NEW FAST RECURSIVE ALGORITHMS FOR SIMULTANEOUS RECONSTRUCTION AND IDENTIFICATION OF AR PROCESSES WITH MISSING OBSERVATIONS

Rawad Zgheib, Gilles Fleury, Elisabeth Lahalle

NEW FAST RECURSIVE ALGORITHMS FOR SIMULTANEOUS RECONSTRUCTION AND IDENTIFICATION OF AR PROCESSES WITH MISSING OBSERVATIONS

Rawad Zgheib, Gilles Fleury, Elisabeth Lahalle

Department of Signal Processing and Electronic Systems, Supelec, Gif-sur-Yvette, France
E-mail : firstname.lastname@supelec.fr

ABSTRACT

This paper deals with the problem of adaptive reconstruction and identification of AR processes with randomly missing observations. The performances of a previously proposed real time algorithm are studied. Two new alternatives, based on other predictors, are proposed. They offer an unbiased estimation of the AR parameters. The first algorithm, based on the *h*-step predictor, is very simple but suffers from a large reconstruction error. The second one, based on the incomplete past predictor, offers an optimal reconstruction error in the least mean square sense.

1. INTRODUCTION

In many practical situations, periodically sampled signals with missing observations may be encountered. This is the case, for example, of errors in transmission, or of temporary unavailability of measurement. It is also the case, in coding of audio signals or images, for compression purposes.

Several methods have already been developed for the processing of autoregressive signals with missing data. They are generally of two types:

- Off line methods that use all available data. They are developped for spectral estimation such as in [5], for identification purposes such as in [6, 9, 14], or for reconstruction such as in [8]. In [9], Jones used a Kalman filter to calculate the exact likelihood function for unequally spaced data. Model's parameters, that most fits the data, are then estimated by a non linear optimization. Isaksson [6], proposed an EM algorithm equivalent to a maximum likelihood algorithm but much faster. In [8], the estimates of the unknown samples are obtained by minimizing the sum of squares of the residual errors that involve estimates of the autoregressive parameters. In [14], the identification of ARX models from incomplete data using least squares is studied.
- Real time adaptive methods such as in [1, 7, 11, 13]. In [11], an LMS-like algorithm for simultaneous reconstruction and identification is developed. In [13], an attempt to recursive identification based on pseudo-linear regression has been derived. In [1], the problem of estimating the output in missing-data situations is addressed. In [7], Isaksson derives a recursive EM algorithm for the identification of AR processes subject to missing data, based on the offline version of the one described in [6]. This algorithm uses a smoothing Kalman filter for the prediction. In addition, at each time, the inversion of a matrix is required to update the parameters. Therefore, it is not a real time algorithm.

In many applications, such as digital communications or systems tracking, on line processing is necessary. We are interested here in on line adaptive reconstruction and identification of autoregressive signals with randomly missing observations. The loss of samples process follows a Bernoulli law independent of the signal. In section 2, we begin by describing and analyzing the performances of a previously proposed LMS-like algorithm [11]. In section 3, new solutions, based on different predictors, are proposed. Finally, the new recursive algorithms are studied and compared in section 4.

2. LMS-LIKE ALGORITHM

2.1 Description

The LMS-like algorithm [11, 12] is based on the stochastic gradient principle. It minimizes a quadratic prediction error to estimate the AR parameters $\mathbf{a} = [a_1, \ldots, a_p]^\top$. Let $\hat{\mathbf{a}}_n$ be the vector of the estimated parameters at time n, it is updated using the following equation [10]:

$$\hat{\mathbf{a}}_{n+1} = \hat{\mathbf{a}}_n - \mu \left. \frac{\partial J_{n+1}}{\partial \mathbf{a}} \right|_{\mathbf{a}=\hat{\mathbf{a}}_n} \tag{1}$$

where, at time n+1, J_{n+1} is the instantaneous quadratic prediction error.

Let y_n be an AR process of order p subject to missing observations, and z_n its reconstruction. The principle of the reconstruction approach used in [5, 6, 8, 11, 12, 13] is to replace each missing observation by its predicted value. Then

$$z_n = \begin{cases} y_n & \text{if } y_n \text{ is available,} \\ \hat{y}_n & \text{otherwise,} \end{cases} \tag{2}$$

where $\hat{y}_n = \sum_{i=1}^{p} a_i z_{n-i}$. Due to missing observations, a sample may be predicted in terms of the prediction of previously lost samples. Its prediction, $\hat{y}_n$, is therefore non linear with respect to the parameters.

The square prediction error cost function is then non linear in terms of the parameters and its derivative is therefore:

$$\frac{\partial J_n}{\partial \mathbf{a}} = -2(y_n - \hat{y}_n)\frac{\partial \hat{y}_n}{\partial \mathbf{a}}. \tag{3}$$

The term $\partial \hat{y}_n / \partial \mathbf{a}$ is a vector of dimension p whose elements are:

$$\frac{\partial \hat{y}_n}{\partial a_k} = \frac{\partial \sum_{i=1}^{p} a_i z_{n-i}}{\partial a_k} = z_{n-k} + \sum_{i=1}^{p} a_i \frac{\partial z_{n-i}}{\partial a_k} \tag{4}$$

where, $\partial z_{n-i}/\partial \mathbf{a} = 0$ if y_{n-i} is available, and $\partial z_{n-i}/\partial \mathbf{a} = \partial \hat{y}_{n-i}/\partial \mathbf{a}$ otherwise. Therefore, the term $\partial \hat{y}_n / \partial \mathbf{a}$ is obtained

by the following recursive equation [12]:

$$\frac{\partial \hat{y}_n}{\partial \mathbf{a}} = \mathbf{z}_n + \left[\frac{\partial z_{n-1}}{\partial \mathbf{a}} \middle| \dots \middle| \frac{\partial z_{n-p}}{\partial \mathbf{a}} \right] \mathbf{a} \tag{5}$$

where $\mathbf{z}_n = [z_{n-1} \dots z_{n-p}]^\top$.

The vector $\mathbf{a}$ is updated only when a new sample is available, whereas the recursion (5) is used at all times. The LMS-like algorithm as proposed by Mirsaidi et Al. [11] will be referred to by Mirsaidi's algorithm in the following.

2.2 Performance analysis

One of the major interest of Mirsaidi's algorithm is its simplicity. Unfortunately, simulations show that it yields biased estimations of the parameters, for AR signals of order 2 and above. The biases increase with q, the Bernoulli's probability of a sample to be lost. Moreover, the biases seem to be independent of the parameters initialization, which tends to prove that Mirsaidi's algorithm converges toward a global minimum of the cost function. This assumption is confirmed by Wallin et al. [15]. According to their work, randomly missing data should not cause a multiple optima problem in the AR parameters estimation. Besides, they maximize the likelihood probability function to estimate the parameters. In the case of AR processes (Gaussian process), it is equivalent [4] to minimize the mean square prediction error as in [11].

3. PROPOSED ALTERNATIVES

In order to avoid the observed bias problem, we propose new LMS-like alternatives based on two other predictors.

3.1 Prediction

3.1.1 H-step predictor

The best linear combination of $1, y_1, \dots, y_n$ for predicting y_{n+h} is $\widehat{y}_{n+h} = \boldsymbol{P}_n y_{n+h}$, where $\boldsymbol{P}_n$ denotes the orthogonal projection mapping onto the subspace $\mathscr{S}$ generated by $1, y_1, \dots, y_n$. According to the projection theorem, $\widehat{y}_{n+h}$ thus defined is the unique element of $\mathscr{S}$ for which the distance $\|y_{n+h} - \widehat{y}_{n+h}\|$ is minimal and it is the best mean square predictor of y_{n+h} in $\mathscr{S}$ [4]. The recursive equation of the best h-step predictor for an AR(p), deduced from [4], is:

$$\widehat{y}_{n+h} = \boldsymbol{P}_n y_{n+h} = \sum_{i=1}^{p} a_i \boldsymbol{P}_n y_{n+h-i}. \tag{6}$$

For an AR(p), $\hat{y}_{n+1} = \boldsymbol{P}_n y_{n+1} = \sum_{i=1}^{p} a_i y_{n+1-i}$, so using the recursive equation (6) we get that $\widehat{y}_{n+h}$ is a linear combination of $y_{n-p+1}, \dots, y_n$. In the case of missing observations, a sample (at time $n+h$) is predicted in terms of the p preceeding consecutive available samples. Hence, minimizing the square prediction error in terms of the parameters leads to solve p equations for p variables, whatever the observation pattern is. If all the data between n and $n+h$ are missing, i.e. $z_{n+h-i} = \widehat{y}_{n+h-i}$ for $i = 1, ..., h$, this predictor is equivalent to the previous one.

3.1.2 Incomplete past predictor

To predict y_n, we shall assume that the data $y_{n-n_1}, \dots, y_{n-n_L}$ are missing with $0 < n_1 < \dots < n_s < \dots < n_L$ and let $M = \{n-n_1, \dots, n-n_L\}$. Bondon [3] has proved that $\widehat{y}_n$, given by the equation (7), is the orthogonal projection of y_n onto the space generated by the previous available observations. Let (ε_n) be the innovation process of (y_n),

$$y_n - \widehat{y}_n = -\sum_{s=0}^{L} \psi_s \sum_{j=0}^{n_s} a_{n_s-j} \varepsilon_{n-j}, \tag{7}$$

where the coefficients (ψ_s) satisfy the matrix equation

$$U(\psi_0, \psi_1, \dots, \psi_L)^\top = (1, 0, \dots, 0)^\top \tag{8}$$

U is the nonsingular $(L+1) \times (L+1)$ matrix with elements

$$u_{v,w} = \sum_{j=0}^{n_v \wedge n_w} a_{n_v-j} a_{n_w-j} \quad v, w = 0, \dots, L. \tag{9}$$

Since (y_n) is supposed to have an AR representation, the predictor $\widehat{y}_n$ has an AR representation for any finite set of missing data. This representation is unique [3] and is given by

$$\widehat{y}_n = \sum_{k \in \mathbb{N}-M} r_k y_{n-k}, \tag{10}$$

with

$$r_k = \delta_k - \sum_{s=0}^{L} \psi_s \sum_{j=0}^{n_s \wedge k} a_{n_s-j} a_{k-j}. \tag{11}$$

In the case of an AR(p) process, $\widehat{y}_n = \sum_{k=0, k \notin M}^{n_L+p} r_k y_{n-k}$.

3.1.3 Example

Let us consider, for example, the following observation pattern $\{1011010\}$ where 0 stands for a missing observation. We predict y_7, for an AR(2) process, using the different predictors presented in this paper.

1. The predictor used in [11]:

$$\hat{y}_7 = a_1 y_6 + a_1 a_2 y_4 + a_2^2 y_3. \tag{12}$$

2. The h-step predictor:

$$\hat{y}_7 = (a_1^3 + 2a_1 a_2) y_4 + (a_1^2 a_2 + a_2^2) y_3. \tag{13}$$

3. The incomplete past predictor:

$$\begin{aligned} \hat{y}_7 = &\left(a_1 + \frac{a_1 a_2}{1 + a_1^2} \right) y_6 + \left(a_1 a_2 - \frac{a_1^3 a_2 + a_1 a_2^2}{1 + a_1^2} \right) y_4 \\ &+ \left(a_2^2 - \frac{a_1^2 a_2^2}{1 + a_1^2} \right) y_3. \end{aligned} \tag{14}$$

The predictor used in [11] and the incomplete past predictors predict the missing observation using all the previous available data up to and including the last bloc of p (in the case of an AR(p)) consecutive available observations. Hence, they use more information than the h-step predictor. The orthogonal projection of y_n on its incomplete past has a unique AR representation of coefficients r_k given by the incomplete past predictor [3]. As we can see from this example, the predictor used in [11] has a different AR representation than the incomplete past predictor. The coefficients r_k obtained with the incomplete past predictor contain additive

corrective terms compared to those obtained with the predictor used in [11]. We conclude that the predictor used in [11] is not an orthogonal projection on the incomplete past. Thus it is not optimal in the least mean squares sense. In [7], it was noticed that the expectations of the state based on the observed data will not just replace a partially unknown state by its predicted value but there will also be a correction term based on the prediction error covariance matrix. Without this correction, it would correspond to an ordinary least square solution on reconstructed data which was the prediction used in [11], and this will typically converge to biased estimates. The same problem of bias is present with the pseudo linear RLS proposed in [13]. The problem of identification ARX models with missing observations using the least squares has been studied by Wallin et al. [15]. They showed that least squares estimate of the parameters using the predictor used in [11] is biased. They calculated an expression of that bias and concluded that the bias is zero for any optimal predictor in the least squares sense. This explains the bias of the estimated AR parameters with Mirsaidi's algorithm, and will lead us to use the *h*-step and the incomplete past predictors as predictors in LMS-like algorithms.

3.2 Proposed Algorithms

The difference between the three proposed methods relies on the used predictor and consequently on the computation of $\partial J_n/\partial \mathbf{a}$ to update the model parameters.

3.2.1 LMS-like using the h-step predictor

The *h*-step predictor is equivalent to the one used in [11] when all the samples between the last block of p consecutive available samples, $y_{n-p+1},\ldots,y_n$, and the current sample y_{n+h}, are lost. This algorithm is therefore simply deduced from Mirsaidi's algorithm. The unique difference is to consider, for the next predictions, an observed sample as missing until a new block of p consecutive available samples is formed. So, in the subsequent steps, $\hat{y}_{n+h}$ is used for the prediction and $\partial\hat{y}_{n+h}/\partial\mathbf{a}$ is used in (5), instead of 0. In opposition, if at time $n+h$, a new block of p consecutive available samples, $y_{n+h-p+1},\ldots,y_{n+h}$, is formed, their observed values are used for the next predictions. The next samples are predicted in terms of these p samples so, $\partial z_{n+h}/\partial\mathbf{a},\ldots,\partial z_{n+h-p+1}/\partial\mathbf{a}$ are equal to 0.

This algorithm is very simple and leads to unbiased estimated parameters. On the other hand, this method is highly dependent on the observation pattern and h can become arbitrarily large. In particular, in the case of high-order AR signals with missing observations, blocks of p consecutive available samples are rarely formed. This may lead to a large mean square reconstruction error and consequently to a large variance on the parameter estimation, increasing with q.

3.2.2 LMS-like using the incomplete past predictor

Since the incomplete past predictor is optimal in the least mean square sense, we propose to use it as a predictor in a LMS-like algorithm. We consider its AR representation given by equations (10) and (11). The cost function to optimize is then $J_n=(y_n-\sum_{k\in\mathbb{N}-M} r_k y_{n-k})^2$. Its partial derivative with respect to (a_i) is,

$$\begin{aligned}\frac{\partial J_n}{\partial a_i} &= \sum_{s=0}^{L}\frac{\partial\psi_s}{\partial a_i}\sum_{j=0}^{n_s\wedge k} a_{n_s-j}a_{k-j} \\ &\quad+\sum_{s=0}^{L}\psi_s\left[A(i,k,n_s)+A(i,n_s,k)\right]\end{aligned} \tag{15}$$

with $A(i,k,l) = \begin{cases} a_{k-l+i} & \text{if } 0<l-i<l\wedge k \\ 0 & \text{otherwise}\end{cases}$

Note here that $a_0=1$ and $a_j=0$ for $j>p$.

To calculate $\partial\psi_s/\partial a_i$, we differentiate the matrix system (8) with respect to a_i, which gives,

$$U\frac{\partial(\psi_0,\ldots,\psi_L)^\top}{\partial a_i} = -\frac{\partial U}{\partial a_i}(\psi_0,\ldots,\psi_L)^\top. \tag{16}$$

The matrix $\partial U/\partial a_i$ is formed by the derivative of each element of U with respect to the parameter a_i. Thus, we get p matrices corresponding to the derivative of U with respect to each of the p parameters. In the same way as in (15),

$$\frac{\partial u_{vw}}{\partial a_i} = A(i,n_v,n_w)+A(i,n_w,n_v) \tag{17}$$

In opposition to the two previous algorithms, the prediction and the calculation of $\partial J_n/\partial a_i$, for this algorithm, are not recursive. Consequently the systems (8) and (16) must be solved at each time, which is computationally intensive.

4. COMPARISON

The test signal, used to compare the performances of the three algorithms, is an AR(2) process of parameters [1.5, -0.7] generated over 10^5 samples. The Bernoulli's probability of sample loss is $q=0.3$. Due to the missing samples, μ must have smaller values than for the classical LMS. Simulations are done with $\mu=7.10^{-5}$. Algorithms Misaidi, A and B stand, respectively, for Mirsaidi's algorithm and its alternatives using the *h*-step predictor and the incomplete past predictor. The figure 1 shows the estimation of the parameter a_1 for the three algorithms, and figure 2 shows the behaviour after the convergence.

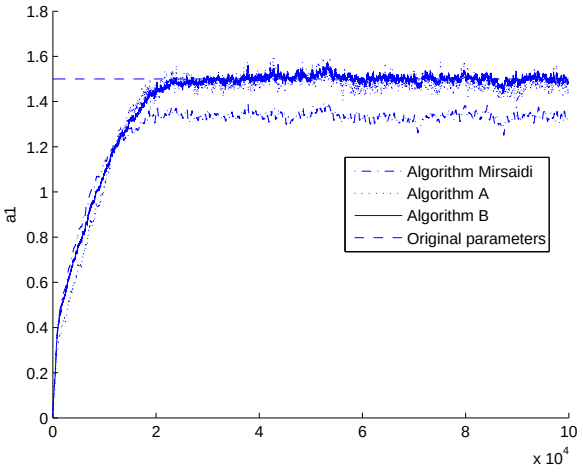

Figure 1: Estimation of a_1 with the three algorithms.

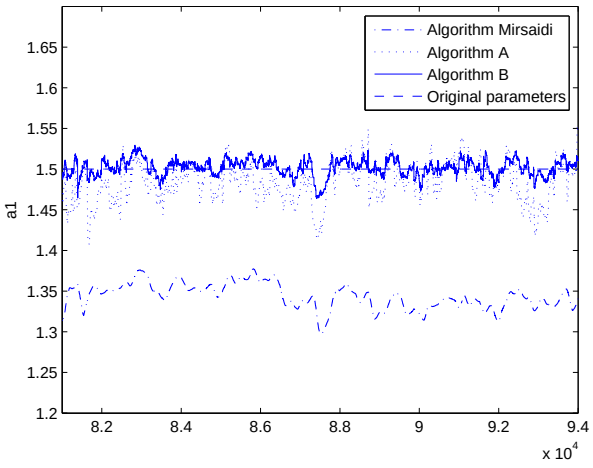

Figure 2: Estimation of a_1 after the convergence of the three algorithms.

The performances of these algorithms are summarized in table 1. b_i is the bias existing in the estimation of a parameter and σ_i is the standard deviation of a parameter estimation. The MQRE is the normalized mean quadratic reconstruction error. The normalization is done with respect to the power of the signal. For a signal generated over N samples, it is given by the following equation:

$$MQRE = \frac{\sum_{k=1}^{N}(y_k - z_k)^2}{\sum_{k=1}^{N} y_k^2}. \qquad (18)$$

The CPU is the computation time, in seconds, needed to simulate each one of the three algorithms for the above test signal, using MATLAB on a processor of 2.4 GHz.

Table 1: Comparison of the three algorithms performances

Algorithm	b_1	b_2	σ_1	σ_2
Mirsaidi	0.166	0.137	0.019	0.018
A	0.026	0.014	0.029	0.0298
B	0.004	0.009	0.014	0.016

Algorithm	MQRE	CPU
Mirsaidi	0.07	5
A	0.11	4
B	0.061	38

Table 1 and figures 1 and 2 show that Mirsaidi's algorithm converges toward biased values of the parameters while the estimation with the two other algorithms is not biased. Moreover, referring to the table 1, the algorithm B is computationally expensive and simulations show that the CPU time increases with q, this may be due to the increasing size of the matrices to be inverted. However, the algorithms Mirsaidi and A are much faster and the CPU time does not increase with q. On the other hand, the algorithm B offers the smallest prediction error comparing to the others while the algorithm A has the largest one.

Moreover, figure 3 shows, for each algorithm, an empirical distribution for the estimator of a_1. For 2000 generations of the signal and of the observation scheme, the parameter a_1 is estimated with the three algorithms. For each one, the empirical distribution of a_1 at a time t_0 after the convergence is deduced.

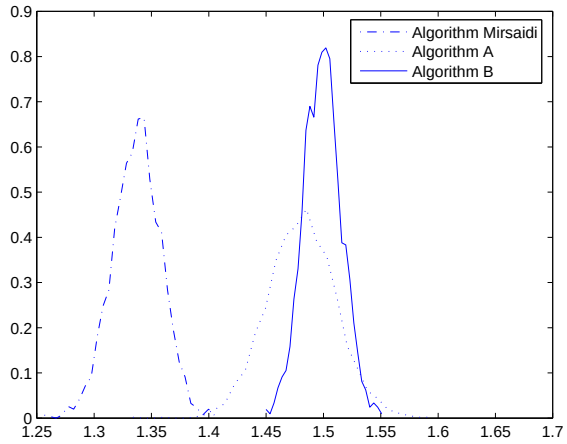

Figure 3: Empirical estimator distribution of a_1 with the three algorithms.

Figure 3 shows that the empirical distribution for the estimator of a_1 using algorithm B is centered on 1.5 which confirms that this estimator is not biased. The estimation of a_1, using the algorithm A, is slightly biased since the empirical distribution is centered on 1.49. This is not normal since this estimator is supposed to be unbiased, this may be the result of the high variance of this estimator and of the empirical estimation of the distribution (it is limited to 2000 generations). However, for Mirsaidi's algorithm, the empirical distribution is centered on 1.34 which shows that this estimator is strongly biased.

According to table 1 and figures 2 and 3, for all LMS-like algorithms, the variance of the parameters estimator is related to the signal prediction error. Indeed, the estimator of the algorithm B offers the smallest variances σ_1^2 and σ_2^2, however, in the case of the algorithm A, they have the largest values. This observation was expected since the update of the parameters using equations (1) and (3) is proportional to the prediction error.

Finally, figure 4 shows the evolution of the MQRE in terms of q, for the three algorithms and for the same test signal as above. Referring to figure 4, the MQRE obtained for Misaidi's algorithm is still close to the optimal one, even for large q.

5. CONCLUSION

Two new adaptive algorithms for reconstruction and identification of an AR process with missing observations are described. The first one based on the h-step predictor, which is very simple, offers an unbiased parameter estimation, but presents a large quadratic reconstruction error. The second one is based on the incomplete past predictor. This algorithm allows, at the same time, an unbiased estimation of the parameters, and an optimal reconstruction in the least mean square sense. It can be noticed from figure 4, that the quadratic reconstruction error for algorithm B is, effectively,

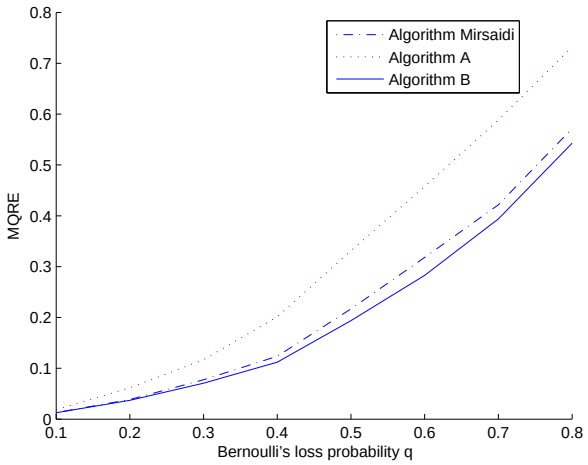

Figure 4: MQRE in terms of the loss probability q.

minimal comparing to the other techniques. However, this algorithm is quite time consuming.

In conclusion, the analysis of the three algorithms shows that Mirsaidi's algorithm, while not suited for identification, is useful for real time reconstruction of a signal. On the other hand, algorithm A is more suited for real time identification. However, for applications where computation time does not present a limitation, algorithm B presents the best performances.

REFERENCES

[1] P. Albertos and R. Sanchis and A. Sala, "Output prediction under scarce data operation: control application", *Automatica, 35*, pp. 1671–1681, 1999.

[2] K. Asswad and E. Lahalle and J. Oksman, "Reconstruction en temps rel de signaux bidimensionnels chantillons manquants," *Signal Processing*, vol. 18, No. 2, 2001.

[3] P. Bondon, "Prediction with incomplete past of a stationnary process," *Elsevier, Stochastic processes and their applications*, pp. 67–76, 2002.

[4] P. Brockwell and R. Davis, *Time Series: Theory and methods*. Springer-Verlag, 1991.

[5] A. Ferrari and G. Alengrin, "AR Spectral analysis with randomly missing observations," *IEEE Proceedings on Statistical Signal and Array Processing*, September 1998, pp. 320–323.

[6] A. Isaksson, "Identification of ARX Models Subject to Missing Data," *IEEE Transactions on Automatic Control*, vol. 38, pp. 813–819, 1993.

[7] A. Isaksson, "A Recursive EM Algorithm for identification subject to missing data," in *SYSID*, Copenhagen, Denmark, 1994, vol. 2, pp. 679–684.

[8] A. Janssen and R. Veldhuis and L. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 45, No. 6, 1997.

[9] R. H. Jones, "Fitting a continous time autoregression to discrete data," *Academic, Applied Time Series Analysis II*, 1981.

[10] O. Macchi, *Adaptive processing: The least mean squares approach with applications in transmission*. Wiley, 1995.

[11] S. Mirsaidi and G. Fleury and J. Oksman, "LMS Like AR modeling in the case of missing observations," *IEEE Transactions on Signal Processing*, vol. 45, No. 6, 1997.

[12] S. Mirsaidi and J. Oksman, "A class of real-time AR identification algorithms in the case of missing observations," in *Proc. Eighth European Signal Processing Conference, EUSIPCO*, Trieste, Italy, 1996, pp. 803–806.

[13] R. Sanchis and P. Albertos, "Recursive identification under scarce measurements - convergence analysis," *Automatica, 38*, pp. 535–544, 2002.

[14] R. Wallin and A. Isaksson and L. Ljung, "An iterative method for identification of ARX models from incomplete data," in *Proceedings of the* 39^th *IEEE Conference on Decision and Control*, Sydney, Australia, Dec. 2000, pp. 203–208.

[15] R. Wallin and A. Isaksson, "Multiple Optima in Identification of ARX Models Subject to Missing Data," *EURASIP Journal on Applied Signal Processing*, vol. 1, pp. 30–37, 2002.