

ANDDR WITH NOVEL GROSS ERROR DETECTION AND SMART TRACKING SYSTEM

Mazyar B. Laylabadi, James H. Taylor

*Department of Electrical and Computer Engineering,
University of New Brunswick,
Fredericton, N.B., Canada
mazyar.L@unb.ca - jtaylor@unb.ca*

Data reconciliation is a well-known method in on-line process control engineering aimed at estimating the true values of corrupted measurements under constraints. An adaptive nonlinear dynamic data reconciliation (ANDDR) method is proposed that includes the application to processes with an unknown statistical model. ANDDR enables gross error detection (GED) as well. Finally, a novel smart tracking system is combined to ameliorate the problem of delay seen in both the original and later NDDR methods. This package with its smart tracking features is suggested for use in distributed control systems (DCSs) for process control and manufacturing applications such as paper making. *Copyright © 2006 IFAC*

Keywords: Adaptive Algorithms, Dynamic Systems, Error Detection, Nonlinear Systems, Estimation Algorithms.

1. INTRODUCTION

Data reconciliation (DR) is a well-known method in process control aimed at estimating the true values of corrupted measurements, taking into account constraints on dynamic behaviour, material and energy balance, etc. Usually it is presumed that the corrupted measurements are free of gross errors or, more specifically, that they only contain zero-mean random noise. Gross errors, however, exist normally in manufacturing processes due to malfunctioning sensors (e.g., data drop-out) or can be process-related (such as process leaks). It is crucial to detect and identify gross errors first or in some cases simultaneously with DR. There have been some new GED and statistical model identification approaches developed recently and combined with original NDDR method in order to remove the negative effects of gross errors. Among these methods there are very few approaches which address the situation where a statistical model is not available. However, they cannot handle either nonlinearity or dynamic behaviour of the processes which make them unsuitable for many manufacturing processes with common nonlinear and dynamic characteristics.

In this study an adaptive nonlinear dynamic data reconciliation (ANDDR) method is proposed that includes the application to processes with unknown statistical models. Then, a novel smart tracking

system is combined with this adaptive method, which ameliorates the problem of delay seen in existing NDDR methods. These new adaptive features permit solving the GED problem as well. Therefore, the combined ANDDR and GED algorithms together with the smart tracking system produce a package that is suitable for most manufacturing process applications. The proposed package has been successfully applied to the continuously stirred tank reactor (CSTR) model cited commonly in the literature. This package with its smart tracking features is suggested for use in distributed control systems (DCSs) for process control and manufacturing applications such as paper making, in order to improve process monitoring and lessen operator load work.

A brief overview of DR methods and algorithms is outlined in section 2. In section 3 the original NDDR problem formulation and basic theory of ANDDR, which is proposed as an enhancement, is presented. Then, this ANDDR is combined with a novel GED and identification algorithm. Next in section 4, the smart tracking system is proposed which removes the delay seen in the results of original and existing NDDR methods. In section 5 the results of the implementation of the whole package on a simulated CSTR model are depicted. Finally, in section 6 the conclusions and future work are discussed.

2. DATA RECONCILIATION BACKGROUND

The DR problem was first introduced by Kuehn and Davidson (1961) for linear steady state models. There has been a great deal of research conducted in the area of steady state and linear processes, while NDDR has received less attention (Crowe, 1996). On the other hand, as far as engineering processes are concerned they often operate dynamically in highly nonlinear regions where traditional methods such as the Kalman filter or extended Kalman filter (EKF) may be ineffective (Liebman and Edgar, 1988).

The necessity of developing NDDR methods was proposed by Liebman and Edgar (1988), and the advantages of using nonlinear programming (NLP) over traditional steady state DR methods were demonstrated. In the next step Liebman et al. (1992) developed their main NDDR algorithm. Their approach was based on simultaneous optimization and solution techniques where efficient state estimation was performed. There was no GED and identification included in the NDDR approach of Liebman et al. (1992), but, since then, to complete their work, there have been some new approaches which are capable of detection of gross errors, or, more specifically, identification of gross errors as well. For instance, Soderstrom et al. (2001) proposed an approach to simultaneously tackle the problem of GED and identification together with DR. Abu-el-zet et al. (2002) proposed a combined method of bias and outlier identification in dynamic DR where a history of all previous bias detection and identification methods is briefly presented.

There have been studies to address the estimation of the measurement error covariance matrix but either applied only to linear processes or limited to stationary processes (Alici, 2001). Although several authors have stated the need for covariance estimation for DR, none of them has proved the effectiveness of using this matrix in DR except for Alici (2001) who demonstrated its necessity and briefly discussed the effects that covariance matrix estimation has on DR. She also addressed the combination of dynamic model identification (DMI) with NDDR.

3. PROBLEM FORMULATION AND SOLUTION STRATEGY

3.1 The original NDDR problem formulation and solution strategy

The original NDDR formulation was developed in Liebman et al. (1992), and it may be re-stated as follows:

$$\min_{\hat{y}} \phi(\tilde{y}, \hat{y}(t); \sigma), \quad (1)$$

subject to:

$$f\left(\frac{d}{dt}\hat{y}, \hat{y}\right) = 0, \quad (2)$$

$$h(\hat{y}) = 0, \quad (3)$$

$$g(\hat{y}) \geq 0, \quad (4)$$

Here the corrupted measurements \tilde{y} and reconciled estimates \hat{y} include both state variables and input variables. The first constraint, Eq. (2), represents the process dynamics, the second, Eq. (3), may describe energy and/or material balance, and the third, Eq. (4), may impose process variable limits. For more details one can refer to the paper by Liebman et al. (1992).

For most applications the objective function is weighted squared error (WSE):

$$\phi(\tilde{y}, \hat{y}(t); \sigma) = \sum_{j=0}^c \frac{1}{2} (\hat{y}(t_j) - \tilde{y}_j)^T V^{-1} (\hat{y}(t_j) - \tilde{y}_j), \quad (5)$$

\tilde{y}_j = measured (corrupted) values at discrete time t_j

\hat{y}_j = estimated values at discrete time t_j

V = variance-covariance matrix where each diagonal element V_{ii} is σ_i^2

t_0 = initial time, t_c = current time

The solution adopted by Liebman et al. (1992) is a moving horizon approach which enables the user to utilize all the information at hand (process measurements) from start-up to the current time. This moving horizon window (MHW) approach has the advantage of reduced optimization problem size while giving the user the advantage of having only one tuning parameter, the window horizon (H), compared to other nonlinear approaches such as the EKF where more tuning parameters need to be adjusted. Another advantage is its capability of handling constraints such as equalities and inequalities, whereas other approaches such as the EKF cannot. Figure 1 shows the basic idea of the MHW approach for NDDR. In this method after collecting the process measurements up to t_c , ϕ is optimized over $t_c - H$ to t_c , the current time. Then one point $\hat{y}(t_c)$ is saved and the procedure is repeated at the next time step (Liebman et al., 1992).

In order to solve the optimization problem in Eqs. (1) to (4) the nonlinear model presented as the first constraint in Eq. (2) needs to be discretized. In this study the fourth order Runge-Kutta method has been chosen to simulate and discretize the model.

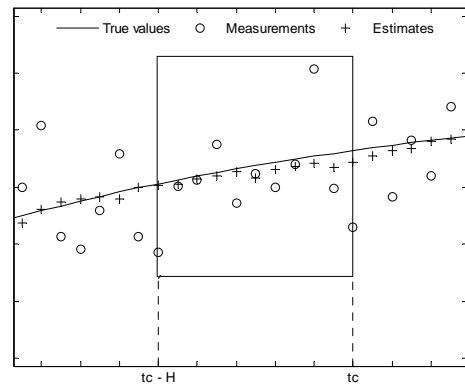


Fig. 1. History horizon for NDDR

Once the discretization is implemented, Eqs. (1) to (4) can be rewritten as the following NLP problem:

$$\min_{\hat{y}} \sum_{i=1}^{n_i+n_s} \eta_i \sum_{j=c-H}^c \left(\frac{\hat{y}_{ij} - \tilde{y}_{ij}}{\sigma_i} \right)^2 \quad (5)$$

subject to:

$$f\left(\frac{d}{dt}\hat{y}, \hat{y}\right) = 0, \quad (6)$$

$$h(\hat{y}) = 0, \quad (7)$$

$$g(\hat{y}) \geq 0, \quad (8)$$

where $f(\hat{y})$, $h(\hat{y})$ and $g(\hat{y})$ now represent the constraints obtained through discretization, η is a vector of weights and n_i and n_s are the numbers of inputs and states (outputs), respectively.

3.2 ANDDR and GED methodology

As mentioned earlier, most NDDR techniques today are based on two major assumptions: 1) having known dynamic and statistical models, and 2) having gross-error-free measurements. The novel ANDDR + GED approach is suitable for cases where one does not have a statistical model for noise, or, in other words, standard deviation σ or covariance matrix V is not known, and where *isolated* outliers may occur. In this methodology the same moving window approach proposed in Liebman et al. (1992) is used and σ is estimated as each measurement variable is processed. The method can briefly be described as follows: the moving window provides us with H measurements at each time step. Assume that $H \geq 10$ in this discussion; if this is not true, then a longer window may be used for estimating σ . Using these measured variables the sample variance is used to estimate σ for each variable. It is known that (Taylor, 1981) the σ of the sample variance for a Gaussian process is:

$$\sigma_{\hat{\sigma}} = \sqrt{2\frac{V}{H}} \quad (9)$$

So for $H \geq 10$ the estimate is adequate for a threshold test that is usually conservatively chosen, e.g., $T = 3\hat{\sigma}$. This is the basis of the σ estimation method which justifies using $\hat{\sigma} = \sqrt{\hat{V}}$ in solving the ANDDR problem.

To perform GED the previous mean, Eq. (10), is used to derive the difference d_c for each time step, and this is compared with the threshold based on the previous $\hat{\sigma}$, $\hat{\sigma}_{c-1}$, as indicated (Eq. 12). If d_c exceeds this threshold, then the algorithm suspects the existence of an outlier and waits to receive the next measurement. Based on the new d , d_{c+1} , it decides if the previous measurement was an outlier or a step change. This comparison is continuously conducted at each time step.

$$\tilde{m}_{c-1} = \sum_{j=c-H}^{c-1} \left(\frac{\tilde{y}_{ij}}{H} \right) \quad (10)$$

$$d_c = \|\tilde{y}_c - \tilde{m}_{c-1}\| \quad (11)$$

$$\text{If } d_c(i) > 3\hat{\sigma}_{c-1}(i), \text{ (suspect outlier in } \tilde{y}_c(i)) \quad (12)$$

Attention must be focused on points in time near set-point changes. If the algorithm is not smart, then such a change can simply be taken as a gross error. In our development, a sample point where an error threshold is exceeded is designated as a *possible* outlier and the next point is processed to decide if a set-point change occurred or the previous point contained an outlier (in which case it is edited out, e.g., by interpolation). This is effective under the assumption that outliers are isolated (do not happen in successive samples); if this cannot be assumed then the algorithm would have to be modified to wait several samples before the outlier/set-point-change decision can be made. This logic causes a one or several time-step delay in a real-time setting; thus there is a trade-off between fast or robust detection. The former needs a GED algorithm based on an assumption of isolated gross errors, and the latter, which enables us to handle a number of successive gross errors, requires more detailed logic.

Figure 2 is a flowchart that describes step by step the procedure of implementation of proposed ANDDR and GED. This flowchart is inspired by Alici et al. (2002).

4. ANDDR AND ITS SMART TRACKING SYSTEM

Studying the results of the original NDDR implementation on the simulated CSTR model presented by Liebman et al. (1992) clearly shows a significant delay for the input estimation when a step occurs, which also causes some delay in the estimation of output variables. This delay, especially for the input estimation, is the main drawback of the original NDDR and later enhanced approaches. This delay is caused by the assumption that the input is constant over the entire moving window. Therefore, a complete window length is required to reach steady state at the new set-point. Our new algorithm includes set-point change detection as part of the GED logic, and it tracks the set-point change by introducing a second level. When the set-point change is sensed, the second level tracks the new set-point. In this way the whole delay is removed, producing estimates \hat{y} that are significantly more accurate. In the next section the results for the input estimation show satisfactory results of this smart tracking system implementation.

5. CASE STUDY

In this paper a simulated CSTR model studied commonly in the literature is chosen. In order to produce the comparable results with those of Liebman et al. (1992), the same assumptions and parameters values for the model are used here. The normalized model can be presented as follows:

$$\frac{dA}{dt} = \frac{q}{V}(A_0 - A) - \alpha_d kA \quad (13)$$

$$\frac{dT}{dt} = \frac{q}{V}(T_0 - T) + \alpha_d \frac{\Delta H_r A_r}{\rho C_p T_r} kA - \frac{UA_R}{\rho C_p V}(T - T_c) \quad (14)$$

$$k = k_0 \exp\left(\frac{-E_A}{TT_r}\right) \quad (15)$$

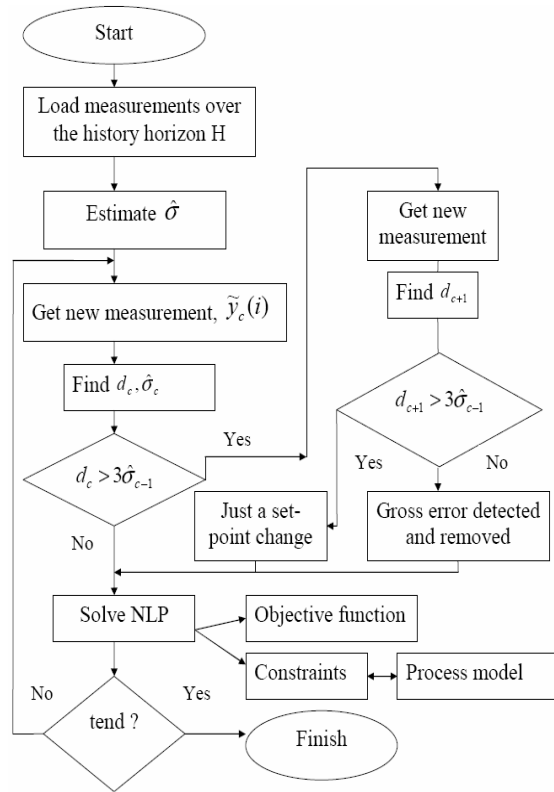


Fig. 2 ANDDR and GED algorithm flowchart

where the input stream feed concentration A_0 and feed temperature T_0 are input variables and concentration A and temperature T are output variables. There are two simple constraints on both input and output variables as follows:

$$0 \leq A, A_0 \leq 20.0 \quad (16)$$

$$0 \leq T, T_0 \leq 10.0 \quad (17)$$

The values of other constants in this model are presented in Liebman et al. (1992). In this example both the two inputs and two states (outputs) are being estimated. The measurement noise is assumed to be Gaussian with σ equal to 5% of the actual value and zero mean. The time step is assumed to be 2 sec and the simulation is run for 100 samples with window width of $H=10$. Obviously, the first estimation is achieved at time step 10 where the first window of measurements is available.

5.1 Comparison of original NDDR results with ANDDR + GED in the presence of outliers

Here it is assumed that isolated outliers exist for each variable. The number of gross errors over the entire horizon is 5 for each variable.

In Figs. 3 to 6 the estimation results considering a step change at time step 30 for the first input, A_0 , from 6.5 to 7.5, and another step change from 7.5 to 4.5 at time step 70, are presented. In each figure the complete time-history is shown, plus a zoomed section for better viewing. The solid lines in these figures show the true values, circles show the corrupted measurements, stars present the proposed ANDDR + GED estimation results and plus signs mark the original NDDR data. As the figures show, the gross errors have been detected and successfully

removed, and the estimation has not been corrupted. Observe that the outliers cause significant corruption of the NDDR data.

5.2 Comparison of original NDDR with proposed smart tracking system

As discussed in section 4, using the original NDDR method (and later extended versions of it) it takes a full moving window length to get to the exact new set-point after an input step change occurs. One can see this delay in Fig. 5 in the estimation of feed concentration A_0 when the step inputs are applied at time points 30 and 70, which also causes a slight delay in both states' estimation. Using the smart tracking system, however, the estimates do not show any delay (Fig. 5). To present this feature of the smart tracking system, another situation is considered where no gross errors exist on any measurement variable. Here again the same two steps are applied on feed concentration. Fig. 7 shows the estimation of feed concentration which represents no delay. Also the slight delay seen in both NDDR output estimations are eliminated (Figs. 8, 9).

Note that in all the estimation results of ANDDR + GED the assumption was that the statistical model, σ , is not available, and only for the original NDDR implementation the σ is assumed to be known. Therefore, Figs. 3 to 9 prove that the ANDDR + GED package has successfully estimated the σ and there is no accuracy degradation seen in the main NLP problem solution. Also the noise reduction ratio of ANDDR was compared with that of the original NDDR, and no difference was seen.

Another improvement observed in this study is the ability to detect the peak for the output estimations. For instance, Figs. 8, 9 show the peak (at time step 70) has been detected clearly but the results given by Liebman et al. (1992) demonstrate the output estimates do not follow the peak when the set-point is changed.

6. CONCLUSION AND FUTURE WORK

The proposed package has been successfully implemented and applied to the CSTR model of Liebman et al. (1992). These results demonstrate the performance improvements possible in applications where set-point step changes occur and where the variance/covariance matrix is not known. The next step will be an evaluation on a more realistic model of a pilot plant facility at the College of North Atlantic, which is involved in the Petroleum Application of Wireless Systems (PAWS) project, followed by studies with actual data acquired from the plant. This package with its smart tracking features can be successfully used in a DCS in order to improve process monitoring and lessen operator load work. Adding a DMI feature to the proposed approach may be a useful extension for this study. As stated in section 2, DMI has been addressed in Alici et al. (2002), but the selection of a DMI approach requires some attention, since it can be application dependent. Therefore, the combination of a suitable

DMI method and our new package for ANDDR and GED is suggested as future work.

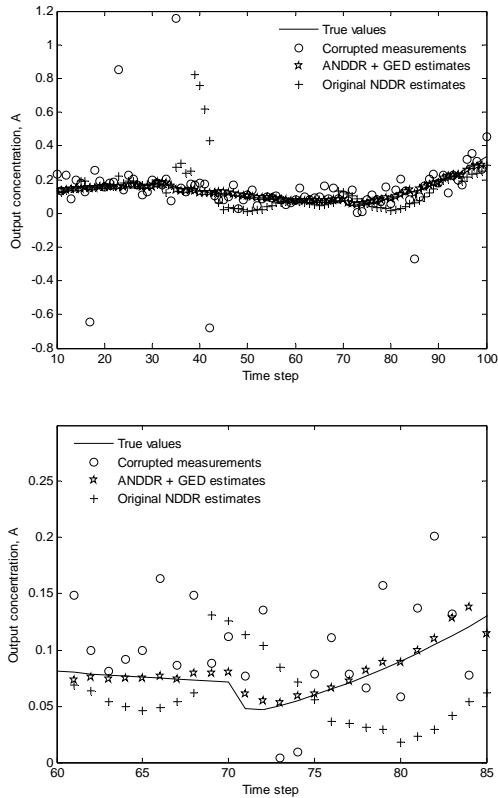


Fig. 3 Output concentration, A , estimation

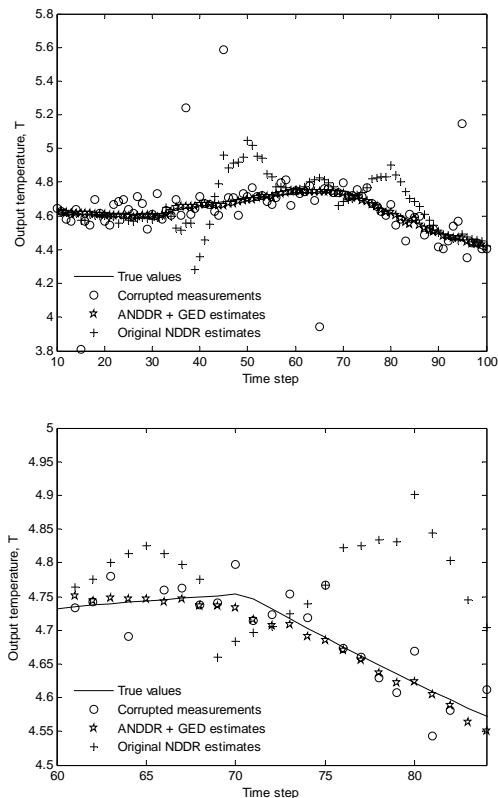


Fig. 4 Output temperature, T , estimation

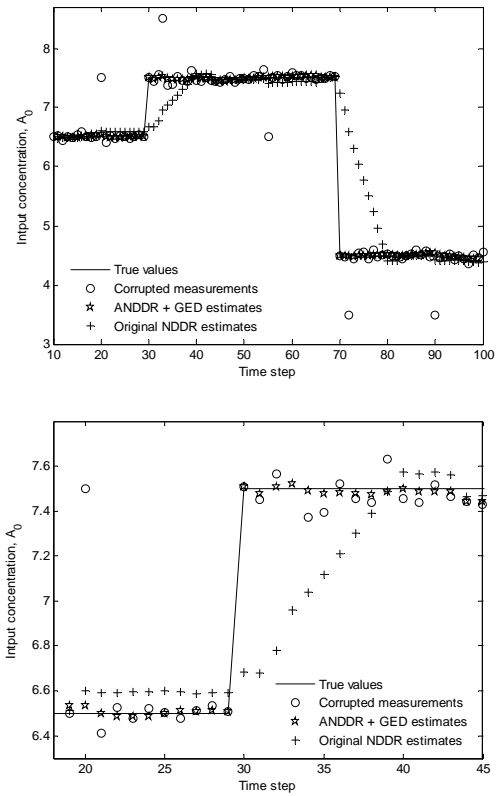


Fig. 5 Feed concentration, A_0 , estimation

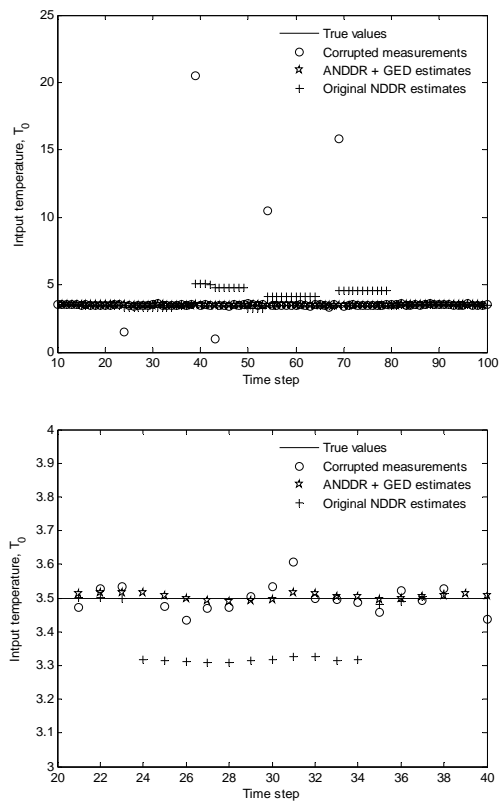


Fig. 6 Feed temperature, T_0 , estimation

7. ACKNOWLEDGEMENT

This project is supported by Atlantic Canada Opportunities Agency (ACOA) under the Atlantic Innovation Fund (AIF) program. The authors gratefully acknowledge that support and the collaboration of the Cape Breton University, the

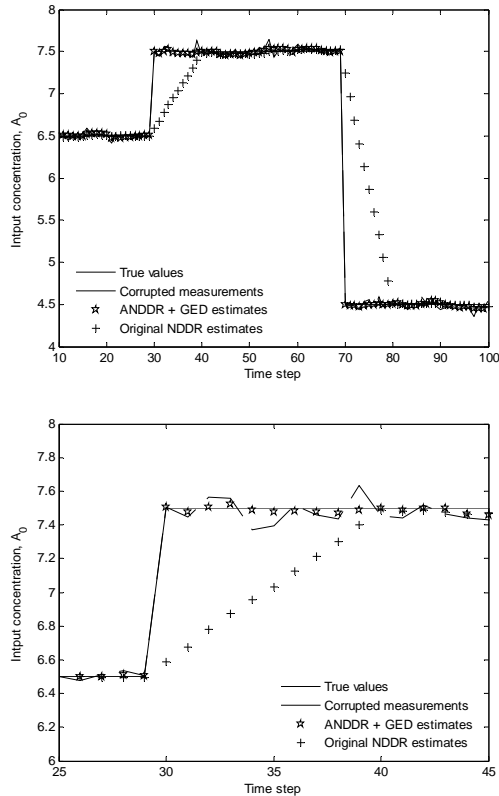


Fig. 7 Feed concentration, A_0 , estimation

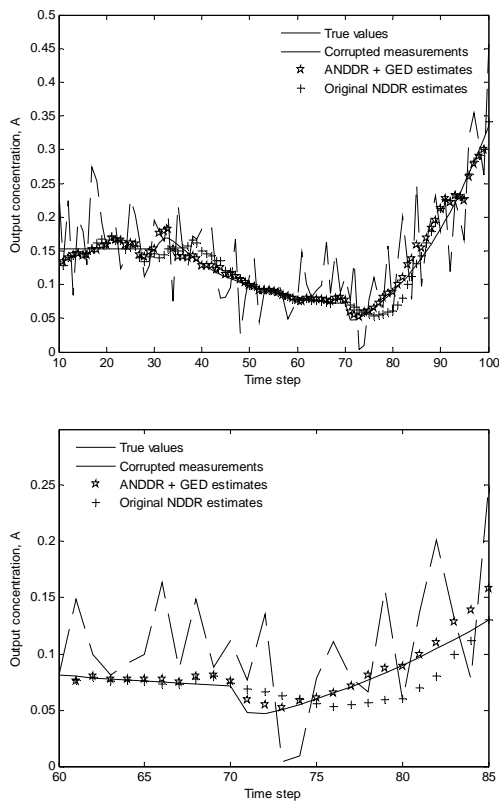


Fig. 8 Output concentration, A , estimation

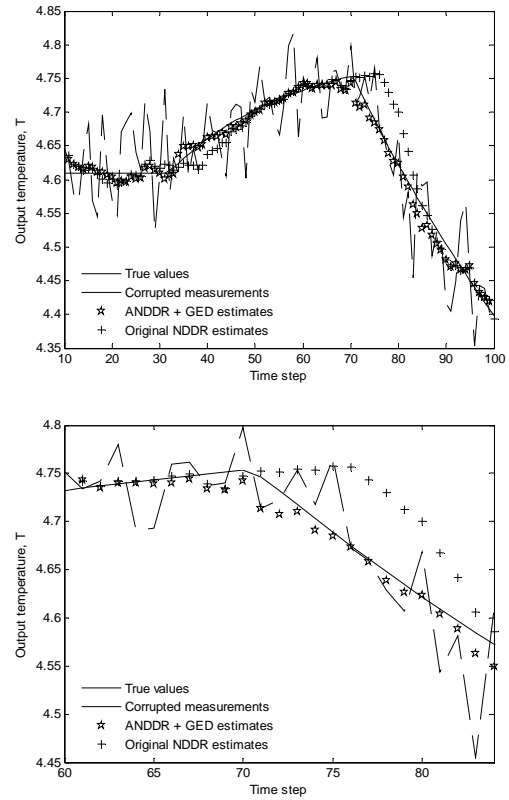


Fig. 9 Output concentration, T , estimation

REFERENCES

- Abu-el-zet Z.H., Becerra V.M. and Roberts P.D. (2002). Combined bias and outlier identification in dynamic data reconciliation. *Comp. Chem. Engng*, **26**, pp. 921-935.
- Alici S. and Edgar T.F. (2002). Nonlinear dynamic data reconciliation via process simulation software and model identification tools. *Ind. Engng Chem. Research*, **41**, pp. 3984-3992.
- Alici S. (2001). Dynamic data reconciliation using process simulation software and model identification tools, *Ph.D. Th.*, U of Texas Austin.
- Crowe C. M. (1996). Data reconciliation progress and challenges. *Jour. of Proc. Cont.*, **6**, pp. 89-98.
- Kuehn D.R. and Davidson H. (1961). Computer Control. *Chem. Engng Prog.*, **57**, pp. 44-47.
- Liebman M.J., Edgar T.F. and Lasdon L.S. (1992). Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques. *Comp. and Chem. Engng*, **16**, pp. 963-986.
- Liebman M.J. and Edgar T.F. (1988). Data reconciliation for nonlinear processes. *Proc. of AIChE Annual Meeting*, Washington DC.
- Soderstrom T.A., Himmelblau D.M. and Edgar T.F. (2001). A mixed integer optimization approach for simultaneous data reconciliation and identification of measurement bias. *Cont. Engng Practice*, **9**, pp. 869-876.
- Taylor J.H. (1981). Statistical performance analysis of nonlinear stochastic systems by the monte carlo method. *Math. & Comp in Simul.*, **XXIII**.