Journal of Process Control xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Process Control

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

A geometric method for batch data visualization, process monitoring and fault detection

Ray Wang [a], Thomas F. Edgar [a,b], Michael Baldea [a,c,*], Mark Nixon [d], Willy Wojsznis [d], Ricardo Dunia [d]

[a] McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, TX 78712, United States
[b] Energy Institute, The University of Texas at Austin, Austin, TX 78712, United States
[c] Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX 78712, United States
[d] Emerson Process Management, Round Rock, TX 78681, United States

ARTICLE INFO

Article history:
Received 14 June 2016
Received in revised form 15 May 2017
Accepted 15 May 2017
Available online xxx

Keywords:
Batch processes
Process monitoring
Data visualization
Fault detection

ABSTRACT

Batch processing is used extensively in the production of high value products, and there are strong economic incentives for developing methodologies for ensuring the successful completion of batches via process monitoring and fault detection. Building on our previous work using time-explicit Kiviat diagrams for continuous processes, this paper introduces data visualization, data-driven process monitoring and fault detection for batch systems. Handling batch data, including unfolding and alignment are addressed as well. The proposed methodology is demonstrated on data obtained from a benchmark bioreactor simulator and a semiconductor etching process.

1. Introduction

Batch processing is extensively used in the production of high-value, typically low-volume materials including pharmaceuticals and microelectronics. The economic cost of lost process performance is typically high, and has motivated extensive research in batch process monitoring, fault detection and control. Data-driven methods have an important role in this area. The operating data sets used for model-building purposes comprise measurements of the process variables collected from many batch runs, and are thus often of the "big data" class.

A primary challenge for batch process monitoring techniques is to define "normal" operation, i.e., the "yardstick" by which operating cycles are evaluated and can be identified as successful or failed. The states of a batch system are constantly changing and the system may go through multiple phases involving multiple unit operations [20]. As a consequence, there is no nominal steady-state to refer to (as in the case of continuous systems). Rather, the control, monitoring and performance evaluation methodologies developed for batch processes must account for their transient, dynamic nature.

Secondly, batch durations are not fixed, meaning that the duration of the phases of a batch process can change between runs, which in turn hinders the use of models that assume that the process will or should be in a specific state at a particular time instant.

Finally, there are two different types of variability to consider: intra-batch variations (variations that occur within a single batch run) and inter-batch (or run-to-run) variations (variations that occur across runs). Equivalently, there are two time scales to consider: a fast time scale which is the order of magnitude of the rate of evolution of a batch, and a slower time scale that spans the time horizon of a production campaign that involves multiple batches.

The body of literature on data-driven batch process monitoring (which, amongst others, attempts to address the above challenges) is vast [6]. One of the most widespread approaches is multi-way principal component analysis (MPCA), introduced by Wold et al. [25] and popularized by Nomikos and MacGregor [18,19]. The key idea behind MPCA is the batch-wise unfolding of the three-dimensional (Time × Variables × Batches) batch data matrix into a specific two-dimensional matrix that captures the variation of the data across batches, followed by conducting principal component analysis (PCA) on these "unfolded" data. Several variations of MPCA methods were introduced later: Yoo et al. [28] and Lee et al. [15] used multi-model MPCA and kernel MPCA to monitor the different phases of the batch operation. Li et al. used a recursive approach with PCA to perform adaptive monitoring of batch processes to

* Corresponding author at: McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, TX 78712, United States.
E-mail address: mbaldea@che.utexas.edu (M. Baldea).

http://dx.doi.org/10.1016/j.jprocont.2017.05.011

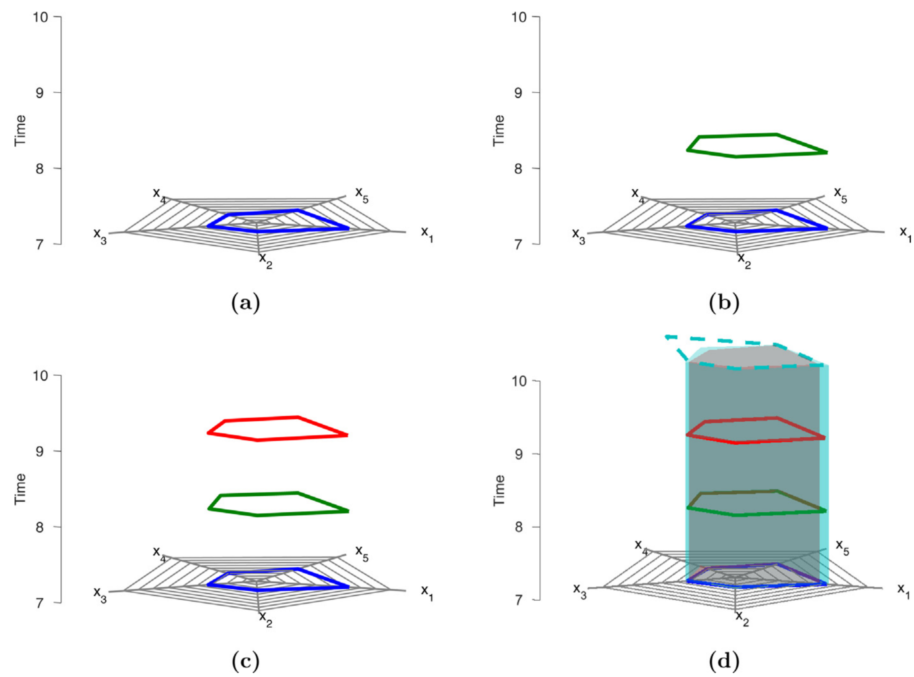

Fig. 1. Representing multi-dimensional data in time-resolved cylindrical coordinates. A five-dimensional data set with one-minute sampling time is used for illustration purposes. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

update the model [16]. Yoo et al. also explored using independent component analysis (ICA) to analyze unfolded batch data [27]. Hu et al. [9] and Kourti et al. [13,12] have also applied the same unfolding method, but using partial least squares (PLS) instead to relate the batch quality variables to the batch process variables. Wurl et al. used PLS to specifically monitor batch startup [26].

In a different vein, Meng et al. [17] proposed the use of parallel factor analysis to perform online batch monitoring, with the benefit that the method can handle three-dimensional data directly without the need for unfolding. A comparison of the MPCA and PARAFAC approaches can be found in Westerhuis et al. [23].

An alternate approach to online multivariate statistical process monitoring can be found in Ündey et al. [21]. Ündey et al. employ variable-wise unfolding for online process monitoring as it can handle unequal batch lengths easily and avoids the need to account for uneven batches during data processing. Later works by Yu et al. [29] and Camacho et al. [3] further develop the use of variable-wise unfolding in online process monitoring.

Dynamic principal component analysis (DPCA), initially developed for the analysis of transient continuous processes [14] has also been applied to batch data, in this case to capture intra-batch variations [4].

In this paper, we present a novel approach for monitoring and fault detection in batch processes, based on visualization of batch operating data. Our work is based on the framework we recently introduced for visualization and visualization-based fault detection for continuous systems [22]. The novelty of the present contribution consists of an extension of our previous results to account for the inherent transients present in the operating cycles of batch systems, while dealing with variability in cycle durations. We propose a *time-wise unfolding* (as opposed to the conventional *batch-wise unfolding*) rearrangement of the data collected from multiple runs of a batch process. We utilize the resulting "flattened" batch data to construct reference batch trajectories and the corresponding time-varying confidence intervals for process monitoring and fault detection.

The paper is organized as follows: the next section provides a brief overview of the geometric framework used for data visualization. The unfolding method used in preprocessing the data is introduced, followed by the method for obtaining the batch trajectory that captures the "normal" operation of the batch process. A discussion of the batch alignment methods available in the literature and the method of choice for this paper follows. Case study results are then presented and discussed. Finally a conclusion and potential directions for future work are provided.

2. Preliminaries

2.1. Time-resolved radial plots for representing multivariate time series data

In our previous work [22], we introduced a framework for representing multi-dimensional time series data in time-resolved radial coordinates, which we termed time-explicit Kiviat diagrams. In this framework, each (appropriately normalized and scaled) data sample is represented using a radial (Kiviat) plot. The Kiviat plots for successive samples are stacked equidistantly along a vertical time axis, thereby creating a spatial representation of a multivariate time-dependent data set.

We briefly illustrate these concepts using a five-dimensional data set with one-minute sampling time for illustration purposes. The first sample is represented in radial coordinates and a time axis normal to the plane of the plot is added (Fig. 1a). Subsequent samples are added as radial plots aligned along the time axis (b and c). The plot can be updated by adding such "data slices" (and removing older ones) in a first-in, first-out fashion.

This framework affords opportunities for performing fault detection from both a univariate and a multivariate perspective. From a univariate standpoint, upper and lower confidence limits can be defined for each variable, and these bounds can be used to construct inner and outer convex hulls for the graph. Data samples whose Kiviat plots fall outside the space defined by the two hulls are then labeled as corresponding to a faulty operating state. This is illustrated in Fig. 1d: assuming that samples for $t \leq 3$ min represent normal operation, the *normal* operating region is defined as the cylindrical shell (red) between the inner and outer convex

Please cite this article in press as: R. Wang, et al., A geometric method for batch data visualization, process monitoring and fault detection, J. Process Control (2017), http://dx.doi.org/10.1016/j.jprocont.2017.05.011