Subspace Identification for Open and Closed Loop Petrochemical Plants

Elder M. Hemerly, hemerly@ita.br
Technological Institute of Aeronautics - Electronics Division, Systems and Control Department
Campo Montenegro - Vila das Acácias 12228-900 São José dos Campos/ SP – Brazil

Rafael P. Breckenfeld, rafael.breckenfeld@chemtech.com.br
Rua Werner Siemens, 111, prédio 20, 2º andar
Lapa – 05069-010 São Paulo/SP-Brazil

Valmir J. Camolesi, camolesi@petrobras.com.br
Rod. Presidente Dutra, km 143
Jd. Diamante – 12223-900 São José dos Campos/SP- Brazil

Abstract. *This paper addresses the experimental identification of petrochemical plants by using subspace identification methods (SIM). The 4 main peculiarities of these plants are: a) MIMO characteristics; b) unavailability in general of a structure for the model; c) low level excitation for security reasons, and d) typical need for linearized models, which are used for implementing predictive control strategies. The first 2 characteristics rule out, for instance, the use of identification based on the prediction error method, due to the parameterization problem. However, the SIM seems to be efficient here, since they are tailored to provide a MIMO black box linear model of reduced order. An implementation of a subspace identification method which deals with the feedback case is carried out and compared with alternative implementations, with simulated and real plants (located at Petrobras-REVAP). The effect of noise, feedback and bad data scaling, which are usual in the industrial scenario, are investigated and discussed.*

Keywords*: system identification, subspace identification, petrochemical plants*

1. INTRODUCTION

Petrochemical plants have some particularities, such as multivariable structure, complex model and low level of excitation for security reasons. Moreover, in some cases it is not possible to open the loop for performing system identification, also for security matters. Besides, in general models of petrochemical plants are sought for in the standard black box state space linear form, since the main use for these models is to enable the design of a predictive controller.

By leaving temporarily aside the problem raised by feedback operation, it can then be concluded that the subspace identification method is a good candidate for performing identification of petrochemical plants, since this method is tailored for obtaining MIMO black box linear state space models. Indeed, for black box models the alternative method, namely the prediction error method, suffers from the parameterization problem: inadequate selection of the initial parameter estimates can lead to instability. This difficulty is ameliorated, but not solved entirely, by using the approach in (search for the paper), which proposes a parameterization with the smallest number of parameters in the dynamic matrix. In this work, simulations were performed with this minimum parameterization and using the prediction error method, with the general conclusion that instability can arise due to inadequate initial parameter estimates. It must be stressed, however, that the prediction error method is efficient for identifying grey box models, for the system structure is available.

In contrast to the prediction error method for black box models, the SIM do not need any initial parameter estimate, and the model structure, namely the number of states, is also obtained from the data, by means of a singular value analysis. For further details about the SIM and the several alternative implementations, the reader is referred to Qin (2006).

Identification of general industrial processes were considered by Favoreel *et al.* (2000), where the subspace identification method called N4SID was compared to the PEM (Prediction Error Method), with respect to computational complexity and prediction error value. It must be stressed, as the authors did, that the SIM and PEM are not competing methods, but actually complementary. For instance, initial order and system parameter estimates can be obtained via SIM, and then be refined by the PEM, by using these estimates as initial conditions. A total of 10 data sets were considered for performance evaluation and the conclusions were: 1) the PEM, when not trapped in a local minimum, provides smaller prediction errors, as expected, with an improving factor of 2 with respect to the N4SID, in the worst case, and 2) the computational complexity of the N4SID is much smaller, with an improvement factor of 35 with respect to the PEM, in the best case. The overall conclusion is that the SIM are fast, since no recursive optimization is used, and are sufficiently accurate in practical applications.

The identification of petrochemical plants has been addressed recently, as in Meshksar *at al.* (2007) and Logist *et al.* (2009). In Meshksar *et al.* (2007), a grey model structure was first obtained, by using mass and energy balance. An output error method, based on a genetic algorithm, was then employed for parameter estimation. The method was applied to a pilot distillation column, and the prediction capability for the 2 outputs, top plate boiler and reflux temperatures, presented.

In Logist *at al.* (2009) a pilot scale distillation column is considered and the commercial software INCA (IPCOS, Leuen-Belgium, see www.ipcos.com for further details) is employed for black box identification. Two parameterizations are compared: finite impulse response (FIR) and state space. For the state space model, the SIM is used for structure and parameter identification. The conclusion was that the state space parameterization provided better prediction capability than the FIR one, with smoother behavior. Second order state space model was then used in the identification procedure.

This paper describes identification experiments carried out with petrochemical plants located at Petrobras-REVAP, by using subspace identification method. The implemented method, called SIM_gen is compared with 2 alternative proprietary codes: N4SID, from Matlab, and SMC, from SMCA (SETPOINT Multivariable Control Architecture). The first one is also a subspace method, and the second employs an ARX parameterization. Therefore, the performance comparison is comprehensive. Besides, whereas in Meshksar *et al.* (2007) and Logist *et al.* (2009) only plants with smaller number of inputs and outputs are considered, in this work truly MIMO systems are investigated.

Another important aspect of petrochemical plants has to do with the fact that sometimes, due to potential instability or poor damping, the loop can not be opened for a classical identification procedure. However, it is well known that under feedback the model can be biased. Therefore, a procedure for data decorrelation is also implemented, which enables closed loop identification. This is imperative, for instance, for enabling improvement of a predictive controller by obtaining a better plant model through identification, without the need for opening the loop. Indeed, the controller in this case does need the plant open loop model. Simulated data is considered for the closed loop operation case, in order to better illustrate the effect of noise.

2. SUBSPACE IDENTIFICATION

System identification concerns structure and parameters determination from experimental data. Regarding the structure, there are 2 possibilities, namely a) grey box: balance relations can be used to obtain the system mathematical model, regardless of the data, and b) black box: both structure and parameters must be obtained from the experimental data. For the grey box case, as in the aeronautical area, in the time domain the parameters can be estimated by using several approaches, for either liner or nonlinear models, such as Equation Error Method, Output Error Method, Filter Error Method and Filtering Approach. For further details, see, for instance, Jategaonkar (2006).

For linear black box models, the aforementioned approaches have difficulty with the MIMO case: the inadequate selection of the model structure and parameterization can lead to a numerically ill-conditioned problem, due to poor identifiability (Katayama, 2005). Besides, these methods are based on recursive optimization, hence can be trapped at a local minimum.

An alternative for black box MIMO systems identification are the SIM (Subspace Identification Methods), which are based on the realization theory. This approach sidesteps 3 main difficulties in the MIMO identification, namely there is no parameterization problem (since the model structure is also obtained from the experimental data), the solution is non recursive, basically amounting to the solution of 1 or several SVD (Singular Value Decomposition) problem, which is numerically stable and reliable, and there is no need for an initial state estimate. For some examples of SIM implementations, such as CVA (Canonical Variate Analysis) (Larimore, 1990), N4SID (Numerical algorithm for Subspace State-Space System Identification) (van Overschee and De Moor, 1994) and MOESP (Multivariable Output-Error State-Space) (Verhaegen and Dewilde, 1992) see Quin (2006).

All the comments so far concern the open loop operation scenario. However, in some applications, such as in the petrochemical are, the plant model must be obtained under closed loop operation, mainly for security reasons. This brings a complication, namely the correlation between the output and the unmeasured noise. The simplest way forward here would be to identify the closed loop dynamics and then, by using knowledge about the controller structure, to obtain the plant model. This however suffers from a basic limitation: even if the controller structure were know, the mapping from the closed to the open loop dynamics may not be unique. Therefore, a better approach is to decorrelate the input and measurement noise, so as to avoid parameter estimate bias, as in Jansson (2003). This is achieved by means of an ARX modeling of the future outputs. With this proviso, the method can be used indifferently for open or closed loop operation.

The state space discrete time model employed in this paper has *n* states, *m* outputs and *l* inputs and is written in the so called innovation form, i.e.,

$$\boldsymbol{x}(t+1)=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{B}\boldsymbol{u}(t)+\boldsymbol{K}\boldsymbol{e}(t) \tag{1}$$

$$\boldsymbol{y}(t)=\boldsymbol{C}\boldsymbol{x}(t)+\boldsymbol{D}\boldsymbol{u}(t)+\boldsymbol{e}(t) \tag{2}$$

where the matrices have appropriate dimensions, and (1) can also be written in the predictor form as

$$\boldsymbol{x}(t+1)=\bar{\boldsymbol{A}}\boldsymbol{x}(t)+\bar{\boldsymbol{B}}\boldsymbol{u}(t)+\boldsymbol{K}\boldsymbol{y}(t) \tag{3}$$
$$\boldsymbol{y}(t)=\boldsymbol{C}\boldsymbol{x}(t)+\boldsymbol{D}\boldsymbol{u}(t)+\boldsymbol{e}(t) \tag{4}$$

which is convenient, see Quin (2006), for identification in closed loop, since $\bar{\boldsymbol{A}}$ guaranteed stable, even if the original dynamic matrix $\boldsymbol{A}$ is unstable. For the relationships between the 3 representation forms (process, innovation and predictor), see Quin (2006).

The identification of (3)-(4) via a subspace method amounts to the estimation of the dimension of the state space, namely, the value of *n*, and all the corresponding 5 matrices $\{\boldsymbol{A},\boldsymbol{B},\boldsymbol{K},\boldsymbol{C},\boldsymbol{D}\}$ in (1)-(2). A brief summary for the closed loop case is presented below. It is based in the approach proposed by Janssonn (2003), but the procedures for solving the ensuing least squares sub problems are different. As a matter of fact, for better efficiency, the subspace identification method implemented here does not employ LQ (transpose of QR) decomposition and requires a total of 7 singular value decompositions. Besides, the coding is completely C++ based, for portability and fast operation, and a complete application was implemented, so as to assist the user in selecting the system order, estimate the system matrices and validate the model. See Section 4 for details.

By supposing N readings are available and by packing $f(>n)$ outputs as (f stands for future horizon)

$$\boldsymbol{Y}_f=\left[\boldsymbol{y}_f(f)\;\; \boldsymbol{y}_f(f+1)\cdots \boldsymbol{y}_f(f+N-1)\right] \text{ (Hankel form)} \tag{5}$$
where
$$\boldsymbol{y}_f(f)=\left[\boldsymbol{y}(f)\;\; \boldsymbol{y}(f+1)\cdots\; \boldsymbol{y}(2f-1)\right]^T \tag{6}$$

based on the prediction form (3)-(4), the extended state space model is obtained, i.e.,

$$\boldsymbol{Y}_f=\boldsymbol{M}_1\boldsymbol{X}_f+\boldsymbol{M}_2\boldsymbol{U}_f+\boldsymbol{M}_3\boldsymbol{Y}_f+\boldsymbol{E}_f \tag{7}$$

The matrices in (7) are given by

$$\boldsymbol{M}_1=\begin{bmatrix}\boldsymbol{C}\\ \boldsymbol{C}\bar{\boldsymbol{A}}\\ \vdots\\ \boldsymbol{C}\bar{\boldsymbol{A}}^{f-1}\end{bmatrix} \text{ (extended observability matrix)} \tag{8}$$

$$\boldsymbol{M}_2=\begin{bmatrix}\boldsymbol{D} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{C}\bar{\boldsymbol{B}} & \boldsymbol{D} & \cdots & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{C}\boldsymbol{A}\bar{\boldsymbol{B}} & \boldsymbol{C}\bar{\boldsymbol{B}} & \cdots & \boldsymbol{0} & \boldsymbol{0}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \boldsymbol{C}\bar{\boldsymbol{A}}^{f-2}\bar{\boldsymbol{B}} & \boldsymbol{C}\bar{\boldsymbol{A}}^{f-3}\bar{\boldsymbol{B}} & \cdots & \boldsymbol{C}\bar{\boldsymbol{B}} & \boldsymbol{D}\end{bmatrix},\; \boldsymbol{M}_3=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{C}\boldsymbol{K} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{C}\bar{\boldsymbol{A}}\boldsymbol{K} & \boldsymbol{C}\boldsymbol{K} & \cdots & \boldsymbol{0} & \boldsymbol{0}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \boldsymbol{C}\bar{\boldsymbol{A}}^{f-2}\boldsymbol{K} & \boldsymbol{C}\bar{\boldsymbol{A}}^{f-3}\boldsymbol{K} & \cdots & \boldsymbol{C}\boldsymbol{K} & \boldsymbol{0}\end{bmatrix} \text{ (Toeplitz matrices)} \tag{9}$$

and by following the same approach, a relation for past values similar to (8) can also be obtained. From these relations, if follows that $\boldsymbol{X}_f$ depends on $\boldsymbol{X}_p$ by $\bar{\boldsymbol{A}}^f$. Since $\bar{\boldsymbol{A}}$ is stable, for large f it is possible to approximate $\boldsymbol{x}(t)$ by previous value of inputs and outputs. This way, $\boldsymbol{Y}_f$ in (7) can be rewritten as function of past and future inputs and outputs. Moreover, from (4) and past values of inputs and outputs a least squares problem can be solved for estimating the matrices $\hat{\boldsymbol{M}}_2$ and $\hat{\boldsymbol{M}}_3$. The regression, as function of the states,

$$\boldsymbol{Z}=\boldsymbol{Y}_f-\hat{\boldsymbol{M}}_2\boldsymbol{U}_f-\hat{\boldsymbol{M}}_3\boldsymbol{Y}_f \tag{10}$$

has uncorrelated noise, hence can be solved via several methods, see Quin (2006) for details. The system order can then be estimated by selecting the largest n singular values. Under low noise operation, the singular values plot displays a clear decaying pattern, which enables a straightforward selection of the system order. However, under high noise, this task is harder and some tests must be performed, having the prediction error as selecting criterion. Once the system order is selected, the state estimates $\hat{\boldsymbol{X}}_f$ can be obtained and from this point onwards the system matrices can be obtained by solving least squares problems.

In Di Ruscio (2009) is presented a bootstrap subspace identification method for dealing with the closed loop case, but it is a recursive procedure. Hence it may be susceptible to convergence problem cause by inadequate initial estimates. Indeed, for the 2x2 MIMO system with 3 states considered there, a biased model was obtained. Therefore, this approach is not considered in this work.

3. PERFORMANCE EVALUATION UNDER FEEDBACK OPERATION

In this section the MIMO plant used by Di Ruscio (2009) is employed. However, in contrast to the investigation carried out there, namely the effect of the number of data points, here we consider the impact of the state and output noises and also on the excitation of the input signal, since they have higher influence on the model quality.

Consider then the closed loop MIMO system described in the process form

$$\boldsymbol{x}(t+1)=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{B}\boldsymbol{u}(t)+\boldsymbol{w}(t) \tag{11}$$

$$\boldsymbol{y}(t)=\boldsymbol{C}\boldsymbol{x}(t)+\boldsymbol{v}(t) \tag{12}$$

where

$$\boldsymbol{A}=\begin{bmatrix}1.5 & 1 & 0.1\\ -0.7 & 0 & 0.1\\ 0 & 0 & 0.85\end{bmatrix}, \quad \boldsymbol{B}=\begin{bmatrix}0 & 0\\ 0 & 1\\ 1 & 0\end{bmatrix}, \quad \boldsymbol{C}=\begin{bmatrix}3 & 0 & -0.6\\ 0 & 1 & 1\end{bmatrix} \tag{13}$$

and the loop is closed via

$$\boldsymbol{u}(t)=\boldsymbol{G}(\boldsymbol{y}_{ref}(t)-\boldsymbol{y}(t)) \quad , \quad \boldsymbol{G}=\begin{bmatrix}0.2 & 0\\ 0 & 0.2\end{bmatrix} \tag{14}$$

A total of $N=1000$ samples are considered in the simulations, and the problem is to estimate the system matrices $\{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}\}$ in (11)-(12). Three subspace identification method are used and compared, namely SIMF (implemented in this work), N4SID (Matlab) and SMC (SETPOINT). In order to evaluate the estimated model, the step response is used, since the SMC only provides this information to the user. The conclusions were the following: for open loop operation and small noises, all the 3 methods provide similar results, since the data is well conditioned. By the way, this scenario explains why the large majority of reported simulated results on SIM are so good. Needless to say that in the industrial case, this favorable combination (low noise and well-conditioned data) is far-fetched and unrealistic. The addition of output noise brings some problem, but not much. However, more severe is the presence of state noise, as expected. Indeed, as it is well known, in this case the best solution for the identification problem is to employ the Filter Error Method (Ravindra, 2006), which naturally brings the Kalman Filter in, for dealing with the state noise by means of optimal state estimation.

But the focus in this section lies in the feedback operation. Several simulations were performed by varying the noises intensities and also the input excitation. A typical result, evaluated by using the step responses as index of performance, is presented in Fig. 1. It should be stressed that the results for the SMC method is not inserted in Fig. 1, because the results were even worse that those displayed by the N4SID. Actually, the inadequate results exhibited by the N4SID and SMC were somehow expected, since these methods do not have any mechanism for dealing with feedback, hence the parameter estimates can be considerably biased. The N4SID and SMC performances degrade even further if the excitation of the input signal $\boldsymbol{y}_{ref}(t)$ is decreased. Thus is bound to happen in the case of petrochemical plants, where large excitation is in general not allowed, for production quality preservation and stability reasons.

The plain conclusion of these simulations is: the usual SIM reported in the literature only exhibit good performance in operation under low noise, adequate excitation levels and open loop operation. All these favorable conditions may not be met by petrochemical plants. As a matter of fact, suppose a control system under predictive control is not exhibiting good performance. Then the recommended procedure for petrochemical plants would be to improve the system model, by performing identification in closed loop, and then updating the system model employed by the predictive controller. This is an example in which it is not realistic to open the loop for identification, for stability reasons.

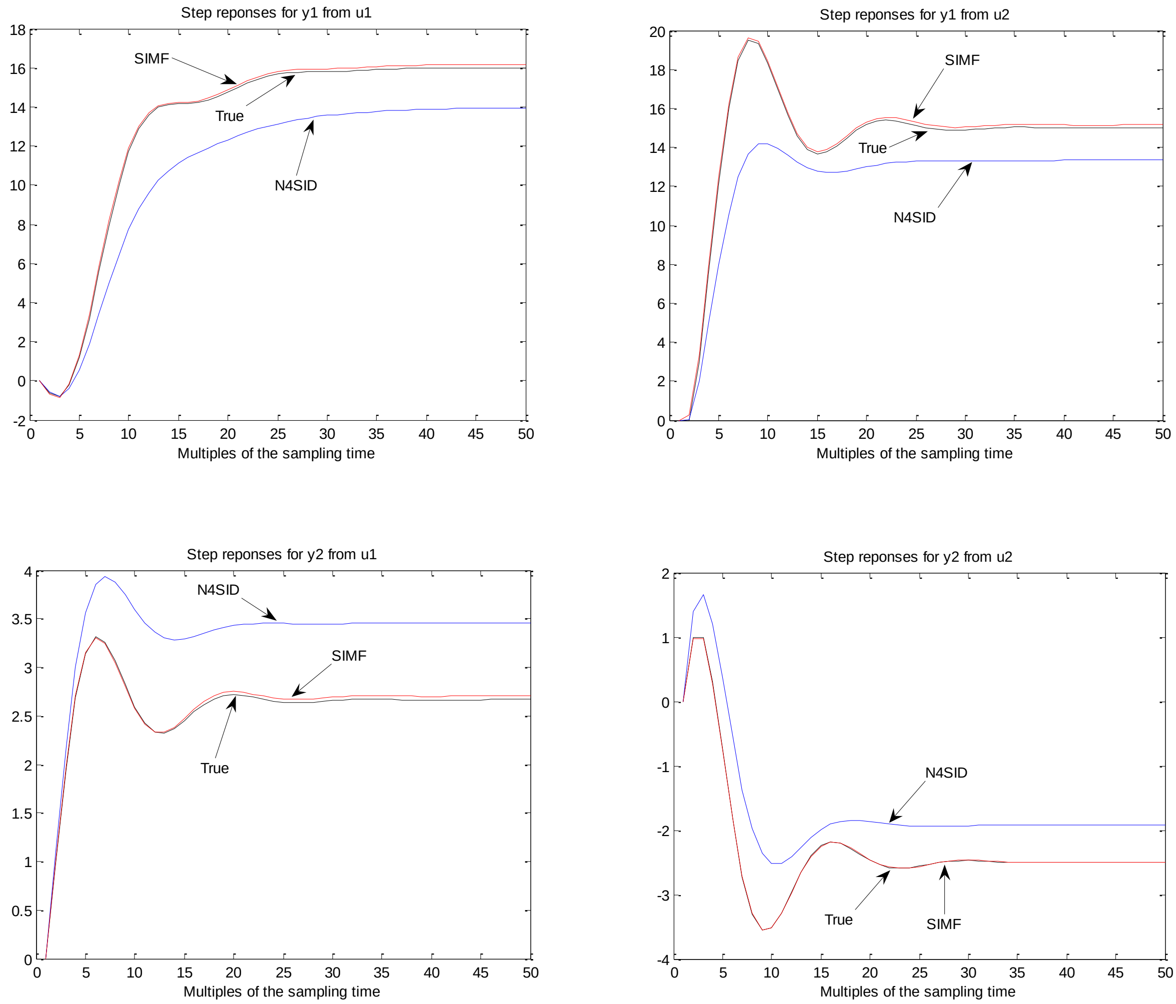

Figure 1. Performance evaluation for the 2x2 MIMO system, under feedback.

To summarize, based on Fig. 1 it is concluded that the SIMF, implemented in this work, can cope with the closed loop operation, by performing data decorrelation.

4. PERFORMANCE EVALUATION WITH REAL PLANTS

Two plants, located at Petrobras-Revap, are now considered. This investigation is relevant because the plants present the unfavorable scenario bound to be met with real systems: high measurement noise, low excitation and possibly badly conditioned data.

4.1- Plant 1- Hydrogen generation unit: This plant has 3 inputs and 8 outputs. In Fig. 2 are shown: a) the measured and predicted signals for the 6-th output, comparing the performance of the SIMF and SMC, and b) the ratio between the variance of the prediction errors for all outputs.

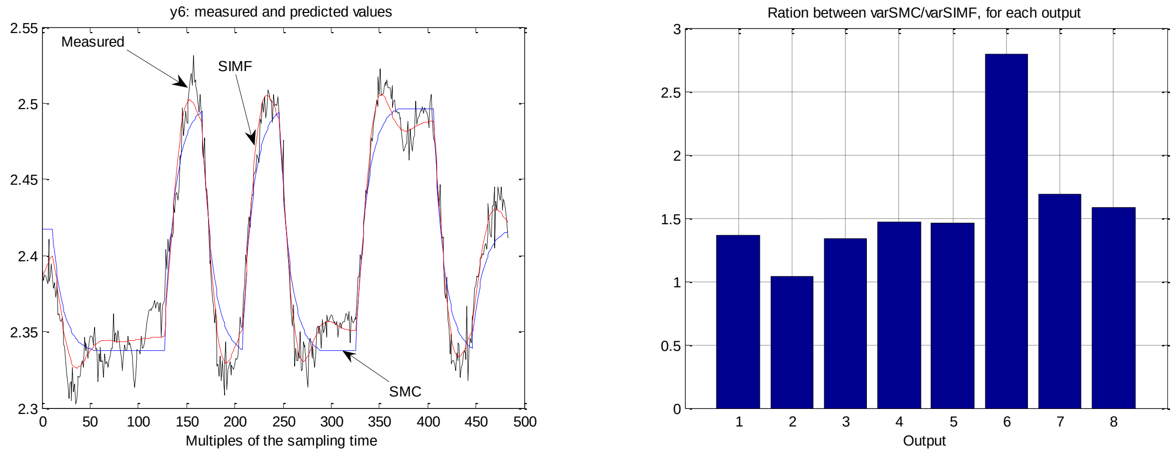

Figure 2. Measured and predicted signals for the 6-th output and ratio between the prediction error variances, for each output, for plant 1.

From Fig. 2 it is concluded that the implemented method outperforms the SMC. The results for N4SID are not presented because they are similar to those from SMC.

4.2- Plant 2- Diesel hydrotreatment unit: The plant has 2 inputs and 5 outputs. This is an example of bad conditioned data, since the 4-th output ranges from 332 to 348, whereas the 2-nd output ranges from -3 to 6. The measured and predicted signals for the 4-th output is shown in Fig. 3.

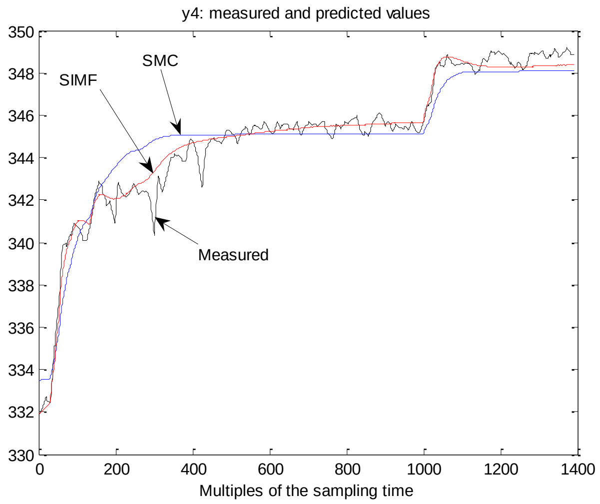

Figure 3. Measured and predicted signals for the 4-th, for plant 2.

For the signals in Fig. 3, the ratio between varSMC/varSIMF is around 4. Therefore, the implemented method outperforms SMC. The result from N4SID is not presented because it is similar to that exhibited by SMC.

5. CONCLUSIONS

The identification of petrochemical plants has been investigated by means of the subspace identification method, taking into account some peculiarities of these plants: high order MIMO system, presence of measurement noise, potentially bad data conditioning and eventually impossibility of opening the loop for identification. The implemented method, called SIMF, was compared, via simulation and with real plants, with 2 other methods. Good results were obtained in both cases, particularly under feedback operation. The model obtained from the identification procedure can be used, for instance, to improve the performance of a predictive controller.

6. REFERENCES

Di Ruscio, D., (2009), "A Bootstrap Subspace Identification Method: Comparing Methods for Closed Loop Subspace Identification by Monte Carlo Simulations", Modeling, Identification and Control, Vol. 30, No. 4, pp. 203-222.

Favoreel, W., De Moor, B., and Van Overschee, P., (2000), "Subspace state space system identification for industrial processes", Journal of Process Control, 10:149–155.

Janssonn,M., (2003), "Subspace Identification and ARX Modelling", Proceedings of the 13th IFAC SYSID Symposium.

Jategaonkar, R.V., (2006), "Flight Vehicle System Identification", AIAA Inc., Reston-VA, USA.

Katayama, T., (2005), "Subspace Methods for System Identification", Springer-Verlag, London.

Larimore, W.E., (1990), "Canonical variate analysis in identification, filtering and adaptive control", Procedings of the 29-th Conference on Decision and Control, pp. 596-604.

Logist, F., Huyck, B., Fabré, M., Verwerft, M., Pluymers, B., De Brabanter, J., De Moor, B., Van Impe, J., (2009), "Identification and Control of a Pilot Scale Binary Distillation Column", Proceedings of the European Control Conference, Budapest, Hungary.

Meshksar, S., Khayatian, A., Jenabali, M., Hashemi, M., (2007), "Modeling and Identification of a Pilot Distillation Column for Control and Estimation using Genetic Algorithms", The 33rd Annual Conference of the IEEE-IECON, Taipei, Taiwan.

Qin, S. J., (2006), "An overview of subspace identification", Computers & Chemical Engineering, 30, pp. 1502-1513.

van Overschee, P., De Moor, B., (1994), " N4SID: Subspace algorithms for identification of combined deterministic-stochastic systems", Automatica, 30, pp. 75-93.

Verhaegen, M., Dewilde, P., (1992), "Subspace model identification, part i: The output-error state-space model identification class of algorithms", International Journal of Control, 56, pp. 1187-1210.

7. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.