18th European Symposium on Computer Aided Process Engineering – ESCAPE 18
Bertrand Braunschweig and Xavier Joulia (Editors)

Comparative Analysis of Robust Estimators on Nonlinear Dynamic Data Reconciliation

Diego Martinez Prata, José Carlos Pinto, Enrique Luis Lima

Programa de Engenharia Química/COPPE/UFRJ, Cidade Universitária, Rio de Janeiro - CP 68502, CEP 21945-970, Brasil.

Abstract

This paper presents a comparative performance analysis of various robust estimators used for nonlinear dynamic data reconciliation process subject to gross errors. Robust estimators based on cost functions derived from robust probability theory reduce the effect of gross errors on the reconciled data, avoiding the traditional iterative requirement procedures. The following robust probability functions were compared in this paper: Cauchy, Fair, Hampel, Logistic, Lorentzian, Normal Contaminated and Welsch. As a benchmark for this study it was adopted a nonlinear CSTR frequently reported in the process data reconciliation literature. The comparative analysis was based on the ability of the reconciliation approaches for reducing gross errors effect. Although the presence of constant biases has represented a problem for all the analyzed estimators, Welsch and Lorentzian cost functions, in this order, have shown better global performance.

Keywords: Nonlinear dynamic data reconciliation, robust estimation and gross error.

1. Introduction

Nowadays, data reconciliation (DR) represents an important step for many engineering activities in chemical processes as for example real time optimization and control implementations. It adjusts the measurement data, usually assumed associated to normally distributed random errors, to satisfy process constraints. However, to obtain satisfactory estimates, the negative influence of less frequently gross errors should be eliminated. This class of errors can be considered measurements that do not follow the statistical distribution of the bulk of the data. Gross errors can be divided in two classes: *outliers* and *bias*. The first class may be considered to include some abnormal behavior of measurement values as for example process leaks or malfunctioning instruments. The second class refers to the situation in which the measurement values are systematically too high or too low. A number of approaches have been proposed to deal with gross errors, mainly related to their detection and elimination. The traditional methods include serial elimination, compensation, and combinatorial ones, however these approaches are based on the assumption that the measurements are normally (*Gaussian*) distributed in which case

Corresponding author: enrique@peq.coppe.ufrj.br

Weighted Least Squares (WLS) is the maximum likelihood estimator. As gross errors do not satisfy this ideal assumption an iterative sequential procedure is necessary for gross error detection and elimination, increasing computational effort. Tjoa and Biegler (1991) proved that using the *Contaminated Normal* estimator instead of the *WLS* one, any outlier present in the measurements could be replaced with reconciled values, without requiring iterative detection and elimination procedures. Johnston and Kramer (1995) reported the feasibility and better performance of robust estimators when used to cope with DR problems in the presence of gross errors. Subsequently, different types of robust estimators and their performance on DR were reported (Table 1). These studies have shown the potential of robust statistics to perform DR in the presence of gross errors, resulting in robust estimators that are insensitive to deviations from ideal assumptions, tending to look at the bulk of the measured data and ignoring atypical values.

Table 1 Examples of Robust Estimators used for Data Reconciliation

Author (Year)	Estimator Applied
Tjoa and Biegler (1991)	*Normal Contaminated*
Johnston and Kramer (1995)	*Normal Contaminated and Lorenzian*
Zhang *et al*. (1995)[R]	*Normal Contaminated*
Albuquerque and Biegler (1996)[D]	*Normal Contaminated and Fair*
Chen *et al.* (1998)[R]	*Fair and Lorenzian*
Bourouis et al. (1998)[R]	*Normal Contaminated*
Arora and Biegler (2001)[D]	*Hampel*
Özyurt and Pike (2004)[R]	*Normal Contaminated, Cauchy, Fair, Logistic, Lorenzian and Hampel*
Wongrat *et al.* (2005)	*Hampel*
Zhou *et al.* (2006)	*Huber*

[R] Works applied on real plant data (steady state conditions). [D] Works applied in NDDR.

In our knowledge robust estimators have not been applied in nonlinear dynamic real plant data yet. The first comparative study among some robust estimators in DR has been presented by Özyurt and Pike (2004). They conclude that the estimators based on *Cauchy* and *Hampel* distributions give promising results, however did not consider dynamic systems. Other earlier studied has been accomplished by Basu and Paliwal (1989) in autoregressive parameter robust estimation issues, showing that for their case the *Welsch* estimator produced the best results.

This work presents a comparative performance analysis among some robust estimators (all estimators reported by Özyurt and Pike, 2004, and *Welsch* estimator) for nonlinear dynamic data reconciliation (NDDR in the presence of gross errors.

2. Problem formulation

The most important robust estimators for data reconciliation belong to the class of M-estimators, which are generalizations of the maximum likelihood estimator. Assuming uncorrelated measurement data their covariance matrix becomes diagonal and the generalized DR problem has the form,

$$\min \sum_i \rho\left(\frac{z_i - y_i}{\sigma_i}\right) = \min \sum_i \rho(\xi_i) \tag{01}$$

s. t.

$$\begin{aligned} & f\left[\frac{dy(t)}{dt}, y(t)\right] = 0 \\ & b\left[y(t)\right] = 0 \\ & g\left[y(t)\right] \geq 0 \end{aligned} \tag{02}$$

where ρ is any reasonable monotone function used for DR formulation, σ_i and ξ_i are, respectively, the standard deviation and the standard error of the discrete measured variable z_i, y is the vector of estimated functions y_i (reconciled measurements, model parameters and non-measured variables), f is a vector of dynamic constraints, h and g are, respectively, vectors of equality and inequality algebraic constraints.

As an example, using the generalized formulation the ρ functions for the weighted least squares and Welsh estimators take the following forms,

WLS $$\rho_{WLS}(\xi) = \frac{1}{2}\xi_i^2 \tag{03}$$

Welsch $$\rho_W(\xi, c_W) = \frac{c_W^2}{2}\left\{1 - \exp\left[-\left(\frac{\xi_i}{c_W}\right)^2\right]\right\} \tag{04}$$

where c_W is a tuning parameter related to asymptotic efficiency (Rey,1988).

Methods used to measure the robustness of an estimator involve an influence function (IF) that can be summarized by the effect of an observation on the estimates obtained (Arora and Biegler, 2001). The *Welsch* M-estimator introduced by Dennis and Welsch (1976) is a soft redescending estimator that, as the *Cauchy* estimator, presents an IF asymptotically approaching zero for large $|\xi|$. The 95% asymptotic efficiency on the standard normal distribution is obtained with the tuning constant $c_W = 2.9846$.

Figures 1 and 2 show, respectively, the effect of the standard error on the standardized ρ functions and influence functions for the *WLS* and *Welsch* estimators. It can be observed in both figures that the robust estimator is much less influenced by large errors.

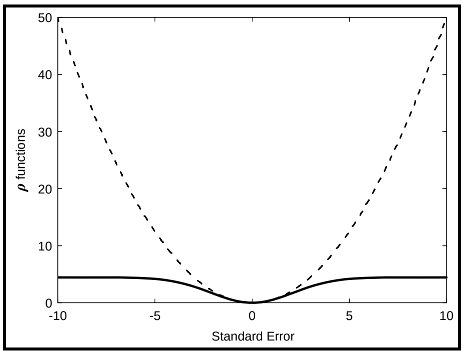

Fig. 1. ρ: *WLS* (---) and *Welsch* (—).

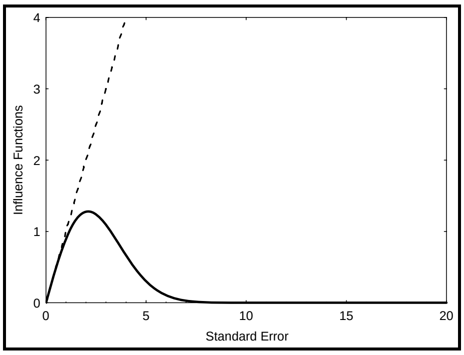

Fig. 2. IF: *WLS* (---) and *Welsch* (—).

Several strategies have been proposed to solve constrained nonlinear dynamic programming problems (Biegler and Grossman, 2004). In this work a sequential strategy is applied to a time moving window (size = 5). For every sample time the differential equations of the dynamic constraints and the nonlinear programming optimization problem are solved sequentially using the measured data over the window until convergence is reached. The optimization problem is solved using the Gauss-Newton based solver ESTIMA (Noronha *et al*., 1993).

3. Illustration example

The performance of the robust estimators has been tested on the same CSTR used by Liebman *et al*. (1992) where the four variables in the system were assumed to be measured. The two input variables are the feed concentration and temperature while the two state variables are the output concentration and temperature. Measurements for both state and input variables were simulated at time steps of 1 (scaled time value corresponding to 2.5 s) adding *Gaussian* noise with a standard deviation of 5% of the reference values (see Liebman *et al.*, 1992) to the "true" values obtained from the numerical integration of the reactor dynamic model. Same *outliers* and a *bias* were added to the simulated measurements. The simulation was initialized at a scaled steady state operation point (feed concentration = 6.5, feed temperature = 3.5, output concentration = 0.1531 and output temperature = 4.6091). At time step 30 the feed concentration was stepped to 7.5.

Due to space limitations, only results of output concentration and temperature for the *WLS* and *Welsch* estimators are presented in Figures 3, 4, 5 and 6. The symbols (—), (○) and (●) represent the "true", simulated and reconciled data, respectively. The output temperatures plotted have been magnified to emphasize the effect of the bias on their estimates.

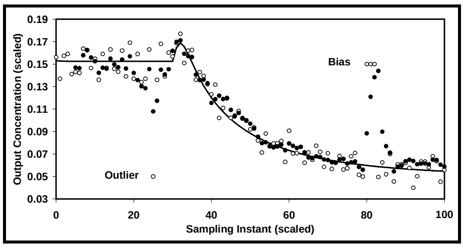

Fig. 3. ***WLS*****:** Output Concentration.

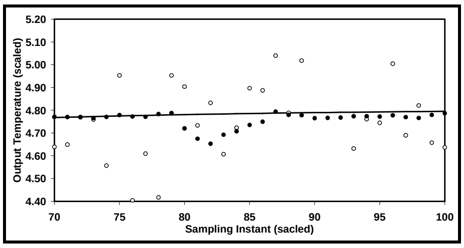

Fig. 4. WLS: Output Temperature.

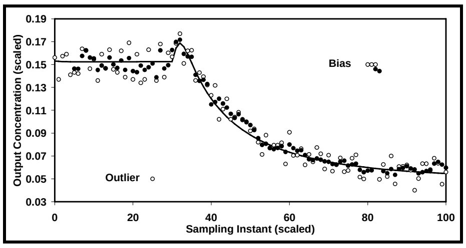

Fig. 5. Welsch: Output Concentration.

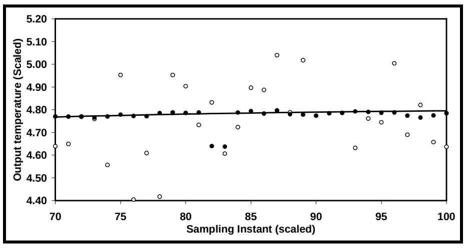

Fig. 6. Welsch: Output Temperature.

Comparing Figures 3 and 5 it can be seen that in the presence of an outlier in sampling time 25 the reconciled output concentrations using the robust *Welsch* estimator are better that the ones using the *WLS* estimator, which presents *smearing* values around this sampling time. However, even a robust estimator can result in biased estimates in the presence of a bias as can be seen around sampling times 80-82. In this work the time varying window always corresponds to measured values. However if the time varying window is built with the measured values at the current sample time and the already reconciled values at past sample times the effect of a bias will be minimized. Figures 4 and 6 show the effect of bias measurements in the reconciled values of the output temperature, and again the *WLS* estimator results in worst estimates.

Looking for a fair comparison among the estimators it was used the TER (Total Error Reduction) criteria proposed by Serth *et al*. (1987) that can be applicable when the “true” values are known. Table 2 summarizes the results obtained and shows best results for the *Welsch* and *Lorentzian* estimators.

Table 2. TER analysis results for the estimators studied.

Estimator Applied	Output Concentration	Output Temperature
WLS	0.2040	0.9501
Normal Contaminated	0.2885	0.9635
Cauchy	0.3667	0.9631
Fair	0.4072	0.9632
Hampel	0.3953	0.9622
Logistic	0.3633	0.9628
Lorentzian	0.4290	0.9655
Welsch	0.4724	0.9657

4. Conclusions

In this work a comparative analysis of the capacity of robust estimators to reduce the negative effect of gross errors on nonlinear dynamic data reconciliation was accomplished. The results obtained have shown that among the studied cases the *Welsch* and *Lorentzian* robust estimators produced better reconciled values, but they also have shown that, although the robust estimators were more efficient in reducing the effect of biases, this problem still deserves more investigation.

References

Albuquerque, J. S., Biegler, L. T., 1996. Data Reconciliation and Gross-Error Detection for Dynamic Systems. AIChE J. 42, 2841.

Arora, N., Biegler, L. T., 2001. Redescending estimators for Data Reconciliation and Parameter Estimation. Comput. Chem. Eng. 25, 1585.

Basu, A., Paliwal, K. K., 1989. Robust M-Estimates and Generalized M-estimates for Autoregressive Parameter Estimation. TENCON 89, 4th IEEE region 10 Int. conf., Bombay.

Biegler, L. T., Grossmann, I. E., 2004. Retrospective on optimization. Comput. Chem. Eng., 28, 1169.

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

Chen, X., Pike, R. W., Hertwing, T. A., Hopper, J. R., 1998. Optimal Implementation of On-Line Optimization. Comput. Chem. Eng. 22, S435.

Dennis, J. E., Welsch, R. E., 1976. Techniques for Nonlinear Least Squares and Robust Regression. Proc. Amer. Statist. Ass. 83-87.

Johnson, L. P. M., Kramer, M. A., 1995. Maximum Likelihood Data Rectification: Steady-State Systems. AIChE J. 41, 2415.

Liebman, M. J., Edgar, T. F., Lasdon, L. S., 1992. Efficient Data Reconciliation and Estimation for Dynamic Processes Using Nonlinear Programming Techniques. Comput. Chem. Eng. 16, 963.

Noronha, F.B., Pinto, J.C., Monteiro, J.L., *et al.*. 1993. Um Pacote Computacional para Estimação de Parâmetros e Projeto de Experimentos, Technical Report, PEQ/COPPE/UFRJ.

Özyurt, D. B., Pike, R. W., 2004. Theory and practice of simultaneous data reconciliation and gross error detection for chemical process. Comput. Chem. Eng. 28, 381.

Rey, W. J. J., 1983. *Introduction to Robust and Quasi-Robust Statistical Methods*. Springer-Verlang, Berlin/ new York.

Serth, R. W., Valero, C. M., Heenan, W. A., 1987. Detection of gross errors in nonlinearly constrained data: a case study. Chem. Eng. Commun. 51, 89.

Tjoa, I. B., Biegler, L. T., 1991. Simultaneous Strategy for Data Reconciliation and Gross Error Detection of Nonlinear Systems. Comput. Chem. Eng. 15, 679.

Wongrat, M., Srinophakun, T. Srinophakun, P., 2005. Modified genetic algorithm for nonlinear data reconciliation. Comput. Chem. Eng. 29, 1059.

Zhang, Z., Pike, R.W., Herting, T., 1995. Source reduction from chemical plants using on-line optimization. Waste Management. 15, 183.

Zhou, L., Su, H., Chu, J., 2006. A new method to solve robust data reconciliation in nonlinear process. Chinese J. Chem. Eng. 14, 357.

Acknowledgements

The authors would like to thank CNPq and CAPES for financial support.