Computers and Chemical Engineering 75 (2015) 120–134

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Correntropy based data reconciliation and gross error detection and identification for nonlinear dynamic processes

Zhengjiang Zhang[a,b], Junghui Chen[b,*]

[a] College of Physics and Electronic Information Engineering, Wenzhou University, Wenzhou 325035, People's Republic of China
[b] Department of Chemical Engineering, Chung Yuan Christian University, Chung Li District, Taoyuan, Taiwan 32023, Republic of China

ARTICLE INFO

Article history:
Received 22 September 2014
Received in revised form 2 December 2014
Accepted 9 January 2015
Available online 16 January 2015

Keywords:
Chemical processes
Correntropy
Data reconciliation
Instrumentation
Optimization
Systems engineering

ABSTRACT

Measurement information in dynamic chemical processes is subject to corruption. Although nonlinear dynamic data reconciliation (NDDR) utilizes enhanced simultaneous optimization and solution techniques associated with a finite calculation horizon, it is still affected by different types of gross errors. In this paper, two algorithms of data processing, including correntropy based NDDR (CNDDR) as well as gross error detection and identification (GEDI), are developed to improve the quality of the data measurements. CNDDR's reconciliation and estimation are accurate in spite of the presence of gross errors. In addition to CNDDR, GEDI with a hypothesis testing and a distance–time step criterion identifies types of gross errors in dynamic systems. Through a case study of the free radical polymerization of styrene in a complex nonlinear dynamic chemical process, CNDDR greatly decreases the influence of the gross errors on the reconciled results and GEDI successfully classifies the types of gross errors of the measured data.

1. Introduction

Accurate process data is important for the evaluation of the process performance and to justify current process data requires large capital expenditures. Also, process control and optimization schemes rely on accurate process data monitoring for trustworthy assessments. However, process data are often inaccurate or inconsistent with the mass balances, energy balances, and their constraints of the process systems. The inaccuracy in the process data may come from the measurement information usually corrupted by random measurement errors and systematic errors. Random measurement errors can be small perturbations from the true values. However, systematic errors, which are so called gross errors, can be quite large. The primary concerns are the gross errors usually caused by malfunctioning instruments, measurement device biases or process deficiencies. The presence of random errors decreases the precision of measurement information while gross errors introduce inaccurate information. As the improvement of the raw data set would increase the process performance and maintenance efficiency, data reconciliation (DR), which could rectify the errors in the raw data, would be very important. It uses the redundancies in the measurements to improve the accuracy and precision of measurement information to reduce the influence of measurement errors.

Kuehn and Davidson (1961) were the first to address DR. They focused on the DR problem in steady-state chemical engineering processes. Their proposed method was the solution to an optimization problem. It minimized a weighted least-squares objective function of errors between the measured and the estimated values of process variables under static material and energy balance constraints. Since then, several researchers have developed many other approaches. Romagnoli and Stephanopoulos (1981) proposed a systematic strategy for the location of the source and the rectification of gross errors in a chemical process. Their strategy can efficiently reduce the size of the DR problem and conform to the general process of variable monitoring in a chemical plant. Several researchers also proposed different strategies to enhance the solution to DR problem (Serth and Heenan, 1986; Narasimhan and Mah, 1987; Tong and Crowe, 1995; Rollins et al., 1996; Arora and Biegler, 2001; Martinez Prata et al., 2010; Zhang et al., 2010; Chen et al., 2013). In the study of dynamic data reconciliation (DDR), Kalman Filter (KF) had been effectively used to smooth measurement data (Sage and Melsa, 1971). KF estimates possess the desirable statistical properties of being unbiased. KF can also obtain the minimum variance under the assumption of the Gaussian distribution. For the dynamic nonlinear system, Stanley and Mah (1977) tackled the DDR problem in a dynamic nonlinear process using extended Kalman filter (EKF) (Narasimhan and Jordache, 2000). Their research showed that the reliability of EKF-based

* 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.compchemeng.2015.01.005
0098-1354/

approaches often decreases while the nonlinear complexities and modeling uncertainties of the system increase. Large errors and divergence of the filter might occur (Romanenko and Castro, 2004; Romanenko et al., 2004). Therefore, a model should be properly selected in order to reduce complexity. Furthermore, when the state and/or measurement equations were highly nonlinear and the posterior distribution of the states was non-Gaussian, KF or EKF based DDR would yield unsatisfactory reconciled and estimated results in a number of applications (Chen et al., 2005, 2008). The particle filtering (PF) technique, which is served as a general filter in the nonlinear and non-Gaussian state-space systems, was recently applied to DDR problems (Chen et al., 2008). However, it was restricted to the use of process state-space models, and it was not able to deal with inequality constraints, such as lower and upper bounds on the states (Bai et al., 2007; Nicholson et al., 2014).

In the study of nonlinear dynamic processes, Leibman et al. (1992) and later Ramamurthi et al. (1993) formulated the nonlinear dynamic data reconciliation (NDDR) problem and proposed solution strategies by neglecting the random noise disturbances in the state transition equations. The NDDR formulation included the manipulated input variables as part of the objective function. It was more general than the model used in filtering, whose manipulated inputs are assumed to be known exactly (Narasimhan and Jordache, 2000). This formulation could deal with inequality constraint, and it was widely used by many researchers (Chen and Romagnoli, 1998; Kong et al., 2000; Martinez Prata et al., 2010). However, the NDDR problem was still formulated as a weighted least-squares objective function which is the sum of squared measurement errors in each time step. The function was minimized subject to the process dynamic model. It was very sensitive to large measurement errors, and it would lead to unsatisfactory reconciliation and estimation in the presence of the gross errors.

Gross errors are random or deterministic errors without the relation with the true values. In the original DR study, it was assumed that the noise that affected the variables was randomly distributed with zero mean. However, in practice, gross errors may occur. The presence of gross errors will affect the results of DR if the large errors are not sufficiently eliminated or corrected. As a result of smearing, both the reconciled measurements and the estimates of states may become distorted. Gross error detection and identification (GEDI) is generally considered as a crucial technique within the DR framework. In order to avoid corrupted adjustments, the GEDI problem has received considerable attention in the past few decades and a number of strategies have been developed. The classical hypothesis testing strategies are the first methods used for GEDI, including the global test (Almasy and Sztano, 1975), the nodal test (NT) (Mah et al., 1976) and the measurement test (MT) (Mah and Tamhane, 1982). Serth and Heenan (1986) proposed several tests, including the iterative measurement test (IMT) and the modified IMT. They were more efficient than MT and NT in terms of performance. Other methods, such as generalized likelihood ratio methods (Narasimhan and Mah, 1987), maximum power test methods (Crowe, 1992), principal component test methods (Tong and Crowe, 1995), etc., were also developed for GEDI. A general survey of gross error detection with data reconciliation approaches was given by Özyurt and Pike (2004). However, most of the above strategies were developed to solve the DR problems in steady-state chemical processes.

After DR, the methods that identified gross errors in dynamic systems were also developed because the process model error was considered as an important contributing factor in the estimation of the measurement bias and process state variables. McBrayer and Edgar (1995) used the NDDR formulation to derive the resulting difference between the measured and the reconciled values, and they developed a method for bias detection in nonlinear dynamic processes. Bagajewicz and Jiang (1997) proposed a new statistic method to detect bias in the linear dynamic systems. Chen and Romagnoli (1998) used the moving horizon concept and the cluster analysis techniques to successfully distinguish outliers from normal measurements in dynamic chemical processes. Bai et al. (2007) developed an algorithm to deal simultaneously with bias correction and DR in dynamic processes. Xu and Rong (2010) proposed a new framework for DR and measurement bias identification in generalized linear dynamic systems. Gonzalez et al. (2011) proposed a Bayesian approach to determine the inconsistency of sensors. They used the modified principal components for factor analysis to determine the initial value, and then estimated sensor variance and gross errors by means of the Bayesian estimation. In 2012, they developed an online algorithm to detect and estimate gross errors from measurement data under mass and energy balance constraints (Gonzalez et al., 2012). By applying filtering techniques, Singhal and Seborg (2000) proposed a probabilistic formulation that combined EKF and the expectation-maximization (EM) algorithm in the measurement reconciliation. The new EKF-EM method removed the outliers and reduced noise effects. Later, Chen et al. (2008) used the PF technique for the NDDR problem and used a mixture model comprising two Gaussian distributions to address the effect of outliers. The outlier detection was more efficient than the EKF-EM method in terms of performance. The strategies mentioned above for GEDI problems only deal with outlier or bias detection without considering different types of gross errors even if there were mixed types of gross errors.

GEDI is also considered as sensor fault detection and isolation problems in the area of fault detection and diagnosis (FDD). Many different FDD approaches were developed to detect and isolate sensor faults. Those approaches are mainly classified into two categories: the model-based approaches and the knowledge-based or the data-driven approaches. The data-driven approaches include the traditional multivariate statistical-based methods (such as the principal component analysis and partial least-squares methods) and many other improved data-based methods (such as independent component analysis, Gaussian mixture models, neural networks, support vector machines, and support vector data description) (Ge et al., 2013). Those FDD methods of sensor fault detection and isolation generally train models from data rather than relying on accurate prior models which are not often available in practice. The techniques recently developed for data based learning models contain closed loop identification technique (Wei et al., 2010), neural networks (Samy et al., 2011; Sadough Vanini et al., 2014), expert systems (Silva et al., 2012), fuzzy logic (Zhang et al., 2013), and adaptive estimation (Zhang, 2011). Those FDD methods can optimally exploit information on sensor faults whose corresponding data are stored in the historical database of the plant. In model-based approaches, the multiple-model (MM) approaches are more flexible and powerful. The term “MM” covers a wide range of approaches whose common goal is to propose an architecture (or hierarchy) for a bank of estimators or filters for isolation and identification of faults. The choice of the application domains in the MM FDD schemes in implementing the Kalman filter (Wei et al., 2010; Pourbabaee et al., 2013) for linear dynamic system, the extended Kalman filters (An and Sepehri, 2005) and particle filters (Alrowaie et al., 2012) for nonlinear dynamic system. Those filters are used as state estimators. Model-based approaches are by nature more powerful and popular if a perfect analytical model can be created and utilized. Many FDD approaches that detect and isolate sensor faults are focused on permanent sensor bias faults (Pourbabaee et al., 2013; Zhang, 2011; Samy et al., 2011) or sensor saturation (Zhang et al., 2013). However, the strategies for detecting and isolating mixed types of gross errors in sensor faults are rarely considered.

In fact, only a handful of researchers have addressed the GEDI strategies for the mixed types of gross errors in dynamic chemical processes. Abu-El-Zeet et al. (2002) proposed a novel technique