Contents lists available at ScienceDirect

Chemical Engineering Research and Design

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

Detecting faults in process systems with singular spectrum analysis

S. Krishnannair[a,b], C. Aldrich[b,c,*], G.T. Jemwa[b]

[a] University of Zululand, Private Bag X1001, Kwadlangezwa 3886, South Africa
[b] Department of Process Engineering, University of Stellenbosch, Private Bag X1, Matieland 7602, Stellenbosch, South Africa
[c] Department of Mining Engineering and Metallurgical Engineering, Western Australian School of Mines, Curtin University, GPO Box U1987, 6845, WA, Australia

ARTICLE INFO

Article history:
Received 31 January 2016
Received in revised form 7 July 2016
Accepted 17 July 2016
Available online 22 July 2016

Keywords:
Process monitoring
Singular spectrum analysis
Dissimilarity

ABSTRACT

In this study, process monitoring based on signal decomposition by use of singular spectrum analysis (SSA) is considered. SSA makes use of adaptive basis functions to decompose a time series into multiple components that may be periodic, aperiodic or random. Two variants of SSA are considered in this investigation. In the first, the conventional approach is used based on latent variables extracted from the covariances of the lagged trajectory matrix of the process variables. The second approach is identical to the first approach, except that the covariances of the lagged trajectory matrices are replaced by Euclidean distance dissimilarities to decompose the variables into additive components. These components are subsequently monitored and the merits of the two approaches are considered on the basis of two case studies using simulated nonlinear data and data from the benchmark Tennessee Eastman process.

1. Introduction

Several key drivers in modern chemical and metallurgical industries have led to increased interest and adoption of data driven innovations in process control and monitoring technology. These include enhanced process safety, process efficiency, and improved performance in achieving product quality specifications through the management of process variation. As a result, large volumes of data are routinely logged and stored in data warehouses, and can be used in fault detection and diagnostic tasks in process operations. Owing to the highly correlated nature of the observed process variables, the challenge associated with these process monitoring schemes is the development of efficient modeling techniques that can account for the relationships between process variables in order to detect faults in the process operations (Chiang et al., 2001; Qin, 2003).

The most common multivariate statistical process monitoring (MSPM) methods are principal component analysis (PCA) and partial least squares (PLS) (Qin, 2003). However, PCA and PLS assume linear relationships between variables and Gaussian latent variables. Hence, PCA models can give misleading information when applied to highly nonlinear systems with large numbers of variables to monitor.

As a result, a plethora of nonlinear extensions to PCA have been proposed to handle nonlinear processes, such as principal curves (Dong and McAvoy, 1996; Harkat et al., 2003; Shi et al., 2013), independent component analysis (Kano et al., 2003; Wang and Shi, 2010; Hsu et al., 2010), kernel methods (Jemwa and Aldrich, 2005; Zhang and Qin, 2007; Deng et al.,

* *Corresponding author at*: Department of Mining Engineering and Metallurgical Engineering, Western Australian School of Mines, Curtin University, GPO Box U1987, 6845, WA, Australia.
E-mail address: chris.aldrich@curtin.edu.au (C. Aldrich).
http://dx.doi.org/10.1016/j.cherd.2016.07.014

2013) and neural networks (Jia et al., 1998; Aradhye et al., 2002) to name but a few. Another practical limitation in the application of these multivariate statistical methods is that process operations are nonstationary and dynamic in nature, which may not be captured adequately with a single scale representation of the measurements. Likewise, classical MSPM can also fail when measurements are autocorrelated (Ku et al., 1995).

In some instances, a more effective approach would take advantage of multiple representations of measurements, that is, the representation of features that occur with different localization in time, space and frequency (Aradhye et al., 2003). Such multiscale process control methods combining PCA and wavelets have been proposed by Kosanovich and Piovoso (1997) and Bakshi (1998). In general, nonlinear and/or multiscale extensions to PCA can be expected to give improved performance in targeted use-cases. Even so, these methods require large data sets to calibrate reference models, and the added computational cost is challenging (Yunus and Zhang, 2014).

In other approaches, decomposition of the variables is aimed at reconstruction of a common subspace, a specific subspace and a residual subspace. Monitoring is consequently performed in every subspace (Zhang et al., 2013; Zhang and Li, 2013; Zhang and Zhang, 2014). This can be particularly effective when monitoring multimodal processes and the transitions between modes.

To better summarize multivariate data in a lower dimension than the conventional approach inspired the development of MSPM frameworks using multidimensional scaling (Auret and Aldrich, 2010; Yunus, 2012; Yunus and Zhang, 2014). In that framework any measure of inter-distance scaling can be modified and converted into a set of correlation matrix based eigenvectors, which could subsequently be used in PCA to project the variables to a lower-dimensional plane of latent variables.

The same approach is followed in this paper via the use of squared dissimilarity matrices in lieu of correlation matrices as used conventionally in PCA. This approach can also handle nonlinear correlations in the data like other nonlinear methods in process monitoring (Yunus and Zhang, 2010).

In the area of chemical process monitoring, singular spectrum analysis (SSA) has become a promising tool to prefilter process data before the application of multivariate statistical process control (MSPC) tools for fault detection in the process (Aldrich and Barkhuizen, 2003a, 2003b, 2003c; Barkhuizen and Aldrich, 2003, 2004). Moreover, a multimodal method for process monitoring based on SSA has previously been proposed for the simultaneous extraction of complex trends and periodicities with varying amplitude in the process data (Krishnannair, 2010; Aldrich et al., 2007).

In this paper, these studies are extended by considering two variants of SSA as a basis for process monitoring. The first is based on the use of classical SSA (cSSA) (Golyandina et al., 2001), while the second comprises modification of the SSA decomposition stage based on dissimilarity scales, instead of the variance-covariance association to quantify the relationships between variables. This modification provides modeling frameworks with SSA by using Euclidean distances as the dissimilarity measure to develop scores in a reduced dimensional space and will be referred to as dissimilarity based SSA (DSSA).

The application of DSSA is demonstrated using simulated data and the Tennessee Eastman Challenge process and the paper is organized as follows. Section 3 presents the methodology of basic SSA. In Section 3, the concept of classical multidimensional scaling (CMDS) is introduced as a building block for DSSA. The concept of DSSA is subsequently outlined in Section 4. Sections 5 and 6 briefly explain the procedures of cSSA and DSSA. The monitoring results of the proposed DSSA and its comparison with cSSA and PCA using simulated data and the Tennessee Eastman process are discussed in Section 7. Section 8 concludes the paper.

2. Singular spectrum analysis

In this section the basic steps involved in SSA is briefly explained with its applications in time series analysis.

2.1. *Singular spectrum analysis of univariate time series*

Singular spectral time series analysis prefilters the original time series into a sum of series that contains components such as a trend, periodic or quasi-periodic components, or noise. This is done by the singular value decomposition (SVD) of a trajectory or lagged covariance matrix obtained from the original time series, followed by reconstruction of the series using subsets of eigenfunctions and corresponding principal components. Standard PCA is performed on the trajectory matrix of the time series and hence the mathematical and statistical properties of PCA extend to SSA.

The time series is first embedded into an M-dimensional space known as the *trajectory matrix*. Singular value decomposition is then applied to decompose the trajectory matrix into a sum of elementary matrices. Subsequently, the elementary matrices that contribute to the norm of the original matrix are grouped, with each group giving an approximation of the original matrix. Finally, the smoothed approximation of the time series is recovered by diagonal averaging of the elementary matrices obtained from the decomposing the trajectory matrix.

An outline of the basic SSA methodology (Golyandina et al., 2001) and the procedural steps involved in SSA are discussed in more detail below (Jemwa and Aldrich, 2006; Hassani and Thomakos, 2010; Krishnannair, 2010).

2.1.1. *Step 1: Embedding*

Given a time series $\mathbf{y}_N = (y(1), y(2), \ldots, y(N))$, that is embedded with a window of length $2 \le M \le N$ to construct $K = N - M + 1$ lagged vectors: $\mathbf{x}_i \in \mathbb{R}^M$

$$\mathbf{x}_i = [y(i), y(i+1), \ldots, y(i+M-1)]^T, \quad 1 \le i \le K = N - M + 1 \tag{1}$$

These embedded vectors $\mathbf{x}_i$ are then collected into an array or multidimensional time series known as a trajectory matrix: $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K)$

$$\mathbf{X} = (x_{ij})_{i,j=1}^{K,M} = \begin{bmatrix} y(1) & y(2) & \cdots & y(M) \\ y(2) & y(3) & & y(M+1) \\ \vdots & & \ddots & \vdots \\ y(K) & y(K+1) & \cdots & y(N) \end{bmatrix} \tag{2}$$

The trajectory matrix $\mathbf{X} \in \mathbb{R}^{K \times M}$ is a Hankel matrix, that is the matrix has equal elements on the antidiagonals, where $i + j =$ constant, or

$$x_{ij} = x(i + j - 1), \quad 1 \le i \le K, \quad 1 \le j \le M \tag{3}$$