Contents lists available at ScienceDirect

Chemical Engineering Research and Design

IChemE

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

# Probabilistic combination of local independent component regression model for multimode quality prediction in chemical processes

*Zhiqiang Ge*[a,*], *Zhihuan Song*[a,*], *Peiliang Wang*[b]

[a] *State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Department of Control Science and Engineering, Zhejiang University, Hangzhou 310027, Zhejiang, PR China*
[b] *School of Information and Engineering, Huzhou Teachers College, Huzhou, Zhejiang, PR China*

ABSTRACT

In this paper, a probabilistic combination form of the local independent component regression (ICR) model is proposed for quality prediction of chemical processes with multiple operation modes. Through the introduction of the Bayesian inference strategy, the posterior probabilities of the data sample in different operation modes are calculated upon two monitoring statistics of the independent component analysis (ICA) model. Then, based on the combination of local ICR models in different operation modes, a probabilistic multiple ICR (MICR) model is developed. Meanwhile, the operation mode information of the data sample is located through posterior analysis of the new model. To evaluate the multimode quality prediction performance of the proposed method, two case studies are provided.


*Keywords:* Independent component regression; Monitoring statistic; Bayesian inference; Multimode quality prediction

## 1. Introduction

In the chemical process industry, it is well-known that some important quality variables or key indices, such as product composition in distillation columns and concentration of reaction mass in chemical reactors, are difficult to measure online. This is due to the limitation of the process itself or measurement techniques. Therefore, these key variables of the process are often determined by offline analyses in the laboratory or by some online analyzers. However, both offline analyses and online analyzers are expensive or time-consuming which may introduce a large delay to the control system. To this end, Joseph and Brosilow first introduced inference control method to estimate the key variables by using some easy-measured secondary process variables (Joseph and Joseph, 1978). Since then, online quality estimation and prediction methods have been widely studied both in academy and industrial areas (Tham et al., 1991; Kresta et al., 1994; Dochain, 2003). Particularly, with the wide utilization of the distributed control system (DCS) in modern industrial processes, a huge number of process data have been collected, upon which the data-based modeling methods have gained much attentions in recent years (Fortuna et al., 2007; Kano and Nakagawa, 2008; Kadlec et al., 2009; Zhang and Ma, 2012; Zhang et al., 2012; Ge et al., 2013; Liu et al., 2013; Lee et al., 2013). Different from the traditional model-based method which is greatly relied on the first-principle model of the process, the data-based method rarely need any process knowledge or experiences. Instead, it tries to extract useful information from the process data, which can characterize the process condition, operation behavior, and variable correlations.

To date, various data-based methods have been developed for quality prediction purpose, such as principal component regression (PCR) (Xie and Kalivas, 1997; Hartnett et al., 1998; Keithley et al., 2009; Ge et al., 2011), partial least squares (PLS) (Kruger et al., 2001; Zhang and Zhang, 2009), artificial neural network (ANN) (Mandic and Chambers, 2001; Lee et al., 2005; Gonzaga et al., 2009), and support vector machine (SVM) (Vapnik, 1995; Suykens et al., 2002; Bylesjo et al., 2008; Liu et al., 2009). According the recent review paper on the

* *Corresponding author*. Tel.: +86 571 87951442.
E-mail addresses: gezhiqiang@zju.edu.cn, gezhiqiang@gmail.com (Z. Ge), zhsong@iipc.zju.edu.cn (Z. Song).
Received 13 December 2012; Received in revised form 21 August 2013; Accepted 21 September 2013
0263-8762/$ – see front matter 
http://dx.doi.org/10.1016/j.cherd.2013.09.010

data-based soft sensor topic by Kadlec et al. (2009), PCR/PLS are two of the most widely used methods for quality prediction in industrial process. Even though successful studies have demonstrated the efficiency of PCR and PLS based quality prediction methods, as Gustafsson pointed out, neither of these two methods can generally recover a true underlying linear latent model for the process data (Gustafsson, 2005). Here, the linear latent model is referred to those data-based models which are driven by latent components, such as PCR and PLS. In addition, PCR/PLS can only model the first and the second order statistics for the process data, higher order statistical information has been ignored. As we know, the first and second order statistics can only describe the process data which is Gaussian distribution. For non-Gaussian process data, high order statistic is necessary for information extraction and interpretation. Fortunately, a further development of the traditional PCA method which is called independent component analysis has been proposed (Hyvarinen and Oja, 2000). The main idea of ICA is to recover statistically independent source signals from the process data. In the ICA method, the extracted component is assumed to be mutually independent instead of merely uncorrelated. Through probability interpretation, independence is a much stronger condition than uncorrelatedness, which can make use of higher order statistical signals. Motivated by the advantage of ICA, it has been successfully used for process monitoring in last several years (Li and Wang, 2002; Kano et al., 2003; Lee et al., 2004; Ge and Song, 2007, 2008; Liu et al., 2008). More recently, ICA has also been used for regression purpose, which is termed as independent component regression (ICR) (Chen and Wang, 2001; Westad, 2005; Ahn et al., 2007; Kaneko et al., 2008; Zhao et al., 2010). Compared to PCR, several research studies have demonstrated that ICR can recover the true underlying sources much better, depending on which an improved statistical interpretation of the process data can be obtained.

However, the traditionally developed ICR method can only perform efficiently under a single stationary operation region. When the process condition changes or multiple products are intended to be produced, a single ICR model may not function very well. For quality prediction of those processes, a multiple ICR model is proposed in this paper. In this method, an important issue is how to select the corresponding local ICR model for the present data sample. Without any process knowledge, it is difficult to select the appropriate local ICR model. Nevertheless, the mode information of the data sample is important for quality prediction, as well as process control and optimization.

To address the mode localization problem, a monitoring statistic based Bayesian inference approach is proposed in this paper. After the local ICR models have been built in different operation modes, two monitoring statistics $I^2$ and SPE can be constructed. Depending on these two monitoring statistics, the posterior probability of the present data sample in different operation modes can be calculated through the Bayesian inference approach. Therefore, the final quality prediction result can be obtained by combing local prediction results in different operation modes through their corresponding posterior probabilities. Meanwhile, based on the calculated posterior probabilities, the mode information of the present data sample can be easily located, detailed information of mode localization is provided in Section 3.

The rest of this paper is organized as follows. In Section 2, the principle of the traditional ICR method is introduced, which is followed by the detailed demonstration of the proposed local ICR based quality prediction method in the next section. In Section 4, two case studies are provided to evaluate the performance of the proposed method. Finally, conclusions are made.

## 2. Independent component regression (ICR)

Generally, the ICR model can be built through two steps: independent component extraction and linear regression. In the ICA algorithm, it is assumed that the measured process variables $\mathbf{x} \in R^{m\times 1}$ can be expressed as linear combinations of $r(\leq m)$ unknown independent components $\mathbf{s} \in R^{r\times 1}$, the relationship between them is given by (Hyvarinen and Oja, 2000)

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{e} \tag{1}$$

where $\mathbf{A} \in R^{m\times r}$ is the mixing matrix, $\mathbf{e} \in R^{m\times 1}$ is the residual vector. The basic problem of ICA is to estimate the original component $\mathbf{s}$ and the mixing matrix $\mathbf{A}$ from $\mathbf{x}$. Due to this end, the objective of ICA is to calculate a separating matrix $\mathbf{W}$ so that the components of the reconstructed data matrix $\hat{\mathbf{s}}$ become as independent of each other as possible, given as(2)$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$

After the independent components have been estimated from the process data, the linear regression can be carried out between two datasets: the independent component dataset $\hat{\mathbf{S}} = [\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2, \ldots, \hat{\mathbf{s}}_n]^T \in R^{n\times r}$ and the quality variable dataset $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]^T \in R^{n\times p}$. Therefore, the linear regression matrix can be calculated as

$$\mathbf{Q} = (\hat{\mathbf{S}}^T\hat{\mathbf{S}})^{-1}\hat{\mathbf{S}}^T\mathbf{Y} \tag{3}$$

If we denote the dataset of process variables as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T \in R^{n\times m}$, and combine the two steps of ICR modeling procedures, the ICR regression matrix can be determined as

$$\mathbf{R}_{ICR} = \mathbf{Q}^T\mathbf{W} \tag{4}$$

## 3. Method development

In this section, detailed description of the proposed method is provided. First, local ICR models are developed in different operation modes. Second, two monitoring statistics are constructed for the mode localization purpose. Finally, the probabilistic prediction model is developed by combining local prediction results in different operation modes.

### 3.1. *Local ICR model derivation*

Denote the process datasets in different operation modes as $\{\mathbf{X}_i, \mathbf{Y}_i\}_{i=1,2,L,T}$, where $T$ is the number of operation modes. Through the ICA modeling step, different local ICA models can be formulated as follows

$$\mathbf{X}_i = \mathbf{A}_i\mathbf{S}_i + \mathbf{E}_i \tag{5}$$

where the independent component dataset can be estimated by the separating matrix $\mathbf{W}_i$ of the corresponding ICA model, thus(6)$\hat{\mathbf{S}}_i = \mathbf{W}_i\mathbf{X}_i$