Chemical Engineering Science 163 (2017) 223–233

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

Markovian and Non-Markovian sensitivity enhancing transformations for process monitoring

Tiago J. Rato, Marco S. Reis *

CIEPQPF, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima, 3030-790 Coimbra, Portugal

HIGHLIGHTS

- Process diagnosis requires causal information to be successful completed.
- Current process monitoring methods are based on acausal models.
- We propose a plug-in approach that integrates causal information into standard process monitoring.
- Smearing-out effect is reduced without compromising fault detection and diagnosis.
- Results obtained recommend the use of the static Non-Markovian SET as pre-processing for the Hotelling's T^2 methodology.

ARTICLE INFO

Article history:
Received 9 July 2016
Received in revised form 23 November 2016
Accepted 19 January 2017
Available online 1 February 2017

Keywords:
Principal component analysis
Variables transformation
Fault detection
Fault diagnosis

ABSTRACT

Process monitoring is a key activity in modern industrial processes. Even though abnormality detection can be rather effectively done with resort to acausal correlation models of the variables normal operating conditions associations, fault diagnosis and troubleshooting do require causal information. In this article, we propose a new plug-in approach that brings the causal network structure into a classical monitoring scheme based on the Hotelling's T^2 methodology. The modular plug-in nature associated to a well-known monitoring scheme aims at facilitating the access to the benefits of using more information about the system structure in fault analysis and diagnosis. The pre-processing module consists of a Sensitivity Enhancing Transformation (SET) that incorporates the network structure inferred from normal operation data, which has recently conducted to significant improvements for monitoring the correlation structure of industrial processes. Additionally, we consider both Markovian and Non-Markovian network structures in the development of the SET. The proposed methodology was tested with two simulated case studies (a CSTR and the Tennessee Eastman benchmark) and compared with several alternative approaches. The results obtained recommend the use of the static Non-Markovian SET as pre-processing for the Hotelling's T^2 methodology.

1. Introduction

Statistical Process Monitoring is an activity of critical importance in today's highly complex industrial processes. Not only this activity should provide a reliable and robust source of earlier alarms of evolving process upsets or abrupt faults, but also to assist, as much as possible, in the subsequent stage of fault diagnosis and troubleshooting. This last aspect of a process monitoring platform has been growing in importance, as the time spent in diagnosing the abnormality is usually much larger than the delay in detecting its occurrence. The importance of this difference, which can span orders of magnitude (for instance detection can be done in seconds or minutes, while a proper diagnosis may require hours to days), has not yet been fully acknowledged by the scientific community, taking into consideration the dominant emphasis in metrics related to speed of detection in comparison studies, such as the average run length (ARL) or the related average time to signal (ATS), when compared to accuracy rates in diagnosis. We believe it would be highly desirable, especially to practitioners and process owners, that both perspectives are considered on an equal footing, reflecting their actual role in industrial process monitoring.

However, fault detection and fault diagnosis, even though highly connected in the workflow of process monitoring, are fundamentally different tasks. Fault detection aims at signaling any significant deviation from normal operation behavior, by analyzing

* Corresponding author.
E-mail address: marco@eq.uc.pt (M.S. Reis).

http://dx.doi.org/10.1016/j.ces.2017.01.047
0009-2509/

data routinely collected from the process. Fault diagnosis, on the other hand, is dedicated to finding out the origin of the abnormality, so that the root cause can be isolated and analysed for its criticality, from which a decision will be made on the actions to take (e.g., stop the process immediately; continue with the process without any changes; accommodate the fault and continue with the operation; etc.). Underlying these two distinct goals, are different modelling and analysis premises. For process detection, a normal operation conditions (NOC) model describing the main associations and regularities between process variables is sufficient. It does not have to be causal, i.e., it does not have to provide the directionality of effects propagation. Instead, it just has to faithfully reflect the normal structure of associations, in an efficient and robust way, so that any existing discrepancy can be rapidly detected and signaled. This is the true nature of the NOC models used in multivariate and megavariate (or high-dimensional) statistical process control, where the multivariate normal distribution or principal component analysis (PCA) provide the acausal (also referred as non-causal) probabilistic model structures that support fault detection. On the other hand, fault diagnosis and troubleshooting, imply starting from the observed effects or symptoms, and then tracing back the causes that may have originated them. For such, some type of causal mapping, of a quantitative or qualitative nature, is required to conduct the analysis. Relying on acausal NOC models for fault diagnosis, is prone to ambiguous and sometime erroneous findings. This is a consequence of questioning a model that only contains acausal association information, expecting that it will reveal, without modification, cause-effect relationships justifying the abnormal observed patterns. A well-known manifestation of this misuse is the smearing-out effect in PCA-based statistical process monitoring schemes, which causes non-faulty variables to have significant contributions due to fault propagation (Van den Kerkhof et al., 2013). The different nature of detection and diagnosis tasks and associated modelling requirements is another aspect that has also been quite often overlooked in process monitoring studies, a possible consequence of which being a certain underappreciation of the limitations of some current data-driven diagnosis tools.

In order to circumvent the diagnosis limitations of classic statistical process monitoring (SPM) schemes, several structured approaches have been proposed in the literature, namely using transfer entropy (Bauer et al., 2007; Shu and Zhao, 2013), time delay analysis (Bauer and Thornhill, 2008), Granger causality (Yuan and Qin, 2012) and causal maps (Chiang and Braatz, 2003; Cheng et al., 2008; Thambirajah et al., 2009). These structured SPM approaches, insert, through different means and with different extents, information about the causal structure of the system in the monitoring procedure. They usually end up constituting a new monitoring scheme, with novel statistics and algorithms. This represents a challenge to practitioners willing to improve the monitoring activities in their processes, for which they must first fully realize the conceptual benefits of the new proposals, and secondly they must be able to program the algorithms and implement the methods using the available resources. Given the limited nature of such resources in practice, most proposals remain untested and vastly unexplored. Therefore, in this article, we opt to implement a plug-in approach to structure integration in process monitoring, where a specific pre-processing module brings the causal structure required for diagnosis. This module is then easily integrated (plugged-in) in a standard monitoring scheme based on the Hotelling's T^2 statistic. The module regards a specific type of pre-processing, called Sensitivity Enhancing Transformation (SET), that incorporates the network structure of the process variables. SET's have already conducted to significant improvements for monitoring the correlation structure of industrial processes (Rato and Reis, 2014a,b, 2015a,b,c). They consist in whitening each variable by regressing it onto the set of variables with a direct effect on it (its causal parents), following a Markovian approach. In the present work, we also consider for the first time to regress each target variable on all variables with indirect, but causally related effects, leading to a Non-Markovian modelling approach.

In the following subsections, the current fault detection and diagnose modes of PCA and related techniques that are relevant for the purposes of this article, are described. Then, the proposed SPM methodology is presented in detail. In the next sections, this approach is tested and comparatively assessed against current benchmarks, and the results obtained discussed. The paper ends with a conclusions section summarizing the main findings of this work.

2. PCA-based process monitoring for detection and diagnosis and the MTY decomposition

PCA is the *de facto* standard platform for performing multi- and megavariate (or high-dimensional) statistical process monitoring (SPM) of industrial processes and any new proposal in this domain should consider not only this technique as benchmark, but also its related developments for handling more properly any particular characteristics of the system under consideration, such as dynamics, non-linearity, non-stationarity, etc., in other to conduct a fair, consistent and unbiased comparison. In this subsection the background of process monitoring via PCA modelling is provided. Focus is given to the cases of static and dynamic PCA as these are the application scenarios more often encountered in practice. An introduction to contribution plots for the purpose of fault diagnosis will be also provided. Finally, a multivariate diagnosis procedure proposed by Mason, Tracy and Young (henceforth designated by MTY) (Mason et al., 1995) is reviewed. Even though this methodology, without modification, is of limited application in large scale (megavariate or high-dimensional) scenarios, it is opportune to consider it here for better contextualizing the proposed methodology, which shares some algorithmic similarities and concepts.

2.1. Principal component analysis

Principal component analysis (PCA) is a latent variable methodology focused on reducing the data dimensionality by finding a subspace around which the majority of data variability is concentrated. By doing so, the original data matrix, $\mathbf{X}$, with n observations and m variables is decomposed as:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{1}$$

where $\mathbf{T}_{n\times p}$ is the matrix of PCA scores, $\mathbf{P}_{m\times p}$ is the matrix with the PCA loadings, and $\mathbf{E}_{n\times m}$ is the residual matrix. p stands for the number of retained principal components (PC). 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 the analysis. The most common pre-processing is to center all variables to zero mean and scale them to unit variance (defined as "autoscaling"), but many other approaches are available (Martens and Naes, 1989; Naes et al., 2002). In Section 3, another pre-processing procedure will be put forward, which is based on the causal network linking the observed variables.

By application of PCA, the original data is effectively decomposed into two complementary subspaces, which will be monitored separately. As the number of retained principal components is low, say p, and they are uncorrelated by design, the Hotelling's T^2 procedure can be applied without limitations to monitor the PCA subspace (Jackson, 1959; Jackson and Mudholkar, 1979). Therefore, the following monitoring statistic is applied for monitoring the PCA subspace: