Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Multirate Data Assimilation in a Cultivation Process

Barrero, O. * **Petersen, J.** ** **Jorgensen, S.B.** ***

* *Industrial Automation Department, Universidad de Ibagué, Carrera 22 Calle 67 Ibagué - Tolima, Colombia (e-mail: oscar.barrero@unibague.edu.co)*
** *Seniorchemist at Novo Nordisk A/S, DK 2880 Bagsvrd, Denmark (e-mail: jpet@novonordisk.com)*
*** *Chemical Engineering Department, Technical University of Denmark, DK 2800 Lyngby, Denmark (e-mail: sbj@kt.dtu.dk)*

Abstract: Infrequent and delayed output measurements are a commonly found in industrial processes. Hence, multirate estimation has gained some attention from researchers in the past decades to provide solutions to receiving data with widely differing sampling intervals. In this paper we present a discrete-time multirate estimator which combines ideas from Data Assimilation and the Extended Kalman filter. The performance of the estimator is demonstrated through a cultivation process study.

1. INTRODUCTION

In order to improve the efficiency of industrial processes, it is desirable to monitor and control the physical dynamic variables which are involved in such processes. A problem often encountered in the pharmaceutical industry is that some of the quality variables cannot be measured on-line, but only off-line with a significant time delay of one or more days due to the analytical methods employed. Both the irregular sampling times and the long analysis delays raises important issues when a monitoring and/or a control system is be deployed. Soft sensors have received significant attention during the recent decades, these sensors are basically used to estimate the value of a variable based on a dynamical model where the input data are on-line measurements of other related process variables. There exists many approaches to design soft sensors, i.e. neural networks, fuzzy logic, system identification, etc. [Fortuna , 2007, Lin , 2006]. The precise choice of approach depends on the process complexity and knowledge as well as on the available measurement data. However even for soft sensors the issue of infrequent and delayed data constitute an important problem to be dealt with in an appropriate manner. This paper addresses precisely these problems.

The Kalman filter (KF) [Anderson , 1979] results from the linear-quadratic Gaussian estimation problem, i.e. estimating the instantaneous state of a linear dynamic system perturbed by a Gaussian random process noise, and using measurements linearly related to the state but corrupted also by a Gaussian random measurement noise. The resulting estimator is statistically optimal for any quadratic function of estimation error [Grewal , 2001]. The Kalman filter was originally designed for linear systems. However it has been extended to the nonlinear case by many authors. As a result, several different algorithms are found in literature; the Extended Kalman Filter (EKF) [Anderson , 1979, Grewal , 2001], the Ensemble Kalman Filter (EnKF) [Evensen , 1994, Penland , 2003], and the Unscented Kalman filter (UKF) [Wan , 2001], just to mention some of them. Now, depending upon the model complexity and the order of the system, it is possible to choose the most appropriated one. For instance, to deal with large scale systems, the EnKF is very efficient, while if the system has strong nonlinearities the UKF works better, or when the model is small scale with weak nonlinearities, the EKF give good results.

An important assumption behind the standard KF in order to guarantee an optimal estimation is that inputs and outputs measurements are assumed available at each time instant. This assumption is often not fulfilled for quality variables in processes. One approach to tackle this problem is multirate estimation [Lee , 1992, Lu , 2004, Zhang , 2005] however this approache still requires regular sampling of the slowly sampled data. This assumption may be expensive too fulfil when slowly sampled data stem from chemical analytical analysis where long and irregular delays may occur. Therefore a new multirate estimator called Multirate Data Assimilating Kalman Filter (MDA-EKF) is presented in this paper. This estimator is based on a technique used in research areas e.g. weather forecasting, where observation data is combined with outputs from a numerical model to produce an optimal estimate of the evolving state of the system [Barrero , 2006, Cohn , 1996, Evensen , 2003]. Then, by assimilating data into a model via the KF, at the time instants when measurement data actually is ready, it is possible to obtain a convergent but suboptimal state estimate.

Moreover, the Kalman filter can be used to estimate not just dynamical variables but also model parameters, recursively. For instance, when process operation changes, due to either physical modifications on the plant, or to changes in operational/environmental conditions. Then the model parameters can be adjusted on line using the Extended Kalman Filter thus potentially rendering the estimation more robust.

978-1-1234-7890-2/08/$20.00 © 2008 IFAC 10.3182/20080706-5-KR-1001.3700

The purpose of the present paper is to develop and demonstrate the applicability of a multirate data assimilating extended Kalman filter which can handle long delays and variations in initial conditions. The paper is organized as follows, in the first section the problem is formulated, then in the second section Multirate Data assimilation using the EKF is introduced. In the section three, a case study concerning estimation of biomass and product concentration in an industrial cultivation process is given, and finally the conclusions are presented.

2. PROBLEM FORMULATION

Given a discrete time nonlinear system

$$\begin{aligned} \boldsymbol{x}_{k+1} &= f(\boldsymbol{x}_k, \boldsymbol{u}_k) + g(\boldsymbol{x}_k, \boldsymbol{u}_k)\boldsymbol{w}_k \\ \boldsymbol{y}_k &= c(\boldsymbol{x}_k, \boldsymbol{u}_k, t_k) + \boldsymbol{v}_k \end{aligned}$$

where $f(\cdot)$, $g(\cdot)$, and $c(\cdot)$ are nonlinear functions of the state variable $\boldsymbol{x}_k \in \mathbb{R}^n$ and the input $\boldsymbol{u}_k \in \mathbb{R}^p$, with output vector $\boldsymbol{y}_k \in \mathbb{R}^m$. Where $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are zero mean, white Gaussian processes. In practice often the quality measurements are delayed and irregular:

$$\boldsymbol{z}_k = \begin{bmatrix} y^1_{k-d_{11}} & y^1_{k-d_{12}} & \cdots & y^1_{k-d_{1i}} \\ y^2_{k-d_{21}} & y^2_{k-d_{22}} & \cdots & y^2_{k-d_{2j}} \\ \vdots & \vdots & \vdots & \vdots \\ y^m_{k-d_{m1}} & y^m_{k-d_{m2}} & \cdots & y^m_{k-d_{ml}} \end{bmatrix} \tag{1}$$

where $\boldsymbol{z}_k$ are infrequent and delayed output measurements with $0 \le \boldsymbol{d}_{ml} \le k$. Notice that the d_{ml} values are not ordered sequentially in time. Thus d_{12} can be larger than d_{11} and less than d_{13}.

Given this problem setting then an optimal estimator, e.g. in the H_2 norm, for $\boldsymbol{x}_k$ based on $\boldsymbol{z}_k$ and $\boldsymbol{u}_k$ is needed.

3. MULTIRATE DATA ASSIMILATING EXTENDED KALMAN FILTER - MDA-EKF

Data assimilation is a technique whereby observation data is combined with outputs from a numerical model to produce an optimal estimate of the evolving state of the system [Evensen , 1994, Penland , 2003]. Note that this definition is similar to that of a standard 'state space observer' in the systems theory literature [Franklin , 2002, Anderson , 1979], the difference is that in a state observer it is assumed that measurements are available at each sampling time while in data assimilation this assumption is not always fulfilled. The methodology developed in this paper is to assimilate data into first-principles based models to render maximal usage of the data and to improve forecasting. The data assimilation scheme provides a structure where multirate measurements can be incorporated easily into the estimator [Myers , 1996]. As a result, a suboptimal estimator, in H_2 norm, that can handle infrequent and delayed output measurements is developed.

Basically, the Extended Kalman filter operation is summarized in the time and the measurement updates:

(1) Time Update

$$\begin{aligned} \hat{\boldsymbol{x}}_{k|k-1} &= f(\hat{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{u}_{k-1}) \\ A_{k-1} &= \left.\frac{\partial f}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_{k-1|k-1}} \\ P_{k|k-1} &= A_{k-1}P_{k-1|k-1}A^T_{k-1} + Q_{k-1} \end{aligned}$$

(2) Measurement Update

$$\begin{aligned} C_k &= \left.\frac{\partial f}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_{k|k-1}} \\ K_k &= P_{k|k-1}C^T_k(C_kP_{k|k-1}C^T_k + R_k)^{-1} \\ \hat{\boldsymbol{x}}_{k|k} &= \hat{\boldsymbol{x}}_{k|k-1} + K_k(\boldsymbol{y}_k - C_k\hat{\boldsymbol{x}}_{k|k-1}) \\ P_{k|k} &= (I_n - K_kC_k)P_{k|k-1}. \end{aligned}$$

Where $P_{k|k-1} \in \mathbb{R}^{n\times n}$ is the error covariance matrix at time k based on information obtained at time $k-1$, where n is the order of the system. $Q_k \in \mathbb{R}^{n\times n}$ the process noise covariance matrix, $R_k \in \mathbb{R}^{m\times m}$ the measurement noise covariance matrix, where m is the number of measured outputs, and $K_k \in \mathbb{R}^{n\times m}$ is the Kalman filter gain.

The above description of the extended Kalman filter assumes that measured outputs $\boldsymbol{y}_k$ are available each sampling instant k, which guarantees an optimal convergence of the estimation error to zero [Anderson , 1979]. It often occur in real processes, e.g. cultivation processes, that quality variable measurements are infrequent and delayed. The data assimilation framework, consists of computing the measurement update only for available output measurements, otherwise, just a time updated is carried out. Thus for delayed and infrequent measurements Kalman filtering can still be used, but with careful management of measurement and time update when the infrequent and delayed measurements become available.

Consequentially, the measurement update will consist of two parts: a) infrequent measurements update: which is carried out by redefining C_k according to the available output measurements, and b) delayed measurement update: which is carried out by recomputing the estimate at time k using the available measurements from $k - d_i$ up to k. A flow diagram of the Multirate Data Assimilating EKF (MDA-EKF) with infrequent and delayed output measurements is shown in Figure 1.

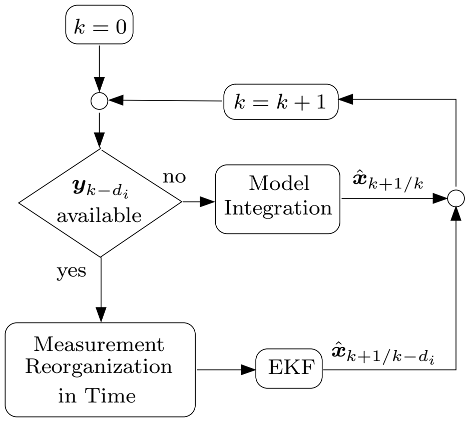

Fig. 1. Multirate Data Assimilating Extended Kalman Filter (MDA-EKF) Scheme with Infrequent and Delayed Output Measurements

4. CASE STUDY

In this section the MDA-EKF estimator is designed for an industrial cultivation process where monitoring is of significant concern. This estimator should to infer product and biomass concentrations in a bioreactor during fed-batch and continuous operation based on on-line input measurements and infrequent and delayed quality output measurements. It is a typical case often found in bio-chemical processes where samples of bioreactor contents are taken to be analyzed in lab, then, after several hours the results become available. Additionally, these delayed measurements also have different time delays for each variable as described above in section 1.

First the collected data are described as follows: The input measured data are mass flow rate of nutrients $\dot{m}_{in}$, mass of culture broth m_b, and ammonia mass flow rate $\dot{m}_{NH_3}$. This data set have a relatively short constant sampling interval of $T_s = 0.0208$ units of time. The off-line measured quality variables are product and biomass concentrations, this data set is sampled infrequently and delayed in time, i.e. one unit of time for X and two for P. Henceforth, since a model-based estimator is to be designed, a gray-box stochastic modelling approach is proposed because firstly, the general model dynamics is known and secondly measured data is available. In the next subsection the model structure is described.

4.1 First Principles Based Model

The model structure selected for this case is one based on reactions in a well stirred bioreactor [Nielsen , 2003]. The general dynamic mass and product balance equations are,

$$\frac{dX}{dt} = r_x - DX \tag{2}$$

$$\frac{dP}{dt} = r_p - DP, \tag{3}$$

where $D = F_{in}/V$ is the dilution rate, r_x the biomass growth rate, r_p the product production rate, P the product concentration, and X the biomass concentration. Now in [Thaysen , 2005], section 9.4, is shown that r_p is proportional to r_x during the continuous stage of the process, hence, we can rewrite (3) as follows

$$\frac{dP}{dt} = Y_{xp} r_x - DP, \tag{4}$$

where Y_{xp} is the yield coefficient of product on biomass.

The Biomass Concentration Model According to [Lei , 2001], for this type of processes, the biomass production rate can be assumed dependent on the addition of ammonia and substrate or, more precisely, the pH difference between them. Consequently, a mass balance for the proton concentration $[H^+]$ during fed-batch and batch operation, yields

$$\begin{aligned} V\frac{d[H^+]}{dt} =& F_{in}[H^+]_s - F_e[H^+]_e \dots \\ &+ F_{H^+,gen} - F_{NH3} - [H^+]\frac{dV}{dt} \end{aligned} \tag{5}$$

where F_{in} is the feed flow rate of nutrients into the reactor, F_e the effluent flow rate, and F_{NH3} the feed flow rate of ammonia. The reactor volume is modelled as (assuming a negligible ammonia feed rate)

$$\frac{dV}{dt} \approx F_{in} - F_e, \tag{6}$$

then substituting (6) into (5) yields

$$\begin{aligned} V\frac{d[H^+]}{dt} =& F_{in}[H^+]_s - F_e[H^+]_e + F_{H^+,gen} \dots \\ &- F_{NH3} - [H^+](F_{in} - F_e). \end{aligned} \tag{7}$$

Assuming perfect mixing, i.e. $[H^+]_e = [H^+]$ yields:

$$V\frac{d[H^+]}{dt} = F_{in}([H^+]_s - [H^+]) + F_{H^+,gen} - F_{NH3},$$

Now, assuming a constant pH level in the reactor, the balance becomes

$$F_{in}([H^+]_s - [H^+]) + F_{H^+,gen} - F_{NH3} = 0$$

from which the following relation is obtained

$$F_{H^+,gen} = F_{NH3} - F_{in}q_{s,H^+}$$

The value of q_{s,H^+} has been determined experimentally to $q_{s,H^+} = 36\ mmole/L$ by [Thaysen , 2005]. Thus the volumetric proton production rate r_{H^+} can be described by

$$r_{H^+} = \frac{F_{H^+,gen}}{V} = \frac{F_{NH3} - F_{in}q_{s,H^+}}{V}, \tag{8}$$

the proton production rate can be used to compute the biomass production rate r_x as follows

$$r_x = \frac{M_{DW} r_{H^+}}{Y_{xH}}. \tag{9}$$

Where M_{DW} is the molecular weight of C-mole dry weight biomass and Y_{xH} is the yield coefficient of mole protons produced per C-mole biomass produced. Inserting (8) into (9) yields a more useful formulation of the biomass production rate:

$$r_x = \alpha r_{H^+} = \alpha\frac{F_{NH3} - F_{in}q_{s,H^+}}{V}, \tag{10}$$

where α represents the product between physiological (M_{DW}, Y_{xH}) and vessel properties. For more details, the reader is referred to chapter 6 in [Thaysen , 2005]. As a result, replacing (10) into (2) and (4) yield the relatively simple model based upon process knowledge:

$$\begin{aligned} \frac{dX}{dt} &= \alpha\frac{F_{NH3} - F_{in}q_{s,H^+}}{V} - DX \\ \frac{dP}{dt} &= Y_{xp}\ \alpha\frac{F_{NH3} - F_{in}q_{s,H^+}}{V} - DP. \end{aligned} \tag{11}$$

4.2 Model Parameter Estimation

After having derived a model structure (11), the data available can be used to tune the model by estimating the unknown parameters. First of all, rewrite (11) as a function of the measured inputs; $\dot{m}_{in}$, $\dot{m}_{NH3}$, and m_b.

$$\begin{aligned} \frac{dX}{dt} &= \frac{\rho_b}{m_b}\left[\alpha\left(\frac{\dot{m}_{NH3}}{M_{NH3}} - q_{s,H^+}\frac{\dot{m}_{in}}{S_{in}}\right) - \frac{\dot{m}_{in}}{S_{in}}X\right] \\ \frac{dP}{dt} &= \frac{\rho_b}{m_b}\left[Y_{xp}\ \alpha\left(\frac{\dot{m}_{NH3}}{M_{NH3}} - q_{s,H^+}\frac{\dot{m}_{in}}{S_{in}}\right) - \frac{\dot{m}_{in}}{S_{in}}P\right], \end{aligned} \tag{12}$$

where the known parameters are: the density of the broth $\rho_b = 1.03\ kg/L$, the concentration of the feed flow $S_{in} = 1.1\ kg/L$, the molar mass of ammonia $M_{NH3} =$

$17.03e^{-3}$ $kg/mole$, and the number of proton equivalents $q_{s,H^+} = 0.036$ $mole/L$. Then, the parameters that have to be estimated are: the vessel and physiological constant α, and the yield coefficient of product on biomass Y_{xp}.

To estimate these unknown parameters a computer program for performing Continuous Time Stochastic Modelling (CTSM version 2.3) is used. This program has been developed at Informatics and Mathematical Modelling (IMM), and CAPEC at the Technical University of Denmark (DTU). Continuous time stochastic modelling means semi-physical modelling of dynamic systems based on stochastic differential equations. Stochastic differential equations contain a diffusion term to account for random effects, but are otherwise structurally similar to the above ordinary differential equations. Therefore conventional modelling principles can be applied to set up the model structure. With the model structure given, the program provides methods for model validation and estimating unknown parameters of the model from data, including the parameters of the diffusion term [Kristensen , 2004]. Consequently (12) can be rewritten as follows,

$$
\begin{aligned}
dX &= \left(\frac{\rho_b}{m_b}\left[\alpha\left(\frac{\dot{m}_{NH3}}{M_{NH3}} - q_{s,H^+}\frac{\dot{m}_{in}}{S_{in}}\right) - \frac{\dot{m}_{in}}{S_{in}}X\right]\right) dt \ldots \\
&\quad + \sigma_X(u_t, t, \theta)dw \\
dP &= \left(\frac{\rho_b}{m_b}\left[Y_{xp}\ \alpha\left(\frac{\dot{m}_{NH3}}{M_{NH3}} - q_{s,H^+}\frac{\dot{m}_{in}}{S_{in}}\right) - \frac{\dot{m}_{in}}{S_{in}}P\right]\right) dt \ldots \\
&\quad + \sigma_P(u_t, t, \theta)dw,
\end{aligned}
$$

where $\theta = [\alpha\ Y_{xp}]$ is a vector of unkown parameters, w are standard Wiener processes, and $\sigma_X(\cdot)$ and $\sigma_P(\cdot)$ are nonlinear functions. The outputs are

$$
\begin{aligned}
y_1 &= X + e_x \\
y_2 &= P + e_p,
\end{aligned}
$$

where e_x and e_p are measurement white noise processes $N(0, \sigma_e)$.

The data collected for 10 batches out of 15 are used for estimating the parameters θ with CTSM, using a Maximum Likelihood (ML) method. The remaining batches are used for validation. The results obtained are shown in Table 1, where it is observed that the values of the parameters Y_{xp}, α, initial conditions X_0, P_0, and σ_{xx}, and σ_{e_p} are reliable as indicated by the standard deviation and the t-score analysis. However the results for σ_{pp}, and σ_{e_x} are not good, these parameters are important for learning about the process noise of P, and measurement noise of X this information will be used later for the tuning of the MDA-EKF.

In Figure 2, the Root Mean Square Error (RMSE) for

Table 1. Paramater Estimation Results using CTSM

	Estimate	σ	**t-score**	**p(> |t|)**
X_0	1.87e-2	3.121e-3	6.00	0.00
P_0	5.11e-3	2.34e-3	2.18	0.03
α	1.95e-1	1.35e-3	144.38	0.00
Y_{xp}	16.73e-1	1.41e-2	119.05	0.00
σ_{xx}	2.66e-3	1.37e-4	19.47	0.00
σ_{pp}	6.88e-10	7.42e-9	0.09	0.93
σ_{e_x}	3.03e-16	3.04e-14	0.01	0.99
σ_{e_p}	1.96e-4	1.06e-5	18.41	0.00

the model estimation is shown with black bars. It can be seen that even though the fitting of the model is good for most cases, there are some batches were the RMSE is quite large. This is due to the fact that, in case of a perfect model, the model estimation depends mainly on the initial conditions. Consequently if the initial conditions for some batches change for any reason the estimation of the model will be biased. Therefore an estimator depending upon measurements such as the Kalman filter can be used to resolve this sensitivity problem.

4.3 Biomasss and Product Estimation using MDA-EKF

An important feature of the classical Kalman filter, and in general of the state space observers, is that the estimation is based on input and output process measured data. Consequently, the problem of unknown initial conditions can be solved by recursively adding a correcting factor to the system based on the difference between the estimated and measured outputs.

Hence, to improve the estimation performance of the model, the discrete-time MDA-EKF described in section 3, with sample time $T_s = 0.0208$ time units , is used. Accounting for the information obtained in the previous section (Table 1) about the process and measurement noises, σ_{xx}, σ_{pp}, and σ_{e_x}, σ_{e_p}, respectively, the process and measurement noise covariance matrices are set as follows,

$$
\begin{aligned}
Q &= diag[\sigma_{xx}\sigma_{xx}^{\mathrm{T}}\ \sigma_{pp}\sigma_{pp}^{\mathrm{T}}] \\
R_k &= diag[\sigma_{e_x}\sigma_{e_x}^{\mathrm{T}}\ \sigma_{e_p}\sigma_{e_p}^{\mathrm{T}}],
\end{aligned}
$$

Next, the system (12) can be rewritten as

$$
\begin{aligned}
\frac{dX}{dt} &= \frac{\rho_b}{m_b}\left[\alpha\left(\frac{\dot{m}_{NH3}}{M_{NH3}} - q_{s,H^+}\frac{\dot{m}_{in}}{S_{in}}\right) - \frac{\dot{m}_{in}}{S_{in}}X\right] + \mathrm{w}_X(t) \\
\frac{dX}{dt} &= \frac{\rho_b}{m_b}\left[Y_{xp}\ \alpha\left(\frac{\dot{m}_{NH3}}{M_{NH3}} - q_{s,H^+}\frac{\dot{m}_{in}}{S_{in}}\right) - \frac{\dot{m}_{in}}{S_{in}}P\right] + \ldots \\
&\quad + \mathrm{w}_P(t),
\end{aligned}
$$

where $\mathrm{w}_X(t)$ and $\mathrm{w}_P(t)$ are Gaussian random processes with $N(0, \sigma_{\mathrm{w}_X} = \sigma_{xx})$ and $N(0, \sigma_{\mathrm{w}_P} = \sigma_{pp})$, respectively. Note that the system has to be discretized in time in order to apply MDA-EKF.

Figure 2 shows the RMSE for MDA-EKF estimation with white bars. Generally speaking, the results show that the gain from the application of the discrete-time MDA-EKF is limited. This is caused by a) the outputs measured data have long delays,i.e. one unit of time for X and two for P, hence the estimation correction starts only after the first output measurements are made available, and b) unknown variations in the inputs and model parameters.

To facilitate accommodation to biased initial conditions and other process disturbances it is interesting to investigate whether some parameters may vary in time. This is investigated for nutrient concentration in the feed flow S_{in}, and the yield coefficient of product on biomass Y_{xp}. Thus, an augmented system is proposed to include this dynamics into the model. This can be incorporated by assuming Y_{xp} and S_{in} as random walk processes, with the initial conditions given in Table 1. Thus,

$$
\begin{aligned}
\frac{dS_{in}}{dt} &= \mathrm{w}_{S_{in}}(t) \\
\frac{dY_{xp}}{dt} &= \mathrm{w}_{Y_{xp}}(t),
\end{aligned}
$$

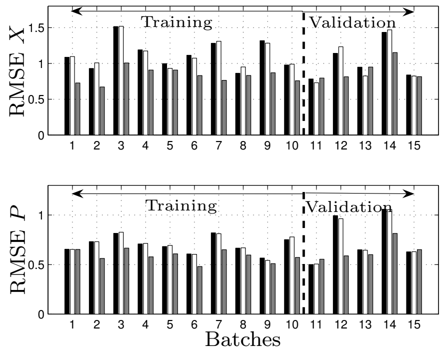

Fig. 2. RMSE Comparison Between Model and MDA-EKF. The black bars are the RMSE for model estimation, the white ones the RMSE for MDA-EKF estimation, and the gray ones the RMSE for MDA-EKF estimation with augmented system.

where $\mathrm{w}_{S_{in}}(t)$ and $\mathrm{w}_{Y_{xp}}(t)$ are Gaussian random processes with $N(0, \sigma_{\mathrm{w}})$.

To tune the MDA-EKF covariance matrices for this augmented system, the values in Table 1 were taken as initial conditions. Then by trial and error these were refined to obtain better estimation results.

Figure 2 shows with gray bars the RMSE for MDA-EKF estimation with the augmented system. Clearly a reduction of the RMSE is obtained for most of the batches. It is noteworthy that when the RMSE is relatively large the benefit obtained by estimating Y_{xp} and S_{in} is rather important. Furthermore, Figure 3 shows two examples of the model performance and MDA-EKF estimation. In the first case for batch 11, the estimation is reasonably good, and just a slight retuning of Y_{xp} and S_{in} is needed. In the second case for batch 14, the estimation is performing poorly during the fed-batch part, but after assimilating some delayed measurements into the model, the estimator starts reducing the estimation error. In addition in the bottom of the figure it is shown how the parameters Y_{xp} and S_{in} are retuned helping to improve the performance of the filter. This case demonstrates the adaptivity of the MDA-EKF estimator with the augmented system when the process is operating under conditions which deviate from standard operation.

5. CONCLUSIONS

In this paper a new estimator is developed to handle infrequent and delayed process measurements where Multirate Data Assimilation is performed using the Extended Kalman Filter (MDA-EKF). This estimator is capable of assimilating infrequent and delayed output measurements into the model by minimizing, in a suboptimal way, the H_2 norm of the estimation error. An industrial case study illustrates the model development and the tuning the estimator. A first principle based model structure was chosen where the parameters were estimated with data obtained from the process by means of Continuous Time Stochastic Modelling [Kristensen , 2004]. The results obtained showed a good fitting of the model with the validation data, however in some cases the estimation error was large

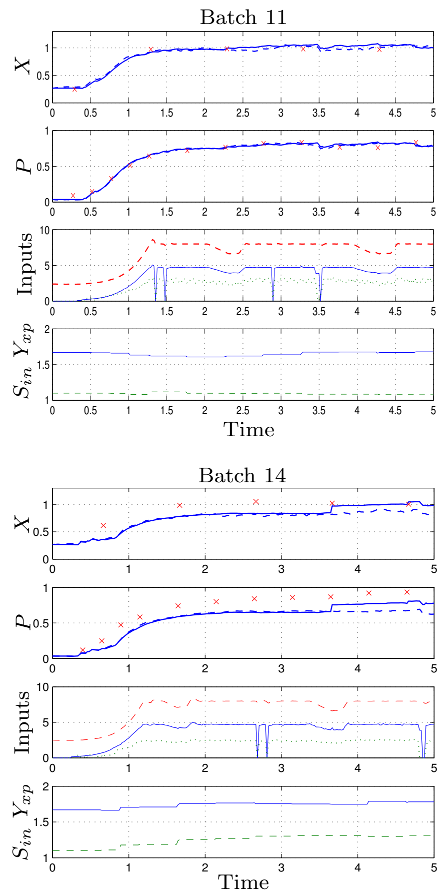

Fig. 3. Time sequence comparison Between Model and MDA-EKF Estimation for batch 11 (top four subfigures) and bacth 14 (bottom four subfigures). Within each of the two sets of four subfigures in the first and second subfigures the crosses represent the real measurements, the solid line the MDA-EKF estimation, and the dashed line the model estimation. In the third figure, the solid line is $\dot{m}_{in}$, the dashed line m_b, and the dotted line $\dot{m}_{NH_3}$. Finally in the fourth figure, the solid line is Y_{xp}, and the dashed line S_{in}. Notice that the data has been normalized.

presumably due to unknown disturbances. In order to improve the on-line performance of the model, a multirate data assimilating Kalman filter estimator was proposed, the MDA-EKF.

The first results obtained with the MDA-EKF estimator did not improve the performance of the filter very much, showing that the large estimation error in those batches was caused by variations in the model parameters as well as model inputs more than in the initial conditions. Therefore, after analyzing the model parameters which potentially could vary in time, the yield coefficient of product on biomass Y_{xp} and the nutrients concentration in the feed flow S_{in} were incorporated as augmented state variables.

With the augmented system, the MDA-EKF estimator significantly reduced the estimation error in the batches which previously showed poor fit. The resulting fit is very similar to batches with good fit, confirming the above hypothesis. Investigating the behavior of the augmented state variables, Y_{xp} and S_{in}, it was found that for batches with bad fit, these parameters were strongly retuned while for the ones with good fit only slight retuning occurred. Subsequently it was revealed by further analysis that a specific sensor in the batches with poor fit needed recalibration. In conclusion, the MDA-EKF has been developed into an estimator which can handle infrequent and delayed output process measurements and adapt to variations in model and input parameters.

6. ACKNOWLEDGEMENTS

The first author has been financially supported with a grant from the European Commission ALFA office, within the Project II-0407-FA, "Latin American BioProcess Control".

The authors are grateful to Novo Nordisk A/S for permission to use their data for this work.

REFERENCES

B.D.O. Anderson, and J.B. Moore. Optimal Filtering. *Prentice-Hall*, 1979.

O. Barrero, B.L.R. De Moor, and D.S. Bernstein. Data Assimilation for Magneto-Hydrodynamics Systems. *Journal of Computational and Appied Mathematics*, volume 189, pages 242–259, 2006.

O. Barrero, and S.B. Jorgensen. Multirate Process Data Assimilation using the Extended Kalman Filter. *Internal Report, Department of Chemical Engineering - CAPEC, Technical University of Denmark*, 2007.

S. Cohn, and R. Todling. Approximate Data Assimilation Schemes for Stable and Unstable Dynamics. *Meteorological Society of Japan*, volume 74, pages 63–75, 1996.

G. Evensen. Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using M,onte C,arlo Methods to Forecast Error Statistics. *Journal of Geophysical Research*, volume 99(C5), pages 10,143–10,162, 1994.

G. Evensen. The Ensemble K,alman Filter: Theoretical Formulation and Practical Implementation. *Ocean Dynamics*, volume 53, pages 343–367, 2003,

L. Fortuna, A. Rizzo, S. Graziani, and M.G. Xibilia. Soft Sensors for Monitoring and Control of Industrial Processes Series: Advances in Industrial Control. *Springer*, 2005.

G.F. Franklin, J.D. Powell, and A. Emami-Naeni. Feedback Control of Dynamic Systems. *Prentice Hall*, edition 4th, 2002.

S.M. Grewal, and A.P. Andrews. Kalman Filtering: Theory and Practice Using Matlab. *Wiley Inter-Science*, edition 2nd, 2001.

R.D. Gudi, and S.L. Shah, and M.R. Gray. Multirate State and Parameter Estimation in an Antibiotic Fermentation with Delayed Measurements. *Biotechnology and Bioengineering*, volume 44, pages 2451–1464, 1994.

R.D. Gudi, S.L. Shah, and M.R. Gray. Adaptive Multirate State and Parameter Estimation Strategies with Application to a Bioreactor. *AIChE*, volume 41, pages 1271–1278, 1995.

N.R. Kristensen, H. Madsen, and S.B. Jorgensen. Parameter Estimation in Stochastic Grey-Box Models. *Automatica*, volume 40, pages 225–237, 2004.

J.H. Lee, M. Morari. Robust Inferential Control of Multi-Rate Sampled-Data Systems. *Chemical Engineering Science*, volume 47, pages 865–885, 1992.

F. Lei. Dynamics and non-linear phenomena in continuous cultivations of S,acharomyces C,erevisiae. *Ph.D. Thesis, Department of Chemical Engineering, Technical University of Denmark*, 2001.

B. Lin, B. Recke, J.K.H Knudsen, and S.B. Jorgensen. A Systematic Approach for Soft Sensor Development. *Computers & Chemical Engineering*, In Press, 2006.

X. Lu, H. Zhang, and W. Wang. Optimal Filtering for Discrete Time-Varying Systems with Multiple Time-Delay Measurements. *Proceedings of International Conference on Control, Automation, Robotics and Vision, Kunming, China*, pages 1735-1740, 2004.

M.A. Myers, S. Kang, and R.H. Luecke. State Estimation and Control for Systems with Delayed Off-Line Measurements. *Computers & Chemical Engineering*, volume 20(5), pages 585–588, 1996.

J. Nielsen, J. Villadsen, and G. Liden. Bioreactor Engineering Principle. *Kluwer Academic/Plenum*, edition 2nd, 2003.

C. Penland. A Stochastic Approach to Nonlinear Dynamics: a review. *Bulletin of the American Meteorological Society*, volume 84, pages 921–925, 2003.

M. Thaysen. Hybrid Modeling of Enhanced Bioreactor Performance. *Ph.D. Thesis, Department of Chemical Engineering, Technical University of Denmark*, 2005.

E.A. Wan, and R. Van der Merwe. Kalman Filtering and Neural Networks. *Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control*, 2001.

H. Zhang, M.V. Basin, and M. Skliar. Kalman Filter for Continuous State-Space System with Continuous, Multirate, Randomly Sampled and Delayed Measurements. *Proceedings of American Control Conference, Portland, OR, USA*, pages 3462–3467, 2005.