IDENTIFICATION OF STATE-SPACE MODELS FOR PROCESSES WITH IRREGULARLY SAMPLED OUTPUTS

Harigopal Raghavan [*,1] **Arun K. Tangirala** [*,2]
Sirish L. Shah [*,3]

[*] *Department of Chemical & Materials Engineering, University of Alberta, Edmonton, AB, Canada.*

Abstract: In many processes, variables which indicate product quality are irregularly sampled. Often, the inter-sample behavior of these quality variables can be inferred from manipulated variables (MV) and other process variables which are measured frequently. When the quality variables are irregularly sampled, Maximum Likelihood Estimation (MLE) can be performed using the Expectation Maximization (EM) approach. The initial model required for the EM algorithm can be obtained using a realization-based subspace identification technique. We describe such a model identification method and present its application on simulation and industrial case studies.

Keywords: Irregular output sampling, Maximum Likelihood Estimation, Expectation maximization.

1. INTRODUCTION

In many applications outputs, such as composition, molecular weight or a product quality variable, may not be available as frequently as would be desired for satisfactory closed-loop control. The relationship used to predict quality variables from other process variables is called an inferential sensor. For a number of such applications, it might be possible to identify dynamic models and design state estimators.

Traditional system identification techniques for sampled-data systems with uniformly spaced sampling intervals include MLE, the closely related Prediction Error Methods (PEM), Instrumental Variable Techniques (Ljung, 1999) and Subspace Identification (VanOverschee and DeMoor, 1996). MLE has been popular because of its theoretical optimality properties and well researched practical issues such as variance and bias distributions. However, MLE suffers from a number of practical problems. In general, MLE may involve solving a non-convex optimization problem. This is generally tackled using a gradient-based iterative search strategy, which can suffer from serious numerical issues when a canonical state-space parameterization is chosen (Deistler, 2000). Subspace Identification avoids these parameterization problems and identifies models in an arbitrary state-space basis (VanOverschee and DeMoor, 1996). However many statistical properties of these techniques have not yet been established.

In many processes, we can distinguish three classes of measurements; inputs manipulated at a fast-rate (*e.g.,* control-valves), outputs measured at a fast-rate (*e.g.,* temperatures) and outputs measured at a slow rate (*e.g.,* compositions). Pro-

[1] Currently with Honeywell Technology Solutions Ltd., Bangalore, India.
[2] Currently with Dept. of Chemical Engineering, Indian Institute of Technology Madras, Chennai, India.
[3] Corresponding Author. Email: sirish.shah@ualberta.ca

cesses with differing sample rates are known as *multirate* processes and model identification for such processes is of great practical interest to the process industry. We call the fastest sample rate the *base* sample rate and unavailable data points in the slow measurements, *missing* data.

The problem of identifying optimal models with some irregularly sampled variables has been studied using the EM approach (Dempster *et al.*, 1977; Shumway and Stoffer, 1982; Ninness and Gibson, 2002). The traditional gradient-based MLE algorithm has been implemented by Isaksson (1993) with the observed data likelihood function calculated using a modified form of the Kalman filter (Ansley and Kohn, 1983) to account for missing observations and it is shown that the presence of missing-data can aggravate the numerical problems faced by these techniques resulting in slow convergence. While the work in Isaksson (1993) is restricted to ARX models, we present the missing-data identification problem for state-space models. In Ninness and Gibson (2002), state-space identification using the EM approach has been addressed, but the missing-data case has not been considered. Shumway and Stoffer (1982, 2000) and Tanaka and Katayama (1990) use the EM algorithm for state-space identification for time series with missing-data. However, they assume that the state-output transformation matrix is known and give no guidelines on choosing the parameters for the initial model which is an important step in the MLE procedure.

In this paper, we perform state-space model identification from data which has irregularly sampled outputs. We use a realization-based subspace identification technique to obtain an initial model (Kung, 1978) which is then used in the EM algorithm. Expressions for the state-space matrices obtained in the Maximization step, in the presence of irregular observations, are provided. The rest of this paper is organized as follows. In Sec. 2, we present the problem description. We present the gradient-based MLE procedure in Sec. 3.1 and the EM-based state-space model identification procedure in Sec. 3.2. The initial model identification procedure is described in Sec. 3.3. Expressions for the model parameters at the end of each iteration and the modifications in the case of irregular output measurements are provided in Sec. 3.4 and illustrative case studies are provided in Sec. 4.

2. PROBLEM DESCRIPTION

2.1 Model Structure and Assumptions

Let us assume that the process is represented by the following discrete-time state-space model:

$$\begin{aligned}\mathbf{x}_t &= \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_{t-1} + \mathbf{w}_t \\ \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t + \mathbf{v}_t \end{aligned} \quad (1)$$

$\mathbf{x}_t \in R^n$ is the state vector, $\mathbf{u}_t \in R^m$ is the manipulated input, $\mathbf{y}_t \in R^p$ is the output, and $\mathbf{w}_t \in R^n$ and $\mathbf{v}_t \in R^p$ are the state and measurement noise vectors. Assume that $\mathbf{x}_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{R}_w), \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{R}_v)$. For simplicity, $E\left(\mathbf{v}_t\mathbf{w}_t^T\right) = \mathbf{0}$, $E\left(\mathbf{x}_0\mathbf{w}_t^T\right) = \mathbf{0}$ and $E\left(\mathbf{x}_0\mathbf{v}_t^T\right) = \mathbf{0}$. The outputs are classified into fast-sampled outputs ($\mathbf{y_1} \in R^{p_1}$) and slow-sampled outputs ($\mathbf{y_2} \in R^{p_2}$). $\mathbf{y}_t = [\mathbf{y}_{1,t}^T\ \mathbf{y}_{2,t}^T]^T, \mathbf{C} = [\mathbf{C}_1^T\ \mathbf{C}_2^T]^T$ and $\mathbf{v}_t = [\mathbf{v}_{1,t}^T\ \mathbf{v}_{2,t}^T]^T$. We assume that the fast-sample interval (T_{s_1}) is fixed, the slow-sample interval (T_{s_2}) is variable and the slow-sample interval is an integral multiple of the fast-sample interval. We assume that, $E\left(\mathbf{v}_{1,t}\mathbf{v}_{2,t}^T\right) = \mathbf{0}$. We define, $\mathbf{Y}_s \equiv \{\mathbf{y}_1, \ldots, \mathbf{y}_s\}$, $\mathbf{U}_s \equiv \{\mathbf{u}_1, \ldots, \mathbf{u}_s\}$ and $\mathbf{Z}_s \equiv \{\mathbf{Y}_s, \mathbf{U}_s\}$. We define the conditional expectations, $\mathbf{x}_t^s = E(\mathbf{x}_t|\mathbf{Z}_s)$ and $\mathbf{P}_{t_1,t_2}^s = E((\mathbf{x}_{t_1} - \mathbf{x}_{t_1}^s)(\mathbf{x}_{t_2} - \mathbf{x}_{t_2}^s)^T|\mathbf{Z}_s)$. For convenience, $\mathbf{P}_{t,t}^s$ is written as $\mathbf{P}_t^s$. We are interested in optimally estimating the model parameters $\hat{\boldsymbol{\Theta}} \equiv (\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \mathbf{A}, \mathbf{B}, \mathbf{C}_1, \mathbf{C}_2, \mathbf{R}_w, \mathbf{R}_{v1}, \mathbf{R}_{v2})$ using all the available samples in the identification data set. The data comprises N samples of $\mathbf{u}$ and $\mathbf{y}_1$ and N_o unequally spaced samples of $\mathbf{y}_2$, $N_o \leq N$.

3. EM-BASED STATE SPACE MODEL IDENTIFICATION

MLE-based algorithms are usually implemented using gradient-based numerical optimization techniques. EM was developed for MLE from data sets with missing observations. It can also be used for state-space model identification. We briefly summarize these techniques.

3.1 Maximization of the likelihood function

The parameters in Eq. 1 can be estimated by maximizing the *likelihood* function of the observed data ($\mathbf{Z}_N \equiv \{\mathbf{Y}_N, \mathbf{U}_N\}$), which is written in terms of *one-step-ahead* prediction errors:

$$\begin{aligned}\boldsymbol{\epsilon}_t &\equiv \mathbf{y}_t - E\left(\mathbf{y}_t|\mathbf{Z}_{t-1}\right) = \mathbf{y}_t - \mathbf{C}\mathbf{x}_t^{t-1} \\ \boldsymbol{\epsilon}_t &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_t),\ \boldsymbol{\Sigma}_t = \mathbf{C}\mathbf{P}_t^{t-1}\mathbf{C}^T + \mathbf{R}_v\end{aligned} \quad (2)$$

The innovations form of the likelihood function (Schweppe, 1965) can be written as:

$$L(\boldsymbol{\Theta}|\mathbf{Z}_N) = \prod_{t=1}^{N} \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}_t|^{\frac{1}{2}}} \exp\left(\frac{-\boldsymbol{\epsilon}_t^T\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\epsilon}_t}{2}\right) \quad (3)$$

$|\boldsymbol{\Sigma}_t|$ refers to the determinant of the matrix $\boldsymbol{\Sigma}_t$. The quantities $\mathbf{x}_t^{t-1}$ and $\mathbf{P}_t^{t-1}$ are calculated using the Kalman filter recursions. These, and the Kalman smoother recursions required for the

EM algorithm can be derived along the lines of Shumway and Stoffer (2000) and are not provided here due to space constraints. Due to the monotonicity of the *log* function, MLE can be performed by minimizing the *negative log-likelihood* function. Due to the non-convexity of this problem, numerical techniques are used (Gupta and Mehra, 1974). In order to account for missing data, the observed data likelihood function can be calculated using a modified Kalman filter (Ansley and Kohn, 1983). However, these techniques suffer from a number of problems. The likelihood function need not increase from iteration to iteration (Gupta and Mehra, 1974). On the other hand, in the EM algorithm the likelihood always increases and convergence to a stationary point is guaranteed (Wu, 1983). Isaksson (1993) shows some surprising results through simulations which indicate that the EM can be much faster than the gradient-based techniques when there is a significant amount of missing data.

3.2 EM algorithm

The EM algorithm is a simple and efficient alternative for MLE from incomplete data records. Denote the observed data set as $\mathbf{Z}_N$ and the unobserved data set as $\mathbf{X}_N$. Assume that the *complete* data set consists of both $\mathbf{Z}_N$ and $\mathbf{X}_N$. The distribution of the complete data can be factored as,

$$f_{zx}(\mathbf{Z}_N, \mathbf{X}_N|\mathbf{\Theta}) = f_z(\mathbf{Z}_N|\mathbf{\Theta}) f_x(\mathbf{X}_N|\mathbf{Z}_N, \mathbf{\Theta}) \quad (4)$$

Hence, the log-likelihood can be decomposed as,

$$\ell(\mathbf{\Theta}|\mathbf{Z}_N, \mathbf{X}_N) = \ell(\mathbf{\Theta}|\mathbf{Z}_N) + \log f_x(\mathbf{X}_N|\mathbf{Z}_N, \mathbf{\Theta}) \quad (5)$$

This can be rewritten as,

$$\ell(\mathbf{\Theta}|\mathbf{Z}_N) = \ell(\mathbf{\Theta}|\mathbf{Z}_N, \mathbf{X}_N) - \log f_x(\mathbf{X}_N|\mathbf{Z}_N, \mathbf{\Theta}) \quad (6)$$

Consider the expected value of Eq. 6 conditioned on the observed data $\mathbf{Z}_N$, and an estimate of the parameters, $\mathbf{\Theta}'$.

$$\ell(\mathbf{\Theta}|\mathbf{Z}_N) = Q(\mathbf{\Theta}|\mathbf{\Theta}') - H(\mathbf{\Theta}|\mathbf{\Theta}') \quad (7)$$

where,

$$\begin{aligned} Q(\mathbf{\Theta}|\mathbf{\Theta}') &\equiv E\left(\log f_{zx}(\mathbf{Z}_N, \mathbf{X}_N|\mathbf{\Theta})|\mathbf{Z}_N, \mathbf{\Theta}'\right) \\ H(\mathbf{\Theta}|\mathbf{\Theta}') &\equiv E\left(\log f_x(\mathbf{X}_N|\mathbf{Z}_N, \mathbf{\Theta})|\mathbf{Z}_N, \mathbf{\Theta}'\right) \end{aligned} \quad (8)$$

Theorem 1. The EM algorithm increases $\ell(\mathbf{\Theta}|\mathbf{Z}_N)$ at each iteration, *i.e.,*

$$\ell(\mathbf{\Theta}^k|\mathbf{Z}_N) \geq \ell(\mathbf{\Theta}^{k-1}|\mathbf{Z}_N) \quad (9)$$

Proof: This theorem is a key result of Dempster *et al.* (1977). Consider a sequence of iterates, $\ell(\mathbf{\Theta}^0), \ell(\mathbf{\Theta}^1), \ldots$, where $\ell(\mathbf{\Theta}^k) = M(\ell(\mathbf{\Theta}^{k-1}))$ for some function $M(\ )$. The difference in the values of $\ell(\mathbf{\Theta}|\mathbf{Z}_N)$ at successive iterates is,

$$\begin{aligned} &\ell(\mathbf{\Theta}^k|\mathbf{Z}_N) - \ \ell(\mathbf{\Theta}^{k-1}|\mathbf{Z}_N) \\ &\quad = [Q(\mathbf{\Theta}^k|\mathbf{\Theta}^{k-1}) - Q(\mathbf{\Theta}^{k-1}|\mathbf{\Theta}^{k-1})] \\ &\quad -[H(\mathbf{\Theta}^k|\mathbf{\Theta}^{k-1}) - H(\mathbf{\Theta}^{k-1}|\mathbf{\Theta}^{k-1})] \end{aligned} \quad (10)$$

The EM algorithm chooses $\mathbf{\Theta}^k$ to maximize $Q(\mathbf{\Theta}|\mathbf{\Theta}^{k-1})$. Hence, the first part of the RHS of Eq. 10 is positive. For the second part, we have the following result which is established using Jensen's inequality (Rao, 2001).

$$H(\mathbf{\Theta}|\mathbf{\Theta}^{k-1}) \leq H(\mathbf{\Theta}^{k-1}|\mathbf{\Theta}^{k-1}) \quad (11)$$

■

In effect, the EM procedure reduces to finding the expected value of the Q-function at each iteration conditioned on all the available data and the estimated parameters from the previous iteration, followed by a multivariate regression. From an application point-of-view, the convergence can be monitored by calculating the negative-log-likelihood (NLL) function.

The EM algorithm can be summarized as follows: Start with an initial estimate of the parameter vector, $\mathbf{\Theta}^0$ and carry out the following steps at each iteration, k, until convergence:

- **Expectation (E-step)**: Find the expected value of the complete data log likelihood function(Q-function) given the observed data set, $\mathbf{Z}_N$ and the previously estimated parameter vector, $\mathbf{\Theta}^{k-1}$. This conditional expectation is obtained using Kalman smoothers.
- **Maximization (M-step)**: Maximize the Q-function with respect to the parameter vector. Expressions for calculating these parameters are provided in Sec. 3.4.

The above steps ensure that the NLL function decreases at every iteration. Therefore, the EM algorithm is guaranteed to converge to a local minimum of the likelihood function.

3.3 Identification of the initial model

We now present a realization-based subspace identification technique (Kung, 1978) which can be used to identify the initial model. Let us represent the process described in Eq. 1 as a multivariate FIR model:

$$y_{t,a} = \sum_{b=0}^{m} \sum_{c=0}^{2s-1} h_{abc} u_{t-c,b} + \nu_{t,a}, \ \ \forall\, a = 1, \ldots, p \quad (12)$$

where, ν represents an arbitrary noise process. The impulse response coefficients in Eq. 12 can be estimated even in the presence of some missing observations. Following the estimation of the

impulse response coefficients $\hat{h}_{abc}$ we can form a set of matrices of the impulse response coefficients, $\hat{H}_r \in R^{p\times m}$ corresponding to the lags $r = 0, \ldots, 2s-1$. Using the matrices $\hat{H}_r$ we can form a Hankel matrix $\hat{\mathbf{H}}$ as shown below:

$$\hat{\mathbf{H}} = \begin{bmatrix} \hat{H}_1 & \hat{H}_2 & \cdots & \hat{H}_s \\ \hat{H}_2 & \hat{H}_3 & \cdots & \hat{H}_{s+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{H}_s & \hat{H}_{s+1} & \cdots & \hat{H}_{2s-1} \end{bmatrix} \in R^{ps\times ms} \quad (13)$$

The matrix $\mathbf{H}$ can be factorized as, $\mathbf{H} = \boldsymbol{\Gamma}_s\boldsymbol{\Omega}_s$, where $\boldsymbol{\Gamma}_s$ is the extended observability matrix and $\boldsymbol{\Omega}_s$ is the extended controllability matrix. Estimates of these matrices can be obtained by performing a singular value decomposition of $\hat{\mathbf{H}}$.

$$\hat{\mathbf{H}} = \hat{\mathbf{Q}}\hat{\mathbf{S}}\hat{\mathbf{V}}^T = \begin{bmatrix} \hat{\mathbf{Q}}_s & \hat{\mathbf{Q}}_n \end{bmatrix} \begin{bmatrix} \hat{\mathbf{S}}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{V}}_s^T \\ \hat{\mathbf{V}}_n^T \end{bmatrix}$$

$$\hat{\boldsymbol{\Gamma}}_s = \hat{\mathbf{Q}}_s\hat{\mathbf{S}}_s^{\frac{1}{2}}, \quad \hat{\boldsymbol{\Omega}}_s = \hat{\mathbf{S}}_s^{\frac{1}{2}}\hat{\mathbf{V}}_s^T.$$

We can choose the appropriate system order by looking at a plot of the singular values. The $\hat{\mathbf{B}}$ and $\hat{\mathbf{C}}$ matrices can be read out from the first block column of $\hat{\boldsymbol{\Omega}}_s$ and the first block row of $\hat{\boldsymbol{\Gamma}}_s$ respectively. $\hat{\mathbf{A}}$ can be estimated from the shift invariant structure of either $\hat{\boldsymbol{\Gamma}}_s$ or $\hat{\boldsymbol{\Omega}}_s$ (VanOverschee and DeMoor, 1996).

Remark: Using an FIR model structure for the initial model restricts it to the output-error class of models. Hence the identification result yields only the deterministic sub-system. In order to include stochastic dynamics, it is necessary to fit a pre-whitening filter to the irregularly spaced residuals using the EM algorithm.

3.4 Expressions for new model parameters in the M-step

The new model parameters at the end of each iteration are obtained by solving the optimization problem in the M-step, which can be written as:

$$\boldsymbol{\Theta}^k = \arg\min_{\boldsymbol{\Theta}}\{-2Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{k-1})\} =$$
$$\arg\min_{\boldsymbol{\Theta}} [\log|\boldsymbol{\Sigma}_0| + N\log|\mathbf{R}_w| + N\log|\mathbf{R}_v| +$$
$$tr\{\boldsymbol{\Sigma}_0^{-1}\left(\mathbf{P}_0^N + (\mathbf{x}_0^N - \boldsymbol{\mu}_0)(\mathbf{x}_0^N - \boldsymbol{\mu}_0)^T\right)\} +$$
$$tr\{\mathbf{R}_w^{-1}(\beta_1 - \beta_2[\mathbf{A}|\mathbf{B}]^T - [\mathbf{A}|\mathbf{B}]\beta_2^T + [\mathbf{A}|\mathbf{B}]\beta_3[\mathbf{A}|\mathbf{B}]^T)\}$$
$$+tr\{\mathbf{R}_v^{-1}\left(\beta_4 - \beta_5\mathbf{C}^T - \mathbf{C}\beta_5^T + \mathbf{C}\beta_1\mathbf{C}^T\right)\}]$$

β_1 to β_5 are functions of the observed data and smoothed estimates. $\beta_1 = \sum_{t=1}^N(\mathbf{P}_t^N + \mathbf{x}_t^N(\mathbf{x}_t^N)^T)$

$$\beta_2 = \begin{bmatrix} \sum_{t=1}^N(\mathbf{P}_{t,t-1}^N + \mathbf{x}_t^N(\mathbf{x}_{t-1}^N)^T) & \sum_{t=1}^N(\mathbf{x}_t^N\mathbf{u}_{t-1}^T) \end{bmatrix}$$

$$\beta_3 = \begin{bmatrix} \sum_{t=1}^N(\mathbf{P}_{t-1}^N + \mathbf{x}_{t-1}^N(\mathbf{x}_{t-1}^N)^T) & \sum_{t=1}^N(\mathbf{x}_{t-1}^N\mathbf{u}_{t-1}^T) \\ \sum_{t=1}^N(\mathbf{x}_{t-1}^N\mathbf{u}_{t-1}^T)^T & \sum_{t=1}^N\mathbf{u}_{t-1}\mathbf{u}_{t-1}^T \end{bmatrix}$$

$\beta_4 = \sum_{t=1}^N(\mathbf{y}_t\mathbf{y}_t^T) \quad \beta_5 = \sum_{t=1}^N(\mathbf{y}_t(\mathbf{x}_t^N)^T)$.

The solution we obtain by setting the first derivatives of $-2Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{k-1})$ to zero is given by:

$\boldsymbol{\mu}_0 = \mathbf{x}_0^N, \boldsymbol{\Sigma}_0 = \mathbf{P}_0^N, [\mathbf{A} \mid \mathbf{B}] = \beta_2\beta_3^{-1}$,
$\mathbf{C} = \beta_5\beta_1^{-1}, \mathbf{R}_w = \frac{1}{N}(\beta_1 - \beta_2\beta_3^{-1}\beta_2^T)$,
$\mathbf{R}_v = \frac{1}{N}(\beta_4 - \beta_5\beta_1^{-1}\beta_5^T)$.

Expressions for the missing-data case: Instead of the standard Kalman filter expressions, the E-step involves the Kalman smoother which uses the missing-data Kalman filter (Ansley and Kohn, 1983). Also, the expressions for estimating the $\mathbf{C}$ and $\mathbf{R}_v$ matrices in the M-step change. Assume that there are N_m instances in which $\mathbf{y}_2$ is not measured and $N_o = N - N_m$ observations in which $\mathbf{y}_2$ is measured. The term involving the $\mathbf{C}$ and $\mathbf{R}_v$ matrices can be written as:

$E\left((\mathbf{y}_t - \mathbf{C}\mathbf{x}_t)^T\mathbf{R}_v^{-1}(\mathbf{y}_t - \mathbf{C}\mathbf{x}_t)|\{\mathbf{Y}_N^{obs}, \mathbf{U}_N\}, \boldsymbol{\Theta}^{k-1}\right)$
$= N\log|\mathbf{R}_v| + tr\{\mathbf{R}_v^{-1}(\boldsymbol{\Delta}_1 - \mathbf{C}\boldsymbol{\Delta}_2 - \boldsymbol{\Delta}_2^T\mathbf{C}^T + \mathbf{C}\boldsymbol{\Delta}_3\mathbf{C}^T)\}$, where $\mathbf{y}_t^{obs} = \mathbf{D}_t\mathbf{y}_t$, $\mathbf{D}_t = \mathbf{I}_p$ if $\mathbf{y}_{2,t}$ is measured and $\mathbf{D}_t = [\mathbf{I}_{p_1} \quad \mathbf{0}]$ if $\mathbf{y}_{2,t}$ is missing where, $\mathbf{I}_p$ and $\mathbf{I}_{p_1}$ refer to the $p \times p$ and $p_1 \times p_1$ identity matrices respectively and,

$$\boldsymbol{\Delta}_1 = \begin{bmatrix} \sum_{t=1}^N\mathbf{y}_{1,t}\mathbf{y}_{1,t}^T & \sum_{N_o}\mathbf{y}_{1,t}\mathbf{y}_{2,t}^T \\ \sum_{N_o}\mathbf{y}_{2,t}\mathbf{y}_{1,t}^T & \sum_{N_o}\mathbf{y}_{2,t}\mathbf{y}_{2,t}^T + N_m\mathbf{R}_{v2}^{k-1} \end{bmatrix}$$

$$\boldsymbol{\Delta}_2 = \begin{bmatrix} \sum_{t=1}^N\mathbf{x}_t^N\mathbf{y}_{1,t}^T & \sum_{N_o}\mathbf{x}_t^N\mathbf{y}_{2,t}^T \end{bmatrix}$$

$$\boldsymbol{\Delta}_3 = \begin{bmatrix} \sum_{t=1}^N\mathbf{P}_t^N + \mathbf{x}_t^N(\mathbf{x}_t^N)^T & \sum_{N_o}\mathbf{P}_t^N + \mathbf{x}_t^N(\mathbf{x}_t^N)^T \\ \sum_{N_o}\mathbf{P}_t^N + \mathbf{x}_t^N(\mathbf{x}_t^N)^T & \sum_{N_o}\mathbf{P}_t^N + \mathbf{x}_t^N(\mathbf{x}_t^N)^T \end{bmatrix},$$

where $\mathbf{R}_{v2}^{k-1}$ refers to the value of $\mathbf{R}_{v2}$ at the previous iteration. Notice that some of these terms involve summations over N_o observations. The expressions for the $\mathbf{C}$ and $\mathbf{R}_v$ matrices become, $\mathbf{C} = \boldsymbol{\Delta}_2^T\boldsymbol{\Delta}_3^{-1}$, $\mathbf{R}_v = \frac{1}{N}(\boldsymbol{\Delta}_1 - \boldsymbol{\Delta}_2^T\boldsymbol{\Delta}_3^{-1}\boldsymbol{\Delta}_2)$.

4. ILLUSTRATIVE APPLICATIONS

4.1 Simulated Case-study: 3^{rd} order underdamped system

In this example, we use the proposed EM-based strategy for identifying an underdamped system in the presence of missing data. The simulated system is defined by the state-space matrices:

$$A = \begin{bmatrix} 0.3688 & 0.4767 & 0.0114 \\ -0.5976 & 0.6095 & -0.5408 \\ -0.0156 & -0.0686 & 0.0422 \end{bmatrix} \quad B = \begin{bmatrix} 0.34 \\ 0.56 \\ 0.78 \end{bmatrix}$$

$$C = \begin{bmatrix} 0.5 & 0.3 & 0.1 \\ 1.2 & 0.96 & 1.5 \end{bmatrix} \quad R_v = \begin{bmatrix} 0.0398 & 0 \\ 0 & 0.0398 \end{bmatrix}$$

$$R_w = \begin{bmatrix} 0.0407 & 0.0001 & 0.0015 \\ 0.0001 & 0.0407 & -0.0020 \\ 0.0015 & -0.0020 & 0.0428 \end{bmatrix}$$

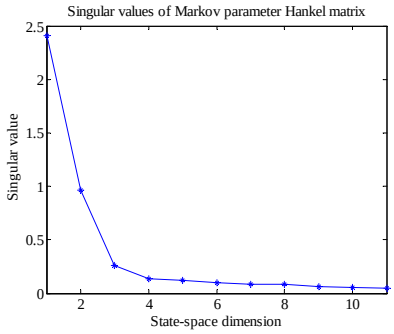

Fig. 1. Singular value plot for order selection

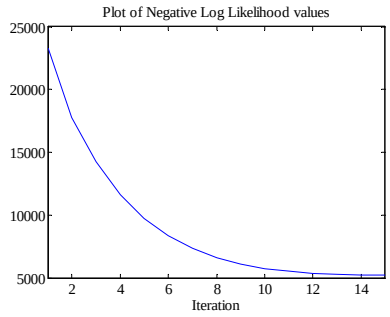

Fig. 2. Monitoring convergence using the Negative log likelihood function

The system is excited using $N = 5000$ samples of a random binary input signal. To simulate irregular sampling, samples of y_2 at instants which are multiples of 3 and 5 were dropped. Hence 2333 samples out of the 5000 samples of y_2 were dropped. The FIR model was identified with 20 FIR coefficients. The singular value plot was used for selecting the order of the initial model (Fig. 1). Following this, the EM-algorithm was applied and the NLL function was displayed for monitoring convergence (Fig. 2). A comparison of the step responses of the true system and the identified model (Fig. 3) shows that the identified model is close to the true system. In addition, a plot showing a small portion of the true output values, measured observations and the Kalman filter predictions is given in Fig. 4.

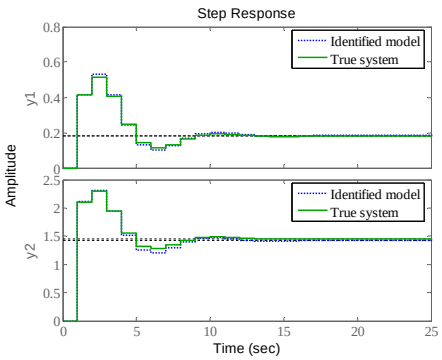

Fig. 3. Comparison of step responses of the underdamped system

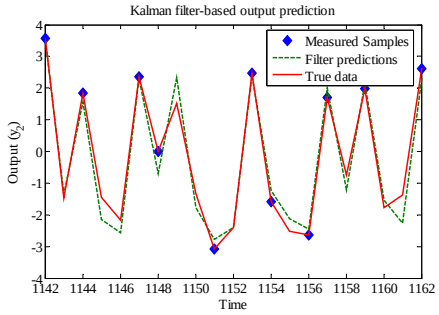

Fig. 4. Comparison of Kalman filter predictions with true output in the underdamped system

4.2 Industrial Case-study: Bleaching unit in a BCTMP mill

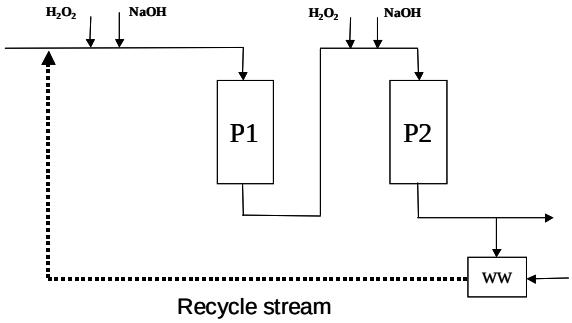

Fig. 5. Simplified flow sheet of Mechanical Pulp mill

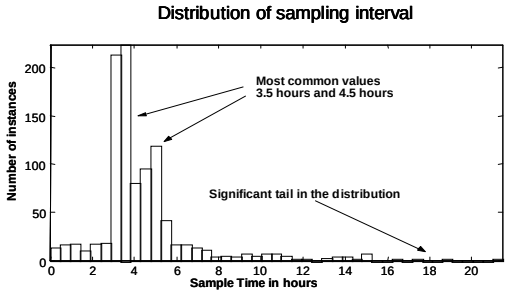

Fig. 6. Distribution of the output sampling intervals

We present the results of an industrial application of the proposed identification strategy. In this application, model identification was performed for the bleaching operation of a Bleached-Chemi Thermo-Mechanical Pulp process at Millar Western, Whitecourt, AB, Canada. The unit consists of 2 towers, has 4 MV's and 2 disturbances. We show the results for one of the output variables, the pulp brightness. A simplified process schematic is presented in Fig. 5. The MV's were chemical add-rates to the two towers, sampled every 10 minutes. The process has plug-flow characteristics. Forward path dynamics can be captured by low-order delay dominant models, with significant chemical recycle. The recycle effect can be taken into account using input terms lagged by the delay in the loop. The distribution of the output sampling intervals (Fig. 6) shows significant

variation in the sample times between consecutive samples, i.e. the outputs are irregularly sampled. We had to use routine operating data for model identification because plant tests could not be economically justified. However, it was feasible to use operating data in this case, because it contained enough excitation. The models obtained have to conform to *a priori* process knowledge which can be summarized as, fast forward-path dynamics, significant recycle and positive gains. The predictions from the proposed identification strategy are shown in Fig. 7 where the output values have been re-scaled for confidentiality reasons, and the corresponding step responses are shown in Fig. 8. The correlation coefficient (CC) between the predicted and measured values is 0.984 and the root-mean-squared error (RMSE) value is 0.87. The predictions are good. More importantly, the step responses conform to our qualitative knowledge about the process dynamics.

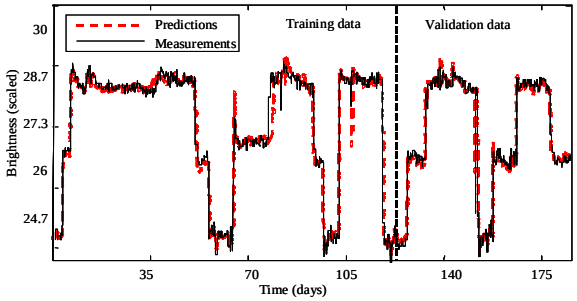

Fig. 7. Brightness predictions using EM-based strategy

5. CONCLUSIONS

An EM-based strategy for identification of processes with irregularly sampled outputs has been presented. The initial model required for the EM algorithm is obtained from FIR coefficients through an SVD procedure. Applications of the proposed approach to simulation and industrial case-studies have been presented and these show that the EM-based identification strategy is useful for data-based identification of state-space models even when output observations are missing, at regular or irregular intervals.

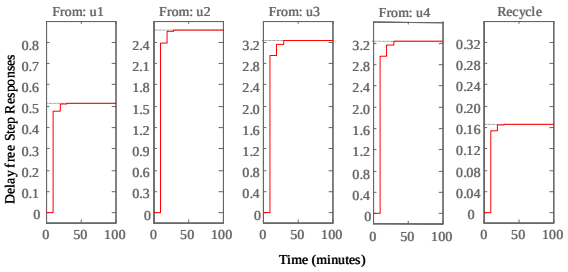

Fig. 8. Step responses using EM-based strategy

REFERENCES

Ansley, C. F. and R. Kohn (1983). Exact Likelihood of Vector Autoregressive-Moving Average Process with Missing or Aggregated Data. *Biometrika* **70**(1), 275–278.

Deistler, M. (2000). Model Identification and Adaptive Control. In: *System Identification - General Aspects and Structure*. Springer-Verlag.

Dempster, A.P., N.M. Laird and D.B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38.

Gupta, N.K. and R.K Mehra (1974). Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations. *IEEE Trans. Auto. Cont.* **19**, 774–783.

Isaksson, A.J. (1993). Identification of ARX-models subject to missing data. *IEEE Trans. Auto. Cont.* **38**, 813–819.

Kung, S.Y. (1978). A new identification and model reduction algorithm via singular value decomposition. In: *12th Asilomar Conference on Circuits, Systems and Computers*. Pacific Grove, CA. pp. 705–714.

Ljung, Lennart (1999). *System Identification: Theory for the User*. Prentice-Hall, Inc.. Englewood Cliffs, New Jersey. Second Edition.

Ninness, B. and S. Gibson (2002). Robust and simple algorithms for maximum likelihood estimation of multivariable systems. In: *Proceedings of the 2002 IFAC World Congress*. Barcelona.

Rao, C. Radhakrishna (2001). *Linear Statistical Inference and its Applications*. John Wiley and Sons. New York. Second Edition.

Schweppe, F.C. (1965). Evaluation of likelihood functions for gaussian signals. *IEEE Trans. Inform. Theo.* **IT-4**, 294–305.

Shumway, R.H. and D.S. Stoffer (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis* **3**(4), 253–264.

Shumway, R.H. and D.S. Stoffer (2000). *Time Series Analysis and Its Applications*. Springer-Verlag. New York.

Tanaka, Masahiro and Tohru Katayama (1990). Robust Identification and Smoothing for Linear System with Outliers and Missing Data. In: *Proceedings of the 1990 IFAC World Congress*. Tallinn.

VanOverschee, P. and B. De-Moor (1996). *Subspace Identification for Linear Systems-Theory, Implementation, Applications*. Kluwer Academic Publisher, Boston, MA.

Wu, C.F. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics* **11**, 95–103.