ROBUST DATA RECONCILIATION AND GROSS ERROR DETECTION IN AN INDUSTRIAL PREHEAT TRAIN

José Loyola-Fuentes[1], Tommaso Acerbi[1,2], Carlos Planelles[1], Emilio Diaz-Bejarano[1], Pierantonio Facco[2], Francesco Coletti,[1,3*]

[1]Hexxcell Ltd., Foundry Building, 77 Fulham Palace Rd, London W6 8AF, UK
[2]CAPE-Lab - Computer-Aided Process Engineering Laboratory, Department of Industrial Engineering, University of Padova, via Marzolo 9, IT-35131, Padova, Italy
[3]Department of Chemical Engineering, Brunel University London, Kingston Lane, Uxbridge UB8 3PH, UK

ABSTRACT

There are several challenges associated with the use of raw measurements for process monitoring, maintenance planning and operational and safety studies. One of the main causes is the typical low quality of industrial data which are often corrupted by numerous types of errors that can be broadly classified in random or systematic. Such errors hinder the quality of the data analysis that can be performed on the measurements, the quality of the models that can be developed, and thus the conclusions and related actions that can be taken at a plant level. It is therefore paramount to develop and implement systematic methods capable of dealing with these varied sources of measurement error. The need for these methods are more noticeable in complex, interconnected systems, such as Heat Exchanger Networks (HENs), whose performance are of vital importance and multiple measurements are taken to monitor such performance. Data Reconciliation (DR) and Gross Error Detection (GED) are techniques that complement each other. The former deals with the minimisation of random errors, whereas the latter deals with the detection and the mitigation of the effect of systematic errors. DR solutions can be improved using Robust Data Reconciliation (RDR), which exploits concepts from Robust Statistics to reduce the sensitivity of reconciled solutions with respect to the magnitude of gross errors (when present). This work presents the implementation of RDR and GED in a real heat exchanger network using plant data and a simplified HEN simulation model. Two different robust estimators and heuristic rules for GED are compared using the data with the aim of identifying potentially beneficial reconciliation algorithms, for the given context. The extension of this work aims at the integration of accurate HEN modelling with the use of RDR and GED as a pre-processing method.

1. INTRODUCTION

A reliable analysis of industrial thermal systems for monitoring, optimisation or maintenance, greatly depends on the number and quality of measured variables. To account for this dependency, it is paramount to ensure the validity and accuracy of any type of collected data. However, there is a limited accuracy one can expect, as process data are inevitably corrupted by measurement errors. These errors affect data by producing deviations from their true values, affecting further operations via error propagation [1]. The sources of measurement errors are generally categorised into two classes: random errors, which are caused by arbitrary fluctuations (environment, transmission, etc.); and gross errors, which are mainly caused by non-random events that mainly occur systematically (sensor bias or malfunctioning, measurement drifts, etc.). Thus, once a proper amount of measurements is set, suitable data-treatment methods are needed to mitigate the effect of measurement errors.

Data reconciliation is a well-known methodology that reduces the effect of random errors by adjusting the process measurements to satisfy specific process constraints, such as mass and energy conservation laws [1]. In this context, it is important to note that the relevance of accurate process models is significant, as these are used to calculate missing measurements and adjust existing ones. These estimated values (*i.e.* reconciled values) can then be used to estimate process KPIs and thus make

*Corresponding Author: f.coletti@hexxcell.com

important operating decisions. Nevertheless, the single use of DR does not always guarantee well-adjusted solutions, as process measurements could also contain gross errors. To account for their presence, a complementary technique known as Gross Error Detection [1] is capable of identifying the presence, and (in some cases) the value of systematic errors. A variety of methods are available for this task, where the use of statistical tests, integrated with serial or combinatorial approaches, are more abundant [2]. Classic data reconciliation (CDR) approaches integrate the use of GED to find reconciled measurements; however, when these detections fail, their effect propagates to measurements that do not contain gross errors. To avoid this, the use of estimators based on robust statistics, commonly known as robust data reconciliation (RDR) is a convenient alternative. These robust estimators are alternative functions that replace the weighted-least-square (WLS) estimator that is used in CDR [3]. These alternative functions present less sensitivity to gross errors than that of CDR. Nevertheless, when persistent gross errors are present (*i.e.* data drifting or time-persistent biases), the performance of RDR is expected to decrease (although not as much as in CDR). To overcome this issue, GED methods for RDR are also available [4]. These methods vary from robust statistical tests to heuristic rules based on the behaviour of the chosen robust estimator.

The benefits of RDR and GED are more noticeable in larger, more complex systems, such as HENs, where numerous streams and units are interconnected [7]. Relevant process variables such as flow rates, temperatures and/or pressures are continuously used by plant-personnel. For example, in monitoring activities, fouling resistance is estimated using plant measurements, where measurement errors can critically damage the quality of fouling resistance estimations. Moreover, process data are used in the training of prediction models, where the more contaminated the data are, the less accurate the prediction model will be. Hence, data treatment before these activities provides long-term benefits in many aspects of plant management.

To demonstrate the effects of the approaches above in real-case scenarios, this work presents an industrial case study, where the measurements from a crude preheat train are subjected to RDR and GED. Based on a comparative analysis of the results, two different robust estimators are selected as accurate alternatives. Two heuristic GED criteria, which are suitable for the selected robust estimators, are compared in terms of the number of gross errors detected, and their relative context within the data set. The use of these techniques show potential benefits for later calculations and decision-making processes.

2. CLASSIC AND ROBUST DR & GED

A reconciled value should generally comply with two conditions: i) the reconciled value should be as close as possible to the measured value (in the absence of gross error) and ii) the reconciled value should satisfy the process model [1]. Consequently, an optimisation problem can be formulated so that reconciled values satisfy these two conditions. Both CDR and RDR problems are based on optimisation approaches, where the main difference lies within the function to minimise. Additionally, in both reconciliation problems, the process model, along with specific restrictions such as practical ranges for measured values or empirical correlations, are used as constraints in the optimisation problem. More contrasts are found when comparing techniques for GED, such as the use of heuristics in RDR.

2.1 Classic Data Reconciliation

In this formulation, the objective is to minimise function $\rho(\varepsilon)$, which represents the squared of the differences between measurements and reconciled values, as a function of the standardised measurement error vector ε. These differences are weighted using a covariance matrix (σ_m), to penalise large differences. This method is commonly known as Weighted Least Squares (WLS). The general formulation for this CDR problem is given in Equation (1), where x_m is the vector of measured values and x_r represents the vector of reconciled values. Unmeasured variables (when present) are contained in the vector x_u. The minimisation problem in Equation (1) is subject to a set of equality constraints f, which can be of linear or nonlinear nature, and another set of inequality constraints g. Normally, the

handling of these two types of constraints depends on the type of solution, namely linear/nonlinear DR or steady-state/dynamic DR.

$$\min_{x_r} \rho(\varepsilon) = \varepsilon^2 = \left(\frac{x_m - x_r}{\sigma_m}\right)^2 \tag{1}$$

$$f(x_r, x_u) = c \tag{2}$$

$$g(x_r, x_u) \geq 0 \tag{3}$$

2.2 Robust Data Reconciliation

RDR replaces the WLS function of CDR in Equation (1) and uses instead functions that belong to what are called *M-estimators* [3]. These estimators have the distinction of being almost insensitive to significant increases in the standardised measurement error ε. This means that the effect of gross errors in other measurements is alleviated, so better estimates of reconciled values are achieved. Two different robust estimators, namely the Welsch and the Correntropy estimators are shown in Equations (4) and (5) respectively. Both these estimators have been used in industrial applications in the past, exhibiting promising results [5].

$$\rho_W(\varepsilon) = c_W^2 \left\{1 - exp\left[-\left(\frac{\varepsilon}{c_w}\right)^2\right]\right\} \tag{4}$$

$$\rho_{co}(\varepsilon) = -\frac{1}{c_{co}\sqrt{2\pi}} exp\left[-\left(\frac{\varepsilon^2}{2c_{co}^2}\right)\right] \tag{5}$$

Both estimators present a tuning parameter, in this case ρ_W and ρ_{CO} for Welsch and Correntropy estimators, respectively. Known values for these parameters are available, such that they can be compared based on similarities [5]. In this case, these parameters are tuned so that they perform at equivalent accuracy. This criterion deals with how robust each estimator is, and a convenient indicator for this robustness is what is called an *influence function* (IF), which describes the effect of changes in measurements on the estimator's value [3]. For the robust estimators in Equations (4) and (5), the influence function is proportional to the first derivative of $\rho(\varepsilon)$ with respect to ε. This fact facilitates the comparison for different robust estimators.

2.3 Gross Error Detection

In this work, only measurement biases are considered. In CDR, these gross errors are found and estimated via the deployment of statistical tests together with error models that attempt to capture the location and value of gross errors. On the other hand, in RDR, the information provided from the IF can be exploited to determine threshold values in the standardised measurement error, below which no gross errors are present in the data. These threshold values are called *cut points*. This method has the advantage of not assuming a prior probability distribution for the measurement error [3]. A measurement bias is then found when its corresponding adjustment after reconciliation ε is greater than the selected threshold value. In this work, the maximum of the influence function, along with the inflection point of such function are chosen as cut points, based on previous studies [3].

3. RESULTS AND DISCUSSIONS

The preheat train used in the work by Coletti and Macchietto [6] is further utilised in this case study. A simplified diagram is shown in Figure 1. Five years of operation, in daily averaged data of flow rates and temperatures around the network, were subject to both RDR and GED. Furthermore, a simplified steady-state version of an integrated fouling/heat exchanger network simulation model [7] is used to represent the process constraints. A performance indicator is created to assess the data reconciliation,

and gross error detection results are assessed by comparing the results from different cut points. The selection of the two robust estimators used in this work was carried out previously, where the simulation model was utilised to generate synthetic data and assess the performance of RDR and select the best estimators (Welsch and Correntropy in this case).

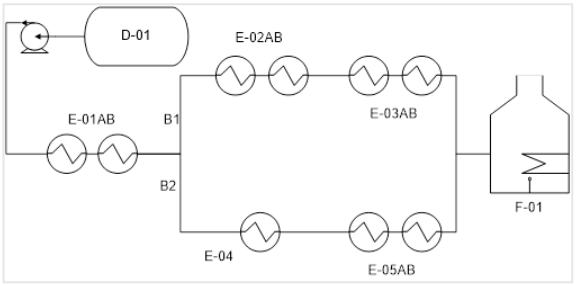


Figure 1: Simplified diagram of preheat train from Coletti and Macchietto [3].

The reconciliation performance is assessed by estimating the percentage of variance variation (σ^2_{var}), relative to the measured values (*e.g.* before RDR). The variance of each measured variable was estimated and stored before and after RDR (σ^2_m and σ^2_r respectively). The percentage of variance variation is defined in Equation (6). In this Equation, the absolute difference between variances is used, so negative values will indicate a variance decrease, whereas positive values would indicate an increase. The magnitude of this indicator is also relevant, as relatively large values (greater or less than zero) indicate either the presence of a gross error, or an issue with the RDR formulation.

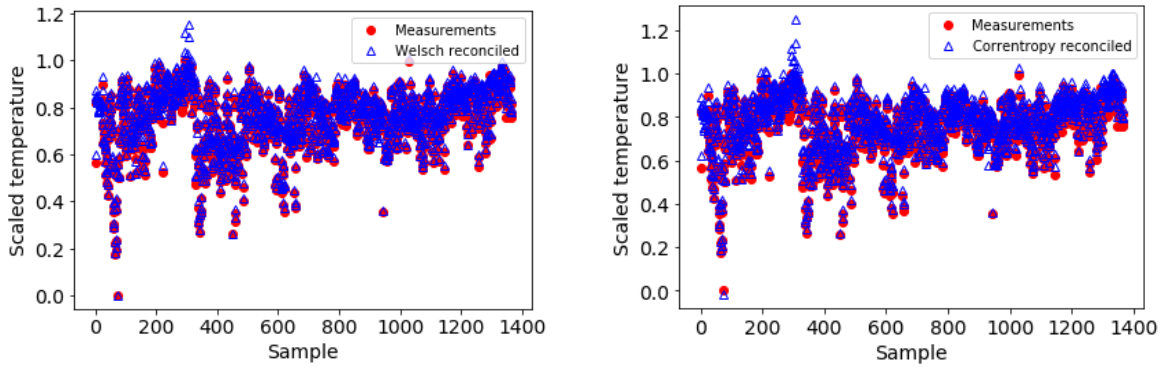
$$\sigma^2_{var} = 100\left(\frac{\sigma^2_r - \sigma^2_m}{\sigma^2_m}\right) \tag{6}$$

Table 1: Percentage of variance variation and gross error detection results for Welsch and Correntropy estimators.

	RDR		GED			
Variable	Welsch	Correntropy	Welsch IF max. point	Correntropy IF max point	Welsch IF inflection. point	Correntropy IF inflection point
FC001	1.71	1.712	-	-	-	-
FC002	4.05	4.05	0	1	0	0
FC003	-10.25	-10.25	-	-	-	-
FC004	0	0	-	-	-	-
FC005	0	0	-	-	-	-
FC006	0	0	-	-	-	-
FC007	0	0	-	-	-	-
TI001	-0.79	-1.31	-	-	-	-
TI002	-0.49	0.59	2	3	0	0
TI003	-19.33	-18.23	5	5	4	5
TI004	2.88	2.94	2	4	0	0
TI005	2.89	4.70	3	4	0	0
TI006	-5.97	-5.86	4	5	0	0
TI007	0.42	0.35	-	-	-	-
TI008	3.27	3.14	-	-	-	-
TI009	0.01	0.24	-	-	-	-
TI010	-7.87	-3.12	-	-	-	-
TI011	6.36	10.52	0	1	0	0
TI012	-40.57	-38.01	9	9	6	6
TI013	-0.49	1.27	7	6	0	0
TI014	3.07	5.87	0	1	0	0
TI015	0.64	0.87	3	3	0	0
TI016	-67.43	-67.50	14	17	2	2

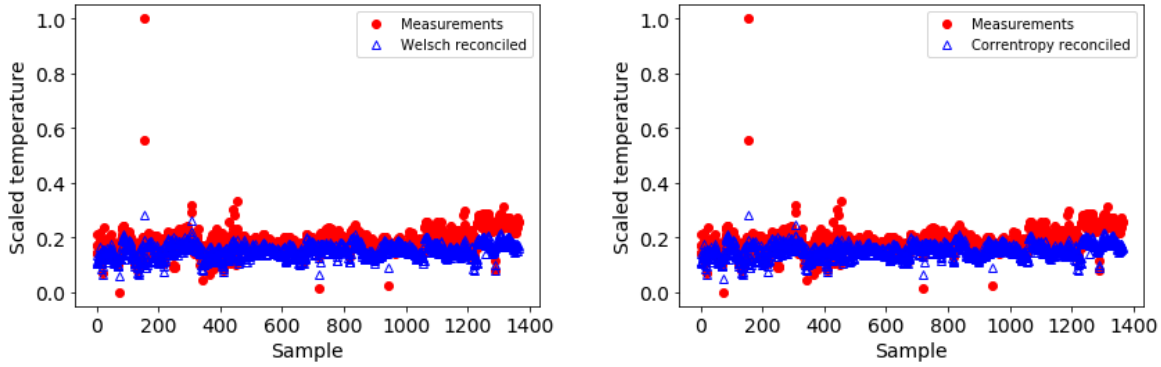
The estimation of these variation percentages are shown in the left-hand side of Table 1.Here, the Welsch estimator presents the lowest amount of measured variables with increasing variance after RDR, with 10 out of the 22 measured values. The Correntropy estimator presents 12 measured variables with increasing variance. Variance increases are usually not expected, and they reflect the fact that a simplified model, in steady state, has been used as a process model. This can also be seen in the variation percentage of sensor TI016, which measures the furnace (or coil) inlet temperature (CIT). This sensor presents the highest variance reduction after RDR. The source of this value is the lack of complexity of the simulation model when dealing with such measurement, which usually interacts with the complex dynamics of the furnace. This illustrates the importance of an accurate process model during RDR. In general, the largest variance increase was around 10%, provided by the Correntropy estimator in sensor T011. Note that the same sensor provides the maximum variation in the Welsch estimator as well, with a value of 6.36%. To the authors' knowledge, these values represent a typical uncertainty range when working with measured data. Another set of measurements presents null variance variation, which suggests that said measured values are non-redundant, and the system can only be reconciled when those values are not adjusted.

To further evaluate and compare the reconciliation performance, the trajectories of both measured and reconciled values are plotted in Figure 2 and Figure 3 for each robust estimator, and for each of the extreme cases discussed above (*i.e.* sensors T011 and T016), respectively. Note that for confidentiality reasons, the values of all measurements have been scaled down. Figure 2(a) and Figure 2(b) show the reconciliation of sensor T011 performed by the Welsch and Correntropy estimators respectively, where the former seems to implement more consistent adjustments to the reconciled values, with respect to the original measurements. In both cases, the trend of reconciled values follows continuously that of the measured values, indicating good reconciliation performance, with the exception of those values whose variance was increased during reconciliation. Similar results are shown for sensor T016. The reconciled values using both robust estimators are depicted in Figure 3(a) and Figure 3(b). Here, the difference between reconciled and measured values trend is more noticeable, compared to T011. Also, the reconciled values corresponding to those out-of-trend data points are not consistent with each other, not satisfying the first condition for being a reconciled value (*i.e.* reconciled and measured values are not close to each other). In both sensors, the Welsch estimator provides a better reconciliation performance compared to the Correntropy estimator, given its lower amount of sensors with increased variance and its capability of adjusting better to the original data-trend.



(a) (b)

Figure 2: Measured and reconciled values for T011 with Welsch (a) and Correntropy (b) estimators



(a) (b)

Figure 3: Measured and reconciled values for T016 with Welsch (a) and Correntropy (b) estimators

From the GED results in Table 1, for both estimators, the maximum point usage as a cut point detects a larger number of gross errors than that of the inflection point value, suggesting than the former is more conservative, offering a suitable alternative when no or little knowledge about the nature of the data is at hand. There are no major differences in the number of detected gross errors between the two estimators . At this point, a convenient method for assessing the differences in the number of gross error detected for each estimator is to account for the interaction among process variables, as measurement biases or unexpected peaks in the data are usually different from changes in operating conditions, as such changes are consistent among interconnected variables. An example is flow rate and temperature changes due to changes in working fluid (*i.e.* changes in density). Overall, the Welsch estimator seem to provide more promising results.

4. CONCLUSIONS

In this work, the benefits of robust data reconciliation and gross error detection in heat exchanger networks have been assessed. Two different robust estimators, along with two more threshold values for gross error detection have been tested and compared for choosing the best combination. The results show that the use of the Welsch estimator, along with the inflection point value of its influence function bring the best reconciliation performance and the most detection of systematic errors. Further extensions for this approach include the integration of more detailed heat transfer models and a complementary strategy for validating the gross error detection results.

REFERENCES

[1] S. Narasimhan and C. Jordache, *Data reconciliation and gross error detection: An intelligent use of process data* Gulf Publishing Company, (1999), Texas.

[2] J.A. Romagnoli and M.C. Sanchez, *Data processing and reconciliation for chemical process operations*. Academic Press, (2000), San Diego, CA.

[3] D. B. Özyurt and R. W. Pike, Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes. *Computers & Chemical Engineering*, **28(3)** (2004) 381–402.

[4] C.E. Llanos, M.C. Sanchez and R.A. Maronna, Classification of systematic measurement errors within the framework of robust data reconciliation. *Ind. Eng. Chem. Res.*, **56(34)** (2017) 9617–9628.

[5] C.E. Llanos, M.C. Sanchez and R.A. Maronna, Robust Estimators for Data Reconciliation. *Ind. Eng. Chem. Res.*, ***54*** (2015) 5096–5105.

[6] F. Coletti and S. Macchietto, Refinery Pre-Heat Train Network Simulation Undergoing Fouling: Assessment of Energy Efficiency and Carbon Emissions. *Heat Transfer Engineering*, **32(3–4)** (2011) 228–236.

[7] J. Loyola-Fuentes and R. Smith, Data reconciliation and gross error detection in crude oil pre-heat trains undergoing shell-side and tube-side fouling deposition. *Energy*, **183** (2019) 368–384.