Chemometrics and Intelligent Laboratory Systems 137 (2014) 110–119

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Methodology of data reconciliation and parameter estimation for process systems with multi-operating conditions

Zhengjiang Zhang [a,b], Ying-Yu Chuang [b], Junghui Chen [b,*]

[a] College of Physics and Electronic Information Engineering, Wenzhou University, Wenzhou 325035, Zhejiang, PRC
[b] Department of Chemical Engineering, Chung Yuan Christian University, Chungli 320, Taiwan, ROC

ARTICLE INFO

Article history:
Received 4 April 2014
Received in revised form 2 June 2014
Accepted 12 June 2014
Available online 21 June 2014

Keywords:
Clustering
Data reconciliation
Maximum-correntropy estimate
Multi-operating conditions
Parameter estimation
Principal component analysis

ABSTRACT

The economic performance of the real time optimization and process control is influenced by the accuracy of the process model. Data reconciliation and parameter estimation (DRPE) is a crucial technique to obtain reliable process models. In real industrial processes with multi-operating conditions, the effects of contaminated measured data, nonlinear characteristics of model parameters with operating conditions and different types of gross errors increase the difficulty to tune the process models. This paper focuses on the influence of those factors on DRPE problems. A practical DRPE methodology is proposed for the process system with multi-operating conditions to decrease the impact of those factors. The methodology contains the principal component analysis (PCA) based steady state detection, the clustering of multi-operating conditions and the maximum-correntropy estimate based DRPE with multiple data sets. PCA based steady state detection, which is a novel method for steady state detection, is used to choose useful and reliable measured data for DRPE. Clustering partitions the steady state data sets into multi-operating conditions. Maximum-correntropy estimate based DRPE for the data of each operating condition is used to reconcile the measured process data. The proposed methodology is finally applied to a typical real industrial process with multi-operating conditions: the air separation process. The effectiveness of the proposed methodology can be demonstrated by the results of DRPE.

1 . Introduction

Model-based real time optimization and advanced process control techniques have been widely used to track the economic optimum of the industrial processes. Reliable process models are fundamental and crucial for the implementation of those techniques. Since most rigorous process models often include undetermined parameters, which have to be estimated based on the measured data, data reconciliation and parameter estimation (DRPE) is considered a crucial technique to obtain reliable parameters for model fitting, validation, real time optimization and process control in chemical industries.

In most general form, data reconciliation is a minimization of measurement errors to satisfy the constraints of the process models. Parameter estimation is the step after data reconciliation. In this step, the reconciled values of the measured variables are used to estimate the values of the model parameters [1,2]. This two-step approach is inefficient and it has led to the development of the simultaneous strategies for DRPE [3–7]. The most commonly used formulation of the DRPE problem is to minimize least squares errors in measurements and to reconcile values of the measured variables at the same time subject to model constraints and bounds. Based on the rigorous process models, special attention was paid to the influence of measurement errors and computation strategies of DRPE. The objective function is based on the assumption that measurements have random errors with normal distribution. However, measurement errors generally contain both random and gross errors. When gross errors are present in the measured data, they can lead to incorrect estimations and severe bias reconciliations of the other measurements. Many robust estimators are proposed to reduce the effect of gross errors and yield less biased estimates, such as the fair function, the contaminated normal objective function, the redescending estimator, the quasi-weighted least squares estimator, and other robust estimators [3,4,8–14]. Another important issue is the convergence of the DRPE problem. Many computation strategies are proposed to improve the convergence or to speed up computation, such as the reduced successive quadratic programming strategy, the nested three-stage computation framework, the particle swarm optimization, and the parallel calculation method [5,6,15,16]. However, most of those papers assumed that process model parameters are constants in spite of the change of operating conditions. In fact, in most industrial processes, product demand is not fixed but periodical, stepwise and intermittent, leading to the product or load demand fluctuation in response to the changing product demand. The product or load demand fluctuation impacts the operating condition. Thus, any model will have a limited range. It may be restricted by the modeling assumptions for

* Corresponding author. Tel.: +886 3 2654107; fax: +886 3 2654199.
E-mail address: jason@wavenet.cycu.edu.tw (J. Chen).

http://dx.doi.org/10.1016/j.chemolab.2014.06.007

physical models. This means that the local model with constant parameters is only fit for a limited range of validity.

Many methods estimate model parameters based on the real-time measured data [10,17]. The measured data are acquired by the sensors installed on process devices, but the measured data are contaminated, either by white noise from the measurement equipment or the effects of sensor failure, process leaking etc. As hundreds or thousands of process measurements are simultaneously measured and stored in the massive storage media in industrial processes, the multiple data sets often contain multi-operating conditions in the operation processes. Besides, the measured data acquired from the plant are at steady states and dynamic states. To properly select the useful and reliable historical measured data and present an operation condition are important for solving the DRPE problem. Moreover, many different steady states in the process system may contain multi-operating conditions. With the review of DRPE methods in literature, most cases based on the simulated data are used for DRPE [4,9,13]. The data pre-processing combined with the DRPE problem is lacking of consideration before. As operating conditions are changed frequently in a wide range, the model parameters may change nonlinearly. The simultaneous DRPE problem based on the individual operation condition is solved. Thus, the proper clustering data should be considered.

The presence of gross errors in the measured data affects the results of conventional data reconciliation since the large errors are not sufficiently corrected. As a result of smearing, the estimates may become less accurate than the collected measurements. It is difficult but necessary to eliminate or decrease the effect of different types of gross errors on the results of the DRPE problem. Those problems exist in the real industrial process system. The use of M estimators has been proposed in the DRPE problem with the gross error detection [4,9,18,19]. M estimators are robust and they yield unbiased results insensitive to departures from ideal statistical distributions. Among different M estimators, correntropy can measure both the uncertainty and dispersion and it is a good optimality criterion in the estimation problems. By appropriate use of the kernel width, it can estimate the unknown function from data contaminated with outliers [19,20]. Because of the outstanding characteristics, correntropy is used as the objective function of the simultaneous DRPE problem in this paper. Therefore, a whole design methodology of the DRPE problem for the process system with multi-operating conditions is proposed to decrease the impact of the factors, including contaminated measured data, nonlinear characteristics of model parameters with multi-operating conditions and different types of gross errors.

The remainder of this paper is organized as follows. The DRPE problem of the process system with the multi-operating conditions is described in the next section. In Section 3, the methodology of DRPE, which contains the PCA based steady state detection, the clustering of multi-operating conditions and the maximum-correntropy estimate based DRPE, is proposed. The PCA based steady state detection is used to choose useful and reliable measured data for DRPE. The clustering method is used to partition the steady state data sets into multi-operating conditions. The maximum-correntropy estimate based DRPE for the data of each operating condition is used to reconcile the measured process data. And then, the real industrial application in the air separation process is developed in Section 4. Finally, conclusions are drawn in Section 5.

2 . DRPE problem of process system with multi-operating conditions

In many markets, product demand is not fixed but periodical, stepwise and intermittent, leading to the product or load fluctuation. The product or load demand fluctuation impacts the entire process system as the material and energy balances readjust the flow and purities. An accurate process model is usually used to decide the feed conditions when switching from the current operating condition to another operating condition, because the model is often used for optimization and process control to track the economic optimum. In the last few years, there has been an increasing interest in applying real time optimization and advanced control based on the rigorous model. If the rigorous model is not consistent with the real process plant, there would be offset between the true plant optimum and the predicted optimum. As a reliable technique, it is important to use DRPE to make sure that the rigorous model is consistent with the real plant. A general formulation of DRPE for the process system in multi-operating conditions is shown as follows:

$$\min J(X,Y)=\sum_{j=1}^{NC}\sum_{i=1}^{N_j}\left(\mathbf{y}_j-\mathbf{x}_{j,i}\right)^T\mathbf{W}_\mathbf{y}^{-1}\left(\mathbf{y}_j-\mathbf{x}_{j,i}\right) \tag{1}$$

subject to

$$\begin{aligned}&\mathbf{g}_j\left(\mathbf{y}_j,\mathbf{u}_j,\boldsymbol{\theta}\right)=0\\&\mathbf{h}_j\left(\mathbf{y}_j,\mathbf{u}_j,\boldsymbol{\theta}\right)\geq 0\\&\mathbf{y}^L\leq\mathbf{y}_j\leq\mathbf{y}^U\qquad j=1,\ldots,NC\\&\mathbf{u}^L\leq\mathbf{u}_j\leq\mathbf{u}^U\\&\boldsymbol{\theta}^L\leq\boldsymbol{\theta}\leq\boldsymbol{\theta}^U\end{aligned} \tag{2}$$

where NC is the number of operating conditions, and N_j is the number of sampled measured data in the operation condition j. $\mathbf{x}_{j,i}\in\Re^N$ is the ith measured data of the measured variables in the jth operating condition and $\mathbf{y}_j\in\Re^N$ is the corresponding reconciled data in the jth operating condition; $\mathbf{u}_j\in\Re^{N_u}$ is the unmeasured variables in the jth operating condition; $\mathbf{W}$y are the covariance matrices of the measurement errors; $\boldsymbol{\theta}\in\Re^{N_p}$ is the set of parameters in the process model; $\mathbf{g}$ is the vector of the model equations; $\mathbf{h}$ is the vector of the inequality constraints; $(\mathbf{y}^L,\mathbf{y}^U)$, $(\mathbf{u}^L,\mathbf{u}^U)$ and $(\boldsymbol{\theta}^L,\boldsymbol{\theta}^U)$ are the lower and upper bounds of the measured variables, the unmeasured variables and the parameters, respectively.

The general formulation of the DRPE problem Eqs. (1) and (2) assumes that process model parameters are constant despite the change of the operating conditions. In the real industrial process system with multi-operating conditions, as operating conditions are switched frequently in a wide range, the model parameters change nonlinearly with the operating conditions. Thus, the models have a limited range and the assumption for Eqs. (1) and (2) may not be reasonable. Moreover, the measured process data generally have both random and gross errors. Gross errors may have different types when the operating conditions change frequently in a wide range, such as gross errors caused by bias, outliers, complete failure and precision degradation in measurements, drifts and even errors by modeling. It is necessary to eliminate or decrease the effect of the nonlinear change of model parameters and gross errors on the results of the DPRE problem. Besides, measured data acquired from the plant generally contain data information with the steady states and the dynamic states. How to choose useful and proper measured data for the DPRE problem is another important issue. Therefore, to successfully solve the DPRE problem, an efficient pre-processing method for the collected data is strongly needed.

3. Scheme of DRPE for process systems with multi-operating conditions

In this paper, a DRPE scheme for process systems with multi-operating conditions is proposed (Fig. 1). In Fig. 1, many process measurements are simultaneously measured from the operating industrial process with multi-operating conditions. The scheme contains three parts.

(1) A continuous process operated in a state of steadiness is useful, and steady state models are being used to solve DRPE problems. The term steady state condition implies that the process is operated around some stable points or within some stationary region. When some processes are steady and others are unsteady, knowing the status of the steady state in time can also help to screen