On the preestimation technique and its application to identification of nonstationary systems

Maciej Niedźwiecki, Artur Gańcza and Marcin Ciołek

Abstract— The problem of noncausal identification of a nonstationary stochastic FIR (finite impulse response) system is reformulated, and solved, as a problem of smoothing of preestimated parameter trajectories. Three approaches to preestimation are critically analyzed and compared. It is shown that optimization of the smoothing operation can be performed adaptively using the parallel estimation technique. The new approach is computationally attractive and yields estimation results that are comparable or better than those provided by the state-of-the-art local basis function approach and the multi-resolution wavelet approach.

I. Introduction

Preestimates are rough, unbiased but very noisy, estimates of time-varying parameters of an identified nonstationary system. Due to their unbiasedness property preestimates provide interesting insights into the structure of system time-variation, and they do so without making any assumptions about the functional form and speed of parameter variation. On the qualitative level the idea of preestimation, proposed in [1], is similar to the concept of preperiodogram in nonparametric nonstationary spectrum estimation [2], [3], [4], [5], and to the concept of maximum bandwidth estimators considered in [6]. In all cases, due to large variability, the statistically reliable results can be obtained by smoothing the preestimated quantities. Interestingly (and somewhat surprisingly) the parameter tracking capabilities of the smoothed preestimates are comparable with those provided by the state-of-the-art (and computationally much more involved) local basis function (LBF) estimators proposed in [7]. When the impulse response of the linear smoothing filter is chosen appropriately, the filtered preestimates are almost indistinguishable from the LBF estimates obtained as a result of local approximation of parameter trajectories by linear combinations of basis functions [8].

In the current contribution we compare two variants of preestimation, applicable to identification of a FIR system with a time-varying impulse response. Then, we propose a new preestimation formula, which can be regarded as a combination of the first two ones, and which yields even better results than those reported in [7].

II. System

We will consider a time-varying FIR system governed by

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + e(t) \tag{1}$$

where $t = \ldots, -1, 0, 1, \ldots$ denotes discrete (normalized) time, $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_n(t)]^{\mathrm{T}}$ is the vector of unknown time-varying parameters (coefficients of the impulse response of the identified system), $\boldsymbol{\varphi}(t) = [u(t-1), \ldots, u(t-n)]^{\mathrm{T}}$ denotes the regression vector made up of the previous samples of the input signal $u(t)$, and $e(t)$ denotes white measurement noise.

One of the possible applications, which admits such problem formulation, is identification of multi-path (e.g.,mobile radio) channels. In this particular case the regression vector $\boldsymbol{\varphi}(t)$ is made up of past input (transmitted) symbols, $y(t)$ is the received baseband signal and $\boldsymbol{\theta}(t)$ is the vector of time varying impulse response coefficients of the channel. The values of FIR coefficients depend on the strength of "natural reflectors" and their time variation is caused by the receiver motion [9] - [10].

To obtain analytical results we will assume that

(A1) $\{u(t)\}$ is a zero-mean wide sense stationary Gaussian sequence, persistently exciting of order at least n, with an exponentially decaying autocorrelation function.

(A2) $\{e(t)\}$, independent of $\{u(t)\}$, is a sequence of zero-mean independent and identically distributed random variables.

(A3) $\{\boldsymbol{\theta}(t)\}$ is a uniformly bounded sequence, independent of $\{u(t)\}$ and $\{e(t)\}$.

III. Direct preestimates

Suppose that the covariance matrix of the regression vector $\boldsymbol{\Phi}_0 = \mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)] > 0$ is known. Then the natural form of a preestimate of $\boldsymbol{\theta}(t)$ is given by

$$\boldsymbol{\theta}^{\sharp}(t) = \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)y(t). \tag{2}$$

Actually, combining (1) and (2), one obtains

$$\boldsymbol{\theta}^{\sharp}(t) = \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)e(t) \tag{3}$$

which leads to

$$\begin{aligned} \mathrm{E}[\boldsymbol{\theta}^{\sharp}(t)] &= \boldsymbol{\Phi}_0^{-1}\mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]\boldsymbol{\theta}(t) \\ &+ \boldsymbol{\Phi}_0^{-1}\mathrm{E}[\boldsymbol{\varphi}(t)e(t)] = \boldsymbol{\theta}(t) \end{aligned} \tag{4}$$

where, here and later, the expectation is over $\Omega(t) = \{\boldsymbol{\varphi}(t), e(t)\}$.

According to (4), $\boldsymbol{\theta}^\sharp(t)$ is an unbiased estimator of $\boldsymbol{\theta}(t)$. To evaluate the covariance matrix of $\boldsymbol{\theta}^\sharp(t)$, note that

$$\Delta\boldsymbol{\theta}^\sharp(t) = \boldsymbol{\theta}^\sharp(t) - \boldsymbol{\theta}(t) = [\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) - \mathbf{I}_n]\boldsymbol{\theta}(t)$$
$$+ \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)e(t) \quad (5)$$

and, under assumption (A2),

$$\mathrm{cov}[\boldsymbol{\theta}^\sharp(t)] = \mathrm{E}\{\Delta\boldsymbol{\theta}^\sharp(t)[\Delta\boldsymbol{\theta}^\sharp(t)]^{\mathrm{T}}\}$$
$$= \mathrm{E}\{[\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) - \mathbf{I}_n]\boldsymbol{\theta}(t)\boldsymbol{\theta}^{\mathrm{T}}(t)[\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) - \mathbf{I}_n]\}$$
$$+ \sigma_e^2\mathrm{E}\{\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\Phi}_0^{-1}\}$$
$$= \boldsymbol{\Phi}_0^{-1}\mathrm{E}\{\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t)\boldsymbol{\theta}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\}\boldsymbol{\Phi}_0^{-1}$$
$$- \boldsymbol{\theta}(t)\boldsymbol{\theta}^{\mathrm{T}}(t) + \sigma_e^2\boldsymbol{\Phi}_0^{-1}. \quad (6)$$

Using the well-known properties of higher order moments of Gaussian variables [11], [12], one can easily derive the following formula which holds for zero-mean jointly normally distributed random vectors $\mathbf{x}, \mathbf{y}, \mathbf{w}, \mathbf{v}$ and a symmetric matrix $\mathbf{A}$

$$\mathrm{E}[\mathbf{x}\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{w}\mathbf{v}^{\mathrm{T}}] = \mathrm{E}[\mathbf{x}\mathbf{y}^{\mathrm{T}}]\mathbf{A}\mathrm{E}[\mathbf{w}\mathbf{v}^{\mathrm{T}}] + \mathrm{E}[\mathbf{x}\mathbf{w}^{\mathrm{T}}]\mathbf{A}\mathrm{E}[\mathbf{y}\mathbf{v}^{\mathrm{T}}]$$
$$+ \mathrm{E}[\mathbf{x}\mathbf{v}^{\mathrm{T}}]\mathrm{E}[\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{w}]. \quad (7)$$

Using this formula, one obtains

$$\mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t)\boldsymbol{\theta}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)] = 2\boldsymbol{\Phi}_0\boldsymbol{\theta}(t)\boldsymbol{\theta}^{\mathrm{T}}(t)\boldsymbol{\Phi}_0$$
$$+ \boldsymbol{\Phi}_0\boldsymbol{\theta}^{\mathrm{T}}(t)\boldsymbol{\Phi}_0\boldsymbol{\theta}(t). \quad (8)$$

Combining (6) and (8), one arrives at

$$\mathrm{cov}[\boldsymbol{\theta}^\sharp(t)] = \boldsymbol{\theta}(t)\boldsymbol{\theta}^{\mathrm{T}}(t) + \boldsymbol{\theta}^{\mathrm{T}}(t)\boldsymbol{\Phi}_0\boldsymbol{\theta}(t)\boldsymbol{\Phi}_0^{-1} + \sigma_e^2\boldsymbol{\Phi}_0^{-1}. \quad (9)$$

When the covariance matrix $\boldsymbol{\Phi}_0$ is not known *a priori*, which is a more realistic assumption, one can replace it in (2) with the following local, exponentially weighted estimate (to simplify analytical results, we will assume that an infinite observation history is available at the instant t)

$$\widehat{\boldsymbol{\Phi}}_\lambda(t) = \frac{1}{N_\lambda}\mathbf{R}_\lambda(t), \quad N_\lambda = \sum_{i=0}^{\infty}\lambda^i = \frac{1}{1-\lambda}$$
$$\mathbf{R}_\lambda(t) = \sum_{i=0}^{\infty}\lambda^i\boldsymbol{\varphi}(t-i)\boldsymbol{\varphi}^{\mathrm{T}}(t-i) \quad (10)$$

where $\lambda, 0 < \lambda < 1$ denotes the so-called forgetting constant and N_λ is the effective width of the exponential window. Replacing $\boldsymbol{\Phi}_0$ with $\widehat{\boldsymbol{\Phi}}_\lambda(t)$ in (2), one obtains the following preestimation formula

$$\boldsymbol{\theta}^\dagger(t) = \widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)y(t). \quad (11)$$

Note that $\widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t) = N_\lambda\mathbf{P}_\lambda(t)$, where $\mathbf{P}_\lambda(t) = \mathbf{R}_\lambda^{-1}(t)$ is a recursively computable matrix [12]

$$\mathbf{P}_\lambda(t) = \frac{1}{\lambda}\left[\mathbf{P}_\lambda(t-1) - \frac{\mathbf{P}_\lambda(t-1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{P}_\lambda(t-1)}{\lambda + \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{P}_\lambda(t-1)\boldsymbol{\varphi}(t)}\right]. \quad (12)$$

One can show that under assumption (A1) it holds that [13]

$$\lim_{\lambda \mapsto 1}\widehat{\boldsymbol{\Phi}}_\lambda(t) = \boldsymbol{\Phi}_0 \quad (13)$$

where the convergence takes place in the mean square sense. This justifies the following approximation valid for sufficiently large values of N_λ

$$\widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t) \cong \boldsymbol{\Phi}_0^{-1}. \quad (14)$$

Combining (1) and (11), one arrives at

$$\mathrm{E}[\boldsymbol{\theta}^\dagger(t)] = \mathrm{E}\left[\widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right]\boldsymbol{\theta}(t). \quad (15)$$

Since

$$\mathrm{E}\left[\widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right] \cong \boldsymbol{\Phi}_0^{-1}\mathrm{E}\left[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right]$$
$$\cong \mathbf{I}_n \quad (16)$$

one obtains

$$\mathrm{E}[\boldsymbol{\theta}^\dagger(t)] \cong \boldsymbol{\theta}(t). \quad (17)$$

Using similar arguments, one can show that

$$\Delta\boldsymbol{\theta}^\dagger(t) = \boldsymbol{\theta}^\dagger(t) - \boldsymbol{\theta}(t)$$
$$= [\widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) - \mathbf{I}_n]\boldsymbol{\theta}(t) + \widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)e(t)$$
$$\cong [\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) - \mathbf{I}_n]\boldsymbol{\theta}(t) + \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)e(t) \quad (18)$$

and hence

$$\mathrm{cov}[\boldsymbol{\theta}^\dagger(t)] \cong \mathrm{cov}[\boldsymbol{\theta}^\sharp(t)]. \quad (19)$$

Fig. 1 shows preestimation results obtained for a nonstationary two-tap FIR system governed by

$$y(t) = \theta_1(t)u(t-1) + \theta_2(t)u(t-2) + e(t) \quad (20)$$

where the input signal was governed by the first-order autoregressive equation $u(t) = 0.8u(t-1)+v(t)$, $\mathrm{var}[v(t)] = 1$, and $\{v(t)\}$ denotes white noise independent of $\{e(t)\}$. Parameter $\theta_1(t)$ was changing in a chirp-like way, and parameter $\theta_2(t)$ was piecewise constant. The forgetting constant λ was set to 0.9.

IV. INDIRECT PREESTIMATES

The indirect scheme, proposed in [7], is based on inverse filtering of exponentially weighted least squares (EWLS) estimates of $\boldsymbol{\theta}(t)$, which in the steady state take the form

$$\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) = \mathbf{R}_\lambda^{-1}(t)\mathbf{r}_\lambda(t) \quad (21)$$

where $\mathbf{R}_\lambda(t)$ is a regression matrix given by (10) and

$$\mathbf{r}_\lambda(t) = \sum_{i=0}^{\infty}\lambda^i\boldsymbol{\varphi}(t-i)y(t-i). \quad (22)$$

The EWLS estimates can be computed recursively using the following algorithm [12]

$$\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) = \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1) + \mathbf{R}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)\varepsilon(t) \quad (23)$$

where

$$\varepsilon(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)$$

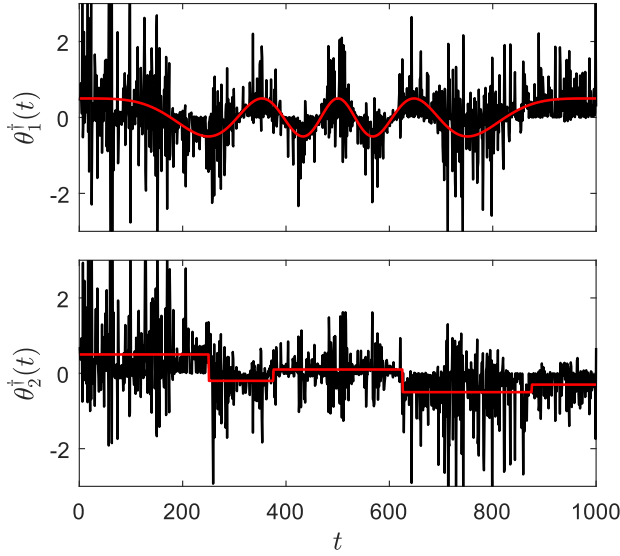is a one-step-ahead prediction error.

Fig. 1: Preestimated parameter trajectories of a two-tap FIR system ($\lambda = 0.9$). Direct preestimates (black lines) are superimposed on true parameter trajectories (red lines).
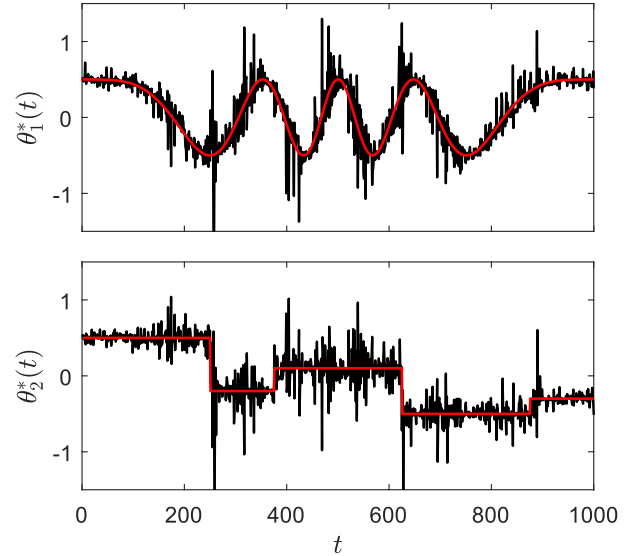

Fig. 2: Preestimated parameter trajectories of a two-tap FIR system ($\lambda = 0.9$). Indirect preestimates (black lines) are superimposed on true parameter trajectories (red lines).

The preestimation formula proposed in [7] has the form

$$
\boldsymbol{\theta}^*(t) = \frac{\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) - \lambda\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)}{1-\lambda}
$$
$$
= \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1) + \widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)\varepsilon(t). \quad (24)
$$

Combining (24) with (1), one obtains

$$
\boldsymbol{\theta}^*(t) = \widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t)
$$
$$
+ [\mathbf{I}_n - \widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)
$$
$$
+ \widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(e)e(t) \quad (25)
$$

which leads to

$$
\mathrm{E}[\boldsymbol{\theta}^*(t)] \cong \boldsymbol{\theta}(t) \quad (26)
$$

and

$$
\Delta\boldsymbol{\theta}^*(t) = \boldsymbol{\theta}^*(t) - \boldsymbol{\theta}(t)
$$
$$
= [\widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) - \mathbf{I}_n][\boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)]
$$
$$
+ \widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)e(t). \quad (27)
$$

It is instructive to compare the estimation error expressions derived for the direct and indirect preestimates. Note that the first term on the right hand side of (18) depends on $\boldsymbol{\theta}(t)$, while the analogous term in (27) depends on $\boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)$. Since typically $||\boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)|| \ll ||\boldsymbol{\theta}(t)||$, one can expect variability of indirect preestimates to be much smaller than variability of direct preestimates. Simulation results, depicted in Fig. 2 (obtained for the same FIR system which was analyzed in Section 3), fully confirm this claim.

It was noticed that the best preestimation results can be obtained for small values of N_λ, such as $N_\lambda = 10$ ($\lambda = 0.9$) – even if the number of estimated coefficients is large. It is not difficult to explain this. Since $\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t)$ is a causal estimator, incorporating only past data samples $\{y(i), u(i), i \leq t\}$, when N_λ is increased the mean square deviation of $\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t)$ from $\boldsymbol{\theta}(t)$ becomes quickly dominated by the bias error caused mainly by the fact that the estimated parameter trajectory lags behind the true trajectory. As shown in [13], the resulting estimation delay is roughly equal to N_λ sampling intervals. Hence, to achieve a good bias-variance trade-off, the value of N_λ should be relatively small.

V. Postfiltering

Due to their very large variability, preestimates can be regarded only as intermediate estimates which should be further processed to trade-off the bias and variance components of the mean square parameter estimation error. The postfiltering scheme proposed in [7] is based on the local basis function approximation of the estimated parameter trajectory. Denote by $T_k(t) = [t-k, t+k]$ the local analysis interval of width $2k+1$, centered at t, and by

$$
G_{m|k} = \{g_{1|k}(i), \ldots, g_{m|k}(i), i \in I_k\}, \quad I_k = [-k, k] \quad (28)
$$

the set of m linearly independent functions of time, called basis functions – such as harmonic functions or powers of time. In the latter case

$$
g_{j|k}(i) = \left(\frac{i}{k}\right)^{j-1}, \quad j = 1, \ldots, m, \quad i \in I_k. \quad (29)
$$

Finally, denote by $w_k(i) \geq 0$, $i \in I_k$, $w_k(0) = 1$, the symmetric, bell-shaped window, such as the cosinusoidal window $w_k(i) = \cos\frac{\pi i}{2k}$, used to put more emphasis on data gathered at instants close to t.

Assume that the true parameter trajectory can be modeled as a linear combination of basis functions, and that the preestimated trajectory can be regarded as a true trajectory corrupted by additive noise. Applying the method of weighted least squares, one obtains the following expression for the local estimate of $\boldsymbol{\theta}(t)$, further referred to as fast local basis function (fLBF) estimate

$$\widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{fLBF}}(t) = \sum_{i=-k}^{k} h_{m|k}(i)\boldsymbol{\theta}^*(t+i) \qquad (30)$$

where $\{h_{m|k}(i), i \in I_k\}$ is the impulse response of the smoothing filter. This impulse response can be expressed in terms of the w-orthonormal basis set of $G_{m|k}$, namely

$$h_{m|k}(i) = w_k(i)\mathbf{f}_{m|k}^{\mathrm{T}}(0)\mathbf{f}_{m|k}(i), \quad i \in I_k \qquad (31)$$

where $\mathbf{f}_{m|k}(i) = [f_{1|k}(i), \ldots, f_{m|k}(i)]^{\mathrm{T}}$, and the functions $f_{1|k}(i), \ldots, f_{m|k}(i)$, $i \in I_k$, obtained by means of orthogonalization of $G_{m|k}$, obey

$$\sum_{i=-k}^{k} w_k(i)\mathbf{f}_{m|k}(i)\mathbf{f}_{m|k}^{\mathrm{T}}(i) = \mathbf{I}_m. \qquad (32)$$

The term "fast LBF" refers to the fact that the estimates obtained using the very simple formula (30) (convolution can be efficiently computed using the FFT-based procedure) are hardly distinguishable from the estimates yielded by a computationally much more demanding LBF approach (generalized Savitzky-Golay filtering) described in [8] – see Fig. 3. The asymptotic approximate equivalence of fLBF and LBF estimators was shown in [7].

To obtain satisfactory tracking results, i.e., to minimize the mean square parameter estimation error, the values of m (the number of basis functions) and k (the size of the analysis interval) should be chosen so as to match the type and speed of parameter variations. As shown in [8], increasing m and/or decreasing k, one decreases the bias component of MSE, but increases its variance component. Decreasing m and/or increasing k has the opposite effect. Since MSE is the sum of its bias and variance components, some sort of bias-variance compromise should be reached to guarantee good performance. The solution proposed in [7] is based on parallel estimation. In this case not one but several fLBF algorithms, equipped with different design parameters $m \in \mathcal{M}$, $k \in \mathcal{K}$, are run simultaneously and compared. At each time instant t only one of the estimates is selected, i.e., the estimated parameter trajectory has the form

$$\widehat{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\theta}}_{\widehat{m}(t)|\widehat{k}(t)}^{\mathrm{fLBF}}(t) \qquad (33)$$

where

$$\{\widehat{m}(t), \widehat{k}(t)\} = \arg \min_{\substack{m \in \mathcal{M} \\ k \in \mathcal{K}}} J_{m|k}(t) \qquad (34)$$

and $J_{m|k}(t)$ denotes the local decision statistic. Using the cross-validation principle [14], one can set

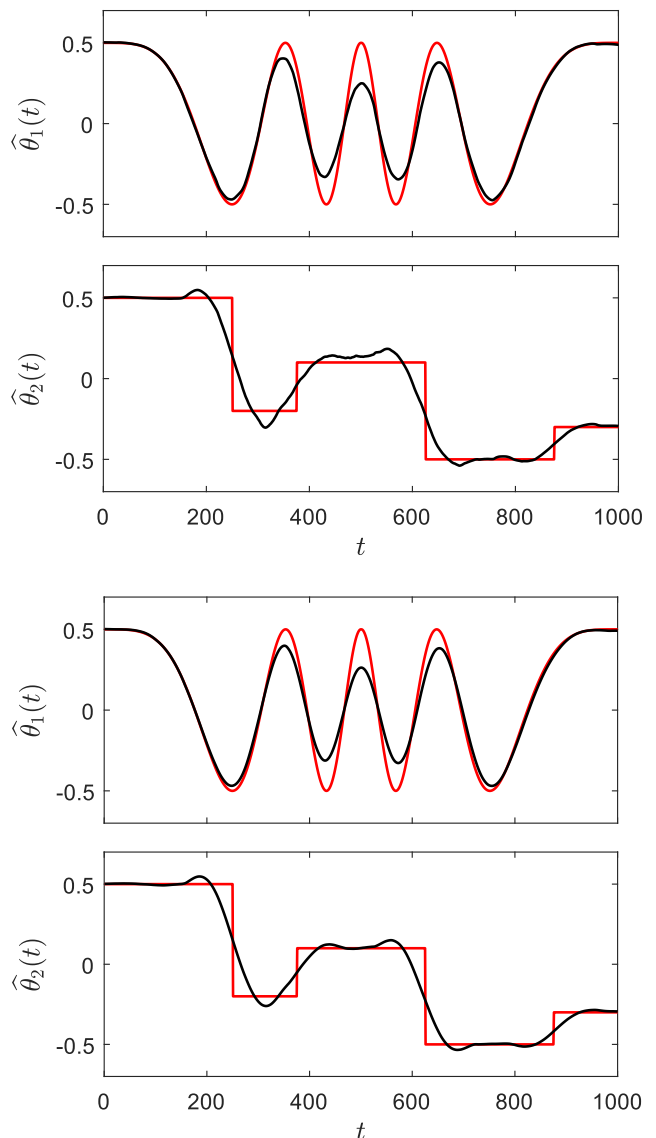$$J_{m|k}(t) = \sum_{i=-L}^{L} \eta_{m|k}^2(t+i) \qquad (35)$$

Fig. 3: Comparison of LBF estimates (two upper plots) and fLBF estimates (two lower plots) for $k = 100$, $m = 3$ and cosinusoidal window; fLBF estimates were obtained by means of smoothing indirect preestimates shown in Fig. 2. Estimated trajectories (black lines) are superimposed on true trajectories (red lines).

where $\eta_{m|k}(t)$ denotes deleted residual (leave-one-out interpolation error of the system output). If needed, the order n of the FIR system can be selected adaptively at the preestimation stage using the modified (localized) version of the Akaike's final prediction error (FPE) criterion [15].

VI. ENHANCED PREESTIMATES

The estimate of parameter trajectory $\widehat{\boldsymbol{\theta}}(t)$ obtained by competitive smoothing of indirect preestimates can be used to generate enhanced preestimates with reduced variability

$$\widetilde{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\theta}}(t) + \widehat{\boldsymbol{\Phi}}_{\lambda}^{-1}(t)\boldsymbol{\varphi}(t)[y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}(t)]. \qquad (36)$$
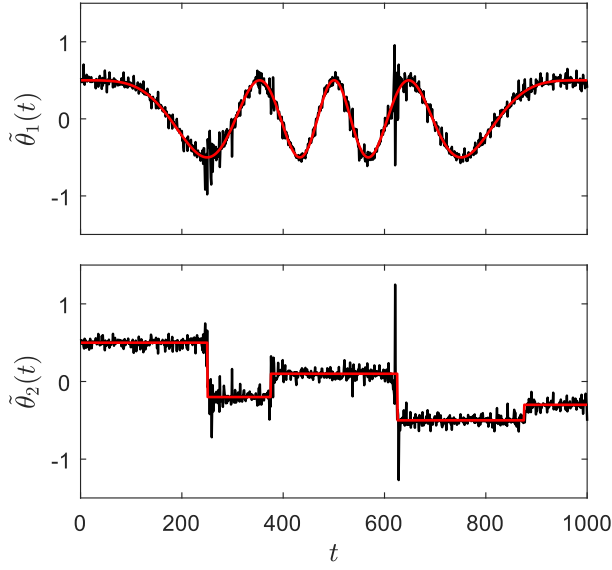
Fig. 4: Preestimated parameter trajectories of a two-tap FIR system ($\lambda = 0.9$). Enhanced preestimates (black lines) are superimposed on true parameter trajectories (red lines).
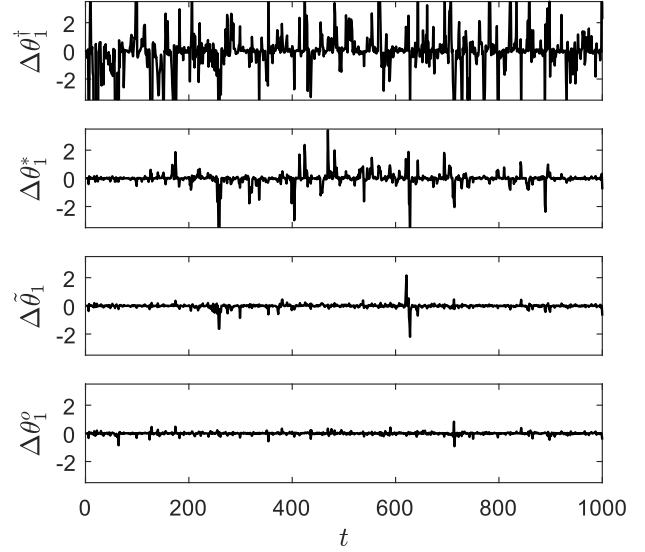

Fig. 5: Comparison of parameter estimation errors yielded by direct preestimates $\Delta\theta_1^\dagger(t)$, indirect preestimates $\Delta\theta_1^*(t)$ and enhanced preestimates $\Delta\widetilde{\theta}_1(t)$, with oracle preestimation errors $\Delta\theta_1^\circ(t)$.

In accordance with (1),

$$\widetilde{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t)$$
$$+ [\mathbf{I}_n - \widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]\widehat{\boldsymbol{\theta}}(t)$$
$$+ \widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)e(t) \qquad (37)$$

and hence [cf. (25) - (26)]

$$\mathrm{E}[\widetilde{\boldsymbol{\theta}}(t)] \cong \boldsymbol{\theta}(t). \qquad (38)$$

Furthermore, it holds that

$$\Delta\widetilde{\boldsymbol{\theta}}(t) = \widetilde{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)$$
$$= [\widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) - \mathbf{I}_n][\boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}(t)]$$
$$+ \widehat{\boldsymbol{\Phi}}_\lambda^{-1}(t)\boldsymbol{\varphi}(t)e(t). \qquad (39)$$

Note that (36) can be rewritten in the form

$$\widetilde{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\theta}}(t) + \boldsymbol{\theta}^\dagger(t)\left[1 - \frac{\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}(t)}{y(t)}\right] \qquad (40)$$

i.e., the enhanced preestimates can be regarded as a combination of direct preestimates $\boldsymbol{\theta}^\dagger(t)$ and smoothed indirect preestimates $\widehat{\boldsymbol{\theta}}(t)$.

It is instructive to compare (39) with (27). Since N_λ must be kept small to reduce the bias of the EWLS estimator incorporated in (24), under typical operating conditions it holds that $||\boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}(t)|| < ||\boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)||$. Consequently, the enhanced preestimates usually have smaller variability than the indirect preestimates $\boldsymbol{\theta}^*(t)$ – see Fig. 4.

Fig. 5 compares parameter estimation errors yielded by direct preestimates $\Delta\boldsymbol{\theta}^\dagger(t)$, indirect preestimates $\Delta\boldsymbol{\theta}^*(t)$ and enhanced preestimates $\Delta\widetilde{\boldsymbol{\theta}}(t)$, with the oracle (minimum attainable) preestimation errors defined as $\Delta\boldsymbol{\theta}^\circ(t) =$

$\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)e(t)$. All plots show preestimation errors observed – for the same realization of the input signal and measurement noise – for the parameter $\theta_1(t)$.

Enhanced preestimates can be further postprocessed. Using the competitive Savitzky-Golay [16] smoothing approach (described in Section 5) to denoise enhanced preestimates $\{\widetilde{\boldsymbol{\theta}}(t)\}$, one can achieve further reduction of the mean square parameter estimation error. Note that the resulting four-step procedure

1) evaluate indirect preestimates $\{\boldsymbol{\theta}^*(t)\}$
2) smooth $\{\boldsymbol{\theta}^*(t)\}$
3) evaluate enhanced preestimates $\{\widetilde{\boldsymbol{\theta}}(t)\}$
4) smooth $\{\widetilde{\boldsymbol{\theta}}(t)\}$

is still computationally very attractive compared with the original LBF approach (which requires inversion of $mn \times mn$ generalized regression matrices every time step t).

VII. SIMULATION RESULTS

To compare different estimation schemes described in the paper, the two-tap FIR system (20) was simulated for 3 rates of parameter variation and 2 signal-to-noise ratios. The simulated parameter trajectories were generated by sampling the same prototype analog parameter trajectories at different sampling rates. The resulting lengths of the simulation interval T_s were equal to 1000, 2000 and 4000 for fast, medium speed and slow changes, respectively (fast parameter variations are depicted in Figs. 1–4).

Table I shows the mean square parameter estimation errors obtained for 9 fLBF/LBF estimators corresponding to different choices of design parameters k (50, 100, 200) and m (1, 3, 5), and for the adaptive parallel estimation schemes based on cross-validation (A). All results were obtained by means

TABLE I: Mean square parameter estimation errors obtained for 9 fLBF/LBF estimators corresponding to different choices of design parameters k (50, 100, 200) and m (1, 3, 5), and for the adaptive parallel estimation schemes based on cross-validation (A). All averages were computed for 100 process realizations, 3 speeds of parameter variation and 2 average signal-to-noise ratios (SNR). The best results in each case are shown in boldface.

fast variations

Settings		Smoothed direct preestimates			Smoothed indirect preestimates			Smoothed enhanced preestimates			LBF estimates		
SNR	$k\backslash m$	1	3	5	1	3	5	1	3	5	1	3	5
15	50	3.15E-02	2.52E-02	3.36E-02	1.37E-02	5.35E-03	5.45E-03	1.35E-02	4.94E-03	4.74E-03	1.58E-02	5.12E-03	4.84E-03
	100	7.19E-02	3.01E-02	2.29E-02	5.78E-02	1.34E-02	6.25E-03	5.78E-02	1.32E-02	5.97E-03	6.08E-02	1.42E-02	6.19E-03
	200	1.21E-01	8.83E-02	5.41E-02	1.12E-01	7.85E-02	4.02E-02	1.12E-01	7.85E-02	4.01E-02	1.15E-01	8.07E-02	4.19E-02
	A	1.95E-02			4.87E-03			**4.46E-03**			4.46E-03		
25	50	3.08E-02	2.41E-02	3.21E-02	1.31E-02	4.20E-03	3.76E-03	1.28E-02	3.78E-03	3.02E-03	1.51E-02	3.92E-03	2.98E-03
	100	7.14E-02	2.93E-02	2.21E-02	5.75E-02	1.27E-02	5.40E-03	5.74E-02	1.25E-02	5.08E-03	6.06E-02	1.35E-02	5.34E-03
	200	1.20E-01	8.80E-02	5.36E-02	1.12E-01	7.82E-02	3.96E-02	1.12E-01	7.82E-02	3.95E-02	1.15E-01	8.06E-02	4.14E-02
	A	1.84E-02			3.62E-03			3.02E-03			**2.97E-03**		

medium speed variations

Settings		Smoothed direct preestimates			Smoothed indirect preestimates			Smoothed enhanced preestimates			LBF estimates		
SNR	$k\backslash m$	1	3	5	1	3	5	1	3	5	1	3	5
15	50	1.95E-02	2.27E-02	3.22E-02	4.18E-03	3.23E-03	3.64E-03	4.14E-03	3.09E-03	3.33E-03	4.93E-03	3.16E-03	3.49E-03
	100	2.92E-02	1.80E-02	1.94E-02	1.27E-02	4.01E-03	3.45E-03	1.27E-02	3.97E-03	3.36E-03	1.40E-02	4.13E-03	3.44E-03
	200	7.08E-02	2.79E-02	1.87E-02	5.70E-02	1.25E-02	5.24E-03	5.70E-02	1.25E-02	5.21E-03	5.87E-02	1.31E-02	5.39E-03
	A	1.55E-02			2.75E-03			2.66E-03			**2.65E-03**		
25	50	1.83E-02	2.06E-02	2.91E-02	3.56E-03	1.94E-03	1.70E-03	3.50E-03	1.82E-03	1.48E-03	4.24E-03	1.88E-03	1.51E-03
	100	2.83E-02	1.68E-02	1.78E-02	1.24E-02	3.35E-03	2.47E-03	1.24E-02	3.29E-03	2.38E-03	1.37E-02	3.47E-03	2.48E-03
	200	7.03E-02	2.70E-02	1.76E-02	5.69E-02	1.22E-02	4.75E-03	5.69E-02	1.21E-02	4.70E-03	5.87E-02	1.28E-02	4.89E-03
	A	1.39E-02			1.54E-03			**1.42E-03**			1.44E-03		

slow variations

Settings		Smoothed direct preestimates			Smoothed indirect preestimates			Smoothed enhanced preestimates			LBF estimates		
SNR	$k\backslash m$	1	3	5	1	3	5	1	3	5	1	3	5
15	50	1.64E-02	2.21E-02	3.17E-02	2.15E-03	2.22E-03	2.72E-03	2.14E-03	2.15E-03	2.56E-03	2.39E-03	2.22E-03	2.74E-03
	100	1.75E-02	1.60E-02	1.84E-02	3.69E-03	2.27E-03	2.18E-03	3.70E-03	2.26E-03	2.15E-03	4.12E-03	2.30E-03	2.19E-03
	200	2.83E-02	1.61E-02	1.55E-02	1.24E-02	3.51E-03	2.73E-03	1.24E-02	3.52E-03	2.73E-03	1.31E-02	3.61E-03	2.78E-03
	A	1.31E-02			1.68E-03			1.68E-03			**1.64E-03**		
25	50	1.54E-02	2.04E-02	2.91E-02	1.57E-03	1.05E-03	9.67E-04	1.56E-03	1.00E-03	8.71E-04	1.83E-03	1.04E-03	9.07E-04
	100	1.67E-02	1.50E-02	1.70E-02	3.38E-03	1.68E-03	1.30E-03	3.37E-03	1.66E-03	1.26E-03	3.84E-03	1.74E-03	1.32E-03
	200	2.77E-02	1.53E-02	1.46E-02	1.22E-02	3.21E-03	2.28E-03	1.22E-02	3.20E-03	2.26E-03	1.29E-02	3.33E-03	2.35E-03
	A	1.23E-02			8.36E-04			**7.90E-04**			8.07E-04		

of combined ensemble averaging (over 100 independent process realizations) and time averaging (over the simulation interval T_s). The basis consisted of powers of time (29) and the applied window was cosinusoidal. Typical identification results, yielded by the adaptive parallel estimation scheme combining smoothed enhanced preestimates (cf. Fig. 4), are shown in Fig. 6.

Examination of the results shown in Table I lead to the following conclusions:

1) For all choices of m and k the smoothed enhanced preestimates yield better results than the smoothed indirect preestimates, and the smoothed indirect preestimates yield better results than smoothed direct preestimates (for all SNR and parameter variation rates).

2) In all cases considered the results yielded by the smoothed enhanced preestimates only marginally differ (sometimes up, and sometimes down) from those provided by their, computationally much more demanding, LBF "prototypes".

3) In all cases considered the parallel estimation schemes yield better results than those provided by the best fLBF/LBF algorithms with fixed settings.

Simulation experiments were next repeated for a different pattern of window sizes (60, 90, 135, 200), different sets of basis functions (harmonic, Slepian), and different shapes of the weighting sequence (rectangular, Hann). The obtained results, not shown here because of the lack of space, were very similar to those summarized in Table I, which confirms that all design choices mentioned above are by no means critical, i.e., the proposed parallel estimation scheme is robust.

The aim of our second experiment was to compare the competitive fLBF estimator described above (the variant based on smoothing enhanced preestimates) with the estimator obtained using the multi-resolution wavelet (MW) estimation scheme described in [17]–[19]. Following recommendations given in [19], the multi-wavelet approach involved third, fourth and fifth order cardinal B-splines and
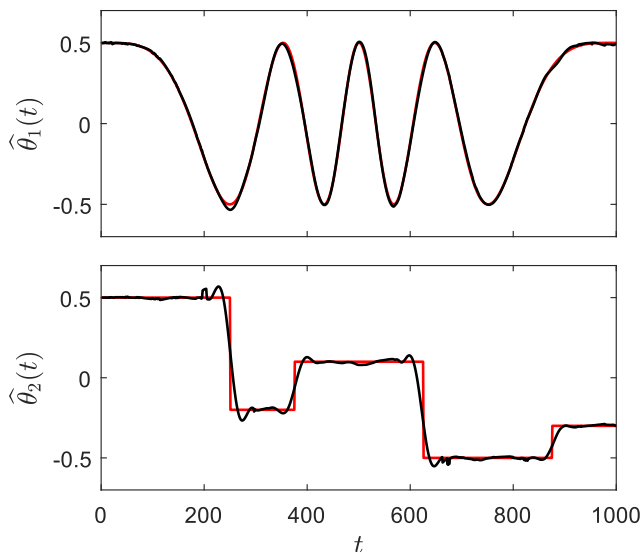
Fig. 6: Estimated parameter trajectories of a two-tap FIR system. Smoothed enhanced preestimates (black lines) are superimposed on true parameter trajectories (red lines).

the resolution level was set to 3. The best subset of 33 approximating wavelets was selected by the orthogonal least squares algorithm and the number of wavelets - by the well-known Akaike's information criterion (AIC). The approximation was carried out in intervals of length 501 using the overlap-add approach (50 % overlap, Hann synthesis window).

The results of comparison are shown – separately for $\theta_1(t)$ and $\theta_2(t)$ – in Table II. According to the simulation evidence summarized in Table II, the proposed approach yields results that are better (in 8 cases out of 12) or comparable with those provided by the MW approach.

$\theta_1(t)$

SoV		Fast	Medium	Slow
SNR	Method			
15 dB	fLBF	**1.14E-03**	**7.69E-04**	**5.36E-04**
	MW	1.29E-03	8.01E-04	7.35E-04
25 dB	fLBF	3.74E-04	**1.37E-04**	**1.11E-04**
	MW	**3.39E-04**	1.79E-04	1.36E-04

$\theta_2(t)$

SoV		Fast	Medium	Slow
SNR	Method			
15 dB	fLBF	**3.32E-03**	**1.89E-03**	**1.14E-03**
	MW	3.74E-03	2.28E-03	1.47E-03
25 dB	fLBF	2.65E-03	1.28E-03	6.79E-04
	MW	**1.94E-03**	**1.14E-03**	**5.99E-04**

TABLE II: Comparison of tracking results obtained – for different values of SNR and different speeds of parameter variation (SoV) – using adaptive fLBF approach (smoothed enhanced preestimates) with the analogous results yielded by the multi-resolution wavelet (MW) approach. In each case the best result is shown in boldface.
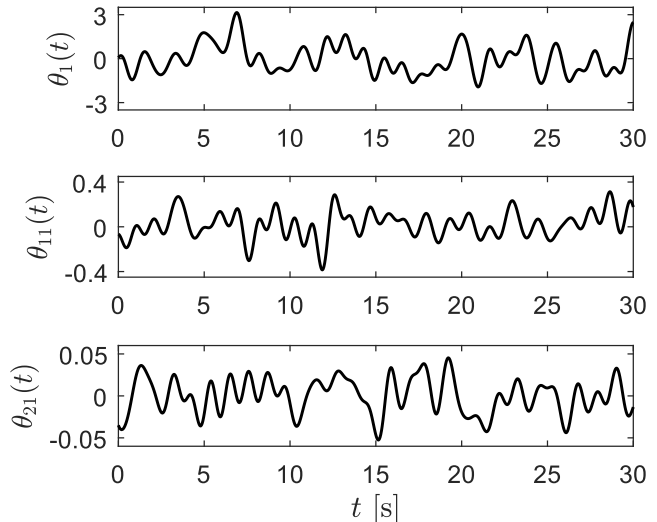

Fig. 7: Trajectories of the first, eleventh and twenty first parameter of the simulated underwater acoustic channel.

In the third experiment, we compared performance of the presented algorithms when applied to simulated data similar to observations that can be gathered during the self-interference cancellation in full-duplex underwater acoustic (UWA) communication systems [20]. For simplicity, all signals and parameters were assumed to be real-valued (extension to the complex-valued case is straightforward). Experimental data were generated following the procedures described in [20]. The UWA channel was modeled and identified as a 50-tap FIR filter ($n = 50$) and λ was set to 0.96. System parameters $\{\theta_i(t)\}$ were generated as mutually independent stationary random Gaussian processes with a power spectral density $c_i G(f)$, where c_i denotes the variance of the n-th tap of the time-varying filter, and $G(f)$ is uniform within the frequency interval $[-f_{\max}, f_{\max}]$. Due to the spreading and absorption loss, the variance of consecutive parameters decays exponentially: $c_i = e^{-\gamma(i-1)}, \quad i \in [1, n]$. Following [20], γ was chosen so that the variance decrease between the first and the last arrival was equal to 80 dB. The sampling frequency was set to 1 kHz and $f_{\max}$ – to 1 Hz (which corresponds to fast changes in the UWA case). Fig. 7 shows trajectories of selected parameters, while Fig. 8 displays a snapshot of the system impulse response. The generated input signal was white binary $u(t) \in \{-1, 1\}$.

Identification results, gathered in Table III, confirm our earlier findings – for low and moderate values of SNR (≤ 30 dB) fLBF estimators are an attractive alternative to computationally much more demanding LBF estimators. The MW approach was excluded from the comparison because the time needed to process even a single realization of the generated data was unacceptably long.

VIII. CONCLUSION

A new method of identification of time-varying FIR systems subject to (locally) stationary excitation was described.

TABLE III: Mean square parameter estimation errors obtained for 3 fLBF/LBF estimators corresponding to $k = 150$ and different choices of design parameter m (1, 3, 5), and for the adaptive parallel estimation schemes based on cross-validation (A). All averages were computed for 100 process realizations and 3 average signal-to-noise ratios (SNR). The best results in each case are shown in boldface.

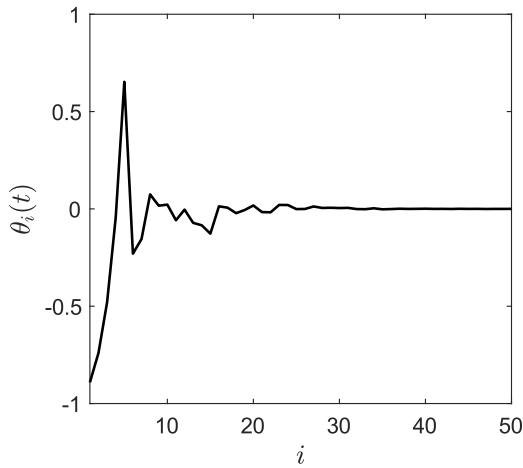| SNR	m	Smoothed direct preestimates	Smoothed indirect preestimates	Smoothed enhanced preestimates	LBF estimates
15 dB	1	1.02E+00	3.44E-02	**3.07E-02**	3.79E-02
	3	1.11E+00	5.92E-02	4.40E-02	5.39E-02
	5	1.27E+00	8.57E-02	5.43E-02	1.52E-01
	A	1.02E+00	4.66E-02	5.07E-02	3.98E-02
30 dB	1	1.01E+00	1.04E-02	7.89E-03	1.87E-02
	3	1.10E+00	1.08E-02	5.22E-03	**1.72E-03**
	5	1.25E+00	1.49E-02	6.13E-03	4.81E-03
	A	1.01E+00	9.54E-03	5.37E-03	1.72E-03
45 dB	1	1.01E+00	9.62E-03	7.16E-03	1.81E-02
	3	1.10E+00	9.29E-03	4.01E-03	**6.78E-05**
	5	1.25E+00	1.27E-02	4.64E-03	1.52E-04
	A	1.01E+00	8.44E-03	4.06E-03	6.86E-05

Fig. 8: A snapshot of the time-varying impulse response of the simulated underwater communication channel.

The proposed approach is based on smoothing of preestimated parameter trajectories obtained by inverse filtering of short-memory EWLS parameter estimates, and yields results that are better, or at least comparable with those provided by the more sophisticated and computationally much more demanding methods based on local basis function approximation and multi-resolution wavelet approximation. Optimization of the smoothing operation can be carried out adaptively using parallel estimation and cross-validation.

REFERENCES

[1] M. Niedźwiecki and T. Kłaput "Fast recursive basis function estimators for identification of nonstationary systems," *IEEE Transactions on Signal Processing*, vol. 50, 1925–1934, 2002.

[2] M.H. Neumann and R. von Sachs, "Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra," *The Annals of Statistics* vol. 25, pp. 38–76, 1997.

[3] R. Dahlhaus, "Local inference for locally sationary time series based on the empirical spectral measure," *Journal of Econometrics*, vol. 151, pp. 101–112, 2009.

[4] R. Dahlhaus, "Locally stationary processes," *Handbook Statist.*, vol. 25, pp. 1–37, 2012.

[5] A. van Delft and M. Eichler, "Data-adaptive estimation of time-varying spectral densities," *Journal of Computational and Graphical Statistics,* vol. 28, pp. 244–255, 2019.

[6] J.R. Bellegarda and D.C. Farden, "Constrained time-varying system modelling," *Proc. 1988 American Contol Conference*, Atlanta, USA, pp. 1295–1300, 1988.

[7] M. Niedźwiecki, M. Ciołek and A. Gańcza, "A new look at the statistical identification of nonstationary systems," *Automatica*, vol. 118, no. 109037, 2020.

[8] M. Niedźwiecki and M. Ciołek, "Generalized Savitzky-Golay filters for identification of nonstationary systems," *Automatica*, vol. 108, no. 108522, 2019.

[9] M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *Int. J. Adaptive Contr. Signal Process.*, vol. 10, pp. 159-176, 1996.

[10] G. B. Giannakis and C. Tepedelenlioĝlu, "Basis expansion models anddiversity techniques for blind identification and equalization of time-varying channels," *Proceedings of IEEE*, vol. 86, pp. 1969–1986, 1998.

[11] L. Isserilis, "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika*, vol. 12, pp. 134-139, 1918.

[12] T. Söderström and P. Stoica, *System Identification*, Englewood Cliffs NJ: Prentice-Hall, 1988.

[13] M. Niedźwiecki, *Identification of Time-varying Processes.* New York: Wiley, 2000.

[14] D.M. Allen, D.M., "The relationship between variable selection and data augmentation and a method for prediction," *Technometrics*, vol. 16, pp. 125–127, 1974.

[15] M. Niedźwiecki and M. Ciołek, "Akaike's final prediction error criterion revisited," 40th International Conference on Telecommunications and Signal Processing (TSP), Barcelona, Spain, pp. 237-242, 2017

[16] R.W. Schafer, "What is a Savitzky-Golay filter?," *IEEE Signal Process. Mag.*, vol. 28, pp. 111-117, 2011.

[17] H.L. Wei, J. Liu and S.A. Billings, "Identification of time-varying systems using multi-resolution wavelet models," *International Journal of Systems Science*, vol. 33, pp. 1217-1228, 2002.

[18] H.L. Wei, S.A. Billings, and J.J. Liu "Time-varying parametric modelling and time-dependent spectral characterisation with applications to EEG signals using multiwavelets," *International Journal of Modelling, Identification and Control*, vol.9, pp. 215 – 224, 2010.

[19] Y. Li, H.L., Wei, S.A. Billings, and P. Sarrigiannis, "Time-varying model identification for time-frequency feature extraction from EEG data," *Journal of Neuroscience Methods*, vol. 196, pp. 151-8, 2011.

[20] L. Shen, Y. Zakharov, B. Henson, N. Morozs and P. Mitchell, "Adaptive filtering for full-duplex UWA systems with time-varying self-interference channel," Submitted to *IEEE J. Oceanic Eng.*, TechRxiv. https://doi.org/10.36227/techrxiv.11967813.v1, 2019.