the process mean [66-69], having some limited ability to address process variability, in particular concerning changes in the process correlation structure. In fact, only a comparatively low number of methodologies were proposed for explicitly addressing the problem of detecting changes in correlation.

However, monitoring process dispersion and correlation is also a relevant issue in process monitoring, since a process failure may not manifest itself so notoriously as a deviation from the nominal mean values, especially due to the action of control systems fighting to maintain key process variables close to their target values. In multivariate processes the monitoring of process dispersion is usually achieved by monitoring statistics based on the

generalized variance [70-72] or through the implementation of supervision schemes based on successive likelihood ratio tests [69, 75-77]. Yet, these procedures tend to consider only the marginal covariance and ignore the variables inner associations. Even the SPC methodologies based on state-space models (such as CVA and subspace system identification), are based on the marginal cross-covariance between the variables [65, 95, 110]. On the other hand, partial correlations have the potential to describe such local associations, even though in a non-causal way, being already applied to retrieve the structural information underlying collected data [130, 133, 136]. In this chapter it will be showed that monitoring statistics based on partial correlations have indeed the potential to enhance the detection of structural changes. Furthermore, it will be demonstrated that structural changes are easier to detect on variables that are initially uncorrelated. To obtain such uncorrelated variables, procedures based on the Cholesky decomposition were initially explored in the present work, following a similar approach as Hawkins and Maboudou-Tchao (2008) [82]. However, this transformation presents limitations, especially regarding its inability to properly model the data structure and its dependency of a suitable variable ordering in the data matrix, which is usually arbitrary. To mitigate these issues, a new class of variable transformation based on the network reflecting the inner associations between variables, which is obtained through the analysis of partial correlations, was proposed in Chapter 9. Additionally, a monitoring statistic specifically designed to detect changes in the variance will be proposed, since the sole monitoring of partial correlations is not enough to capture all the process dispersion features. Departures from the system's normal operation conditions (NOC) model are also detected by this new monitoring statistic. Therefore, used in combination with the partial correlation monitoring statistics, it becomes possible to detect a wide spectrum of faults specifically related with changes in the process correlation structure.

Some related approaches have already been proposed mostly for fault diagnosis, based on the analysis of fault propagation pathways over an existent causal network [125, 126]. However, these applications are dedicated to the diagnostic stage, assuming the existence of a previous detection phase, and also assuming the availability of *a priori* knowledge regarding the variables' connectivity. The same applies to model-based approaches to fault detection, where a process dynamic model is always assumed to be available [137-139].

The effects of the proposed sensitivity enhancing transformations and the performance of the monitoring statistics based on partial correlations are here tested and compared against the monitoring statistics based on marginal covariance in a set of multivariate systems, for the case of off-line monitoring.

10.1 Statistical Process Control Based on Partial Correlations

The current monitoring statistics typically resort to measures of multivariate dispersion based on marginal information. It is well-known that such information leads to correlations between variables that are not directly associated, as long as they share some common inducing variation sources or are part of the same causality chain. Therefore, these methodologies are in principle unable to detect subtle and localized changes on the process structure. One way to improve the detection capability to faults specifically related with the local process structure relies on the use of partial correlations. Partial correlations, have the ability to identify variables that are directly associated or connected, a feature that was already extensively explored to reconstruct interaction networks [130, 131, 133, 136] as well as in classification problems [128, 129, 132]. In this context, it is expected that monitoring partial correlations can bring the benefits of increasing sensitivity to more subtle changes in the process structure (due to the finer description of the NOC structure) and to speed up the diagnosis stages (due to the localized nature of the information they provide), which potentially leads to a significant reduction in the total time to detect, diagnose and accommodate or fix, a given process upset. With these aims in mind, in the following sections, statistics for monitoring the fine process structure based on the information provided by partial correlations will be presented, along with their complementary statistics dedicated to monitor changes in variance.

10.1.1 $R0MAX$ and $R1MAX$

One way to detect changes in the process structure is by using sequential hypothesis tests in order to verify if the partial correlations for new samples remain close to their respective NOC values. This is a problem showing some similarities with network

reconstruction. In network reconstruction, the partial correlation coefficients are tested for the hypothesis H_0: $\rho = 0$ *vs* H_1: $\rho \neq 0$ in order to identify significant relationships between variables. In the present situation, the goal is rather to detect changes on their nominal values since they are associated with changes on the variables dependencies. To formalize the hypotheses tests to be applied in this situation, one should first discuss the probability distribution of the correlation coefficients and their extension to partial correlations.

When the number of observations (n) is large, the distribution of the correlation coefficients (0^{th} order partial correlation), transformed according to Equation (10.1), tend to be normally distributed with zero mean and unit variance [46].

$$w_1 = \frac{\sqrt{n-1}\left(r-\rho\right)}{1-\rho^2} \tag{10.1}$$

This tendency to normality can be strengthen by the use of the Fisher's z transformation, defined as,

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \tag{10.2}$$

resulting in,

$$w_2 = \frac{\sqrt{n-1}}{2}\left[\ln\left(\frac{1+r}{1-r}\right) - \ln\left(\frac{1+\rho}{1-\rho}\right)\right] \tag{10.3}$$

In both cases, the underlying distribution for hypothesis testing of $\rho = \rho_0$ against the alternative of $\rho \neq \rho_0$, corresponds to the standardized normal distribution.

Since the distribution of the q^{th} order partial correlation coefficients based on n observations is the same as the correlation coefficients based on $(n-q)$ observations, its hypothesis test is exactly the same, except that n is replaced by $(n-q)$ [46].

If the transformation w_1 or w_2 is applied to all the partial correlations, then they will present the same distribution and the same probability of exceeding a certain limit. As only one partial correlation coefficient needs to exceed the control limits to consider that a change in the process structure has occurred, it is proposed to monitor the maximum

norm (i.e., the maximum, in absolute value) of the normalized partial correlations, defined as,

$$R0MAX = \left\| w(\mathbf{r}_0) \right\|_\infty = \max\left\{ \left| w(\mathbf{r}_0) \right| \right\} \tag{10.4}$$

for the correlation coefficients (0^{th} order partial correlations, i.e., the marginal correlations) and as,

$$R1MAX = \left\| w(\mathbf{r}_1) \right\|_\infty = \max\left\{ \left| w(\mathbf{r}_1) \right| \right\} \tag{10.5}$$

for the 1^{st} order partial correlations. In these Equations, $\mathbf{r}_0$ is the $(m(m-1)/2)\times 1$ column vector containing all distinct correlation coefficients (0^{th} order partial correlations) and $\mathbf{r}_1$ is the $(m(m-1)(m-2)/2)\times 1$ column vector of 1^{st} order partial correlation coefficients. Moreover, $w(\cdot)$ stands for either one of the transformations presented in Equations (10.1) and (10.3), or any other that guarantees that the transformed partial correlations follow approximately a standard normal distribution. For instance, in cases where it cannot be assume that partial correlation follow a normal distribution, an estimation of the density function can be applied to normalize them, for instance based on Kernel density estimation, which can then be used to normalize the partial correlations by matching the empirical distribution quantiles with those of the standard normal distribution.

10.1.2 *VnMAX*

As stated earlier, the sole monitoring of the partial correlation coefficients is not enough to detect all faults, as some events may pass undetected, such as those affecting process variance. This is particularly relevant when transformed variables are used, since a model mismatch can be translated into changes on the transformed variables' variance. In this context, the *VMAX* statistic proposed by Costa and Machado (2009) [84] (see Section 4.1.5) is the one offering more potential for dealing with this problem but, as it is defined, it is only able to detect a significant increase in the variance. Of course, in some situations this may be all that is necessary, but in the present context it is of interest to detect *any* departure from normal operation conditions. Therefore, a modification of this statistic,

based on the inference properties of the variance in a normal distribution will be proposed here.

The test statistics for the hypothesis test that a sample variance obtained from n observations of a random normally distributed variable is equal to σ_0^2, is given by [48],

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2} \tag{10.6}$$

where χ_0^2 follows a chi-squared distribution with $(n-1)$ degrees of freedom. The one sided-test is already performed with the *VMAX* statistic, for detecting an increase in the variance (Section 4.1.5). In order to set the conditions for maintaining the use of a single control limit, the probability of exceeding a symmetric UCL and the LCL must be the same, which can easily be achieved by taken the absolute value of a symmetric random variable. In this particular case, one can apply the Wilson-Hilferty transformation to the chi-squared distribution [140]. The use of this particular transformation also facilitates future constructions of a combined statistic to monitor both variance and correlation coefficients.

The Wilson-Hilferty transformation states that a chi-squared distribution with v degrees of freedom is well approximated by:

$$\chi_{\alpha,\nu}^2 \approx \nu\left[z_\alpha \left(\frac{2}{9\nu}\right)^{0.5} + \left(1-\frac{2}{9\nu}\right)\right]^3 \tag{10.7}$$

where z_α is the $100\times\alpha\%$ upper quantile from the standard normal distribution.

Given this relation, it is proposed to use the following transformation:

$$w_s = \frac{\left(\frac{s^2}{\sigma_0^2}\right)^{1/3} - \left(1-\frac{2}{9(n-1)}\right)}{\sqrt{\frac{2}{9(n-1)}}} \tag{10.8}$$

which is approximately distributed as $N(0,1)$. Consequently, the modified (normalized) *VMAX* statistic becomes:

$$VnMAX = \left\| w_s(\mathbf{v}) \right\|_\infty = \max\left\{ \left| w_s(\mathbf{v}) \right| \right\} \tag{10.9}$$

where **v** is a ($m \times 1$) column vector containing the variables' variance.

This modified version is able to detect both increases and decreases in the process variance. It is also capable to detect changes in the process structure, when any of the sensitivity enhancing transformations described in Chapter 9 is applied as a result of model mismatches.

10.1.3 Control Charts for Partial Correlations Based on MSPC-PCA

MSPC-PCA is a non-causal approach with the capability to monitor, in an implicit way, the variables correlation structure, namely through the residual statistic [15-17]. This approach uses two monitoring statistics. One of them follows the variability in the PCA subspace estimated with NOC data, corresponding to a Hotelling's T^2 statistic applied to the retained principal components (PCs). The other monitoring statistic monitors the complementary variability, around the PCA subspace, through a lack of fit or residual statistic usually called Q or *SPE* (squared prediction error). In this way, both orthogonal data variability complements (in the PCA subspace and around it) are efficiently followed. More specifically, the MSPC-PCA monitoring statistics are defined as follows,

$$T^2_{PCA} = \sum_{i=1}^{p} \frac{t_i^2}{\lambda_i} = \mathbf{x}^{\mathrm{T}} \mathbf{P} \boldsymbol{\Lambda}_p^{-1} \mathbf{P}^{\mathrm{T}} \mathbf{x} \tag{10.10}$$

$$Q = \mathbf{e}^{\mathrm{T}}\mathbf{e} = \left(\mathbf{x} - \hat{\mathbf{x}}\right)^{\mathrm{T}} \left(\mathbf{x} - \hat{\mathbf{x}}\right) = \mathbf{x}^{\mathrm{T}} \left(\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}}\right) \mathbf{x} \tag{10.11}$$

where **P** is a matrix containing the first p eigenvectors, $\boldsymbol{\Lambda}_p = \mathrm{diag}\left(\lambda_1, \ldots, \lambda_p\right)$ is a diagonal matrix with the respective first p eigenvalues in the main diagonal, $\hat{\mathbf{x}}$ is the projection of **x** onto the PCA model, p is the number of retained PCs (the pseudorank) and **I** is an identity matrix of proper size. For more details on this monitoring procedure please refer to Section 3.3. As this is an approach extensively used to monitor the processes state, a preliminary analysis of its performance when applied to the problem of monitoring the process multivariate dispersion is presented in Section 10.2.1, where one can verify that the direct implementation of this well-known scheme to the observed

variables (**x**) is unsuitable for detecting small structural changes. For this reason, MSPC-PCA was excluded from the main study.

The results presented in Section 10.2.1 show that MSPC-PCA based on process data (**x**) is not efficient in the detection of local structural changes. However, using the $\mathbf{r}_0$ and $\mathbf{r}_1$ vectors of correlation coefficients instead of **x** in the MSPC-PCA procedure, can lead to significant improvements in the detection sensitivity. This result is very useful for improving the process monitoring performance through a MSPC-PCA procedure, using the associated T^2_{PCA} and Q statistics. For such, the respective $\mathbf{r}_q$ (q = 0,1) covariance matrix must be first estimated, in order to construct the NOC, PCA model. The covariance matrix of $\mathbf{r}_q$ can be obtained from k subgroups with n observations each, leading to k observations of $\mathbf{r}_q$ (see Table 10.1). The number of subsets (k) required to construct the PCA model is one of the disadvantages of this method, since it may translate into an overall large number of observations, especially when the number of process variables is substantial (the estimation of a full rank covariance matrices require $k > m(m-1)/2$ for $\mathbf{r}_0$ and $k > m(m-1)(m-2)/2$ for $\mathbf{r}_1$). In the current study it is always assumed that a sufficient amount of data is available for estimating the PCA model.

Table 10.1 Algorithm to determine the PCA model for the q^{th} order partial correlation.

1. Obtain reference data:
   a. For i = 1 to k,
      i. Collect n observations (**X**);
      ii. Determine the sample covariance matrix:
      $$\mathbf{S}_i = \frac{1}{n-1}\sum_{j=1}^{n}\left(\mathbf{x}_j - \bar{\mathbf{x}}\right)\left(\mathbf{x}_j - \bar{\mathbf{x}}\right)^{\mathrm{T}}$$
      iii. Compute $\mathbf{r}_{q,i}$ from $\mathbf{S}_i$, using Equations (8.1) to (8.3);
   b. Determine the sample mean ($\bar{\mathbf{r}}_q$) and sample covariance ($\mathbf{S}_{\mathbf{r}_q}$) of $\mathbf{r}_q$:
   $$\bar{\mathbf{r}}_q = \frac{1}{k}\sum_{i=1}^{k}\mathbf{r}_{q,i}$$
   $$\mathbf{S}_{\mathbf{r}_q} = \frac{1}{k-1}\sum_{i=1}^{k}\left(\mathbf{r}_{q,i} - \bar{\mathbf{r}}_q\right)\left(\mathbf{r}_{q,i} - \bar{\mathbf{r}}_q\right)^{\mathrm{T}}$$
2. Construct PCA model for the partial correlations ($\mathbf{r}_{q,i}$):
   a. Perform the spectral decomposition: $\mathbf{S}_{\mathbf{r}_q} = \mathbf{\Gamma\Lambda\Gamma}^{\mathrm{T}}$;
   b. Determine the number of principal components to retain (p);
   c. Define the loading matrix, **P**, as the first p columns of $\mathbf{\Gamma}$.

In practice, the pair of T^2_{PCA} and Q statistics can be merged into a single combined index [141] or combined through a logical gate "or". In the current study, the later approach was used to construct a combined statistic, C, that signals an alarm when either T^2_{PCA} or Q, or both, fall beyond their control limits. In the comparison study, the control limits of T^2_{PCA} and Q were adjusted to the same false alarm rate and also in such a way that C presents the desired global false alarm rate.

10.2 An Extensive Comparative Assessment of Methodologies for Off-line Monitoring of the Process Correlation Structure

By applying the monitoring statistics proposed in Section 10.1 to both original and the transformed variables obtained by the sensitivity enhancing transformations introduced in Chapter 9, the complete set of new statistics under analysis in this section is obtained, which is summarized in Table 10.2. The general workflow of the proposed procedure for defining these monitoring statistics is also represented in Figure 10.1. The main stages include data Pre-Processing, where original variables are transformed according to one of the sensitivity enhancing transformations described in Chapter 9, and Process Monitoring, performed by either MSPC-PCA or *RMAX* on the vectors of marginal or partial correlation coefficients. As an example, $R1MAX_{ChExt}$ results from applying $R1MAX$ to data transformed according to the transformation, T_{ChExt}. The performance of all these monitoring statistics will be assessed and compared in the next sections, together with the monitoring statistics already proposed in the literature, namely the ones described in Section 4.1.

The analysis of the process monitoring statistics and sensitivity enhancing transformations is here divided in three parts. In the first part (Section 10.2.1) a preliminary analysis of the MSPC-PCA performance is conducted in order to justify and explain its absence on the remaining of the study. In the second part (Section 10.2.2), focus is given to the performance assessment of the monitoring statistics on the same causal network model, considering different degrees of complexity in the underlying process dynamics: (i) linear stationary (without dynamics); (ii) linear dynamic system; (iii) non-linear stationary system. As the purpose of this study is to assess the capability

of methods to detect changes in variables correlation structure, the use of a network system of reasonable size provides a suitable testing scenario, offering the flexibility to perform and analyze various types of local and global perturbations of different kinds and magnitudes. Finally, in the third part (Section 10.2.3), the study is orientated towards the analysis of the effects of the different types of sensitivity enhancing transformations when applied to the more complex cases of dynamic non-linear systems. In this case, the purpose is to move the test scenario to typical industrial systems, once the methods were characterized in detail under well controlled and easily interpretable conditions.

Table 10.2 Definition of the monitoring statistics proposed in this work.

SET	Monitoring Statistic				
	VnMAX	RMAX		MSPC-PCA	
	$\mathbf{v}$	$\mathbf{r}_0$	$\mathbf{r}_1$	$\mathbf{r}_0$	$\mathbf{r}_1$
$\mathbf{T_X}$	$VnMAX_X$	$R0MAX_X$	$R1MAX_X$	$C_{r_0,X}\begin{cases}T^2_{r_0,X}\\Q_{r_0,X}\end{cases}$	$C_{r_1,X}\begin{cases}T^2_{r_1,X}\\Q_{r_1,X}\end{cases}$
$\mathbf{T_{Ch}}$	$VnMAX_{Ch}$	$R0MAX_{Ch}$	$R1MAX_{Ch}$	$C_{r_0,Ch}\begin{cases}T^2_{r_0,Ch}\\Q_{r_0,Ch}\end{cases}$	$C_{r_1,Ch}\begin{cases}T^2_{r_1,Ch}\\Q_{r_1,Ch}\end{cases}$
$\mathbf{T_{ChExt}}$	$VnMAX_{ChExt}$	$R0MAX_{ChExt}$	$R1MAX_{ChExt}$	$C_{r_0,ChExt}\begin{cases}T^2_{r_0,ChExt}\\Q_{r_0,ChExt}\end{cases}$	$C_{r_1,ChExt}\begin{cases}T^2_{r_1,ChExt}\\Q_{r_1,ChExt}\end{cases}$
$\mathbf{T_{Net}}$	$VnMAX_{Net}$	$R0MAX_{Net}$	$R1MAX_{Net}$	$C_{r_0,Net}\begin{cases}T^2_{r_0,Net}\\Q_{r_0,Net}\end{cases}$	$C_{r_1,Net}\begin{cases}T^2_{r_1,Net}\\Q_{r_1,Net}\end{cases}$
$\mathbf{T_{NetCh}}$	$VnMAX_{NetCh}$	$R0MAX_{NetCh}$	$R1MAX_{NetCh}$	$C_{r_0,NetCh}\begin{cases}T^2_{r_0,NetCh}\\Q_{r_0,NetCh}\end{cases}$	$C_{r_1,NetCh}\begin{cases}T^2_{r_1,NetCh}\\Q_{r_1,NetCh}\end{cases}$

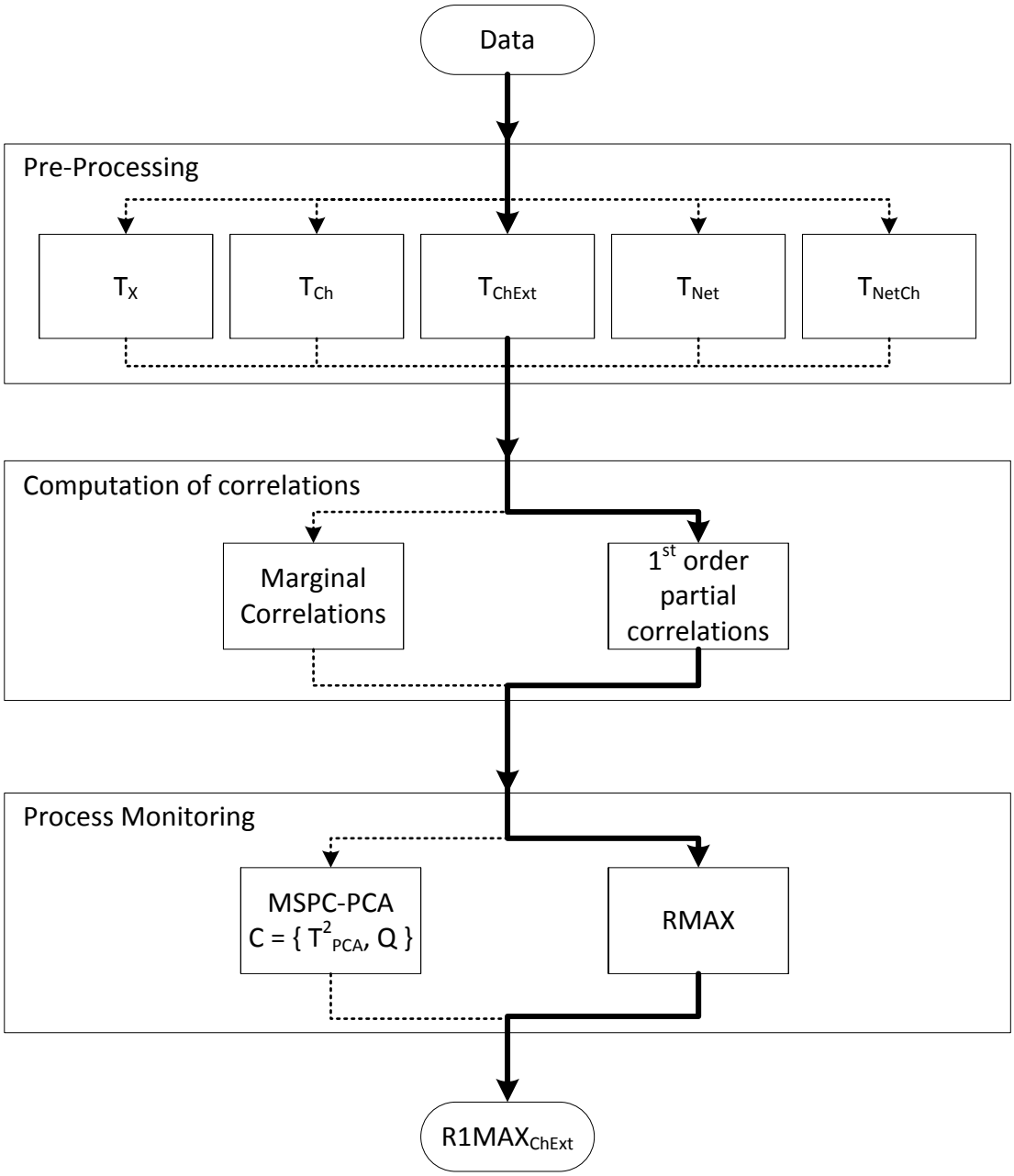


Figure 10.1 Schematic representation of the main blocks that compose the proposed monitoring procedures. The modules involved in the construction of the $R1MAX_{ChExt}$ statistic are highlighted as an example.

Throughout this study, the performance of the monitoring statistics is summarized by a performance index (*N*) based on the area under the fault detection rate curve (i.e., the integral of the curve detection rate *versus* fault's magnitude). This index condenses the results obtained by a given method for different fault magnitudes, which enables the comparison of a large number of monitoring statistics. Under normal operation conditions, the fault detection rate is close to the pre-established false detection rate. As the magnitude of the fault increases, the detection rates tend to 1. Therefore, a monitoring

statistic with a greater area under the detection rate curve tends to 1 more rapidly and will present higher detections rates. This index was computed for each fault and normalized so that its values fall in the range $[0,1]$, where 1 represents the best performance observed (corresponding to the greatest area under the fault detection rate curve). More details on the computation of this performance index will be provided in Section 10.2.1.

The detections rates of each monitoring statistics were also compared using a permutation test in order to verify if the differences obtained were statistically significant. The permutation test assesses the null hypothesis that *n* observations of two variables came from the same distribution. Under such assumption, the variables involved become interchangeable and the difference (**d**) between the two would not depend on the variables order if the null hypothesis is correct. In this case, one way to test this hypotheses is by considering the test statistic $T = \sum_{i=1}^{n} d_i$, whose distribution is obtained by performing *b* permutations of the sign (+) or (–) to each difference [104]. In this study, the detection rates of each pair of monitoring statistics were compared, for all the combinations of faults and magnitudes considered, using 10000 permutations, after which the corresponding *p*-value was computed.

10.2.1 Preliminary Assessment of MSPC-PCA for the Monitoring of the Process Correlation Structure

MSPC-PCA is a well-established and extensively used procedure to monitor the process operation status, not only due to its ability to explain most of the process variability with a reduced number of variables but also for its performance on the detection of deviations from the mean levels and even structural changes. However, this procedure presents some limitations in detecting changes in the local correlation structure when compared to monitoring statistics specifically designed for that purpose. This situation will be exemplified in this section, where the performance of MSPC-PCA is compared against the current *W* statistic for the stationary linear system described in Section 10.2.2.1.

The faults simulated in this system are localized and are exclusively related with changes in the correlation between variables. The sample covariance matrices to be monitored by the *W* statistics were computed from 3000 observations. Likewise, the average of these

observations was also used to determine the T^2_{PCA} and Q statistic, which were here combined through a logical gate "or" (MSPC-PCA). The same procedure was conducted with transformation T_{Ch}, resulting in the MSPC-PCA$_{Ch}$ monitoring scheme. The control limits were set to a false detection rate of 1% for all the monitoring procedures and the detection rates were determined for the same set of faults used in Section 10.2.2.1, with magnitudes in the range of ±50%. The simulations were repeated 100 times in order to assess the consistency of the results.

For faults A, B and C (see Table 10.3), the detection rates of both MSPC-PCA procedures are always lower than 0.10 under the range of simulated faults, while the W statistic presents detection rates of 1 for deviations even smaller than ±10%. For fault D, the obtained results are represented in Figure 10.2. Analyzing the detection curves (Figure 10.2 (a)), it is observed that both MSPC-PCA procedures are able to detect the simulated faults, but a more rapid detection is obtained when the proposed transformation T_{Ch} is applied. Still, the W statistics performs better, correctly detecting all the samples as faulty even for small magnitude perturbations. Furthermore, the smallest fault's magnitude needed to obtain a full detection rate with MSPC-PCA (±40%) is considerable greater than the one used on the main study, which is ±20%. Therefore, most of the other studied monitoring statistics dedicated to detect correlation changes will be able to attain a detection rate of 1 long before MSPC-PCA attain such level.

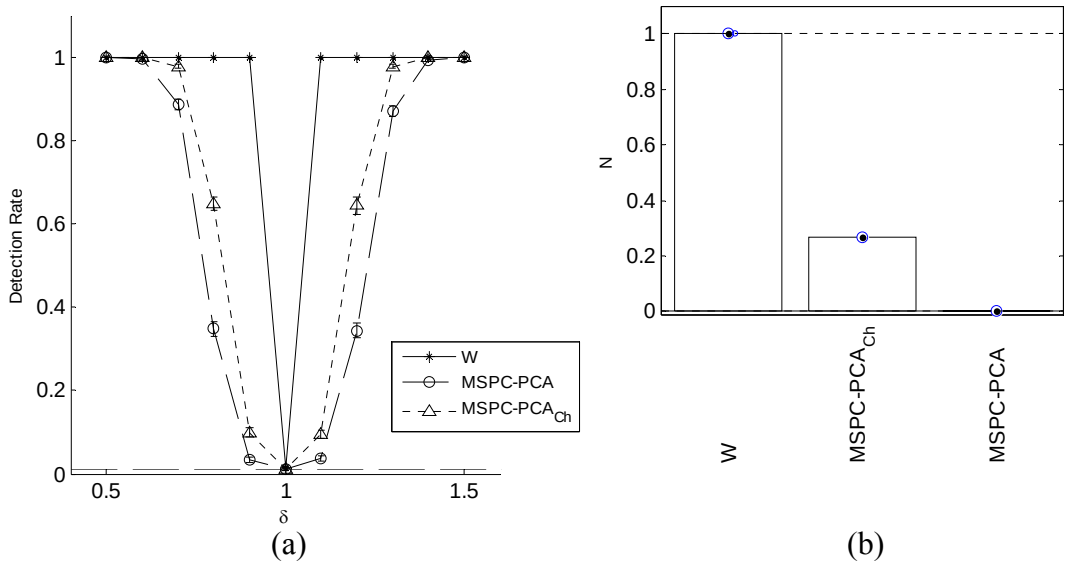


Figure 10.2 Performance comparison of MSPC-PCA and the W statistic in fault D. (a) Fault detection curve, where δ is a multiplicative factor that changes the model parameters (under NOC, $\delta = 1$); (b) box-plots of the performance index N based on the area under the fault detection curve. Bar heights correspond to the associated mean values.

These results are corroborated by the analysis of the performance index. This index aims to facilitate the analysis of results by condensing the outcomes for the different magnitudes of faults into a single value of the index, thus providing a suitable way for comparing a large number of monitoring statistics. This index is computed from the area under the detection rate curve (i.e., the integral of the curve detection rate *versus* fault's magnitude) of each monitoring statistic as described earlier. For this case, the areas under the curve, for one of the replications, were 0.9017 for W, 0.5492 for MSPC-PCA and 0.6448 for MSPC-PCA$_{Ch}$. By normalizing these values, the performance index, in this replication, becomes 1 for W, 0 for MSPC-PCA and 0.2712 for MSPC-PCA$_{Ch}$. The distribution of the obtained indices is depicted of Figure 10.2 (b) for 100 replicates. This representation clearly shows that the W statistics is consistently ranked as the best monitoring statistics and tends toward a full detection rate of 1 more rapidly (note, that the fault detection curve increases monotonically with the increase of the fault's magnitude). It also underlines the extent to which the monitoring procedures differ. For instance, the performance of MSPC-PCA$_{Ch}$ (with transformed variables) is closer to MSPC-PCA (with original variables) than to W. This relative positioning is easily observable by a simple inspection of the detection rate curve in Figure 10.2 (a). However, when the number of monitoring statistics becomes large, it is necessary to summarize the results in a meaningful way, since the plot of the detection rate curve for all monitoring statistics lead to very dense graphs and confusing representations.

From these results it can be concluded that even though the current MSPC-PCA is able to detect changes in the correlation structure of process data, the magnitude after which it becomes significant is relatively larger when compared with most of the current monitoring statistics (in most simulations, MSPC-PCA presented a detection rate lower than 0.10, regardless of the transformation used) and therefore presents a lower performance under the range of fault's magnitude studied. The only situation where MSPC-PCA leads to better results happened in the case studies referred in Section 10.2.3. Yet, the faults simulated in such systems were not completely confined to structural changes, and often led to deviations in the mean value of the measured variables, which explains the better performance obtained by MSPC-PCA. Therefore, in order to reduce the amount of monitoring statistics to be compared, MSPC-PCA will not be included in the comparison studies of the following sections. Finally, it is worth noticing that when

MSPC-PCA is applied to single observations, which is the most frequent situation, the observed detection rates become even lower.

10.2.2 Performance Assessment of the Proposed Monitoring Statistics

To assess the performance of the various methodologies under study, they were applied to a modified version of the artificial network originally presented by Tamada *et al.* (2003) [134]. This network is composed by 16 nodes (or variables) causally related according to the representation provided in Figure 10.3. In each scenario considered in the following examples, 1000 sample covariance matrices were computed based on 3000 observations each, taken at regular intervals of time. The sample covariance matrices were then used to determine the monitoring statistics and to compute the corresponding fault detection rates (true detection and false alarm rates). The same procedure was repeated 10 times in order to estimate the confidence levels of the performance indicators (detection rates). The control limits for all the monitoring statistics were preliminarily adjusted, by trial and error, so that all monitoring statistics present the same false detection rate of 1% under normal operation conditions (NOC). This approach to determine the control limits was taken since the direct use of their theoretical expressions often fails to produce comparable control limits. This situation happens because the underlying assumptions behind the monitoring statistics, such as *i.i.d.* conditions or normal distribution of the data, are not met in all data sets. Therefore, by taking such an approach, it is guaranteed that a fair and sound comparative assessment of the detection performances of the various methods under analysis is conducted.

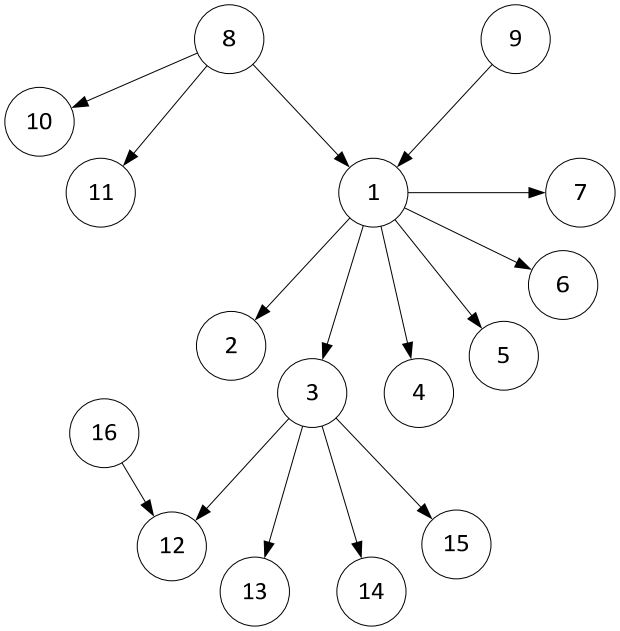


Figure 10.3 Graphical representation of the network structure used in this work.

10.2.2.1 Stationary Linear System

In this case study, the original variables relationships were linearized according to Equations (10.12), where ε_i is a white noise sequence with a signal-to-noise ratio of 10 dB (see Equation (5.8)). This system was then subjected to a set of perturbations, as described in Table 10.3, where δ represents a multiplicative factor that causes a change on the model's parameter on the range of ±20%. For the purposes of comparing the monitoring statistics, the original variables order was kept for transformation T_{Ch}. Likewise, the real causal network was also considered in the construction of the sensitive enhancing transformation T_{Net}, since it can be easily reconstructed from the data. In addition to these SET, the original variables were also used in combination with the proposed monitoring statistics.

$$
\begin{aligned}
&g_8 = \varepsilon_8,\ g_9 = \varepsilon_9,\ g_{16} = \varepsilon_{16} \\
&g_{10} = 1 + 0.40 g_8 + \varepsilon_{10} \\
&g_{11} = 0.56 + 0.15 g_8 + \varepsilon_{11} \\
&g_1 = 1.2 g_8 + 0.80 g_9 + \varepsilon_1 \\
&g_2 = 0.60 g_1 + \varepsilon_2 \\
&g_3 = 0.05 + 0.22 g_1 + \varepsilon_3 \\
&g_4 = 1 + 0.4 g_1 + \varepsilon_4 \\
&g_5 = 0.062 + 0.16 g_1 + \varepsilon_5 \\
&g_6 = 0.60 g_1 + \varepsilon_6 \\
&g_7 = 0.70 g_1 + \varepsilon_7 \\
&g_{12} = 0.80 g_{16} + 0.51 g_3 + \varepsilon_{12} \\
&g_{13} = 1.30 g_3 + \varepsilon_{13} \\
&g_{14} = 1 + 0.40 g_3 + \varepsilon_{14} \\
&g_{15} = 0.028 + 1.30 g_3 + \varepsilon_{15}
\end{aligned}
\tag{10.12}
$$

Table 10.3 Definition of the faults and respective variables involved in the network system without dynamics or non-linearities. δ is a multiplicative factor that changes the model parameters (under NOC, $\delta = 1$).

Fault	Variables relation changed
A	$g_8 \rightarrow g_1$ $(g_1 = 1.2\delta g_8 + 0.80 g_9 + \varepsilon_1)$
B	$g_1 \rightarrow g_3$ $(g_3 = 0.05 + 0.22\delta g_1 + \varepsilon_3)$
C	$g_8 \rightarrow g_{10}$ $(g_{10} = 1 + 0.40\delta g_8 + \varepsilon_{10})$
D	$g_3 \rightarrow g_{14}$ $(g_{14} = 1 + 0.40\delta g_3 + \varepsilon_{14})$

The distribution of the comparison index (N) over each type of fault is represented in Figure 10.4 and the permutation tests for each pair of monitoring statistics are presented in Table S1 of the supplementary material (available in the CD attached to this thesis). Among the monitoring statistics already proposed in the literature, the W and G statistics are the ones presenting the best performances, while the determinant of the sample covariance matrix, $|\mathbf{S}|$, is unable to detect any change. The E statistic only performs well on fault A, where the perturbation occurs close to the root node of the network and is propagated to almost all variables. As the number of variables affected by a fault decreases, the performance of the E statistic also degrades. This was an expected result, since the E statistic is focused in detecting changes in variance for a fixed correlation structure. From these results, the W statistic is chosen as a benchmark against which the new proposed statistics will be compared. However, it is important to point out that the W

statistic requires the inversion of the covariance matrix (see Equation (4.1)), which in some situations may be ill-conditioned or even singular. In such case, a pseudo-inverse or a linear transformation of the variables can be performed without changing the statistic value, since it is invariant to non-singular linear transformations [82]. The only observable change of this modification was on the asymptotic distributional behavior of the statistic, which seems to follow now a non-central chi-square distribution. This situation reinforces the choice of setting the control limits by trial and error adjustment, instead of using the theoretical control limits.

Regarding the results for the proposed monitoring statistics, one can notice that the sensitivity enhancing transformation proved to be a relevant factor in the monitoring task, as it significantly increases the statistics performance. In fact, without such transformation, the statistics performance is worse than for the *W* statistic. It is also noticeable that without the transformation the use of the partial correlations tends to decrease the detection capability of the proposed statistics, especially the ones based on MSPC-PCA.

After applying the sensitivity enhancing transformations most of the proposed statistics outperform the current *W* statistic, with the exception of $C_{r0,Ch}$ (MSPC-PCA applied to the marginal correlation of the transformed variables T_{Ch}) which has a similar performance (*p*-value of 0.048). This shows that the use of partial correlations has a relevant effect on the MSPC-PCA based statistics, especially for changes of small magnitude and when few variables are involved in the fault (faults C and D), i.e., when the fault is rather localized in the process. However, these statistics require a significant amount of data in order to construct the PCA model for normal operation conditions.

Regarding the *RMAX* statistics, it can be concluded that the variables' transformation also improves their performance. The performance of the *RMAX* statistics with transformed variables was always found to be better than the *W* statistic and also presented a better detection capability than the MSPC-PCA statistics. Curiously, the use of partial correlations seems to have no effect on the performance of these statistics after variables transformation, as both $R0MAX_{Net}$ and $R1MAX_{Net}$ (*RMAX* applied to marginal and partial correlations of the transformed variables T_{Net}, respectively), present a performance index (*N*) consistently close to one, i.e. meaning that higher detection rates are being achieved (see Figure 10.4). Moreover, both linear transformations (T_{Ch} and T_{Net}) conducted to

similar performances, with detection rates slightly higher on the $R0MAX_{Net}$ and $R1MAX_{Net}$ statistics. The performance of $R0MAX_{ChExt}$ and $R1MAX_{ChExt}$ (*RMAX* applied to marginal and partial correlations of the transformed variables T_{ChExt}, respectively) was not assessed, since in this case study they are essentially equal to their stationary counterparts ($R0MAX_{Ch}$ and $R1MAX_{Ch}$).

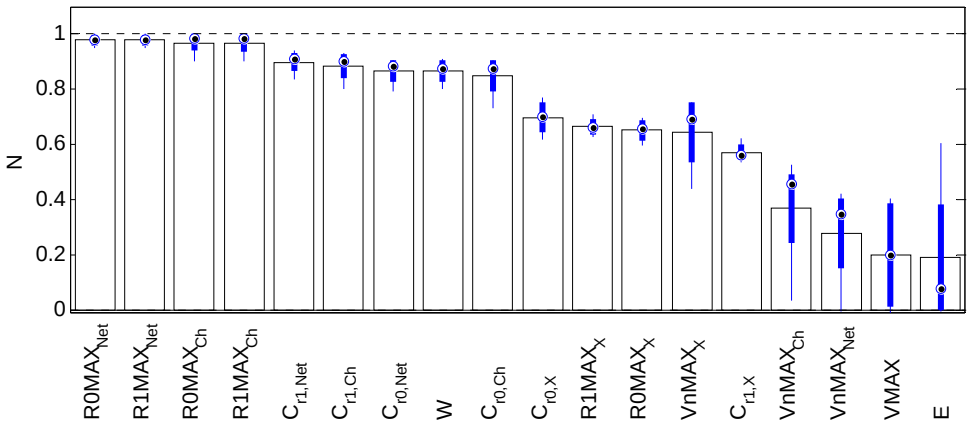


Figure 10.4 Comparison of the leading statistics performances for the network system with no dynamics and no non-linearities, with a sample covariance matrix computed from 3000 observations: box-plots of the performance index *N* based on the area under the fault detection curve obtained for all perturbations. Bar heights correspond to the associated mean values.

A similar analysis was performed with sample covariance matrices obtained from 100 observations. The results from this case study are presented in Figure 10.5 and clearly show that the number of observations used to estimate de covariance matrix has a great impact on the monitoring statistics capability to detect changes on the process structure. All the studied statistics presented significant reductions in their performance. This decrease is more noticeable on the MSPC-PCA-based statistics, which remains close the *W* statistic but are no longer comparable to any of the *RMAX* statistics with transformed variables. In fact, it is visible that the *RMAX* statistics continue to be ranked as the best monitoring statistics in all faults. Moreover, by cross comparison of Figure 10.4 and Figure 10.5 it is observed that the relative performance of the *RMAX* statistics is also much larger than that for the other monitoring schemes. Thus, the *RMAX* statistics remain capable to efficiently detect changes in the correlation structure when the number of observation is small and is also less sensitive to variations in this design parameter.

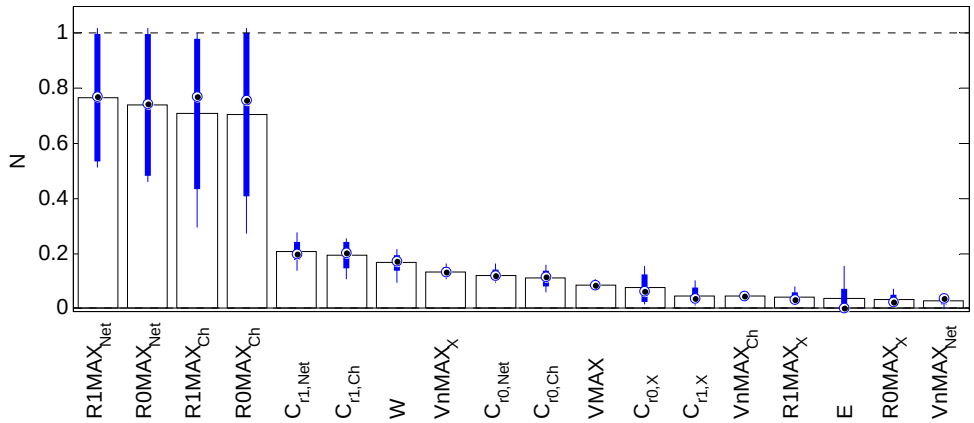


Figure 10.5 Comparison of the leading statistics performance for the network system with no dynamics and no non-linearities with a sample covariance matrix computed from 100 observations: box-plots of the performance index *N* based on the area under the fault detection curve obtained for all perturbations. Bar heights correspond to the associated mean values.

10.2.2.2 Dynamic Linear System

A similar analysis was performed with a dynamic version of the same network system with the addition of a multivariate time series dependency between variables according to Equation (10.13), where ε_i is a white noise sequence with a signal-to-noise ratio of 10 dB. The recovery of the causal network of this system was already addressed in Section 9.2, from where only the relationship between variables g_1 and g_4 was wrongly directed. However, as this situation can be mitigated by the use of an addition Cholesky decomposition of the residuals in transformation T_{NetCh}, in this study the correct causal network is used to construct transformation T_{Net}. Thus, only transformations T_{Ch}, T_{ChExt} and T_{Net} are considered in this study. For the dynamic transformations 3 lags were used. All transformations were determined based on a data set of NOC observations. This process was subjected to the perturbations presented in Table 10.4, where δ was changed to cause perturbations in the range of ±20%.

The results involving the original variables and transformation T_{Ch} are consistent with the ones obtained for the stationary case, i.e., the use of uncorrelated variables improves the detection capability of the monitoring methods. The main difference is the better performance of $R1MAX_{Ch}$ relatively to $R0MAX_{Ch}$, even though they show a similar performance for the W statistic (see Figure 10.6 and Table S2 of the supplementary material). In general, transformation T_{Ch} does not contribute to a significant increase on the performance of the monitoring statistics, as it does not handles the variables' time dependency, failing to detect changes on the variables dynamics.

$$
\begin{aligned}
&g_{8,t} = \varepsilon_{8,t},\ g_{9,t} = \varepsilon_{9,t},\ g_{16,t} = \varepsilon_{16,t} \\
&g_{10,t} = 1 + 0.40\left(g_{8,t} + 0.60g_{8,t-1} - 0.30g_{8,t-2}\right) + \varepsilon_{10,t} \\
&g_{11,t} = 0.56 + 0.15\left(g_{8,t} + 0.40g_{8,t-1} + 0.60g_{8,t-2}\right) + \varepsilon_{11,t} \\
&g_{1,t} = 1.2\left(g_{8,t} + 0.60g_{8,t-1} + 0.30g_{8,t-2}\right) + 0.80g_{9,t} + \varepsilon_{1,t} \\
&g_{2,t} = 0.60\left(g_{1,t} + 0.50g_{1,t-1} + 0.20g_{1,t-2}\right) + \varepsilon_{2,t} \\
&g_{3,t} = 0.05 + 0.22\left(g_{1,t} - 0.40g_{1,t-1} - 0.20g_{1,t-2}\right) + \varepsilon_{3,t} \\
&g_{4,t} = 1 + 0.4\left(g_{1,t} - 0.20g_{1,t-1} - 0.10g_{1,t-2}\right) + \varepsilon_{4,t} \\
&g_{5,t} = 0.062 + 0.16\left(g_{1,t} + 0.40g_{1,t-1} + 0.60g_{1,t-2}\right) + \varepsilon_{5,t} \\
&g_{6,t} = 0.60\left(g_{1,t} + 0.80g_{1,t-1} + 0.10g_{1,t-2}\right) + \varepsilon_{6,t} \\
&g_{7,t} = 0.70\left(g_{1,t} + 0.40g_{1,t-1} + 0.20g_{1,t-2}\right) + \varepsilon_{7,t} \\
&g_{12,t} = 0.80\left(g_{16,t} + 0.60g_{16,t-1} + 0.30g_{16,t-2}\right) + 0.51g_{3,t} + \varepsilon_{12,t} \\
&g_{13,t} = 1.30\left(g_{3,t} + 0.50g_{3,t-1} + 0.50g_{3,t-2}\right) + \varepsilon_{13,t} \\
&g_{14,t} = 1 + 0.40\left(g_{3,t} + 0.40g_{3,t-1} + 0.60g_{3,t-2}\right) + \varepsilon_{14,t} \\
&g_{15,t} = 0.028 + 1.30\left(g_{3,t} + 0.60g_{3,t-1} - 0.30g_{3,t-2}\right) + \varepsilon_{15,t}
\end{aligned}
\tag{10.13}
$$

Table 10.4 Definition of the faults and respective variables involved in the network system with linear dynamics. δ is a multiplicative factor that changes the model parameters (under NOC, $\delta = 1$).

Fault	Variables relation changed
A	$g_8 \rightarrow g_1$ $(g_{1,t} = 1.2\delta\left(g_{8,t} + 0.60g_{8,t-1} + 0.30g_{8,t-2}\right) + 0.80g_{9,t} + \varepsilon_{1,t})$
B	$g_1 \rightarrow g_3$ $(g_{3,t} = 0.05 + 0.22\delta\left(g_{1,t} - 0.40g_{1,t-1} - 0.20g_{1,t-2}\right) + \varepsilon_{3,t})$
C	$g_8 \rightarrow g_{10}$ $(g_{10,t} = 1 + 0.40\delta\left(g_{8,t} + 0.60g_{8,t-1} - 0.30g_{8,t-2}\right) + \varepsilon_{10,t})$
D	$g_3 \rightarrow g_{14}$ $(g_{14,t} = 1 + 0.40\delta\left(g_{3,t} + 0.40g_{3,t-1} + 0.60g_{3,t-2}\right) + \varepsilon_{14,t})$

On other hand, the use of a transformation with dynamic components (T_{ChExt} and T_{Net}) overcomes the deficiencies of transformation T_{Ch}, by the simultaneous consideration of both cross- and auto-correlation. This feature, leads to an increase of the monitoring statistics detection capabilities, especially on the case of the *RMAX* statistics. Again, $R0MAX_{Net}$ and $R1MAX_{Net}$ presented a consistently better performance than the other monitoring statistics (see Figure 10.6), showing the importance of applying a transformation that efficiently breaks the variables' correlation by taking into account the underlying system's structure.

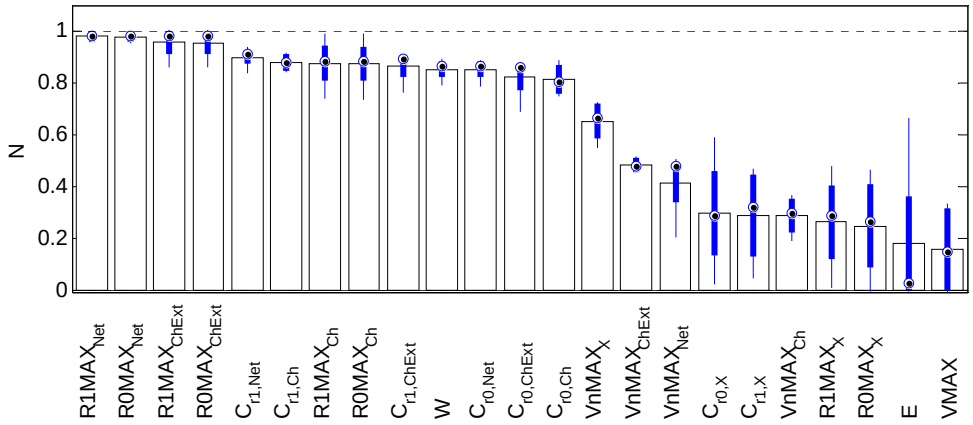


Figure 10.6 Comparison of the leading statistics performance for the network system with linear dynamics with a sample covariance matrix computed from 3000 observations: box-plots of the performance index *N* based on the area under the fault detection curve obtained for all perturbations. Bar heights correspond to the associated mean values.

The impact of the number of observations on the statistics monitoring performance was also studied by estimating the sample covariance matrices with 100 observations. The rest of the procedure was the same. As in the stationary linear system, the monitoring statistics performance decreased as can be seen in Figure 10.7. Again, the $R0MAX_{Net}$ and $R1MAX_{Net}$ statistics presented the best overall performance with no significant difference between each other. However, the relative performance of the other monitoring statistics is greatly diminished. Note that with estimates based on 3000 observations most of the monitoring statistics have a similar performance (see Figure 10.6), while with 100 observations only the *RMAX*-base statistics remain on top (see Figure 10.7). This indicates that the number of observations has a smaller impact on the *RMAX* statistics, making them more suitable for monitoring procedure based on a reduced number of observations per sample.

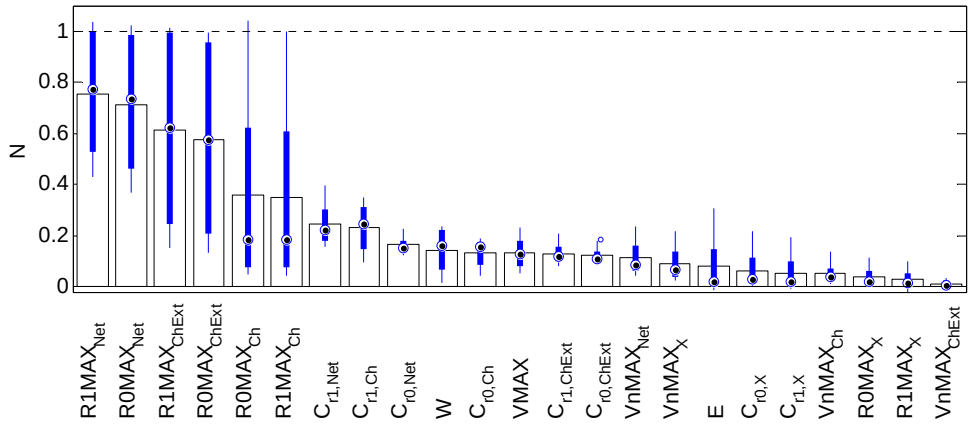


Figure 10.7 Comparison of the leading statistics performance for the network system with linear dynamics with a sample covariance matrix computed from 100 observations: box-plots of the performance index *N* based on the area under the fault detection curve obtained for all perturbations. Bar heights correspond to the associated mean values.

10.2.2.3 Stationary Non-linear System

In this case study, the original non-linear structure of the network system was approximated by polynomial relationships according to Equations (10.14), where ε_i is a white noise sequence with a signal-to-noise ratio of 10 dB. By application of the identification and causal directionality inference algorithms referred in Chapter 9 to a data set with NOC data, the original causal network was obtained (full reconstruction). The subsequent variables transformation T_{Net} was then carried out, taken this inferred structure into account, and a 3rd order polynomial model was fitted for each variable using the identified variables' parents as regressors. Transformation T_{ChExt} was also obtained from the same reference data set by adding polynomial terms to the data matrix. These transformed variables, along with the original ones, were then used to monitor the process structure. The system was perturbed with the faults presented in Table 10.5, where the magnitude of change, δ, was set to cause deviations on the model's parameter on the range of ±10%. For each fault the performance index N was obtained, resulting in the global distribution shown in Figure 10.8. Similarly, the results of the permutations tests are summarized in Table S3 of the supplementary material.

$$
\begin{aligned}
&g_8 = \varepsilon_8,\ g_9 = \varepsilon_9,\ g_{16} = \varepsilon_{16} \\
&g_{10} = 0.020 g_8^2 + 0.44 g_8 + 0.82 + \varepsilon_{10} \\
&g_{11} = -0.053 g_8^3 - 0.00068 g_8^2 + 0.52 g_8 + 0.50 + \varepsilon_{11} \\
&g_1 = 1.20 g_8 + 0.80 g_9 + \varepsilon_1 \\
&g_2 = 0.60 g_1 + \varepsilon_2 \\
&g_3 = \left(\delta g_1 - 4\right)\left(g_1 + 4\right) + \varepsilon_3 \\
&g_4 = 0.020 g_1^2 + 0.44 g_1 + 0.82 + \varepsilon_4 \\
&g_5 = -0.057 g_1^3 - 0.077 g_1^2 + 0.52 g_1 + 0.22 + \varepsilon_5 \\
&g_6 = 0.60 g_1 + \varepsilon_6 \\
&g_7 = 0.70 g_1 + \varepsilon_7 \\
&g_{12} = 0.80 g_{16} + 0.60 g_3 + \varepsilon_{12} \\
&g_{13} = 1.30 g_3 + \varepsilon_{13} \\
&g_{14} = 0.020 g_3^2 + 0.44 g_3 - 0.82 + \varepsilon_{14} \\
&g_{15} = 1.40 g_3 + \varepsilon_{15}
\end{aligned}
\tag{10.14}
$$

Table 10.5 Definition of the faults and respective variables involved in the network system with a non-linear model structure. δ is a multiplicative factor that changes the model parameters (under NOC, $\delta = 1$).

Fault	Variables relation changed
A	$g_8 \rightarrow g_1 \ (g_1 = 1.20\delta g_8 + 0.80 g_9 + \varepsilon_1)$
B	$g_1 \rightarrow g_3 \ (g_3 = (\delta g_1 - 4)(g_1 + 4) + \varepsilon_3)$
C	$g_8 \rightarrow g_{10} \ (g_{10} = 0.02 g_8^2 + 0.44\delta g_8 + 0.82 + \varepsilon_{10})$
D	$g_3 \rightarrow g_{14} \ (g_{14} = 0.020\delta g_3^2 + 0.44 g_3 - 0.82 + \varepsilon_{14})$
E	$g_8 \rightarrow g_{11} \ (g_{11} = -0.053\delta g_8^3 - 0.00068 g_8^2 + 0.52 g_8 + 0.50 + \varepsilon_{11})$
F	$g_3 \rightarrow g_{15} \ (g_{15} = 1.40\delta g_3 + \varepsilon_{15})$

Among the current monitoring statistics, only the *E* statistics present a good performance in detecting fault A, while for the remaining faults, all current statistics (including *E*) present detection rates of less than 0.05. These poor detection capabilities may be partially associated with the kind of perturbations simulated, which were made on non-linear terms or in network boundary nodes. This also explains why the *E* statistic was capable to detect fault A, which occurs close to the root node of the network and is propagated to the rest of the network nodes. Given these results, the *E* statistics was chosen as the benchmark for this case study.

On the other hand, the proposed statistics were generally able to detect the simulated faults. Again, the sensitivity enhancing transformations and the use of partial correlations increase the detection capability of both MSPC-PCA and *RMAX* based monitoring statistics. Note that even the linear transformation, T_{Ch}, leads to great improvements in the monitoring statistic based on it (namely, $R0MAX_{Ch}$, $R1MAX_{Ch}$, $C_{r0,Ch}$ and $C_{r1,Ch}$), while the *W* statistic would remain unchanged since it is invariant to any non-singular linear transformation [82]. However, due to the system's non-linear nature, transformations T_{Ch} and T_{ChExt} are not suitable for the *RMAX* statistics, as the partial correlations do not follow their assumed NOC probability distribution. For this case, a more complex transformation is required or, alternatively, an estimation of the partial correlations distribution for describing their variability. Yet, the use of transformation T_{Net} is preferable because the actual variables' relationships are explicitly considered in its construction. Moreover, $R0MAX_{Net}$ and $R1MAX_{Net}$ can detect efficiently most of the simulated faults, with the exception of fault D which is detected by $VnMAX_{Net}$. Hence, transformation T_{Net} is adequate for modeling and monitoring this non-linear system, since either $RMAX_{Net}$ or

$VnMAX_{\text{Net}}$ rapidly detect any structural change, while the current monitoring statistics remain mostly under their control limits.

The MSPC-PCA based statistics are less sensitive to the partial correlations distribution, as most of the scores would still approximately follow a normal distribution due to the central limit theorem effect induced in the computation of the scores. Nevertheless, the *RMAX* statistics proved to have potential to be applied on non-linear systems and, when properly scaled, outperform all the studied statistics.

As in the previous case studies, the same monitoring procedures were conducted with samples covariance matrices estimated from 100 observations. For this case, the performance index *N* remains in line with the ones obtained with 3000 observations. Yet, it is still visible that while the *RMAX* statistics (namely $R0MAX_{Net}$ and $R1MAX_{Net}$) are ranked as the best, the performance of the MSPC-PCA-based statistics is reduced (for instance, $C_{r1,Net}$ is ranked in the 3rd position in Figure 10.8 and in 7th position in Figure 10.9).

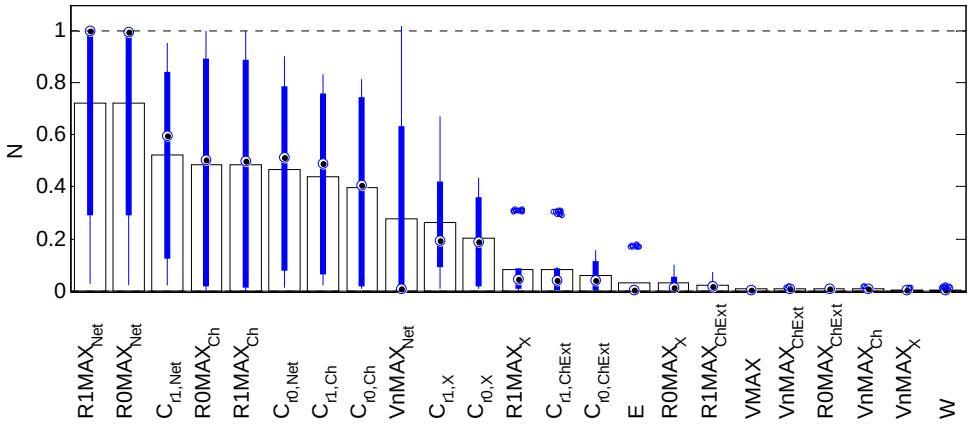


Figure 10.8 Comparison of the leading statistics performance for the network system with a non-linear model structure with a sample covariance matrix computed from 3000 observations: box-plots of the performance index *N* based on the area under the fault detection curve obtained for all perturbations. Bar heights correspond to the associated mean values.

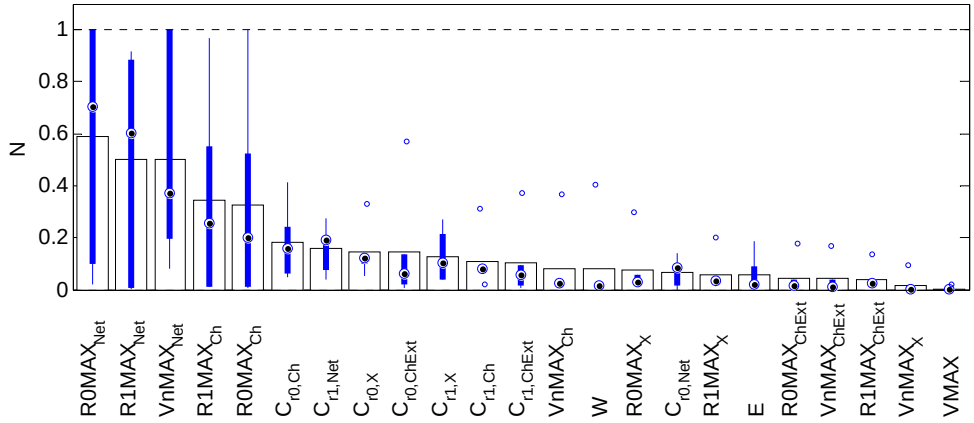


Figure 10.9 Comparison of the leading statistics performance for the network system with a non-linear model structure with a sample covariance matrix computed from 100 observations: box-plots of the performance index *N* based on the area under the fault detection curve obtained for all perturbations. Bar heights correspond to the associated mean values.

10.2.3 Analysis of the SET in Non-linear Dynamic Systems

In this section, the comparative assessment of the monitoring statistics performance is conducted with resource to systems presenting dynamic non-linear model structures. This class of model structures can be found in real processes and present additional complexity features regarding the ones studied before, such as close-loops and bidirectional dependencies. In the case studies considered in this section, 500 sample covariance matrices were computed and collected for each fault. Each sample covariance matrices was calculated from 3000 observations. The fault detection rates were subsequently adjusted to a fault detection rate of 1% under normal operation conditions. This procedure was repeated 5 times in order to obtain approximate confidence intervals for the fault detection rates.

10.2.3.1 Gene Network Model

The test system considered in this case study was proposed by Fuente *et al.* (2004) [130]. It regards the dynamic modeling of transcription levels (T_i) in a gene network, and consists of the following set of differential equations:

$$\begin{aligned}
\frac{dT_1}{dt} &= \frac{V_1}{\left(1+K_{T_1}/T_4\right)} - k_1T_1 + \theta_1T_1 \\
\frac{dT_2}{dt} &= \frac{V_2}{\left(1+T_1/K_{T_1}\right)} - k_2T_2 + \theta_2T_2 \\
\frac{dT_3}{dt} &= \frac{V_3}{\left(1+K_{T_2}/T_2\right)} - k_3T_3 + \theta_3T_3 \\
\frac{dT_4}{dt} &= \frac{V_4}{\left(1+T_2/K_{T_2}\right)} - k_4T_4 + \theta_4T_4 \\
\frac{dT_5}{dt} &= \frac{V_5}{\left(1+T_3/K_{T_3}\right)\left(1+K_{T_4}/T_4\right)} - k_5T_5 + \theta_5T_5
\end{aligned} \tag{10.15}$$

where V_i are maximal transcription rates, k_i degradation rate constants, K_i inhibition or activation constants and θ_i error terms designed to simulate biological variability. Parameters V_i, k_i and K_i were set to unity and the error terms (θ_i) were sampled from a normal distribution with zero mean and standard deviation of 0.01. The underlying network is represented in Figure 10.10. The recovery of the undirected network from data was already addressed by Fuente *et al.* (2004) [130] through the use of partial correlations. As the directions of the edges can also be obtained by analysis of the cross-correlation between the variables, the complete network was considered in the construction of transformations T_{Net} and T_{NetCh}.

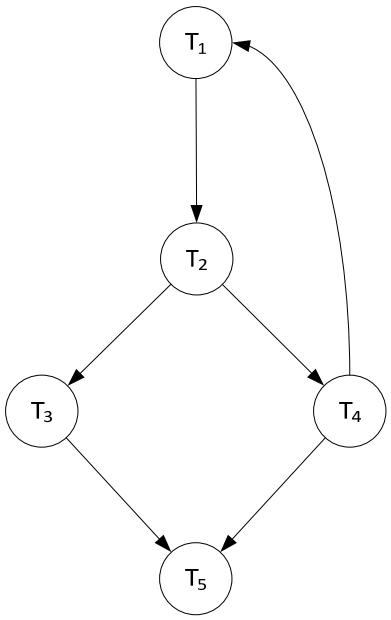


Figure 10.10 Directed network representation of the gene network model.

The system was subjected to perturbations on V_1, V_5, k_1 and k_5 in a magnitude range of ±20% and the results obtained appear summarized by the performance index N in Figure 10.11. The permutation tests are available in Table S4 of the supplementary material. Analyzing these results, it is possible to verify that, for this particular system, the current monitoring statistics presented higher detection rates than MSPC-PCA and *RMAX* based statistics, regardless of the transformation used. Special emphasis is given to W and G, which had the best performances, and to the E statistic, which also indicates that fault detection is primarily done at the variables' variance level. These results are corroborated by the raw data where only small changes in the correlation matrix are observed. The most significant changes occurred in the process mean, since the faults drive the process to a new steady state, and in a lesser extent in the process' variance. Therefore, as none of the partial correlation based statistics account for changes in variance, all of them show poor detection performances under these simulated scenarios. These results stress the need for the adoption of a complementary monitoring statistic dedicated to supervising process variance.

For detecting a change on the variance, the *VMAX* statistic was extended to detect both an increase and a decrease in process variance. Its direct extension without transformation, $VnMAX_X$, showed to be slightly worse than *VMAX* in detecting increases in variance, due to the increase in the control limit. However, when the variance decreases, $VnMAX_X$ was capable to signal such change while *VMAX* is unable to detect them at all. Consequently $VnMAX_X$ is globally better than *VMAX* when the detection of both an increase and decrease is of interest, as can be seen in Figure 10.11. The performance of *VnMAX* can be further improved by use of an appropriated transformation, namely transformations T_{ChExt}, T_{Net} and T_{NetCh}, which incorporate dynamic dependencies.

It is important to note that perturbations detected by *VnMAX* are not only related to changes in the variance, but also to changes in the process' structure when transformed variables are used. In this case, if a change in the process' structure occurs, the transformation no longer describes the variables' relationships correctly. Therefore, non-accounted contributions will lead to deviations on the transformed variables' variance, which will ultimately be captured by *VnMAX*. This behavior shows the importance of monitoring the variance in order to also detect structural changes, when using SET.

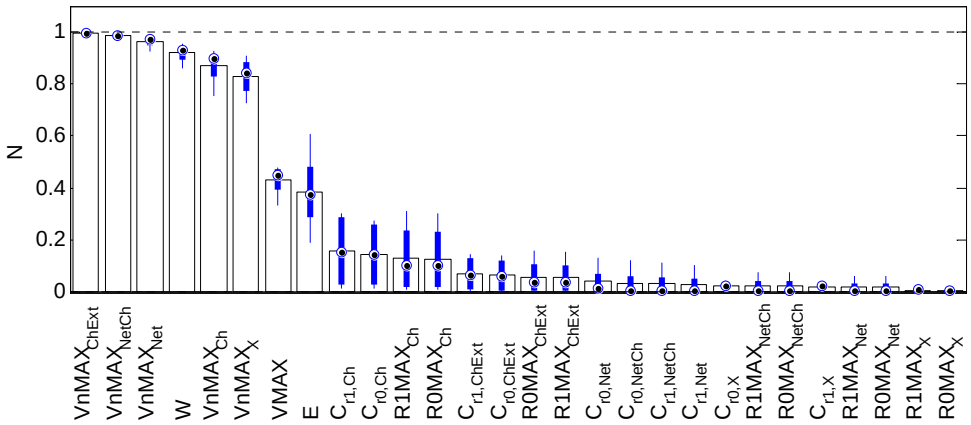


Figure 10.11 Comparison of the statistics performance on the gene network model: box-plots of the performance index *N* based on the area under the fault detection curve obtained for all perturbations. Bar heights correspond to the associated mean values.

10.2.3.2 Biologic Production of Ethanol

This dynamic non-linear system is based in the ethanol production from glucose fermentation by *Zymomonas mobilis* bacteria. One of the models proposed to describe the dynamics of this process is the Jöbses' model [142], given by the following equations:

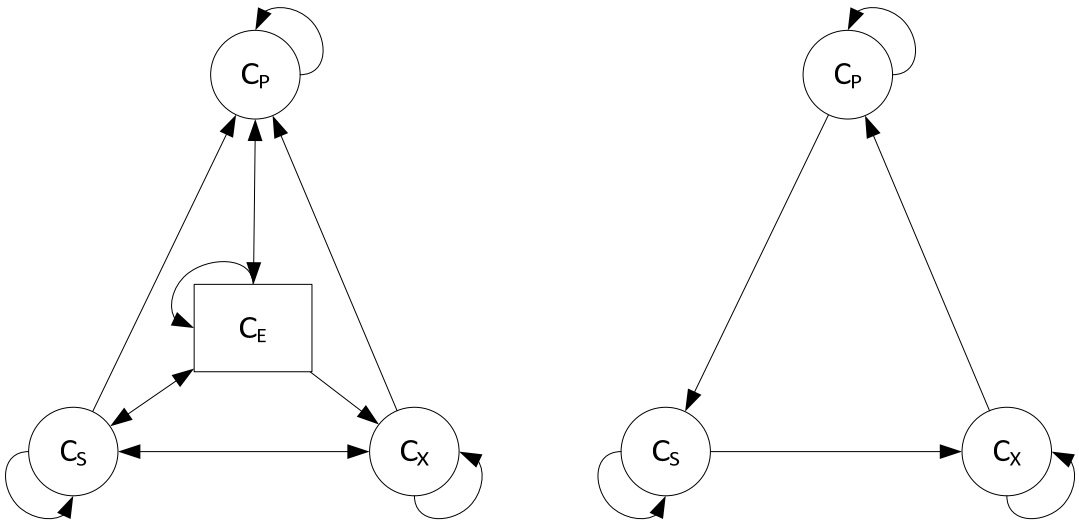
$$\begin{aligned}
\frac{dC_S}{dt} &= -\left(\frac{\mu_{\max} C_S C_E}{Y_{SX}\left(K_S + C_S\right)}\right) - m_S C_X + D\left(C_{S0} - C_S\right) \\
\frac{dC_X}{dt} &= \left(\frac{\mu_{\max} C_S C_E}{K_S + C_S}\right) + D\left(C_{X0} - C_X\right) \\
\frac{dC_E}{dt} &= K_E\left(C_P - c_1\right)\left(C_P - c_2\right)\left(\frac{C_S C_E}{K_S + C_S}\right) + D\left(C_{E0} - C_E\right) \\
\frac{dC_P}{dt} &= -\left(\frac{\mu_{\max} C_S C_E}{Y_{PX}\left(K_S + C_S\right)}\right) - m_P C_X + D\left(C_{P0} - C_P\right)
\end{aligned} \tag{10.16}$$

where C_S is the substract (glucose) concentration, C_X is the biomass (*Zymomonas mobilis*), C_P is the product (ethanol) concentration, and C_E is an auxiliary variable used to account for the lagged effect of ethanol concentration in the kinetic model. The variables C_{S0}, C_{X0}, C_{E0} and C_{P0} complete the mass balance representing the input concentrations in the reactor, where normally only C_{S0} (substract inlet) is non-zero. The dilution rate (D) is the inverse of the resident time. The reactor volume was kept constant and all the remaining parameters are listed in Table 10.6. Only C_S, C_X and C_P were considered to be measured variables.

Table 10.6 Jöbses' model parameters.

Parameters	Values	Unit
K_E	0.00383	$m^6\ kg^{-2}\ h^{-1}$
c_1	59.2085	$kg\ m^{-3}$
c_2	70.5565	$kg\ m^{-3}$
K_S	0.500	$kg\ m^{-3}$
m_S	2.160	$kg\ kg^{-1}\ h^{-1}$
m_P	1.10	$kg\ kg^{-1}\ h^{-1}$
Y_{SX}	0.02445	$kg\ kg^{-1}$
Y_{PX}	0.05263	$kg\ kg^{-1}$
μ_{max}	1.0	h^{-1}

The system model represented by Equation (10.16), can be translated into the causal network presented in Figure 10.12 (a). This representation clearly shows a complex relationship between variables, as they are all directly or indirectly linked by close-loops. The reconstruction of the correct dependency between the variables connective structure has an impact in the monitoring statistics, since the derivation of the sensitivity enhancing transformations depends on this map.



(a) (b)

Figure 10.12 Jöbses' model network: (a) the original causal network, (b) the estimated causal network. Circles represent measured variables and rectangles unmeasured variables.

All the variable transformations described in Chapter 9 were determined based on a data set collected under normal operation conditions. In what concerns to transformation T_{ChExt}, no particular order was considered for the variables. As for transformations T_{Net} and T_{NetCh}, the algorithm developed for identifying the variables edges and dependencies (see Chapter 9) led to the graph shown in Figure 10.12 (b). The differences between the reconstructed network and the real one (presented in Figure 10.12 (a)), lie essentially in the fact that only one direction is considered together with the absence of the unmeasured variable, C_E. However, even with these structural mismatches, the transformations are able to break the major cross- and auto-correlations present in the original data.

In order to compare the monitoring statistics performance, the system was subjected to changes on c_1, K_S and Y_{SX}. The range of the perturbations was of ±25% for c_1 and Y_{SX} and ±50% for K_S. As in the previous case studies, the performance index N (Figure 10.13) and permutations tests (Table S5 in supplementary material) were computed and saved for analysis.

Looking to the current monitoring statistics, the W and G statistics were the ones that presented the best performance. However none of the current monitoring statistics successfully detect faults on K_S and Y_{SX}, leading to fault detections rates below 0.05. On the other hand, most of the proposed monitoring statistics are able to signal all the simulated faults. Only the statistics based on the original variables (T_X) presented poor detection capabilities. This result is expected as the T_X variables are highly correlated, and therefore changes in their values are difficult to detect. The effect of cross- and auto-correlation is eliminated through the use of the transformation T_{ChExt}, which includes time-shifted variables on the Cholesky decomposition. This is the simplest among the three proposed sensitivity enhancing transformations for dynamic systems and consequently both T_{Net} and T_{NetCh} transformations present even better performances, since they make use of more information about the variables' relationships. Regardless of the type of transformation adopted, the MSPC-PCA based monitoring statistics have in general better performances than *RMAX*. However, different monitoring statistics tend to have similar performances when computed over the same transformation, indicating that in this case study, the type of transformation is more relevant than the monitoring statistics used.

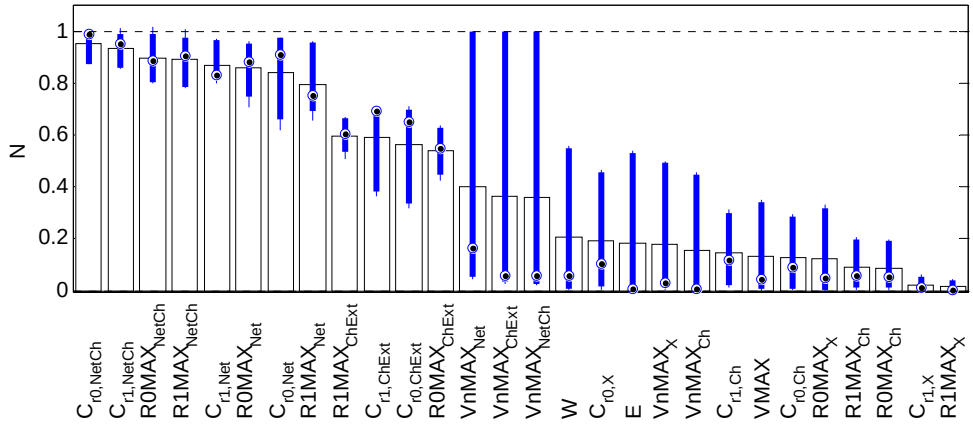


Figure 10.13 Comparison of the statistics performance on the Jöbses' model: box-plots of the performance index *N* based on the area under the fault detection curve obtained for all perturbations. Bar heights correspond to the associated mean values.

The main difference between transformations T_{Net} and T_{NetCh} is the additional Cholesky decomposition performed over the final residuals in T_{NetCh}. This extra step ensures that the obtained variables are indeed uncorrelated and accounts for any relationship that is not being considered by T_{Net}. This characteristic gives some robustness to the proposed methodology, since it accommodates certain limitations in the estimation of the true underlying causal network. As stated before, the retrieved causal network for this case study is substantially different from the real one. Yet, with transformation T_{NetCh}, the effects of the missed interactions are mitigated and an increase in the monitoring statistics performance is observed (see Figure 10.13). This shows that the proposed monitoring statistics based on this type of SET presents a reasonable degree of robustness, a useful feature in real world practical applications.

10.2.3.3 Continuous Stirred-Tank Reactor model

A dynamic model of a continuous stirred-tank reactor (CSTR) with a heating jacket and under feedback control was used to further assess the monitoring statistics performance. In this system, an endothermic reaction of the type $A \rightarrow B$ takes place in the CSTR with free discharge flow. The system is schematically represented by Figure 10.14. In this representation, C_{A0} (feed stream concentration) T_0 (feed stream temperature) and T_{j0} (heating fluid inlet temperature) are the system inputs. Likewise, h (CSTR level), C_A (concentration of compound A), T (CSTR temperature) and T_j (heating fluid outlet temperature) are the system outputs.

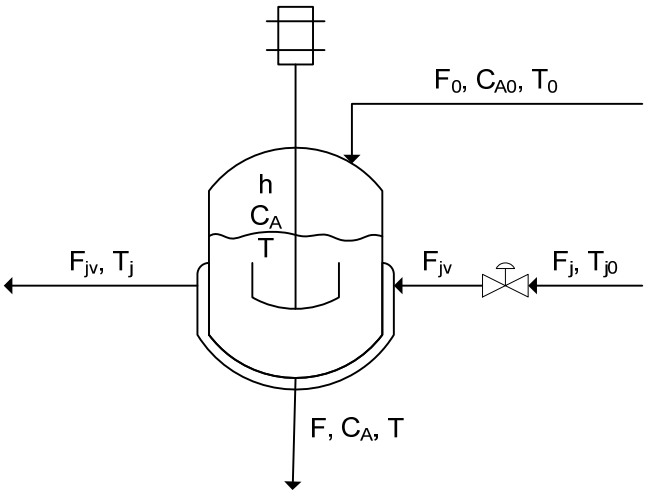
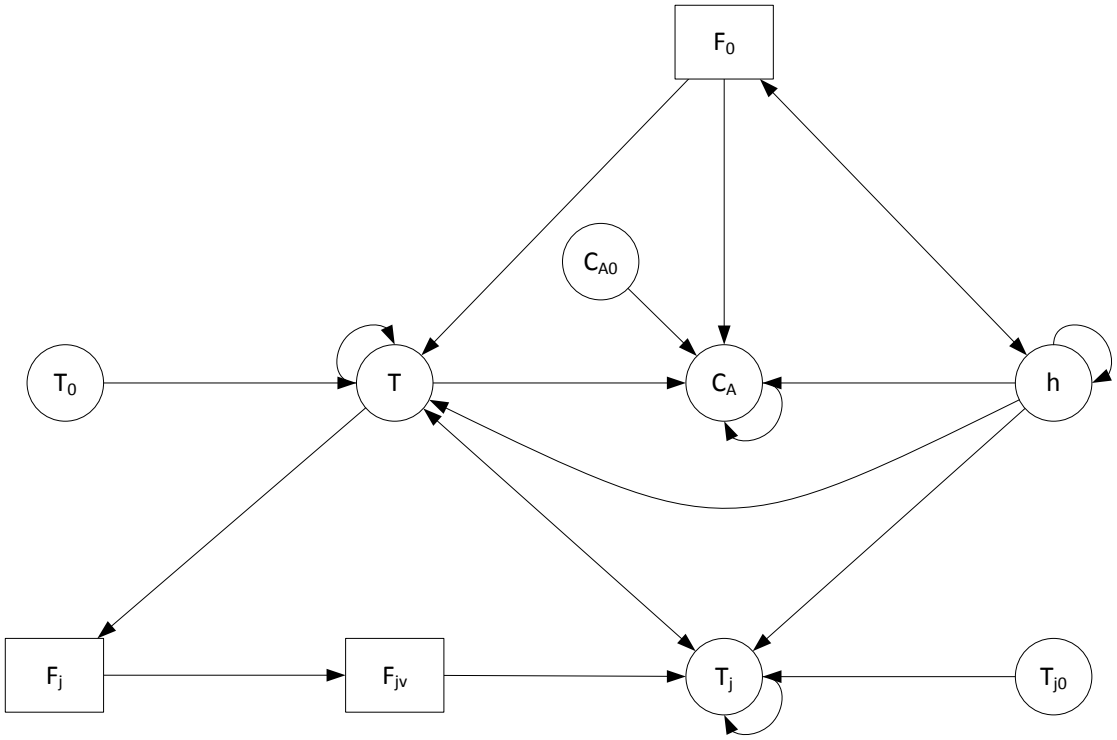


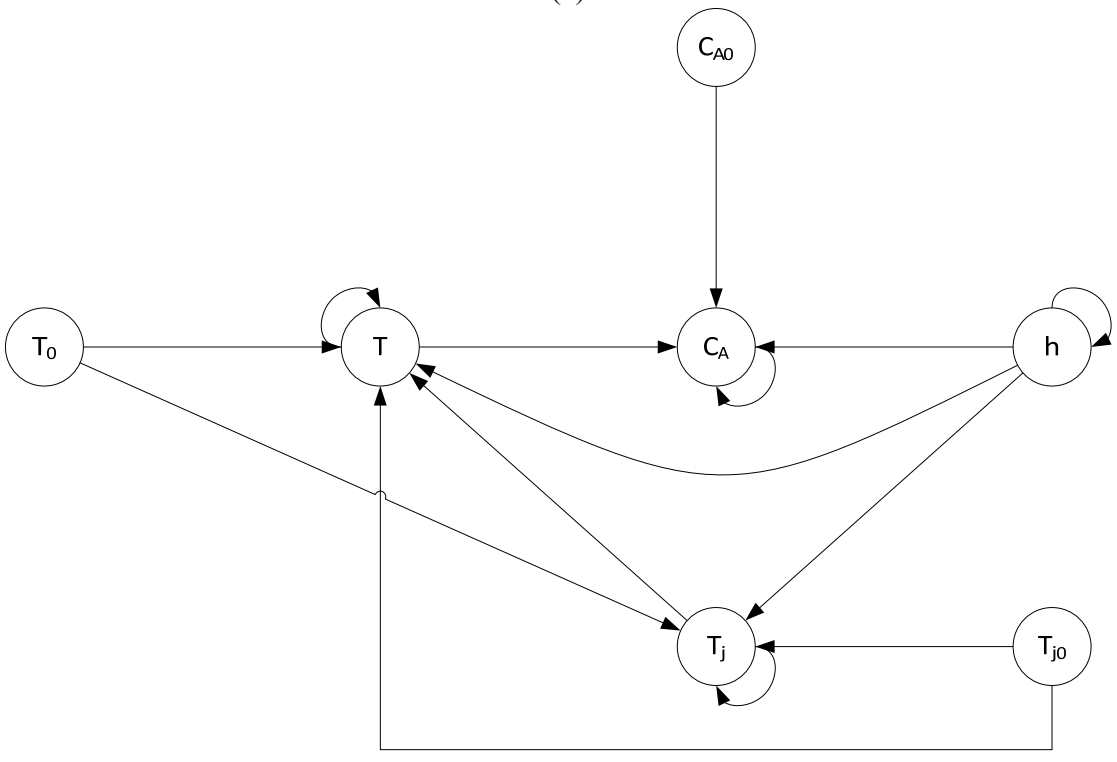
Figure 10.14 Process flow diagram for the CSTR system.

The transformations were determined based on a reference data set where the process was operating under normal conditions. Transformation T_{ChExt} with 2 lags for all variables was employed in order to model the dynamic features of the data. For transformations T_{Net} and T_{NetCh}, the identification of the linked variables and their directionality was determined by the data-driven algorithms referred in Chapter 9, even though the real variables connections were known in this case from the process flowsheet (Figure 10.15 (a)). The connections identified are presented in Figure 10.15 (b). These connections lead to a transformation T_{Net} that breaks most of the cross- and auto-correlation. Yet some residual correlations are observed which justifies the implementation of the Cholesky decomposition over the residuals (transformation T_{NetCh}).

In order to compare the performance of the monitoring statistics, the system was subjected to several perturbations, namely on the heat transfer coefficient, discharge coefficient, pre-exponential factor and valve's time constant (for which a 1st order dynamic model was assumed). The results are presented in Figure 10.16 by the performance index *N*. The permutation test results are given in Table S6 of the supplementary material.



(a)



(b)

Figure 10.15 Causal network for the CSTR system: (a) the original causal network, (b) the estimated causal network. Circles represent measured variables and rectangles unmeasured variables.

In this case study, the current monitoring statistics were only capable to detect faults on the discharge coefficient and decreases on the heat transfer coefficient while the proposed monitoring statistics based on transformations T_{ChExt}, T_{Net} and T_{NetCh} detected most of the faults, even for changes of small magnitude. Transformation T_{Net} led to the lowest performance statistics due to its difficulty to detect changes on the heat transfer coefficient and valve's time constant. On the first case (heat transfer coefficient), the MSPC-PCA and *RMAX* showed rather inconsistent results, as the fault detection curve was not monotonically increasing. However, the complementary $VnMAX_{Net}$ statistic leads to good results, indicating that this particular fault affects primarily the variance of the transformed variables. Recall that when transformed variables are used, the *VnMAX* statistic is also capable to detect changes in the process' structure as a result of model mismatch. This situation also reveals the necessity of this complementary statistic: a combination of *RMAX* and *VnMAX* with the transformation T_{Net}, leads to performances that are similar to other transformations.

Transformation T_{ChExt} is the second best transformation, yet it fails to detect faults on the valve's time constant. These perturbations were not detected by most of the studied monitoring statistics. Since this fault is related to the time that the control valve takes to respond to a control action, changes on the valve's time constant are translated into hard to detect time delays. Therefore, the detection of this fault relies in the transformation capability to model the system dynamics. Transformation T_{NetCh} (which differs from transformation T_{Net} only by an additional Cholesky decomposition) was the only one leading to monitoring statistics with good detection capabilities to changes on the valve's time constant, improving substantially their performance on the other faults. The monitoring statistics based on Transformation T_{NetCh} were also consistently better, namely $R0MAX_{NetCh}$ and $R1MAX_{NetCh}$ (see Figure 10.16), which confirms the potential of partial correlations to detect structural changes, especially when coupled with suitable variables transformation.

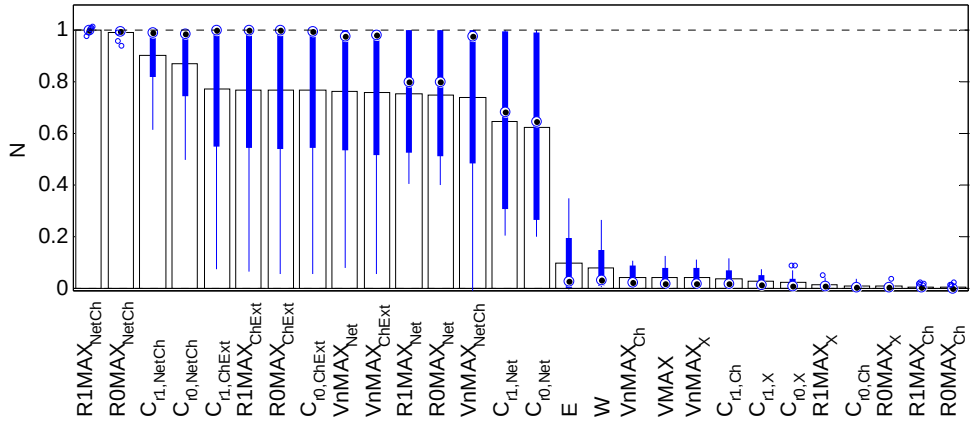


Figure 10.16 Comparison of the statistics performance on the CSTR system: box-plots of the performance index *N* based on the area under the fault detection curve obtained for all perturbations. Bar heights correspond to the associated mean values.

10.3 Discussion

The statistical process control of process multivariate dispersion is mostly done by analysis of either the generalized variance, likelihood test ratio or other dispersion measures that only considers the marginal distributions of process data. Therefore, the resulting monitoring statistics have a small resolution of the process inner structure and are unable to detect localized changes. This feature becomes more evident when the system cannot be assumed to be fairly linear around its operation state, a situation where such methods can perform quite poorly (e.g., for dynamic non-linear systems). To improve the ability to detect such structural faults in complex systems, the use of partial correlations as a measure of the variables local association was explored. This study showed that monitoring schemes based on partial correlation information obtained from process data can in fact improve the ability to detect changes in the process' structure.

However, the performance of the resulting monitoring statistics is dependent upon several factors. The most important one is the type of data transformation implemented. Therefore, the use of a sensitivity enhancing transformations for monitoring the variables' correlation structure is one of the most relevant contributions of this work. The application of these sensitivity enhancing transformations lead to a new set of uncorrelated variables that account from the process structure and, due to their uncorrelated nature, make the fault detection easier. Moreover, as the system is subjected to a fix model, any drift from the reference model can also be perceived as a structural

change. These types of faults are essentially translated into changes in variance and can be detected by an extended version of the current monitoring statistic *VMAX*. The proposed extension, *VnMAX*, is capable to detect both increases and decreases in variance, complementing the *RMAX* statistics in such a way that any structural change is detected by at least one of them. A particular example of this situation is illustrated in the case study of Section 10.2.2.3, where $R0MAX_{Net}$ and $R1MAX_{Net}$ do not detect fault D, while $VnMAX_{Net}$ was able to detect it. The way *VnMAX* is constructed also facilitates its combination with *RMAX* into a single monitoring statistic of the form $SMAX = \max\{RMAX, VnMAX\}$, since the variance and partial correlations are transformed into the same reference distribution, i.e., a normal distribution with zero mean and variance one. The combined use of these statistics will enable to monitor the full spectrum of structural changes conveyed by both *VnMAX* and *RMAX*. Yet, on this study, they were treated separately to clarify and understand their complementariness.

Both approaches for monitoring the process structure (MSPC-PCA-based and *RMAX*) had their performance improved when a sensitivity enhancing transformation based on partial correlations was used. The effect of the partial correlations was more noticeable on the MSPC-PCA-based statistics. Regarding the *RMAX* approach, the monitoring of the marginal correlation by *R*0*MAX* or of the 1st order partial correlations by *R*1*MAX*, seems to be statistically similar after application of a sensitivity enhancing transformation. This situation arises from the close relationship between these two measures but most importantly, because of the fact that the transformed variables are inherently uncorrelated, which means that there is no effect in controlling any pair of variables by a third one. Still, the inner process structure is explored, since the transformed variables result from the application of a model based on partial correlation, which are well-known by their capability to uncover the variables relationships. Furthermore, the whole concept of partial correlations is based on the regression residuals obtained by controlling two variables by a set of other. Consequently, the marginal correlation of the transformed variables can, to some extent, be considered as partial correlations of the original variables.

The major disadvantage of the proposed monitoring statistics is the amount of data required to construct the NOC PCA model (MSPC-PCA-based approach) and the requirement that all partial correlations follow the same distribution (*RMAX*). The latter

issue can be solved through the use of a proper data sensitivity enhancing transformation or by adopting an estimation procedure for defining the empirical distribution of the coefficients.

The number of observations used to estimate the sample covariance matrix is another relevant factor. For all the monitoring statistics, the performance decreases as the number of observations decrease, but their relative performance is maintained in general. The only exception regards the MSPC-PCA-based statistics which no longer outperform the *W* statistic when a small number of observations is used. The *RMAX* statistics showed to be less sensitive to this factor. This feature makes the *RMAX* family of statistics suitable candidates to on-line monitoring when coupled with an adequate sensitivity enhancing transformation.

Regarding fault diagnosis based on marginal and partial correlations, a simple analysis of the marginal correlations falling outside some pre-established threshold can give an indication of the variables involved and fault location. However, since the marginal correlation does not distinguish between directly from indirectly related variables, the variables involved may not be correctly identified in this way. This situation is exemplified in Figure 10.17 (a) for the stationary linear system of Section 10.2.2.1, regarding fault A (change in the relationship between variables 1 and 8) with a multiplicative factor δ of 1.20. To improve the diagnostic properties, 1st order partial correlations can be used. To do so, the number of partial correlations with values above the threshold is counted, with the i^{th} variable controlled. The controlled variable that gives the lowest count can then be considered as being related with the fault, based on the reasoning that if the faulty variable is controlled, then the corresponding partial correlations will remain under control, as a result of removing the faults origin. For the same fault described earlier the distribution of the number of partial correlations obtained by this procedure unequivocally selects the correct cause in about 90% of the cases, as can be seen in Figure 10.17 (b). Similar results are obtained in other faults, such as fault B represented in Figure 10.18. This fault is a result of a relationship change between variables 1 and 3, yet, by analysis of the marginal correlation, the correlations involving variables 13, 14 and 15 are above the threshold more frequently (see Figure 10.18 (a)). Even though these variables are directly dependent on variable 3, they do not identify the correct root case. On the other hand, variable 1 is identified as the root cause when partial correlations are employed (see Figure 10.18 (b)). These results, clearly suggest that

partial correlations can help in the isolation of the root cause more efficiently, as a result of conveying more localized information. Furthermore, in order to apply such procedure, it is recommended to use $R1MAX$ to monitor the process, since it guarantees that at least one partial correlation coefficient is out-of-control, if a threshold equal to the monitoring statistic UCL is used.

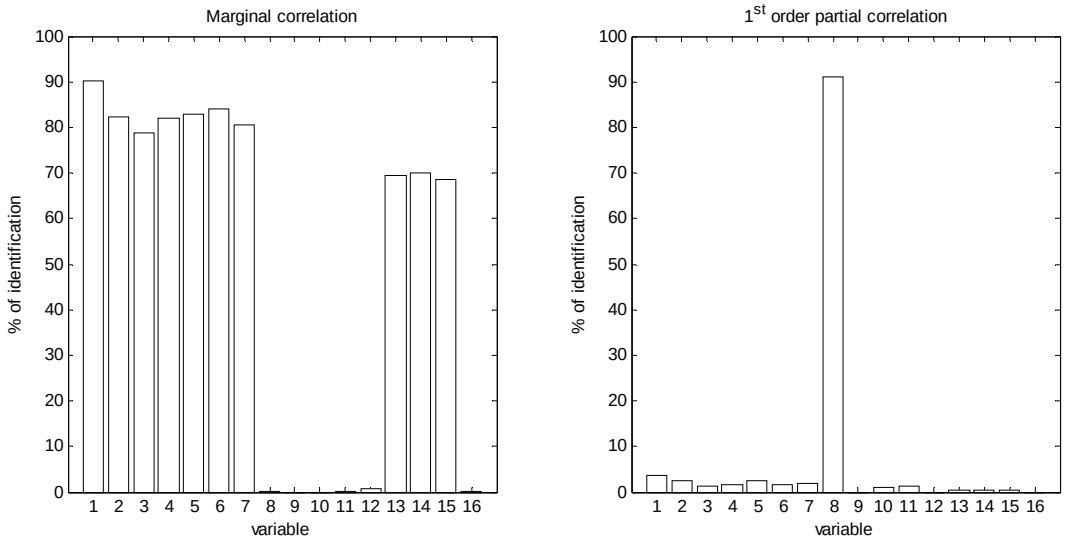


Figure 10.17 Percentage of instances that each variable was identified as the faults' root case on fault A (i.e., on the change of the relationship between variables 1 and 8), with $\delta = 1.20$, for the stationary linear system, in a total of 1000 cases. The thresholds used by both methods were set for the same statistical significance of 0.01. The plot on the left (a) presents the identification results for the diagnosis based on marginal correlation, whereas the plot on the right (b) regards the use of the proposed procedure based on partial correlations.

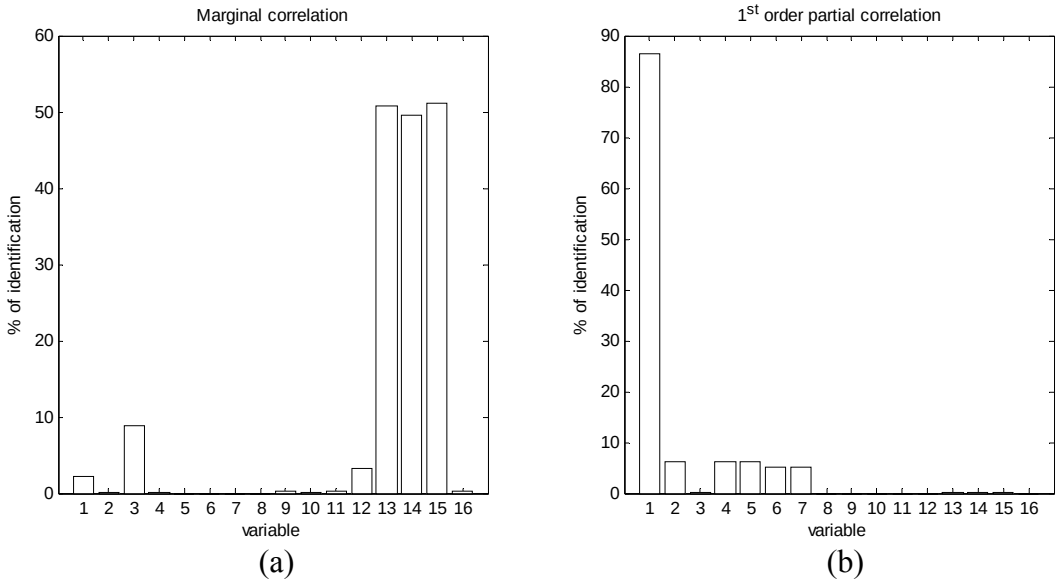


Figure 10.18 Percentage of instances that each variable was identified as the faults' root case on fault B (i.e., on the change of the relationship between variables 1 and 3), with $\delta = 1.10$, for the stationary linear system, in a total of 1000 cases. The thresholds used by both methods were set for the same statistical significance of 0.01. The plot on the left (a) presents the identification results for the diagnosis based on marginal correlation, whereas the plot on the right (b) regards the use of the proposed procedure based on partial correlations.

10.4 Conclusions

In the present study several monitoring statistics based on the use of partial correlations were proposed in order to detect changes in the process' structure. These monitoring statistics were applied to systems with different degrees of complexity, including linear, dynamic and non-linear systems and compared with the current statistics for monitoring process multivariate dispersion. Furthermore, several sensitivity enhancing transformations were considered with the goal of improving the methods performance regarding their ability to detect several types of process changes.

In general, the proposed statistics present higher detection sensitivities for the same false alarm rate, especially when only few variables are affected by the faults, making more difficult their detection. It was also showed that the use of sensitivity enhancing transformations that break the variables' correlation based on their causal structure is a key element for improving the fault detection capability.

The characteristics of the *RMAX* family of statistics make them suitable to detect changes in the process' structure even when few observations are collected. However, they do not account for changes in the variance and therefore should be complemented with a monitoring statistic that follows this particular feature. This can be done by application of the proposed *VnMAX* statistic. This monitoring statistic is a generalization of the current *VMAX* statistic, and it is able to detect both changes in variance and in the process structure, since model deviations lead to changes in the transformed variables variance. This result is another indication that the sensitivity enhancing transformations increase the sensitivity to detect changes in correlation, through *RMAX*, and simultaneously assess the model validity, through *VnMAX*. These two monitoring statistics (*RMAX* and *VnMAX*) can be easily combined into a single monitoring scheme, able to detect both types of deviations and therefore consolidate the detection of structural changes through a unified monitoring scheme.

The proposed monitoring statistics also proved to be statistically superior when compared to the current monitoring statistics tested, which makes this approach a feasible solution worthwhile considering in practical applications. The development of similar monitoring schemes for on-line monitoring will be considered in Chapter 11.

11 On-line Monitoring of the Process Correlation Structure

As mentioned in the previous chapter, the few approaches for monitoring multivariate process' variability currently available are based on subgroups and usually employ either the generalized variance (the determinant of the covariance matrix) (e.g. [70-72]) or likelihood ratio tests (e.g. [69, 75, 76]). Other related approaches based on the conditional entropy [73], vector variance (which is the sum of the squares of all eigenvalues of the sample covariance matrix) [74] and independent components obtained from the decomposition of the sample covariance matrix [78] have also been proposed. As these methodologies are strictly based on the marginal covariance, they have little resolution to detect small localized deviations in the intrinsic relationships between process variables. Their on-line versions (i.e., based on individual observations) are also subjected to the same constrains, since they are also based on estimates of the marginal covariance.

In the case of on-line monitoring, the sample covariance matrix can be estimated by either employing a moving windows approach or an EWMA recursion scheme as proposed by Yeh *et al.* (2005) [79] and Huwang *et al.* (2007) [80]. The latter approach is more commonly used and is the foundation of several multivariate monitoring schemes that simple analyze the trace of the covariance matrix [80], squared deviations from target [79, 87] or their likelihood ratio statistic [82].

More sophisticated monitoring schemes were proposed by Reynolds and Cho (2006) [81], who monitors the Mahalanobis distance of the EWMA recursion applied to the squared deviations from target and by Bodnar *et al.* (2009) [88], where the estimate of the covariance matrix is decomposed into a set of vectors that can then be monitored by typical multivariate control charts for process location.

Nevertheless, all the monitoring statistics mentioned above only use the information conveyed by the marginal covariance matrix and do not take into consideration the inner connectivity structure of the process. For the case of off-line monitoring, addressed in Chapter 10, it has been demonstrated the usefulness of partial correlations in the detection of structural changes, namely through a control chart that takes the maximum partial

correlation, in absolute value, as a monitoring statistic (*RMAX* statistic), properly complemented with the monitoring statistics for the variance (*VnMAX* statistic). In this chapter, the application scope of the *RMAX* and *VnMAX* statistics is extended to on-line monitoring through the use of an EWMA recursion scheme to update the estimate of the covariance matrix using each new multivariate observation collected from the process.

11.1 The *RMAX* and *VnMAX* Statistics for On-line Monitoring

The *RMAX* and *VnMAX* statistics were first introduced in Chapter 10 in the context of off-line monitoring of the process correlation structure. It was demonstrated that both *RMAX* and *VnMAX* presented good monitoring characteristics, which makes them natural candidates to consider in the development of improved monitoring methodologies for on-line use. Among these characteristics, the higher detection performance and lower dependency on the number of observations used in the estimation of the sample covariance matrix, should be highlighted. The latter one is perhaps the most important when an on-line monitoring scheme is considered, since the requirement of a large amount of observations to compute the sample covariance also means higher delays in fault detection. The number of observations also plays a major role in the normalization functions that transform the observed partial correlations and variances into approximate *i.i.d.* $N(0,1)$ distributions.

Although a moving windows approach would be the natural extension of the proposed methodology, an on-line estimation of the covariance matrix through an EWMA recursive updating scheme is preferable, since it does not require the storage of a set of n past observations and its implementation is also considerably faster. Moreover, the EWMA recursion scheme is extensively applied to estimate the marginal covariance, as can be seen from the literature review presented in Section 4.2. Therefore, in the proposed on-line procedures, the sample covariance will be estimated by an EWMA recursion procedure. One of such recursion was proposed by Yeh *et al.* (2005) [79] and is given by,

$$\mathbf{S}_t = \lambda \mathbf{x}_t \mathbf{x}_t^{\mathrm{T}} + (1-\lambda)\mathbf{S}_{t-1} \tag{11.1}$$

where $0 < \lambda < 1$ is a forgetting factor and it is assumed that the mean is constant and equal to zero. From this definition, it can be shown that $\mathbf{S}_t$ is a positive definite matrix when

$t \geq m$ [79]. Alternatively, the approach of Huwang *et al.* (2007) [80] can be employed. In this case, the process mean is assumed to vary over time and therefore an estimation of its level should also be considered, resulting in an estimation of the covariance matrix given by,

$$\mathbf{V}_t = \lambda\left(\mathbf{x}_t - \mathbf{y}_t\right)\left(\mathbf{x}_t - \mathbf{y}_t\right)^{\mathrm{T}} + \left(1-\lambda\right)\mathbf{V}_{t-1} \tag{11.2}$$

$$\mathbf{y}_t = \omega\mathbf{x}_t + \left(1-\omega\right)\mathbf{y}_{t-1} \tag{11.3}$$

where $0 < \lambda < 1$, $\mathbf{V}_0 = \left(\mathbf{z}_1 - \mathbf{y}_1\right)\left(\mathbf{z}_1 - \mathbf{y}_1\right)^{\mathrm{T}}$, $0 < \omega < 1$ and $\mathbf{y}_0 = \mathbf{0}$. For $t \geq m$, $\mathbf{V}_t$ is a positive definite matrix and $E\left(\mathbf{V}_t\right) \to 2(1-\omega)^2/(2-\omega)\boldsymbol{\Sigma}$ as $t \to \infty$. Therefore $(2-\omega)/[2(1-\omega)^2]\mathbf{V}_t$ can be used to estimate the covariance matrix [80].

Even though an EWMA recursion scheme is adopted in this study, it is worth noticing that an equivalence relationship was found and developed for the *RMAX* and *VnMAX* statistics between a moving window of length *n* (either non-overlapping windows or moving windows) and an EWMA recursion scheme with a forgetting factor λ. This equivalence relationship will be described in detail in Chapter 12. From this relation, it becomes possible to establish a conceptual scheme that allows the transfer of most of the monitoring statistics properties between the off-line and on-line monitoring scenarios. Moreover, a new normalization function can be derived for the partial correlations and variances as follows,

$$w_r = \frac{r}{\sqrt{\lambda/2}} \tag{11.4}$$

$$w_s = \frac{\left(\frac{s^2}{\sigma_0^2}\right)^{1/3} - \left(1 - \frac{2}{9}\frac{\lambda}{2-\lambda}\right)}{\sqrt{\frac{2}{9}\frac{\lambda}{2-\lambda}}} \tag{11.5}$$

where *r* is the sample partial correlation coefficient, s^2 represent the samples variance, σ_0^2 is the in-control variance and λ is the forgetting factor of the EWMA recursion used to estimate the correlation matrix. By application of Equations (11.4) and (11.5) the partial

correlations and variances become approximately distributed as $N(0,1)$ and can then be monitored by,

$$R0MAX = \left\| w_r \left(\mathbf{r}_0 \right) \right\|_\infty = \max \left\{ \left| w_r \left(\mathbf{r}_0 \right) \right| \right\} \tag{11.6}$$

$$R1MAX = \left\| w_r \left(\mathbf{r}_1 \right) \right\|_\infty = \max \left\{ \left| w_r \left(\mathbf{r}_1 \right) \right| \right\} \tag{11.7}$$

$$VnMAX = \left\| w_s \left(\mathbf{v} \right) \right\|_\infty = \max \left\{ \left| w_s \left(\mathbf{v} \right) \right| \right\} \tag{11.8}$$

where $\mathbf{r}_0$ is the $(m(m-1)/2)\times 1$ column vector of correlation coefficients, $\mathbf{r}_1$ is the $(m(m-1)(m-2)/2)\times 1$ column vector of 1st order partial correlation coefficients and $\mathbf{v}$ is a $(m\times 1)$ column vector containing the variables' variance.

As two different EWMA recursions can be used ($\mathbf{S}_t$, from Equation (11.1) and $\mathbf{V}_t$ from Equation (11.2)), as well as four different sets of variables, depending on the SET presented in Chapter 9 (original variables, T_Z; transformed variables based on the process causal network using a stationary model, T_{NetLin}; transformed variables based on the process network using a dynamic model, T_{NetDyn}; and transformed variables based on the process network using a dynamic model with an additional Cholesky decomposition of the residuals, $T_{NetDynCh}$), twenty four monitoring statistics are obtained. Moreover, the effects of the SET is also assessed when applied in combination with the M_1Z^2 monitoring procedure proposed by Reynolds and Cho (2006) [81]. Thus, a total of twenty seven new monitoring schemes were studied in this work. A summary of these monitoring statistics is given in Table 11.1.

Table 11.1 On-line monitoring statistics considered in this study.

Base statistic	EWMA recursion	Sensitivity enhancing transformation			
		$\mathbf{T_Z}$	$\mathbf{T_{NetLin}}$	$\mathbf{T_{NetDyn}}$	$\mathbf{T_{NetDynCh}}$
M_1Z^2	–	–	$M_1Y^2_{NetLin}$	$M_1Y^2_{NetDyn}$	$M_1Y^2_{NetDynCh}$
$VnMAX$	$\mathbf{S}_t$	$VnMAX_{S,Z}$	$VnMAX_{S,NetLin}$	$VnMAX_{S,NetDyn}$	$VnMAX_{S,NetDynCh}$
	$\mathbf{V}_t$	$VnMAX_{V,Z}$	$VnMAX_{V,NetLin}$	$VnMAX_{V,NetDyn}$	$VnMAX_{V,NetDynCh}$
$R0MAX$	$\mathbf{S}_t$	$R0MAX_{S,Z}$	$R0MAX_{S,NetLin}$	$R0MAX_{S,NetDyn}$	$R0MAX_{S,NetDynCh}$
	$\mathbf{V}_t$	$R0MAX_{V,Z}$	$R0MAX_{V,NetLin}$	$R0MAX_{V,NetDyn}$	$R0MAX_{V,NetDynCh}$
$R1MAX$	$\mathbf{S}_t$	$R1MAX_{S,Z}$	$R1MAX_{S,NetLin}$	$R1MAX_{S,NetDyn}$	$R1MAX_{S,NetDynCh}$
	$\mathbf{V}_t$	$R1MAX_{V,Z}$	$R1MAX_{V,NetLin}$	$R1MAX_{V,NetDyn}$	$R1MAX_{V,NetDynCh}$

11.2 An Extensive Comparative Assessment of On-line Monitoring Methodologies for the Process Correlation Structure

In this section, the proposed monitoring statistics are compared against their current on-line counterparts described in Section 4.2, based on the evaluation of their average run length (ARL), i.e., the average number of samples between the fault's start and is detection. Several case studies were considered, representing different degrees of complexity. Since it is expected that the ARL decreases monotonically with the increase of the faults magnitude, a performance index based on the area under the ARL curve (i.e., the integral of the curve ARL versus fault's magnitude) is considered as a suitable measurement to summarize the relative performance of the studied monitoring statistics. This index in normalized so that it falls in the range of [0 1] where 1 represents the best performance observed.

In each perturbation studied in the following examples, the ARL were determined based on 3000 simulations, leading to 3000 run lengths. For the MCUSUM based statistics, the reference parameter k was set as [0.34 0.68 1.02 1.36 1.70] and for the MEWMA based statistics the forgetting factor λ was [0.01 0.02 0.03 0.04 0.05] (for the EWMA recursion with correction of the mean value, ω was set to 0.20, as this value was the original value recommended by the authors and also used in other published studies). This range of values allows the determination of which of them is more suitable to detect faults with a given monitoring statistic. The final performance comparison is conducted with the parameter that led to the best results in each monitoring statistic.

The control limits for all the monitoring statistics were set by trial and error adjustment based on a NOC data set, to an in-control ARL_0 of 370. This approach to determine the control limits was taken because the monitoring statistics often violate their underlying assumptions, and therefore the theoretical limits usually do not correspond to the observed ones.

11.2.1 Artificial Network

To assess the performance of the proposed statistics, they were applied to adaptations of the artificial network presented by Tamada *et al.* (2003) [134]. From Chapter 10, recall

that this network is composed by 16 variables related according to the representation shown in Figure 10.3. The mathematical expressions describing the systems' variability were set in order to obtain three distinct scenarios with different degrees of complexity (stationary linear, dynamic linear and stationary non-linear). These systems, along with the main results, are fully described in the following sections.

11.2.1.1 Stationary Linear System

In this section, a stationary linear version of the base system is considered, as previously described in Section 10.2.2.1. For this case study, only the sensitivity enhancing transformation T_{NetDyn} was used, since the system is already linear and the obtained transformed variables are completely uncorrelated without the need of an additional Cholesky decomposition. This SET was determined based on NOC data and the real causal network, since it was found that it can be easily recovered from the data by application of the algorithms presented in Chapter 9. The system was then subjected to the set of perturbation described on Table 11.2, where δ represents a multiplicative factor that introduces a change on the original relationship. This factor was varied in the range of 0.80 to 1.20 for all faults. The results obtained are summarized in the performance comparison index N represented in Figure 11.1. The permutation tests between each pair of monitoring statistics studied in this case study are displayed in Table S7 of the supplementary material.

Table 11.2 Definition of the faults and respective variables involved in the stationary linear system. δ is a multiplicative factor that changes the model parameters (under NOC, $\delta = 1$).

Fault	Variables relation changed
A	$g_8 \rightarrow g_1 \ (g_1 = 1.2\delta g_8 + 0.80 g_9 + \varepsilon_1)$
B	$g_1 \rightarrow g_3 \ (g_3 = 0.05 + 0.22\delta g_1 + \varepsilon_3)$
C	$g_8 \rightarrow g_{10} \ (g_{10} = 1 + 0.40\delta g_8 + \varepsilon_{10})$
D	$g_3 \rightarrow g_{14} \ (g_{14} = 1 + 0.40\delta g_3 + \varepsilon_{14})$

Among the current monitoring statistics, the $MEWMA_t$ and M_1Z^2 were the ones that presented the best performance, having in general a similar behavior as can be seen in Figure 12.2. For most faults, M_1Z^2 outperforms all the other monitoring statistics.

However, in fault A, M_1Z^2 was unable to detect the decreases in correlation in an efficient way. This particular fault occurs close to the root node, and most of the variables are affected by it, presenting both changes in correlation and variance. Furthermore, the perturbations induced similar deviations in the variance, in both increase and decrease of magnitude. Therefore, as M_1Z^2 is primarily looking for changes in variance, through their Mahalanobis distance, it should react in a symmetric way in both directions. Yet, this was not the observed behavior, indicating that the performance of M_1Z^2 is dependent on the fault's location and direction (namely decreases in variance, as reported by Reynolds and Cho (2006) [81]), and therefore unsuitable for some types of faults. A more evident case were the M_1Z^2 statistic also fails in the detection task, is when a structural change occurs without any change in the overall variables' variance, as will be exemplified later.

On the other hand, $MEWMA_t$ presented a consistent performance, regardless of the fault location or direction. The performance of the $MEWMA_t$ statistics is rapidly followed by the *RMAX* proposed monitoring statistics, namely the ones that use transformed variables (for instance, $R0MAX_{S,NetDyn}$). Special notice is given for fault A, where $R0MAX_{S,NetDyn}$ presented a better performance than the other monitoring statistics. Again, in this case, there are multiple variables affected by the fault, and therefore multiple correlation coefficients experience a substantial change, which increases the probability of $R0MAX_{S,NetDyn}$ to signal a fault. Regarding the proposed monitoring statistics, it is also worth pointing out the poor performance of $VnMAX_{S,NetDyn}$ and $M_1Y^2_{NetDyn}$. Since these monitoring statistics are directly related to the transformed variables' variance, their low capability to detect faults indicate that the transformed variables have small changes at the variance level, which supports the idea that the fault is occurring on the process' structure and that a simple replacement of variables used on the current schemes (explicitly M_1Z^2) is not enough. The performance of the proposed monitoring statistics was also similar when either the 0^{th} order correlation or 1^{st} order partial correlations were used. This feature results from the application of the sensitivity enhancing transformation, which in this case, uses the process structure to obtain uncorrelated variables. Therefore, as the process structure is considered to construct linear regressions, the 0^{th} order correlation of the transformed variables can, in some extent, be considered as partial correlations of the original variables, which justifies the similar behavior found for 0^{th} order correlation or 1^{st} order partial correlations.

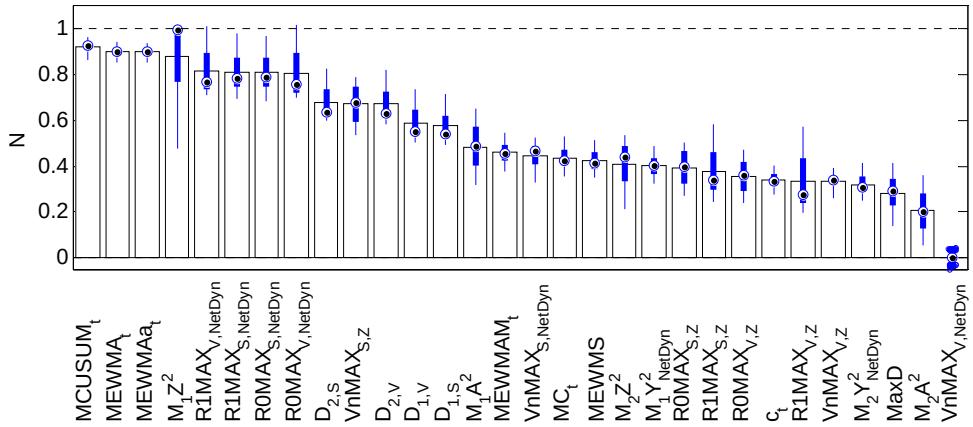


Figure 11.1 Comparison of the statistics performance on the stationary linear system: box-plots of the performance index *N* based on the area under the ARL curve obtained on all perturbation, superimposed to the bars with heights corresponding to the associated mean values.

Another important observation is the lower performance of monitoring statistics that actually consider the full marginal covariance matrix (*MaxD* and c_t) and uncorrelated variables (*MEWMS*, M_1A^2 and M_2A^2) to detect faults. These results highlight the need for a proper variable transformation and the benefits of considering partial correlations to measure process structural changes.

To better exemplify the robustness of the proposed monitoring statistics, an additional fault was studied, by creating a perturbation on the relationship between variables g_1, g_8 and g_9 in such a way, that any change on the relation $g_8 \rightarrow g_1$ is compensated by $g_9 \rightarrow g_1$ in order to maintain the variance of g_1 close to its target value (by defining $g_1 = 1.2\delta g_8 + 0.80\sqrt{1+2.25\left(1-\delta^2\right)}g_9 + \varepsilon_1$). The results for these perturbations are presented in Figure 11.2. From the analysis of Figure 11.2 it becomes clear that M_1Z^2 is unable to detect this type of fault. In fact, the monitoring statistics that presents the best performance is M_1A^2, which uses regressed-adjusted variables to obtain uncorrelated variables. Yet, this transformation cannot efficiently detect the other faults studied previously. The same happens with the proposed transformation in the $M_1Y^2_{NetDyn}$ statistic. Consequently, none of the M_1Z^2 based procedures can detect the overall variety of faults related with structural changes. The only monitoring statistics that maintained their capability to detect this type of fault were $MEWMA_t$ and the proposed $R0MAX_{S,NetDyn}$ statistics, the latter one presenting a clearly more rapid response.

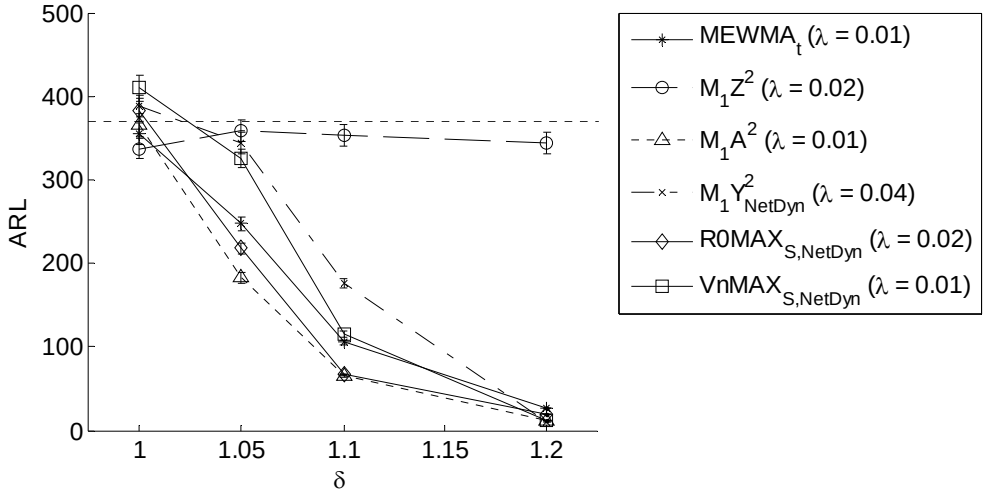


Figure 11.2 ARL of the most relevant monitoring statistics for a change on the relationship between g_1, g_8 and g_9 while maintaining the variance of g_1 close to its target value. δ represents the magnitude of the change caused on the variables relationships.

These results indicate that, although the *RMAX* based statistics have a slightly lower performance, their detection capability remain consistent and symmetric to both increases and decreases throughout multiple faults with different characteristics. Furthermore, *RMAX* is not subjected to mathematical and numerical instabilities that may arise on the other monitoring statistics, due to, for instance, the inversion of the covariance matrix, as happens on the M_1Z^2 and $MEWMA_t$ statistics.

11.2.1.2 Dynamic Linear System

As most systems present dynamic features, the second modification of the base network relationships includes time dependency by addition of moving average (MA) terms between the variables. These modifications gave rise to the mathematical expressions already presented in Section 10.2.2.2. Uncorrelated variables were obtained after application of the sensitivity enhancing transformations T_{NetLin}, T_{NetDyn} and $T_{NetDynCh}$. These SET were determined based on the full causal network and a data set of NOC observations, with up to three lags for the case of dynamic transformations. This system was subjected to the faults indicated on Table 11.3 for changes in the multiplicative factor δ on the range of 0.80 to 1.20. The subsequent results are summarized in Figure 11.3, where the distribution of the performance index N is depicted, while the permutation test results are presented in Table S8 of the supplementary material.

Table 11.3 Definition of the faults and respective variables involved in the network system with linear dynamics. δ is a multiplicative factor that changes the model parameters (under NOC, $\delta = 1$).

Fault	Variables relation changed
A	$g_8 \rightarrow g_1$ $(g_{1,t} = 1.2\delta\left(g_{8,t} + 0.60g_{8,t-1} + 0.30g_{8,t-2}\right) + 0.80g_{9,t} + \varepsilon_{1,t})$
B	$g_1 \rightarrow g_3$ $(g_{3,t} = 0.05 + 0.22\delta\left(g_{1,t} - 0.40g_{1,t-1} - 0.20g_{1,t-2}\right) + \varepsilon_{3,t})$
C	$g_8 \rightarrow g_{10}$ $(g_{10,t} = 1 + 0.40\delta\left(g_{8,t} + 0.60g_{8,t-1} - 0.30g_{8,t-2}\right) + \varepsilon_{10,t})$
D	$g_3 \rightarrow g_{14}$ $(g_{14,t} = 1 + 0.40\delta\left(g_{3,t} + 0.40g_{3,t-1} + 0.60g_{3,t-2}\right) + \varepsilon_{14,t})$

Regarding the current monitoring statistics, M_1Z^2, $MEWMA_t$ and $D_{2,S}$ are the ones presenting the best performance through all the simulated faults in this case study. The proposed $RMAX_{NetDyn}$ and $RMAX_{NetDynCh}$ statistics only outperform the current ones on fault B, showing a slightly worse performance on the remaining faults. The performance of these monitoring statistics is generally followed by the use of *RMAX* with the stationary linear transformation ($RMAX_{NetLin}$) and only afterwards by the remains schemes (both current and proposed). These performances ranking suggests that the variables transformation can indeed enhance fault detection. However, they require a new procedure to monitor the correlation, since the combination of transformed variables and M_1Z^2 (applied on the monitoring statistic $M_1Y^2_{NetDynCh}$) was not particularly suitable for such task, since it even lowers the detection power of the M_1Z^2 scheme (see Figure 11.3). This is however an expected result, since the simulated faults are related with the process correlation, which is not captured by this monitoring statistic. Likewise the *VnMAX* base monitoring statistics also presented a low detection capability, indicating that on the transformed variables, the faults are more visible as structural deviations rather than in deviations from their target value. This type of behavior allows for a better isolation of the root case, since *RMAX* is only affected by changes in the process structure, while M_1Z^2 depends on correlation, variance and mean deviations. Note that $VnMAX_{S,Z}$ also shows a relatively high performance, which sustains the idea that significant changes on the variance of the original variables is also occurring.

The sensitivity enhancing transformations were found to be the most relevant factor contributing to the monitoring statistics performance, with the higher performances obtained with T_{NetDyn} and $T_{NetDynCh}$. These two transformations include time-shifted variables that allow for a better description of the systems and therefore a better detection of structural changes than the monitoring statistics based on the stationary transformation

T_{NetLin}. The use of the variables transformations also lead to similar results when either 0th order or 1st order partial correlations are used, since this transformations produce uncorrelated variables, making the use of 1st order partial correlations redundant for detection purposes.

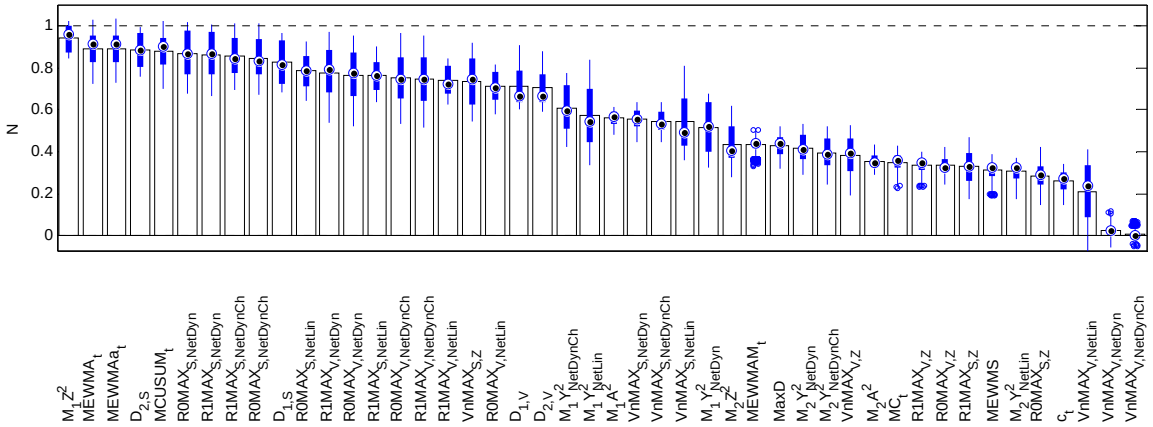


Figure 11.3 Comparison of the statistics performance on the dynamic linear system: box-plots of the performance index *N* based on the area under the ARL curve obtained on all perturbation, superimposed to the bars with heights corresponding to the associated mean values.

11.2.1.3 Stationary Non-linear System

To better represent the original non-linear structure of the studied system, the variables relationships were approximated by polynomials according to the description provided in Section 10.2.2.3. The system was then subjected to the faults presented in Table 11.4, where δ was changed between 0.85 and 1.15 in order to enable a good assessment of the detection capabilities of each monitoring statistic. A resume of the obtained results are presented on Figure 11.4 through the performance index *N* based on the area under the ARL curve. For this case study, the sensitivity enhancing transformations T_{NetLin} and T_{NetDyn} were used to analyze the SET impact in the monitoring schemes. T_{NetLin} is composed only by linear terms, while T_{NetDyn} uses polynomial terms up to the 3rd order.

Table 11.4 Definition of the faults and respective variables involved for the network system with a non-linear model structure. δ is a multiplicative factor that changes the model parameters (under NOC, $\delta = 1$).

Fault	Variables relation changed
A	$g_8 \rightarrow g_1 \ (g_1 = 1.20\delta g_8 + 0.80 g_9 + \varepsilon_1)$
B	$g_1 \rightarrow g_3 \ (g_3 = (\delta g_1 - 4)(g_1 + 4) + \varepsilon_3)$
C	$g_8 \rightarrow g_{10} \ (g_{10} = 0.02 g_8^2 + 0.44\delta g_8 + 0.82 + \varepsilon_{10})$
D	$g_3 \rightarrow g_{15} \ (g_{15} = 1.40\delta g_3 + \varepsilon_{15})$

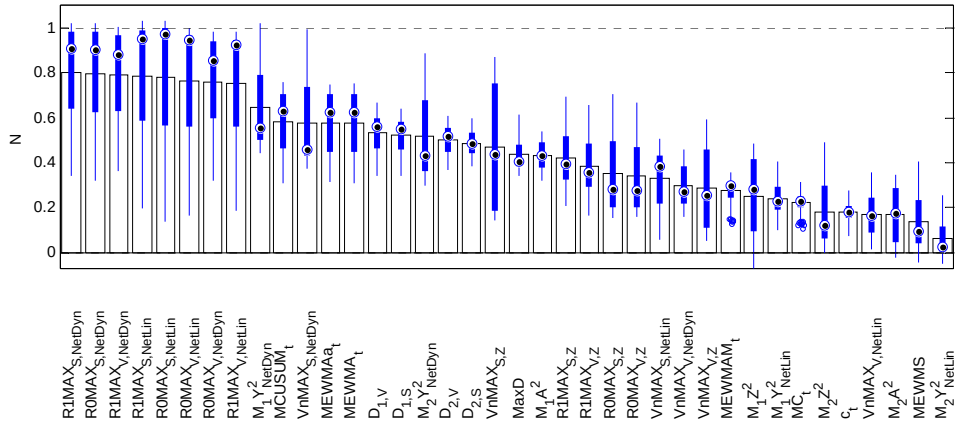


Figure 11.4 Comparison of the statistics performance on the stationary non-linear system: box-plots of the performance index *N* based on the area under the ARL curve obtained on all perturbation, superimposed to the bars with heights corresponding to the associated mean values.

From the analysis of the results it can be observed that $MEWMA_t$, $D_{1,S}$, $D_{2,S}$, $D_{1,V}$ and $D_{2,V}$ were the current monitoring statistics that gave the best results with generally similar detection capabilities. Yet, $MEWMA_t$ remains statistically superior than these monitoring statistics as can be seen from Table S9 of the supplementary material, having the lowest ARL for most of the faults. On the other hand, M_1Z^2, which showed one of the best performances on the previous case studies, was not able to detect some of the faults, namely the ones related with decreases in the faults magnitude. This behavior is mitigated by replacing the original variables by the ones based on variable transformations, namely T_{NetDyn} ($M_1Y^2_{NetDyn}$). The resulting $M_1Y^2_{NetDyn}$ statistic leads to an increase in performance of the M_1Z^2 procedure and showed to be capable to detect changes related with the transformed variables' variance. Nevertheless, $M_1Y^2_{NetDyn}$ is strictly related to the process variance, and when a structural change that does not affect the variance occurs, its performance is diminished. Furthermore, it is rapidly followed by $VnMAX_{NetDyn}$, which has a closer relationship with the *RMAX* scheme, making it a more favorable choice since it allows for their easy combination into a single combined monitoring statistic.

The current monitoring statistics also presented better results when the faults happen closer to the root node. These types of faults propagate to most of the process variables and are therefore easier to detect, since a larger number of variables are affected and leave their normal operation values. An example of such case is fault B on Figure 11.5, which leads to lower ARL for $MEWMA_t$ and for most of the others. However, when a fault related with a leaf node (e.g. fault C) occurs, the current monitoring statistics tend to

perform poorly, while $R0MAX_{NetDyn}$ maintains the capability to perform rapid detections (see Figure 11.6). The good performance of $R0MAX_{NetDyn}$ is not restricted to this particular case, being also observed on Faults A and B. On fault D, the detection is mostly done by $VnMAX_{NetDyn}$, which highlight the complementariness between these two monitoring statistics. The performance of *RMAX* and *VnMAX* based on a linear transformation (T_{NetLin}) also showed to be capable to detect most of the faults and presented a similar, yet generally worst, performance than the non-linear transformation T_{NetDyn}. This result, along with the $M_1Y^2_{NetDyn}$ behavior mentioned earlier, proves that the obtained performances are not simple a result of using a more complex non-linear transformation but are indeed related with the proposed monitoring procedure. This remark can be further confirmed by the results of the permutation test presented in Table S9 of the supplementary material, where it can be verified that the *RMAX* and *VnMAX* procedures are statistically superior to the others, regardless of the variables transformation used.

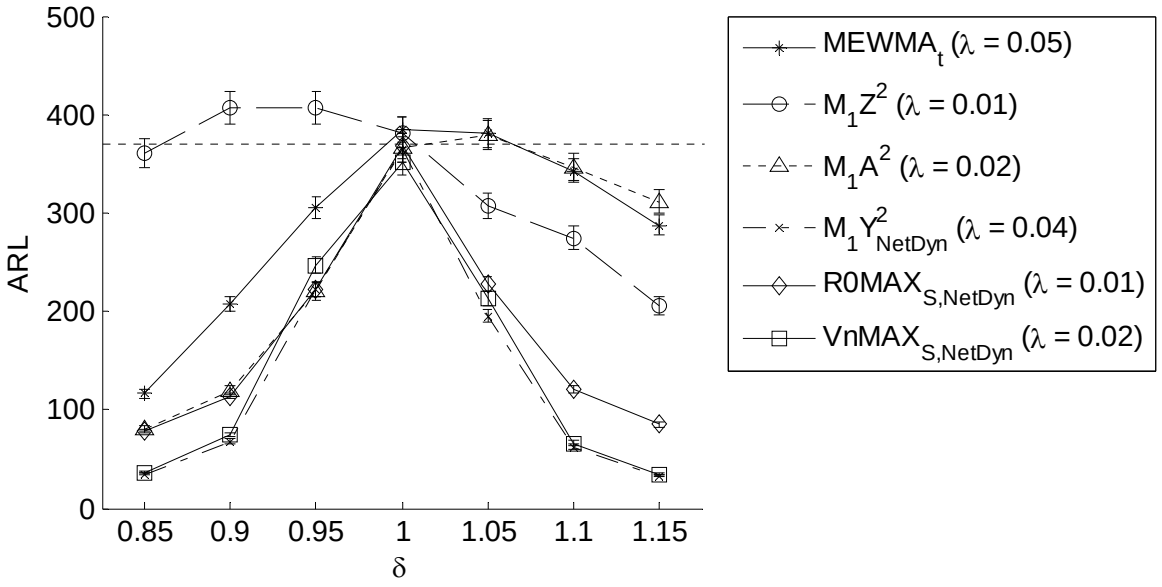


Figure 11.5 ARL of the most relevant monitoring statistics on the stationary non-linear system for fault B. δ represents the magnitude of the change caused on the variables relationships.

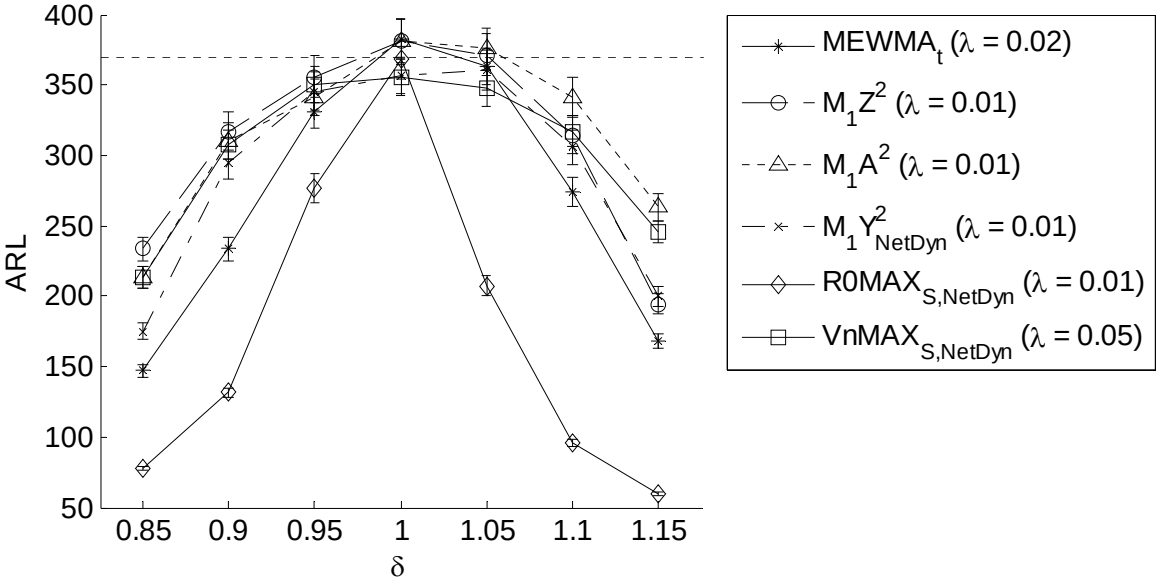


Figure 11.6 ARL of the most relevant monitoring statistics on the stationary non-linear system for fault C. δ represents the magnitude of the change caused on the variables relationships.

11.2.2 Dynamic Non-linear Systems

In this section the performance of the monitoring statistics are assessed on dynamic non-linear systems. These systems represent more realistic processes with added complexity on their inner causal network. For these case studies, the underlying process structure is assumed to be unknown and therefore it has to be estimated through the network reconstruction algorithms presented in Chapter 9. The problem of reconstruct the networks of these systems was already addressed in Section 10.2.3, when they were used to compare the off-line monitoring statistics. Thus, only brief descriptions will be given here. In all of the following case studies, the sensitivity enhancing transformations T_{NetLin}, T_{NetDyn} and $T_{NetDynCh}$ were based on such estimated causal networks and determined by application of regression models using a data set of NOC observations.

11.2.2.1 Gene Network Model

The first of the dynamic systems treated in this section is the one proposed by Fuente *et al.* (2004) [130]. A full description of the systems dynamics and causal network is provided in Section 10.2.3.1. Based on this previous study, the system was subjected to faults on V_1, K_1 and K_2 on a range between [0.985 0.995]. From these faults, the ARL

were determined from where the performance index N was computed, given raise to the results summarized in Figure 11.7. The results of the permutations tests are given in Table S10 of the supplementary material.

The studied faults produce mostly deviations on the variables mean values, as a result of a new steady state produced by changes on the process parameters. To a lesser extent, some changes were also observed in the variance with generally small changes on the correlation coefficients. Consequently, most of the detections are due to changes in the mean values of the variables. An evidence of this situation is clearly shown by the low detection performance of the monitoring statistics based on the EWMA recursion with correction of the mean (e.g. $R1MAX_{V,NetDynCh}$). For changes in the mean this may be an interesting feature, since it allows for the rapid differentiation between structural and mean changes. Moreover, in this particular system, the structural changes are detected by *VnMAX*, which captures model's mismatches between the sensitivity enhancing transformation and the observed values. One of these statistics is $VnMAX_{V,NetDynCh}$. Even though it presents a relatively lower performance when compared to the other monitoring statistics, its detections are less affected by the mean change and are more related with changes in the process structure or variables' variance.

For the cases were the mean's effect is not corrected (i.e., the monitoring statistics based on estimation $\mathbf{S}_t$), the statistics based on transformations T_{NetDyn} and $T_{NetDynCh}$ also present low detection capabilities. This happens because these transformations describe more accurately the system dynamics, leading to better predictions of the new values, and therefore produces variables with smaller deviations on the mean. On the other hand, transformation T_{NetLin}, which does not accounts for the dynamic behavior of the data, maintained mean deviations similar to the ones of the original variables and presented the greatest performances through the monitoring statistic $M_1Y^2_{NetLin}$, followed by $VnMAX_{S,NetLin}$. The monitoring statistics $VnMAX_{S,Z}$ (based on the original variables) also presented one of the best performances. Note, that the *VnMAX* based statistics are directly linked to the variance and, if not corrected in the EWMA recursion scheme, also with the process' mean. Therefore, if a fault occurs in one of these parameters, they are easily detected by the *VnMAX* statistics. Regarding the current monitoring statistics, M_1Z^2 is the one presenting the best performance. However, as it measures the squared deviations from the target value, its detection capability is also affected by deviations in the mean values.

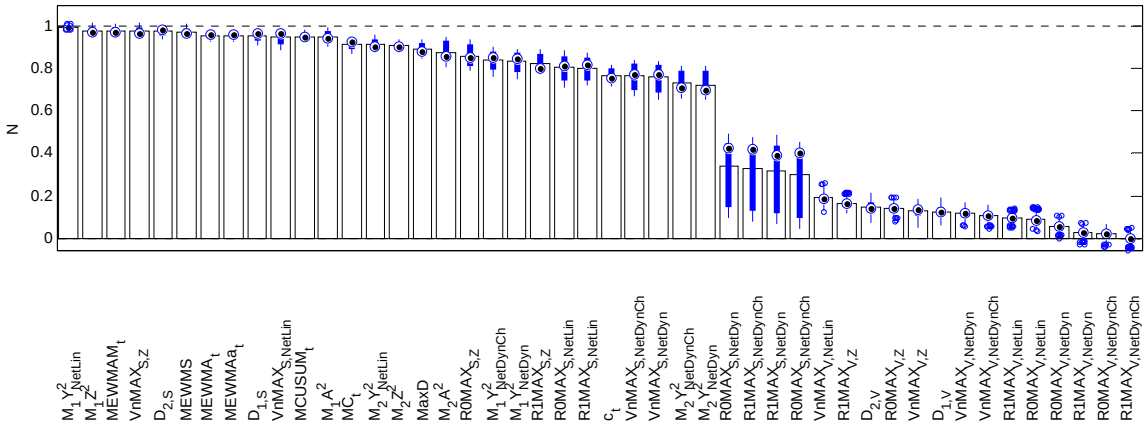


Figure 11.7 Comparison of the statistics performance on the gene network model: box-plots of the performance index *N* based on the area under the ARL curve obtained on all perturbation, superimposed to the bars with heights corresponding to the associated mean values.

These results suggest that even without the mean correction, the proposed procedures are less sensitive to these deviations, focusing their detection on structural changes, which in this case are much smaller than the mean deviations. Furthermore, the detection of faults due to changes on the mean value arises from a violation of the assumption that the mean is constant during the estimation of the covariance matrix by the EWMA recursion. The effect of the mean is also the reason why the on-line monitoring statistics are detecting faults with smaller magnitude than the ones on the off-line scenario.

11.2.2.2 Biologic Production of Ethanol

The second dynamic system is based on the ethanol production from glucose fermentation by *Zymomonas mobilis* bacteria. The model used in this case study, was proposed by Jöbses *et al.* (1987) [142] as described in Section 10.2.3.2. As in the previous off-line study, only C_S, C_X and C_P were considered as measured variables and faults were introduced on c_1, K_S and Y_{SX}. The faults magnitude were set in the range of [0.994 1.006] for c_1, [0.94 1.06] for K_S and [0.985 1.015] for Y_{SX}. The main results are presented in Figure 11.8 in the form of the performance index *N*. As in the previous case studies, the permutation tests are represented in Table S11 of the supplementary material.

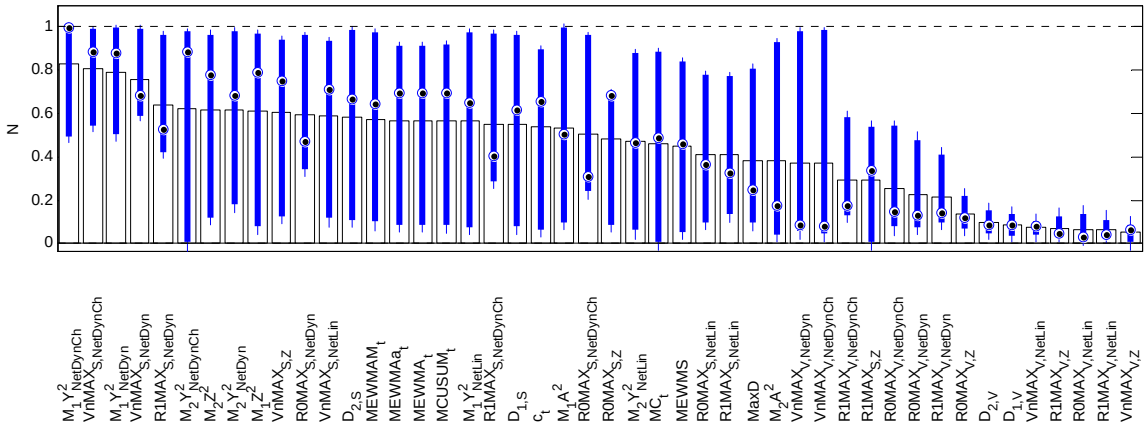


Figure 11.8 Comparison of the statistics performance on the Jöbses' model: box-plots of the performance index *N* based on the area under the ARL curve obtained on all perturbation, superimposed to the bars with heights corresponding to the associated mean values.

From Figure 11.8 it can be seen that most of the monitoring statistics have a fairly wide performance distribution. Yet, it also shows that the monitoring statistics based in $T_{NetDynCh}$ and T_{NetDyn}, namely $VnMAX_{S,NetDynCh}$ and $R1MAX_{S,NetDyn}$, are more consistent. In fact, they were among the few monitoring statistics capable to detect faults on K_S as can be seen in Figure 11.9 and immediately detect small faults on c_1 as represented in Figure 11.10. Nevertheless, these rapid detections are mostly an effect of change in the mean value, which is an undesired side effect. The *VnMAX* statistics are more subjected to this effect, since they are based on the variables dispersion that increases with the square of the mean deviation. On the other hand, *RMAX* is insensitive to mean changes as long as only one variable is affected. This phenomena is better explain for the case of moving windows, where the sample covariance would be estimated by,

$$\begin{aligned} c &= \frac{1}{n}\sum_{i=1}^{n}\left(x_i+\varepsilon_i\right)\left(y_i+\varphi_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(x_i y_i + x_i\varphi_i + \varepsilon_i y_i + \varepsilon_i\varphi_i\right) \end{aligned} \tag{11.9}$$

where ε and φ are deviations from the mean, which is assumed to be known and equal to zero. Assuming that the deviations are independent, for step deviations the expected value of c is given by,

$$E(c) = \operatorname{cov}(x,y) + E(\varepsilon)E(\varphi) \tag{11.10}$$

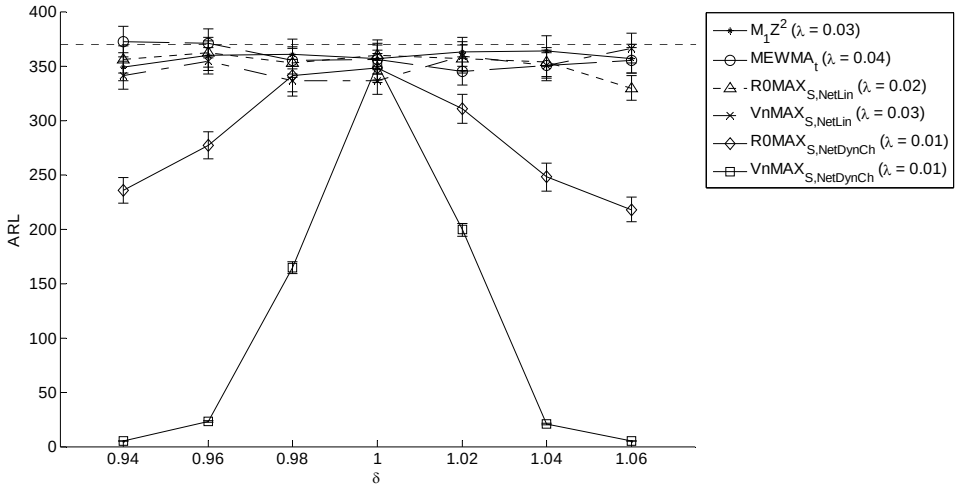


Figure 11.9 ARL of the most relevant monitoring statistics for step deviations of magnitude δ time the K_S parameter of the Jöbses' model.

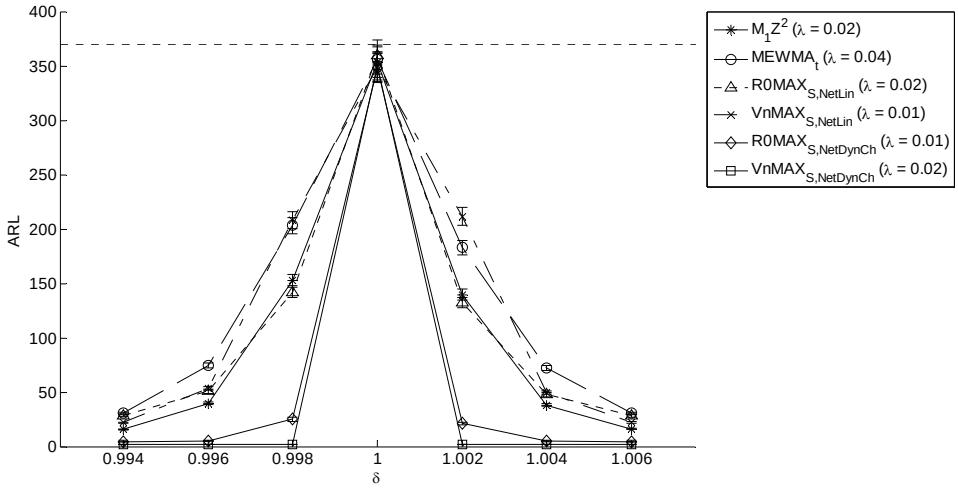


Figure 11.10 ARL of the most relevant monitoring statistics for step deviations of magnitude δ time the c_1 parameter of the Jöbses' model.

Thus, if only one deviation on the mean is observed, the expected value of the covariance does not change. However, the variation in each variable does increase, implying an effective decrease in the correlation coefficients obtained from the data. A similar behavior can also be observed for other types of faults.

By analyzing the raw data, even for faults of larger magnitudes than the ones studied here, the major observable deviation occurs in the mean value of only one variable. In these cases, the *RMAX* statistics are, by design, robust to such deviations, causing them to

react only to changes in the correlation. However, as the remaining monitoring statistics also detect changes on the mean, they tend to signal such fault earlier and therefore they present smaller ARL.

Regarding the monitoring statistics with mean correction, it is clear that the *RMAX* and *VnMAX* with dynamic transformations (either T_{NetDyn} or $T_{NetDynCh}$) performed better than $D_{1,V}$ and $D_{2,V}$. When a stationary linear transformation is used (T_{NetLin}), they tend to be statistically identical. Moreover, it was observed that for low magnitude faults $D_{1,V}$ and $D_{2,V}$ return to their in-control values after the mean effect is corrected, while the *VnMAX* statistics remain out-of-control as exemplified in Figure 11.11.

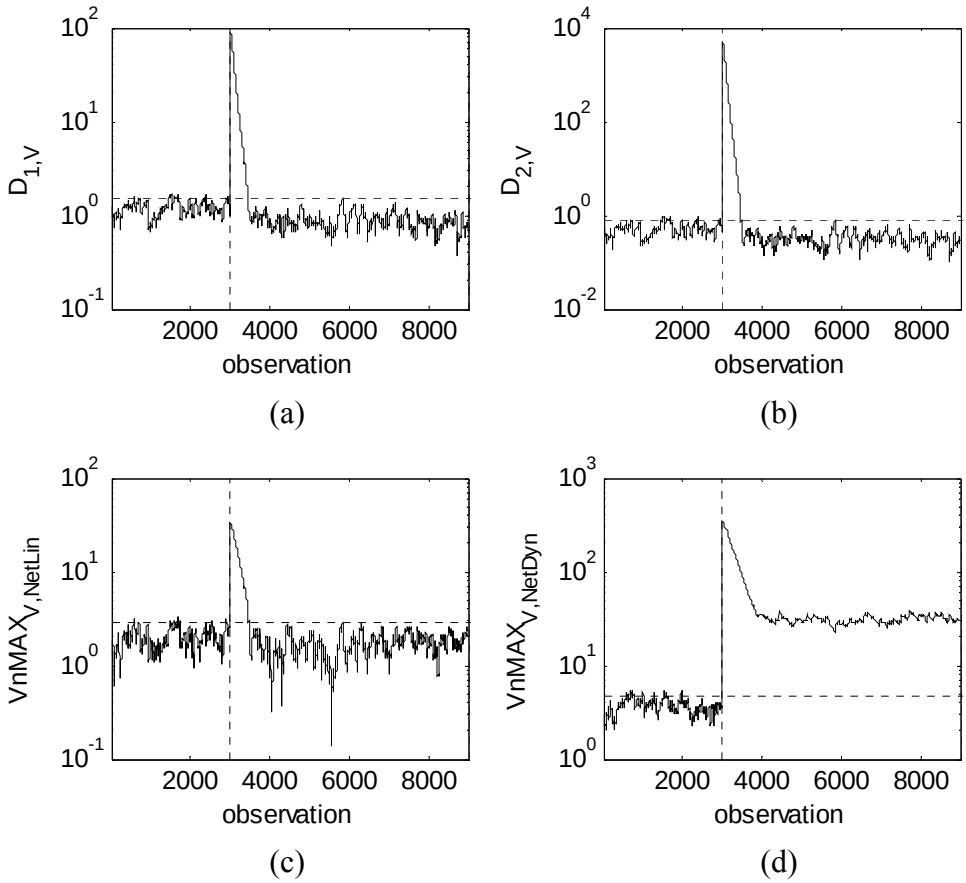


Figure 11.11 Behavior of the monitoring statistics with correction of the mean value (λ = 0.01 and ω = 0.20): (a) current $D_{1,V}$ monitoring statistic; (b) current $D_{2,V}$ monitoring statistic; (c) proposed $VnMAX_{V,NetLin}$; (d) proposed $VnMAX_{V,NetDyn}$. Data obtained from an increase of 20% in c_1 at observation 2000. The fault persists until the end of the simulation.

Although the monitoring statistics with correction of the mean presented the lowest performance, they are the ones that are less sensitive to it. In fact, it is this low dependency to the mean value that originates lower detection capabilities on the simulated faults. Comparing the monitoring statistics with mean correction it is possible to observe than they only signal a fault when the system is subjected to larger perturbations. These results are in line with the ones detected on the off-line study performed in Chapter 10. In general, all the studied methods detect the fault very rapidly due to the mean change which is not immediately corrected. However, after some observations, the mean effect starts to be eliminated and the monitoring statistics return to their normal values. The only exceptions are the $R0MAX_{V,NetDyn}$, $R1MAX_{V,NetDyn}$ and $VnMAX_{V,NetDyn}$ statistics, which consistently signal an out-of-control state.

11.2.2.3 Continuous Stirred-Tank Reactor

The third dynamic non-linear system studied is based on the model of a continuous stirred-tank reactor (CSTR) with a heating jacket and under feedback control. This system represents a more realistic process with added complexity due to close-loops, bidirectional dependencies, unmeasured variables and with faults that can affect multiple variables at once. The main unit of the system is a CSTR with free discharge flow, where an endothermic reaction of the type $A \rightarrow B$ takes place. The system is under PI control in order to maintain the temperature and fluid level close to their set-points. The process structure behind this system was already presented in Section 10.2.3.3 for both the real (Figure 10.15 (a)) and estimated (Figure 10.15 (b)) causal network. Therefore, it will not be further discussed here. In order to construct the sensitivity enhancing transformations (T_{NetLin}, T_{NetDyn} and $T_{NetDynCh}$), only the estimated causal network was used with up to two lags.

To compare the performance of the monitoring statistics, the system was subjected to faults on the heat transfer coefficient, discharge coefficient, pre-exponential factor and activation energy. The results are presented in Figure 11.12 by the performance index *N*, while the permutation tests between each monitoring statistics are given in Table S12 of the supplementary material.

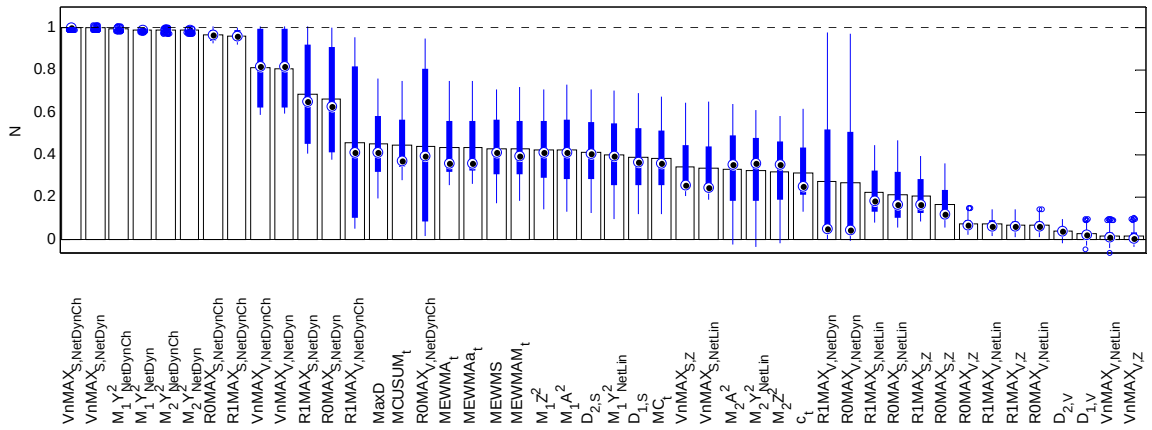


Figure 11.12 Comparison of the statistics performance on the CSTR system: box-plots of the performance index *N* based on the area under the ARL curve obtained on all perturbation, superimposed to the bars with heights corresponding to the associated mean values.

From the analysis of the current monitoring statistics, it is observed that *MaxD* is the one presenting the best performance. Still, the current monitoring statistics have generally similar performances, with the exception of $D_{1,V}$ and $D_{2,V}$. These two statistics are the only ones that account for changes on the mean, reducing their detection capabilities when compared to others, such as $D_{1,S}$ and $D_{2,S}$, which do not correct for the mean effect. This feature is an indication that the current monitoring statistics are detecting both changes in structure and in the mean, which can be easily corroborated through simple inspection of the raw data, where a significant deviation in the mean value can be observed.

A similar behavior is observed with the proposed monitoring statistics, namely those based on $\mathbf{V}_t$, which have worse performances than those based on $\mathbf{S}_t$. However, they depart from the current monitoring schemes by presenting better and more consistent performances in all the studied faults, with abrupt decreases on the ARL_1 to near 1, even when the mean correction is used ($VnMAX_{V,NetDynCh}$). This result is mainly a consequence of the applied transformation $T_{NetDynCh}$ that encompasses the dynamic behavior of the system and the variables connection structure leading to considerable deviations on the mean, variance and correlation when a fault occurs. Even though $VnMAX_{V,NetDynCh}$ is also affected by the mean during the first faulty observations, it is capable to maintain an out-of-control signal after the effect of the mean has been correct. This situation is not observed in other monitoring statistics as illustrated in Figure 11.13. Finally, it is worth

noticing that a SET without dynamic components, i.e., transformation T_{NetLin}, leads to monitoring statistics with a performance in line with the current ones.

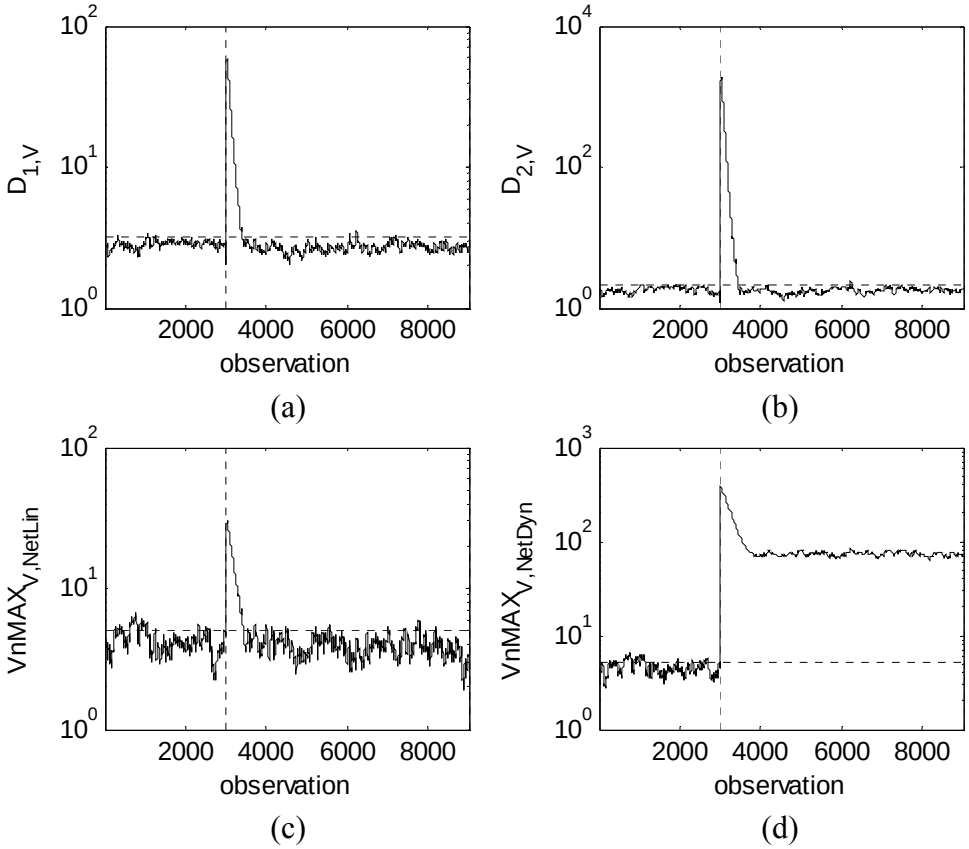


Figure 11.13 Behavior of the monitoring statistics with correction of the mean value (λ = 0.01 and ω = 0.20): (a) current $D_{1,V}$ monitoring statistic; (b) current $D_{2,V}$ monitoring statistic; (c) proposed $VnMAX_{V,NetLin}$; (d) proposed $VnMAX_{V,NetDyn}$. Data obtained from a decrease of 40% in the heat transfer coefficient at observation 2000. The fault persists until the end of the simulation.

11.3 Discussion

The monitoring of structural changes based on individual observations is a task that raises many challenges. First of all, the correlation matrix is not defined in such cases, and therefore a moving window or EWMA recursive estimation of the correlation matrix is required in order to proceed with the monitoring task. Such estimation has the immediate disadvantage of introducing autocorrelation and time delays on the detection procedures. Furthermore, depending on their underlying assumptions, the estimated covariance matrix can also be affected by mean deviations, which may not be related to changes on the process structure.

From the current studied monitoring statistics, only $D_{1,V}$ and $D_{2,V}$ correct the mean effect on the estimation on the covariance matrix. Consequently, these monitoring statistics tend

to have a lower performance, even though they present an initial response to the mean (since it takes some observations until it is properly corrected). Furthermore, for small deviations they usually return to an in-control state after the mean effect is corrected, which may lead to a false sense of normality. These monitoring statistics are also only related with the process variance, and therefore some structural changes may pass undetected which significantly decreases their utility in general.

The M_1Z^2 and $MEWMA_t$ procedures were the ones that presented better results in the case studies analyzed. Regarding the M_1Z^2 procedure, it did not present a consistent behavior in all faults, and is even prone to miss some of them, especially faults that do not cause deviations in the variance or mean. Moreover, the simple replacement of the original variables by transformed ones is not enough to improve its performance (see for instance M_1Z^2 and $M_1Y^2_{NetDyn}$ in Figure 11.1). M_1Z^2 and $MEWMA_t$ also require at some point of their procedure an inversion of a correlation matrix, which can cause computational problems if it is ill-conditioned, as it is often the case with current processes and data acquisition systems. This is particularly more relevant for $MEWMA_t$ which decompose the data into m sets of variables, each of them dependent of the inversion of a reduced form of the covariance matrix. On the other hand, the proposed variables transformations can require heavy computation during the modeling stage. Yet, the pre-selection of the related variables, through network reconstruction techniques, reduces the amount of regressors to a meaningful set of variables that indeed contribute to the model. After this stage, the correlation matrix is used directly without any mathematic manipulation. Therefore, the proposed monitoring schemes are considerable simpler and more stable than the current ones.

The proposed procedure is also consistently ranked as one of the best monitoring statistics, by either *RMAX* or *VnMAX*, regardless of the fault location. This consistency is one of its major advantages, since it performs well on different faults, both in correlation and variance, while the current monitoring schemes fail in some of these situations. Depending on the covariance estimation procedure (with or without mean correction), *RMAX* and *VnMAX* are also sensitive to mean deviations. This feature is clearer on the *VnMAX* statistic since it is directly related with the variables dispersion.

Another major contribution to the increase in the detection abilities is the sensitivity enhancing transformation. The transformations based on the inference of the process'

network allow for the modeling of the variables dependency in a meaningful way and creates a new set of uncorrelated variables where detections of correlation changes are maximized. This type of faults is monitored by *RMAX* and is generally well detected by it. Deviation from the model can also be interpreted as structural changes and are efficiently detected by *VnMAX*. Therefore, the combined used of *RMAX* and *VnMAX* can detect a wide range of faults related with structural changes. The effects of the sensitivity enhancing transformations become even more relevant on dynamic and non-linear systems since the correlation only accounts for the linear relationship between the variables. This situation is visible on the CSTR case study, where a large difference between the performance of $VnMAX_{S,NetDyn}$ (with dynamics) and $VnMAX_{S,NetLin}$ (without dynamics) can be clearly observed.

After application of the sensitivity enhancing transformations, the use of either marginal or 1st order partial correlations seems to have no significant effect on the global performance of the *RMAX* statistics. This is because the marginal correlations of the transformed variables can be interpreted as partial correlations of the original variables, since they are being conditioned on another set of variables (in this particular case, its causal parents), which is the core of the partial correlations concept. However, in cases where the transformed variables are not truly uncorrelated, the use of 1st order partial correlation may be justifiable.

The forgetting factor λ is also an importance aspect on the final performance of the monitoring statistics. Its selection is a compromise between detection speed and certainty, since higher values give more weight to recent observations but also increase the uncertainty on the correlations estimations, since less information is being used. A forgetting factor of 0.01 seems to lead to good results, representing about 200 observations in a moving window procedure. A lower forgetting factor can be used to enhance the sensitivity, with the adverse effect of increasing the detection delay for faults with larger magnitudes.

11.4 Conclusions

In this chapter, the previous proposed monitoring procedures based on partial correlations were extended to the case of process monitoring with individual multivariate observations

(on-line scenario). To do so, EWMA recursive updating schemes for the correlation matrix were employed. This type of approach is also widely employed in other methodologies already proposed in the literature against which the proposed procedures were compared.

The proposed monitoring statistics generally presented a performance similar to the ones obtained with the current approaches when a linear transformation was used. However, when more complex transformations were applied, it was observed an increase in performance. This situation is both due to the sensitivity enhancing transformation and also to the monitoring statistics employed, since the simple integration of the transformed variables on the current procedures does not lead to better results by itself. Furthermore, the current procedures with the best performance, M_1Z^2 and $MEWMA_t$, are dependent of the inversion of the covariance matrix, which may be ill-conditioned. On the other hand, the sensitivity enhancing transformation requires the construction of a system model by performing a linear regression on each variable. Yet, as it pre-selects the variables to be used in the model (with resort to *a prior* knowledge of the process structure or by a network reconstruction methodology) no major problems are expected on the model fitting stage.

The final performance of the proposed *RMAX* and *VnMAX* procedures is also dependent of the forgetting factor used to estimate the covariance matrix. To increase their sensitivity to smaller deviations, a lower forgetting factor is required. However, this implies that more weight is given to past observations, and therefore the ARL also increases. On this regard, a forgetting factor of 0.01 seems to be a good compromise between these two factors. Finally, it is worth noticing that this parameter is related with an equivalent number of observations that would lead to a similar performance, as it will be demonstrated in Chapter 12.

12 Relationship between Off-line and On-line Methodologies for Monitoring the Process Correlation Structure

The first scenario addressed in this thesis regarding the monitoring of the process correlation structure was relative to the off-line case, where non-overlapping moving windows of length n were employed. The natural extension of the schemes proposed for this scenario is based on applying the same principle of moving windows, but now with overlap, i.e., in the form of sliding window of constant length that always contains the last observations. The relation between these two approaches is easy to establish, since they both share the same mechanistic concepts and an observation's window clearly defined. However, in the cases of on-line monitoring, an EWMA recursion is usually applied to estimate the covariance matrix, instead of overlapping moving windows. Examples of the application of EWMA recursion schemes can be found in [79, 80, 82]. Even though the covariance matrix can be efficiently computed from moving windows, namely through the use of updating schemes as the one presented by Wang *et al.* (2005) [143], they still require the storage of all the past n observations, which can be avoided by resorting to an EWMA recursion scheme.

The EWMA updating approach is parameterized by a forgetting factor λ that balances the relative importance of recent observations and those from the past. Analyzing the weighting profile applied by the EWMA scheme to all observations, it is not clear how many observations are indeed significantly contributing to the current estimate of the covariance matrix. However, a relationship between the number of observations and the forgetting factor would be very useful, not only as a way to select it in a more informed way, but also because it would allow to establish a bridge between both approaches to process monitoring: one based on a moving windows and the other on a recursive scheme parameterized by a forgetting factor.

These two parameters (n for the moving window approach and λ for the EWMA recursion scheme) have a major effect not only in the correlation coefficients and variance

distributions, but also on the monitoring statistics performance. In both cases, these effects are a direct consequence of the variability (uncertainty) in the sampling estimators for the correlations and variances. Therefore, an equivalent relationship between the number of observations n and the forgetting factor λ is here derived based on matching the standard deviation of the estimators for the correlation coefficients, σ_r, and sample variance, $\sigma_{v^{1/3}}$ (note that the variance is preliminarily transformed as $v^{1/3}$, in order to have a result consistent with the *VnMAX* statistic presented in Section 10.1.2, which also uses the same transformation), obtained by both procedures.

In order to perform such comparison, numerical simulations were conducted to determine the nature of the empiric relationship. A theoretical approach was also considered, from where an analytic relationship was obtained. Both expressions point to the same equivalent relationship between the number of observation and the forgetting factor, which consolidate its validity. The application example given at the end of this chapter illustrates the interest and relevancy of this expression in the selection of the control charts parameters.

12.1 Numerical Simulation

To obtain the desired relationship, the sample standard deviations obtained from moving windows of length n and EWMA recursions with a forgetting factor λ were matched under normal operation conditions for multiple process with m uncorrelated variables. The use of uncorrelated variables corresponds to the general case obtained after application of a suitable sensitivity enhancing transformation as this is the case where the higher sensitivities to changes on the correlations coefficients are obtained. In this study, m was set in the range of 2 to 300, λ was varied between 0.001 and 1 (with increment of 0.001) and the sample standard deviations ($\hat{\sigma}_r$ and $\hat{\sigma}_{v^{1/3}}$) were determined from 3000 observations.

The sample standard deviations obtained for different number of variables are presented in Figure 12.1. The first result is the apparent invariance to the number of variables as they all lead to the same pattern. In fact, the number of variables only seems to limits the maximum λ from which an equivalent number of observations can be obtained (given the

limitation in estimating a covariance matrix when the number of observations is less than the number of variables). The difference between the sample standard deviations obtained with the correlation coefficients and transformed variance is consistent with their theoretical expressions (Equations (10.1) and (10.8)), differing only by a factor of $\sqrt{9/2}$. This evidence was then used to construct a relationship between the standard deviation and the forgetting factor. These two variables appears to be related by a power model, defined as,

$$y = ax^b \tag{12.1}$$

Taken x as the forgetting factor, and y as the standard deviation, the least squares regression of this model led to the parameters presented in Table 12.1.

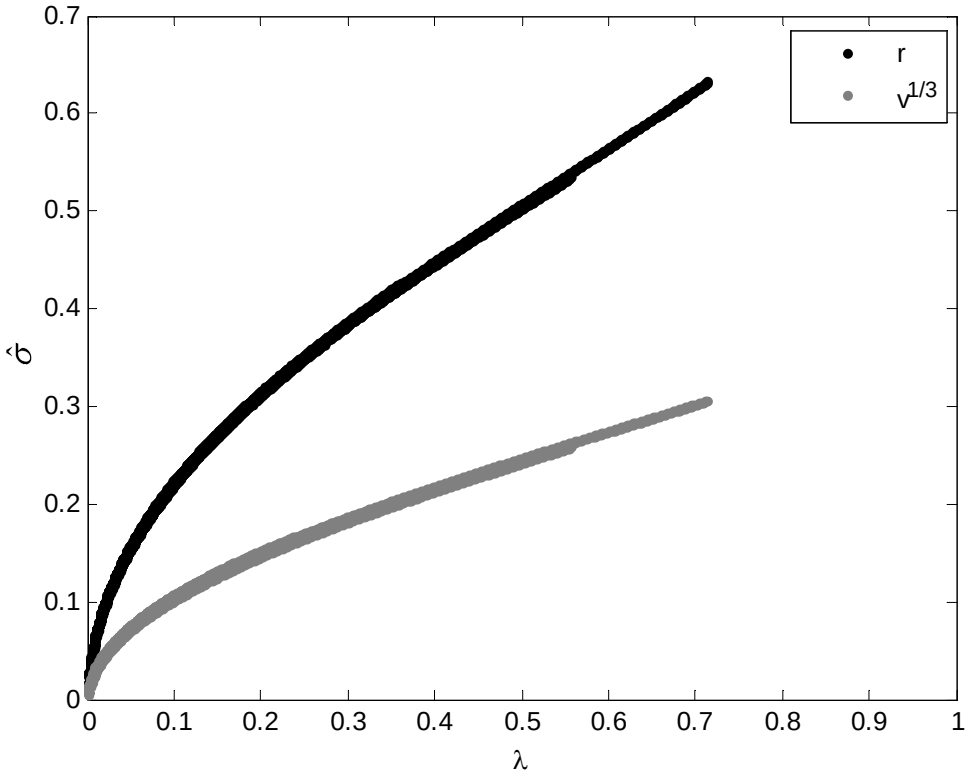


Figure 12.1 Sample standard deviation of the correlation coefficients (r) and transformed variance ($v^{1/3}$) as a function of the EWMA recursion forgetting factor (λ).

Table 12.1 Model parameters estimates for the power model relating the standard deviation as a function of the forgetting factor, using data obtained from the numerical simulations carried out. The coefficient of determination (R^2) is 0.9987.

Parameter	Estimated values	95% Confidence bounds
a	0.7356	[0.7351, 0.7361]
b	0.5212	[0.5207, 0.5216]

Recalling that the standard deviation of the correlation coefficients of two uncorrelated variables is asymptotically given by [46],

$$\sigma_r = \frac{1}{\sqrt{n-1}} \tag{12.2}$$

an equivalent number of observation (n_{eq}) can be obtain from,

$$n_{eq} = \left(\frac{1}{a\lambda^b} \right)^2 + 1 \tag{12.3}$$

It is worth noticing that the model parameters can be approximated, without a significant loss of information by,

$$n_{eq} \approx \frac{2}{\lambda} \tag{12.4}$$

Given the simulation data used in this study, Equation (12.4) is valid for $n > m$ with $m \leq 300$.

12.2 Analytical Derivation

In order to obtain the analytical relationship for the equivalence between λ and the number of observations, *n*, the theoretical variances of the sample correlation and variance based on a window of length *n* and the their analogous expression obtained from the EWMA recursion, were compared.

For the case of moving windows of length *n*, the asymptotical sample variance of the correlation coefficients for uncorrelated variables is given as [46],

$$\mathrm{var}(r) = \frac{1}{n-1} \tag{12.5}$$

For deriving the corresponding variance based on EWMA recursions the following considerations were taken.

The univariate EWMA recursion is generically given by,

$$e_t = \lambda x_t + (1-\lambda) e_{t-1} \tag{12.6}$$

where its mean and asymptotic variance are given by [48],

$$E(e_t) = E\left(\lambda \sum_{i=0}^{t-1} (1-\lambda)^i x_{t-i} + (1-\lambda) x_0 \right) = E(x_t) \tag{12.7}$$

$$\text{var}(e_t) \approx \text{var}(x_t) \left(\frac{\lambda}{2-\lambda} \right) \tag{12.8}$$

For the case of the marginal correlation coefficient (r_t), due to the action of the sensitivity enhancing transformation, x_t is the product of two uncorrelated variables (y_t and u_t) distributed as $N(0,1)$, therefore, assuming that these variables are also independent [144],

$$\begin{aligned} \text{var}(x_t) = \text{var}(y_t u_t) &= E(y_t)^2 \, \text{var}(u_t) + E(u_t)^2 \, \text{var}(y_t) + \text{var}(y_t) \, \text{var}(u_t) \\ &= 1 \end{aligned} \tag{12.9}$$

and consequently,

$$\text{var}(r_t) \approx \frac{\lambda}{2-\lambda} \tag{12.10}$$

Comparing Equations (12.5) and (12.10) it is straightforward to obtain that the equivalent number of observation (n_{eq}) of an EWMA recursion scheme with forgetting factor λ, in what concerns the estimate of the sample covariance is given by,

$$n_{eq} = \frac{2}{\lambda} \tag{12.11}$$

The same procedure can be conducted with the variables variance. In this case, x_t in Equation (12.6) is given by the square of a random variable (y_t) normally distributed as $N(0,1)$. Therefore, and considering that [144],

$$\text{var}(z) = E(z^2) - E(z)^2 \tag{12.12}$$

the variance of x_t is given by,

$$\mathrm{var}\left(x_t\right) = \mathrm{var}\left(y_t^2\right) = E\left(y^4\right) - E\left(y^2\right)^2 \tag{12.13}$$

Since the second and forth central moments of the normal distribution are [145],

$$\begin{aligned} E\left(y^2\right) &= \mathrm{var}\left(y_t\right) \\ E\left(y^4\right) &= 3\,\mathrm{var}\left(y_t\right) \end{aligned} \tag{12.14}$$

and that y_t has variance equal to 1, the variance of x_t becomes,

$$\mathrm{var}\left(x_t\right) = 2 \tag{12.15}$$

Combining this result with Equation (12.8), the variance of the sample variance (v_t) recursively estimated by the EWMA recursion is approximately given by,

$$\mathrm{var}\left(v_t\right) \approx 2\left(\frac{\lambda}{2-\lambda}\right) \tag{12.16}$$

Additionally, from Equations (12.7) and (12.14), the mean of v_t is obtained by,

$$E\left(v_t\right) = E\left(x_t\right) = E\left(y_t^2\right) = \mathrm{var}\left(y_t\right) = 1 \tag{12.17}$$

However, we are really interested in monitoring $v_t^{1/3}$ in order to maintain a consistent approach with the off-line scenario (see Equation (10.8)) and detect both increases and decreases in variance based on a variable that is normally distributed. Therefore it is required to obtain the mean and variance of this transformation.

For a generic function $g(z)$, with $g'(z)$ and $g''(z)$ as first and second derivative, its mean and variance are approximately given by [144],

$$\begin{aligned} E\left(g\left(z\right)\right) &= g\left(E\left(z\right)\right) + \frac{1}{2}\mathrm{var}\left(z\right)g''\left(E\left(z\right)\right) \\ \mathrm{var}\left(g\left(z\right)\right) &= \mathrm{var}\left(z\right)\left[g'\left(E\left(z\right)\right)\right]^2 \end{aligned} \tag{12.18}$$

as long as $g(z)$ and $g(z)^2$ are well approximated by a linear function on the operation interval of the transformation. For this case, $g(z) = z^{1/3}$, from where it results that,

$$E\left(v_t^{1/3}\right) \approx 1 - \frac{2}{9}\frac{\lambda}{2-\lambda}$$
$$\text{var}\left(v_t^{1/3}\right) \approx \frac{2}{9}\frac{\lambda}{2-\lambda} \tag{12.19}$$

When compared against Equation (10.8), from where it is recall that,

$$E\left(\left(s^2\right)^{1/3}\right) = 1 - \frac{2}{9(n-1)}$$
$$\text{var}\left(\left(s^2\right)^{1/3}\right) = \frac{2}{9(n-1)} \tag{12.20}$$

where s^2 is the sample variance based on n observation of a random variable distributed as $N(0,1)$, both expressions point to an equivalent number of observation given by Equation (12.11).

12.3 Consequences of the Equivalence Relationship between Off-line and On-line Monitoring

As the equivalence relationship derived in the previous sections relates the standard deviation of the correlation coefficients obtained by both moving windows and EWMA recursion schemes, both procedures will have the same uncertainty on the estimates of the sample correlation coefficients when the respective parameters satisfy this expression. Moreover, as the monitoring statistics' detection power is related with such uncertainty, both approaches will also present similar properties (such as, control limits and detection rates). The same behavior is observed for the sample variance.

An immediate application of this equivalence is in the normalization function used to obtain correlation coefficients and variances distributed as $N(0,1)$. These new normalization functions are applied in combination with the EWMA recursion scheme and can be written as,

$$w_r = \frac{r}{\sqrt{\lambda/2}} \tag{12.21}$$

$$w_s = \frac{\left(\frac{s^2}{\sigma_0^2}\right)^{1/3} - \left(1 - \frac{2}{9}\frac{\lambda}{2-\lambda}\right)}{\sqrt{\frac{2}{9}\frac{\lambda}{2-\lambda}}} \tag{12.22}$$

Another useful application of this equivalence is in the selection of the forgetting factor λ. In the case of non-overlapping windows, the number of observations can be selected using power analysis, i.e., defining the fault magnitude $\Delta\rho$ to be detected, with a $(1-\beta)$ power at a given significance level α. It can be shown, that under the same false detection rate (α), the EWMA recursive version of *RMAX* will have a similar performance as a non-overlapping window of length n, where n is given by the equivalence Equation (12.4). Therefore, the selection of the forgetting factor can be made based on the choice of a simpler and intuitive parameter (n). As an example of this relation, the detection rates obtained in a system with 5 uncorrelated variables are presented in Figure 12.2 along with the calibration curves for non-overlapping moving windows (off-line scenario) in the same system with a false detection rate of 1%. The values depicted are the average of 1000 detection rates associated with deviations on one of the correlation coefficient and are plotted against the forgetting factors' equivalent number of observations. For most of the simulated faults, the EWMA recursion scheme presents detection rates that fall very close to the corresponding values for the moving window approach. This confirms that, on average, both methodologies are related through the false detection rate and the equivalent number of observations, thus providing a good starting point to select the forgetting factor λ. Moreover, the in-control ARL of the on-line EWMA recursive scheme can be obtained as a function of the false detection as shown in Table 12.2. Therefore, as for the off-line scenario, a control chart with the desired characteristics (i.e. false alarm and detection power) can be easily constructed by analysis of calibration curves similar to the one presented in Figure 12.2.

To illustrate this parameter selection procedure, let us consider that one is interested in constructing an on-line control chart for 5 variables with a false detection rate of 1%, capable to detect a deviation in correlation of 0.15 with a detection power of 30%. From inspection of the off-line calibration curve in Figure 12.2, these specifications lead to a non-overlapping window of about 320 observations. From the equivalence relationship, a EWMA recursion with $\lambda = 2/n = 2/320 \approx 0.0063$ should then be used in order to obtain the desired performance. The detection rates of this monitoring scheme are presented in

Table 12.3. These results show that not only the desired detection power is indeed obtained (i.e., a detection rate of 0.30 for deviations of ±0.15) but also that the remaining deviations are in line with the off-line calibration curve.

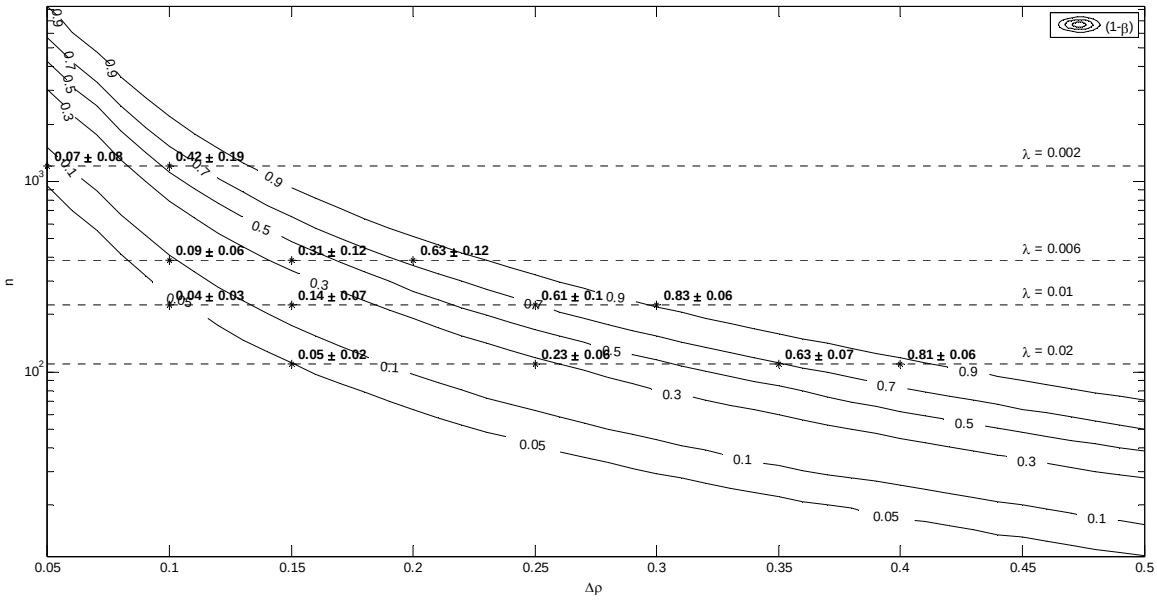


Figure 12.2 *R*0*MAX* calibration curves for non-overlapping windows of length *n* for a 5 variables system with a false detection rate of 1%. The detection rates of *R*0*MAX* based on an EWMA recursion procedure are superimposed in bold for their corresponding equivalent number of observations.

Table 12.2 False detection rates (α) for systems with *m* uncorrelated variables and $ARL_0 = 370$.

	m			
λ	5	10	15	20
0.002	0.157	0.126	0.117	0.108
0.004	0.090	0.071	0.064	0.058
0.006	0.062	0.050	0.044	0.041
0.008	0.048	0.038	0.034	0.032
0.010	0.039	0.030	0.028	0.026
0.012	0.033	0.026	0.024	0.024
0.014	0.029	0.022	0.021	0.020
0.016	0.026	0.020	0.019	0.018
0.018	0.023	0.018	0.017	0.016
0.020	0.021	0.017	0.016	0.015

Table 12.3 Detection rates of an on-line control chart design based on the off-line calibration curves. The results are averages of 10 simulations of 5000 observation. The associated standard deviations are also presented.

$\Delta\rho$	Detection Rate	$\Delta\rho$	Detection Rate
		0	0.0087 ± 0.0098
-0.05	0.0210 ± 0.0124	0.05	0.0268 ± 0.0168
-0.10	0.0664 ± 0.0318	0.10	0.0866 ± 0.0463
-0.15	0.3200 ± 0.0582	0.15	0.2685 ± 0.0547
-0.20	0.5805 ± 0.0585	0.20	0.6502 ± 0.1120
-0.25	0.8876 ± 0.0291	0.25	0.8542 ± 0.0499

The similarity in behavior in not only restricted to off-line and on-line monitoring. The same relation can be used to compare the on-line monitoring based on moving windows and the EWMA recursion scheme. On this regard, it is observed that control charts related by the equivalence equation present a similar pattern as exemplified in Figure 12.3. In Figure 12.3, *RMAX* statistics are plotted for both EWMA recursion estimation ($RMAX_{EWMA}$) and moving windows approach ($RMAX_{MW}$). Even though these monitoring statistics do not actually overlap, since the observations have different weights depending on the monitoring scheme, they present similar trends and detect the fault roughly at the same time. Moreover, the upper control limit is the same for these two monitoring statistics.

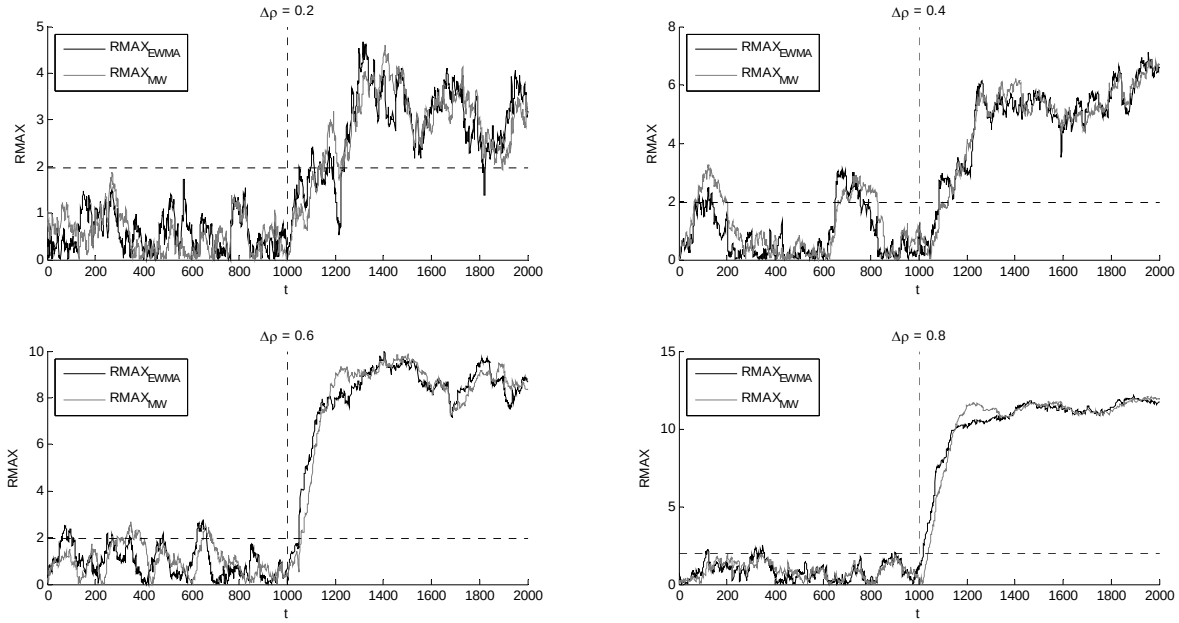


Figure 12.3 *RMAX* statistics for a two variables system with a step deviation on the correlation matrix of magnitude $\Delta\rho$ introduced at $t = 1001$. $RMAX_{EWMA}$ is based on the EWMA recursion with $\lambda = 0.01$ and $RMAX_{MW}$ is based on a moving window of 200 observations. The UCL is set to a false alarm rate of 0.01.

12.4 Summary

The parameters of the off-line monitoring (number of observations, n) and on-line monitoring (forgetting factor, λ) are related through an equivalence equation that allows the construction of a transversal interpretation of the monitoring performance of both approaches. This situation was demonstrated for several application scenarios, showing the usefulness of the equivalence relation.

The obtained equivalence relationship relies on the principle of matching the uncertainties of the estimated monitored characteristics (i.e., the sample correlation and sample variance). Therefore, as both procedures will give estimates with similar uncertainties, their performances will be also similar. In this respect, it was found that the off-line and on-line schemes related by the equivalence relation and with the same false alarm rate, do have the same detection power. Note however, that their ARL is not the same, since in the off-line case the monitoring statistics are serial independent, while in the on-line scenario they become autocorrelated due to the overlapping of information inherent to the EWMA recursion. Still, the proposed equivalence allows for a better interpretation of the on-line procedure and links its performance with the more simple case of non-overlapping windows, where calibration curves are easier to obtain. Thus, the selection of the EWMA forgetting factor can be based on the results of the more simple case of non-overlapping windows.

13 Multiscale and Megavariate Monitoring of the Process Correlation Structure: M2NET

Most current industrial processes encompass several underlying phenomena that take place simultaneously and span over different regions of the time-frequency plane. This is a consequence of the system's multiscale nature, which is also reflected in the patterns exhibited by data they generate. Consequently, classical monitoring methods are not effective for this type of processes, as they were design to detect changes occurring at a single and rather specific scale [5, 58]. Moreover, they are also mostly dedicated to detect deviations from the process mean levels [66-69]. Therefore, it is both opportune and important to extend the monitoring procedures for the correlation structure proposed previously in this thesis, to multiscale dynamical scenarios. This will be addressed in this chapter.

Perhaps the first successful implementation of a multiscale approach for process monitoring, was the one proposed by Bakshi (1998) [9]. This approach is based on decomposing the original observations into a set of wavelet coefficients at different scales, which are then simultaneously monitored through parallel MSPC-PCA schemes. From the monitoring of the wavelet coefficients, the relevant (abnormal) events at each scale can be detected and selected to be used in the reconstruction of the signal, in what effectively corresponds to performing a feature extraction of the fault signature. Then, the reconstructed signal is again subjected to a confirmatory assessment of the actual state of the process. Other approaches based on this procedure have also been proposed [146-148], all of them sharing the same methodological backbone.

As to the monitoring of the process multivariate dispersion, the current approaches relay on the information conveyed by the marginal covariance through the application of successive likelihood ratio tests [70, 75, 76, 82] or with resort to the generalized variance [71, 72]. Other approaches can be found in the literature for both subgroup (off-line) and individual (on-line) observations monitoring as described in Chapter 4. Amount the current procedures, the M_1Z^2 monitoring statistic proposed by Reynolds and Cho (2006) [81], based on the Mahalanobis distance of an EWMA recursion applied to the squared

deviations from target, and the Bodnar *et al.* (2009) [88] transformation of the estimated covariance matrix into a set of Gaussian vectors, were the ones that presented the most promising results (see Chapter 11). However, the proposed on-line *RMAX* and *VnMAX* statistics (Section 11.1) showed to have similar performances being also less sensitive to the recurrent problems faced in megavariate process monitoring, namely the inversion of the ill-conditioned covariance matrices, which is not required in the proposed *RMAX* and *VnMAX* statistics. Moreover, the current monitoring schemes also present severe implementation issues that hinder their extension to multiscale contexts. For instance, the M_1Z^2 statistic is based on an EWMA recursion scheme, which cannot be straightforwardly applied at the data reconstruction level, since the scales included in the reconstructions are not always the same. This situation also implies that an EWMA update is not a feasible solution to estimate the covariance matrix of the reconstructed data in a multiscale procedure analogous to the one proposed by Bakshi (1998) [9]. However, from the equivalency relationship, the *RMAX* and *VnMAX* procedures, can also be applied to moving windows of length n, with the same detection properties than an EWMA recursion with $\lambda = 2/n$ (see Chapter 12). For these reasons, the *RMAX* and *VnMAX* statistics based on moving windows were selected for the development of a multiscale monitoring scheme of the process correlation structure.

Given these considerations, in this chapter, a combined multiscale procedure involving the *RMAX*/*VnMAX* statistics and wavelet decomposition is proposed which will be tested, validated and compared with its single-scale counterpart. The effects of the sensitivity enhancing transformations (SET) presented in Chapter 9 in the final detection performance of the monitoring schemes are also considered. The entire methodology will be called the M2NET (Megavariate and Multiscale monitoring of the process NETworked structure).

13.1 Multiscale *RMAX* and *VnMAX*

For the purpose of monitoring the structure of megavariate processes with multiscale features, the general multiscale procedure proposed by Bakshi (1998) [9] was adapted and combined with monitoring statistics for the process correlation structure. With this goal in mind, the *RMAX* and *VnMAX* monitoring statistics for on-line monitoring, as well as the

sensitivity enhancing transformations (SET) were used to develop a multiscale monitoring scheme, as follows.

The first step of the proposed multiscale methodology involves the decomposition of the current observation into a set of wavelet coefficients. As the procedure should be implemented on-line, a dyadic window composed by the last $2^{J_{max}}$ observations (where J_{max} is the wavelet decomposition depth) is used to compute the current values of the wavelet coefficients at each observation. This new set of wavelet coefficients at each scale constitute the data that will be subjected to monitoring using *RMAX*/*VnMAX* statistics. Therefore, at each scale, a SET is applied in order to decorrelate the data and maximize the detection of structural changes. Then, the sample correlation matrix and variance are computed based on the last *n* transformed wavelet coefficients. Afterwards, the sample correlations at each scale are monitored by *RMAX* and the sample variances by *VnMAX*. Contrary to the procedure proposed by Bakshi (1998) [9], where the T^2_{PCA} and Q statistics are used through a logical gate "or" in order to determine the scales where special events are detected, in the current monitoring scheme, *RMAX* and *VnMAX* are treated separately in order to better isolate the nature of the features arising from each scale. Therefore, the rest of the procedure is here described solely for *RMAX*, since the same sequence of stages would entirely apply for the *VnMAX* statistic.

Given the current value of the *RMAX* statistic at each scale ($RMAX^{(j)}$, $j = 1, 2, \ldots, J_{max}+1$), the ones that have a *RMAX* greater than a pre-specified UCL ($\mathrm{UCL}^{(j)}$, $j = 1, 2, \ldots, J_{max}+1$) are considered as relevant and used to reconstruct the data at the original scale. To determine the UCLs, the false alarm rate of the control chart at each scale (α_A) is corrected by the Bonferroni inequality, $\alpha_A = \alpha / \left(J_{max}+1\right)$, where α is the desired overall false alarm rate of the monitoring procedure. Thus, the effects of using multiple parallel control charts at the wavelets levels are mitigated in this way.

In the reconstruction stage, the last *n* observations are reconstructed using the transformed wavelet coefficients (i.e., the SET wavelet coefficients) of the relevant scales determined earlier. The need for reconstructing the last *n* observations comes from the requirements associated with the estimation of the covariance matrix. These *n* reconstructed observations are then used to estimate the correlation coefficients and finally the *RMAX* statistic of the reconstructed data ($RMAX^{(rec)}$). Note that no SET is applied to the

reconstructed data, since the reconstruction of uncorrelated wavelet coefficients also leads to uncorrelated reconstructions.

A resume of the multiscale *RMAX* (*MS-RMAX*) methodology is given in Table 13.1 in the form of a pseudo-code. The graphic representation of the procedure is also provided in Figure 13.1. The same monitoring scheme is applied for *VnMAX* to monitor the sample variances.

Table 13.1 Pseudo-code for the proposed M2NET methodology (for the case of the *RMAX* statistic).

1. Get dyadic window containing the current observation (length equal to $2^{J_{max}}$);
2. Obtain the current wavelet coefficients up to scale $J_{max} + 1$;
3. Apply a sensitivity enhancing transformation to each scale;
4. For each scale:
    a. Compute the correlation of the last *n* transformed wavelet coefficients;
    b. Implement *RMAX*-based MSPC, using $RMAX^{(j)}$ statistics, and verify if the scale is relevant (i.e. if $RMAX^{(j)} \geq ULC\left(RMAX^{(j)}\right)$).
5. If at least one scale was found to be relevant, reconstruct the last *n* observations using the transformed variables from the relevant scales (by doing so, the reconstructed data is also uncorrelated);
6. Compute the correlation of the data reconstructed in Step 5;
7. Implement *RMAX*-based MSPC to the correlation at the original scale;
8. Obtain the reconstruction UCL from the database. If such an UCL is not available:
    a. Reconstruct validation data with the same relevant scales;
    b. Compute *RMAX* statistics for the validation data;
    c. Determine the UCL based on the *RMAX* for validation data;
    d. Save UCL on database.
9. Compare $RMAX^{(rec)}$ against its UCL.

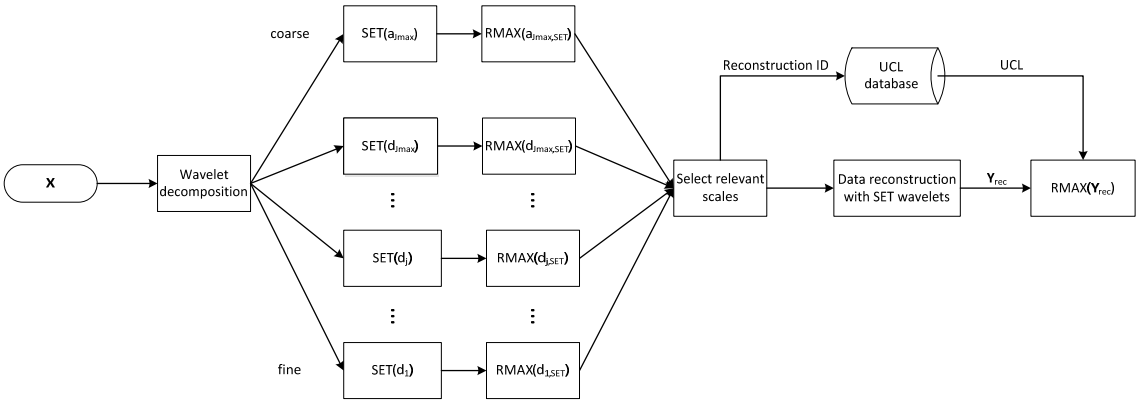


Figure 13.1 Schematic representation of the proposed M2NET methodology (for the case of *RMAX*).

13.2 Comparative Study

The performance of the proposed M2NET methodology through the statistics *MS-RMAX* and *MS-VnMAX*, is compared against their single-scale versions. Two SET were considered, namely: (i) a stationary linear transformation with a Cholesky decomposition of the residuals $T_{NetChLin}$ ($R0MAX_{NetChLin}$ and $VnMAX_{NetChLin}$) and (ii) a dynamic linear transformation with a Cholesky decomposition of the residuals $T_{NetChDyn}$ ($R0MAX_{NetChDyn}$ and $VnMAX_{NetChDyn}$). For the multiscale approach the same type of transformations were considered for all scales and reconstructions. Thus, for transformation $T_{NetChLin}$ the monitoring statistics $MS\text{-}R0MAX_{NetChLin}$ and $MS\text{-}VnMAX_{NetChLin}$ are defined. Likewise, for transformation $T_{NetChDyn}$ the monitoring statistics $MS\text{-}R0MAX_{NetChDyn}$ and $MS\text{-}VnMAX_{NetChDyn}$ were employed. Other sensitivity enhancing transformation schemes could be considered, such as different transformations at different scales, depending on the data features. However, in order to obtain a sound comparison of the monitoring methodologies, sensitivity enhancing transformations of the same category were used. Therefore, the effects of the wavelet decomposition in the monitoring procedure are the main source of variability in the results and the effects of SET are minimized. All the SET used were determined based on a 5000 observations NOC data set. For the case of the multiscale transformations, the same NOC data set was used to obtain the wavelet coefficients and an estimation of the causal network at each scale.

The case study selected to conduct the performance comparison study is the continuous stirred-tank reactor (CSTR) with a heating jacket, under feedback control. The study of the multiscale features of this particular system was already addressed in Section 7.2,

from where it is recalled that the variability of the input variables (feed stream concentration, C_{A0}, feed stream temperature, T_0, and heating fluid inlet temperature, T_{j0}) is mostly located at the low frequency scales (detail coefficients d_1 to d_3). On the other hand, the system outputs (CSTR level, h, concentration, C_A, temperature, T, and heating fluid outlet temperature, T_j) present different distributions across the scales, especially in regarding the sets of detail coefficients d_3 to d_8. Given these characteristics, a decomposition depth (J_{max}) equal to 8 was selected for implementing the multiscale approach.

The monitoring statistics were evaluated after introducing step perturbations in the heat transfer coefficient, discharge coefficient, activation energy and heat capacity of the heating fluid. These changes were made by application of a multiplicative factor δ to the nominal values of these quantities. For each simulation, 5000 NOC observations, followed by 5000 faulty observations were simulated and saved. These observations were then used to determine the sample covariance matrix based on moving windows containing the last 300 observations. The monitoring procedures are then implemented and their detection rates (both false detection rates and true detection rates) registered for control limits set to an overall false alarm rate (α) of 1% for all the monitoring schemes (i.e. the control limit of each control chart was determined for a significance level of $\alpha_A = \alpha / N_{chart}$, where N_{chart} is the number of control charts involved in the procedure). Each fault was simulated 5 times in order to analyze the consistency of the results.

For the case of deviations in the heat transfer coefficient, the multiplicative factor δ was set to 0.90, 0.85, 0.80 and 0.75 in order to simulate decreases in this process parameter, for instance due to fouling. The detection rates obtained by the various methods tested are given in Table 13.2. For illustration purposes, the monitoring statistics obtained for a fault of magnitude 0.80 are presented in Figure 13.2.

Table 13.2 Detection rates for step perturbations in the heat transfer coefficient for the single-scale and multiscale monitoring statistics.

Statistic	Fault magnitude			
	0.90	0.85	0.80	0.75
$R0MAX_{NetChLin}$	0.021 ± 0.014	0.036 ± 0.014	0.053 ± 0.003	0.060 ± 0.004
$VnMAX_{NetChLin}$	0.088 ± 0.026	0.070 ± 0.007	0.098 ± 0.034	0.111 ± 0.032
$R0MAX_{NetChDyn}$	0.065 ± 0.003	0.090 ± 0.028	0.112 ± 0.024	0.091 ± 0.027
$VnMAX_{NetChDyn}$	0.109 ± 0.035	0.314 ± 0.092	0.894 ± 0.059	1.000 ± 0.000
$MS\text{-}R0MAX_{NetChLin}$	0.172 ± 0.06	0.161 ± 0.039	0.221 ± 0.024	0.353 ± 0.081
$MS\text{-}VnMAX_{NetChLin}$	0.151 ± 0.023	0.240 ± 0.05	0.416 ± 0.128	0.517 ± 0.063
$MS\text{-}R0MAX_{NetChDyn}$	0.197 ± 0.025	0.296 ± 0.022	0.387 ± 0.104	0.491 ± 0.084
$MS\text{-}VnMAX_{NetChDyn}$	0.169 ± 0.030	0.327 ± 0.048	0.677 ± 0.058	0.954 ± 0.026

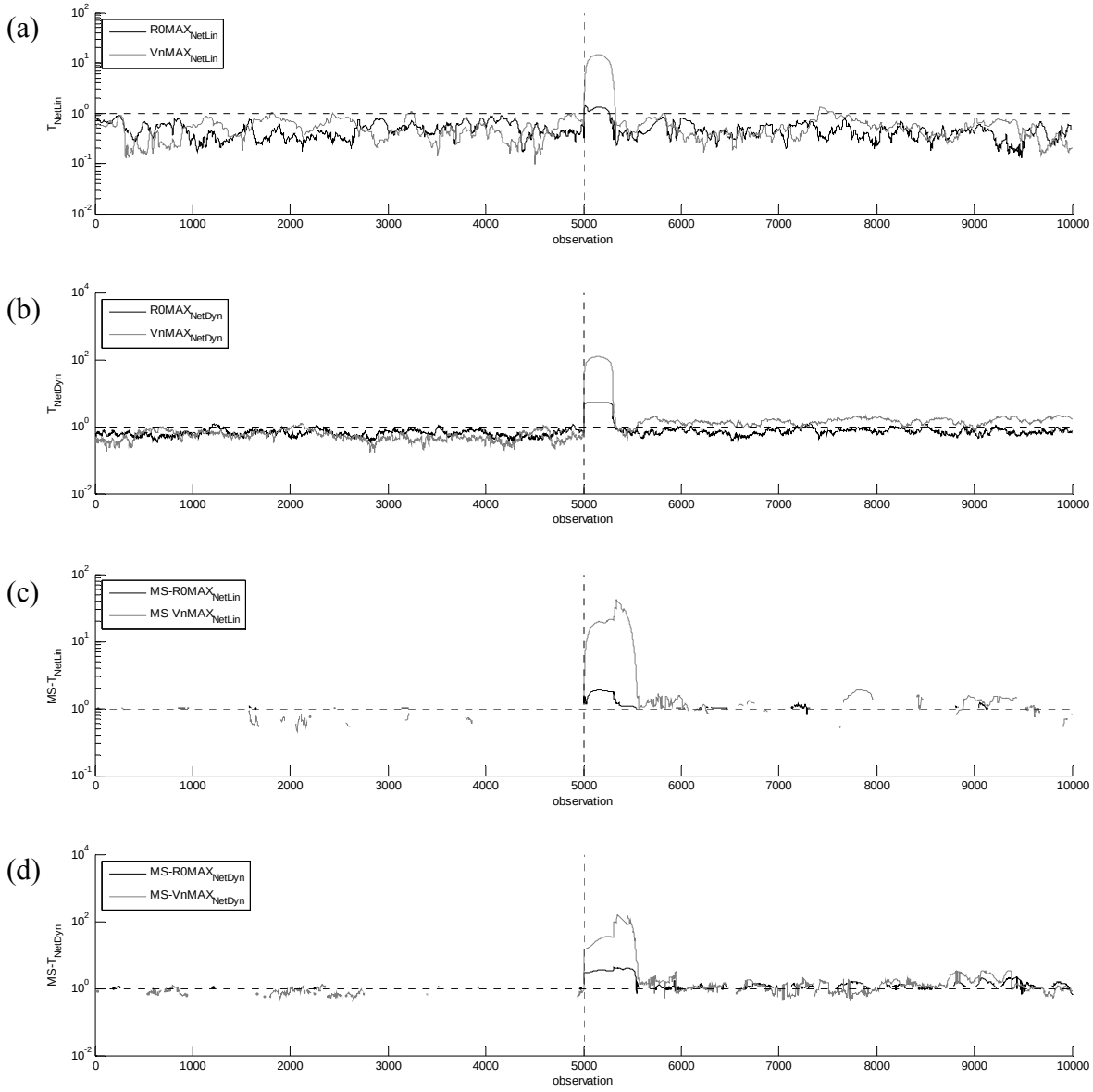


Figure 13.2 *RMAX* and *VnMAX* monitoring statistics for a step deviation in the heat transfer coefficient of magnitude 0.80: (a) single-scale $R0MAX_{NetLin}$ and $VnMAX_{NetLin}$; (b) single-scale $R0MAX_{NetDyn}$ and $VnMAX_{NetDyn}$; (c) multiscale $MS\text{-}R0MAX_{NetLin}$ and $MS\text{-}VnMAX_{NetLin}$; (d) multiscale $MS\text{-}R0MAX_{NetDyn}$ and $MS\text{-}VnMAX_{NetDyn}$. The monitoring statistics presented are normalized by their UCL in order to account for the difference in UCL for each reconstruction.

From the analysis of Figure 13.2 it is visible that all methods remain mostly under control on the first 5000 NOC observations. For the cases of the multiscale approaches, the reconstruction is only performed in few instances, and most of the times the confirmatory test discard them as false alarms. Then, when the fault is introduced, a sudden increase in the *VnMAX* statistics occurs, due to changes in the process mean value, which are not properly accommodated by the monitoring statistics during the first faulty observations. This effect of the mean disappear around observation 5300, which is consistent with the moving window length (i.e., 300 observations). However, for the multiscale approaches, this effect takes more observations to vanish as a result of the delay introduced by the wavelet transform, as can be observed in Figures 13.2 (c) and.13.2 (d). After this initial response to the mean, the monitoring statistics became only sensitive to changes in correlation (detected by *RMAX*) and variance (detected by *VnMAX*). From Figure 13.2 and Table 13.2 it is clear that neither $R0MAX_{NetChLin}$ nor $VnMAX_{NetChLin}$ can efficiently detect the structural changes. This result is mostly a consequence of the SET $T_{NetChLin}$ which does not consider the variables time dependency. Therefore, the system's full structure is not being properly modeled. In fact, the transformed variables still present high autocorrelation, which undermines the detection of structural changes. On the other hand, the SET $T_{NetChDyn}$, removes most of the cross- and auto-correlation present on the data. Consequently, the variables are transformed to a new subspace of uncorrelated variables were detection of structural changes is maximized. This is indeed observed, since the $VnMAX_{NetChDyn}$ statistic has generally higher detections rates for most faults.

As for the multiscale monitoring statistics, from Table 13.2 it can be observed that $MS\text{-}R0MAX_{NetChLin}$ and $MS\text{-}VnMAX_{NetChLin}$ are far behind the single-scale $T_{NetChDyn}$ based statistics ($R0MAX_{NetChDyn}$ and $VnMAX_{NetChDyn}$). For instance, for a fault of $\delta = 0.75$, the average detection rate of $MS\text{-}VnMAX_{NetChDyn}$ is 0.517 while $VnMAX_{NetChDyn}$ considers all observations as faulty. This result arises from the fact that $T_{NetChDyn}$ may present some multiscale features: the regression model for the different variables may span different scales of time, depending on the variables involved. Therefore, multiscale features may be accounted for in this way. Moreover, the transformed variables are essentially white noise and any deviation of this behavior, regardless of the scale, will lead to changes in the process model, which are well captured by the *VnMAX* statistics. Nevertheless, a multiscale decomposition of the data can provide useful information regarding the scale of the fault, as well as simplify the development of the models since a stationary

transformation requires less computation burden than a complex time series identification.

When a dynamic SET, $T_{NetChDyn}$, is used to decorrelate the wavelet coefficients, the performance is greatly increased, as can be seen from Table 13.2 for the monitoring statistics *MS-R0MAX*$_{NetChDyn}$ and *MS-VnMAX*$_{NetChDyn}$. However, the detection rates obtain by this monitoring scheme are still very similar to the ones obtain with the single-scale scheme, namely *VnMAX*$_{NetChDyn}$. Moreover, it is worth pointing that a different dynamic model and a causal network may be required for each scale, which adds significant amounts of complexity to this approach. Nevertheless, depending on the nature of the data, a transformation based on the Cholesky decomposition (see Chapter 9) can be employed in some scales, which simplifies the multiscale procedure. This situation was not explored here, since the goal was to compare the monitoring procedures at their best attainable performance.

The results also show that the detection rates of the multiscale *RMAX* statistics (*MS-R0MAX*$_{NetChLin}$ and *MS-R0MAX*$_{NetChDyn}$) also increase with the faults magnitude. As reported earlier, this effect can be traced back to some specific scales. As this situation encompasses additional information about the perturbation, the fault diagnosis stage can be improved. Nevertheless, it is important to note that structural changes are detected by both *RMAX* and *VnMAX* statistics. Moreover, it was verified that different variable transformations lead to different types of deviations that can be detected by *RMAX* on some transformations and by *VnMAX* on others. Therefore, a fair comparison can only be implemented if both monitoring statistics are considered in simultaneous, since they have a complementary behavior. Even in such case, higher detection abilities are observed in the single-scale approach when a dynamic SET is used.

When the same simulation procedure is applied to a perturbation in the discharge coefficient and heat capacity of the heating fluid the fault detection rates presented in Tables 13.3 and 13.4 are obtained. These results are consistent with the previous analysis, i.e., the fault detection rates of the single-scale and multiscale methodologies with dynamic SET (T_{NetDyn}) are very similar. As for the faults in the activation energy, Table 13.5, *VnMAX*$_{NetDyn}$ clearly has the best performance. These results suggest that an adequate single-scale sensitivity enhancing transformation is able to deal with the multiscale nature of the process.

Table 13.3 Detection rates for step perturbations in the discharge coefficient for the single-scale and multiscale monitoring statistics.

Statistic	Fault magnitude			
	1.05	1.10	1.15	1.20
$R0MAX_{NetChLin}$	0.089 ± 0.011	0.117 ± 0.015	0.127 ± 0.010	0.136 ± 0.011
$VnMAX_{NetChLin}$	0.131 ± 0.025	0.134 ± 0.015	0.165 ± 0.027	0.192 ± 0.030
$R0MAX_{NetChDyn}$	0.113 ± 0.022	0.126 ± 0.023	0.148 ± 0.035	0.132 ± 0.024
$VnMAX_{NetChDyn}$	0.129 ± 0.024	0.365 ± 0.055	0.976 ± 0.023	1.000 ± 0.000
$MS\text{-}R0MAX_{NetChLin}$	0.185 ± 0.040	0.227 ± 0.019	0.342 ± 0.050	0.654 ± 0.067
$MS\text{-}VnMAX_{NetChLin}$	0.174 ± 0.019	0.237 ± 0.057	0.478 ± 0.053	0.822 ± 0.049
$MS\text{-}R0MAX_{NetChDyn}$	0.194 ± 0.044	0.276 ± 0.074	0.279 ± 0.062	0.411 ± 0.043
$MS\text{-}VnMAX_{NetChDyn}$	0.186 ± 0.033	0.437 ± 0.041	0.942 ± 0.036	1.000 ± 0.000

Table 13.4 Detection rates for step perturbations in the heat capacity of the heating fluid for the single-scale and multiscale monitoring statistics.

Statistic	Fault magnitude			
	0.9	0.85	0.8	0.75
$R0MAX_{NetChLin}$	0.003 ± 0.007	0.006 ± 0.005	0.010 ± 0.019	0.021 ± 0.014
$VnMAX_{NetChLin}$	0.078 ± 0.013	0.081 ± 0.010	0.073 ± 0.009	0.110 ± 0.018
$R0MAX_{NetChDyn}$	0.014 ± 0.010	0.096 ± 0.039	0.094 ± 0.035	0.088 ± 0.012
$VnMAX_{NetChDyn}$	0.092 ± 0.035	0.322 ± 0.120	0.882 ± 0.065	0.998 ± 0.003
$MS\text{-}R0MAX_{NetChLin}$	0.144 ± 0.029	0.159 ± 0.017	0.192 ± 0.049	0.326 ± 0.066
$MS\text{-}VnMAX_{NetChLin}$	0.123 ± 0.033	0.133 ± 0.029	0.258 ± 0.028	0.399 ± 0.076
$MS\text{-}R0MAX_{NetChDyn}$	0.111 ± 0.021	0.145 ± 0.034	0.154 ± 0.026	0.201 ± 0.080
$MS\text{-}VnMAX_{NetChDyn}$	0.157 ± 0.029	0.306 ± 0.061	0.759 ± 0.030	0.990 ± 0.013

Table 13.5 Detection rates for step perturbations in the activation energy for the single-scale and multiscale monitoring statistics.

Statistic	Fault magnitude			
	1.15	1.20	1.25	1.30
$R0MAX_{NetChLin}$	0.028 ± 0.013	0.045 ± 0.024	0.058 ± 0.007	0.06 ± 0.015
$VnMAX_{NetChLin}$	0.107 ± 0.029	0.133 ± 0.053	0.145 ± 0.039	0.174 ± 0.044
$R0MAX_{NetChDyn}$	0.086 ± 0.022	0.079 ± 0.016	0.084 ± 0.014	0.116 ± 0.044
$VnMAX_{NetChDyn}$	0.572 ± 0.102	0.907 ± 0.046	0.998 ± 0.005	1.000 ± 0.000
$MS\text{-}R0MAX_{NetChLin}$	0.169 ± 0.020	0.181 ± 0.034	0.254 ± 0.043	0.346 ± 0.082
$MS\text{-}VnMAX_{NetChLin}$	0.147 ± 0.030	0.188 ± 0.050	0.241 ± 0.047	0.311 ± 0.136
$MS\text{-}R0MAX_{NetChDyn}$	0.166 ± 0.030	0.127 ± 0.036	0.154 ± 0.034	0.22 ± 0.027
$MS\text{-}VnMAX_{NetChDyn}$	0.193 ± 0.010	0.243 ± 0.066	0.273 ± 0.089	0.442 ± 0.107

To further assess the relative performance of the single-scale and multiscale monitoring statistics a permutation test [104] was performed to each pair of monitoring statistics. The goal of this permutation test is to verify if the obtained detection rates are significantly different. The results of this test are presented in Table 13.6 where the *p*-value and the

statistics signal is display. From this results it is verified that as the monitoring statistic complexity increases (i.e., from single-scale to multiscale and stationary to dynamic transformations) better results are obtained. Still, the performance of $VnMAX_{NetChDyn}$ showed to give the best results, being statistically equal to $MS\text{-}VnMAX_{NetChDyn}$, with a *p*-value of 0.061. Thus, a single-scale monitoring scheme with a proper SET is capable to model a multiscale system and present detection rates similar to more complex multiscale monitoring schemes.

Table 13.6 *p*-values for the permutation test involving the detection rates obtained in the CSTR system with method A (see first column) and method B (see first line), on all simulated faults, along with the signal of the test statistic, i.e. sign(t_0). For instance, a plus (+) signal, indicates that method A leads to higher detections rates, on average, when compared to method B. The results in bold correspond to the case where method A is statistically superior than method B. The underlined results correspond to the case where both are statistically equal.

A \ B	$R0MAX_{NetChLin}$	$VnMAX_{NetChLin}$	$R0MAX_{NetChDyn}$	$VnMAX_{NetChDyn}$	MS-$R0MAX_{NetChLin}$	MS-$VnMAX_{NetChLin}$	MS-$R0MAX_{NetChDyn}$	MS-$VnMAX_{NetChDyn}$
$R0MAX_{NetChLin}$	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	-	(-)	(-)	(-)	(-)	(-)	(-)	(-)
$VnMAX_{NetChLin}$	**<0.001**	-	0.193	<0.001	<0.001	<0.001	<0.001	<0.001
	(+)	-	(+)	(-)	(-)	(-)	(-)	(-)
$R0MAX_{NetChDyn}$	**<0.001**	0.193	-	<0.001	<0.001	<0.001	<0.001	<0.001
	(+)	(-)	-	(-)	(-)	(-)	(-)	(-)
$VnMAX_{NetChDyn}$	**<0.001**	**<0.001**	**<0.001**	-	**<0.001**	**<0.001**	**0.001**	0.061
	(+)	**(+)**	**(+)**	-	**(+)**	**(+)**	**(+)**	(+)
MS-$R0MAX_{NetChLin}$	**<0.001**	**<0.001**	**<0.001**	<0.001	-	0.585	0.180	<0.001
	(+)	**(+)**	**(+)**	(-)	-	(-)	(+)	(-)
MS-$VnMAX_{NetChLin}$	**<0.001**	**<0.001**	**<0.001**	<0.001	0.585	-	0.151	<0.001
	(+)	**(+)**	**(+)**	(-)	(+)	-	(+)	(-)
MS-$R0MAX_{NetChDyn}$	**<0.001**	**<0.001**	**<0.001**	0.001	0.180	0.151	-	0.001
	(+)	**(+)**	**(+)**	(-)	(-)	(-)	-	(-)
MS-$VnMAX_{NetChDyn}$	**<0.001**	**<0.001**	**<0.001**	0.061	**<0.001**	**<0.001**	**0.001**	-
	(+)	**(+)**	**(+)**	(-)	**(+)**	**(+)**	**(+)**	-

13.3 Conclusions

The sensitivity enhancing transformation is confirmed as a major factor in the detection performance of structural changes. Its impact is very visible on this case study and shows that if a proper transformation is chosen, the resulting monitoring statistics can achieve the same level of detection capability as their multiscale counterparts. For most of the simulated faults (head transfer coefficient, discharge coefficient and heat capacity of the heating fluid) no significant difference in the fault detection rates is observed. Moreover, when the system was subjected to a perturbation in the activation energy, the detection rates of the single-scale $VnMAX_{NetChDyn}$ are undoubtedly higher. In fact, during this study, only under a limited number of situations the multiscale methodology lead to better results than the dynamic singe-scale approach.

Even though the wavelet transform present an intrinsic ability to decorrelate autocorrelated measurements, it may also induce autocorrelation in the wavelet coefficients due to the moving window approach used in the decomposition stage. This added autocorrelation can in some cases be more complex than the original signal, which will then cause problems on the model fitting, especially if a dynamic model is considered. However, even in the cases where only a stationary model was used, a significant improvement is observed when compared with the same single-scale transformation. Therefore, the application of the wavelet decomposition is advantageous for systems that are difficult to model and thus may provide a good solution when a balance must be established between the modeling complexity and performance.

Part V – Conclusions and Future Work

14 Conclusions

In this thesis, the problem of monitoring the location and correlation structure of complex systems was addressed through the development of new monitoring schemes with added value to the field of statistical process control (SPC). For monitoring the process location, the dynamic principal component analysis with decorrelated residuals (DPCA-DR) was proposed in Part III, while for monitoring the process multivariate dispersion a *RMAX*/*VnMAX* procedure was proposed in Part IV. Moreover, during the development of these two main contributions, other important aspects were also explored and solutions put forward, such as lag selection methods for DPCA (Chapter 5), sensitivity enhancing transformations for improving the monitoring of correlation (Chapter 9), network reconstruction algorithms (Chapter 9) and an equivalence relationship between off-line and on-line monitoring (Chapter 12). Even though these parallel developments were mostly a consequence of the main study, some of them are by themselves very useful for implementing current SPC approaches in practice.

The Part III of this thesis is focused in the problem of monitoring the process mean of megavariate dynamic processes. The proper monitoring of these type of processes was not yet completely solved, as the methodology of DPCA proposed by Ku *et al.* (1995) [27] still leads to monitoring statistics with autocorrelation. Moreover, their algorithm to select the lagged structure clearly underestimates the true time dependency of the data. In this regard, new lag selection procedures were proposed. These new methods present better lag estimation properties and one of them is even able to select a different number of lags for each variable. By doing so, a more precise description of the process can be obtained, which greatly affects the final performance of the monitoring schemes based of this lag selection method. This can be easily observed by implementing a combination of the current DPCA method with the new lag selection technique (Section 5.3.1). The results in this situation clearly reveal an improvement on the faults detection abilities, even though the autocorrelation in the monitoring statistics is still present.

To mitigate the autocorrelation effect referred earlier, a total of twenty two monitoring statistics (most of them new) were investigated, as described in Chapter 6. From this study, a class of monitoring statistics defined as dynamic principal component analysis

with decorrelated residuals (DPCA-DR) proved to be consistently superior to the remaining studied monitoring schemes. Moreover, the monitoring statistics based on DPCA-DR have low autocorrelation levels even for complex systems, such as the Tennessee Eastman process, without compromising their detection ability. In fact, greater detection performances are observed in all the case studies analyzed.

The implementation of DPCA-DR with wavelet transforms was also considered to improve the detection of deviations in systems with multiscale features. The procedure proposed in Chapter 7, MS-DPCA-DR, has is foundations in the work of Bakshi (1998) [9] and exploits the combined properties of wavelet transforms (that separate the time-frequency scales that compose the data) and DPCA-DR (that explain the dynamic structure of the processes) in order to construct a scale dependent monitoring scheme. The results obtained in this context proved that MS-DPCA-DR is capable to isolate the time-frequencies components related with the fault and to filter out their contribution from the signal. Thus, the unnecessary information is removed from the signal and a clearer fault pattern can be extracted. However, DPCA-DR was not capable to cope with the autocorrelation induced by the wavelet transform (due to the use of a moving window approach), especially for the lower time-frequency scales.

In Part IV the focus was centered on the monitoring of the process correlation structure. Comparatively to the problem of monitoring the process location, the detection of changes in the correlation structure of multivariate systems has received much less attention. In fact, even the most recent proposals are still related with the base concepts of the likelihood ratio test and generalized variance and use solely the information of the marginal covariance matrix. Thus, in order to improve their detection performance for localized perturbations, the causal structure of the processes was considered in this work. Partial correlation coefficients are a useful measurement of the local association between process variables, even though they do not convey information regarding the variables causality. Nevertheless, the proposed monitoring procedures based on partial correlations and network reconstruction techniques showed to be superior to the current marginal based approaches.

To improve the detection of changes in correlation it was observed that the use of uncorrelated variables greatly enhances the methods sensitivity. This observation is a direct result of the nature of the correlation curve, $\mathrm{corr}(x,z) = k/\sqrt{k^2 + w^2}$ (with

$x = kz + w\varepsilon$, where both z and ε follows a $N(0,1)$), which is maximized at $k/w = 0$, corresponding to corr(x,z) = 0. Therefore, even small deviations from the zero correlation are translated into large faults. However, as the typical data recovered from industrial processes have high levels of correlation, a variable sensitivity enhancing transformation (SET) is required. Still, not all transformations with decorrelation properties, such as the Cholesky decomposition or PCA, are appropriate to be used for this end. Instead, the SET that make use of the process causal structure showed to be the most adequate, as described in Chapter 9. Moreover, a new set of monitoring statistics was developed to take full advantage of the SET properties. The current monitoring schemes showed to be invariant or little affected by such variable transformations.

To monitor uncorrelated variables, the *RMAX* and *VnMAX* monitoring statistics were found to be the most suitable ones. This conclusion was first drawn for the off-line monitoring of the process correlation structure (Chapter 10), where it was observed that they were capable to detect small deviations even when a small number of observations was used to compute the sample covariance matrix. Moreover, it was verified that these two monitoring statistics have complementary features, since *RMAX* detects faults at the correlation level and *VnMAX* at the variance level. *VnMAX* is also related with the model validity, and therefore it is also capable to detect structural changes as a result of model mismatch.

Due to their good performance on the off-line scenario, the *RMAX* and *VnMAX* monitoring schemes were extended to the on-line scenario in Chapter 11. To do so, an EWMA recursion scheme was employed in order to estimate the covariance matrix. Compared to the current on-line approaches, *RMAX* and *VnMAX* were found to have similar performance when stationary linear SET are used and higher performances when a dynamic non-linear SET is employed. The observed improvements in performance are a result of both monitoring statistics (*RMAX* and *VnMAX*) and the SET.

In both off-line and on-line monitoring, the SET proved to be the major factor responsible for the increase in performance. The impact of this transformation is so relevant, that even a single-scale transformation can cope with the multiscale characteristics of the processes and outperform multiscale oriented approaches, as observed in Chapter 13.

15 Future Work

The main focus of the current thesis is the monitoring of multi- and megavariate dynamic continuous systems, which culminated in the development of the approaches summarized in Chapter 14 for monitor their location and multivariate dispersion. These methodologies can also be applied to other research areas, such as the monitoring of batch processes and time-varying processes. Further improvements in the network reconstruction algorithms can also be achieved, since the ones proposed here are mostly oriented to the decorrelation of the process variables in a meaningful way, rather than to extract the real process causal network. The effects of using the sensitivity enhancing transformations in the monitoring of the process location was also not fully explored, even thought it was visible that the *VnMAX* statistics are able to detect such faults. Moreover, a combined monitoring scheme to detect both location and dispersion is also of interest, since it was observed that PCA (and its derivative forms) have a relatively low sensitivity to changes in correlation and the monitoring statistics for the process correlation structure are not appropriate to detect changes in the mean value.

Fault diagnosis is another area of interest for future developments. While for DPCA-DR contributions plots or reconstruction-based contributions can be used [149], for the monitoring schemes based on partial correlations a different approach is required.

Some of these questions will be discussed in the following sections, where more information is given regarding future areas of research in the line of the developments proposed in this thesis.

15.1 Application of DPCA-DR to Batch Processes

Even though the main focus of this thesis was toward the monitoring of continuous processes, the monitoring of batch processes is also a relevant field of research, since they are very common in the chemical and biochemical industry, namely in the production of high added-value materials and products, such as polymers, pharmaceuticals, biochemicals and semiconductors [150, 151]. Furthermore, the on-line monitoring of

batch processes not only allows for the promptly detection and correction of abnormal situations, but also save production time in the cases where it is found that the product specifications will not be attained in the end of the batch.

To conduct such monitoring tasks in batch processes, several approaches based on multi-way principal component analysis (PCA) and partial lest squares (PLS) have been proposed [19, 150, 152]. A characteristic feature of these procedures is the way data is organized, since in batch processes data present three fundamental dimensions (I batches × J variables × K time periods) while the classic latent variables methods are only able to handle two dimensional data tables. To address this issue, the data is usually unfolded in either a batch-wise or variable-wise fashion.

In the case of batch-wise unfolding, a matrix of size ($I \times JK$) is obtained by placing different time periods side by side, as depicted in Figure 15.1 (a). Thus, each row of the unfolded matrix contains the data of a single batch and the columns correspond to the measurements at different times. By doing so and after proper pre-processing, the full linear dynamic behavior around the average trajectory can be taken into account. However, due to its construction, it requires the estimation of the future observations during the monitoring stage. To handle this issue several approaches were proposed for future data imputation, such as by assuming a zero deviation regarding the mean trajectory for future observations, or that the current deviation will be maintained throughout the rest of the batch. Alternatively, these "missing" values may also be estimated by application of missing data techniques [150, 152].

As for the variable-wise unfolding, the data is organized as present in Figure 15.1 (b) into a ($IK \times J$) matrix by stacking the data at each time period. Therefore, each row of the unfolded matrix corresponds to a single time observation. Consequently, this approach only considers the instantaneous cross-correlation between the measurements, ignoring their time dependency.

After unfolding the data in a meaningful way, the typical PCA-based approaches to monitor the process state through the Hotelling's T^2 and Q statistics can be employed. Some examples of these monitoring schemes are described below.

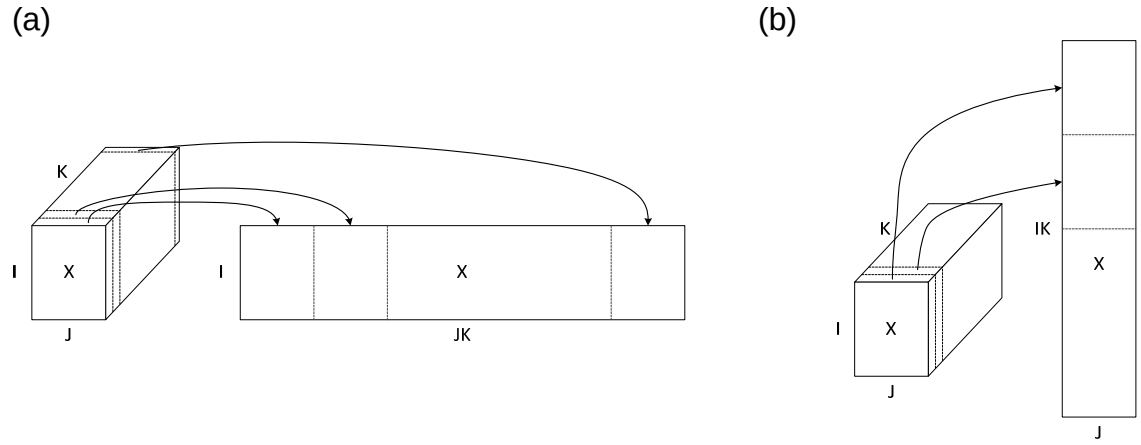


Figure 15.1 Unfolding of a three-way array of *I* batches × *J* variables × *K* time periods: (a) batch-wise unfolding to a *I*×*JK* matrix; (b) variable-wise unfolding to a *IK*×*J* matrix.

The batch-wise multi-way PCA (MPCAB) proposed by Nomikos and MacGregor (1994) [57] uses the batch-wise unfolded data matrix and normalize the data by subtracting the mean trajectory of each variable over time. Given the data organization, this normalization step is equivalent to mean centering each column to its normal operation conditions value. After that, the data is scaled to unit variance and the scores and residuals of the PCA model are computed and monitored as usual. However, as mentioned earlier, due to the unfolding technique, the variables corresponding to future observations are not available and need to be properly estimated.

Wold *et al.* (1998) [153] proposed a different approach by considering a variable-wise multi-way PCA (MPCAV). In this case, data is unfolded in a variable-wise fashion and normalized by taking the variables grand mean. This procedure has the advantage of avoiding the estimation of future observations (as happens in MPCAB), but, as it monitors the variables deviations from the mean value over the batch process, it fails to take the batch dynamics into account. Thus, an usual alternative is to normalize the data around the average trajectory prior to unfolding the data in a variable-wise way [151].

Chen and Liu (2002) [154] proposed a batch dynamic PCA (BDPCA) methodology that extends the DPCA method of Ku *et al.* (1995) [27] to batch processes. The base idea of this method is the construction of an extended data matrix of time-shifted replicates for each batch. That is, for batch *i*, the extended data matrix, $\tilde{\mathbf{X}}^{(i)}$, with *l* lags is given as,

$$\tilde{\mathbf{X}}^{(i)} = \left[\mathbf{X}(0) \quad \mathbf{X}(1) \quad \ldots \quad \mathbf{X}(l)\right] \tag{15.1}$$

where $\mathbf{X}(j)$ is an $((K-l)\times J)$ matrix of variables shifted j times into the past Thus, the sample covariance matrix of batch i is computed as,

$$\mathbf{S}^{(i)} = \frac{\tilde{\mathbf{X}}^{(i)\mathrm{T}}\tilde{\mathbf{X}}^{(i)}}{K-L-1} \tag{15.2}$$

Afterwards, the sample covariance matrices of each batch can be merged into a pooled sample covariance matrix by considering their average, i.e.,

$$\mathbf{S}^{(pool)} = \frac{\left(K-L-1\right)\sum_{i=1}^{I}\mathbf{S}^{(i)}}{I\left(K-L\right)} \tag{15.3}$$

Given $\mathbf{S}^{(pool)}$, the standard PCA model can be obtained, as well as its monitoring statistics. However, as this procedure uses time-shifted replicates of the variables, it can only be employed after the first $l+1$ observations are available. Moreover, this method also requires the selection of the number of lags to construct the PCA model, for which the method of Ku *et al.* (1995) [27] has been employed. Even though Chen and Liu (2002) [154] do not mentioned any particular normalization of the data, the use of a batch-wise normalization is recommended [151].

Instead of monitoring the measured observations, Choi *et al.* (2008) [155] proposed the monitoring of their residuals with resort to an autoregressive PCA (ARPCA) method. Basically, this method fits a multivariate autoregressive model to the data and then uses it to compute one-step-ahead predictions of the current measurements. Afterwards, the residuals are determined and organized variable-wise prior to a PCA analysis.

Among these methodologies, the BDPCA scheme is highlighted, since this procedure is a direct extension of the continuous DPCA scheme of Ku *et al.* (1995) [27]. Therefore, it can be easily adapted to a new batch monitoring procedure based on DPCA-DR. In this case, the DPCA model can be constructed based on $\mathbf{S}^{(pool)}$ and by assumption that the current observations are missing, one-step-ahead predictions of the scores ($\hat{\mathbf{t}}$) and observations can be obtained (see Section 6.2.5). Then, the process state can be monitored by,

$$T^2_{PREV} = \left(\mathbf{t}-\hat{\mathbf{t}}\right)^{\mathrm{T}}\mathbf{S}^{-1}_{\mathbf{t}-\hat{\mathbf{t}}}\left(\mathbf{t}-\hat{\mathbf{t}}\right) \tag{15.4}$$

$$T^2_{RES} = \mathbf{r}^{\mathrm{T}}\mathbf{S}_{\mathbf{r}}^{-1}\mathbf{r} = \left(\mathbf{x} - \mathbf{P}\hat{\mathbf{t}}\right)^{\mathrm{T}} \mathbf{S}_{\mathbf{r}}^{-1} \left(\mathbf{x} - \mathbf{P}\hat{\mathbf{t}}\right) \tag{15.5}$$

where $\mathbf{S}_{\mathbf{t}-\hat{\mathbf{t}}}$ is the sample covariance matrix of the difference between the observed and estimated scores, $(\mathbf{t} - \hat{\mathbf{t}})$, and $\mathbf{S}_{\mathbf{r}}$ is the sample covariance matrix of the residuals in the reconstructed data.

As in the continuous case, these monitoring statistics are expected to present low autocorrelation levels and good detection abilities. Moreover, note that the batch DPCA-DR approach also shares some similarities with ARPCA, namely the fact that both procedures monitor the residuals of one-step-ahead predictions. However, while ARPCA requires a construction of a multivariate time series model, DPCA-DR does it implicitly within the DPCA model. Furthermore, it is also recommended the use of the proposed lag selection methods (Chapter 5), since they give better estimates of the lagged structure of processes than the lag selection method proposed by Ku *et al.* (1995) [27].

Recently, Van den Kerkhof *et al.* (2012) [151], compared the performances of MPCAB, MPCAV, BDPCA and ARPCA based on simulated data from an industrial-scale biochemical process for penicillin fermentation. From this study they observed that ARPCA outperformed the other methods in the fault detection task. However, for diagnosis proposes MPCAB gave the best results. Thus, they conclude that a combined approach for fault detection by ARPCA and fault diagnosis by MPCAB could result in a methodology with fast detection and more accurate diagnosis properties. As the presented batch version of DPCA-DR is, in some extent, comparable to both methodologies, it is considered pertinent to perform a similar comparison study in order to verify the true potential of DPCA-DR in the monitoring of batch processes.

15.2 Application of DPCA-DR to Time-varying Processes

Although the monitoring procedures studied in this work are able to cope with the dynamic and megavariate nature of industrial data, they still assume processes to be stationary. This assumption is often not met in practice. Non-stationarity can be caused by multiple factors such as equipment ageing, sensor drifting, catalyst deactivation, periodic cleaning, among others [63]. As these factors can affect any process, causing changes in

the mean, variance and correlation structure, there is a need to develop methods capable to efficiently handle time-varying features.

One way to deal with this situation is by the use of adaptive methods. These methods are extensions of the commonly used MSPC procedures, namely PCA and PLS models, which now can successively adapt to incoming data over time in order to cope with the time-varying nature of the underlying processes. This updated model can be either obtained by application of a recursive estimation of the model with the aid of a forgetting factor (RPCA), or by a moving window procedure (MWPCA), where the newest samples are included in the model, while the oldest ones are removed. The first approach was proposed by Li *et al.* (2000) [63], who employed a recursive estimation of the correlation matrix to update the PCA model. The eigenvalues and eigenvectors of this correlation matrix are then computed by Lanczos tridiagonalization and used to construct the updated PCA model. The use of moving windows to update the latent model is perhaps the most straightforward approach. However, it requires the storage of a sliding data window from which the model is successively computed. To improve its computational efficiency, and avoid the compromise between computational speed and modeling representativeness, Wang *et al.* (2005) [143] proposed a fast moving window PCA, which recursively updates the correlation matrix through a three step procedure. After updating, the MWPCA model is obtained in the same way as RPCA [63]. Even though RPCA and MWPCA are able to deal with megavariate processes and incorporate time-varying features, the dynamic behavior of the data is not captured in the model. Therefore, these approaches do not fully convey all the data features, which may seriously undermine their reliability and efficiency.

Another way to deal with non- stationarity is by explicitly modeling these characteristics by means of time series, for instance ARIMA (autoregressive integrated moving average) models [156]. The main idea behind these techniques is that the residuals obtained after pretreatment of the data with the estimated model, are time-invariant. Therefore, the traditional control charts can be applied to the residuals. However, the residuals obtained this way, can mask some types of faults due to the differencing procedure inherent to this approach [13]. To overcome this situation, a parametric model of the time dependency can be applied, e.g. [14]. Although this is an appealing solution, which proved to be successful in some applications, the time series modeling of large scale systems is a quite difficult task if possible at all, even for systems of low to moderate size (10-15 variables).

DPCA-DR intrinsically models a VAR process through the inclusion of time-shifted variables and uses it to obtain one-step-ahead predictions by application of missing data techniques. Thus, DPCA-DR is effectively monitoring the model residuals, which gives it the capacity to deal with time-varying processes, as it was documented in the preliminary results presented in [157]. One of the advantages of applying DPCA-DR for such task is the low autocorrelation of the monitoring statistics. This feature is not observed with RPCA or MWPCA and, for this reason, their control limits are difficult to compute, since the analytic expressions are usually invalid for the conditions found in real world scenarios. Moreover, they require an on-line updating of the control limits due to the changes in the model parameters. On the other hand, DPCA-DR uses a fixed model and control limits, which remain valid throughout the monitoring phase even if the in-control mean and variance are no longer equal to the ones used during the modeling stage.

To exemplify the previous reasoning, consider the case of a megavariate process given by the following latent variable model structure:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{15.6}$$

where $\mathbf{X}$ is an $(n \times m)$ matrix of measured variables, $\mathbf{T}$ is a $(n \times p)$ matrix of latent variables, $\mathbf{P}$ is an $(m \times p)$ matrix of orthogonal loadings and $\mathbf{E}$ is an $(n \times m)$ matrix of errors. Furthermore, assume that each latent variable follows an autoregressive integrated (ARI) process of order 1 given as [29],

$$t_i = t_{i-1} + \varphi\left(t_{i-1} - t_{i-2}\right) + \varepsilon_i \tag{15.7}$$

where t_i is the latent variable at instant i, φ is the ARI coefficient and ε_i is white noise.

For this example, the number of measured variables (m) was set to 100, with 5 latent variables (p) following independent ARI(1) processes with φ = 0.90. The $\mathbf{P}$ matrix was randomly generated, but forced to have orthogonal columns.

The sample autocorrelation of the DPCA-DR monitoring statistics in NOC, presented in Figure 15.2, shows that they are serial independent. Thus, an UCL can be easily determined based on the distribution of the NOC data. Afterwards, the monitoring scheme can be employed to new observations without the need to recursively update any of the models parameters. This situation is represented in Figure 15.3 for an independent data set composed by 500 NOC observations followed by another 500 observations with a step

deviation in one of the measurement readings (i.e., a step perturbation in one of the columns of matrix **X**, emulating a sensor failure).

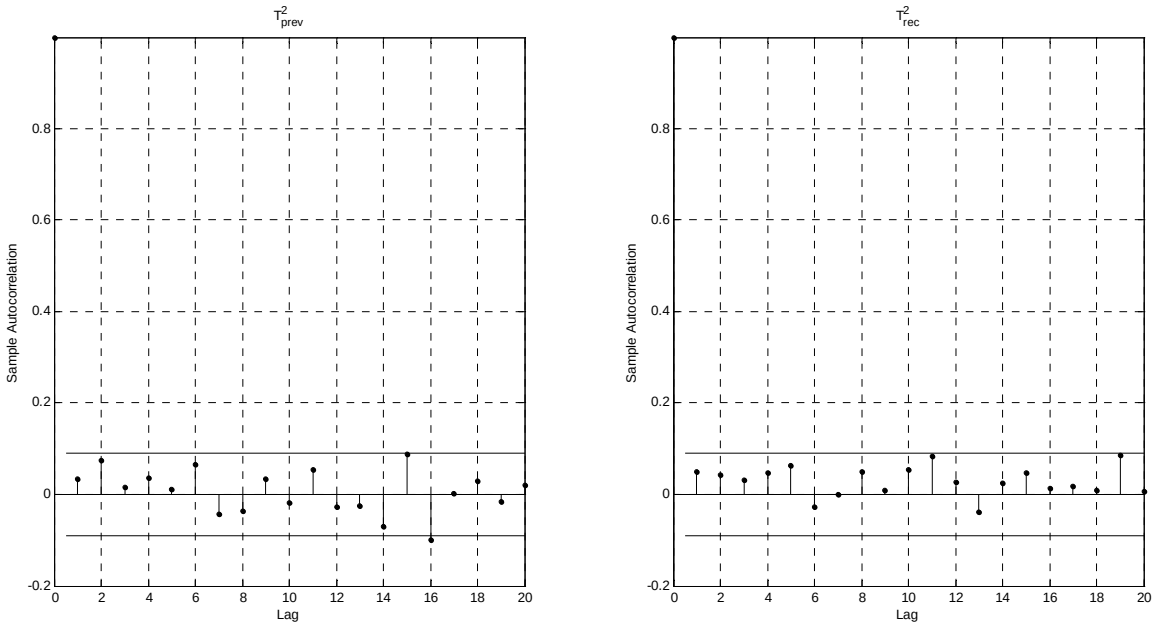


Figure 15.2 Sample autocorrelation of the DPCA-DR monitoring statistics for a megavariate ARI process with $\varphi = 0.90$ under NOC.

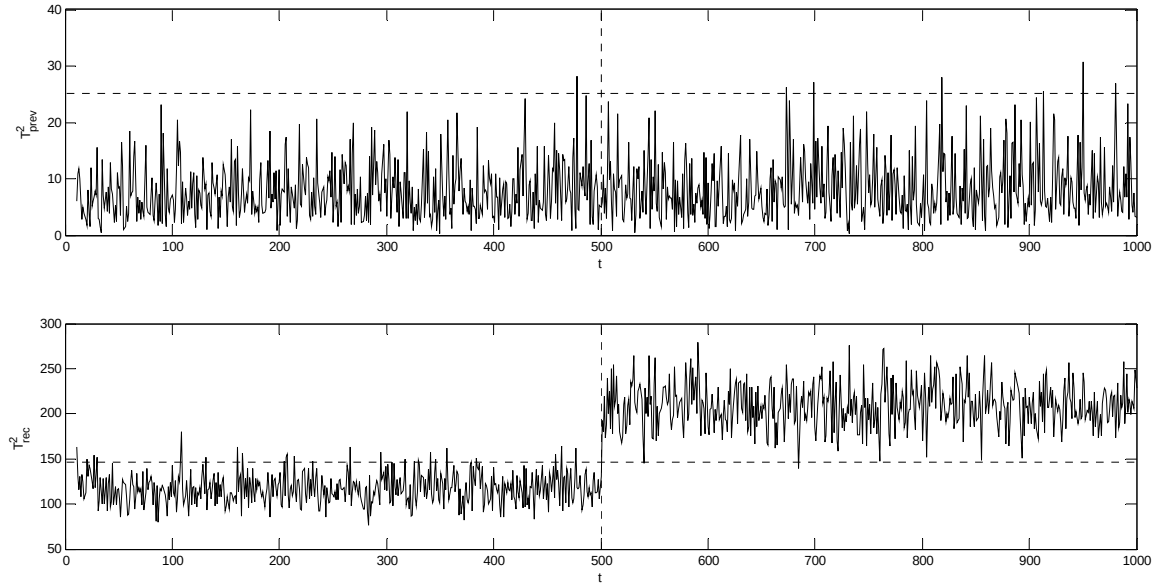


Figure 15.3 DPCA-DR monitoring statistics for a megavariate ARI process with $\varphi = 0.90$. The first 500 observations are collected under NOC and in the last 500 observations the system is subjected to a step deviation on a sensor reading with magnitude 5.

A similar result can be obtained when the underlying latent variables follow a first order integrated moving average (IMA) process [29],

$$t_i = t_{i-1} + \varepsilon_i - \psi\varepsilon_{i-1} \tag{15.8}$$

where t_i is the latent variable at instant i, ψ is the IMA coefficient and ε_i is white noise.

Using the same simulation settings as for the ARI process, with the latent variables determined by Equation (15.8) with $\psi = 0.90$, one obtains the autocorrelation plots for the DPCA-DR monitoring statistics presented in Figure 15.4. Again, they show low autocorrelation levels. On the other hand, from Figure 15.5 one can verify that they present also good detection capabilities for a step perturbation in one of the measured variables.

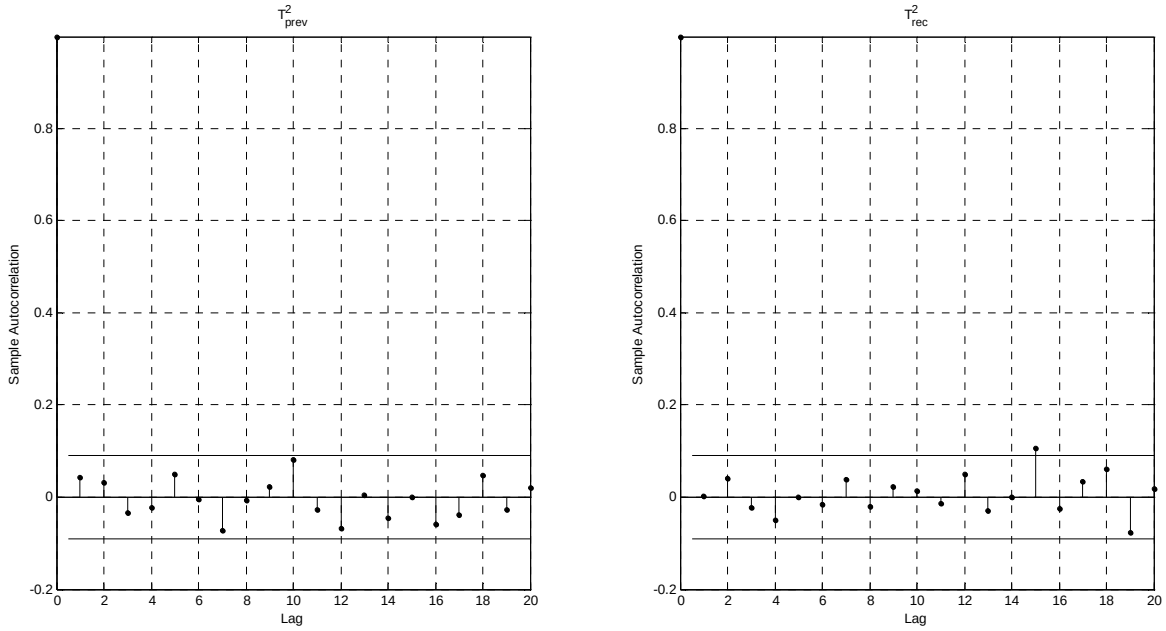


Figure 15.4 Sample autocorrelation of the DPCA-DR monitoring statistics for a megavariate IMA process with ψ = 0.90 under NOC.

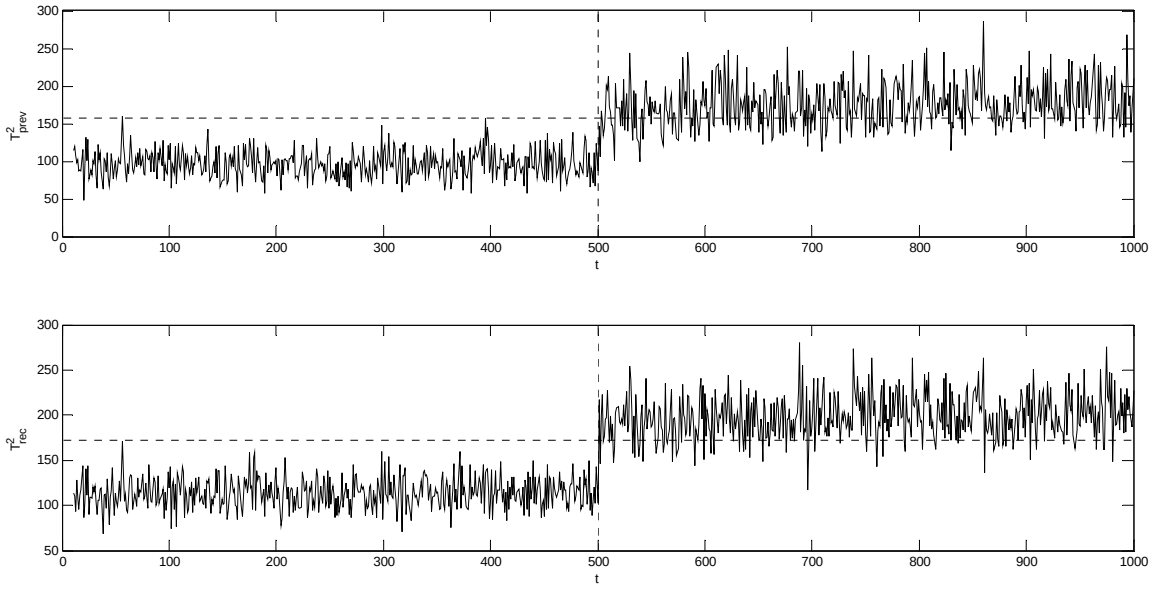


Figure 15.5 DPCA-DR monitoring statistics for a megavariate IMA process with $\psi = 0.90$. The first 500 observations are collected under NOC and for the last 500 observations the system is subjected to a step deviation on a sensor reading (measured variable) with magnitude 0.17.

In both ARI and IMA processes, the traditional PCA and DPCA gave erratic patterns, especially in the Hotelling's T^2 statistics, which clearly show their dependency with the process mean value. RPCA and MWPCA are less subjected to this feature. However they present problems in the selection of the adequate control limits, since the underlying assumptions of the analytic expressions are often violated. Furthermore, even when these methods are used with appropriate control limits they remain unable to detect some particular faults, such as measurement deviations (i.e., deviation in **X**). Amongst the monitoring schemes just discussed, DPCA-DR is the only one capable of detecting such faults, making it particularly suitable for monitoring megavariate processes with non-stationary features. Still, some improvements are required, since the DPCA-DR approach displays insensitivity to some deviations at the latent variables level, mostly due to its ability to estimate them well even if they depart from their nominal value (as it is a residual statistic, the effective departure from nominal conditions is not explicitly taken into account).

15.3 Fault Diagnosis Based on Partial Correlations

After a fault is signaled during the monitoring stage, the next stage is to identify its origin in order to proceed with the necessary corrective actions. Such diagnosis task can be performed by analysis of the partial correlation coefficients, due to their ability to eliminate the effect of the controlled variable upon the others. Therefore, if a change on the variables relationships occurs, it is expected that the partial correlations coefficients controlled by the variables associated with the root cause of the fault remain close to their normal values, since the source of variability is being removed in these circumstances.

Given this property, the following procedure is suggested to diagnose the fault origin based on the analysis of 1st order partial correlation coefficients. To better introduce the concepts behind this procedure, recall that a 1st partial correlation coefficient ($r_{i,j.k}$) is a measure of association between the *pair* of variables (*i*,*j*) when variable *k* is *controlled for*. Consequently, let us define the *pair* as any of the two variables to which the correlation is being measured and *control* as the corresponding controlled variable.

The first step of the proposed procedure involves the mapping of a ($m \times m$) matrix **D** (rows correspond to *control* variables and columns to *pair* variables) where the $d_{k,i}$ element corresponds to the number of times that the i^{th} variable is in a *pair* (*i*,*j*) of 1st order partial correlations controlled by *k* that is found to be above a threshold established from the analysis of their NOC distribution, for all *j* different from *i* and *k*, i.e.,

$$d_{k,i} = \sum_{j \neq i \neq k} f\left(r_{i,j.k}\right), \text{where} f\left(x\right) = \begin{cases} 1 & \text{if } x > \text{threshold} \\ 0 & \text{otherwise} \end{cases} \tag{15.9}$$

By doing so, it can be assessed which *control* variable leads to the lowest number of partial correlations above the threshold by considering each row of **D**. Likewise, one can also assess which *pair* variable is more frequently related with abnormal partial correlations, through the analysis of the columns of **D**. One way to do this is by computing the norm of each row and column of **D**. These quantities are defined as *control* and *pair* distance. Therefore, a variable with a small *control* distance is more related with the root cause, since when this variable is controlled for, the partial correlations tend to be closer to their normal operation values. Similarly, a variable with a large *pair* distance has suffered many changes in correlation with the other variables, and therefore it is expected to be more related to the root cause. As a result of this behavior, variable that comply to

both criteria simultaneously (i.e., lowest *control* distance and higher *pair* distance) are regarded as priority to further analysis. Then, the variables that only present a low *control* distance should be checked and afterwards the variables with high *pair* distance. This is translated in the following variables classification:

1. Variables with lowest *control* distance and highest *pair* distance are marked with code *red* – these are the best candidates for the faults origin, since they satisfy both criteria;

2. Variables with lowest *control* distance but with a *pair* distance lower than the maximum value are marked with code *orange* – when these variables are controlled a small number of partial correlations above the threshold is obtained but their correlation with the other variables does not changes significantly;

3. Variables with highest *pair* distance but with a *control* distance higher than the minimum value are marked with code *yellow* – these variables are related with partial correlations above the threshold, but when controlled, they do not eliminate the fault contribution.

These three decision rules give information about the fault origin with decreasing strength. Therefore, it is expected that most of the times, rule 1 (i.e., *red* variables) identifies the correct variable. The pseudo-code for this procedure is summarized in Table 15.1. In order to implement this algorithm, the maximum possible threshold is the UCL of the $R1MAX$ statistic, otherwise, it is not guaranteed that at least one partial correlation coefficient is above the threshold at the moment of the fault detection. This requirement also implies that the detection stage should be carried out with $R1MAX$.

To compare the diagnostic capabilities of this procedure, a similar scheme was developed for the marginal correlation. In this case, it is only analyzed the number of times that each variable is involved in a correlation coefficient that exceeds a given threshold based on their NOC distribution. Again, as this procedure is related with the correlation, the suitable monitoring statistic would be $R0MAX$, since this is the one that monitors changes in the marginal correlation coefficients.

Table 15.1 Pseudo-code for the fault diagnosis algorithm.

1. Construct a ($m \times m$) counting matrix **D** of variables involved in 1st order partial correlation coefficients exceeding the threshold:
    a. For $k = 1$ to m,
        i. Determine the pair or variables (i,j) that have a 1st partial correlation controlled by k ($r_{ij.k}$) above the threshold. Save this set of pair as P;
        ii. For $i = 1$ to m, count the number of times that the i^{th} variable is present in P and save this value in **D**(k,i).
2. Determine the vector of *control* distances (**c**) as the squared norm of each row of **D**, i.e. $\mathbf{c}(i) = \|\mathbf{D}(i,:)\|^2$;
3. Determine the vector of *pair* distances (**p**) as the squared norm of each column of **D**, i.e. $\mathbf{p}(i) = \|\mathbf{D}(:,i)\|^2$;
4. Determine the lowest *control* distance: $c_{min} = \min\{\mathbf{c}\}$;
5. Determine the highest *pair* distance: $p_{max} = \max\{\mathbf{p}\}$;
6. Classify each variable according to their priority:
    a. For $i = 1$ to m,
        i. If $\mathbf{c}(i) = c_{min}$ and $\mathbf{p}(i) = p_{max}$, mark variable as *red*;
        ii. I f $\mathbf{c}(i) = c_{min}$ and $\mathbf{p}(i) < p_{max}$, mark variable as *orange*;
        iii. If $\mathbf{c}(i) > c_{min}$ and $\mathbf{p}(i) = p_{max}$, mark variable as *yellow*.

For the sake of illustrating the potential of applying partial correlations in the diagnostic of the faults origin, the stationary linear system described on Section 10.2.2.1 was considered for the case of on-line monitoring. The diagnostics procedures based on the marginal correlation and partial correlations were compared by monitoring the system with the respective *RMAX* statistics and whenever these statistics signal an alarm, the respective diagnostic algorithm is run. This is done for three monitoring schemes: (i) marginal correlations of the original variables (monitored by $R0MAX_Z$), (ii) marginal correlations of the transformed T_{Net} variables (monitored by $R0MAX_{Net}$) and (iii) 1st order partial correlation of the transformed T_{Net} variables (monitored by $R1MAX_{Net}$).

As part of the computation of *RMAX* and of the diagnostic procedure, it is required that all correlation coefficients have the same probability to exceed a certain threshold. Therefore, it is necessary to determine the distribution of each correlation coefficient in order to normalize them to the same distribution. For the case of uncorrelated variables, this normalization step can be achieved by application of the equations described in Sections 10.1 and 11.1. For the original variables (which present correlation among each other) similar equations can be used [46]. However, for this comparison, the correlation

coefficients were normalized based on the empirical distribution obtained with 10000 NOC data sets with 200 observations each. The thresholds were then computed so that the corresponding *RMAX* statistics present a false alarm rate of 1%. By doing so, it is guaranteed that all monitoring schemes are comparable and that in the diagnose phase all correlation coefficients have the same probability to exceed the thresholds under NOC. Finally, as the thresholds were determined based on a window of 200 observations, a moving window approach of the same length is considered in the monitoring stage. However, it is worth noticing that an EWMA recursion with $\lambda = 0.01$ could also be used since these two approaches are equivalent (see Chapter 12).

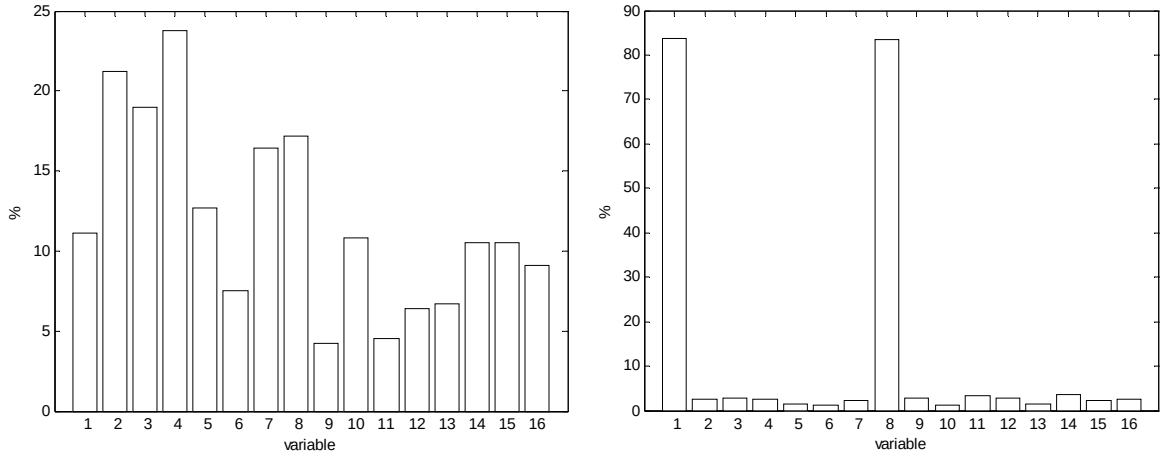
For a change in the relationship between g_1 and g_8 with a fault of magnitude $\delta = 1.10$ the percentage of times that each variable was identified as a possible root cause is represented in Figure 15.6 for the marginal correlations approaches and in Figure 15.7 for the partial correlations procedure. When marginal correlations of transformed variables are used (Figure 15.6 (b)), variable g_1 and g_8 are identified as the root cause in more than 80% of the times, while the other variables are only considered in about 3% of the times. The procedure based on partial correlations (Figure 15.7) identifies the same root causes in about 75% of the times in the *red* criteria. These results clearly show that both procedures are valid approaches to identify the fault's root case. Furthermore, it indicates that both marginal and partial correlations of the transformed variables have similar potential for such task. However, if the original variables are used, the marginal correlation would signal most of the variables as a potential cause (se Figure 15.6 (a)).

Similar results were obtained for the other faults and also for faults with different magnitudes. Therefore, the procedure based on marginal correlations of the transformed variables is regarded as the best method to diagnose the fault's root case. Note however, that since these variables are initially uncorrelated based on the process structure trough linear regressions, they are closely related with the concept of partial correlations. In other words, the marginal correlation of the new transformed variables can be interpreted as partial correlations of the original ones.

The base construction of this diagnosis procedure (hypotheses testing to each marginal or 1st order partial correlation coefficient) requires that, at the moment of detection, at least one correlation coefficient be above the selected thresholds. To ensure that this requirement is attained, it is recommended the use of the corresponding *RMAX* statistic,

since it implicitly performs this hypotheses test during the detection stage. The current monitoring statistics, such as M_1Z^2 and $MEWMA_t$ (see Section 4.2), do not guarantee this prerequisite. Therefore, besides leading to faster detection speeds (see Chapter 11), the *RMAX* procedure can also improve the diagnosis phase.

These preliminary results show that even though the diagnosis based on partial correlation has great potential, there are still some areas that need improvement. For instance, the above procedure often identifies two variables as a possible root cases. This can be reduced to a single root cause by consideration of the process causal network, by analysis of the variables values (namely mean and variance) at the detection moment or by re-estimation of the regression coefficients used in the SET.



(a) (b)

Figure 15.6 Percentage of times that each variable was considered as the faults' root cause by the marginal correlation procedure on fault A (i.e., on the relationship between variable 1 and 8), with $\delta = 1.10$, for the stationary linear system in a total of 1000 cases. The threshold used by both methods was set for the same statistical significance of 0.01: (a) marginal correlation of the original variables; (b) marginal correlation of the transformed variables.

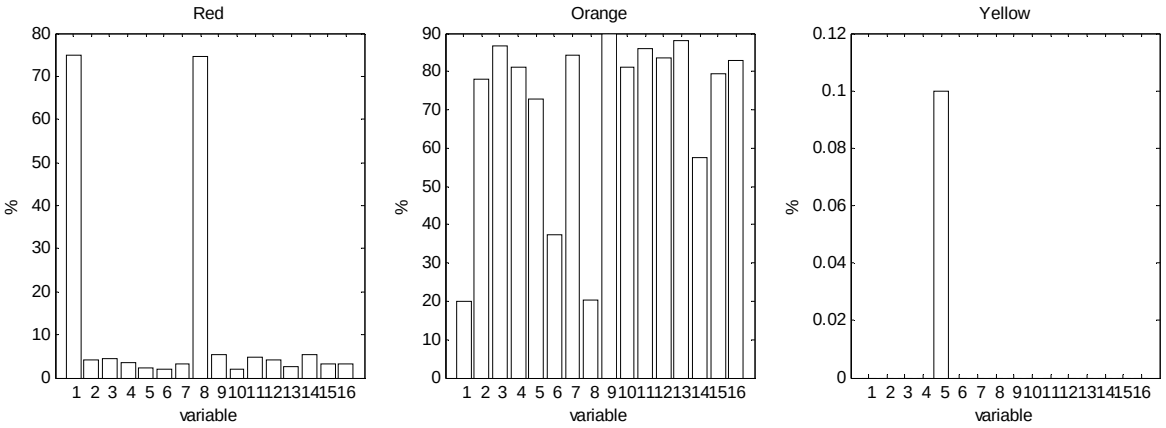


Figure 15.7 Percentage of times that each variable was considered as the faults' root cause by the partial correlations procedure on fault A (i.e., on the relationship between variable 1 and 8), with $\delta = 1.10$, for the stationary linear system in a total of 1000 cases. The threshold was set for the same statistical significance of 0.01.

References

1. Jackson, J.E., *A User's Guide to Principal Components*. 1991, New York: John Wiley & Sons, Inc.
2. Martens, H. and Naes, T., *Multivariate Calibration*. 1989, Chichester: Wiley.
3. Liu, X., Xie, L., Kruger, U., Littler, T., and Wang, S., *Statistical-Based Monitoring of Multivariate Non-Gaussian Systems.* AIChE Journal, 2008. **54**(9): p. 2379-2391.
4. Smilde, A.K., Bro, R., and Geladi, P., *Multi-way Analysis with Applications in the Chemical Sciences*. 2004, Chichester, UK: Wiley.
5. Reis, M.S., Saraiva, P.M., and Bakshi, B.R., *Multiscale statistical process control using wavelet packets.* AIChE Journal, 2008. **54**(9): p. 2366-2378.
6. Reis, M.S. and Saraiva, P.M., *Generalized Multiresolution Decomposition Frameworks for the Analysis of Industrial Data with Uncertainty and Missing Values.* Industrial & Engineering Chemistry Research, 2006. **45**: p. 6330-6338.
7. Reis, M.S. and Saraiva, P.M., *Multiscale Statistical Process Control of Paper Surface Profiles.* Quality Technology and Quantitative Management, 2006. **3**(3): p. 263-282.
8. Reis, M.S. and Saraiva, P.M., *Multiscale Statistical Process Control with Multiresolution Data.* AIChE Journal, 2006. **52**(6): p. 2107-2119.
9. Bakshi, B.R., *Multiscale PCA with Application to Multivariate Statistical Process Monitoring.* AIChE Journal, 1998. **44**(7): p. 1596-1610.
10. Reis, M.S. and Saraiva, P.M., *A Comparative Study of Linear Regression Methods in Noisy Environments.* Journal of Chemometrics, 2004. **18**(12): p. 526-536.
11. Reis, M.S. and Saraiva, P.M., *Heteroscedastic Latent Variable Modelling with Applications to Multivariate Statistical Process Control.* Chemometrics and Intelligent Laboratory Systems, 2006. **80**: p. 57-66.
12. Montgomery, D.C., *Introduction to Statistical Quality Control*. 6th ed. 2008: John Wiley & Sons.
13. Ketelaere, B.D., Mertens, K., Mathijs, F., Diaz, D.S., and Baerdemaeker, J.D., *Nonstationarity in statistical process control — issues, cases, ideas.* Applied Stochastic Models in Business and Industry, 2011. **27**(4): p. 367-376.
14. Mertens, K., Vaesen, I., Löffel, J., Kemps, B., Kamers, B., Zoons, J., Darius, P., Decuypere, E., De Baerdemaeker, J., and De Ketelaere, B., *An intelligent control chart for monitoring of autocorrelated egg production process data based on a synergistic control strategy.* Computers and Electronics in Agriculture, 2009. **69**(1): p. 100-111.
15. Jackson, J.E. and Mudholkar, G.S., *Control Procedures for Residuals Associated With Principal Component Analysis.* Technometrics, 1979. **21**(3): p. 341-349.
16. Jackson, J.E., *Quality Control Methods for Several Related Variables.* Technometrics, 1959. **1**(4): p. 359-377.
17. Kresta, J.V., MacGregor, J.F., and Marlin, T.E., *Multivariate Statistical Monitoring of Process Operating Performance.* The Canadian Journal of Chemical Engineering, 1991. **69**(1): p. 35-47.
18. MacGregor, J.F. and Kourti, T., *Statistical process control of multivariate processes.* Control Engineering Practice, 1995. **3**(3): p. 403-414.
19. Kourti, T. and MacGregor, J.F., *Process analysis, monitoring and diagnosis, using multivariate projection methods.* Chemometrics and Intelligent Laboratory Systems, 1995. **28**: p. 3-21.

20. Jolliffe, I.T., *Principal Component Analysis*. 2nd ed. 2002, New York: Springer.
21. Nelson, P.R.C., Taylor, P.A., and MacGregor, J.F., *Missing data methods in PCA and PLS: Score calculations with incomplete observations.* Chemometrics and Intelligent Laboratory Systems, 1996. **35**: p. 45-65.
22. Arteaga, F. and Ferrer, A., *Dealing with missing data in MSPC: several methods, different interpretations, some examples.* Journal of Chemometrics, 2002. **16**: p. 408-418.
23. Walczak, B. and Massart, D.L., *Dealing with missing data: Part I* Chemometrics and Intelligent Laboratory Systems, 2001. **58**(1): p. 15-27.
24. Walczak, B. and Massart, D.L., *Dealing with missing data: Part II* Chemometrics and Intelligent Laboratory Systems, 2001. **58**(1): p. 29-42.
25. Naes, T., Isaksson, T., Fearn, T., and Davies, T., *A User-Friendly Guide to Multivariate Calibration and Classification*. 2002, Chichester (UK): NIR Publications.
26. Härdle, W. and Simar, L., *Applied Multivariate Statistical Analysis*. 2nd ed. 2007, Berlin: Springer-Verlag.
27. Ku, W., Storer, R.H., and Georgakis, C., *Disturbance detection and isolation by dynamic principal component analysis.* Chemometrics and Intelligent Laboratory Systems, 1995. **30**(1): p. 179-196.
28. Ljung, L., *System Identification - Theory for the User*. 2nd ed. 1999, Upper Saddle River, NJ: Prentice Hall.
29. Box, G.E.P., Jenkins, G.M., and Reinsel, G.C., *Time Series Analysis: Forecasting and Control*. 3rd ed. 1994, New Jersey: Prentice-Hall.
30. Helland, I.S., *On the Structure of Partial Least Squares Regression.* Commun. Statist.-Simula., 1988. **17**(2): p. 581.
31. Höskuldsson, A., *PLS Regression Methods.* Journal of Chemometrics, 1988. **2**: p. 211-228.
32. Wold, S., Sjöström, M., and Eriksson, L., *PLS-Regression: A Basic Tool of Chemometrics.* Chemometrics and Intelligent Laboratory Systems, 2001. **58**: p. 109-130.
33. Geladi, P. and Kowalski, B.R., *Partial Least-Squares Regression: a Tutorial.* Analytica Chimica Acta, 1986. **185**: p. 1-17.
34. Little, R.J.A. and Rubin, D.B., *Statistical Analysis with Missing Data*. 2nd ed. Wiley Series in Probability and Statistics2002, Hoboken (NJ): Wiley.
35. Tsay, R.S., *Analysis of Financial Time Series*. 2nd ed. 2005, New Jersey: John Wiley & Sons, Inc.
36. Lütkepohl, H., *New Introduction to Multiple Time Series Analysis*. 2005, Berlin: Springer-Verlag.
37. Chaterfield, C., *The Analysis of Time Series: An Introduction*. 5 ed. 1996, New York: Chapman & Hall.
38. Cinar, A., Palazoglu, A., and Kayihan, F., *Chemical Process Performance Evaluation*. 2007, Boca Raton: CRC Press.
39. Montgomery, D.C. and Runger, G.C., *Applied Statistics and Probability for Engineers*. 2003: John Wiley & Sons, Inc.
40. Shewhart, W.A., *Economic Control of Quality of Manufactured Product*. 1931, New York: D. Van Nostrand Company, Inc.
41. Page, E.S., *Continuous inspection schemes.* Biometrika, 1954. **41**: p. 100-114.
42. Roberts, S.W., *Control Charts Tests Based on Geometric Moving Averages.* Technometrics, 1959. **1**(3): p. 239-250.

43. Hotelling, H., *The Generalization of Student's Ratio.* The Annals of Mathematical Statistics, 1931. **2**(3): p. 360-378.
44. Crosier, R.B., *Multivariate Generalizations of Cumulative Sum Quality-Control Schemes.* Technometrics, 1988. **30**(3): p. 291-303.
45. Lowry, C.A., Woodall, W.H., Champ, C.W., and Rigdon, S.E., *A Multivariate Exponentially Weighted Moving Average Control Chart.* Technometrics, 1992. **34**(1): p. 46-53.
46. Anderson, T.W., *An Introduction to Multivariate Statistical Analysis*. 3 ed. 2003, New Jersey: Wiley.
47. Tracy, N.D., Young, J.C., and Mason, R.L., *Multivariate Control Charts for Individual Observations.* Journal of Quality Technology, 1992. **24**(2): p. 88-95.
48. Montgomery, D.C., *Introduction to Statistical Quality Control*. 5th ed. 2005: Wiley.
49. Woodall, W.H. and Ncube, M.M., *Multivariate CUSUM Quality-Control Procedures.* Technometrics, 1985. **27**(3): p. 285-292.
50. Pignatiello, J.J. and Runger, G.C., *Comparisons of Multivariate CUSUM Charts.* Journal of Quality Technology, 1990. **22**(3): p. 173-186.
51. Hawkins, D.M., Choi, S., and Lee, S., *A General Multivariate Exponentially Weighted Moving-Average Control Chart.* Journal of Quality Technology, 2007. **39**(2): p. 118-125.
52. Kramer, H.G. and Schmid, L.V., *Ewma charts for multivariate time series.* Sequential Analysis, 1997. **16**(2): p. 131-154.
53. Yumin, L., *An improvement for mewma in multivariate process control.* Computers & Industrial Engineering, 1996. **31**(3–4): p. 779-781.
54. Bersimis, S., Psarakis, S., and Panaretos, J., *Multivariate statistical process control charts: an overview.* Quality and Reliability Engineering International, 2007. **23**(5): p. 517-543.
55. Qin, S.J., *Statistical process monitoring: basics and beyond.* Journal of Chemometrics, 2003. **17**: p. 480-502.
56. MacGregor, J.F., Jaeckle, C., Kiparissides, C., and Koutoudi, M., *Process Monitoring and Diagnosis by Multiblock PLS Methods.* AIChE Journal, 1994. **40**(5).
57. Nomikos, P. and MacGregor, J.F., *Monitoring Batch Processes Using Multiway Principal Component Analysis.* AIChE Journal, 1994. **40**(8): p. 1361-1375.
58. Aradhye, H.B., Bakshi, B.R., Strauss, R.A., and Davis, J.F., *Multiscale SPC using wavelets: Theoretical analysis and properties.* AIChE Journal, 2003. **49**(4): p. 939-958.
59. Daubechies, I., *Orthonormal bases of compactly supported wavelets.* Communications on Pure and Applied Mathematics, 1988. **41**(7): p. 909-996.
60. Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P., *Numerical Recipes: The Art of Scientific Computing*. 3rd ed. 2007: Cambridge University Press.
61. MacGregor, J.F., Jaechle, C., Kiparissides, C., and Koutoudi, M., *Process Monitoring and Diagnosis by Multiblock PLS Methods.* AIChE Journal, 1994. **40**(5): p. 826-838.
62. Wise, B.M. and Gallagher, N.B., *The process chemometrics approach to process monitoring and fault detection.* Journal of Process Control, 1996. **6**(6): p. 329-348.
63. Li, W., Yue, H.H., Valle-Cervantes, S., and Qin, S.J., *Recursive PCA for adaptive process monitoring.* Journal of Process Control, 2000. **10**: p. 471-486.

64. Negiz, A. and Çinar, A., *PLS, balanced, and canonical variate realization techniques for identifying VARMA models in state space.* Chemometrics and Intelligent Laboratory Systems, 1997. **38**: p. 209-221.
65. Negiz, A. and Çinar, A., *Statistical Monitoring of Multivariable Dynamic Processes with State-Space Models.* AIChE Journal, 1997. **43**(8): p. 2002-2020.
66. Ghute, V.B. and Shirke, D.T., *A Multivariate Synthetic Control Chart for Process Dispersion.* Quality Technology & Quantitative Management, 2008. **5**(3): p. 271-288.
67. Yeh, A.B., Lin, D.K.J., and McGrath, R.N., *Multivariate Control Charts for Monitoring Covariance Matrix: A Review.* Quality Technology & Quantitative Management, 2006. **3**(4): p. 415-436.
68. Abbasi, B., Niaki, S., Abdollahian, M., and Hosseinifard, S., *A transformation-based multivariate chart to monitor process dispersion.* The International Journal of Advanced Manufacturing Technology, 2009. **44**(7): p. 748-756.
69. Yen, C.-L., Shiau, J.-J.H., and Yeh, A.B., *Effective Control Charts for Monitoring Multivariate Process Dispersion.* Quality and Reliability Engineering International, 2012. **28**(4): p. 409-426.
70. Alt, F.B., *Multivariate Quality Control*, in *The Encyclopedia of Statistical Sciences*, N.L. Johnson, S. Kotz, and C.R. Read, Editors. 1984, Wiley: New York. p. 110-122.
71. Aparisi, F., Jabaloyes, J., and Carrión, A., *Generalized Variance Chart Design With Adaptive Sample Sizes. The Bivariate Case.* Communications in Statistics - Simulation and Computation, 2001. **30**(4): p. 931–948.
72. Djauhari, M.A., *Improved Monitoring of Multivariate Process Variability.* Journal of Quality Technology, 2005. **37**(1): p. 32-39.
73. Guerrero-Cusumano, J.-L., *Testing variability in multivariate quality control: A conditional entropy measure approach.* Information Sciences, 1995. **86**(1–3): p. 179-202.
74. Djauhari, M.A., Mashuri, M., and Herwindiati, D.E., *Multivariate Process Variability Monitoring.* Communications in Statistics - Theory and Methods, 2008. **37**(11): p. 1742-1754.
75. Alt, F.B. and Smith, N.D., *Multivariate process control*, in *Handbook of Statistics*, P.R. Krishnaiah and C.R. Rao, Editors. 1988, Elsevier. p. 333-351.
76. Levinson, W.A., Holmes, D.S., and Mergen, A.E., *Variation Charts for Multivariate Processes.* Quality Engineering, 2002. **14**(4): p. 539–545.
77. Yen, C.-L. and Shiau, J.-J.H., *A Multivariate Control Chart for Detecting Increases in Process Dispersion.* Statistica Sinica, 2010. **20**: p. 1683-1707.
78. Tang, P.F. and Barnett, N.S., *Dispersion Control for Multivariate Processes.* Australian Journal of Statistics, 1996. **38**(3): p. 235-251.
79. Yeh, A.B., Huwang, L., and Wu, C.-W., *A multivariate EWMA control chart for monitoring process variability with individual observations.* IIE Transactions, 2005. **37**(11): p. 1023-1035.
80. Huwang, L., Yeh, A.B., and Wu, C.-W., *Monitoring Multivariate Process Variability for Individual Observations.* Journal of Quality Technology, 2007. **39**(3): p. 258-278.
81. Reynolds, M.R. and Cho, G.-Y., *Multivariate Control Charts for Monitoring the Mean Vector and Covariance Matrix.* Journal of Quality Technology, 2006. **38**(3): p. 230-252.
82. Hawkins, D.M. and Maboudou-Tchao, E.M., *Multivariate Exponentially Weighted Moving Covariance Matrix.* Technometrics, 2008. **50**(2): p. 155-166.

83. Kramer, C.Y. and Jensen, D.R., *Fundamentals of Multivariate Analysis: Part II. Inference about Two Treatments.* Journal of Quality Technology, 1969. **1**(3): p. 189-204.
84. Costa, A. and Machado, M., *A new chart based on sample variances for monitoring the covariance matrix of multivariate processes.* The International Journal of Advanced Manufacturing Technology, 2009. **41**(7): p. 770-779.
85. Machado, M., Costa, A., and Marins, F., *Control charts for monitoring the mean vector and the covariance matrix of bivariate processes.* The International Journal of Advanced Manufacturing Technology, 2009. **45**(7): p. 772-785.
86. Costa, A.F.B. and Machado, M.A.G., *A control chart based on sample ranges for monitoring the covariance matrix of the multivariate processes.* Journal of Applied Statistics, 2011. **38**(2): p. 233-245.
87. Memar, A.O. and Niaki, S.T.A., *New control charts for monitoring covariance matrix with individual observations.* Quality and Reliability Engineering International, 2009. **25**(7): p. 821-838.
88. Bodnar, O., Bodnar, T., and Okhrin, Y., *Surveillance of the covariance matrix based on the properties of the singular Wishart distribution.* Computational Statistics & Data Analysis, 2009. **53**(9): p. 3372-3385.
89. Hunter, J.S., *The Exponentially Weighted Moving Average.* Journal of Quality Technology, 1986. **18**(4): p. 203-210.
90. Vermaat, M.B., Does, R.J.M.M., and Bisgaard, S., *EWMA Control Chart Limits for First- and Second-Order Autoregressive Processes.* Quality and Reliability Engineering International, 2008. **24**: p. 573–584.
91. Harris, T.J. and Ross, W.H., *Statistical Process Control Procedures for Correlated Observations.* The Canadian Journal of Chemical Engineering, 1991. **69**: p. 48-57.
92. Montgomery, D.C. and Mastrangelo, C.M., *Some Statistical Process Control Methods for Autocorrelated Data.* Journal of Quality Technology, 1991. **23**(3): p. 179-204.
93. Lee, C., Choi, S.W., and Lee, I.-B., *Sensor fault identification based on time-lagged PCA in dynamic processes.* Chemometrics and Intelligent Laboratory Systems, 2004. **70**(2): p. 165-178.
94. Li, W. and Qin, S.J., *Consistent dynamic PCA based on errors-in-variables subspace identification.* Journal of Process Control, 2001. **11**(6): p. 661-678.
95. Russell, E.L., Chiang, L.H., and Braatz, R.D., *Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis.* Chemometrics and Intelligent Laboratory Systems, 2000. **51**(1): p. 81-93.
96. Makis, V., Wu, J., and Gao, Y., *An application of DPCA to oil data for CBM modeling.* European Journal of Operational Research, 2006. **174**(1): p. 112-123.
97. Perry, M.A., Wynn, H.P., and Bates, R.A., *Principal components analysis in sensitivity studies of dynamic systems.* Probabilistic Engineering Mechanics, 2006. **21**(4): p. 454-460.
98. Brillinger, D.R., *A frequency approach to the techniques of principal components, factor analysis and canonical variates in the case of stationary time series*, in *Royal Statistical Society Conference*, 1964: Cardiff.
99. Breitung, J. and Eickmeier, S., *Dynamic factor models.* Allgemeines Statistisches Archiv, 2006. **90**(1): p. 27-42.

100. Mancino, M.E. and Renò, R., *Dynamic Principal Component Analysis of Multivariate Volatility via Fourier Analysis.* Applied Mathematical Finance, 2005. **12**(2): p. 187-199.
101. Tsung, F. and Apley, D.W., *The dynamic T^2 chart for monitoring feedback-controlled processes.* IIE Transactions, 2002. **34**(12): p. 1043-1053.
102. Wachs, A. and Lewin, D.R., *Improved PCA Methods for Process Disturbance and Failure Identification.* AIChE Journal, 1999. **45**(8): p. 1688-1700.
103. Guerfel, M., Othman, K.B., and Benjeb, M. *On the structure determination of a dynamic PCA model using sensitivity of fault detection*. in *7th IFAC International Symposium on Advanced Control of Chemical Processes*. 2009. Istanbul.
104. Pesarin, F. and Salmaso, L., *Permutation Tests for Complex Data: Theory, Applications and Software*. 2010: John Wiley & Sons Ltd.
105. Luyben, W.L., *Simple method for tuning SISO controllers in multivariable systems.* Industrial & Engineering Chemistry Process Design and Development, 1986. **25**(3): p. 654-660.
106. Roffel, B. and Betlem, B., *Process Dynamics and Control: Modeling for Control and Prediction*. 2006, Padstow: John Wiley & Sons Ltd.
107. Downs, J.J. and Vogel, E.F., *A plant-wide industrial process control problem.* Computers and Chemical Engineering, 1993. **17**(3): p. 245-255.
108. Russell, E.L., Chiang, L.H., and Braatz, R.D., *Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes*. 2000: Springer.
109. Simoglou, A., Martin, E.B., and Morris, A.J., *Dynamic Multivariable Statistical Process Control using Partial Least Squares and Canonical Variate Analysis.* Computers and Chemical Engineering, 1999. **Supplement** (S277-S280).
110. Treasure, R.J., Kruger, U., and Cooper, J.E., *Dynamic multivariate statistical process control using subspace identification.* Journal of Process Control, 2004. **14**(3): p. 279-292.
111. Komulainen, T., Sourander, M., and Jämsä-Jounela, S.-L., *An online application of dynamic PLS to a dearomatization process.* Computers & Chemical Engineering, 2004. **28**(12): p. 2611-2619.
112. Kaspar, M.H. and Harmon Ray, W., *Dynamic PLS modelling for process control.* Chemical Engineering Science, 1993. **48**(20): p. 3447-3461.
113. Lakshminarayanan, S., Shah, S.L., and Nandakumar, K., *Modeling and Control of Multivariable Processes: Dynamic PLS Approach.* AIChE Journal, 1997. **43**(9): p. 2307-2322.
114. Schneider, T. and Neumaier, A., *Algorithm 808: ARFIT—A Matlab Package for the Estimation of Parameters and Eigenmodes of Multivariate Autoregressive Models.* ACM Transactions on Mathematical Software, 2001. **27**(1): p. 58–65.
115. Wood, R.K. and Berry, M.W., *Terminal composition control of a binary distillation column.* Chemical Engineering Science 1973. **28**(9): p. 1707-1717.
116. Lin, W., Qian, Y., and Li, X., *Nonlinear dynamic principal component analysis for on-line process monitoring and diagnosis.* Computers and Chemical Engineering, 2000. **24**: p. 423-429.
117. Chiang, L.H. and Braatz, R.D., *Process monitoring using causal map and multivariate statistics: fault detection and identification.* Chemometrics and Intelligent Laboratory Systems, 2003. **65**(2): p. 159-178.
118. Odiowei, P.P. and Cao, Y., *State-space independent component analysis for nonlinear dynamic process monitoring.* Chemometrics and Intelligent Laboratory Systems, 2010. **103**(1): p. 59–65.

119. Yin, S., Ding, S.X., Haghani, A., Hao, H., and Zhang, P., *A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process.* Journal of Process Control, 2012. **22**(9): p. 1567-1581.
120. Lyman, P.R. and Georgakis, C., *Plant-wide control of the Tennessee Eastman problem.* Computers and Chemical Engineering, 1995. **19**(3): p. 321-331.
121. Reis, M., *Data-Driven Multiscale Monitoring, Modelling and Improvement of Chemical Processes*, 2005, University of Coimbra: Coimbra.
122. Alawi, A., Morris, A.J., and Martin, E.B., *Statistical performance monitoring using state space modelling and wavelet analysis*, in *Computer Aided Chemical Engineering*, P. Luis and E. Antonio, Editors. 2005, Elsevier. p. 1459-1464.
123. Rosen, C., *A Chemometric Approach to Process Monitoring and Control - With Applications to Wastewater Treatment Operation*, 2001, Lund University: Lund.
124. Sokal, R. and Rohlf, F.J., *Biometry: The Principles and Practice of Statistics in Biological Research*. 3rd ed. 1995, New York: W. H. Freeman & Co.
125. Bauer, M., Cox, J.W., Caveness, M.H., Downs, J.J., and Thornhill, N.F., *Finding the Direction of Disturbance Propagation in a Chemical Process Using Transfer Entropy.* Control Systems Technology, IEEE Transactions on, 2007. **15**(1): p. 12-21.
126. Yuan, T. and Qin, S.J. *Root Cause Diagnosis of Plant-Wide Oscillations Using Granger Causality*. in *8th IFAC International Symposium on Advanced Control of Chemical Processes*. 2012. Singapore.
127. Hawkins, D.M., *Regression Adjustment for Variables in Multivariate Quality Control.* Journal of Quality Technology, 1993. **25**(3): p. 170-182.
128. Melissa, A.S., Raghuraj, K.R., and Lakshminarayanan, S., *Partial correlation metric based classifier for food product characterization.* Journal of Food Engineering, 2009. **90**(2): p. 146-152.
129. Rao, R. and Lakshminarayanan, S., *Variable interaction network based variable selection for multivariate calibration.* Analytica Chimica Acta, 2007. **599**(1): p. 24-35.
130. Fuente, A.D.L., Bing, N., Hoeschele, I., and Mendes, P., *Discovery of meaningful associations in genomic data using partial correlation coefficients.* Bioinformatics, 2004. **20**(18): p. 3565-3574.
131. Reverter, A. and Chan, E.K.F., *Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks.* Bioinformatics, 2008. **24**(21): p. 2491-2497.
132. Rao, K.R. and Lakshminarayanan, S., *Partial correlation based variable selection approach for multivariate data classification methods.* Chemometrics and Intelligent Laboratory Systems, 2007. **86**(1): p. 68-81.
133. Pellet, J.-P. and Elisseeff, A., *A partial correlation-based algorithm for causal structure discovery with continuous variables*, in *Proceedings of the 7th international conference on Intelligent data analysis*, 2007, Springer-Verlag: Ljubljana, Slovenia. p. 229-239.
134. Tamada, Y., Kim, S., Bannai, H., Imoto, S., Tashiro, K., Kuhara, S., and Miyano, S., *Estimating gene networks from gene expression data by combining Bayesian network model with promoter element detection.* Bioinformatics, 2003. **19**(1): p. ii227-i236.
135. Ottestad, P., *Component Analysis: An Alternative System.* International Statistical Review, 1975. **43**(1): p. 83-107.

136. Kalisch, M. and Bühlmann, P., *Robustification of the PC-Algorithm for Directed Acyclic Graphs.* Journal of Computational and Graphical Statistics, 2008. **17**(4): p. 773-789.
137. Frank, P.M., *Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy - a survey and some new results.* Automatica, 1990. **26**(3): p. 459-474.
138. Isermann, R., *Model-based fault-detection and diagnosis – status and applications.* Annual Reviews in Control, 2005. **29**(1): p. 71-85.
139. Isermann, R., *Process Fault Detection Based on Modeling and Estimation Methods - A Survey.* Automatica, 1984. **20**(4): p. 387-404.
140. Wilson, E.B. and Hilferty, M.M., *The distribution of chi-square*, in *Proceedings of the National Academy of Sciences of the United States of America*, 1931. p. 684-688.
141. Yue, H.H. and Qin, S.J., *Reconstruction-Based Fault Identification Using a Combined Index.* Industrial & Engineering Chemistry Research, 2001. **40**(20): p. 4403-4414.
142. Jöbses, I.M.L., Hiemstra, H.C.H., and Roels, J.A., *Fermentation kinetics of Zymomonas mobilis near zero growth rate.* Biotechnology and Bioengineering, 1987. **29**(4): p. 502-512.
143. Wang, X., Kruger, U., and Irwin, G.W., *Process Monitoring Approach Using Fast Moving Window PCA.* Industrial & Engineering Chemistry Research, 2005. **44**(15): p. 5691-5702.
144. Blumenfeld, D.E., *Operations Research Calculations Handbook*. 2001: CRC Press.
145. Papoulis, A., *Probability, Random Variables and Stochastic Processes*. 1991: McGraw-Hill.
146. Teppola, P. and Minkkinen, P., *Wavelet–PLS regression models for both exploratory data analysis and process monitoring.* Journal of Chemometrics, 2000. **14**(5-6): p. 383-399.
147. Misra, M., Yue, H.H., Qin, S.J., and Ling, C., *Multivariate process monitoring and fault diagnosis by multi-scale PCA.* Computers & Chemical Engineering, 2002. **26**(9): p. 1281-1293.
148. Rosen, C. and Lennox, J.A., *Multivariate and multiscale monitoring of wastewater treatment operation.* Water Research, 2001. **35**(14): p. 3402-3410.
149. Alcala, C.F. and Qin, S.J., *Reconstruction-based contribution for process monitoring.* Automatica, 2009. **45**(7): p. 1593-1600.
150. Nomikos, P. and MacGregor, J.F., *Multivariate SPC Chart for Monitoring Batch Processes.* Technometrics, 1995. **37**(1): p. 41-59.
151. Van den Kerkhof, P., Gins, G., Vanlaer, J., and Van Impe, J.F.M., *Dynamic model-based fault diagnosis for (bio)chemical batch processes.* Computers & Chemical Engineering, 2012. **40**(0): p. 12-21.
152. García-Muñoz, S., Kourti, T., and MacGregor, J.F., *Model Predictive Monitoring for Batch Processes.* Industrial & Engineering Chemistry Research, 2004. **43**(18): p. 5929-5941.
153. Wold, S., Kettaneh, N., Fridén, H., and Holmberg, A., *Modelling and diagnostics of batch processes and analogous kinetic experiments.* Chemometrics and Intelligent Laboratory Systems, 1998. **44**(1–2): p. 331-340.
154. Chen, J. and Liu, K.-C., *On-line batch process monitoring using dynamic PCA and dynamic PLS models.* Chemical Engineering Science, 2002. **57**(1): p. 63-75.

155. Choi, S.W., Morris, J., and Lee, I.-B., *Dynamic model-based batch process monitoring.* Chemical Engineering Science, 2008. **63**(3): p. 622-636.
156. Box, G.E.P. and Paniagua-Quiñones, C., *Two Charts: Not One.* Quality Engineering, 2007. **19**(2): p. 93-100.
157. Ketelaere, B.D., Rato, T., Reis, M., and Schmitt, E. *A Systematic Comparison of Statistical Process Monitoring Methods for High-dimensional, Time-dependent Processes*. in *13th Annual Conference of the European Network for Business and Industrial Statistic*. 2013. Ankara.
158. Seborg, D., Edgar, T., and Mellichamp, D., *Process Dynamics and Control*. 1989, New York: J. Wiley & Sons.

Appendices

Apendix A. Mathematical Model of an Endothermic CSTR under Feedback Control

The set of differential equations used to describe a continuous stirred-tank reactor (CSTR) are here derived based on first principles. In this model, the thermo-physical properties of the materials are constant (compounds A and B and thermic fluid), namely their densities and heat capacities. The description of the symbols and their respective values are presented in Table A.1.

Table A.1 Parameters of the CSTR model.

Symbol	Description	Values	Unit
h	Reactor level	0.45	m
C_A	Outlet concentration of compound A	4.47	kmol m^{-3}
T	Reactor temperature	318.48	K
T_j	Outlet temperature of thermal fluid	330.42	K
F_{jv}	Valve outlet volumetric flow rate of the thermal fluid	0.008	m^3 min^{-1}
F_0	Inlet volumetric flow of the reactor feed stream	0.008	m^3 min^{-1}
C_{A0}	Concentration of compound A in the reactor feed stream	6	kmol m^{-3}
T_0	Temperature of the reactor feed stream	299.15	K
T_{j0}	Inlet temperature of the thermal fluid	348.15	K
F_j	Valve inlet volumetric flow rate of the thermal fluid	0.008	m^3 min^{-1}
c	Discharge coefficient	0.012	$m^{5/2}$ min^{-1}
D	Reactor diameter	0.50	m
V_b	Volume at the reactor's base	0.005	m^3
k_0	Pre-exponential factor	2×10^{10}	min^{-1}
E_a	Activation energy	8677.0	J
ρ	Density	1000	kg m^{-3}
C_p	Heat capacity	4184	J kg^{-1} K^{-1}
ΔH_r	Enthalpy of reaction	-30.0	J $kmol^{-1}$
U	Heat transfer coefficient	48000	J min^{-1} m^{-2} K^{-1}
ρ_j	Density of the heating fluid	950	kg m^{-3}
C_{pj}	Heat capacity of the heating fluid	4800	J kg^{-1} K^{-1}
V_j	Heating jacket volume	0.025	m^3
τ_v	Valve's time constant	0.0833	min

A.1 Global Mass Balance to the CSTR

The global mass balance to the reactor is given by:

$$\frac{dm}{dt} = \rho F_0 - \rho F \tag{A.1}$$

$$\Leftrightarrow \frac{d(\rho V)}{dt} = \rho F_0 - \rho F \Leftrightarrow$$
$$\Leftrightarrow \rho \frac{dV}{dt} = \rho F_0 - \rho F \Leftrightarrow$$

$$\Leftrightarrow \frac{dV}{dt} = F_0 - F \tag{A.2}$$

As the reactor volume is given by:

$$V = V_b + A_b h$$

it is obtained that $dV = A_b dh$ (assuming that V_b is constant). Thus,

$$A_b \frac{dh}{dt} = F_0 - F \Leftrightarrow$$

$$\Leftrightarrow \frac{dh}{dt} = \frac{F_0 - F}{A_b} \tag{A.3}$$

where

$$F = c\sqrt{h}\ . \tag{A.4}$$

A.2 Partial Mass Balance to Compound A

The mass balance of compound A is given by,

$$\frac{dm_A}{dt} = M_A C_{A0} F_0 - M_A C_A F - rVM_A \tag{A.5}$$

In this case, the reaction kinetics is considered to be described according to the Arrhenius law, defined as,

$$r = k_0 e^{-\frac{E_a}{RT}} C_A \tag{A.6}$$

Expanding Equation (A.5),

$$\frac{d\left(M_A C_A V\right)}{dt} = M_A C_{A0} F_0 - M_A C_A F - rVM_A \Leftrightarrow$$
$$\Leftrightarrow \frac{d\left(C_A V\right)}{dt} = C_{A0} F_0 - C_A F - rV \Leftrightarrow$$

$$\Leftrightarrow V\frac{dC_A}{dt} + C_A \frac{dV}{dt} = C_{A0} F_0 - C_A F - rV \tag{A.7}$$

By replacement of Equation (A.2) into Equation (A.7),

$$V\frac{dC_A}{dt} + C_A\left(F_0 - F\right) = C_{A0} F_0 - C_A F - rV \Leftrightarrow$$
$$\Leftrightarrow V\frac{dC_A}{dt} = F_0\left(C_{A0} - C_A\right) - rV \Leftrightarrow$$

$$\Leftrightarrow \frac{dC_A}{dt} = \frac{F_0\left(C_{A0} - C_A\right)}{V} - r \tag{A.8}$$

A.3 Global Energy Balance to the Reactor

The total energy of a thermodynamic system is the result of three components: (i) kinetic energy, (ii) potential energy and (iii) internal energy. As the contributions of the kinetic energy and potential energy are relatively small in comparison to the internal energy in the system, only the contributions of the latter are considered in this model [158]. Moreover, for pure liquids, at moderate and constant pressures, the internal energy is approximately equal to the enthalpy, from where it can be shown that [158]:

$$d\hat{E} = dH = C_p dT \tag{A.9}$$

where $\hat{E}$ is the energy per unit mass, H is the enthalpy per unit mass, C_p is the heat capacity at constant pressure and T is the temperature.

From the integration of Equation (A.9), it is obtained that:

$$H - H_{ref} = C_p\left(T - T_{ref}\right) \tag{A.10}$$

Without loss of generality, it can be assumed that at the reference temperature, the reference enthalpy is zero [158]. Taking in account this considerations, the energy balance becomes:

$$\frac{dE}{dt} = \rho F_0 H_0 - \rho F H + rV\left(-\Delta H_r\right) + Q \tag{A.11}$$

where $Q = UA_t\left(T_j - T\right)$.

The heat transfer area (A_t) is given as the area of the reactor's base plus the lateral area in contact with the heating fluid. Therefore, A_t is given as,

$$A_t = A_b + \pi D h \Leftrightarrow$$

$$\Leftrightarrow A_t = \pi\frac{D^2}{4} + \pi D h \Leftrightarrow$$

$$\Leftrightarrow A_t = \pi\frac{D}{2}\left(\frac{D}{2} + 2h\right) \tag{A.12}$$

Expanding Equation (A.11),

$$\frac{d\left(mH\right)}{dt} = \rho F_0 H_0 - \rho F H + rV\left(-\Delta H_r\right) + Q \Leftrightarrow$$

$$\Leftrightarrow H\frac{dm}{dt} + m\frac{dH}{dt} = \rho F_0 H_0 - \rho F H + rV\left(-\Delta H_r\right) + Q \tag{A.13}$$

By replacement of Equation (A.1) into Equation (A.13),

$$H\left(\rho F_0 - \rho F\right) + m\frac{dH}{dt} = \rho F_0 H_0 - \rho F H + rV\left(-\Delta H_r\right) + Q \Leftrightarrow$$

$$\Leftrightarrow m\frac{dH}{dt} = \rho F_0\left(H_0 - H\right) + rV\left(-\Delta H_r\right) + Q \Leftrightarrow$$

$$\Leftrightarrow \rho V C_p \frac{dT}{dt} = \rho C_p F_0\left(T_0 - T\right) + rV\left(-\Delta H_r\right) + Q \Leftrightarrow$$

$$\Leftrightarrow \frac{dT}{dt} = \frac{F_0\left(T_0 - T\right)}{V} + \frac{r\left(-\Delta H_r\right)}{\rho C_p} + \frac{Q}{\rho V C_p} \tag{A.14}$$

A.4 Global Energy Balance to the Heating Jacket

Taking into consideration the assumptions described in Appendix A.3, the following energy balance is obtained for the heating jacket:

$$\frac{dE_j}{dt} = \rho_j F_j H_{j0} - \rho_j F_j H_j - Q \Leftrightarrow$$

$$\Leftrightarrow \frac{d\left(m_j H_j\right)}{dt} = \rho_j F_j H_{j0} - \rho_j F_j H_j - Q \Leftrightarrow$$

$$\Leftrightarrow H_j\frac{dm_j}{dt} + m_j\frac{dH_j}{dt} = \rho_j F_j H_{j0} - \rho_j F_j H_j - Q \Leftrightarrow$$

$$\Leftrightarrow \rho_j V_j C_{pj}\frac{dT_j}{dt} = \rho_j C_{pj} F_0\left(T_{j0} - T_j\right) - Q \Leftrightarrow$$

$$\Leftrightarrow \frac{dT_j}{dt} = \frac{F_j\left(T_{j0} - T\right)}{V_j} - \frac{Q}{\rho_j V_j C_{pj}} \tag{A.15}$$