Sequential Parameter Estimation of Time-varying Non-Gaussian Autoregressive Processes

Petar M. Djurić, Jayesh Kotecha, Fabien Esteve, and Etienne Perret

Abstract

Parameter estimation of time-varying non-Gaussian autoregressive processes can be a highly nonlinear problem. The problem gets even more difficult if the functional form of the time variation of the process parameters is unknown. In this paper we address parameter estimation of such processes by particle filtering, where posterior densities are approximated by sets of samples (particles) and particle weights. These sets are updated as new measurements become available using the principle of sequential importance sampling. From the samples and their weights one can compute a wide variety of estimates of the unknowns. In absence of exact modeling of the time variation of the process parameters, we exploit the concept of forgetting factors so that recent measurements affect current estimates more than older measurements. We investigate the performance of the proposed approach on autoregressive processes whose parameters change abruptly at unknown instants and with driving noises which are Gaussian mixtures or Laplacian processes.

I. Introduction

In on-line signal processing, a typical objective is to process incoming data sequentially in time and extract information from them. Applications vary and include system identification [30], equalization [31], [32], echo cancelation [11], blind source separation [22], beamforming [20], [23], blind deconvolution [21], time-varying spectrum estimation [20], adaptive detection [38], and digital enhancement of speech and audio signals [15]. These applications find practical use in communications, radar, sonar, geophysical explorations, astrophysics, biomedical signal processing, and financial time series analysis.

The task of on-line signal processing usually amounts to estimation of unknowns and tracking them as they change with time. A widely adopted approach to addressing this problem is the Kalman filter, which is optimal in cases when the signal models are linear and the noises are additive and Gaussian [30]. The framework of the Kalman filter allows for derivation of all the recursive least-squares (RLS) adaptive filters [34]. When nonlinearities have to be tackled, the extended Kalman filter becomes the tool for estimating the unknowns of interest [2], [20], [24]. It has been shown in the literature that in many situations the extended Kalman filter, due to the implemented approximations, can diverge in the tracking of the unknowns and in general can provide poor performance [16]. Many alternative approaches to overcome the deficiencies of the extended Kalman filter have been tried including Gaussian sum filters [1], approximations of the first two moments of densities [12], evaluations of required densities over grids [27], and the unscented Kalman filter [25].

Another approach to tracking time-varying signals is particle filtering [7]. The underlying approximation implemented by particle filters is the representation of densities by samples (particles) and their associated weights. In particular, if $x^{(m)}$ and $w^{(m)}$, $m = 1, 2, \cdots, M$ are the samples and

P. M. Djurić and is with the Department of Electrical and Computer Engineering at Stony Brook University, Stony Brook, NY, 11794, and J. H. Kotecha is with the Department of Electrical and Computer Engineering at the University of Wisconsin at Madison, Madison, WI, 53706. F. Esteve and E. Perret are with the ENSEEIHT/TéSA, 31071 Toulouse, France. E-mail: djuric@ece.sunysb.edu, jkotecha@ece.wisc.edu; fabien.esteve, etienne.perret@etu.enseeiht.fr.

their weights respectively, one approximation of $p(x)$ is given by

$$\hat{p}(x) = \sum_{i=1}^{M} w^{(m)} \delta(x - x^{(m)}) \tag{1}$$

where $\delta(\cdot)$ is Dirac's delta function. The approximation of the densities by particles can be implemented sequentially, where as soon as the next observation becomes available, the set of particles and their weights are updated using the Bayes rule. Some of the basics of this procedure are reviewed in this paper. The recent interest in particle filters within the signal processing community has been initiated by [16], where a special type of particle filters are used for target tracking. Since the particle filtering methods are computationally intensive, the continued advancement of computer technology in the past few years has played a critical role in sustaining this interest. An important feature of particle filtering is that it can be implemented in parallel, which allows for major speed-ups in various applications.

One advantage of particle filters over other methods is that they can be applied to almost any type of problem where signal variations are present. This includes models with high nonlinearities and with noises that are not necessarily Gaussian. In all the work on particle filtering presented in the wide literature, it is assumed that the applied model is composed of a state equation and an observation equation, where the state equation describes the dynamics of the tracked signal (or parameters). Thus, the use of particle filters requires knowledge of the functional form of the signal (parameter) variations. In this paper, we make the assumption that this model is *not* available, that is, we have no information about the dynamics of the unknowns. In absence of a state-equation, we propose to use a random walk model for describing the time variation of the signal (or parameters). We show that the random walk model implies forgetting of old measurements [6], [37]. In other words, it assigns more weight to more recent observations than to older measurements.

In this paper we address the problem of tracking the parameters of a non-Gaussian autoregressive (AR) process whose parameters vary with time. The usefulness of the modeling of time series by autoregressions is well documented in the wide literature [19], [26]. Most of the reported work, however, deals with stationary Gaussian AR processes, and rightfully so because many random processes can be modeled successfully with them. In some cases, however, the Gaussian AR models are inappropriate, as for instance, for processes that contain spikes, that is, samples with large values. Such signals are common in underwater acoustic, communications, oil exploration measurements, and seismology. In all of them, the processes can still be modeled as autoregressions, but with non-Gaussian driving processes, for example, Gaussian mixture or Laplacian processes. Another deviation from the standard AR model is the time-varying AR model where the parameters vary with time [5], [10], [17], [18], [28], [33], [37].

The estimation of the AR parameters of non-Gaussian AR models is a difficult task. Parameter estimation of such models has rarely been reported, primarily due to lack of tractable approaches for dealing with them. In [35], a maximum likelihood estimator is presented and its performance compared to the Cramer-Rao bound. The conditional likelihood function is maximized by a Newton-Raphson search algorithm. This method obviously cannot be used in the setting of interest in this paper. In a more recent publication, [36], the driving noises of the AR model are Gaussian mixtures, and the applied estimation method is based on a generalized version of the expectation-maximization principle.

When the AR parameters change with time, the problem of their estimation becomes even more difficult. In this paper, the objective is to address this problem, and the applied methodology is based on particle filtering. In [9] and [14], particle filters are also applied to estimation of time-varying AR models, but the driving noises there are Gaussian processes.

The paper is organized as follows. In Section II, we formulate the problem. In Section III, we provide a brief review of particle filtering. An important contribution of the paper is in Section IV, where we propose particle filters with forgetting factors. The proposed method is applied to time-varying non-Gaussian autoregressive processes in Section V. In Section VI, we present simulation examples, and in Section VII, we conclude the paper with some final remarks.

II. Problem Formulation

Observed data y_t, $t = 1, 2, \cdots$, represent a time-varying AR process of order K that is excited by a non-Gaussian noise. The data are modeled by

$$y_t = \sum_{k=1}^{K} a_{tk} y_{t-k} + v_t$$

where v_t is the driving noise of the process, and a_{tk}, $k = 1, 2, ..., K$ are the parameters of the process at time t. The values of the AR parameters are unknown, but the model order of the AR process, K, is assumed known. The driving noise process is iid and non-Gaussian, and is modeled as either a Gaussian mixture with two mixands, i.e.,

$$v_t \sim (1-\epsilon)\mathcal{N}(0, \sigma_1^2) + \epsilon\mathcal{N}(0, \sigma_2^2) \tag{2}$$

where $0 < \epsilon < 1$, and $\sigma_2^2 >> \sigma_1^2$, or as a Laplacian, that is,

$$v_t \sim \frac{\alpha}{2} e^{-\alpha|v_t|} \tag{3}$$

where $\alpha > 0$. In this paper we assume that the noise parameters, ϵ, σ_1^2, and σ_2^2 of the Gaussian mixture process and α of the Laplacian noise are known. The objective is to track the AR parameters, a_{tk}, $k = 1, 2, \cdots, K$, $\forall t$.

III. Particle Filters

Many time-varying signals of interest can be described by the following set of equations:

$$\begin{array}{rcl} x_t & = & f_t(x_{t-1}, u_t) \\ y_t & = & h_t(x_t, v_t) \end{array} \tag{4}$$

where $t \in \mathbb{N}$ is a discrete-time index, $x_t \in \mathbb{R}$ is an unobserved signal at t, $y_t \in \mathbb{R}$ is an observation, and $u_t \in \mathbb{R}$ and $v_t \in \mathbb{R}$ are noise samples. The mapping $f_t : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ is referred to as a signal transition function, and $h_t : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$, as a measurement function. The analytic forms of the two functions are assumed known. Generalization of (4) to include vector observations and signals as well as multivariable functions is straightforward.

There are three different classes of signal processing problems related to the model described by (4):
1. filtering: $\forall\, t$, estimate x_t based on $y_{1:t}$,
2. prediction: $\forall\, t$ and some $\tau > 0$, estimate $x_{t+\tau}$, based on $y_{1:t}$, and
3. smoothing: $\forall t$, estimate x_t, based on $y_{1:T}$, $t \in \mathbb{Z}_T = \{1, 2, \ldots, T\}$
where $y_{1:t} = \{y_1, y_2, \cdots, y_t\}$. Another very important objective is to carry out the estimation of the unknowns *recursively* in time.

A key expression for recursive implementation of the estimation is the update equation of the posterior density of $x_{1:t} = \{x_1, x_2, \cdots, x_t\}$, which is given by

$$p(x_{1:t}|y_{1:t}) = \frac{p(y_t|x_t)\, p(x_t|x_{t-1})}{p(y_t|y_{1:t-1})} p(x_{1:t-1}|y_{1:t-1}). \tag{5}$$

Under the standard assumptions that u_t and v_t represent additive noise and are independently and identically distributed according to Gaussian distributions and that the functions $f_t(\cdot)$ and $h_t(\cdot)$ are linear in x_{t-1} and x_t, respectively, the above problems are optimally resolved by the Kalman filter [2]. When the optimal solutions cannot be obtained analytically, one resorts to various approximations of the posterior distributions [2], [24].

The set of methods known as particle filtering methods are based on a very interesting paradigm. The basic idea is to represent the distribution of interest as a collection of samples (particles) from that distribution. One draws M particles, $\mathcal{X}_t = \{x_t^{(m)}\}_{m=1}^M$, from a so called importance sampling distribution $\pi(x_{1:t}|y_{1:t})$. Subsequently, the particles are weighted as $w_t^{(m)} = \frac{p(x_{1:t}^{(m)}|y_{1:t})}{\pi(x_{1:t}^{(m)}|y_{1:t})}$. If $\mathcal{W}_t = \{w_t^{(m)}\}_{m=1}^M$, then the sets $\mathcal{X}_t$ and $\mathcal{W}_t$ can be used to approximate the posterior distribution $p(x_t|y_{1:t})$ as in (1), or

$$\hat{p}(x_t|y_{1:t}) = \sum_{m=1}^{M} w_t^{(m)} \delta(x_t - x_t^{(m)}). \tag{6}$$

It can be shown that the above estimate converges in distribution to the true posterior as $M \to \infty$ [13]. More importantly, the estimate of $E_p(g(x_t))$, where $E_p(g(\cdot))$ is the expected value of the random variable $g(x_t)$ with respect to the posterior distribution $p(x_t|y_{1:t})$, can be written as

$$\hat{E}_p(g(x_t)) = \sum_{m=1}^{M} w_t^{(m)} g(x_t^{(m)}). \tag{7}$$

Thus, the particles and their weights allow for easy computation of minimum mean square error (MMSE) estimates. Other estimates are also easy to obtain.

Due to the Markovian nature of the state equation, we can develop a sequential procedure called sequential importance sampling (SIS), which generates samples from $p(x_{1:t}|y_{1:t})$ sequentially [16], [29]. As new data become available, the particles are propagated by exploiting (5). In this sequential updating mechanism, the importance function has the form $\pi(x_t|x_{1:\,t-1}, y_{1:\,t})$, which allows for easy computation of the particle weights. The ideal importance function minimizes the conditional variance of the weights and is given by [8]

$$\begin{aligned} \pi(x_t|x_{1:\,t-1}, y_{1:\,t}) &= p(x_t|x_{t-1}, y_t) \\ &\propto p(y_t|x_t)p(x_t|x_{t-1}). \end{aligned}$$

The SIS algorithm can be summarized as follows:

1. At time $t = 0$, we generate M particles from $\pi(x_0)$ and denote them $x_0^{(m)}, m = 1, \ldots, M$, with weights

$$w_0^{(m)} = \frac{p(x_0^{(m)})}{\pi(x_0^{(m)})}$$

where $p(x_0)$ is the prior density of x_0.

2. At times $t = 1, \ldots, T$, let $\mathcal{X}_t = \{x_t^{(m)}\}_{m=1}^M$ be the set of particles with weights $\mathcal{W}_t = \{w_t^{(m)}\}_{m=1}^M$. The particles and weights $\{x_{t-1}^{(m)}, w_{t-1}^{(m)})\}_{m=1}^M$ approximate the posterior density $p(x_{t-1}|y_{1:t-1})$ according to (6). We obtain the particles and weights for time t from steps 3, 4 and 5.

3. For $m = 1, \ldots, M$, draw $x_t^{(m)} \sim \pi(x_t|x_{1:t-1}^{(m)}, y_{1:t})$.

4. For $m = 1, \ldots, M$, compute the weights of $x_t^{(m)}$ using [8]

$$\bar{w}_t^{(m)} = w_{t-1}^{(m)} \frac{p(y_t|x_t^{(m)})p(x_t^{(m)}|x_{t-1}^{(m)})}{\pi(x_t^{(m)}|x_{1:t-1}^{(m)}, y_{1:t})}. \tag{8}$$

5. Normalize the weights using

$$w_t^{(m)} = \frac{\bar{w}_t^{(m)}}{\sum_{j=1}^{M} \bar{w}_t^{(j)}}.$$

An important problem that occurs in sequential Monte Carlo methods is that of sample degeneration. As the recursions proceed, the importance weights of all but a few of the trajectories become insignificant [29]. The degeneracy implies that the performance of the particle filter will be very poor. To combat the problem of degeneracy, resampling is used. Resampling effectively throws away the trajectories (or particles) with negligible weights and duplicates the ones having significant weights, in proportion to their weights. Simple random resampling is implemented in the following manner. Let $\{x_t^{(m)}, w_t^{(m)}\}_{m=1}^{M}$ be the weights and particles that are being resampled. Then

1. For $m = 1, \ldots, M$, generate a number $j \in \{1, \ldots, M\}$ with probabilities proportional to $\{w_t^{(1)}, \ldots, w_t^{(M)}\}$, and let $\widetilde{x}_t^{(m)} = x_t^{(j)}$.
2. For $m = 1, \ldots, M$, let $\widetilde{w}_t^{(m)} = 1/M$. Then $\{\widetilde{x}_t^{(m)}, \widetilde{w}_t^{(m)}\}_{m=1}^{M}$ represents the new sets of weights and particles.

Improved resampling in terms of speed can be implemented by using the so called systematic resampling scheme [4] or stratified resampling [3].

Much of the activity in particle filtering in the sixties and seventies was in the field of automatic control. With the advancement of computer technology in the eighties and nineties, the work on particle filters intensified and many new contributions appeared in journal and conference papers. A good source of recent advances and many relevant references is [7].

IV. Particle Filters with Forgetting Factors

In many practical situations, the function that describes the time variation of the signals $f_t(\cdot)$ is unknown. It is unclear then how to apply particle filters, especially keeping in mind that a critical density function needed for implementing the recursion in (5) is missing. Note that the form of the density $p(x_t|x_{t-1})$ depends directly on $f_t(\cdot)$. In [6], we argue that this is possible and can be done in somewhat similar way as with methods known as recursive least-squares with discounted measurements [19]. Recall that the idea there is to minimize a criterion of the form

$$\varepsilon_t = \sum_{n=1}^{t} \lambda^{t-n} e_n^2$$

where λ is known as a forgetting factor with $0 < \lambda \leq 1$, and e_t is an error that is minimized and given by

$$e_t = y_t - d_t$$

with d_t being a desired signal. The tracking of the unknowns is possible without knowledge of the parametric function of their trajectories because with $\lambda < 1$, the more recent measurements have larger weights than measurements taken further in the past. In fact, we apply implicitly a window to our data that allows more recent data to affect current estimates of the unknowns more than old data.

In the case of particle filters, one can replicate this philosophy by introducing a state equation that will enforce "aging" of data. Perhaps the simplest way of doing it is to have a random walk model in the state equation, that is

$$x_t = x_{t-1} + u_t \tag{9}$$

where u_t is a zero mean random sample that comes from a known distribution. Now, if the particles $x_{t-1}^{(m)}$ with their weights $w_{t-1}^{(m)}$ approximate $p(x_{t-1}|y_{1:t-1})$, with (9) the distribution of x_t will be wider due to the convolution of the densities of x_{t-1} and u_t. It turns out that this implies forgetting of old data, where the forgetting depends on the parameters of $p(u_t)$. For example, the larger the variance of u_t, the faster the forgetting of old data [6]. Additional theory on the subject can be found in [37] and the references therein. In the next section we present the details of implementing this approach to the type of AR processes of interest in this paper.

V. Estimation of time-varying non-Gaussian autoregressive processes by particle filters

The observation equation of an AR(K) process can be written as

$$y_t = \mathbf{a}_t^{\mathsf{T}} \mathbf{y}_t + v_t \tag{10}$$

where $\mathbf{a}_t^{\mathsf{T}} \equiv (a_{t1}, \ldots, a_{tK})$ and $\mathbf{y}_t \equiv (y_{t-1}, \ldots, y_{t-K})^{\mathsf{T}}$. Since the dynamic behavior of $\mathbf{a}_t$ is unknown, as suggested in the previous section, we model it with a random walk, i.e.,

$$\mathbf{a}_t = \mathbf{a}_{t-1} + \mathbf{u}_t \tag{11}$$

where $\mathbf{u}_t$ is a known noise process from which we can draw samples easily. It is reasonable to choose the noise process as a zero mean Gaussian with covariance matrix $\mathbf{\Sigma}_{\mathbf{u}_t}$. The covariance matrix $\mathbf{\Sigma}_{\mathbf{u}_t}$ is then set to vary with time by depending on the covariance matrix $\mathbf{\Sigma}_{\mathbf{a}_{t-1}}$. For example, for the AR(1) problem, we choose

$$\sigma_{a_t}^2 = \frac{\sigma_{a_{t-1}}^2}{\lambda}$$

where λ is the forgetting factor. From (11), we get

$$\sigma_{u_t}^2 = \sigma_{a_{t-1}}^2 (\frac{1}{\lambda} - 1). \tag{12}$$

Similarly, for $K > 1$, we can choose

$$\mathbf{\Sigma}_{\mathbf{u}_t} = \mathbf{\Sigma}_{diag} (\frac{1}{\lambda} - 1)$$

where $\mathbf{\Sigma}_{diag}$ is a diagonal matrix whose diagonal elements are equal to the diagonal elements of $\mathbf{\Sigma}_{\mathbf{a}_{t-1}}$.

Now, the problem is cast in the form of a dynamic state space model, and a particle filtering algorithm for sequential estimation of $\mathbf{a}_t$ can readily be applied as discussed in the previous section. An important component of the algorithm is the importance function, $\pi(\mathbf{a}_t|\mathbf{a}_{1:t-1}, y_{1:t})$, which is used to generate the particles $\mathbf{a}_t^{(m)}$.

The algorithm can be outlined as follows:

1. Initialize $\{\mathbf{a}_0^{(m)}\}_{m=1}^M$ by obtaining samples from a prior distribution $p(\mathbf{a}_0)$ and let $\bar{w}_0^{(m)} = 1$ for $m = 1, \ldots, M$. Then for each time step repeat steps 2-6.
2. Compute the covariance matrix of $\mathbf{a}_t$ and obtain the covariance matrix $\mathbf{\Sigma}_{\mathbf{u}_t}$.
3. For $i = 1, \ldots, M$, obtain samples $\mathbf{a}_t^{(m)}$ from the importance function $\pi(\mathbf{a}_t|\mathbf{a}_{1:t-1}^{(m)}, y_{1:t})$. A simple choice of it is $p(\mathbf{a}_t|\mathbf{a}_{t-1}^{(m)})$.
4. For $i = 1, \ldots, M$, update the importance weights by

$$\bar{w}_t^{(m)} = w_{t-1}^{(m)} \frac{p(y_t|\mathbf{a}_t^{(m)}) p(\mathbf{a}_t^{(m)}|\mathbf{a}_{t-1}^{(m)})}{\pi(\mathbf{a}_t^{(m)}|\mathbf{a}_{1:t-1}^{(m)}, y_{1:t})}.$$

If the driving noise is a Gaussian mixture and $\pi(\mathbf{a}_t|\mathbf{a}_{1:t-1}^{(m)}, y_{1:t}) = p(\mathbf{a}_t|\mathbf{a}_{t-1}^{(m)})$, the update is given by

$$\bar{w}_t^{(m)} = w_{t-1}^{(m)} \left((1-\epsilon)\mathcal{N}(\mathbf{a}_t^{(m)\mathsf{T}}\mathbf{y}_t, \sigma_1^2) + \epsilon\mathcal{N}(\mathbf{a}_t^{(m)\mathsf{T}}\mathbf{y}_t, \sigma_2^2) \right).$$

If the noise is Laplacian, the update is done by

$$\bar{w}_t^{(m)} = w_{t-1}^{(m)} e^{-\alpha|y_t - \mathbf{a}_t^{(m)\mathsf{T}}\mathbf{y}_t|}.$$

5. Normalize the weights according to

$$w_t^{(m)} = \frac{\bar{w}_t^{(m)}}{\sum_{i=1}^{M} \bar{w}_t^{(m)}}.$$

6. Resample occasionally or at every time instant from $\{\mathbf{a}_t^{(m)}, w_t^{(m)}\}_{m=1}^{M}$ to obtain particles of equal weights.

VI. Simulation Results

We present, next, results of experiments that show the performance of the proposed approach. In all our simulations we use the $p(\mathbf{a}_t|\mathbf{a}_{t-1})$ as the importance function. First, we show a simple example that emphasizes the central ideas in this paper. We estimated recursively the coefficient of an AR(1) process with non-Gaussian driving noise. The data were generated according to

$$y_t = ay_{t-1} + v_t$$

where v_t was distributed as in (2), with $\epsilon = 0.1$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 100$. Note that a did not vary with time in this experiment, and that its value was fixed to 0.99. A random walk was used as the process equation to impose forgetting of measurements, i.e.,

$$a_t = a_{t-1} + u_t$$

where u_t was zero mean Gaussian with variance $\sigma_{u_t}^2$ chosen according to (12) with forgetting factor $\lambda = 0.9999$. The number of particles was $M = 2000$. For comparison purposes, we applied a recursive least-squares (RLS) algorithm whose forgetting factor was also $\lambda = 0.9999$.[1] One particular representative simulation is shown in Figure 1. Note that a was tracked more accurately using the particle filter algorithm. Similar observations were made in most simulations.

With data generated by this model, we compared the performances of the particle filter and the RLS for various number of particles. The methods were compared by their MSE's averaged over 20 realizations. The results are shown in Figure 2. It is interesting to observe that for $M = 50$ and $M = 100$, the particle filter had worse performance than the RLS filter. As expected, as the number of particles increased, the performance of the particle filter improved considerably.

In Figure 3, we present the evolution of the instantaneous mean-square errors as a function of time of the particle filtering and the RLS methods. The instantaneous mean-square errors were obtained from 20 realizations, and

$$\mathrm{MSE}_i(t) = \sum_{j=1}^{20} (\hat{a}_{j,t} - a_t)^2$$

[1] It should be noted that the RLS algorithm is not based on any probabilistic assumptions, and that it is computationally much less intensive than the particle filtering algorithm.

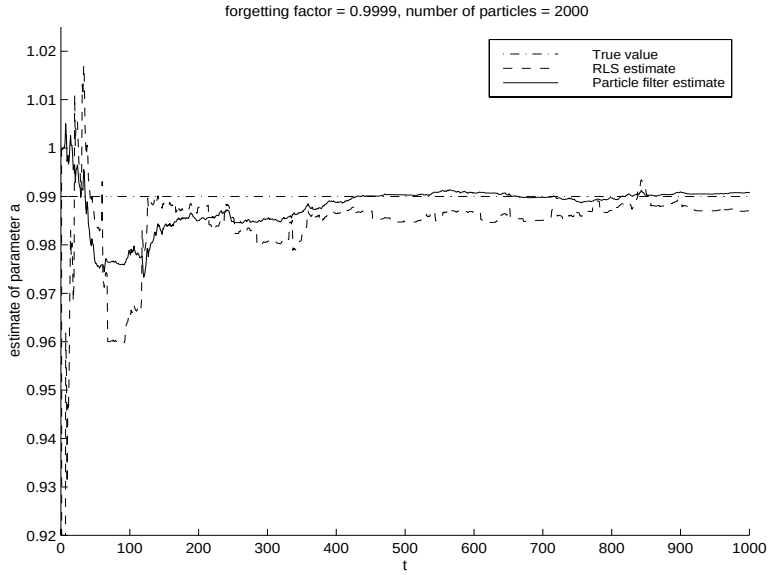


Fig. 1. Estimation of an autoregressive parameter a using the RLS and particle filtering methods. The parameter a was fixed and was equal to 0.99.

where $\hat{a}_{j,t}$ is the estimate of a_t in the $j-$th realization. For the particle filter we used $M = 2000$ particles, and $\lambda = 0.9999$. Clearly, the particle filter performed better. It is not surprising that the largest errors occur at the beginning, since there the methods have little prior knowledge of the true value of the parameter a.

In the next experiment, the noise was Laplacian. There, the parameter α was varied and had values 10, 2, and 1. In Figure 4, we present the MSE's of the particle filter and the RLS estimate averaged over 20 realizations. The particle filter clearly outperformed the RLS for all values of α.

The results of the first experiment with time-varying AR parameters are shown in Figure 5. There, a was attributed a piecewise changing behavior where it jumped from 0.99 to 0.95 at the time instant $t = 1001$, and the driving noise was a mixture Gaussian as in the first experiment. The forgetting factor λ was 0.95. Note that both the RLS and the particle filter follow the jump. However, the particle filter tracks it with higher accuracy and lower variation. Note also that the variation in the estimates in this experiment is much higher since the chosen forgetting factor was much smaller.

Statistical results of this experiment are shown in Figure 6. The figure shows the MSE's of the particle filter and the RLS method averaged over 20 realizations as functions of time. The particle filter outperformed the RLS significantly.

The experiment was repeated for a jump of a from 0.99 to -0.99 at $t = 1001$. Two different values of forgetting factors were used, $\lambda = 0.99$ and $\lambda = 0.95$, and the number of particles was kept at $M = 2000$. In Figures 7 and 8, we plotted MSE(t) obtained from 20 realizations. It is obvious from the figures that the performance of the particle filter was not good for $\lambda = 0.95$. The main reason for this degradation is the importance function of the particle filter. The prior importance function does not expect a change at that time because it does not use observations for generating particles. As a result, the particles at $t = 1001$ are generated around the values of $a_{1000}^{(m)}$, which are all far away from the actual value of a. Moreover, it took the particle filter more than 700 samples to "regroup," and that is a consequence of the relatively high value of the forgetting factor. When this value was decreased to $\lambda = 0.9$, the recovery of the particle filter was much shorter. Note that the price for improvement was a larger MSE during the periods of time when a was constant.

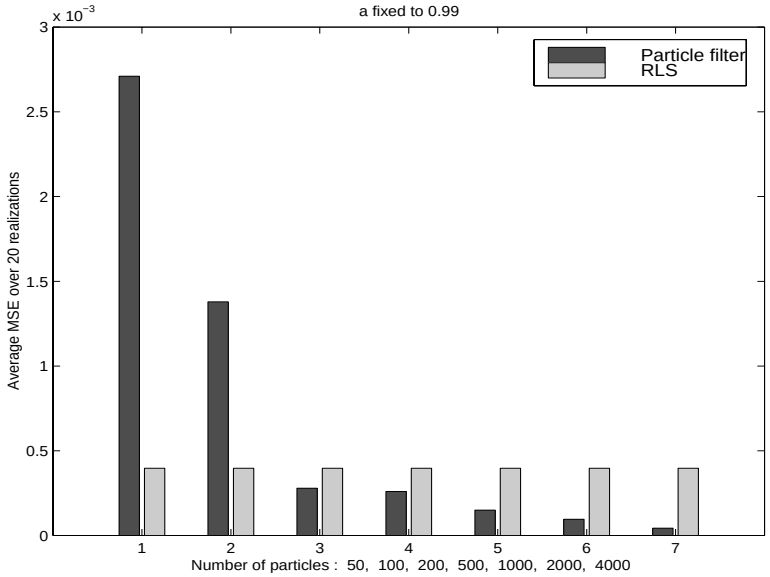


Fig. 2. Mean-square error of the particle filter and the RLS method averaged over 20 realizations. The driving noise was a Gaussian mixture.

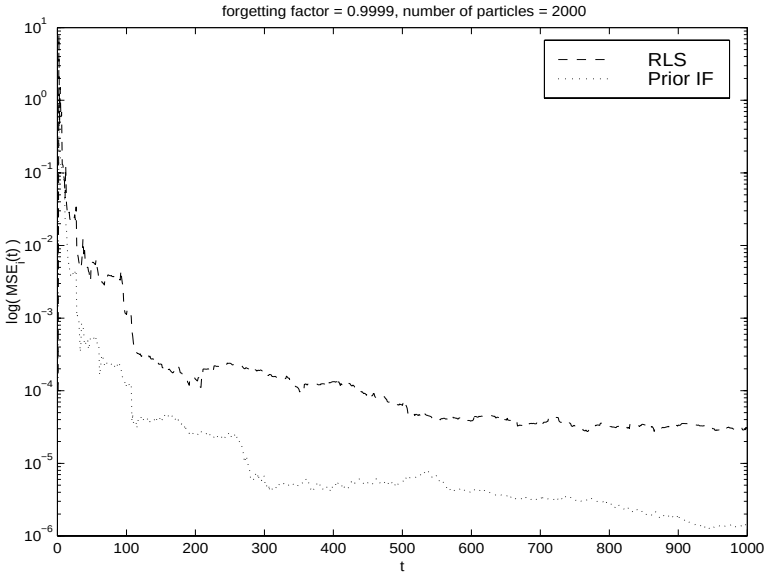


Fig. 3. Evolution of the $\log(\mathrm{MSE}_i(t))$ of the particle filter and the RLS method.

One can enhance the performance of the particle filter by choosing an importance function which explores the parameter space of a better.

In another experiment we generated data with higher order AR models. In particular, the data were obtained by

$$\begin{aligned} y_t &= -0.7348y_{t-1} - 1.8820y_{t-2} - 0.7057 - y_{t-3} - 0.8851y_{t-4} + v_t, && t = 1, 2, \cdots, 500 \\ y_t &= 1.352y_{t-1} - 1.338y_{t-2} + 0.662 + y_{t-3} - 0.240y_{t-4} + v_t, && t = 501, 502, \cdots, 1000 \\ y_t &= 0.37y_{t-1} + 0.56y_{t-2} + v_t, && t = 1001, 1002, \cdots, 1500 \end{aligned}$$

The driving noise was a Gaussian mixture with the same parameters as in the first experiment. The tracking of the parameters by the particle filter and the RLS method from one realization is

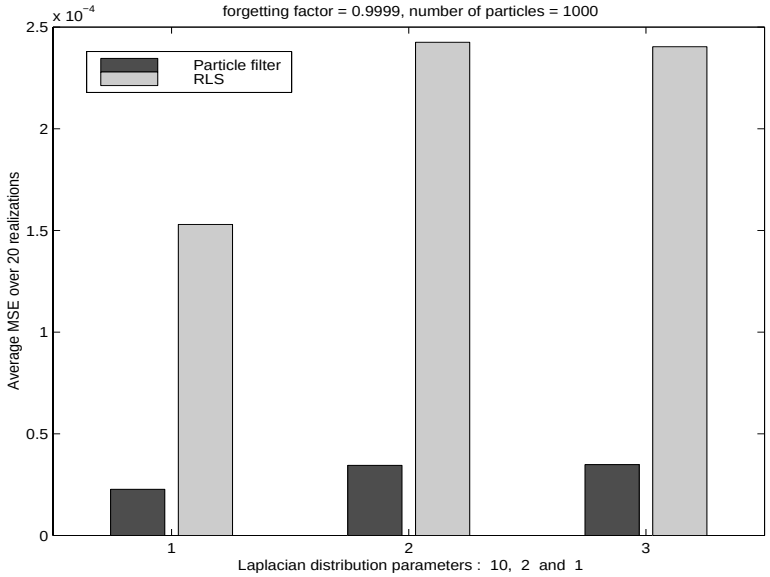


Fig. 4. Mean-square error of the particle filter and the RLS method averaged over 20 realizations. The driving noise was Laplacian.

shown in Figure 8. The number of particles was $M = 2000$ and the forgetting factor $\lambda = 0.9$. In Figure 9, we display the MSE errors of the two methods as functions of time.

Another statistical comparison between the two methods is shown in Figure 10. There we see the average MSE's of the methods presented separately for each parameter and for various forgetting factors. The number of particles $M = 8000$. The particle filter performed better for a_{2t}, a_{3t}, and a_{4t}, but worse for a_{1t}. A reason for the inferior performance of the particle filter in tracking a_{1t} is perhaps due to the big change of values of a_{1t}, which requires smaller forgetting factor than the one used. More importantly, with better importance function the tracking performance of a_{1t} can also be better. Such function, would generate more particles in the region of the new values of the parameters, and thereby would produce a more accurate approximation of their posterior density.

VII. Conclusions

We have presented a method for tracking the parameters of a time-varying AR process which is driven by a non-Gaussian noise. The function that models the variation of the model parameters is unknown. The estimation is carried out by particle filters which produce samples and weights that approximate required densities. The state equation that models the parameter changes with time is a random walk model, which implies discounting of old measurements. In the simulations, the parameters of the process are piecewise constant where the instants of their changes are unknown. The piecewise model is not by any means a restriction imposed by the method, but was used for convenience. Simulation results were presented. The requirement of knowing the noise parameters that drive the AR process can readily be removed [**?**].

References

[1] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximation," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, 1972.

[2] B. D. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, New Jersey, 1979.

[3] E. R. Beadle and P. M. Djurić, "A fast weighted Bayesian bootstrap filter for nonlinear model state estimation," *IEEE Transactions on Aerospace and Electronics Systems*, vol. 33, no. 1, pp. 338–342, 1997.

[4] J. Carpenter, P. Clifford, and P. Fearnhead, "An improved particle filter for non-linear problems," *IEE Proceedings - F: Radar, Sonar and Navigation*, vol. 146, pp. 2–7, 1999.

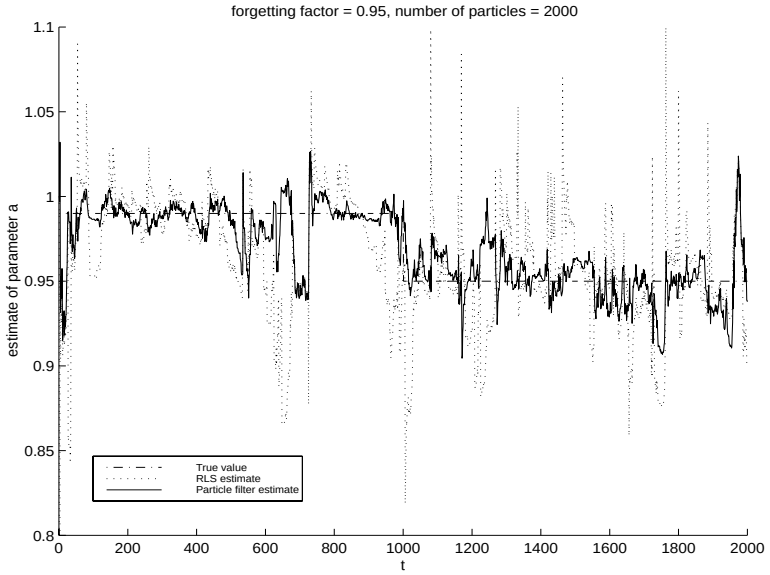


Fig. 5. Tracking performance of the piecewise constant AR parameter a_t with a jump from 0.99 to 0.95 at $t = 1001$.

[5] R. Charbonnier, M. Barlaud, G. Alengrin, and J. Menez, "Results on AR modeling of nonstationary signals," *Signal Processing*, vol. 12, no. 2, pp. 143–151, 1987.

[6] P. M. Djurić, J. Kotecha, J.-Y. Tourneret, and S. Lesage, "Adaptive signal processing by particle filters and discounting of old measurements," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, 2001.

[7] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001.

[8] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, pp. 197–208, 2000.

[9] A. Doucet, S. J. Godsill, and M. West, "Monte Carlo filtering and smoothing with application to time-varying spectral estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Istanbul, Turkey, 2000.

[10] K. B. Eom, "Time-varying autoregressive modeling of hrr radar signatures," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 3, pp. 974–988, 1999.

[11] K. Murano et al., "Echo cancellation and applications," *IEEE Transactions on Communications*, vol. 28, pp. 49–55, 1990.

[12] S. Frühwirth-Schnatter, "Data augmentation and dynamic linear models," *Journal of Time Series Analysis*, vol. 15, pp. 183–202, 1994.

[13] J. Geweke, "Bayesian inference in econometrics models using Monte Carlo integration," *Econometrica*, vol. 57, pp. 1317–1339, 1989.

[14] S. Godsill and T. Clapp, "Improvement strategies for Monte Carlo particle filters," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer, 2001.

[15] S. Godsill and P. Rayner, *Digital Audio Restoration - A Statistical Model Based Approach*, Springer, New York, 1998.

[16] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings-F*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[17] Y. Grenier, "Time-dependent arma modeling of nonstationary signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-31, pp. 899–911, 1983.

[18] M. G. Hall, A. V. Oppenheim, and A. D. Wilsky, "Time varying parametric modeling of speech," *Signal Processing*, vol. 5, pp. 276–285, 1983.

[19] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, Wiley, New York, 1996.

[20] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, third edition, 1996.

[21] S. Haykin, Ed., *Unsupervised Adaptive Filtering: Blind Deconvolution*, vol. II of *Adaptive and Learning Systems for Signal Processing , Communications, and Control*, New York, 2000.

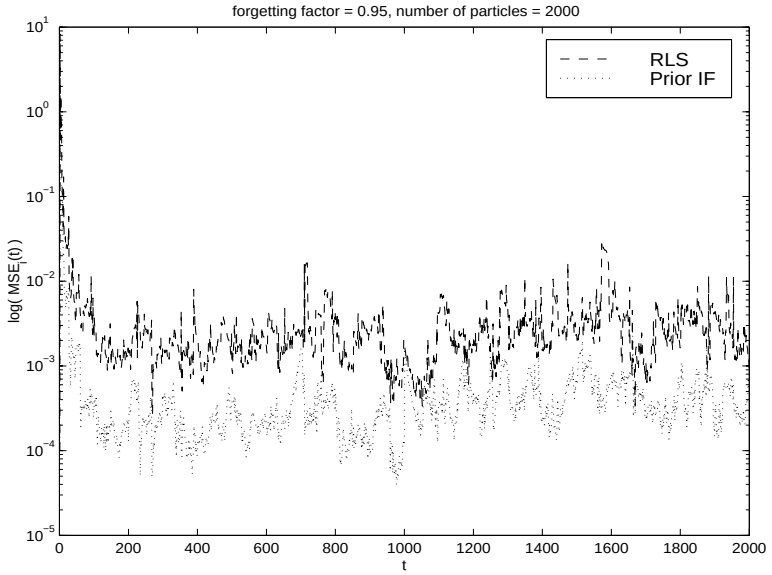


Fig. 6. Evolution of the $\log(\mathrm{MSE}_i(t))$ of the particle filter and the RLS method. The driving noise was mixture Gaussian.

[22] S. Haykin, Ed., *Unsupervised Adaptive Filtering: Blind Source Separation*, vol. I of *Adaptive and Learning Systems for Signal Processing , Communications, and Control*, Wiley Interscience, New York, 2000.

[23] P. W. Howells, "Explorations iun fixed and adaptive resolution at ge and surc," *IEEE Transactions on AP*, vol. AP-24, pp. 575–584, 1975.

[24] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[25] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation and covariances in filters and estimator," *IEEE Transactions on Automatic Control*, , no. 3, pp. 477–482, 2000.

[26] S. M. Kay, *Modern Spectral Estimation*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[27] G. Kitagawa, "Non-Gaussian state-space modeling of nonstationary time series," *Journal of the American Statistical Association*, vol. 82, pp. 1032–1063, 1987.

[28] L. A. Liporace, "Linear estimation of nonstationary signals," *Journal of the Acoustical Society of America*, vol. 58, pp. 1288–1295, 1975.

[29] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Stastistical Association*, vol. 93, no. 443, pp. 1032–1044, September 1998.

[30] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, The MIT Press, Cambridge, MA, 1983.

[31] R. W. Lucky, "Automatic equalization for digital communications," *Bell System Technical Journal*, vol. 44, pp. 547–588, 1965.

[32] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, 3 edition, 1995.

[33] T. S. Rao, "The fitting of nonstationary time serties models with time-dependent parameters," *Journal of the Royal Statistical Society*, vol. Series B, 32, no. 2, pp. 312–322, 1970.

[34] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Magazine*, vol. 11, pp. 18–60, 1994.

[35] D. Sengupta and S. Kay, "Effcient estimation of parameters for non-Gaussian autoregressive processes," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, pp. 785–794, 1989.

[36] S. M. Verbout, J. M. Ooi, J. T. Ludwig, and A. V. Oppenheim, "Parameter estimation for autoregressive Gaussian-mixture processes: The EMAX algorithm," *IEEE Transactions on Signal Processing*, pp. 2744–2756, 1998.

[37] M. West, *Bayesian Forecasting and Dynamic Models*, Springer Verlag, New York, 1997.

[38] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

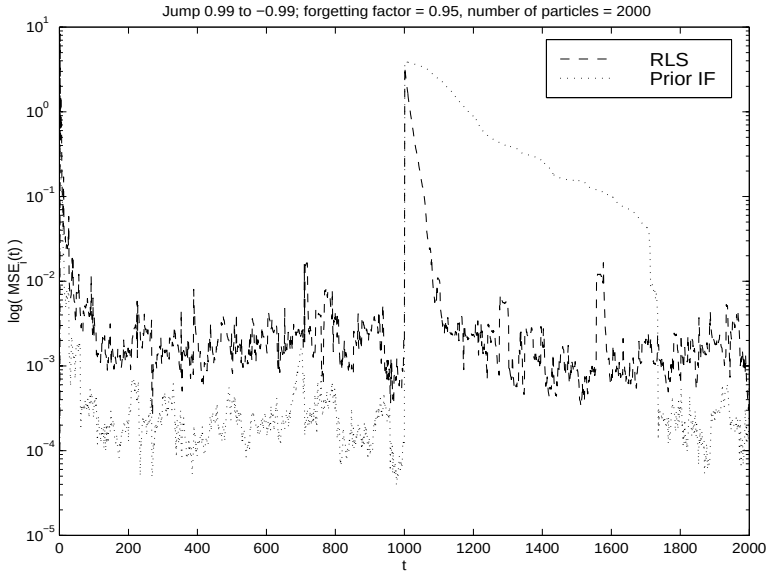


Fig. 7. Evolution of the $\log(\mathrm{MSE}_i(t))$ of the particle filter and the RLS method. The forgetting parameter was $\lambda = 0.95$, and the number of particles was $M = 2000$.

Petar M. Djurić received his B.S. and M.S. degrees in electrical engineering from the University of Belgrade, Yugoslavia, in 1981 and 1986, respectively, and his Ph.D. degree in electrical engineering from the University of Rhode Island, in 1990. From 1981 to 1986 he was Research Associate with the Institute of Nuclear Sciences, Vinča, Belgrade, Yugoslavia. Since 1990 he has been with Stony Brook University, where he is Professor in the Department of Electrical and Computer Engineering. He works in the area of statistical signal processing, and his primary interests are in the theory of modeling, detection, estimation, and time series analysis and its application to a wide variety of disciplines, including telecommunications, bio-medicine, and power engineering. Prof. Djurić has served on numerous Technical Committees for the IEEE and SPIE, and has been invited to lecture at universities in the US and overseas. He was Associate Editor of the IEEE Transactions on Signal Processing, and currently he is the Treasurer of the IEEE Signal Processing Conference Board. He is also Vice Chair of the IEEE Signal Processing Society Committee on Signal Processing - Theory and Methods, and a Member of the American Statistical Association and the International Society for Bayesian Analysis.

Jayesh H. Kotecha received a B.E in electronics and telecommunications from the College of Engineering, Pune, India in 1995, and M.S and Ph.D degrees in electrical engineering from the State University of New York at Stony Brook in 1996 and 2001, respectively. Since January 2002, he has been with the University of Wisconsin at Madison as a post-doctoral researcher. Dr. Kotecha's research interests are chiefly in the fields of communications, signal processing and information theory. In particular, he works on problems related to statistical signal processing, adaptive signal processing and particle filters, physical layer aspects in communications like channel modelling, estimation, detection, and coding in the contexts of single antenna and multi-input multi-output systems.

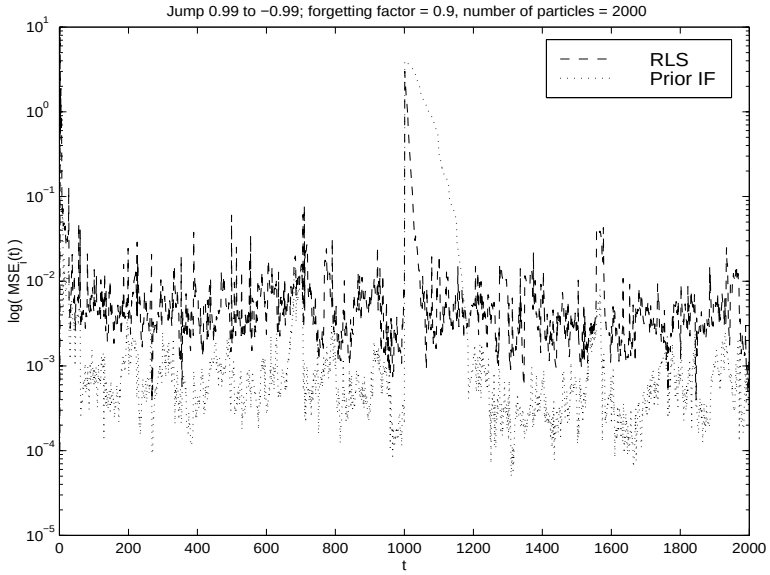


Fig. 8. Evolution of the log(MSE(t)) of the particle filter and the RLS method. The forgetting parameter was $\lambda = 0.9$, and the number of particles was $M = 2000$.

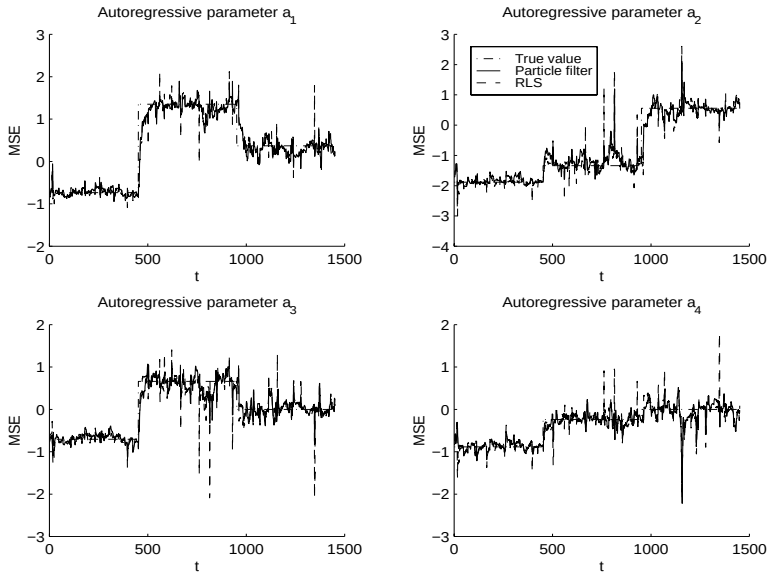
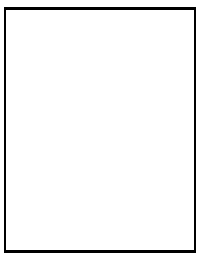


Fig. 9. Tracking of the AR parameters, where the models change at $t = 501$ and $t = 1001$.



Fabien Esteve received the Engineer degree in Electrical Engineering and Signal Processing in 2002 from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT), Toulouse, France.

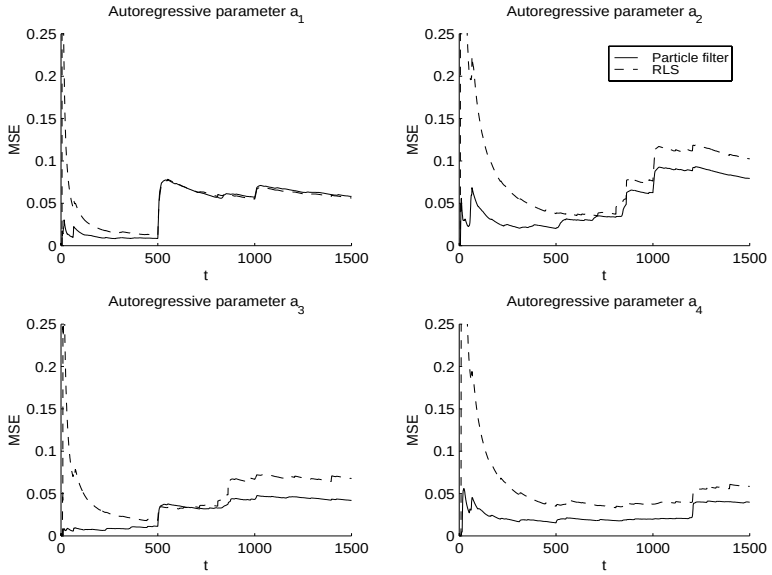


Fig. 10. Evolution of the MSE of each of the AR parameters as a function of time.

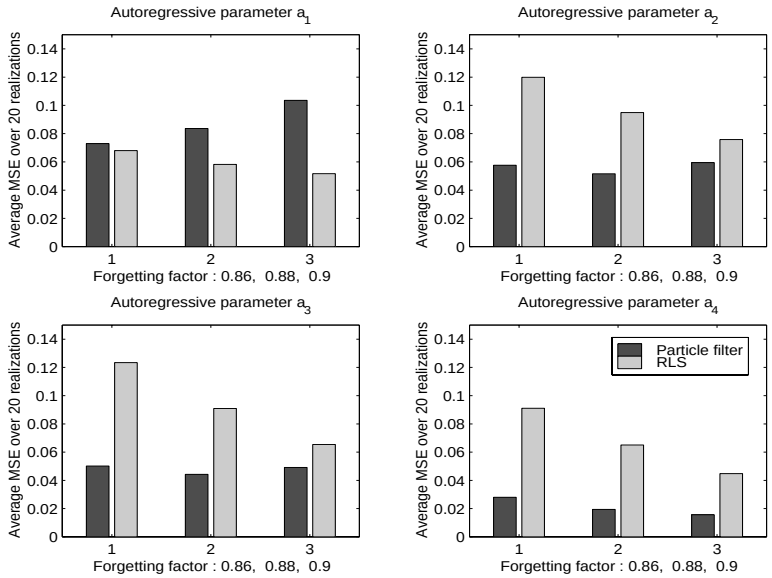


Fig. 11. Mean-square error of each of the AR parameters produced by the particle filter and the RLS method averaged over 20 realizations. The driving noise was mixture Gaussian.

Etienne Perret was born in Albertville, France in 1979. He received his Engineer degree in Electrical Engineering and Signal Processing and Masters degree in Microwave and Optical Telecommunications from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT), Toulouse, France. He is currently working toward his Ph.D. degree at ENSEEIHT. His research interests include digital signal processing and modeling of microwave circuits.