Chemical Engineering Science 166 (2017) 130–143

Contents lists available at ScienceDirect

Chemical Engineering Science

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

Subspace decomposition and critical phase selection based cumulative quality analysis for multiphase batch processes

Yan Qin [a], Chunhui Zhao [a,*], Xuezhong Wang [b], Furong Gao [c]

[a] State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China
[b] Institute of Particle Science and Engineering, School of Process, Environmental and Materials Engineering, The University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK
[c] Department of Chemical and Biomolecular Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong Special Administrative Region

HIGHLIGHTS

- Sequential and definite quality-relevant phases are obtained.
- The type of quality index is prejudged to determine whether it is of cumulative type.
- Subspace of variation is decomposed to explain cumulative quality effect.
- Critical phases are identified for inter-phase cumulative quality analysis.

ARTICLE INFO

Article history:
Received 17 October 2016
Received in revised form 7 January 2017
Accepted 13 March 2017
Available online 15 March 2017

Keywords:
Cumulative quality analysis
Subspace decomposition
Critical phase selection
Batch processes

ABSTRACT

Quality analysis and prediction have been of great significance to ensure consistent and high product quality for chemical engineering processes. However, previous methods have rarely analyzed the cumulative quality effect which is of typical nature for batch processes. That is, with time development, the process variation will determine the final product quality in a cumulative manner. Besides, they cannot get an early sense of the quality nature. In this paper, a quantitative index is defined which can check ahead of time whether the product quality result from accumulation or the addition of successive process variations and cumulative quality effect will be addressed for quality analysis and prediction of batch processes. Several crucial issues will be solved to explore the cumulative quality effect. First, a quality-relevant sequential phase partition method is proposed to separate multiple phases from batch processes by using fast search and find of density peaks clustering (FSFDP) algorithm. Second, after phase partition, a phase-wise cumulative quality analysis method is proposed based on subspace decomposition which can explore the non-repetitive quality-relevant information (NRQRI) from the process variation at each time within each phase. NRQRI refers to the quality-relevant process variations at each time that are orthogonal to those of previous time and thus represents complementary quality information which is the key index to cumulatively explain quality variations time-wise. Third, process-wise cumulative quality analysis is conducted where a critical phase selection strategy is developed to identify critical-to-cumulative-quality phases and quality predictions from critical phases are integrated to exclude influences of uncritical phases. By the two-level cumulative quality analysis (i.e., phase-wise and process-wise), it is feasible to judge whether the quality has the cumulative effect in advance and thus proper quality prediction model can be developed by identifying critical-to-cumulative-quality phases. The feasibility and performance of the proposed algorithm are illustrated by a typical chemical engineering process, injection molding.

1. Introduction

Batch processes have experienced rapid development and become an important manufacturing mode of producing high-value-added products through process repetition in chemical engineering industry (Chen and Liu, 2002). With the nature of quickly responding to changing market demand and customer requirement, batch processes have been widely applied in specialty chemical, biomedical, semiconductor, etc (Choi et al., 2008). Caused by process disturbances and batch to batch variations, low reproducibility brings a great challenge to ensure the consistency of

* Corresponding author.
E-mail address: chhzhao@zju.edu.cn (C. Zhao).

http://dx.doi.org/10.1016/j.ces.2017.03.033

product quality. Besides, for many batch processes, measurements of product quality are in general not available until the end of a batch. Thus, online quality prediction technology is indispensable to guarantee high product quality and improve process efficiency for batch processes.

Benefiting from the development of data acquisition and storage technologies, data-driven multivariate statistical methods (Wold et al., 1987; Zhang et al., 2016, 2013; Yao, 2008; Dayal and Macgregor, 1997; Yu and Macgregor, 2004; Nomikos and MacGregor, 1995; Zhao and Gao, 2017) have attracted increasing attention. Only using process data, multivariate statistical methods have been widely used to develop prediction models, including partial least squares (PLS) (Dayal and Macgregor, 1997), canonical correlation analysis (CCA) (Yu and Macgregor, 2004), etc. These methods overcome the disadvantages of the first principle models, which are time-consuming and require in-depth process knowledge. As an extension of PLS algorithm, multi-way partial least squares (MPLS) (Nomikos and MacGregor, 1995) was proposed for batch processes and it was commonly regarded as a milestone in quality prediction. In MPLS, three-dimensional process data matrix is unfolded batch-wise, which keeps the dimension of batch direction unchanged, and then a prediction model is developed by performing PLS between the batch-wise unfolded process data and corresponding quality data. However, there are two obvious drawbacks in MPLS as pointed out in previous work (Xiao and Wang, 2016; Zhao et al., 2008; Undey and Cinar, 2002). On the one hand, the entire batch-wise unfolded process data are employed for process modeling so that the prediction accuracy heavily relies on the estimation of unavailable future data for the online purpose. On the other hand, it neglects the multiphase characteristics of batch processes, which brings the difficulty for process understanding and may lead to inaccurate prediction. Sequentially, a series of improvements (Randolf et al., 2010; Chiu and Yao, 2013; Yao and Gao, 2008) have been developed. In order to avoid estimating future data, Randolf et al. (2010) suggested the separation of time-slice loading and weight coefficients from batch-wise unfolding based model for online prediction. Aiming at the same problem, Chiu and Yao (2013) proposed a method by performing elastic net on the batch-wise unfolded process data and quality data. The values of calculated coefficients are used to evaluate the importance of sampling times and process variables. However, batch processes, in general, operate in a sequence of physical phases and each phase may have its specific characteristic (Yao and Gao, 2008). It is noticed that process variable correlations, as well as the influences on quality, keep similar within a phase while may have a significant difference between different phases. However, the above mentioned methods (Randolf et al., 2010; Chiu and Yao, 2013) treat entire batch as a single subject without exploring the changes of process characteristics over different phases.

As an important feature of batch processes, the multiphase characteristic has attracted increasing attention (Undey and Cinar, 2002; Smilde et al., 2003; Reinikainen and Hoskuldsson, 2007; Liu and Wong, 2008; Facco et al., 2007; Doan and Srinivasan, 2008; Lu et al., 2004; Lu and Gao, 2005; Zhao et al., 2007; Yao and Gao, 2009; Camacho and Pico, 2006). A class of phase partition methods (Undey and Cinar, 2002; Smilde et al., 2003; Reinikainen and Hoskuldsson, 2007; Liu and Wong, 2008; Facco et al., 2007; Doan and Srinivasan, 2008) based on expert knowledge or process analysis were proposed, such as indicator variable method (Undey and Cinar, 2002), multiblock modeling technique (Smilde et al., 2003) etc. In order to overcome the dependence on process knowledge and partition the phases automatically, data-driven phase partition algorithms have been paid special attention (Lu et al., 2004; Lu and Gao, 2005; Zhao et al., 2007, 2009; Yao and Gao, 2009; Camacho and Pico, 2006; Yu and Qin, 2009; Zhao, 2013; Zhao and Gao, 2011). Lu et al. (2004) put forward the sub-PCA algorithm by clustering those sampling points that have similar variable correlations into one phase. Sequentially, sub-PLS (Lu and Gao, 2005) was proposed for online quality prediction. Considering the between-phase transition patterns, the soft-transition multiple PCA method (Zhao et al., 2007) and an angle based phase partition method (Yao and Gao, 2009) were proposed. For these methods, variable correlations or process correlations are clustered by k-means algorithm, which does not take the time sequential property into consideration. Therefore, initial phase partition results may be discontinuous, which bring heavy burdens of post-processing. Camacho and Pico (2006) recursively divided the batch cycle into phases at the points where the prediction error achieves minimum. Yu and Qin (2009) employed Gaussian mixture model to cluster the sampling times into different classes. As pointed by Zhao et al. (2007), the influences on monitoring performance should also be considered during phase partition as well as changes of variable correlations. Thus, a stepwise sequential phase partition algorithm was proposed for fault detection by checking the changes of monitoring statistics. A quality-relevant sequential phase partition method (QSSPP) was also developed for quality prediction (Zhao, 2013). However, the results of phase partition are greatly influenced by a tunable parameter, i.e. relaxing factor. Besides, the phase-based prediction model isolates the influences of each time on product quality without considering the cumulative quality effect. Influences of process variations on quality are in general increased cumulatively with time evolution which is termed as cumulative quality effect. Zhao et al. (2009) used each phase as the basic analysis object from which the local contributions of different phases are stacked to explore the cumulative quality effect. However, they did not evaluate the significance of different phases. In fact, the significance of different phases is not necessarily in accord with their operation sequence. Uncritical phases should be removed from model development to avoid introducing undesirable disturbances into the model if they have no significant influences on product quality. Zhao and Gao (2011) treated each phase as a single block and analyzed their priority for quality interpretation by checking their different contributions to qualities. However, both work (Zhao et al., 2009; Zhao and Gao, 2011) did not explore how the cumulative effects change time-wise within the same phase. Besides, they did not consider the problem of online application by using the measurement of the entire phase.

From the above analysis, the cumulative quality effect has not been well analyzed time-wise within each phase and explored for online quality prediction. Besides, several problems are noticed. First, phase partition results of QSSPP (Zhao, 2013) are subject to a tunable parameter (relaxing factor), which directly leads to uncertainty of phase partition. Second, the process variations at each time present both similar and dissimilar influences on quality in comparison with previous variations. They explore different quality information and should be well decomposed and separated from each other for analysis. Third, different phases may contribute differently to cumulative quality effects in which only the critical-to-cumulative-quality phases can reliably provide cumulative quality information and thus should be separated from those uncritical ones. In order to solve the above mentioned problems, a subspace decomposition and critical phase selection method is proposed for cumulative quality analysis for multiphase batch processes. First, multiple phases are separated from the perspective of quality analysis in which the tunable parameter is determined by using the fast search and find of density peaks clustering (FSFDP) algorithm (Rodriguez and Laio, 2014). Second, after phase partition, the measurement space at each sampling time is decomposed into several parts to explore their different influences on quality. Here, a quantitative index, termed non-repetitive