Acta Scientiarum

http://www.uem.br/acta
ISSN printed: 1806-2563
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v28i2.28188

Evaluation of robust functions for data reconciliation in thermal systems

Regina Luana Santos de França[1*], Antônio Martins de Oliveira Júnior[1] and Domingos Fabiano de Santana Souza[2]

*[1]Programa de Pós-graduação em Engenharia Química, Universidade Federal de Sergipe, Avenida Marechal Rondon, s/n, Campus Universitário José Aloisio de Campos, 49100-000, São Cristóvão, Sergipe, Brazil.[2]Departamento de Engenharia Química, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil. *Author for correspondence. E-mail: reginaquimica@gmail.com*

ABSTRACT. Process variables regularly control and evaluate industrial processes. Information with gross errors may in some cases not be attenuated by function reconciliation and change the calculation of process balance, leading optimization results towards non-feasible regions or to optimal sites. A promising alternative for reconciling functions is the use of robust functions. Current paper considers the above scenario and evaluates the fitness of some robust functions in solving in steady state chemical processes data reconciliation problems represented by linear and nonlinear systems in the presence of gross errors. Traditional Cauchy, Fair, Contaminated Normal and Logistic robust functions are used in the reconciliation problem where their estimates are compared to those obtained with the use of the latest features, such as New Target and Alarm. Rates for gross errors in tests were limited between 4 and 10σ of the measured current and elaborated a region of outliers. Results showed that New Target and Alarm functions are different from the others as the magnitude of the gross error increases, tending towards true rates specified by set point.

Keywords: set measures, robustness, efficiency, nonlinear systems.

Avaliação de funções robustas para reconciliação de dados em sistemas térmicos

RESUMO. Variáveis de processo são regularmente utilizadas para controlar e avaliar os processos industriais. As informações contendo erros grosseiros podem, em alguns casos não ser atenuadas por funções reconciliadoras e, desta forma, alterar os cálculos de balanços de processo, conduzindo os resultados de otimização a regiões inviáveis ou a ótimos locais. Neste sentido, uma alternativa promissora para função reconciliadora é a utilização das funções robustas. O presente trabalho leva em consideração este cenário e avalia a aptidão de algumas funções robustas na resolução de problemas de reconciliação de dados em processos químicos em estado estacionário representados por sistemas lineares e não lineares na presença de erros grosseiros. As tradicionais funções robustas Cauchy, Fair, Normal Contaminada e Logística são utilizadas, onde, as suas estimativas são comparadas com as obtidas com o uso de funções mais recentes, como a New Target e Alarm. Nos testes realizados, os valores gerados para os erros grosseiros foram limitados entre 4 a 10σ da medida da corrente, configurando assim uma região de grandes desvios. Os resultados indicaram que as funções New Target e Alarm diferenciam-se das demais à medida que aumenta a magnitude do erro grosseiro indicando uma tendência ao valor verdadeiro especificado pelo set-point.

Palavras-chave: ajuste de medidas, robustez, eficiência, sistemas não lineares.

Introduction

Data reconciliation may be seen as a step towards improving the accuracy of data for use in modeling and optimization processes. According to technique precursors Kuehn and Davidson (1961), data reconciliation is a tool that, among a wide range of applications, allows optimum adjustment measures and estimates based on spatial redundancy and model conceived by conservational balances of mass and energy. Its effectiveness for accuracy is achieved even when the estimated standard deviation of errors is different from the actual ones or when errors follow standard distribution (Jiang, Liu, & Li, 2014). However, errors, especially gross ones, cannot be accurately explained or predicted.

Robust functions have been studied by Huber and Ronchetti (2009) with the use of robust statistical tools to find solutions to problems that lacked the normal Gaussian distribution. An important feature in the reconciliation procedure is the low sensitivity of these functions when the data are corrupted with gross errors (Ozyurt & Pike, 2004).

The ability to predict robustness and to process data with gross errors has been dealt with in many research papers. Ozyurt and Pike (2004) tested the robust functions in case studies mentioned in the literature to establish a criterion for error detection and data reconciliation simultaneously. The good results obtained with different functions underscored the efficiency of Hampel, Cauchy and Logistics functions. Prata, Schwaab, Lima, and Pinto (2010) analyzed the ability to predict robust Welsch function for data reconciliation and detection of outliers in a propylene polymerization reactor represented by a nonlinear dynamic model. Zhang, Shao, Chen, Wang, and Qian (2010) analyzed the robustness of the Least Squares Quasi-weighted function by comparing it to the Weighted Least Square, Fair and Redescending functions. Loss of quality of the Weighted Least Squares function was evident when compared to the proposed function. The focus of the study was to analyze the potential of this strategy for the application of online detection of gross errors. Nicholson, Negrete, and Lorenz (2014) used a sequential approach to numerical integration of nonlinear dynamic models and data reconciliation by using Huber, Fair and Hampel redescending robust functions in an advanced method of moving horizon. Process measures were "contaminated" with large errors to demonstrate the ability to reconcile the robust functions used in data reconciliation.

Even with a motivating panorama in this thematic area, there is still no consensus on the criteria for the selection of robust features. In fact, few studies have been proposed to develop and evaluate new robust functions. Jin, Hung, and Liu (2012) proposed a new formulation called New Target to reconcile data. The reconciliation of results obtained with this function was compared to those in the Cauchy functions and Huber and showed that the New Target provided more accurate data especially when there were one or two gross errors in the measurements.

Alamgir, Khan, Khan, and Khalil (2013) proposed a function called Alamgir Redescending M-estimator (Alarm) based on the modification of the hyperbolic tangent function. Even without having been evaluated in a data reconciliation problem, the results obtained by the Alarm function in the detection of outliers were compared to results of such robust functions such as Tukey Biweight, Andrew Sine, Hampel of three parts, Huber and Ordinary Least Squares, and showed that the increased robustness of the other estimators for the elimination of outliers decreased their efficiency, which was not the case with the Alarm function.

Despite the good results obtained from the New Target, the prediction capacity of this new function only coped with two robust functions. In fact, the question is how it behaves when compared with other robust features including Alarm function which showed good performance in the detection of outliers.

In the case of thermal systems, Jiang et al. (2014) proposed to apply data reconciliation and gross errors detection in steam turbine on-line of a coal-fired power generation unit. Results showed whether data reconciliation had contributed towards the reduction of uncertainties of the estimated rate of primary flow rates, steam turbine heat rate and the heat rate sensitivity coefficients. Martínez-Maradiaga, Bruno, and Coronas (2013) studied data reconciliation technique to the steady state operational data for absorption refrigeration systems and a single-effect ammonia-water absorption chiller was observed. Data reconciliation technique was executed with gross error detection procedure and efficient results were obtained in two steps. The identification and removal of gross errors enabled a data reconciliation step consistent with results. Szega and Nowak (2015) simulated a model of the thermal system of power unit for data reconciliation and proposed a mathematical simulation model of a supercritical steam power unit. The data reconciliation step relied on redundancy and observable measurements. Reconciled data show a decrease in the system´s entropy.

Current paper compares the classical Cauchy, Fair, Contaminated Normal and Logistic functions, applicable to data reconciliation problems containing gross errors with the new robust functions, New Target and Alarm. The key aim is to evaluate the potential of these new functions for real time applications. The criterion for selection of functions was based on the average relative error between the amount reconciled and the true rate for the variable. The parameter setting of robust functions was obtained from the literature with a 95% confidence interval. Two cases represented by model processes in steady states, linear and nonlinear, were selected for current analysis. Data were corrupted with gross errors between four and ten times the standard deviation rate randomly generated to simulate real-time behavior.

Methodology

Formulation of data reconciliation problem with robust functions

The data reconciliation problem for processes in steady states represents a class of optimization problem (Sanchez & Romagnoli, 1996) formulated by minimizing an objective function subject to constraints. The optimization problem for data reconciliation may be represented by Equation 1.

$$\begin{aligned}
&Minimize \quad \sum_{i=1}^{N_{med}} \rho(\varepsilon_i) \\
&Subject\ to: \\
&\quad f_k(x,z)=0 \qquad k\in N \\
&\quad g_l(x,z)\geq 0 \qquad l\in N_{des} \\
&\quad x_i^{\inf}\leq x_i\leq x_i^{\sup} \quad i\in N_{med} \\
&\quad z_j^{\inf}\leq z_j\leq z_j^{\sup} \quad j\in N_{unmed}
\end{aligned} \tag{1}$$

where:

N_{med} is the number of measured variables;

N is the number of equality constraints, in this case, the issue number of equations;

N_{des} is the number of inequality constraints;

N_{unmed} is the number of unmeasured variables;

$\rho(\varepsilon_i)$ is the merit function;

f is equality constraint, in this case, the mass balance or energy;

g is the inequality constraint imposed on the problem;

z is the unmeasured variables of the process, estimated concomitantly with reconciliation.

The variables reconciled and unmeasured are limited by $x_i^{\inf}$, $z_j^{\inf}$ (in the lower region) and $x_i^{\sup}$, $z_j^{\sup}$ (in the upper region). In the merit function, ε represents the relative error between the measured and reconciled rates, as shown in Equation 2, with x^m = the measured variables of the process; x^r = the reconciled rates for the variables of the process and σ = standard deviation of the measurements.

$$\varepsilon=\left(\frac{x^m-x^r}{\sigma}\right) \tag{2}$$

In the case of reconciliation data with robust functions, the merit function is modified according to the functional form of interest, i.e. if the robust function is a weighted least square function, the merit function is represented according to Equation 3.

$$\rho(\varepsilon_i)=\frac{1}{2}\cdot\varepsilon_i^2 \tag{3}$$

In an assessment of the functions used in the literature, it became clear that the robust functions employed belonged to the redescending M-estimators families which in turn are based on the maximum likelihood function (Hodouin & Everell, 1980). Due to this agreement, the functions Cauchy, Fair, Contaminated Normal, Logistic and new functions New Target and Alarm were selected to analyze the data reconciliation problem, represented by Equations 4 to 10, under function merit.

Weighted least squares (WLS)

$$\frac{1}{2}\cdot\varepsilon_i^2 \tag{4}$$

Cauchy

$$c_c^2\cdot\ln\left(1+\frac{\varepsilon_i^2}{c_c^2}\right) \tag{5}$$

Fair

$$2\cdot c_f^2\cdot\left[\frac{|\varepsilon_i|}{c_f}-\ln\left(1+\frac{|\varepsilon_i|}{c_f}\right)\right] \tag{6}$$

Contaminated Normal

$$-\ln\left[(1-p_{CN})\cdot e^{\left(-\frac{\varepsilon_i^2}{2}\right)}+\frac{p_{CN}}{b_{CN}}\cdot e^{\left(-\frac{\varepsilon_i^2}{2\cdot b_{CN}^2}\right)}\right] \tag{7}$$

Logistic

$$2\cdot\ln\left(1+e^{\left(\frac{\varepsilon_i}{c_{Lo}}\right)}\right)-\left(\frac{\varepsilon_i}{c_{Lo}}\right) \tag{8}$$

New Target

$$\begin{cases} \dfrac{c_{NT}^2}{6}\cdot\left[1-\left(1-a\cdot\dfrac{\varepsilon_i^2}{c_{NT}^2}\right)\right] & |\varepsilon|\le c_{NT} \\ A\cdot\ln\left(\dfrac{\varepsilon_i^2}{c_{NT}^2}\right)+B & |\varepsilon|\ge c_{NT} \end{cases} \tag{9}$$

Alarm Redescending

$$\begin{cases} \dfrac{2\cdot c_{Al}^2}{3}\cdot\left\{1-\dfrac{2\cdot\left(1+3\cdot e^{\left(\frac{\varepsilon_i}{c_{Al}}\right)^2}\right)}{\left(1+e^{\left(\frac{\varepsilon_i}{c_{Al}}\right)^2}\right)^3}\right\} & |\varepsilon|\le c_{Al} \\ \dfrac{2\cdot c_{Al}^2}{3}\cdot\left\{1-\dfrac{2\cdot\left(1+3\cdot e^1\right)}{\left(1+e^1\right)^3}\right\} & |\varepsilon|\ge c_{Al} \end{cases} \tag{10}$$

In Equation 5 to 10, the variables c_c, c_f, p_{CN}, b_{CN}, c_{Lo}, c_{NT}, a, A, B and c_{Al} are set parameters of robust functions. Parameters may be estimated according to a known distribution (normal) and with specified accuracy, or even estimated together with the reconciliation problem. Rates in Table 1 were used for the study of cases of interest. They were retrieved from Ozyurt and Pike (2004), Jin et al. (2012) and Alamgir etal. (2013), with 95% efficiency.

Table 1. Constants for different ρ functions, with 95% efficiency.

Function ρ	Set parameters
Cauchy	$c_C = 2.3849$
Fair	$c_F = 1.3998$
Contaminated Normal	$b_{CN} = 10$ $p_{CN} = 0.235$
Logistic	$c_{Lo} = 0.602$
New Target	$c_{NT} = 3.$ $A = 0.65$
Alarm	$c_{Al} = 3$

Figure 1 denotes the performance of different robust functions in the analysis of the influence function (IF) representing the sensitivity of the function with respect to error or contamination related to gross errors and outliers (Zhang et al., 2010; Prata, Schwaab, Lima, & Pinto, 2010). The behavior designed to influence the function is that as the error increases, the function converges to a small, constant rate, and becomes indifferent to contamination by big errors. Figure 1 reveals that all robust functions, Contaminated Normal, New Target and Alarm, have the same behavior within the minimum region, when the influence function approaches a value close to zero.

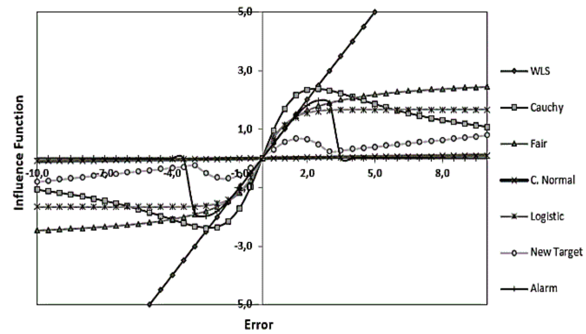

Figure 1. Behavior of influence functions with regard to error (Table 1 setting parameters).

In defining performance criterion for the selection of functions, Equation 11 represents aspects, such as convergence, and reduction of relative error (Zhang et al., 2010). The first aspect indicates whether the function may be employed for real time applications; the latter aspect refers to the ability to function closer to true rates and may be a possible indicator for gross error detection procedures.

$$RER = \frac{\sum_i \left(MRE_i - RRE_i\right)}{\sum_i MRE_i}\times 100 \tag{11}$$

In Equation 11, MRE is the measure´s relative error and RRE is the relative reconciled error, represented by Equation 12 and 13.

$$MRE_i = \frac{\left|x_i - x_i^m\right|}{x_i} \tag{12}$$

$$RRE_i = \frac{\left|x_i - x_i^r\right|}{x_i} \tag{13}$$

where:

x_i is the true rate;
x_i^m is the measured rate;
x_i^r is the reconciled rate.

Cases of interest in current paper have been implemented in a computing environment and reconciled by a nonlinear programming strategy as optimizer (Narashimham & Jordache, 2000). A final tolerance of the objective function 10^{-6} was specified for all.

Results

The two cases analyzed deal with a processing unit studied by Narashimham and Jordache (2000) and Knopf (2012), comprising a mixer, heat exchanger, splitter and recycle controlled by a bypass valve. The first application study is represented by a linear system of equations and has six streams measured with a single gross error. The energy flow is discarded and only mass flow is taken into consideration. The second application study is a modification of the previous case, with energy and mass flow. The problem represented by a nonlinear system of equations has eleven streams measurements (flows and temperatures) and two gross errors. Figure 2a and b illustrate the two cases, especially the nonlinear system in which streams of energy for the heat exchanger may be noted.

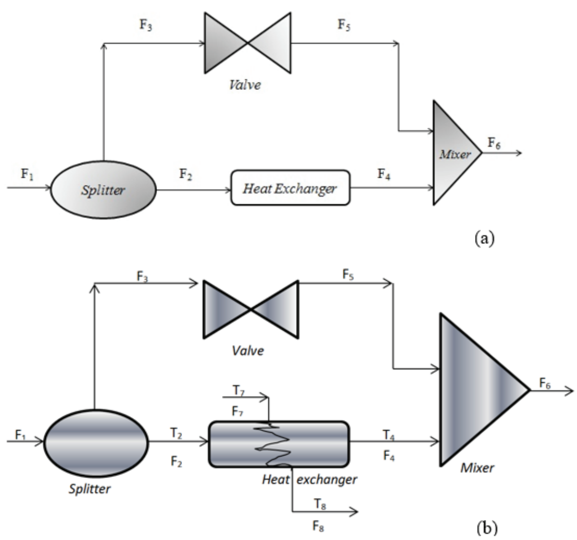

Figure 2. a) Linear model of process and b) nonlinear model of process.

Application study 1: linear model

Flow sheet details are shown in Figure 2a and the information on the measured variables are presented in Table 2 in which a gross error in stream F_2 referring to the exit of the separator and entrance of the heat exchanger, with magnitude 4σ, may be perceived.

Table 2. Information on measured variables in application Study 1, with standard deviation equal to 1.

Streams	True Flow Rates	Measured Flow Rates
F_1	100	101.9
F_2	64	68.45
F_3	36	34.65
F_4	64	64.20
F_5	36	36.64
F_6	100	98.88

Figure 3 demonstrates that robust feature WLS showed a low 19% relative error reduction. Other robust features that have proven performance in the literature, such as Contaminated Normal, Fair and Logistic, were also below 50%. On the other hand, New Target and Alarm functions showed a high 70% error reduction which proved to be more accurate than the other functions.

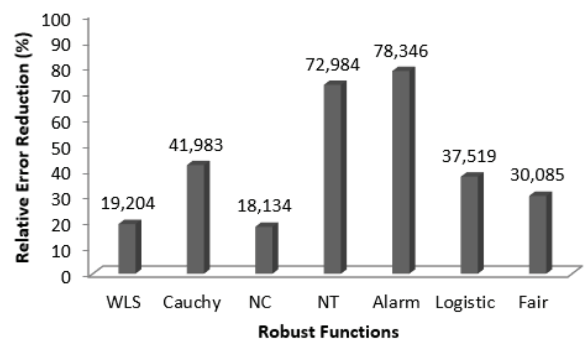

Figure 3. Comparison of the performance of robust functions with regard to reduction of relative error – Application Study1.

Mitigation of inability of Fair and Logistics functions for great errors may be justified by Figure 1. After a certain magnitude of error, the IF rate of the above functions reaches constant and high level. This behavior has not been observed in the New Target and Alarm Functions that tend to zero posterior to the referred error. This aspect confirms the robustness idealized for the new functions especially in the presence of big errors.

An important point to be highlighted was the lack of robustness of the Contaminated Normal function in this case. Although not commented in the literature, the possible loss of robustness may be attributed to the adjustment parameters of the function.

Application study 2: nonlinear model

Figure 2b demonstrates information on energy balances. Table 3 shows the rates of the measured variables. Two gross errors may be observed in the streams F_2 and T_7, with regard to the input stream of the hot flow in the heat exchanger and the inlet temperature of the cold flowof heat exchanger respectively with magnitude 4 and 5σ.

Figure 4 shows that the function New Target has a less than 15% rate to reduce relative error and indicates that the reconciled rate approaches the measured value, albeit not mitigating the gross error. The other functions showed rates close to 30% for the reduction of relative error. However, tests on the magnitude of errors close to those in current test and big errors in the two streams also obtained an error reduction between 26 and 35%, which was considered satisfactory (Zhang et al., 2010).

Table 3. Information on measured variables in application Study 2, with standard deviation equal to 1.

Streams	True rates	Measured rates
F_1	100	101.91
F_2	64	68.45
F_3	36	34.65
F_4	64	64.20
F_5	36	36.44
F_6	100	98.88
F_7	140.6	140.0
T_2	90	90
T_4	51	51
T_7	20	25
T_8	43	43

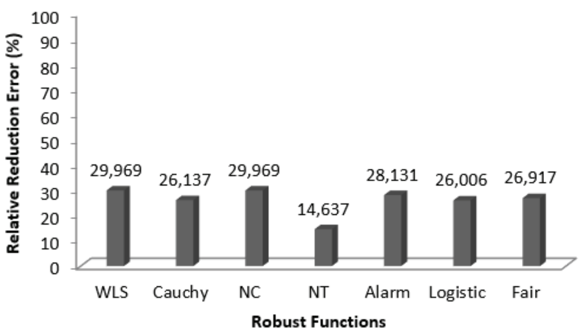

Figure 4. Comparison of performance of robust functions with regard to reduction of relative error for application Study 2.

A new test was performed at a magnitude of gross error 10σ for stream F_2 and T_7. Table 4 gives the new rates.

Table 4. Information on measured variables with gross error of 10.σ with standard deviation equal to 1.

Streams	True rates	Measured rates
F_1	100	101.91
F_2	64	74.20
F_3	36	34.65
F_4	64	64.20
F_5	36	36.44
F_6	100	98.88
F_7	140.6	140.0
T_2	90	90
T_4	51	51
T_7	20	30.41
T_8	43	43

Figure 5 reveals that rise in the magnitude of error in streams F_2 and T_7 resulted in an increase in the reduction of the relative error of functions Alarm and New Target. In this case, a reduction of over 90% occurred.

The behavior by New Target function has been reviewed by Jin et al. (2012) who highlighted increasing function of accuracy according to the increased magnitude of the error. However, the above performance was not followed by the other functions which showed a lower than 10% reduction, with the exception of Contaminated Normal function with an 80% reduction. Although developed for outlier detection, Alarm function also presented an error reduction according to increasing magnitude. In fact, it is a promising alternative for data reconciliation problems. When compared to other functions, Alarm proved to have consistent results, without any loss of accuracy in situations where the magnitude of the error was small.

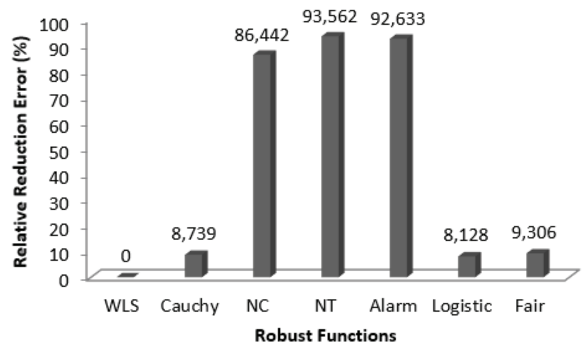

Figure 5. Comparison of performance of robust functions with regard to the reduction of relative error with 10.σ.

Conclusion

A general analysis of robust functions New Target and Alarm with regard to data reconciliation problem in linear and nonlinear systems showed great precision for gross errors between 4 - 5 times the standard deviation. New Target lost efficiency and demonstrated a different behavior to that presented by Jin et al. (2012). In the same context, the Alarm function was shown to have efficiency compatible with the Normal Contaminated function. When gross error increased, New Target and Alarm functions revealed higher results than those of functions and indicated a trend to mitigate major errors.

Acknowledgements

The authors would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) and to the Universidade Federal de Sergipe (UFS) for funding current research.

References

Alamgir, A. A., Khan, S. A., Khan, D. M., & Khalil, U. (2013). A new efficient redescending M-estimator: Alamgir Redescending M-estimator. *Research Journal of Recent Sciences, 2*(8), 79-91.

Hodouin, D., & Everell, M. D. A. (1980). Hierarchical procedure for adjustment and material balancing of mineral processes data. *Journal Mineralogy Processing, 7*(2), 91-116.

Huber, P. J., & Ronchetti, E. M. (2009). *Robust statistics*. Hoboken, NJ: John Wiley & Sons Inc.

Jiang, X., Liu, P., & Li, Z. (2014). A data reconciliation based framework for integrated sensor and equipment performance monitoring in power plants. *Applied Energy, 134*(1), 270-282.

Jin, S., Hung, L. X., & Liu, M. A. (2012). A New Target function for robust data reconciliation. *Industrial & Engineering Chemistry Research, 51*(30), 10220-10224.

Knopf, F. C. (2012). *Modelling, analysis and optimization of process and energy systems*. Baton Rouge, NJ: John Wiley & Sons Inc.

Kuehn, D. R., & Davidson, H. (1961). Computer control. II. Mathematics of control. *Chemical Engineering Progress, 57*(6), 44-47.

Martínez-Maradiaga, D., Bruno, J. C., & Coronas, A. (2013). Steady-state data reconciliation for absorption refrigeration systems. *Applied Thermal Engineering, 51* (1-2), 1170-1180.

Narashimham, S., & Jordache, C. (2000). *Data reconciliation and gross error detection an intelligent use of process data*. Houston, TX: Gulf Publishing Company.

Nicholson, B., Negrete, R. L., & Lorenz, T. B. (2014). On-line state estimation of nonlinear dynamic systems with gross errors. *Computers and Chemical Engineering, 70*(5), 149-159.

Ozyurt, D. B., & Pike, R. W. (2004). Theory and practice of simultaneous data reconciliation and gross error detection for chemical process. *Computers and Chemical Engineering, 28*(3), 381-402.

Prata, D. M., Schwaab, M., Lima, E. L., & Pinto, J. C. (2010). Simultaneous robust data reconciliation and gross error detection through particle swarm optimization for an industrial polypropylene reactor. *Chemical Engineering Science, 65*(17), 4943-4954.

Sanchez, M., & Romagnoli, J. (1996). Use of orthogonal transformations in data classification-Reconciliation. *Computers Chemical Engineering, 20*(5), 483-493.

Szega, M., & Nowak, G. T. (2015). An optimization of redundant measurements location for thermal capacity of power unit steam boiler calculations using data reconciliation method. *Energy, 6*(92), 1-7.

Zhang, Z., Shao, Z., Chen, X., Wang, K., & Qian, J. (2010). Quasi-weighted least squares estimator for data reconciliation. *Computers and Chemical Engineering, 34*(2), 154-162.

Received on June 16, 2015.
Accepted on October 23, 2015.