

Combined Data Reconciliation and Parameter Estimation

Boon Sim Thian, Dept. of Electrical & Computer Engineering, National University of Singapore

Yen Yen Joe, NUS Graduate School, National University of Singapore

Arthur Tay, Dept. of Electrical & Computer Engineering, National University of Singapore

Xuan-Tien Doan, Institute of Chemical and Engineering Sciences (ICES), Singapore

Abstract—This paper presents a nonlinear approach for data reconciliation. The advantages of this approach over the conventional extended Kalman filtering (EKF) are namely, less linearization errors are generated in the process, and secondly the choice of the objective function is flexible. In this work, two probability density functions, namely the Logistic and Lorentz distribution are proposed as the objective function to be optimized in this approach. These two functions are proven to be statistically robust, which is an advantage over the conventional weighted least squares function. The extended Kalman filter and the modified nonlinear approach are implemented and verified via two case studies, namely a simulation case study and more importantly an experimental case study of a heat exchanger. The results obtained from the simulation case study demonstrated a reduction of approximately 50% error of the nonlinear approach over the extended Kalman filter.

Keywords — dynamic data reconciliation, robust estimation, nonlinear estimation

1. INTRODUCTION

Process measurements are fundamentally contaminated by errors during the measurement, processing and transmission of the measured signal [1-3]. The errors can arise from multifarious sources such as instrument degradation and malfunction, changes in ambient conditions, signal conversion noise and so on. The presence of such errors means that the measurements will not obey the laws of conservation. Errors in measured data often lead to considerable deterioration in plant performance; errors of small magnitude can lead to deterioration in the performance of control systems, whereas errors of large magnitude can offset the gains achievable through process optimization.

A simplified understanding of techniques used to process measurement data can be split into three basic steps, namely variable classification, gross error detection and data reconciliation (Figure 1). Variable classification deals with ascertaining which variables are observable or unobservable, redundant or under determined. Gross error detection identifies and treats non-random errors. Variable classification and gross error detection are not within the scope of this work. The third step for processing measurement data is that of data reconciliation. Through utilizing spatial or temporal redundancies, data reconciliation adjusts the process measurements to improve its accuracy. The reconciled estimates are expected to be more accurate than the measurements, and more significantly, are also consistent with the conservation laws and other constraints.

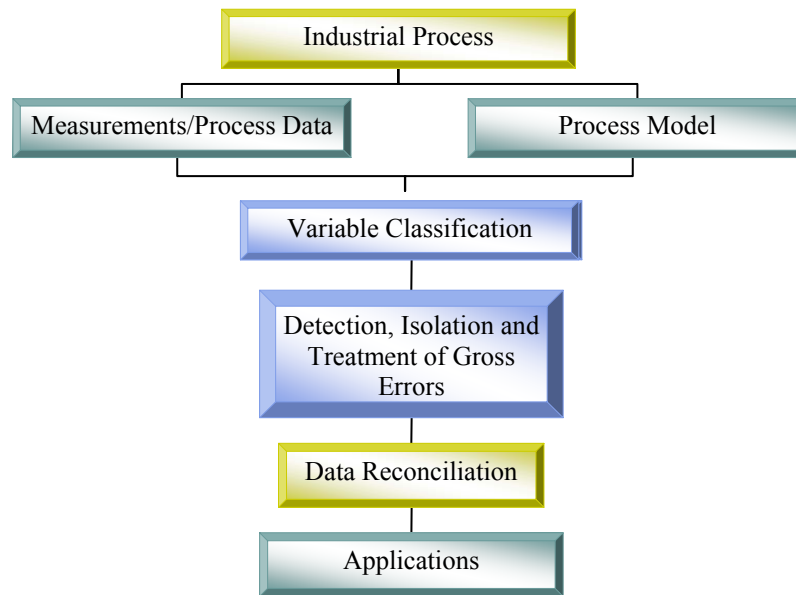


Figure 1. Essential steps for processing measurement data

The most prevalent form of dynamic data reconciliation is Kalman filtering. Kalman filter was originally designed for linear systems and it was proved mathematically to be the optimal estimates for linear systems in the presence of Gaussian noise [4]. Modifications to Kalman filter have been developed to handle nonlinear systems. These adjustments generally include linearizing the nonlinear system equations with first-order Taylor's series expansion. A common modification leads to the extended Kalman filter where the linearization takes place about an estimated trajectory.

Kalman Filter approaches are limited in relevance to data reconciliation (DR) problems for which weighted least squares estimators are appropriate. In cases where the measurement noise is non-Gaussian, the Kalman filter output may be biased. Secondly, the performance of the extended Kalman filter is directly related to the quality of approximations made in the state and covariance estimates. Most chemical engineering processes often function dynamically in highly nonlinear regions. Under such conditions, linearization may generate errors and the extended Kalman filter may produce either biased or divergent estimates.

Moreover, it may be difficult to tune the Kalman filter to achieve a satisfactory performance. Poor initial guess of the covariance matrices can also lead to incorrect estimation. In addition, the Kalman filter approaches do not support the inclusion of variable bounds and inequality constraints. The inclusion of inequality constraints is important because process models may be described in terms of inequalities.

In this paper, a nonlinear dynamic data reconciliation (NDDR) is utilized to overcome the problems of the Kalman filter. The approach undertaken in this paper is similar to that proposed by Liebman et al [3]. The use of other objective functions to replace the weighted least squares (WLS) objective function is also proposed. The proposed approach is extended to the combined data reconciliation and parameter estimation (PE), which results in both parameter and state estimates to be consistent with the process model equations.

The combined data reconciliation and parameter estimation (DRPE) is expected to generate more accurate estimates. This is confirmed by the findings of MacDonald and Howat [5], who examined both the sequential and combined DRPE and concluded that more reliable estimates are obtained using the combined method.

This paper is organized as follows. Section 2 presents the modified nonlinear approach undertaken in this paper. In Section 3, the proposed approach is applied to a simulation case study and the results presented and discussed. In Section 4, the proposed approach is applied to an experimental case study and the results presented and discussed. Finally, conclusion is presented in section 5.

2. NONLINEAR DYNAMIC DATA RECONCILIATION

Since the Kalman Filter approaches are limited in its capability to handle nonlinear processes, this paper undertakes a nonlinear dynamic data reconciliation approach [3]. This approach does not introduce linearization errors and therefore can handle processes with strong nonlinearities. Moreover, the nonlinear approach does not depend on any assumption of measurement error distribution. Inclusion of inequality constraints and variable bounds are also supported.

2.1 Dynamic Data Reconciliation Problem Formulation

The general NDDR problem can be formulated as:

$$\begin{aligned}
 & \min_{\hat{\mathbf{y}}(t)} \mathbf{f}[\mathbf{y}, \hat{\mathbf{y}}(t); \boldsymbol{\sigma}], \\
 & \text{subject to } \mathbf{g}\left[\hat{\mathbf{y}}(t), \frac{d\hat{\mathbf{y}}(t)}{dt}\right] = 0, \\
 & \quad \mathbf{h}[\hat{\mathbf{y}}(t)] = 0, \\
 & \quad \mathbf{r}[\hat{\mathbf{y}}(t)] \geq 0,
 \end{aligned} \tag{1}$$

where $\mathbf{f}[\mathbf{y}, \hat{\mathbf{y}}(t); \boldsymbol{\sigma}]$ =optimization objective function;
 \mathbf{y} =discrete measurements;
 $\boldsymbol{\sigma}$ =measurement noise standard deviations;
 \mathbf{g} =differential equation constraints;
 \mathbf{h} =algebraic equation constraints;
 \mathbf{r} =inequality constraints;

Although for most applications, the weighted least squares objective function is appropriate, there are some situations whereby using the weighted least squares objective function will lead to biased estimates. An important aspect of equation (1) is that under such conditions, other suitable objective functions may be used.

Another important feature equation (1) is the incorporation of inequality constraints and bounds \mathbf{r} . These constraints may become very important when handling unmeasured input estimation or measurement bias.

2.2 Objective Functions

In the nonlinear dynamic data reconciliation approach, the choice of the objective function is important. This paper proposes two statistically robust density functions, namely the Logistic and Lorentz distribution.

The Logistic distribution has the form:

$$f(u) = \frac{\exp(u/\sigma)}{\sigma[1 + \exp(u/\sigma)]^2} \quad (2)$$

The Lorentz distribution has the form:

$$f(u) = \frac{1}{\pi\sigma[1 + (u/\sigma)^2]} \quad (3)$$

where σ is the standard deviation of both functions.

2.3 Statistical Robustness of Estimators

This paper follows the approach taken by Hampel et al [6] in analyzing the statistical robustness of estimators, which is based on the Influence function (IF). The influence function assesses the amount of influence that a residual u_o has on the estimation. The exact derivation and analysis is found in Hampel et al [6]. For maximum likelihood estimators, the influence function can be conveniently taken as $\varphi(u) = \partial / \partial u[\ln(f(u))]$.

In order for the estimator to be robust, the influence function should be bounded such that any single large residual cannot distort the estimation. Additional criteria would be for the influence function to approach zero as the residual gets larger, and to be continuous such that the estimator is well-behaved. The influence function will now be used to analyze the robustness of the WLS, Logistic and Lorentz-based estimator. The graph of their influence function is illustrated in Figure 2.

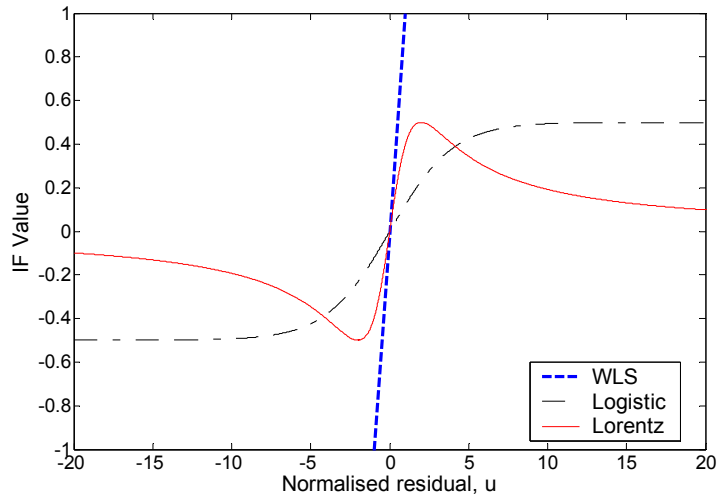


Figure 2. Influence functions of the various estimators

Figure 2 illustrates that the influence function of the WLS estimator is linear and unbounded for large values of residual. Correspondingly, this means that a single gross error can have a proportionally large influence on the estimation, hence leading to biased estimation. Essentially, this implies that the WLS estimator is not a robust estimator.

Conversely, the influence functions of the Logistic and Lorentz-based estimator are bounded even as the value of the residual increases. Hence, this implies that large errors will have bounded influence on the estimation. The influence of the two proposed estimators fulfils the criteria of statistically robust estimator, as described earlier.

In summary, the two proposed objective functions are sufficiently robust and more importantly can be used without a prior assumption about the actual measurement error distributions [7,8]. In terms of statistical robustness and characterization of the errors, the two proposed functions are far more superior to the weighted least squares.

2.4 Solution Strategy

The NDDR problem described in equation (1) involves the optimization of an objective function through the fine-tuning of estimate functions defined by differential equations and inequalities. In this paper, the strategy used for solving the NDDR problem is to optimize the objective function by adjusting the initial conditions $\hat{\mathbf{y}}(0)$ and use numerical integration to solve the differential equality constraints.

The optimization problem generally does not lead to an analytical solution. Many solvers have been proposed, such as the successive linearization [9], the QR factorization for bilinear systems [10], and the nonlinear sequential quadratic programming (SQP) [11,12]. In this paper, the SQP is utilized since it is not limited to linear or bilinear systems, and is not restricted to any form of objective function.

Improvements can be still be made, at the expense of increased computational burden; the use of finite orthogonal collocation [3,13,14] is one such method.

Given measurements, the optimal dynamic data reconciliation approach would utilize all available information from the onset of the process until the current time. However, such a strategy would lead to an optimization problem of ever expanding dimension. A moving window approach can be used to resolve such a problem. The approach is as follows: If the latest available measurements are at time t_1 , a window of size $W\Delta t$ can be defined from $t_1 - W\Delta t$ to t_1 . Only measurements within the window will be reconciled during each optimization (Figure. 3). A possible implementation can be summarized as [3]:

- (a) Collect process measurement data
- (b) Perform optimization over the time horizon, $[t_1 - W\Delta t, t_1)$
- (c) Use the reconciled estimate, $\hat{y}(t)$, for on-line control
- (d) Return to step (a) at the next time step

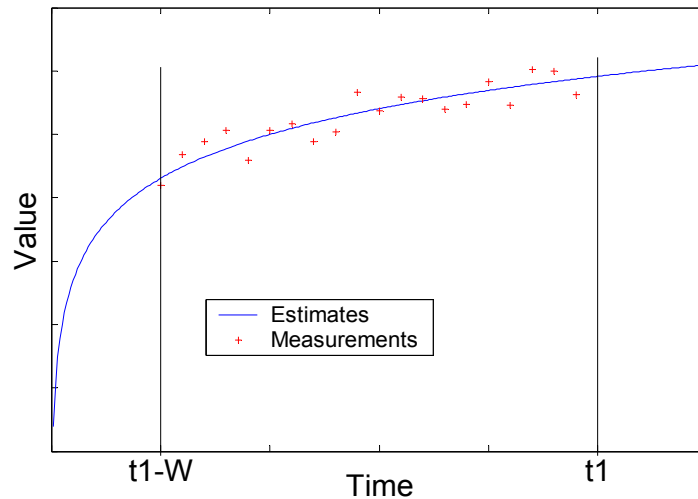


Figure 3. History horizon for an optimization run

3. CASE STUDY 1: SIMULATION OF TWO CSTRS

The nonlinear approach is first applied to a simulation case study of two continuously-stirred tank reactors in series [15]. The system consists of two constant volume reactors, in which an irreversible exothermic reaction $A \rightarrow B$ occurs (Figure. 4). The effluent stream from the first reactor serves as the feed stream for the second reactor. The reactors are cooled by a single coolant stream flowing concurrently with the reaction stream. The process model consists of 4 nonlinear differential equations:

$$\begin{aligned} \dot{C}_{A1} &= q/(C_{Af} - C_{A1})V_1 - k_0 C_{A1} \exp(-E/RT_1) \\ \dot{C}_{A2} &= q/(C_{A1} - C_{A2})V_2 - k_0 C_{A2} \exp(-E/RT_2) \\ \dot{T}_1 &= q/(T_f - T_1)V_1 + (-\Delta H)k_0 C_{A1} \exp(-E/RT_1)/\rho C_\rho \\ &\quad + (\rho_c C_{\rho c} q_c / \rho C_\rho V_1)[1 - \exp(-U_1 A_1 / q_c \rho_c C_{\rho c})](T_{cf} - T_1) \end{aligned}$$

$$\begin{aligned} \dot{T}_2 = & q/(T_1 - T_2)V_2 + (-\Delta H)k_0 C_{A2} \exp(-E/RT_2)/\rho C_p \\ & + \rho_c C_{pc} q_c [1 - \exp(-U_2 A_2 / q_c \rho_c C_{pc})] \\ & \cdot [T_1 - T_2 + \exp(-U_1 A_1 / q_c \rho_c C_{pc})](T_{cf} - T_1)/\rho C_p V_2 \end{aligned} \quad (4)$$

where T_1, T_2, C_{A1}, C_{A2} is the temperature of reactor 1, reactor 2 and the concentration of effluent in reactor 1 and reactor 2 respectively. The other symbols are constants and their values are listed in Table 1.

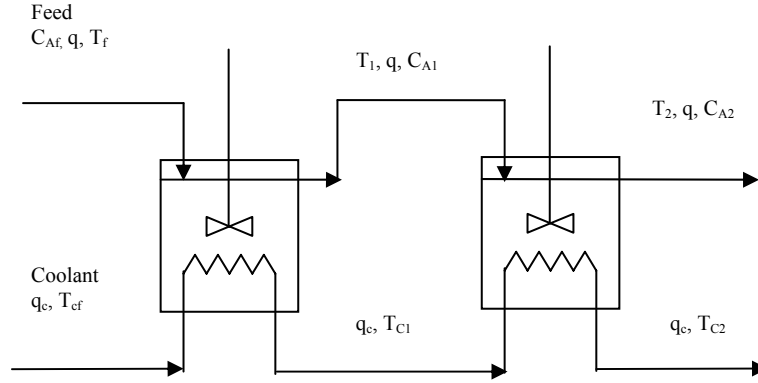


Figure 4. Continuously-stirred reactor system

Table 1. Nominal parameters used in simulation

| Symbol | Description | Nominal Values Used in Simulation |
|--------------------|---|-----------------------------------|
| Q, q_c | Flow rate of feed and coolant | 100 L/min |
| C_{Af} | Concentration of incoming feed | 1 mol/L |
| K_0 | Pre-exponential factor | 7.2×10^{10} /min |
| E/R | Normalized activation energy | 10000 K |
| T_f, T_{cf} | Temperature of incoming feed, coolant | 350 K |
| ΔH | Heat of reaction | -4.78×10^4 J/mol |
| V_1, V_2 | Volume of Reactor 1 and 2 | 100 L |
| C_p, C_{pc} | Specific heat capacity of feed, coolant | 0.239 J/g.K |
| ρ, ρ_c | Mass density of reactor contents, coolant | 1000 g/L |
| $U_1 A_1, U_2 A_2$ | Heat transfer coefficient in Reactor 1, 2 | 1.67×10^5 J/min.K |

3.1 Performance Measures

Two measures will be used to quantify the efficiency of the dynamic data reconciliation: the relative efficiency and the percentage error versus time. The relative efficiency, ε_i , is a normalized parameter and it can be computed as:

$$\varepsilon_i = \frac{\min(MSE)}{MSE_i} \quad (5)$$

where $\min(MSE)$ is the mean-square error of the best estimation method.

The percentage error, p , is computed as:

$$p = \frac{|True\ value - estimate|}{True\ value} \times 100\% \quad (6)$$

The plot of the percentage error over time is an indication of the reliability of the estimation method.

3.2 Results and Analysis

Two results are presented: firstly, the comparison of EKF with the nonlinear approach. Secondly, for the nonlinear approach, the comparison of objective functions is presented.

- *Comparison of EKF with nonlinear approach*

In this comparison study, only Gaussian noise is generated. Figure 5 illustrates the estimate of T_1 , and Figure 6 illustrates the percentage error obtained using the extended Kalman filter and the nonlinear approach. Figure 5 shows that although the estimate obtained using the extended Kalman filter is initially close to the true value of the process variables, the estimates diverge rapidly from the true values (and measurements) and eventually attain a steady-state bias of 50% error (Figure 6). Tuning the extended Kalman filter differently failed to provide reliable estimates. In contrast, the same figure shows that the estimates obtained using the nonlinear approach is very close to the true values. The errors in the estimates are only a mere 2% and more importantly, the estimates displayed the same dynamic trends. This result clearly illustrates the robustness and reliability of using the nonlinear approach over the extended Kalman filter in the presence of process nonlinearities.

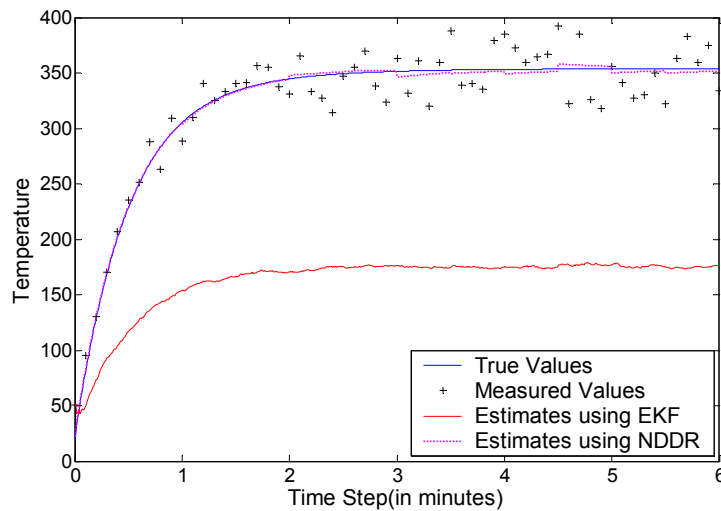


Figure 5. Estimate of T_1 using EKF and NDDR

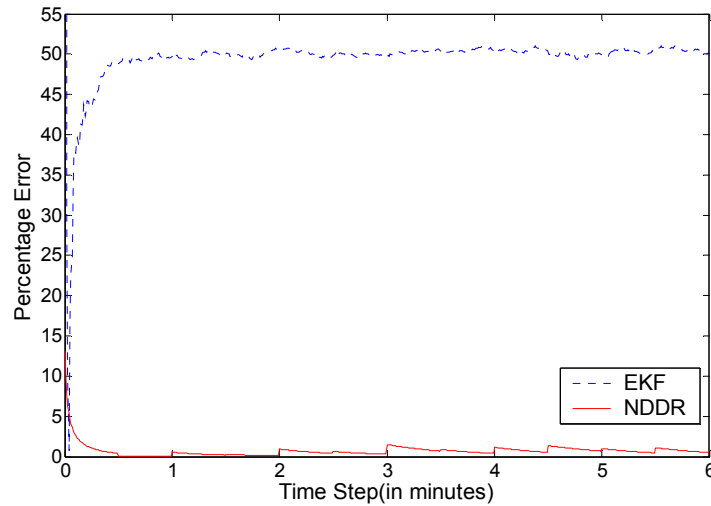


Figure 6. Percentage error using EKF and NDDR

▪ *Nonlinear DR: Comparison of objective functions*

In the second comparison study, non-Gaussian distributed errors are generated. It is important for the proposed nonlinear dynamic data reconciliation to be robust; robustness in this aspect refers to capability to characterize the measurement errors. Figure 7 illustrates the relative efficiency obtained using different objective functions. The maximum relative efficiency is 1.00, which indicates it as the best estimator.

The result shows that the WLS estimator has the lowest relative efficiency and is therefore the least reliable and robust in handling non-Gaussian errors; this is not surprising as it is shown earlier (section 2.3) that the WLS is statistically not robust. In contrast, the 2 proposed functions demonstrated improved efficiency as they have the highest relative efficiency for all the other noise distributions.

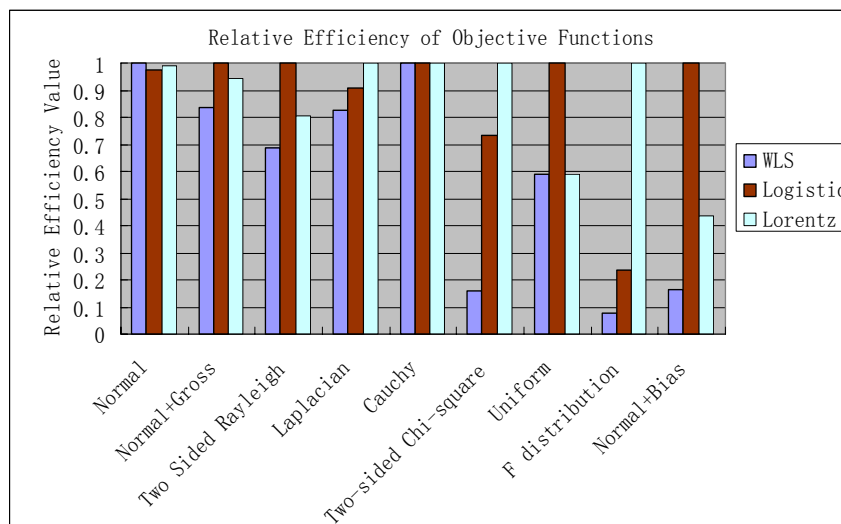


Figure 7. Relative Efficiency of objective functions

4. CASE STUDY 2: HEAT EXCHANGER IMPLEMENTATION

The extended Kalman filter and the nonlinear dynamic data reconciliation strategies are tested on a process plant, PCT23 Trainer from Armfield (Figure 8). The subsystem, as indicated by a solid line, is used for this experimental study.

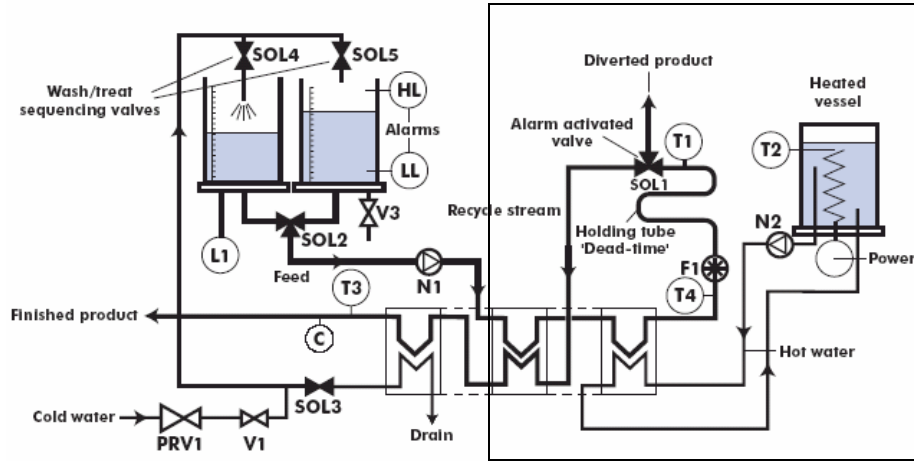


Figure 8. Schematic of the Heat-Exchanger Pilot Plant [16]

A simplified but realistic model of a fluid-fluid heat-exchanger can be expressed as [17]:

$$\begin{aligned} \frac{dT_5}{dt} &= \frac{2}{M} \left[F_2(T_2 - T_5) - \frac{UA(\Delta T_{lm})}{C_{p,hot}} \right] \\ \frac{dT_4}{dt} &= \frac{2}{M} \left[F_1(T_3 - T_4) + \frac{UA(\Delta T_{lm})}{C_{p,cold}} \right] \end{aligned} \quad (7)$$

where $\Delta T_{lm} = \frac{(T_2 - T_4) - (T_5 - T_3)}{\ln\left(\frac{T_2 - T_4}{T_5 - T_3}\right)}$ is the log mean temperature difference, T_h , T_c is the

temperature of the output hot stream and cold stream respectively and the other symbols are constants, except for UA and M , which are unknown parameters to be estimated. The details are listed in Table 2.

Table 2. Nominal values of parameters used in experiment

| Symbol | Description | Nominal values used |
|------------------------|---|---------------------|
| $C_{p,hot} C_{p,cold}$ | Specific heat capacity of hot/cold stream | 4.18 kJ/kg.°C |
| $T_{h,in}$ | Temperature of incoming hot stream | 71.2 °C |
| $T_{c,in}$ | Temperature of incoming cold stream | 22.0 °C |
| F_1 | Volume flowrate of the cold stream | 2.82e-3 kg/s |
| F_2 | Volume flowrate of the hot stream | 4.39e-3 kg/s |
| UA | Overall effective heat transfer coefficient | Unknown |
| M | Mass of hot/cold water held | Unknown |

4.1 Model Verification

The first requirement of any data reconciliation procedure is an accurate process model. To verify the accuracy of the model, the output obtained using the model is compared to the experimental data and is presented in Figure 9. The model output matches the trend found in the data, except for a constant difference and this is expected because a real physical process may differ from a theoretical model; it is therefore appropriate to use the model.

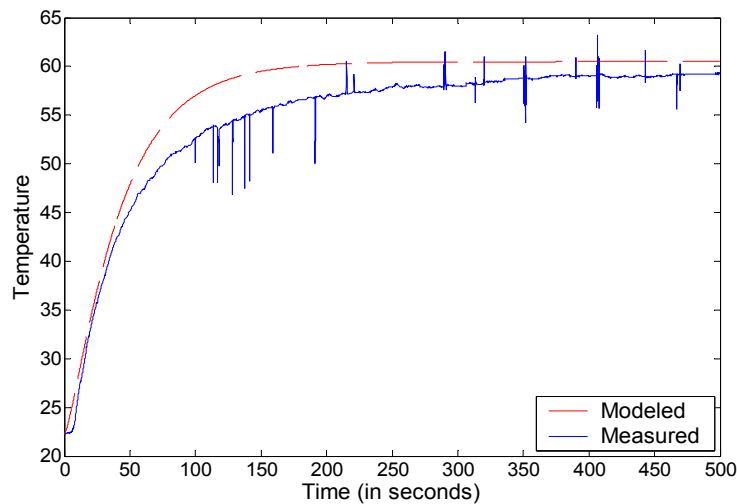


Figure 9. Output results from model and measurement data

4.2 Performance Measures

There is no basis to objectively and quantitatively assess the accuracy of the estimates of the variables and parameters because unlike the simulation case, the true variable and parameter values are unknown. Although some faults are introduced into the measurements, the graph of the variation of the state variables with time demonstrates the general trend of the variables quite obviously. Hence, it is possible to compare estimates with the measurements and qualitatively assess the performance of the different data reconciliation techniques.

4.3 Results and Analysis

The parameter estimates obtained is shown on Table 3 and the estimates of the states are illustrated in Figure 10. It is difficult to draw any conclusion based on the parameter estimates but Figure 10 suggests some analysis of the performance of the estimators.

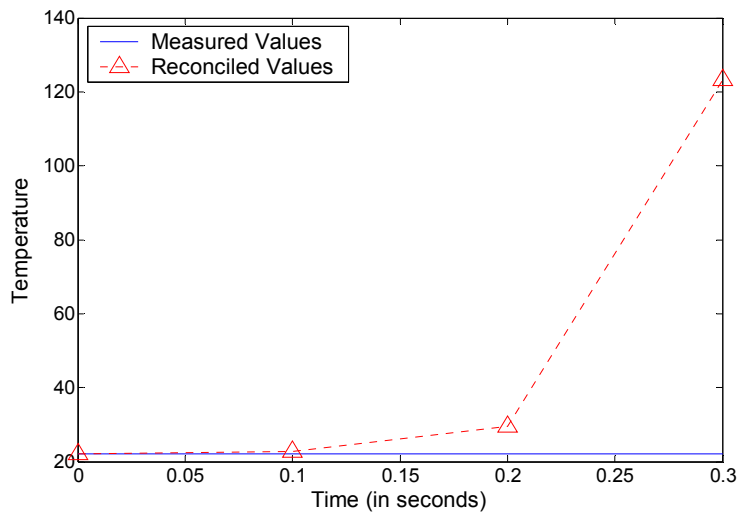
The EKF estimate diverges rapidly and after only 3 samples, the estimates are at least three times larger than the measurements (Figure 10a). In fact, the estimate diverges so rapidly that it reaches the magnitude of a few thousand degrees. This can be explained by the fact that linearization generates errors and hence causes the divergence.

On the other hand, the nonlinear approach generally shows much smoother estimation than the EKF and clearly reflects effective filtering. However, the WLS estimates (Figure 10b) are ‘noiser’ than the other 2 proposed estimators. The estimates obtained using the Logistic and Lorentz estimators are much smoother than the measurements (Figure 10c and 10d).

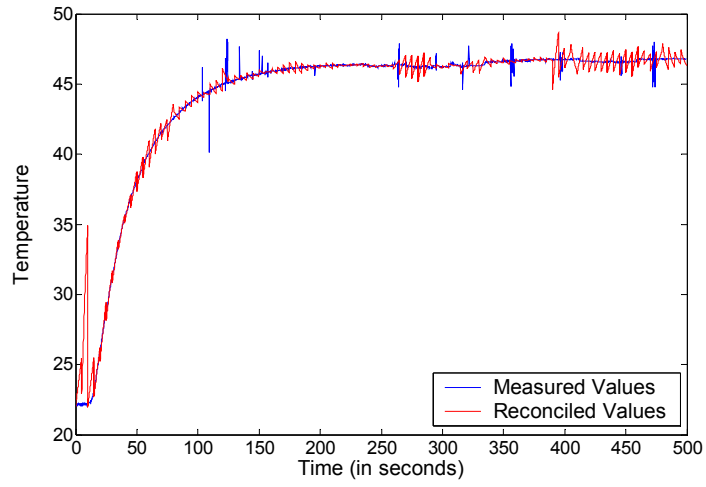
In summary, the Logistic and Lorentz distributions are better choices than the weighted least squares for the objective function when NDDR is used. This is because, as explained earlier, these two functions are at least statistically robust and more reliable than the weighted least squares function.

Table 3. Parameter Estimates

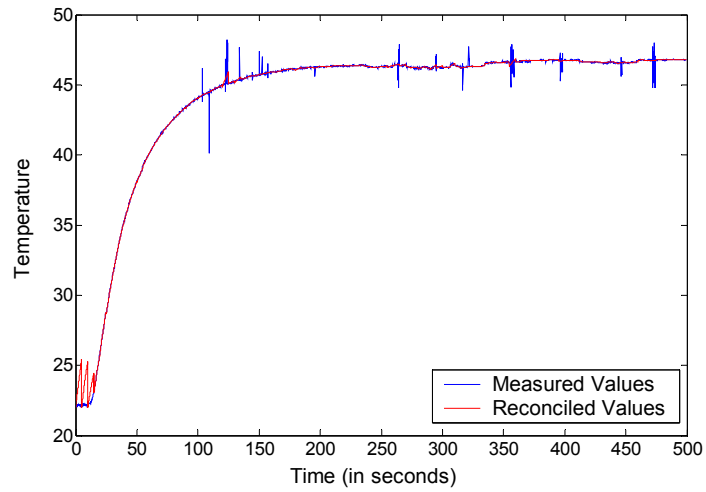
| Methods | Parameters | |
|-----------------------------------|--------------------|--------------------|
| | M | UA |
| Extended Kalman Filter | Failed to estimate | Failed to estimate |
| Nonlinear approach using WLS | 0.35150 | 0.021247 |
| Nonlinear approach using Lorentz | 0.40811 | 0.024113 |
| Nonlinear approach using Logistic | 0.39439 | 0.024093 |



(a)



(b)



(c)

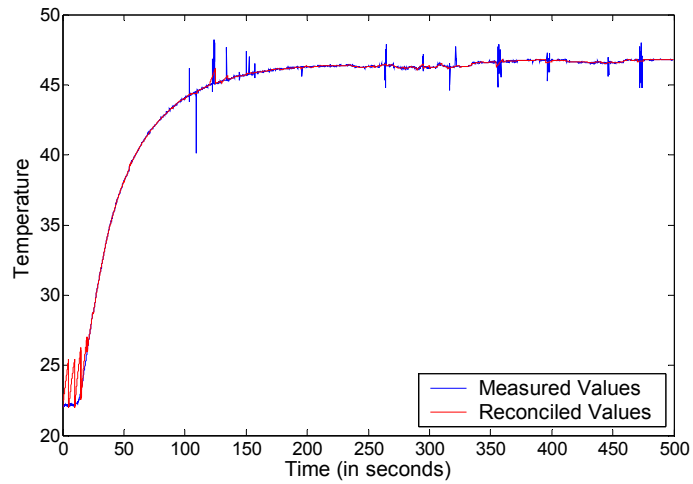


Figure 10. Estimate of T_5 using (a) EKF (b) NDDR WLS (c) NDDR Logistic (d) NDDR Lorentz

5. CONCLUSION

Dynamic data reconciliation is of more significance than the steady state data reconciliation problem because most processes are operating dynamically. In this paper, a nonlinear dynamic data reconciliation algorithm for online estimation is presented. Since most practical processes are nonlinear in nature, the nonlinear method is superior to the extended Kalman filter because it does not generate linearization errors.

In addition to the above-mentioned advantage over the extended Kalmanfilter, the nonlinear technique also allows objective functions other than the weighted least squares function, and supports the inclusion of variable bounds and inequality constraints. The importance of the objective function and the capability to include variable bounds is demonstrated via two case studies.

Two density functions are proposed as viable choices for the objective function in the nonlinear algorithm, namely the Logistic and Lorentz probability density functions. The estimators based on these two functions are statistically robust and hence may be used without any prior assumption of error distribution.

The results obtained from the simulation case study demonstrated a marked reduction of 50% error of the modified approach over the EKF. Also, the results obtained via the experimental case study demonstrated that the proposed Logistic/Lorentz-based objective functions are able to obtain smoother estimates than the weighted least squares objective functions.

6. REFERENCES

- [1] Romagnoli, J. and Sanchez, M., "Data Processing and Reconciliation for Chemical Process Operations", Academic Press (2000)
- [2] Narasimhan, S. and Jordache, C., "Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data", Gulf Publishing Company (2000)
- [3] Liebman, M.J, Edgar, T.F., Lasdon, L.S., "Efficient Data Reconciliation and Estimation for Dynamic Processes using Nonlinear Programming Techniques", *Comp. Chem. Eng.* **16**, 963-986 (1992)
- [4] Gelb, A., "Applied Optimal Estimation", MIT Press, Cambridge, Massachusetts (1974)
- [5] MacDonald, R.J. and Howat, C.S., "Data Reconciliation and Parameter Estimation in plant performance analysis", *AIChE.* **34**, 1-8 (1988)
- [6] Hampel, F.R., Ronchetti, E.M., Rousseuw, P.J. and Stahel, W.A., "Robust Statistics: The Approach Based on Influence Function", Wiley, New York (1986)
- [7] Ozyurt, D.B., Pike, R.W., "Theory and Practice of simultaneous data reconciliation and gross error detection for chemical processes", *Comp. Chem. Eng.* **28**, 381-402 (2004)
- [8] Huber, P.J., "Robust Statistics", Wiley, New York (1981)
- [9] Swartz, C.L.E., "Data Reconciliation for Generalized Flowsheet Applications", *197th National Meeting*, American Chemical Society, Dallas, TX (1989)
- [10] Crowe, C.M., "Reconciliation of process flow rates by matrix Projection. Part II: The nonlinear case", *AIChE. J.* **32**, 616-623 (1986)
- [11] Tjoa, I. and Biegler, L., "Simultaneous Strategies for Data Reconciliation and Gross Error Detection of Nonlinear Systems", *Comp. & Chem. Eng.* **15**, 679-690 (1991)

- [12] Gill, P., Murray, W.A., Saunders, M.A., and Wright, M.H., "User Guide for NPSOL (Version 4.0): A FORTRAN program for Nonlinear Programming, Technical Report" SOL 86-2, Stanford University, Department of Operation Research, Stanford, CA.
- [13] Finlayson, B.A., "The Method of Weighted Residuals and Variational Principles", Academic Press, New York (1972)
- [14] Renfro, J.G., Morshedi and Asbjornsen, O.A., "Simultaneous optimization and solution of systems described by differential/algebraic equations.", *Comp. Chem. Eng.* **11**, 503-517 (1987)
- [15] Henson, A. M. and Seborg, D.E., "Input-Output Linearization of General Nonlinear Processes", *AIChE J.* **36**, 1753-1757 (1990)
- [16] Armfield Limited UK. Product Data Sheet, Process Trainer PCT23 MkII, 2005.
- [17] Alsop, A.W. and Edgar, T.F., "Heat Exchanger Control", *Chem. Eng. Comms.* **75**, 155-170