

ON-LINE PROCESS MONITORING USING A ROBUST STATISTICS BASED METHODOLOGY

C.E. LLANOS[†], R.J. CHÁVEZ GALLETTI[†], M.C. SÁNCHEZ[†] and R. MARONNA[‡]

[†]*Depto de Ing. Química, Univ. Nac. del Sur(UNS), Planta Piloto de Ing. Química-PLAPIQUI(UNS-CONICET), 8000 Bahía Blanca, Argentina. msanchez@plapiqui.edu.ar*

[‡]*Dto. de Matemática, Univ. Nac. de la Plata, 1900 La Plata, Argentina. rmaronna@retina.ar*

Abstract— Robust Data Reconciliation strategies provide unbiased variable estimates in the presence of a moderate quantity of measurement gross errors. Systematic errors which persist in time, as biases or drifts, overcome this quantity causing the deterioration of the estimates. This also occurs due to the presence of process leaks. The fast detection of those faults avoids the use of biased solutions of the data reconciliation procedure, and allows to perform quick corrective actions. In this work, a methodology for leak detection is incorporated into a robust data reconciliation procedure that detects and classifies systematic observation errors. The strategy makes use of the Robust Measurement Test, to detect outliers and leaks, and the Robust Linear Regression of the data contained in a moving window to distinguish between biases and drifts. The methodology is applied for two benchmarks extracted from the literature. Results highlight the performance of the proposed strategy.

Keywords— Measurement Errors, Leak Detection, Robust Data Reconciliation.

I. INTRODUCTION

Taking relevant information from measurements is the key challenge to achieve better results of process control and optimization methodologies (Romagnoli and Sánchez, 2000). With this aim, diverse monitoring techniques have been developed, which allow obtaining reliable variable estimates on-line.

The well-known Data Reconciliation (DR) procedure is a technique that provides precise estimates when measurements follow exactly a certain probability distribution, for example the normal distribution. Typically, the DR uses the Least Squares function to mitigate the discrepancy between measurements with random errors and model equations. But, the presence of atypical observations introduces deviations from the distribution considered that bias the solution of that procedure.

Robust DR (RDR) can provide reliable estimates when data follows approximately a probability distribution model (Maronna *et al.*, 2006). Thus, if a robust estimator is used as objective function of the DR problem instead of the Least Square one, unbiased estimates can be calculated even in the presence of a moderate quantity of systematic errors (SE).

Nevertheless, if SE persist in time the Break Down Point (BDP) of the estimates calculated using robust methodologies is exceeded. Roughly speaking, the BDP of an estimate is the largest amount of contamination (proportion of atypical values) that the data may contain

such that it still gives an unbiased information about the variable value. Thus, an early detection and classification of Systematic Errors that persist in time (SEPT) allows taking corrective actions to reduce the Mean Square Error of the estimates (Llanos *et al.*, 2017).

Different M-estimator functions (ρ), which are generalizations of the Maximum Likelihood Estimator were evaluated and compared when measurements contain outliers (Özyurt and Pike, 2004; Martinez Prata *et al.*, 2010; Chen *et al.*, 2013). Furthermore, procedures which combine the strengths of redescending and monotone M-estimators were presented and compared with others (Llanos *et al.*, 2015). It was concluded that a Simple Methodology, called SiM, was an efficient solution method, taking into account the estimation accuracy and the computational load.

The detection of SEPT applying tools of Robust Statistics was explored by Martinez Prata *et al.* (2010), who worked on the identification of outliers and biases using the Welsch estimator. Moreover, the Correntropy estimator was used to differentiate between outliers, biases and drifts (Zhang and Chen, 2015). In this contribution, the analysis was based on a threshold variance. The aforementioned researches were limited to the classification of SEPT. A wide performance evaluation was presented by Llanos *et al.* (2017). These authors developed a methodology based on the Robust Measurement Test (RMT) that was able to detect and identify observations with SE. Also the Robust Linear Regression (RLR) was used to classify the SEPT, and corrective actions were proposed for faulty measurements. These three works only considered the presence of SE.

Leaks represent model errors and therefore affect the constraint equations (Narasimhan and Jordache, 2000). Narasimhan and Mah (1987) developed a methodology that was able to detect and estimate leaks using the Generalized Likelihood Ratio Test. Others researchers have proposed strategies with the same objective (Sánchez *et al.*, 1999). However, the presence of process leaks was not addressed using RDR.

In this work, the methodology developed by Llanos *et al.* (2017) is extended to detect the occurrence of process leaks. It is proposed to use the RMT for detecting outliers and leaks, and to apply the RLR of the data contained in a moving window for classifying the SEPT as biases or drifts. Two processes under steady state operation, which are extracted from literature, are used to test the behavior of the methodology. A comprehensive analysis of the strategy performance is provided in this article.

II. METHODOLOGY

A. Robust Data Reconciliation Problem

Let define the RDR problem as follows

$$[\hat{\mathbf{x}}_j^R, \hat{\mathbf{u}}_j^R] = \min_{\mathbf{x}_j, \mathbf{u}_j} \sum_{p=j-N+1}^j \sum_{i=1}^I \rho \left(\frac{y_{ip} - x_{ij}}{\sigma_{y,i}} \right), \quad (1)$$

$$s.t. \quad \mathbf{f}(\mathbf{x}, \mathbf{u}) = \mathbf{0}$$

where $[\hat{\mathbf{x}}_j^R, \hat{\mathbf{u}}_j^R]$ are the estimates of the measured (\mathbf{x}) and unmeasured (\mathbf{u}) process variables at the j -th time interval, which are obtained using the measurements contained in a moving data window of length N . Furthermore, y_{ip} is the measured value of the i -th variable ($i=1 \dots I$) at the p -th time interval, $\sigma_{y,i}$ is its standard deviation, and \mathbf{f} represents the nonlinear set of process constraints.

B. Measurement Models

Random errors often follow a Gaussian distribution and are caused by unknown and unpredictable sources. When random errors are present, the measurement model at the j -th time interval can be represented as follows

$$y_{ij} = x_i + e_{ij}, \quad (2)$$

where x_i is the true value of the i -th variable and e_{ij} stands for the random error.

Systematic Errors can be classified as outliers (Out), biases (Bi) or drifts (Di). Let O and B be two scalars and $D(t)$ represent a function of the time. Outliers are isolated errors whose magnitudes are considered equal to $O\sigma_{y,i}$; their detrimental effect on variable estimates can be reduced by applying RDR. In contrast, Bi and Di are SEPT that affect robust estimates because their BDP are exceeded. While the bias magnitude is constant in time ($B\sigma_{y,i}$), the drift one varies in accordance with $D(t)\sigma_{y,i}$. These SE are modeled by the following equation:

$$y_{ij} = x_i + e_{ij} + SE_{ij}, \quad (3)$$

where SE_{ij} is replaced by $O\sigma_{y,i}$, $B\sigma_{y,i}$ or $D(t)\sigma_{y,i}$.

C. Process Leak Model

When a process leak occurs, the set of constraints $\mathbf{f}(\mathbf{x}, \mathbf{u}) = \mathbf{0}$ does not represent the process operation any more. In this work, only a portion of that set is considered to model the leak. In this sense, the mass balance equations are separated from the other ones, and the unmeasured flowrates are eliminated using a linear algebra based method (Sánchez and Romagnoli, 2000). A set of linear reconciliation equations is obtained, which is described by a matrix \mathbf{G} , and the presence of a leak is modeled as follows

$$\mathbf{G}\mathbf{c}_M = \mathbf{L} \times \boldsymbol{\delta}_M, \quad (4)$$

where $\boldsymbol{\delta}_M$ is a vector of zeros except for the row associated to the mass balance of the unit or set of units for which the leak occurs, L refers to its magnitude and \mathbf{c}_M is the vector of mass flowrates.

D. Algorithm steps

Llanos *et al.* (2017) developed an algorithm, which combines the SiM, the RMT and the RLR. This methodology is able to detect Out and suspicious observations, which are then classified as Bi or Di. Next, the steps that are part of the algorithm are briefly described. Furthermore, the

connections between these steps are explained. Finally, the SE and Le classification is presented.

E. Robust Data Reconciliation using SiM

This methodology comprises two sequential steps that take advantage of monotone and redescending M-estimators features and temporal redundancy.

Step 1: At the j -th time interval, the robust median of the i -th variable, $\hat{y}_{i,j}^R$ ($i=1, \dots, I$), is calculated using the data included in a moving window of length N $\{y_{ip}, p=j-N+1, \dots, j\}$:

$$\hat{y}_i = \min \sum_{p=j-N+1}^j \rho_{BW} \left(\frac{y_{ip} - \hat{y}_i}{\sigma_{y,i}} \right), \quad (5)$$

where ρ_{BW} stands for the Biweight M-estimator (BW), which is defined as:

$$\rho_{BW} = \begin{cases} 1 - \left[1 - \left(\frac{a}{c_{BW}} \right)^2 \right]^3 & \text{if } |a| \leq c_{BW} \\ 1 & \text{if } |a| > c_{BW} \end{cases}, \quad (6)$$

where c_{BW} is a scalar and a stands for the standardized error.

Step 2: The RDR problem is solved using as starting point the solution of Step 1:

$$[\hat{\mathbf{x}}_j^{SIM}, \hat{\mathbf{u}}_j^{SIM}] = \min_{\mathbf{x}_j} \sum_{i=1}^I \rho_{HU} \left(\frac{\hat{y}_{ij}^R - x_{ij}}{\sigma_{y,i}} \right), \quad (7)$$

$$s.t. \quad \mathbf{f}(\mathbf{x}, \mathbf{u}) = \mathbf{0}$$

where ρ_{HU} stands for the Huber function (HU), which is defined as:

$$\rho_{HU} = \begin{cases} a^2 & \text{if } |a| \leq c_{HU} \\ 2c_{HU}|a| - c_{HU}^2 & \text{if } |a| > c_{HU} \end{cases}, \quad (8)$$

where c_{HU} is a scalar.

F. Robust Measurement Test

To detect atypical measurements, a robust statistical hypothesis test based on the Measurement Test (Tamhane, 1982) is employed. The RMT relates the vector of robust measurement adjustments, $\mathbf{a}^R \in \mathcal{N}(0,1)$, with the robust estimate of the covariance matrix, $\hat{\mathbf{Q}}^R$:

$$\mathbf{a}_j^R = \mathbf{y}_j - \hat{\mathbf{x}}_j^{SIM}, \quad (9)$$

$$\hat{\mathbf{Q}}_j^R = \hat{\sigma}_a^2 \left\{ \frac{\text{ave}[\psi(\mathbf{A}_j^R)/\hat{\sigma}_a]^2}{(\text{ave}[\psi'(\mathbf{A}_j^R)/\hat{\sigma}_a])^2} \right\}^T, \quad (10)$$

where \mathbf{A}_j^R is a matrix which contains the last $N\mathbf{a}^R$ vectors and $\hat{\sigma}_a$ is a scale estimate vector. The relation between \mathbf{a}^R and the i -th diagonal element of $\hat{\mathbf{Q}}^R|_{ii}$ gives the following statistic:

$$\hat{\tau}_{i,j}^R = \frac{|a_{ij}^R|}{\sqrt{\hat{\mathbf{Q}}_j^R|_{ii}}} \in t_c, \quad (11)$$

that follows the Student distribution, t , with a number of degree of freedom, $df=N-1$. The level of significance of the test is set at $\alpha=0.025$, and it fixes the critical statistic value t_c .

G. Robust Linear Regression

Robust regression is used to estimate the parameters $\hat{\beta}_0^R$ and $\hat{\beta}_1^R$ corresponding to the fit of the bulk of the data, without suffering perturbations by a small proportion of gross measurement errors. Let consider a data set $\{(\mathbf{x}_k, y_k): k=1, \dots, K\}$, where \mathbf{x}_k and y_k are the predictor and re-

sponse variable values and K is the total quantity of measurements utilized for the regression. The following optimization problem is formulated to determine the vector $\hat{\beta} = [\hat{\beta}_0^R, \hat{\beta}_1^R]$:

$$\min \sum_{k=1}^K \rho_{BW} \left(\frac{r_k(\hat{\beta})}{\hat{\sigma}_r} \right), \quad (12)$$

where $r_k = y_k - (\hat{\beta}_0 + \hat{\beta}_1 x_k)$, and the scale estimation $\hat{\sigma}_r$ is calculated as the Normalized Median Absolute Deviation about the Median. The necessary and sufficient condition for solving the problem formulated in Eq. (12) is:

$$\sum_{k=1}^K \rho_{BW} \left(\frac{r_k}{\hat{\sigma}_r} \right) X_k = 0. \quad (13)$$

For the classification of a SEPT as a bias or drift, the following statistical hypotheses are formulated:

$$H_0: \hat{\beta}_1^R = 0$$

$$H_1: \hat{\beta}_1^R \neq 0.$$

To take a decision between both hypotheses, the statistic T_{β_1} is defined as the relation between $\hat{\beta}_1^R$ and its variance

$$T_{\beta_1} = \frac{\hat{\beta}_1^R}{[\text{var}(\hat{\beta}_1^R)]^{1/2}} \in T_c. \quad (14)$$

This statistic follows the Student Distribution with $df=K-2$. The critical value, T_c , is fixed by df and $\alpha=0.05$. If T_{β_1} is lower than T_c , H_0 is not rejected, the SEPT is classified as a bias, and $\hat{\beta}_0^R$ represents its magnitude. In contrast, the SEPT is categorized as a drift.

H. Algorithm description

The RMT is applied for each variable and its statistic is compared with t_c . If $\hat{r}_{i,j-1}^R$ is greater than t_c , an atypical observation is detected for the i -th variable. If its following statistic is lower than t_c , the measurement of the $(j-1)$ -th time interval is classified as an outlier. The Fig. 1 represents a temporal sequence of observations vectors, contained in the matrix $\mathbf{Y}_{ob}(I, N)$, and shows how the classification of an Out is made. The statistics that are lower than t_c are represented with a point, whereas the ones that exceed t_c are symbolized with \otimes .

When four consecutive $\hat{r}_{i,j}^R$ are greater than t_c , the i -th variable is considered a suspicious variable and is saved in an auxiliary vector \mathbf{u}_s (Fig. 2).

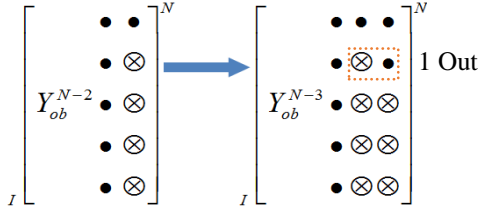


Figure 1. Outlier detection.

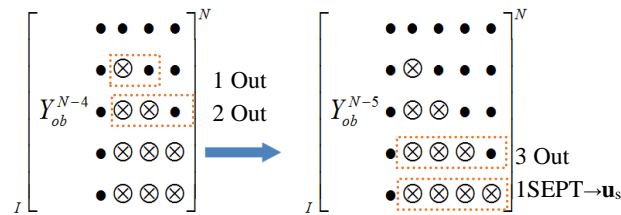


Figure 2. Consecutive Outliers and SEPT detection.

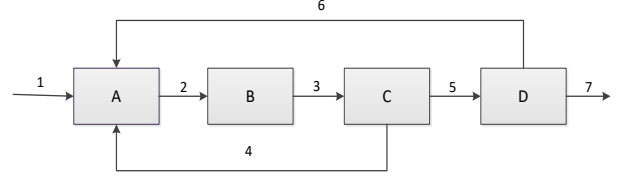


Figure 3. Rosenberg Flowsheet

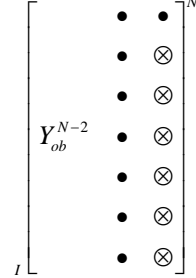


Figure 4. Leak Detection

The measurement variables contained in \mathbf{u}_s are not used for RDR, these ones are replaced by others generated using $\hat{\mathbf{x}}_{j-4}^{SIM}$. The arrival of new measurements is waited before the evaluation of the linear regression model. When $N/2$ observations are received, T_{β_1} is calculated, and the SEPT is classified. This information is sent to the DR stage. When the SEPT is categorized as a drift, faulty measurements are replaced as mentioned before. Otherwise, the observations contaminated with a bias are corrected using its estimated magnitude. This is calculated as the difference between the robust median of the measurements contained in \mathbf{Y}_{ob} and the last reconciled value.

The RMT is also applied to detect the presence of a leak. When a leak occurs, some flowrate measurements contained in the vector \mathbf{c}_M are affected, with the exception of the input flowrates or flowrates not related with the output ones of the faulty unit or set of units. The knowledge of how many observations are affected by a process leak allows to fix the minimum number of statistics that overcome t_c when the event occurs.

Let consider the Rosenberg Flowsheet (ROS), proposed by Rosenberg *et al.* (1985). The process flowsheet is represented in Fig. 3, and it is made up of 7 streams and 4 units. There is only one input stream, and all mass flowrates are measured. The system has a recycle that connects the first unit with the last one, this is why the presence of a leak in any unit affects all the flowrate measurements except the corresponding one to the input stream.

In this example, a fault is detected if six \hat{r}^R exceed t_c at the j -th time interval. The representation of the leak detection is shown in Fig. 4.

This analysis should be done for each specific process. Once a leak is detected all the affected measurements are corrected using previous estimated values.

I. Performance Analysis

Three case studies are proposed for each flowsheet. For Case 1, the methodology is run using simulated measurements with random errors. Case 2 shows the behaviour of

the strategy in the presence of SEPT and leaks. The same data set is used in Case 3 but the detection and classification procedure of SEPT and leaks is not applied.

Fifty thousand simulation trials are run for each case study. Fixed magnitudes of outliers ($O=10$), biases ($B=6$) are considered, and drifts evolve as $D(t) = 1 t$. Once a SEPT is simulated for a certain measured variable, it persists during 100 time intervals, and then this variable is not contaminated with a SEPT during the next 400 time intervals. The leaks simulated have a magnitude equal to 7 times the maximum standard deviation of the flowrates.

The probability of SE both on measurements and the process model is fixed at 0.024. Taking into account the length of SEPT, 10% of the measurements are contaminated with atypical values due to the presence of outliers, faulty sensors or model errors. The set of atypical observation is compound by:

- 93.5 % of Out,
- 5% of SEPT, that could be Bi and Di with equal probability,
- 1.5 % of leaks

The accuracy and detection performance measures are the Mean Square Error (MSE), the Percentages of Total Detection and False Alarms of SEPT (%DT_s -%FA_s), the Percentage of Total Detection of Out (%DT_O) and the Percentage of Total Detection of Leaks (%DT_{LE}).

$$MSE = \frac{1}{I J} \sum_{j=1}^I \sum_{i=1}^I \left(\frac{\hat{x}_{ij}^R - x_i}{\sigma_{y,i}} \right)^2, \quad (14)$$

where J stands for the simulation trials.

$$\%DT_s = \frac{SEPT \text{ Correctly Detected}}{Simulated SEPT} 100 \quad (15)$$

$$\%FA_s = \frac{False SEPT Alarms}{I J} 100 \quad (16)$$

$$\%DT_O = \frac{Outlier \text{ Correctly Detected}}{Simulated Out} 100 \quad (17)$$

$$\%DT_{LE} = \frac{LE \text{ Correctly Detected}}{Simulated LE} 100 \quad (18)$$

The parameters of Out detection and classification gives the same information. This is why just the detection of outlier is calculated.

The classification parameters of SEPT studied are: The Percentage of Correct and Wrong Classification of Bi (%BCC-%BWC) and Drifts (%DCC-%DWC), which are defined as follows:

$$\%XCC = \frac{X \text{ Correctly Classified}}{Simulated X} 100, \quad X \in [B, D] \quad (19)$$

$$\%XWC = \frac{X \text{ Detected and Wrongly Classified}}{Simulated X} 100, \quad X \in [B, D] \quad (20)$$

Different N values are considered in this analysis. The lower value of N is the one that allows detecting the 85% of the SEPT simulated. The upper limit corresponds to the window length for which the MSE starts to deteriorate.

III. RESULTS

Two linear benchmarks are considered to test the methodology. The first one is the ROS, and the second one

comprises the mass balance equations of the Tennessee Eastman Process (TEN) (Downs and Vogel, 1992). The trues flowrate values for both examples are:

$$x_{ROS} = [5 \quad 15 \quad 15 \quad 5 \quad 10 \quad 5 \quad 5]$$

$$x_{TEN} = [22.4 \quad 3664 \quad 4509.3 \quad 6419.4 \\ 8979.7 \quad 48015.4 \quad 48015.4 \quad 30840 \\ 386.5 \quad 16788.9 \quad 14288.6 \quad 48015.4]$$

The standard deviation of the measurements is set at 2.5% of their true values.

A. Rosenberg Flowsheet

The results of the simulation trials for the ROS are presented in Tables 1 to 3. Table 1 shows the MSE for the three case studies, whereas Tables 2 and 3 display the performance measures for Case 2.

Table 1 shows the MSE reduction achieved with the proposed methodology. When N increases the MSE of Case 1 and 2 diminish and the opposite happens with Case 3. For Case 2, the proposed algorithm of detection and classification applies corrective actions to measurements with SEPT and the RDR gives unbiased results. In contrast, the BDP of the estimates is overcome for Case 3.

It can be seen in Table 2 that the smallest windows achieves the best %DT_O and %DT_{LE}. In contrast the %DT_s is 100% for the longest window. When a SEPT evolves, the RMT is not applied because the presence of Out does not affect the RDR. This is why the %DT_O and %DT_{LE} decreases.

Regarding the classification measures of SEPT (Table 3), some observations are reported for $N \geq 30$ next:

- Biases are correctly classified in 92 % of the Bi simulations. The observations of sensors for which a Bi is developing can be corrected. Therefore the instrument still gives information about the variable state until it is repaired when this type of SEPT is present.
- In 100% of Di simulations, the fault is correctly identified, and a message is sent to the maintenance sector.
- The %BWC changes between 5 to 7 %, whereas DWC is zero. A BWC does not introduce inconvenient because the Bi is treated as a Di.

Table 1. MSE vs N (ROS)

N	Case 1	Case 2	Case 3
24	0.0191	1.0079	5.5930
30	0.0170	0.1714	5.7341
40	0.0113	0.1845	7.9245

Table 2. Detection Measures and FAs% (ROS)

N	% DT _O	% DT _{LE}	% DT _s	% FA _s
24	98.48	83.54	89.45	0.81
30	98.23	82.28	99.64	0.00
40	97.74	78.48	100.00	0.00

Table 3. Classification Measures (ROS)

N	BCC %	% BWC	DCC %	% DWC
24	83.33	6.06	88.12	1.39
30	92.42	6.82	100.00	0.00
40	94.7	5.3	100.00	0.00

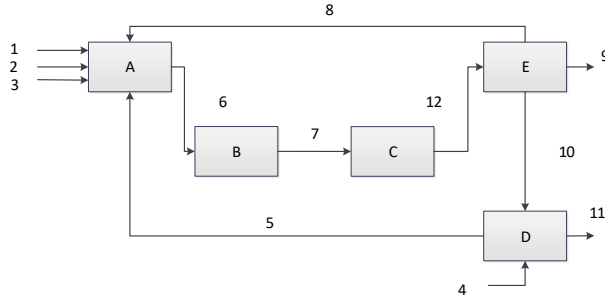


Figure 5. Tennessee Eastman Process

Table 4. MSE vs N (TEN)

N	Case 1	Case 2	Case 3
26	0.0407	5.3613	25.6963
30	0.0361	0.4504	23.8967
40	0.0282	0.3780	24.2219
50	0.0234	0.4677	27.6899

B. Tennessee Eastman Process Mass Balances

The Tennessee Eastman Process (TEN) is a well-known nonlinear benchmark. In this work, only mass balance equations are taking into account to test the performance of the proposed methodology. This process comprises 12 streams and 5 units (Fig. 5). The analysis of the system let fix that a Le is evolving when 8 statistics overcome the t_c .

The results of the simulation trials for the TEN are presented in Tables 4 to 6. Table 4 shows the MSE for the three case studies, whereas Tables 5 and 6 display the performance measures for Case 2.

Table 4 presents the lower and upper MSE (Case 1 and Case 3) and the ones achieved when the proposed methodology is applied (Case 2) for N varying in the range [26 : 50]. The comparison between Case 2 and 3 shows that the RDR is sensitive to the presence of SEPT and leaks.

The worst detection indexes are obtained for $N=26$ (Table 5). When $N \geq 30$ more than the 96% of the SE are detected. Furthermore, corrective actions are applied in 97% of the simulations in which a SEPT is detected. Because of that, their MSE are of one order of magnitude lower than the ones obtained for $N=26$.

When a SEPT is detected the RMT computation is stopped, consequently the $\%DT_O$ and $\%DT_{LE}$ tend to diminish with N . However, the tendency of $\%DT_{LE}$ is not uniform and will be the aim of future research.

Table 6 shows the capability of classification of SEPT. Some observations are reported for $N \geq 30$ next:

- The 92 % of the simulated biases are correctly classified, thus measurements can be corrected and continue providing information about the state of the variables with this SEPT.
- The 100% of Di are correctly categorized. This is an advantage because if a Di is wrongly classified as a Bi, sensor measurements are corrected as if the fault is a Bi.
- The $\%BWC$ changes between 3 to 7 %.

For the analyzed flowsheets, $N=30$ seems to be a good choice to detect faulty sensors and leaks taking into account the tradeoff between $\%DT_O$ - $\%DT_{LE}$ and MSE.

Table 5. Detection Measures and FAs% (TEN)

N	$\%DT_O$	$\%DT_{LE}$	$\%DT_S$	$\%FAs$
26	94.93	86.92	87.42	21.63
30	97.39	93.84	97.94	3.06
40	96.95	88.46	99.79	1.42
50	96.33	90.77	99.17	2.04

Table 6. Classification Measures (TEN)

N	BCC %	$\%BWC$	DCC %	$\%DWC$
26	83.72	5.16	84.54	1.29
30	92.86	3.17	100.00	0.00
40	92.85	6.75	100.00	0.00
50	92.86	5.95	99.57	0.00

IV. CONCLUSIONS

The performance analysis of the proposed methodology of detection and classification of systematic measurement errors integrated with a RDR procedure significantly improves the accuracy of variable estimates. The MSE diminishes when N grows because best percentages of detection of SEPT are achieved. In consequence, appropriate corrective actions are taken to reduce the negative effect of SE.

The analysis of classification indexes demonstrates that for $N > 26$ the observations of almost the 92% of the sensors with Bi can be corrected. Furthermore, the 100% of the Di are detected and well classified for $N > 30$.

This work addresses the detection of leaks by first in the RDR area. In contrast to previous works, model errors are considered and the RMT is used for their detection. A right selection of the windows length allows achieving percentages of leak detection higher than 82%.

ACKNOWLEDGMENTS

The authors thank the financial support of CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), and UNS (Universidad Nacional del Sur, Bahía Blanca, Argentina).

REFERENCES

- Chen, J., Y. Peng and J. Muñoz, "Correntropy Estimator for Data Reconciliation," *Chem. Eng. Sci.*, **104**, 10019-10027 (2013).
- Downs, J.J. and E.F. Vogel, "A plant-wide industrial process control problem," *Comp. Chem. Eng.*, **17**, 245-255 (1993).
- Llanos, C., M. Sánchez and R. Maronna, "Robust Estimators for Data Reconciliation," *Ind. Eng. Chem.*, **54**, 5096-5105 (2015).
- Llanos, C., M. Sánchez and R. Maronna, "Classification of Systematic Measurement Errors within the Framework of Robust Data Reconciliation," *Ind. Eng. Chem.*, **56**, 9617-9628 (2017).
- Maronna, R.A., R.D. Martin and V. Yohai, *Robust Statistics: Theory and Methods*, John Wiley and Sons Ltd., Chichester (2006)
- Martinez Prata, D., M. Schwaab, E.L. Lima and J.C. Pinto, "Simultaneous Robust Data Reconciliation and Gross Error Detection through Particle Swarm

- Optimization for an Industrial Polypropylene Reactor,” *Chem. Eng. Sci.*, **65**, 4943-4954 (2010).
- Narasimhan, S. and C. Jordache, *Data Reconciliation and Gross Error Detection*, Gulf Publishing Company: Houston, 2000.
- Narasimhan, S. and R. S. H. Mah, “Generalized Likelihood Ratio Method for Gross Error Identification,” *AIChE J.*, **33**, 1514-1521 (1987).
- Özyurt, D.B. and R.W. Pike, “Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes,” *Comput. Chem. Eng.*, **28**, 381-402 (2004).
- Romagnoli, J. and M. Sánchez, *Data Processing and Reconciliation for Chemical Process Operations*; Academic Press: San Diego (2000).
- Rosenberg, J., R.S.H. Mah and C. Iordache, “Evaluation of Schemes for Detecting and Identifying Gross Errors in Process Data,” *Ind. Eng. Chem. Res.*, **26**, 555-564 (1987).
- Sánchez, M., J.A. Romagnoli, Q. Jiang and M. Bagajewicz, “M. Simultaneous Estimation of Biases and Leaks In Process Plants,” *Comput. Chem. Eng.*, **23**, 841-857 (1999).
- Tamhane, A.C., “A Note on the Use of Residuals for Detecting an Outlier in Linear Regression”. *Biometrika*, **69**, 488-489 (1982).
- Zhang, Z. and J. Chen, “Correntropy Based Data Reconciliation and Gross Error Detection and Identification for Nonlinear Dynamic Processes,” *Comput. Chem. Eng.* **75**, 120-134 (2015).

Received March 16, 2018.

Sent to Guest Editor March 16, 2018.

Accepted April 19, 2018.

Recommended by Guest Editor Patricia M. Hoch