10th International Symposium on Process Systems Engineering - PSE2009
Rita Maria de Brito Alves, Claudio Augusto Oller do Nascimento and Evaristo Chalbaud Biscaia Jr. (Editors)



# Nonlinear Dynamic Data Reconciliation in Real Time in Actual Processes

Diego Martinez Prata, Enrique Luis Lima, José Carlos Pinto

*Programa de Engenharia Química/ COPPE / Universidade Federal do Rio de Janeiro – Cidade Universitária – CP 68502 – CEP: 21945-972 – Rio de Janeiro –RJ – Brazil*

## Abstract

In this work, some recent developments regarding the real time monitoring of industrial processes based on robust data reconciliation and gross error detection are reviewed. Particularly, robust data reconciliation and gross error detection are performed on line and in real time in a real polymerization process.

**Keywords**: Data Reconciliation, Robust Estimators, Gross Errors, Industrial Data.

## 1. Background

The rigorous monitoring of variables that describe the present and the long-term behavior of industrial processes is one of the first conditions required to guarantee the optimal operation of the plant. For this reason, process measurements are taken and used in process control procedures and for evaluation of process performance. With the advent of modern desktop computers, hundreds or thousands of process measurements are simultaneously measured and stored in massive storage media for the continuous monitoring of the process behavior and for performing process studies, such as the development of robust process models and the online optimization of the process operation. Therefore, modern industrial plants may provide very detailed and rich data sets for development of fundamental modeling studies. However, as information concerning such variables comes from real measurements, numerical values are normally corrupted by different kinds of errors. As a consequence, the collected data generally do not satisfy the process constraints, such as the mass and energy balances. For this reason, implementation of data rectification procedures is essential for obtainment of reliable information about the process behavior.

The data rectification procedure comprises different steps, such as variable classification, gross error detection and data reconciliation (including parameter estimation in complex models). Variable classification is required in order to determine if the available information is sufficient for solving the proposed reconciliation problem and to identify sets of observable and non-observable process variables. Different solutions have been proposed for the variable classification problem during the last four decades and detailed description of published procedures can be found in Romagnoli and Sanchez (2000).

Procedures for gross error detection are required to identify (and possibly remove from the set of measured data) observed values that do not follow the statistical distribution of the measured data, as the presence of gross errors may lead to significant biases during data reconciliation and/or parameter estimation and also to poor decisions. Gross Error Detection (GED) has been (and still is) thoroughly analyzed in the literature using different theoretical and numerical approaches, as described in detail by Narasimhan and Jordache (2000). Finally, during data reconciliation the measured data are adjusted in order to satisfy the required process constraints (or the process model) and obtain reliable estimates for the unknown desired parameters.

Mathematical models play an important role in modern process engineering because they can be used for evaluation of process states, including prediction of non-measured variables (soft sensors). Besides, reliable process models are fundamental for implementation of on-line optimization and advanced process control techniques. As industry is rapidly shifting towards real time optimization and some crucial process parameters (such as heat transfer coefficients and catalyst activities) are likely to change at plant site along the time, estimation of model parameters in real time based on available process measurements is becoming an increasingly important issue.

The scientific literature shows that joint data reconciliation and parameter estimation with simultaneous gross error detection has emerged as a fundamental tool for on-line industrial applications. Thus, the development and implementation of these powerful procedures in real time may be of paramount importance for certain process control and real-time optimization applications. However, the implementation of monitoring procedures in-line and in real time, with the simultaneous data reconciliation, estimation of process parameters, evaluation of non-measured process variables and multiple gross error detection, has seldomly been performed in real world processes due to the existence of some inherent numerical difficulties. Some of these difficulties are related to the necessary treatment of huge amounts of data in short periods of time (for real time applications), while other difficulties are related to the robustness of available numerical techniques (convergence problems are unacceptable in most real time applications), especially when multiple gross errors are present. This work presents a brief overview on data reconciliation and multiple gross error detection applied for real time industrial process applications.

## 2. Background

Data reconciliation (DR) procedures can be used in a broad range of distinct applications, such as process control, real time optimization, process monitoring, safety analysis, improvement of operational efficiency, detection of gross errors and process malfunctions, among many others. As a consequence, DR represents an important step for many engineering activities in industrial process.

The use of DR procedures for analysis of process data was originally proposed in the early sixties, encouraged by the increasing capacity of digital computers. DR may be defined as the adjustment of process measurements and parameters in order to satisfy a certain set of conservation laws and process constraints (as described by a mathematical model), while minimizing some sort of objective function that weighs the deviations between corrected and observed plant values. (In general terms, the estimation of model parameters from available plant data is certainly within this context.) For this reason, it may be said that joint data reconciliation and parameter estimation (DRPE) is useful for the simultaneous analysis of process data (allowing for identification and removal of inconsistent data sets) and process model building. Different solutions have been proposed for the DR problem during the last four decades and good reviews are provided by Crowe (1996a), Romagnoli and Sanchez (2000) and Narasimhan and Jordache (2000).

GED is of fundamental importance for adequate model building and interpretation of process data, as poor input data generally lead to very poor model responses. The commonest methods used for GED are based on statistical hypothesis testing, which requires the definition of a proper statistics for the test, assuming that the error distribution is known *a priori* (Narashimhan and Jordache, 2000). A gross error is detected if the computed statistical test exceeds a critical value, which is calculated with the assumed distribution. It is frequently assumed that measurement errors follow the well-known Normal (Gaussian) distribution, with zero mean. In this particular case, application of the maximum likelihood principle leads to the *Weighted Least Squares* (WLS) DR problem. Although the assumption of normality of the error distribution is usual, it is not necessary for DR and GED (Crowe, 1996b).

When multiple gross errors are present, a strategy (not only a method) for gross error detection may be necessary. Three strategies are normally used to deal with multiple gross error detection (MGED): Serial Elimination, Serial Compensation and Simultaneous (or Collective) Compensation (Narashimhan and Jordache, 2000). Principal component tests have also been used for MGED, but significant improvement of the MGED has not been observed (Jiang *et al.*, 1999). The main disadvantage of these MGED strategies is that they are based on residuals obtained from regression. As residuals may be heavily biased by the presence of gross errors, the implementation of iterative or combinatorial procedures may be necessary, increasing the computational effort. Despite that, these strategies have been widely used for analysis of steady-state or quasi-steady-state process conditions. As real time industrial applications require frequent MGED analysis of nonlinear dynamic models, the usage of iterative and/or combinatorial procedures may be inappropriate. (In practical implementations, it is common to use a time varying moving window in order to capture the recent behavior of the process and reduce the size of the optimization problem to manageable dimensions during the solution of dynamic DR problems.)

In order to deal with the joint nonlinear dynamic data reconciliation (NDDR) and MGED problems, alternative approaches based on cluster analysis (Chen and Romagnoli, 1998; Abu-El-Zeet *et al.*, 2002), artificial neural networks (Vachhani *et al.*, 2001) and robust estimators (Özyurt and Pike, 2004; Prata *et al.*, 2008a) have been proposed and successfully applied in the literature. Among these alternative approaches, robust data reconciliation (RDR) has been used more frequently.

There are several classes of robust estimators. The most important for DR are the M-estimators, which are generalizations of maximum-likelihood estimators. Robust estimators tend to look at the bulk of the data and ignore atypical values (gross errors), due their mathematical structure. Table 1 summarizes the works that applied robust M-estimators for industrial process monitoring, based on RDR procedures. This strategy has been recently applied for in-line and real time monitoring of real industrial nonlinear dynamic systems by Prata *et al.* (2008b,c). Comparative analysis of the performances of distinct robust M-estimators were performed by Özyurt and Pike (2004) and Prata *et al.* (2008a) for steady state and dynamic systems, respectively. Prata *et al.* (2008a) concluded that the Welsch M-estimator (which presents continuous first and second derivatives) shows better global performance, reducing the negative effects of gross errors upon the obtained estimates.

It is important to notice that the implementation of RDR approaches can be difficult sometimes, as some robust estimators are non-convex, requiring very good variable initialization in order to avoid local optima. Besides, some robust estimators present discontinuous first and/or second derivatives. As the computation of first and second derivatives may be necessary during the optimization (depending on the numerical technique implemented for minimization of the objective function), this may lead to lack of numerical stability and require the usage of smoothing functions (making numerical implementation and calculation more complex).

In order to overcome the numerical difficulties, Wongrat *et al*. (2005) proposed the use of Genetic Algorithms (GA) to perform the optimization of RDR. Other non-deterministic optimization methods, such as Tabu Search (TS), Simulated Annealing (SA) and the Particle Swarm (PS), can also be used. These methods present some additional advantages, such as the global character of the search (which avoid local minima), the unnecessary computation of derivatives and the simplicity of the implementation, although they are usually characterized by the high number of objective function evaluations, which may require more CPU time than conventional methods. However, the computed values of the objective function can be used for rigorous statistical analyses of the confidence regions of parameter estimates, which can also constitute an important benefit of these algorithms (Schwaab *et al.*, 2008).

Reimers *et al*. (2008) applied non-deterministic optimization methods (GA, TS and SA) to solve steady-state DR problems in a real industrial case. Prata *et al*. (2008b) also applied a non-deterministic optimization method (PS) to solve the robust NDDR problem in a real industrial plant, in-line and in real time. As a consequence, one may conclude that non-deterministic procedures can be used successfully for solution of DR problems in real time.

**Table 1**: Industrial applications of robust M-estimators for DR and MGED.

| Author (Year) | Industrial Scenario | M-Estimator |
|---|---|---|
| Zhang *et al*. (1995) | Sulfuric acid plant. | Contaminated Normal |
| Chen *et al*. (1998) | Sulfuric acid plant. | Fair and Lorenzian |
| Bourouis *et al*. (1998) | Multistage flash desalination plant. | Contaminated Normal |
| Özyurt and Pike (2004) | Sulfuric acid plant. | Contaminated Normal, Cauchy, Fair, Hampel, Logistic and Lorenzian. |
| Faber *et al*. (2006) | Industrial coke-oven-gas purification process | Li *et al*. (2000) |
| Faber *et al*. (2007) | Industrial coke-oven-gas purification process | Li *et al*. (2000) |
| Schladt and Hu (2007) | Industrial distillation column. | Contaminated Normal |
| Lid and Skogestad (2008) | Industrial naphtha reformer | Contaminated Normal |
| Prata *et al*. (2008b) | Industrial Polymerization Reactor. | Welsch |

It is important to emphasize that both traditional and robust formulations of DR problems depend on the relative magnitudes of the elements of the variance/covariance matrix of measured data. Robust estimators can also be used for the simultaneous characterization of the variance/covariance matrix of measurements and elimination of gross errors (Chen *et al*., 1997; Morad *et al*., 1999), resulting in improved characterization of the statistics of the process data.

## 3. Analysis of Real Industrial Data

Some of the obtained results are presented below for the bulk propylene polymerization performed in a stirred tank reactor (model and process description can be found in Prata *et al*., 2008d). Raw process data were obtained over a time interval of

7h, with a sampling time of 5 min (300s), using a time moving window of 1h (samples collected every 5 min for 22 measured variables, resulting in 286 reconciled variables in each window). The DRPE were conduced in two different computers using the *WLS* and *Welsch* estimators. The symbols (○), (●), (●) e (●) represent the plant measurements, laboratory analyses (obtained with delays of 8h, but synchronized after reconciliation), *WLS* reconciled values and reconciled values by the *Welsch* estimator, respectively. The letter (**E**) represents the gross errors detected by the *Welsch* estimator. The magnitudes of the measurement errors (associated with the instrument) and of the process noise (associated with the operation) were evaluated for all process signals based on the available data, but are not presented here for confidential reasons.

Figures 1 and 2 show measured and reconciled data for two variables: the propane concentration in the recycle stream and the reactor temperature, respectively. It can be observed that the results provided by the *Welsch* estimator are better than the results provided by *WLS* estimator, as the first one is much less sensitive to the presence of gross errors than the second one. It is important to note that deviations from the temperature setpoint (340.15 K) can not be greater than 1K (dot line); otherwise, control actions are taken in order to prevent potentially dangerous operation conditions. Figure 2 shows that some *WLS* reconciled values would violate this general rule, generating unnecessary control actions.

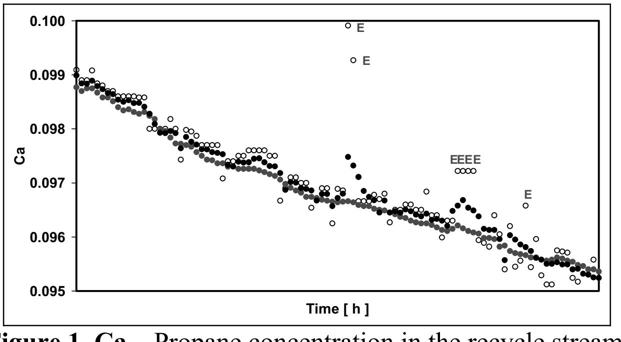


**Figure 1. Ca** – Propane concentration in the recycle stream.

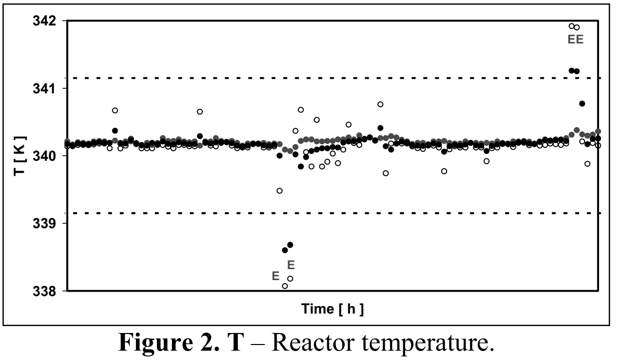


**Figure 2. T** – Reactor temperature.

Figure 3 shows measured, reconciled and laboratory data for the Melt Index of the polymer product. Again, it can be observed that the results obtained with the *Welsch* estimator are better than the results obtained with the *WLS* estimator. The in-line and real time MI inferences provided by the *Welsch* estimator are underpinned by the laboratory analyses (after synchronization), showing that in-line information provided by a process rheometer is biased. It seems clear that the proposed RDR scheme based on *Welsch* estimators can provide better results than the in-line process rheometer, which requires frequently re-calibrations during process operation. Therefore, the RDR scheme can improve the safety and performance of the process operation, aggregating value for decision making.

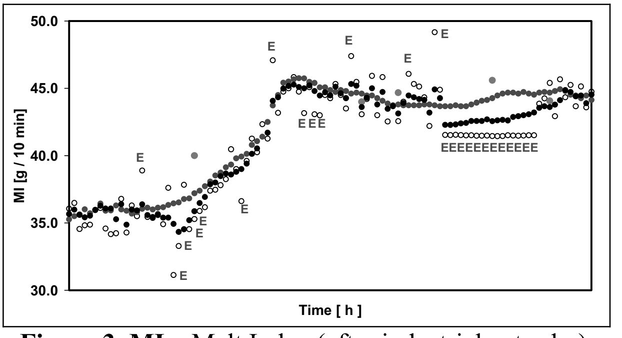


**Figure 3. MI** – Melt Index (after industrial extruder).

It is important to notice that the computational time required to accomplish all the calculations for each moving window (size=1h) using the *Welsch* robust estimator was roughly equal to 260s (250s for solution of the nonlinear DRPE problem with PS and 10s for data handling) in a standard Pentium 4 (single core 3.0 GHz) with 1024 MB memory, allowing for real time implementation of the code.

## 4. Conclusions

Some recent developments regarding the real time monitoring of industrial processes based on robust data reconciliation and gross error detection were reviewed. Particularly, a robust method for the joint nonlinear data reconciliation and parameter estimation with simultaneous gross error detection, based on the particle swarm optimization algorithm, was implemented. The proposed method was used to describe the operation of an industrial bulk propylene polymerization process and was validated with actual industrial data in-line and in real time. The negative influence of gross error was minimized with the help of a *Welsch* estimator. Off-line laboratory analyses confirmed the effectiveness of the proposed procedure, which indicated frequent failures of the process rheometer and allowed for improvement of the process performance. The numerical procedures required 260s of CPU in standard desktop computers, allowing for implementation of advanced process monitoring and control in real time.

## References

Z. H., Abu–El–Zeet, V. M., Becerra, P.D., Roberts, 2002, "Combined Bias and Outlier Identification in Dynamic Data Reconciliation", *Computers and Chemical Engineering*, v.26, pp. 921-935.

M., Bourouis, L., Pibouleau, P., Floquet, *et al*., 1998, " Simulation and data validation in multistage flash desalination plants", *Desalination.,*v.115, pp. 1-14.

J., Chen, A., Bandoni, J.A., Romagnoli, 1997, "Robust estimation of measurement error variance/covariance from process sampling data". *Computers and Chemical Engineering*, v. 21, pp. 593-600.

J., Chen, J. A., Romagnoli, 1998, "A Strategy for Simultaneous Dynamic Data Reconciliation and Outlier Detection", *Computers and Chemical Engineering*, v. 22, pp. 559-562.

X., Chen, R. W., Pike, T. A., Hertwig, J. R., Hopper, 1998, "Optimal Implementation of On-Line Optimization", *Computers and Chemical Engineering*, v.22, pp. S435-S442.

C. M., Crowe, 1996a, "Data Reconciliation - Progress and Challenges", *Journal Process Control*, v. 6, pp. 89-98.

C. M., Crowe, 1996b, "Formulation of Linear Data Reconciliation Using Information Theory", *Chemical Engineering Science*, v. 51, pp. 3359-3366.

R., Faber, H., Arellano-Garcia, P., Li, G., Wozny, 2007, "An optimization framework for parameter estimation of large-scale systems", *Chemical Engineering and Processing*, v. 46, pp. 1085-1095.

R., Faber, B., Li, P., Li, G., Wozny, 2006, "Data reconciliation for real-time optimization of an industrial coke-oven-gas purification process", *Simulation Modelling Practice and Theory*, v. 14, pp. 1121–1134.

Q., Jiang, M., Sánchez, M., Bagajewicz, 1999, "On the Performance of Principal Component Analysis in Multiple Gross Error Identification", *Industrial Engineering and Chemical Research*, v.38, pp.2005-2012.

B., Li, B., Chen, D., Duan, M., Kong, 2000, "Correction coefficient method for gross error detection based on temporal redundancy", *Qinghua Daxue Xuebao, Ziran Kexueban*, v.40, pp. 65-69.

T., Lid, S., Skogestad, 2008, "Scaled steady State Models for Effective On-Line Applications", *Computers and Chemical Engineering*, v.32, pp. 990-999.

K., Morad, W.Y., Svrcek, I., McKAY, 1999, "A robust direct approach for calculating measurement error covariance matrix". *Computers and Chemical Engineering*, v. 23, pp. 889-897.

S., Narashimhan, C., Jordache, 2000, *Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data*, Gulf Professional Publishing. Houston, TX,

D. B., Özyurt, R. W., Pike, 2004, "Theory and practice of simultaneous data reconciliation and gross error detection for chemical process", *Computers and Chemical Engineering*, v. 28, pp. 381-402.

D. M., Prata, E. L., Lima, J. C., Pinto, 2008d, "In-Line Monitoring of Bulk Polypropylene Reactors Based on Data Reconciliation Procedures", *Macromolecular Symposia*, v. 271, p. 26-37.

D. M., Prata, J. C., Pinto, E. L., Lima, 2008a, "Comparative Analysis of Robust Estimators on Nonlinear Dynamic data Reconciliation", *Computer-Aided Chemical Engineering*, v. 25, p. 501-506.

D. M., Prata, J. C., Pinto, E. L., Lima, 2008c, "Robust Data Reconciliation and Gross Error Detection through Particle Swarm Optimization for an Industrial Polymerization Reactor" (*submited*).

D. M., Prata, M., Schwaab, J. C., Pinto, E. L., Lima, 2008b, "Abordagem Híbrida para Detecção de Erros Grosseiros e Reconciliação Robusta de Dados: Aplicação Industrial", In*: Proceedings of the COBEQ 2008*, Recife, Setembro (*in portuguese*).

C., Reimers, J., Werther, G., Gruhn, 2008, "Flowsheet simulation of solids processes Data reconciliation and adjustment of model parameters", *Chemical Engineering and Processing*, v. 47, pp. 138–158.

J. A., Romagnoli, M. C., Sanchez, 2000, *Data Processing and Reconciliation for Chemical Process Operations*, Academic Press. San Diego.

M., Schladt, B., Hu, 2007, "Soft Sensor Based on Nonlinear Steady-State Data Reconciliation in the Process Industry", *Chemical Engineering and Processing*, v.46, pp. 1107-1115.

M., Schwaab, E. C., Biscaia Jr., J. L., Monteiro, J. C., Pinto, 2008, "Nonlinear Parameter Estimation through Particle Swarm Optimization", *Chemical Engineering Science*, v. 63, pp. 1542-1552.

P., Vachhani, R., Rengaswamy, V., Venkatasubramanian, 2001, "A Framework for Integrating Diagnostic Knowledge with Nonlinear Optimization for Data Reconciliation and Parameter Estimation in Dynamic Systems", *Chemical Engineering Science*, v. 56, pp. 2133-2148.

M., Wongrat, T., Srinophakun, P., Srinophakun, 2005, "Modified Genetic Algorithm for Nonlinear Data Reconciliation" *Computers and Chemical Engineering*, v. 29, pp. 1059-1067.

Z., Zhang, R.W., Pike, T., Herting, 1995, “Source Reduction from Chemical Plants Using On-line Optimization”, *Waste Management*, v. 15, pp.183-191.