15th Triennial World Congress, Barcelona, Spain

USE OF MODEL REDUCTION AND IDENTIFICATION TOOLS FOR DYNAMIC DATA RECONCILIATION

Semra Alici * **Thomas F. Edgar** **

* *Air Products and Chemicals, Inc., USA*
** *The University of Texas at Austin, USA*

Abstract: Recent approaches for nonlinear and dynamic data reconciliation suffer from inapplicability and infeasibility for large systems. Because these systems are expressed by differential and algebraic equations, the complete problem definition requires a considerable number of equations that need to be solved simultaneously during the solution of the nonlinear programming problem. One way in avoiding this is to use a commercial software package to model a process and to reduce the size of the model by generating an input-output model from the simulation results. In this research two different approaches are presented to describe dynamics of the system and reduce the size of the model by model identification techniques.

Keywords: Data Acquisition, Dynamic Simulators, Model Reduction, Model Identification, Estimation Theory

1. INTRODUCTION

Data reconciliation is one of the several data analysis methods that utilize a process model to provide more accurate estimates of measured process variables. The majority of the research in data reconciliation has been on the steady state linear systems while most industrial systems are nonlinear and dynamic.

Different approaches have been applied for the extension of data reconciliation to nonlinear dynamic systems. One approach is the application of successive linearization techniques which linearize the nonlinear model about an operating point. Another approach is the solution of the nonlinear and dynamic data reconciliation problem using Nonlinear Programming (NLP) techniques. The nonlinear dynamic data reconciliation formulation using first principles model and applying the latter approach was first developed by (Liebman *et al.*, 1992). This algorithm assumes the process model represents the system accurately. A moving horizon approach can be utilized in order to restrict the size of the problem, and collocation on finite elements method can be applied to convert differential-algebraic equation (DAE) system into a set of algebraic equations (Biegler, 1984).

Liebman's algorithm is not suitable for large systems such as first principles models. For a process which has NS state variables and NI input variables with a history horizon of H and NC collocation points, it requires the solution of a problem with $NS \times (NC-1) \times H$ equations and $NS \times (NC-1) \times H + NI \times H$ variables simultaneously. Large computation time requirement and modeling difficulties make on-line application of this approach almost impossible for large processes. Instead of using first principles, the process model can be developed by simulation software. However, this involves interfacing the software with a reliable optimization package, and reducing the size of the problem by model reduction techniques.

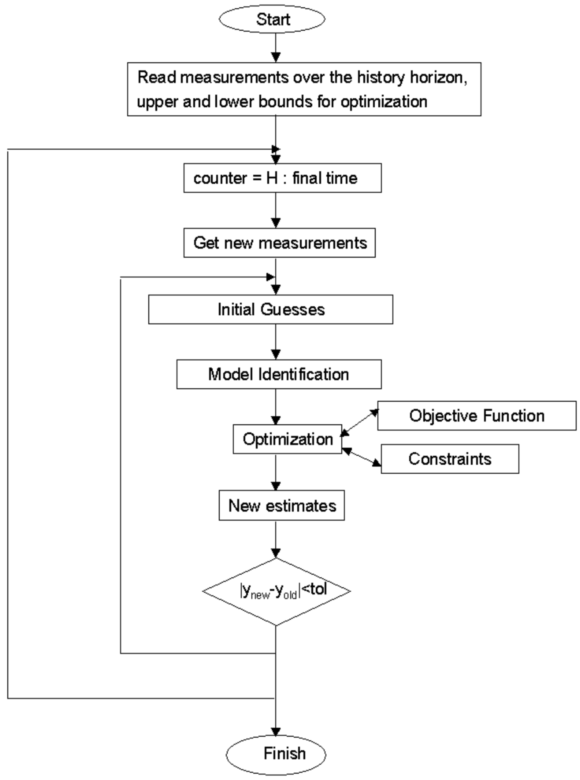

Fig. 1. Nonlinear Data Reconciliation Algorithm

The main objective of this paper is to develop a computationally efficient and industrially feasible approach for the use of commercial simulation software as a modeling tool in data reconciliation. In order to achieve this goal an intermediate model identification step was added to the algorithm as shown in Figure 1. A reduced model was generated by applying input-output modeling approaches to the simulation results at each data reconciliation iteration. Two new approaches have been developed to carry out data reconciliation with dynamic simulation software. In the first approach the dynamics of the system is represented by ordinary differential equations similar to Liebman's Nonlinear Dynamic Data Reconciliation algorithm. A finite difference method is combined with a model identification tool to generate a simplified local model of the process. In the second approach the dynamics of the process is represented by recursive equations. This method utilizes time series analysis generated by fitting a time series model to simulation results. Different model identification tools can be used to fit the data into either finite difference or time series models.

This paper is organized as follows: First, the nonlinear dynamic data reconciliation problem is reviewed. Second, proposed approaches are introduced and formulated. Next, these methods are implemented in two example problems: The first one is a continuous flow stirred tank reactor (CSTR) system with a known model. The second example is a depropanizer process modeled with HYSYS© simulation package. Finally, the validity and the applicability of these approaches are discussed.

2. NONLINEAR DYNAMIC DATA RECONCILIATION OVERVIEW

The nonlinear dynamic data reconciliation problem was expressed as the least squares problem in the following form by (Liebman *et al.*, 1992):

$$\min_{\hat{\mathbf{y}}} \int \frac{1}{2} \left(\hat{\mathbf{y}}(t) - \mathbf{y}(t)\right)^T \mathbf{V}^{-1} \left(\hat{\mathbf{y}}(\mathbf{t}) - \mathbf{y}(t)\right) \quad (1)$$

subject to

$$\begin{aligned} \mathbf{f}\left(\frac{d\hat{\mathbf{y}}(t)}{dt}, \hat{\mathbf{y}}(t)\right) &= 0 \\ \mathbf{h}\left(\hat{\mathbf{y}}(t)\right) &= 0 \\ \mathbf{g}\left(\hat{\mathbf{y}}(t)\right) &\geq 0 \end{aligned} \quad (2)$$

where t is the current time, $\hat{\mathbf{y}}$ is a vector of estimates of the measurements, elements of $\mathbf{y}$ are the measurements, $\mathbf{V}$ is the variance-covariance matrix, $\mathbf{f}$ is a vector of differential equation constraints, $\mathbf{h}$ is a vector of equality constraints, and the $\mathbf{g}$ is a vector of inequality constraints.

In order to restrict the size of the problem the moving horizon approach can be applied over a predefined history horizon. In this formulation the objective function is expressed by the following equation:

$$\min_{\hat{\mathbf{y}}} \sum_{k=t-H}^{t} \frac{1}{2} \left(\hat{\mathbf{y}}(t_k) - \mathbf{y}(t_k)\right)^T \mathbf{V}^{-1} \left(\hat{\mathbf{y}}(\mathbf{t_k}) - \mathbf{y}(t_k)\right) \quad (3)$$

where H is the history horizon or size of the history window.

3. SIMULATION SOFTWARE AND MODEL REDUCTION

One way of using simulation software in data reconciliation calculations is the application of a simultaneous solution/optimization technique (Bequette, 1991). This requires simulation of the model at every line search step of the optimization. This approach will not work with simulation software because the whole model must be simulated for the period of history horizon at every iteration of line search during optimization. The proposed approach in this paper is the use

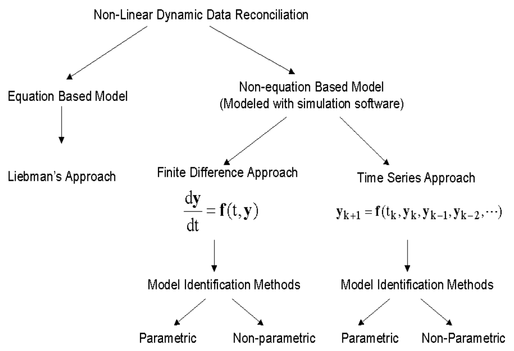

Fig. 2. An Overview of NDDR Algorithms

of model identification techniques which treats the process model as a black box model. Before the optimization step of the data reconciliation algorithm, the process model is simulated and a training set is created from the simulation results. Then, the dynamic model is expressed by ordinary differential equations (ODE) generated by a finite difference method, or as a recursive algorithm such as Autoregressive Time Series model. Next, model parameters are evaluated using a parametric or a non-parametric model identification technique. These methods are illustrated by Figure 2 and explained in preceeding sections in more detail.

The choice of the model identification technique to identify model parameters in the proposed methods is important for the accuracy of the data reconciliation. Parametric or non-parametric modeling techniques can be used for this purpose. An input-output model is identified from the training data set. For the finite difference approach input, data is process variables and output data is the corresponding first derivatives. For the time series analysis approach, input data set is past values of the process variables along history horizon and output data is the current values. The simplest approach is to fit a linear model to data. However, for highly nonlinear systems more sophisticated approaches such as Volterra Models, Recurrent Neural Networks or Multivariate Adaptive Regression Splines methods can be applied (Alici, 2001).

3.1 *Finite Difference Approach*

In the first approach the process simulation software, which is treated as a black box model, is simulated to generate a training data set and then, a finite difference method is utilized to calculate the first derivatives for each corresponding variable at each sampling time. This creates a table of variable values versus corresponding first derivatives. Then, a parametric or nonparametric model identification method is used to evaluate model parameters. This generates a model that is similar in form to 2. Finally, the problem is solved as a traditional NLP problem:

$$\min_{\hat{\mathbf{y}}} \sum_{k=t-H}^{t} \frac{1}{2} \left(\hat{\mathbf{y}}\left(t_k\right) - \mathbf{y}\left(t_k\right)\right)^T \mathbf{V}^{-1} \left(\hat{\mathbf{y}}\left(t_k\right) - \mathbf{y}\left(t_k\right)\right) \tag{4}$$

subject to

$$\begin{aligned} \frac{d\hat{\mathbf{y}}\left(t\right)}{dt} &= \tilde{\mathbf{f}}\left(\hat{\mathbf{y}}\left(t\right)\right) \\ \mathbf{g}\left(\hat{\mathbf{y}}\left(\mathbf{t}\right)\right) &\geq \mathbf{0} \end{aligned} \tag{5}$$

where H is the size of the history horizon, t is the current time, elements of $\hat{\mathbf{y}}$ are estimates of the measurements, elements of $\mathbf{y}$ are the measurements, $\mathbf{V}$ is the variance-covariance matrix, $\tilde{\mathbf{f}}$ is the approximate differential equation system, and the $\mathbf{g}$ are the inequality constraints. All kinetic, thermodynamic and physical properties of the system are embedded in the dynamic simulator in this representation.

3.2 *Time Series Analysis*

The second approach presented in this paper is the expression of the system's dynamic behavior by a time series model. This approach is especially useful when there is a non-monotonic response in the system as shown in Figures 3, or in the presence of process oscillations. These two situations yield multiple first derivatives for some process variables. In this case since the finite difference approach uses the table of process variables versus first derivatives table, multiple values of first derivatives result in a confusion during model identification and inaccurate estimation of model parameters.

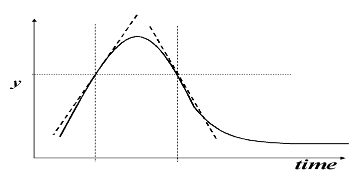

Fig. 3. Multiple $\frac{dy}{dt}$ for a corresponding $\mathbf{y}$ value

Time series models allow filtering and smoothing of process oscillations. When time series analysis is used to identify the dynamic behavior of the system, the constraint equations can be expressed as an autoregressive model where the previous values of all variables along the history horizon are added into the model.

$$\min_{\hat{\mathbf{y}}} \sum_{k=t-H}^{t} \frac{1}{2} (\hat{\mathbf{y}}_k - \mathbf{y}_k)^T \mathbf{V}^{-1} (\hat{\mathbf{y}}_k - \mathbf{y}_k) \quad (6)$$

Subject to

$$\begin{aligned} &\hat{\mathbf{y}}(t_k) = \tilde{f}(\hat{\mathbf{y}}(t_{k-1}), \hat{\mathbf{y}}(t_{k-2}), \hat{\mathbf{y}}(t_{k-3}), \cdots) \\ &h(\hat{\mathbf{y}}(t_k), \hat{\mathbf{y}}(t_{k-1}), \hat{\mathbf{y}}(t_{k-2}), \cdots) = 0 \\ &g(\hat{\mathbf{y}}(t_k), \hat{\mathbf{y}}(t_{k-1}), \hat{\mathbf{y}}(t_{k-2}), \cdots) \geq 0 \end{aligned} \quad (7)$$

The first constraint in Equation (7) replaces the dynamic constraint (ODE) in Equation (5). This equation can be converted into the residuals form to express it as an equality constraint for the solution of the NLP problem:

$$\hat{\mathbf{y}}(t_k) - \tilde{f}(\hat{\mathbf{y}}(t_{k-1}), \hat{\mathbf{y}}(t_{k-2}), \hat{\mathbf{y}}(t_{k-3}), \cdots) = 0 \quad (8)$$

In this equation $\tilde{\mathbf{f}}$ is the approximation to $\mathbf{y}(t_k)$ and evaluated from the previous values of $\hat{\mathbf{y}}$ using model identification tools.

4. RESULTS AND DISCUSSION

In order to validate these approaches proposed algorithms are first applied to a Continuous Flow Stirred Tank Reactor (CSTR) described by the following equations (Liebman *et al.*, 1992):

$$\begin{aligned} \frac{dA}{dt} &= \frac{q}{V}(A_0 - A) - kA \\ \frac{dT}{dt} &= \frac{q}{V}(T_0 - T) + \frac{-\Delta H}{\rho C_p} kA - \frac{UA_R}{\rho C_p V}(T - T_c) \end{aligned} \quad (9)$$

where

$k = k_0 e^{\frac{-E_A}{T}}$, an Arrhenius rate expression
A_0 = concentration of the reactant in the feed
T_0 = temperature of the feed stream
A = concentration of the reactant in the tank
T = tank temperature

There are two state variables, tank concentration and temperature, and two input variables, feed concentration and the feed temperature. Physical properties can be obtained from (Liebman *et al.*, 1992).

A step change was introduced to the system at time=20 seconds by increasing the scaled feed concentration from 6.5 to 7.5. The data reconciliation problem was first solved by the traditional NDDR method for a history horizon of 5. Next, the same problem was solved using the proposed finite difference method under the same conditions. Finally, the time series analysis method was applied to the same problem for a history horizon of 5. Estimation errors standard deviation reduction relative to measurement error standard deviation reductions are tabulated in Table 1 for all cases mentioned above.

Variable	NDDR	FD	TSA
A	76	62	47
T	80	54	53
T_0	66	44	50

Table 1. Standard Deviation Reduction for CSTR Example. (FD is finite difference, TSA is time series analysis.)

The proposed approaches generate results close to the traditional NDDR approach under the same conditions. A non-smooth response is observed for the tank concentration (A). However, since the time steps are small enough, the finite difference approach was capable of catching the dynamic behavior. The time series approach produced less accurate results for the same variable, since it smoothed the data, losing some information on the dynamic behavior. This can be avoided by using a shorter history horizon for the time series approach. Unlike the tank concentration, results for the feed temperature are closer to the traditional NDDR approach because of the smoothing property of the time series approach. Because of the first order filter behavior of the data reconciliation algorithm, step changes are not recognized immediately. The algorithm behaves more cautiously in case of a sudden change considering it can be an outlier data.

FD and TSA approaches were also applied to a more realistic depropanizer system as shown in Figure 4 modeled in HYSYS (AEA, 1998). The model is interfaced with the data reconciliation program using Object Linking Embedding (OLE). Eight variables were chosen to be reconciled. Those variables are: Temperatures of Feed 1, Feed 2, SepVap, LTSExit; molar flow rates of SepVap, SepLiq, TowerInlet; and vapor phase fraction of LTSExit. Among these the first two variables are input variables, and the rest are the state variables.

Measurement data was generated using a Gaussian noise generator, and reconciled using proposed approaches. It is not possible to apply the traditional NDDR method in this case, since there is not any explicit set of model equations. Four step changes are made in the system. Two step changes were introduced to the temperature of Feed 1 at time=800 s and 3000 s by first increasing it from 15.55 C to 25.55 and then decreasing it back to 15.55 C. Two additional step changes were made on the temperature of Feed 2 at time=2100 s and 3000 s by first increasing it from 15.55 C to 25.55 C and then decreasing it back to 15.55 C .

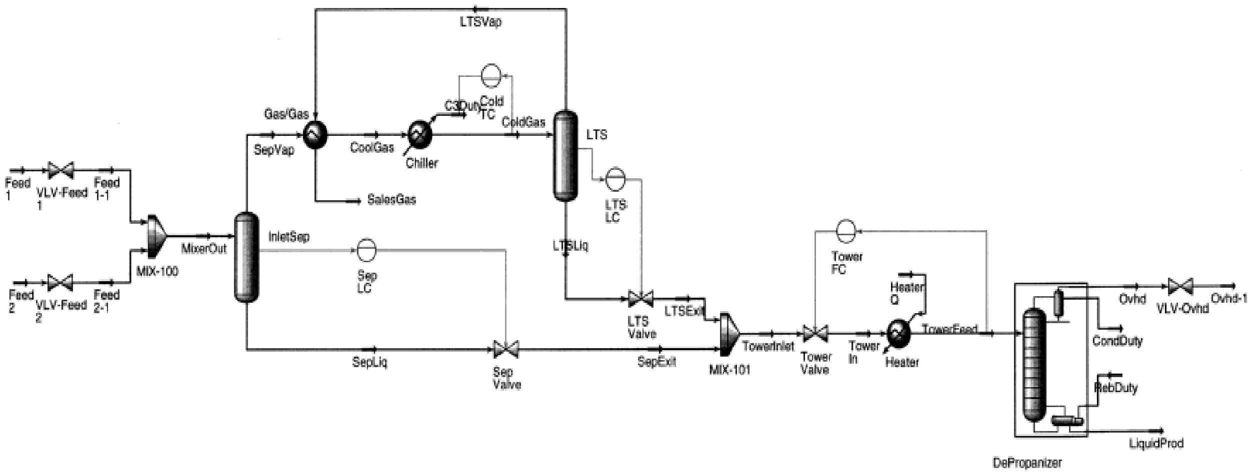

Fig. 4. Example II, Depropanizer Unit

The proposed approaches were applied to this depropanizer system. The standard deviation reductions were tabulated in Table 2 for each variable for these two methods. In this table "-" indicates that since the dynamics of the process is fast at each step change, the data reconciliation algorithm behaves cautiously in order to prevent an unnecessary action for an outlier. Therefore, it shows a slow response. The response time depends on the size of the history horizon. A shorter history horizon generates faster response, but with less accuracy. In Table 2, although the time series analysis method produced less reduction in standard deviation, it was able to catch some of these fast dynamic fluctuations for the temperature of SepVap that the finite difference method missed.

5. CONCLUSIONS

In this paper two new approaches for dynamic data reconciliation was proposed. The process model which is originally defined by many equations in the simulation software was reduced, treating the model as a black box and applying input-output model identification techniques. The application of these proposed techniques into industrial systems enables more accurate data for process control and on-line optimization. The main weakness of the proposed approach is the computational time is spent on model identification. This can be significantly reduced by utilizing previous simulation results during the model identification step of the next data reconciliation iteration, making on-line data reconciliation feasible.

Variable \ Method	FD	TSA
Molar Flow Rate of SepVap	93.2	72.7
Temperature of SepVap	-	12.7
Vapor Phase Fraction of LTSExit	34.3	31.4
Molar Flow Rate of SepLiq	-	-
Temperature of LTSExit	46.4	35.7
Molar Flow Rate of TowerInlet	67.2	30.1

Table 2. Standard Deviation Reduction for Depropanizer Example

The finite difference and time series analysis methods use different approaches to identify the dynamic behavior of the system. Therefore, both have some advantages and disadvantages over another. For oscillatory systems, time series analysis is better suited since it has better smoothing properties than the finite difference approach. Another advantage of time series analysis approach is that it does not use approximations to first derivatives. One approximation of the finite difference approach is that first derivatives are generated from simulation results applying a finite difference method. Another approximation is that since the finite elements approach uses a model in ODE form, differential equations are converted into algebraic ones using collocation on finite elements during optimization. In time series analysis approach, the model equations are already in algebraic form. Therefore, this second approximation is not necessary.

6. REFERENCES

AEA Technology (1998). *HYSYS User's Guide*.

Alici, S. (2001). Dynamic Data Reconciliation Using Process Simulation Software and Model Identification Techniques. PhD thesis. The University of Texas at Austin.

Bequette, B. W. (1991). Nonlinear control of chemical processes: A review. *Ind. Eng. Chem. Res.* **30**, 1391-1413.

Biegler, L. T. (1984). Solution of dynamic optimization problems by using successive quadratic programming and orthogonal collocation. *Comp. Chem. Eng.* **8**(3/4), 243-248.

Finlayson, B. A. (1972). *The Method Of Weighted Residuals And Variational Principles*. Academic Press, Inc.

Jang, S. S., B. Joseph and H. Mukal (1986). Comparison of two approaches to online parameter and state estimation of nonlinear systems. *Ind. Eng. Chem. Process Des. Dev.* **25**, 809-814.

Liebman, M. J. (1991). Reconciliation of Process Measurements Using Statistical and Nonlinear Programming Techniques. PhD thesis. The University of Texas at Austin.

Liebman, M. J., T. F. Edgar and L. S. Lasdon (1992). Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques. *Comp. Chem. Eng.* **16**(11/12), 963-986.

Pearson, R. K. and B. A. Ogunnaike (1997). Nonlinear process identification. *in Nonlinear Process Control, eds. M. A. Henson and D. E. Seborg.*

Ramamurthi, Y., P. B. Sistu and B. W. Bequette (1993). Control-relevant dynamic data reconciliation and parameter estimation. *Comp. Chem. Eng.* **17**(1), 41-59.

Seider, D. W., J. D. Seader and D. R. Lewin (1999). *Process Design Principles: Synthesis, Analysis, and Evaluation.* John Wiley & Sons.

Sistu, P. B., R. S. Gopinath and B. W. Bequette (1993). Computational issues in nonlinear predictive control. *Comp. Chem. Eng.* **17**(4), 361-366.