Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 48-21 (2015) 595–600

# Complex batch processes quality prediction using non-Gaussian dissimilarity measure based just-in-time learning model ★

**Xinmin Zhang*** **Yuan Li*** **Feng Lv****

* *Department of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, P. R. China, (e-mail: zh_xin_min@163.com, li-yuan@mail.tsinghua.edu.cn).*
** *Department of Electronic, Hebei Normal University, Shijiazhuang 050024, P. R. China,(e-mail:lvfeng@mail.tsinghua.edu.cn).*

**Abstract:** In modern batch processes, soft sensors have been widely used for estimating quality variables. However, they do not show superior prediction performance due to the self-limitations of these methods and the unique characteristics of batch processes such as time-varying, nonlinearity, non-Gaussianity, multi-phases and batch-to-batch variations. To cope with these issues, a novel non-Gaussian dissimilarity measure based just-in-time learning (JITL) soft sensor is developed in this paper. Unlike the traditional JITL model which uses the distance-based dissimilarity measure for local modeling, the proposed method uses non-Gaussian dissimilarity measure to evaluate the statistical dependency of the extracted independent components to construct the local model, which can well capture the non-Gaussian features in the process data. Furthermore, a novel relevant samples search strategy is introduced into the JITL framework for local modeling, which not only searches the relevant samples along the direction of time axis but also along the direction of batch-to-batch. The proposed search strategy can guarantee the current query sample and the local modeling data belong to the same phase duration and have the smallest process trajectory variations. Hence, the proposed soft sensor is suitable for uneven-phase and batch-to-batch variations batch processes. Meanwhile, the proposed method can well cope with the changes in process characteristics as well as nonlinearity. The effectiveness of the proposed method is verified on the fed-batch Penicillin Fermentation process.



*Keywords:* Quality prediction, batch processes, independent component analysis, just-in-time learning, non-Gaussian dissimilarity measure.

## 1. INTRODUCTION

In modern chemical processes, batch processes play a critical role in producing low volume and high value-added products due to its high operating flexibility and low capital investment, which have been widely used in the fine chemistry, biochemical and polymer industries. Process monitoring and quality control have become the crucial tasks for batch processes to improve product quality and ensure the process safety(Ge et al. (2013)). Data-driven multivariate statistical process control (MSPC) methods such as Multiway partial least-squares (MPLS) have been widely used for process monitoring and quality prediction in batch processes(Yin et al. (2013)). However, some self-limitations of the conventional MPLS make it do not function well. On the one hand, MPLS is a linear method, which cannot handle the nonlinear characteristic of batch processes. On the other hand, MPLS is a second-order method, which means it takes into account only mean and variance or covariance of the data set, but may not be able to efficient extract higher-order statistical information from the data with non-Gaussian distributions, which is common for the actual industrial processes. Moreover, since the traditional MPLS usually takes the entire batch as an object to build a global model, it also cannot efficiently capture the multi-phase feature of most batch processes.

In order to obtain better prediction performance, several improved approaches have been proposed. For example, to handle the nonlinear problem, a series of nonlinear regression methods such as nonlinear PLS (NLPLSs), artificial neural networks, kernel PLS (KPLS) have been developed. Compared to NLPLS and neural networks, KPLS avoids nonlinear optimization through introducing kernel function, which has recently attracted increasing consideration in many chemical industries. To handle the multi-phase problem, some phase-based PLS modeling methods have been developed to improve the quality prediction ability(Zhao et al. (2012),Ge et al. (2014)), based on the basic idea that batch process operation can be divided into several separate phases with different phase characteristics. Furthermore, they also point out that the transition information between different phases has certain impacts on the final product quality prediction. More recently, independent component analysis (ICA) has been developed for non-Gaussian process control. Based on ICA, independent component regression (ICR), ICA-PLS or ICA-MLR were developed to construct the regression model between the ICs and the quality variables(Kaneko et al. (2009)). It was found that the ICA-based

★ The authors would like to acknowledge the National Natural Science Foundation of China under Grant 61174119, 61034006, 60774070, Liaoning Province Foundation under Grant 2009R47,and Hebei Province Science Foundation under Grant F2014205115.

2405-8963 © 2015, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2015.09.591

regression models give high prediction power and are easy to interpret for non-Gaussian industrial processes.

Generally, batch processes are time-varying due to changes of process characteristics. To cope with this issue, recursive methods, such as recursive PLS, have been proposed to adapt the soft sensor model to a new operation condition recursively(Helland et al. (1992)). However, these methods are easy to adapt the soft sensor excessively and cannot deal with abrupt changes of the process. Alternatively, JITL method was proposed to handle such type of situations. Different from the traditional offline and global modeling methods, the JITL method constructs an online local model. It can cope with the changes of the process as well as nonlinearity. However, its prediction performance is mainly depended upon the samples that are selected for local modeling. Currently, most JITL models select the local modeling samples on the basis of the distance-based similarity measure. However, it does not consider the correlation among process variables. To tackle this issue, Fujiwara et al. (2009) defined a correlation-based similarity index by integrating with statistics of PCA to develop a JITL soft sensor model (Co-JITL). Nevertheless, the correlation-based similarity index is designed based on the assumption that the process data is Gaussian distributed. Hence, the Co-JITL method may have a poor performance for non-Gaussian industrial processes application.

In this study, to explicitly account for the inherent characteristics of batch processes such as non-Gaussianity, time-varying dynamics, multi-phase and take into account the batch-to-batch variation of process trajectories, a novel JITL soft sensor model based on non-Gaussian dissimilarity measure is proposed to enhance the quality prediction performance. The non-Gaussian dissimilarity index is defined by integrating ICA with multidimensional mutual information to measure the statistical independency between two IC subspaces. Compared to the correlation based dissimilarity index, the non-Gaussian dissimilarity index can well capture the non-Gaussian process features. Furthermore, a novel relevant samples search strategy is introduced into the JITL framework for local modeling by using the moving time window. The new strategy can search the relevant samples not only along the direction of time axis but also along the direction of batch-to-batch. It can guarantee the current query sample and the local modeling data belong to the same phase duration and have the smallest process trajectory variations. First, ICA models are built on time-series data subsets of each training batch at each specific time interval to extract the IC subspaces. Second, for online prediction, the dissimilarity indexes between IC subspace of training batches and IC subspace of the current batch at the same time region are calculated at each specific sampling instant by using a moving time window. The IC subspace that minimizes the dissimilarity value is selected to construct the local soft sensor model. Then, the regression relationship between the response variable and the selected ICs can be established by using the partial least squares. From the viewpoint of feature extraction and online prediction performance, the proposed method has the following advantages: 1) The proposed method inherits the merits of JITL modeling, and it can track the changes in process characteristics regardless of abrupt noises and can cope with the process nonlinearity. 2) From the local relevant samples selection to the final regression model construction, the proposed method can efficiently extract the higher-order statistical information, thus it is particularly suitable for quality prediction of non-Gaussian process data. 3) The proposed method takes into account the multi-phase and batch-to-batch variation characteristics, thus it can provide a better prediction performance and give a reasonable interpretation especially for the uneven-phase problem of batch processes. 4) Compared to the conventional MPLS, the proposed method does not need to estimate the future value when predicting an ongoing batch.

The rest of this paper is organized as follows. Section 2 gives a brief overview of the correlation-based dissimilarity measure, ICA. Then the novel non-Gaussian dissimilarity measure based JITL soft sensor model is presented in section 3. In section 4, the proposed method is applied to the Penicillin Fermentation process and its prediction results are compared with MPLS and Co-JITL. The conclusions are given in Section 5.

## 2. PRELIMINARIES

### *2.1 Correlation Based Dissimilarity Measure*

Recently, the correlation-based dissimilarity measure derived from the PCA has been proposed. Consider a dataset $X$, it is decomposed into a score matrix $T$ and a loading matrix $P$ by singular value decomposition.

$$X = \widetilde{X} + E = TP^T + E \tag{1}$$

Based on this, the $Q$ and Hotelling's $T^2$ statistics can be obtained easily.Then a correlation-based dissimilarity measure is defined for data selection by Fujiwara et al. (2009),

$$J = \lambda T^2 + (1-\lambda)Q \tag{2}$$

where $\lambda$ is a weight factor, with $0 \leq \lambda \leq 1$.

### *2.2 ICA*

ICA is a statistical technique for decomposing the observed dataset into linear combinations of statistical independent components. Suppose the observed column vectors $X = [x_1, x_2 \cdots, x_d]^T$ can be expressed as a linear combination of $m (m \leq d)$ unknown ICs $S = [s_1, s_2 \cdots, s_m]^T$, the basic model of ICA can be written as

$$X = AS + E \tag{3}$$

Where $A$ and $E$ are the mixing and residual matrices, respectively. Data whitening is first conducted by PCA. The whitening transformation matrix can be expressed as

$$Z = QX = BS \tag{4}$$

where $Q$ is the whitening matrix and $B$ is the orthogonal matrix. Next, we can calculate the ICs as follows:

$$\widehat{S} = B^T Z = B^T QX \tag{5}$$

$$W = B^T Q \tag{6}$$

The FastICA algorithm is usually used to calculate $B$, where each column vector of $B$ is randomly initialized and updated iteratively such that the $i$-th IC has the maximized non-Gaussianity.

## 3. COMPLEX BATCH PROCESSES QUALITY PREDICTION USING NON-GAUSSIAN DISSIMILARITY MEASURE BASED JUST-IN-TIME LEARNING MODEL

### *3.1 JITL model*

The basic modeling principle of JITL can be summed up in three steps: (1) for a new query sample, relevant samples that