Available online at www.sciencedirect.com
ScienceDirect
IFAC PapersOnLine 51-15 (2018) 933–938

New results on estimation bandwidth adaptation

Maciej Niedźwiecki * **Marcin Ciołek** *

* *Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology ul. Narutowicza 11/12, Gdańsk , Poland e-mail: maciekn@eti.pg.gda.pl, marcin.ciolek@pg.gda.pl*

Abstract: The problem of identification of a nonstationary autoregressive signal using noncausal estimation schemes is considered. Noncausal estimators can be used in applications that are not time-critical, i.e., do not require real-time processing. A new adaptive estimation bandwidth selection rule based on evaluation of pseudoprediction errors is proposed, allowing one to adjust tracking characteristics of noncausal estimators to unknown and/or time-varying degree of signal nonstationarity. The new rule is compared with the previously proposed one, based on the generalized Akaike's final prediction error criterion.

Keywords: identification of nonstationary autoregressive processes, noncausal estimation, parameter tracking, estimation bandwidth selection

1. INTRODUCTION

Consider a nonstationary discrete-time signal $\{y(t), t = \ldots, -1, 0, 1 \ldots\}$, which can be described (or at least well approximated) by the following time-varying autoregressive (AR) model of order n

$$y(t) = \sum_{i=1}^{n} a_i(t)y(t-i) + e(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + e(t) \qquad (1)$$
$$\mathrm{var}[e(t)] = \rho(t)$$

where $\boldsymbol{\theta}(t) = [a_1(t), \ldots, a_n(t)]^{\mathrm{T}}$ denotes the vector of autoregressive coefficients, $\boldsymbol{\varphi}(t) = [y(t-1), \ldots, y(t-n)]^{\mathrm{T}}$ is the regression vector, and $\{e(t)\}$ denotes a squence of independent zero-mean random variables with time-varying variance $\rho(t)$.

Due to their simplicity and good predictive capabilities, AR models have found their way to a large number of practical applications in different fields such as biomedicine [Fabri, Camilleri & Cassar (2011)], [Schlögl (2000)], [Wada, Jinnouchi & Matsumura (1998)], telecommunications [Baddour & Beaulieu (2005)], [Hayes & Ganesh Babu (2004)], and geophysics [Lesage, Glangeaud & Mars (2002)], [Li & Nowack (2004)], among many others. Since parameters of AR models have usually no physical significance, identification of such models is not a goal in itself – it is the means of solving practical problems in the sense that the corresponding solutions depend explicitly on the estimates of model coefficients.

In real-time applications, such as adaptive prediction, the estimate $\boldsymbol{\theta}(t)$ must be causal, i.e., it must rely on the observation history $\mathcal{Y}(t) = \{y(s), s \leq t\}$ available at the instant t. In cases like this, estimation/tracking of $\boldsymbol{\theta}(t)$ can be carried out using the well-known exponentially weighted least squares algorithms – see e.g. Niedźwiecki (2000). On the other hand, many applications, such as predictive coding of signals or parametric spectrum estimation, are not time-critical in the sense that the model-based decisions can be postponed or at least delayed by a certain number of sampling intervals, which means that parameter estimation can be based on past, but also on a certain number of "future" observations. Since noncausal estimation schemes, which take advantage of this fact, can significantly reduce the bias component of the mean square parameter estimation error, their tracking performance is usually much better than that of the comparable causal schemes Niedźwiecki (2000). The aim of this paper is to develop a new technique which allows one to adjust noncausal estimates of $\boldsymbol{\theta}(t)$ to unknown and possibly time-varying rate of parameter variation.

2. NONCAUSAL WEIGHTED LEAST SQUARES ESTIMATORS

The noncausal weighted least squares (NWLS) estimates of $\boldsymbol{\theta}(t)$ and $\rho(t)$ take the form

$$\widehat{\boldsymbol{\theta}}_k(t) = \arg\min_{\boldsymbol{\theta}} \sum_{i=-k}^{k} v_k(i)[y(t+i) - \boldsymbol{\varphi}^{\mathrm{T}}(t+i)\boldsymbol{\theta}]^2 = \mathbf{R}_k^{-1}(t)\mathbf{r}_k(t) \qquad (2)$$

$$\widehat{\rho}_k(t) = \frac{1}{L_k}\sum_{i=-k}^{k} v_k(i)[y(t+i) - \boldsymbol{\varphi}^{\mathrm{T}}(t+i)\widehat{\boldsymbol{\theta}}_k(t)]^2 = \frac{1}{L_k}\,[s_k(t) - \mathbf{r}_k^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_k(t)] \qquad (3)$$

where

$$\mathbf{R}_k(t) = \sum_{i=-k}^{k} v_k(i)\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\mathrm{T}}(t+i)$$

$$\mathbf{r}_k(t) = \sum_{i=-k}^{k} v_k(i)y(t+i)\boldsymbol{\varphi}(t+i)$$

* This work was partially supported by the National Science Centre under the agreement UMO-2015/17/B/ST7/03772. Calculations were carried out at the Academic Computer Centre in Gdańsk.

10.1016/j.ifacol.2018.09.074

$$s_k(t) = \sum_{i=-k}^{k} v_k(i) y^2(t+i)$$

and $\{v_k(i), i = -k, \dots, k\}$, $v_k(0) = 1$, denotes the nonnegative, symmetric, bell-shaped window of width $2k+1$, introduced to make the estimates depend more heavily on the data collected at the instants close to t, and less heavily on measurements coming from the remote past and future. We will assume that $v_k(i) = f(i/k)$ where $f(\cdot)$ is the analog window generating function defined on the interval $[-1, 1]$. Finally, the quantity

$$L_k = \sum_{i=-k}^{k} v_k(i) \cong k \int_{-1}^{1} f(x)dx$$

denotes the effective window width.

Remark

We note that the Hann (raised cosine) window

$$v_k(i) = \frac{1}{2}\left[1 + \cos\frac{\pi i}{k+1}\right] = \frac{1}{2}\left\{1 + \mathrm{Re}\left[e^{j\frac{\pi i}{k+1}}\right]\right\} \quad (4)$$

allows for recursive computation of $\mathbf{R}_k(t)$, $\mathbf{r}_k(t)$ and $s_k(t)$. Actually, let

$$\mathbf{S}_k(t) = \sum_{i=-k}^{k} \boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\mathrm{T}}(t+i)$$

$$\mathbf{T}_k(t) = \sum_{i=-k}^{k} e^{j\frac{\pi i}{k+1}} \boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\mathrm{T}}(t+i) \ .$$

Observe that both quantities defined above are recursively computable

$$\begin{aligned}
\mathbf{S}_k(t+1) &= \mathbf{S}_k(t) - \boldsymbol{\varphi}(t-k)\boldsymbol{\varphi}^{\mathrm{T}}(t-k) \\
&\quad + \boldsymbol{\varphi}(t+k+1)\boldsymbol{\varphi}^{\mathrm{T}}(t+k+1) \\
\mathbf{T}_k(t+1) &= e^{-j\frac{\pi}{k+1}}\mathbf{T}_k(t) + \boldsymbol{\varphi}(t-k)\boldsymbol{\varphi}^{\mathrm{T}}(t-k) \\
&\quad + e^{j\frac{\pi k}{k+1}}\boldsymbol{\varphi}(t+k+1)\boldsymbol{\varphi}^{\mathrm{T}}(t+k+1)
\end{aligned}$$

and that

$$\mathbf{R}_k(t) = \frac{1}{2}\,\mathbf{S}_k(t) + \frac{1}{2}\,\mathrm{Re}[\mathbf{T}_k(t)] \ .$$

The quantities $\mathbf{r}_k(t)$ and $s_k(t)$ can be computed recursively in an analogous way.

Tracking capabilities of NWLS estimators depend on the window size. When the effective width of the window is small, the corresponding estimation algorithm is "fast" (quickly reacts to parameter changes), but "innacurate" (yields estimates with large variability); when it is large, the algorithm is "slow" but "accurate". The estimation bandwidth [Niedźwiecki (2000)], i.e., the frequency range in which signal parameters can be tracked "successfully", is inversely proportional to k. The best tracking results can be obtained when the bandwidth matches the rate of parameter variation, trading off the bias and variance components of the mean square parameter tracking error. When the rate of signal nonstationarity is unknown and/or when it changes with time, bandwidth optimization can be carried out using parallel estimation schemes – see Niedźwiecki (1990), Niedźwiecki (1992). In this approach several estimation algorithms, equipped with different bandwidth settings $k \in \mathcal{K} = \{k_1, \dots, k_K\}$, are simultaneously run and compared. At each time instant only one of the competing algorithms is selected, i.e., the estimated parameter and variance trajectories have the form $\widehat{\boldsymbol{\theta}}_{\widehat{k}(t)}(t)$ and $\widehat{\rho}_{\widehat{k}(t)}(t)$, respectively, where

$$\widehat{k}(t) = \arg\min_{k\in\mathcal{K}} J_k(t) \quad (5)$$

and $J_k(t)$ denotes the local decision statistic.

In our previous paper, Niedźwiecki, Ciołek & Kajikawa (2017), we have shown that the problem of bandwidth selection can be solved using the localized version of the Akaike's final prediction error (FPE) criterion. In this case

$$J_k(t) = \mathrm{FPE}_k(t) = \frac{1 + \frac{n}{N_k}}{1 - \frac{n}{N_k}}\,\widehat{\rho}_k(t) \quad (6)$$

where

$$N_k = \frac{\left[\sum_{i=-k}^{k} v_k(i)\right]^2}{\sum_{i=-k}^{k} v_k^2(i)} \cong k\,\frac{\left[\int_{-1}^{1} f(x)dx\right]^2}{\int_{-1}^{1} f^2(x)dx} \quad (7)$$

denotes the so-called equivalent window width.

In the same paper we have demonstrated that the cross validation approach, which yields $J_k(t)$ in a form of a local sum of squared leave-one-out signal interpolation errors, *does not* provide satisfactory results (most likely because interpolation errors may be strongly correlated). As we will show below, the situation changes substantially if interpolation errors are replaced with the pseudoprediction errors.

To define pseudoprediction errors, we will introduce the notion of a holey estimator of $\boldsymbol{\theta}(t)$. The holey estimator $\widehat{\boldsymbol{\theta}}_k^{\circ}(t)$, introduced in Niedźwiecki (2012) for finite impulse response (FIR) systems of the form

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + e(t),\ \boldsymbol{\varphi}(t) = [u(t-1), \dots, u(t-n)]^{\mathrm{T}}$$

where $\{u(i)\}$ denotes the observable input sequence, can be obtained by excluding from the estimation process the output measurement $y(t)$, collected at the instant t. This leads to the following formula

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_k^{\circ}(t) &= \arg\min_{\boldsymbol{\theta}} \sum_{\substack{i=-k \\ i\neq 0}}^{k} v_k(i)[y(t+i) - \boldsymbol{\varphi}^{\mathrm{T}}(t+i)\boldsymbol{\theta}]^2 \\
&= [\mathbf{R}_k^{\circ}(t)]^{-1}\mathbf{r}_k^{\circ}(t)
\end{aligned} \quad (8)$$

where

$$\mathbf{R}_k^{\circ}(t) = \sum_{\substack{i=-k \\ i\neq 0}}^{k} v_k(i)\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\mathrm{T}}(t+i)$$

$$\mathbf{r}_k^{\circ}(t) = \sum_{\substack{i=-k \\ i\neq 0}}^{k} v_k(i)y(t+i)\boldsymbol{\varphi}(t+i) \ .$$

Even though for AR processes we will adopt the same formula, there is an important difference between the AR case and the FIR case: in the AR case $y(t)$ is a component of regression vectors $\boldsymbol{\varphi}(t+1), \dots, \boldsymbol{\varphi}(t+n)$. Hence, unlike the FIR case, even if the term $v_k(0)[y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}]^2$ is excluded from the sum in (2), the estimate (8) remains a function of $y(t)$.

Based on $\widehat{\boldsymbol{\theta}}_k^{\circ}(t)$, one can compute a pseudoprediction $\widehat{y}_k^{\circ}(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_k^{\circ}(t)$ of $y(t)$ and the corresponding pseudoprediction error

$$\varepsilon_k^{\circ}(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_k^{\circ}(t) \ . \quad (9)$$

The name "pseudoprediction" refers to the fact that the vector of parameter estimates $\widehat{\boldsymbol{\theta}}_k^{\circ}(t)$ depends on $y(t)$, which is the predicted quantity.

The decision statistic based on evaluation of the locally observed pseudoprediction errors (PPE) takes the form

$$J_k(t) = \mathrm{PPE}_k(t) = \sum_{i=-M}^{M} [\varepsilon_k^{\circ}(t+i)]^2 \tag{10}$$

where $M \in [20, 50]$ decides upon the size of the local decision window $[t-M, t+M]$ centered at t.

It is interesting to note that evaluation of (10) does not require implementation of the holey estimator $\widehat{\boldsymbol{\theta}}_k^{\circ}(\cdot)$.

Denote by $\varepsilon_k(t)$ the pseudoprediction error defined in terms of the original (unmodified) estimate $\widehat{\boldsymbol{\theta}}_k(t)$

$$\varepsilon_k(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_k(t) \ . \tag{11}$$

It is straightforward to show that

$$\varepsilon_k^{\circ}(t) = \frac{\varepsilon_k(t)}{1 - b_k(t)} \tag{12}$$

where

$$b_k(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{R}_k^{-1}(t)\boldsymbol{\varphi}(t) \ .$$

Actually, using the well-known matrix inversion lemma [see e.g. Söderström & Stoica (1988)], and exploiting the fact that $v_k(0) = 1$, one arrives at

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_k^{\circ}(t) &= [\mathbf{R}_k(t) - \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]^{-1}[\mathbf{r}_k(t) - y(t)\boldsymbol{\varphi}(t)] \\ &= \left[\mathbf{R}_k^{-1}(t) + \frac{\mathbf{R}_k^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{R}_k^{-1}(t)}{1 - \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{R}_k^{-1}(t)\boldsymbol{\varphi}(t)}\right] \times \\ &\qquad \times [\mathbf{R}_k(t)\widehat{\boldsymbol{\theta}}_k(t) - y(t)\boldsymbol{\varphi}(t)]\end{aligned}$$

which, after substitution into (9), leads to (12).

3. NONCAUSAL WEIGHTED YULE-WALKER ESTIMATORS

NWLS estimators do not guarantee stability of the obtained AR models. While in some applications, such as parametric spectrum estimation, this is not a critical issue, in some other ones, such as predictive coding of signals (where AR model is used to generate a signal that resembles the analyzed one), it is an obvious requirement. The stability problem can be overcome if estimation is carried out using the noncausal weighted Yule-Walker (NWYW) estimators.

First of all, we will show that the NWYW estimates can be interpreted as local least squares estimates obtained for the tapered data sequence

$$y_k(t+i|t) = w_k(i)y(t+i), \quad i \in [-k, k]$$

where $\{w_k(i), i = -k, \dots, k\}$, $w_k(0) = 1$, is the nonnegative, symmetric, bell-shaped data taper. Similarly as in the case of NWLS estimators, we will assume that $w_k(i) = g(i/k)$, where $g(\cdot)$ is the continuous time taper generation function defined on $[-1, 1]$. Suppose that the data sequence $y(t-k), \dots, y(t+k)$ is extended with n zero samples at its beginning and at its end, and that the data taper $w_k(t-k), \dots, w_k(t+k)$ is extended likewise. Note that under such extensions it holds that $y_k(t+i|t) = 0$ for $i \in [-k-n, -k-1]$ and $i \in [k+1, k+n]$. Finally, let $\boldsymbol{\varphi}_k(t+i|t) = [y_k(t+i-1|t), \dots, y_k(t+i-n|t)]^{\mathrm{T}}$.

Consider the following least squares estimates of $\boldsymbol{\theta}(t)$ and $\rho(t)$

$$\begin{aligned}\widetilde{\boldsymbol{\theta}}_k(t) &= \arg\min_{\boldsymbol{\theta}} \sum_{i=-k}^{k+n} [y_k(t+i|t) - \boldsymbol{\varphi}_k^{\mathrm{T}}(t+i|t)\boldsymbol{\theta}]^2 \\ &= \mathbf{Q}_k^{-1}(t)\mathbf{q}_k(t)\end{aligned} \tag{13}$$

$$\begin{aligned}\widetilde{\rho}_k(t) &= \frac{1}{L_k}\sum_{i=-k}^{k+n} [y_k(t+i|t) - \boldsymbol{\varphi}_k^{\mathrm{T}}(t+i|t)\widetilde{\boldsymbol{\theta}}_k(t)]^2 \\ &= \frac{1}{L_k}\,[z_k(t) - \mathbf{q}_k^{\mathrm{T}}(t)\widetilde{\boldsymbol{\theta}}_k(t)]\end{aligned} \tag{14}$$

where

$$\mathbf{Q}_k(t) = \sum_{i=-k}^{k+n} \boldsymbol{\varphi}_k(t+i|t)\boldsymbol{\varphi}_k^{\mathrm{T}}(t+i|t)$$

$$\mathbf{q}_k(t) = \sum_{i=-k}^{k+n} y_k(t+i|t)\boldsymbol{\varphi}_k(t+i|t)$$

$$z_k(t) = \sum_{i=-k}^{k+n} y_k^2(t+i|t)$$

and

$$L_k = \sum_{i=-k}^{k} w_k^2(i) \cong k\int_{-1}^{1} g^2(x)dx \ .$$

It is straightforward to check that the matrix $\mathbf{Q}_k(t)$ is symmetric and Toeplitz

$$\mathbf{Q}_k(t) = \begin{bmatrix} p_{0|k}(t) & \dots & p_{n-1|k}(t) \\ & \ddots & \\ p_{n-1|k}(t) & \dots & p_{0|k}(t) \end{bmatrix}$$

and that

$$\mathbf{q}_k(t) = [p_{1|k}(t), \dots, p_{n|k}(t)]^{\mathrm{T}}, \quad z_k(t) = p_{0|k}(t)$$

where

$$p_{l|k}(t) = \sum_{i=-k+l}^{k} y_k(t+i|t)y_k(t+i-l|t) \ .$$

This means that the estimates $\widetilde{\boldsymbol{\theta}}_k(t)$ and $\widetilde{\rho}_k(t)$, given by (13) and (14), respectively, are solutions of the Yule-Walker equations provided that the true autocorrelation coefficients of $y(t)$ are replaced with their local (tapered) estimates $p_{l|k}(t)/L_k, l = 0, \dots, n$ – see Söderström & Stoica (1988). A rigorous statistical analysis of the benefits of data tapering can be found e.g. in Dahlhaus & Giraitis (1998).

Note that when $n \ll k$, it holds that $\boldsymbol{\varphi}_k(t+i|t) \cong w_k(i)\boldsymbol{\varphi}(t+i)$, and since $y_k(t+i|t) = w_k(i)y(t+i)$, one arrives at $\widetilde{\boldsymbol{\theta}}_k(t) \cong \widehat{\boldsymbol{\theta}}_k(t)$ provided that

$$v_k(i) = w_k^2(i), \quad i \in [-k, k] \ . \tag{15}$$

This means that under the condition (15) the NWYW estimators will yield approximately the same results as the NWLS estimators. Note also that when $v_k(i)$ is a raised cosine window (4), the "equivalent" data taper is

$$w_k(i) = \sqrt{v_k(i)} = \cos\frac{\pi i}{2(k+1)} \ . \tag{16}$$

As shown in Niedźwiecki, Ciołek & Kajikawa (2017), such cosinusoidal taper allows for recursive computation of $\mathbf{Q}_k(t)$ and $\mathbf{q}_k(t)$.

Based on the least squares interpretation of NWYW estimators, one can define the holey estimator $\widetilde{\boldsymbol{\theta}}_k^{\circ}(t)$ in the form

$$\begin{aligned}\widetilde{\boldsymbol{\theta}}_k^{\circ}(t) &= \arg\min_{\boldsymbol{\theta}} \sum_{\substack{i=-k\\ i\neq 0}}^{k+n} [y_k(t+i|t) - \boldsymbol{\varphi}_k^{\mathrm{T}}(t+i|t)\boldsymbol{\theta}]^2 \\ &= [\mathbf{Q}_k^{\circ}(t)]^{-1}\mathbf{q}_k^{\circ}(t)\end{aligned} \tag{17}$$

where

$$\begin{aligned}\mathbf{Q}_k^{\circ}(t) &= \mathbf{Q}_k(t) - \boldsymbol{\varphi}_k(t|t)\boldsymbol{\varphi}_k^{\mathrm{T}}(t|t)\\ \mathbf{q}_k^{\circ}(t) &= \mathbf{q}_k(t) - y_k(t|t)\boldsymbol{\varphi}_k(t|t)\ .\end{aligned}$$

This leads to the following decision statistic

$$J_k(t) = \mathrm{PPE}_k(t) = \sum_{i=-M}^{M} [\eta_k^{\circ}(t+i)]^2 \tag{18}$$

where

$$\eta_k^{\circ}(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widetilde{\boldsymbol{\theta}}_k^{\circ}(t) \tag{19}$$

denotes the pseudoprediction error. One can show that

$$\eta_k^{\circ}(t) = \eta_k(t) + \frac{c_k(t)}{1-d_k(t)}\,\gamma_k(t) \tag{20}$$

where

$$\begin{aligned}\eta_k(t) &= y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widetilde{\boldsymbol{\theta}}_k(t)\\ \gamma_k(t) &= y(t) - \boldsymbol{\varphi}_k^{\mathrm{T}}(t|t)\widetilde{\boldsymbol{\theta}}_k(t)\\ c_k(t) &= \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k(t|t)\\ d_k(t) &= \boldsymbol{\varphi}_k^{\mathrm{T}}(t|t)\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k(t|t)\ .\end{aligned}$$

This allows one to compute pseudoprediction errors without implementing the holey estimation scheme.

When $n \ll k$, it holds that $\boldsymbol{\varphi}_k(t|t) \cong \boldsymbol{\varphi}(t)$, leading to $\eta_k(t) \cong \gamma_k(t)$, $c_k(t) \cong d_k(t)$ and

$$\eta_k^{\circ}(t) \cong \frac{\gamma_k(t)}{1-d_k(t)}$$

which resembles (12).

We note that the holey estimator given by (17) differs from the analogous estimator proposed in our earlier work Niedźwiecki, Ciołek & Kajikawa (2017) – in the case considered there the influence of $y(t)$ on $\widetilde{\boldsymbol{\theta}}_k^{\circ}(t)$ was completely removed, albeit at the expense of introducing extra bias errors.

The alternative solution to the bandwidth selection problem can be obtained using the suitably modified Akaike's final prediction error statistic, i.e., by adopting [Niedźwiecki, Ciołek & Kajikawa (2017)]

$$J_k(t) = \mathrm{FPE}_k(t) = \frac{1+\frac{n}{N_k}}{1-\frac{n}{N_k}}\,\widetilde{\rho}_k(t) \tag{21}$$

where

$$N_k = \frac{\left[\sum_{i=-k}^{k} w_k^2(i)\right]^2}{\sum_{i=-k}^{k} w_k^4(i)} \cong k\,\frac{\left[\int_{-1}^{1} g^2(x)dx\right]^2}{\int_{-1}^{1} g^4(x)dx}\ . \tag{22}$$

Remark

The NWYW approach is computationally attractive. First, the estimates $\widetilde{\boldsymbol{\theta}}_k(t)$ and $\widetilde{\rho}_k(t)$ can be evaluated using the computationally efficient order-recursive Levinson-Durbin algorithm [Söderström & Stoica (1988)]. Second, the square root of the matrix $\mathbf{Q}_k^{-1}(t)$, needed to compute pseudoprediction errors according to (20), can be expressed in terms of the quantities evaluated by the Levinson-Durbin algorithm (autoregressive coefficients of AR models of orders $1,\ldots,n$ and the corresponding reflection coefficients) – see Söderström & Stoica (1988). Given $\mathbf{Q}_k^{-1/2}(t)$, such that $\mathbf{Q}_k^{-1/2}(t)\mathbf{Q}_k^{-\mathrm{T}/2}(t) = \mathbf{Q}_k^{-1}(t)$, one can use the the formulas $c_k(t) = \boldsymbol{\alpha}_k^{\mathrm{T}}(t)\boldsymbol{\beta}_k(t)$ and $d_k(t) = \|\boldsymbol{\beta}_k(t)\|^2$, where

$$\boldsymbol{\alpha}_k(t) = \mathbf{Q}_k^{-\mathrm{T}/2}(t)\boldsymbol{\varphi}(t),\quad \boldsymbol{\beta}_k(t) = \mathbf{Q}_k^{-\mathrm{T}/2}(t)\boldsymbol{\varphi}_k(t|t)\ .$$

4. SYMMETRIZATION

Since the proposed decision statistics (10) and (18) are defined in terms of forward pseudoprediction errors only, the results of bandwidth selection will be in general different if the analysis is performed backward in time. This inconsistency can be eliminated if both forward (−) and backward (+) pseudoprediction errors are incorporated in $J_k(t)$.

The forward/backward AR model can be written down in the form

$$y(t) = \boldsymbol{\varphi}_{\pm}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + e_{\pm}(t)$$

where

$$\boldsymbol{\varphi}_{\pm}(t) = [y(t\pm 1),\ldots,y(t\pm n)]^{\mathrm{T}}\ .$$

Denote by

$$\varepsilon_k^{\pm}(t) = y(t) - \boldsymbol{\varphi}_{\pm}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_k^{\pm}(t)$$

the forward/backward pseudoprediction errors associated with the forward/backward NWLS estimators, respectively. The symmetrized decision statistic can be defined in the form

$$\mathrm{PPE}_k^*(t) = \sum_{i=-M}^{M}\{[\varepsilon_k^-(t+i)]^{\circ}\}^2 + \sum_{i=-M}^{M}\{[\varepsilon_k^+(t+i)]^{\circ}\}^2$$

where

$$\begin{aligned}[\varepsilon_k^{\pm}(t)]^{\circ} &= \frac{\varepsilon_k^{\pm}(t)}{1-b_k^{\pm}(t)}\\ b_k^{\pm}(t) &= \boldsymbol{\varphi}_{\pm}^{\mathrm{T}}(t)[\mathbf{R}_k^{\pm}(t)]^{-1}\boldsymbol{\varphi}_{\pm}(t)\ .\end{aligned}$$

Due to different initial conditions, constituted by the samples $y(t-k-1),\ldots,y(t-k-n)$ for the forward estimator $\widehat{\boldsymbol{\theta}}_k^-(t)$, and by the samples $y(t+k+1),\ldots,y(t+k+n)$ for the backward estimator $\widehat{\boldsymbol{\theta}}_k^+(t)$, it holds that $\mathbf{R}_k^-(t) \neq \mathbf{R}_k^+(t)$, $\mathbf{r}_k^-(t) \neq \mathbf{r}_k^+(t)$ and consequently $\widehat{\boldsymbol{\theta}}_k^-(t) \neq \widehat{\boldsymbol{\theta}}_k^+(t)$, which means that the computational complexity of the joint forward-backward analysis is doubled compared with that of the unidirectional analysis.

In the NWYW case the situation is different. Since it holds that $\mathbf{Q}_k^-(t) = \mathbf{Q}_k^+(t) = \mathbf{Q}_k(t)$, $\mathbf{q}_k^-(t) = \mathbf{q}_k^+(t) = \mathbf{q}_k(t)$ and $\widetilde{\boldsymbol{\theta}}_k^-(t) = \widetilde{\boldsymbol{\theta}}_k^+(t) = \widetilde{\boldsymbol{\theta}}_k(t)$, the computational overhead required to evaluate, instead of (18), the symmetrized decision statistic

$$\mathrm{PPE}_k^*(t) = \sum_{i=-M}^{M}\{[\eta_k^-(t+i)]^{\circ}\}^2 + \sum_{i=-M}^{M}\{[\eta_k^+(t+i)]^{\circ}\}^2$$

where

$$\begin{aligned}[\eta_k^{\pm}(t)]^{\circ} &= \eta_k^{\pm}(t) + \frac{c_k^{\pm}(t)}{1-d_k^{\pm}(t)}\,\gamma_k^{\pm}(t)\\ \eta_k^{\pm}(t) &= y(t) - \boldsymbol{\varphi}_{\pm}^{\mathrm{T}}(t)\widetilde{\boldsymbol{\theta}}_k^{\pm}(t)\\ \gamma_k^{\pm}(t) &= y(t) - [\boldsymbol{\varphi}_k^{\pm}(t|t)]^{\mathrm{T}}\widetilde{\boldsymbol{\theta}}_k^{\pm}(t)\\ c_k^{\pm}(t) &= \boldsymbol{\varphi}_{\pm}^{\mathrm{T}}(t)\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k^{\pm}(t|t)\\ d_k^{\pm}(t) &= [\boldsymbol{\varphi}_k^{\pm}(t|t)]^{\mathrm{T}}\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k^{\pm}(t|t)\\ \boldsymbol{\varphi}_k^{\pm}(t|t) &= [y_k(t\pm 1|t),\ldots,y_k(t\pm n|t)]^{\mathrm{T}},\end{aligned}$$

is noticeably reduced.

If needed, the symmetrized NWLS estimates of $\boldsymbol{\theta}(t)$ and $\rho(t)$ can be obtained from

$$\bar{\boldsymbol{\theta}}_k(t) = \frac{1}{2}\,[\widehat{\boldsymbol{\theta}}_k^-(t) + \widehat{\boldsymbol{\theta}}_k^+(t)], \quad \bar{\rho}_k(t) \quad = \frac{1}{2}\,[\widehat{\rho}_k^-(t) + \widehat{\rho}_k^+(t)]\ .$$

In the NWYW case such symmetrization is not necessary.

5. COMPUTER SIMULATIONS

Performance of the proposed bandwidth selection methods was checked by means of computer simulation. First, two time-invariant "anchor" AR models (A, B), of order $n = 8$, were obtained by performing local identification of an audio signal using the NWYW algorithm. The identified fragments differed in their resonance structures - see Fig. 1. Anchor models were specified in the lattice form $\{\delta_1, \ldots, \delta_n, p_0\}$, where $\delta_i, i = 1, \ldots, n$, denote reflection coefficients, which can be obtained as a byproduct of the Levinson-Durbin algorithm [Söderström & Stoica (1988)], and $p_0 = \sigma_y^2$. The lattice representation of a stable AR model is unique and can be uniquely transformed into the direct representation $\{a_1, \ldots, a_n, \rho\}$.

The time-varying AR model was obtained by morphing the anchor model A into B and *vice versa*. Transition from the model A, described by $\{\delta_1^{\mathrm{A}}, \ldots, \delta_n^{\mathrm{A}}, p_0^{\mathrm{A}}\}$, valid at the instant t_1, to the model B, described by $\{\delta_1^{\mathrm{B}}, \ldots, \delta_n^{\mathrm{B}}, p_0^{\mathrm{B}}\}$, valid at the instant t_2, was realized using the following transformations

$$\begin{gathered}\delta_i(t) = [1-\mu(t)]\delta_i^{\mathrm{A}} + \mu(t)\delta_i^{\mathrm{B}},\\ p_0(t) = [1-\mu(t)]p_0^{\mathrm{A}} + \mu(t)p_0^{\mathrm{B}},\\ i = 1, \ldots, n, \quad t \in [t_1, t_2]\end{gathered}$$

where

$$\mu(t) = \frac{t - t_1}{t_2 - t_1}.$$

Transition from the model B back to the model A was realized in an analogous way. At each time instant $t \in [1, T_s]$, where T_s denotes simulation time, the indirect time-varying lattice parametrization $\{\delta_1(t), \ldots, \delta_n(t), p_0(t)\}$ was transformed into the direct time-varying representation $\{a_1(t), \ldots, a_n(t), \rho(t)\}$, further used for AR process generation. Such a morphing technique guarantees stability of the resultant time-variant model at all times as long as both anchor models are stable (stability is not guaranteed if morphing is applied directly to the autoregressive coefficients).

The applied morphing scenario is symbolically depicted in Fig. 2. The identified model, analyzed in the interval $[1, T_s]$, had 3 periods of time-invariance (A-A, B-B, A-A), each of length l_1, interleaved with 2 periods of nonstationary behavior (A-B, B-A), each of length l_2 ($T_s = 3l_1 + 2l_2$). To check performance of the compared methods under different rates of signal nonstationarity, 3 different values of simulation time T_s were considered (64000, 32000 and 16000), resulting in 3 different speeds of parameter variation (SoV): S_1 (slow, $l_1 = 16000, l_2 = 8000$), S_2 (medium, $l_1 = 8000, l_2 = 4000$), and S_3 (fast, $l_1 = 4000, l_2 = 2000$), respectively. In each case data generation was started 1000 instants prior to $t = 1$ and was continued for 1000 instants after $t = T_s$. The variable-bandwidth schemes were made up of 4 algorithms with bandwidth parameters set to $k_1 = 225$, $k_2 = 337$, $k_3 = 505$ and $k_4 = 757$. The applied windows were given by (4) – for the NWLS algorithms, and by (16) – for the NWYW algorithms.

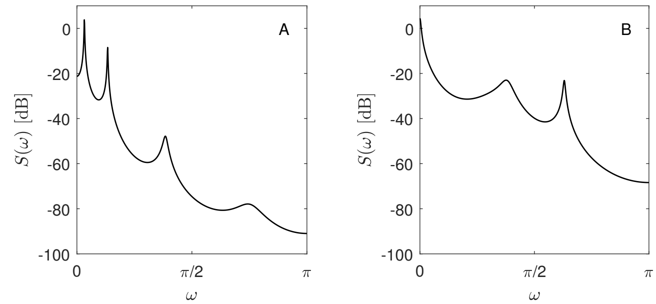

Fig. 1. Power spectra of two stationary anchor processes.

Two performance measures were used to evaluate and compare the instantaneous identification results: the squared parameter tracking error

$$d_{\mathrm{PAR}}(t) = \parallel \boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}(t) \parallel^2$$

and the Itakura-Saito spectral distortion measure

$$d_{\mathrm{IS}}(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{S(\omega,t)}{\widehat{S}(\omega,t)} - \log\frac{S(\omega,t)}{\widehat{S}(\omega,t)} - 1\right]d\omega$$

where

$$S(\omega,t) = \frac{\rho(t)}{|A[e^{j\omega}, \boldsymbol{\theta}(t)]|^2}\ , \quad \widehat{S}(\omega,t) = \frac{\widehat{\rho}(t)}{|A[e^{j\omega}, \widehat{\boldsymbol{\theta}}(t)]|^2}$$

$$A[z, \boldsymbol{\theta}(t)] = 1 - \sum_{i=1}^{n} a_i(t) z^{-i}$$

and $\omega \in (-\pi, \pi]$ denotes the normalized angular frequency. We note that under the local stationarity framework developed by Dahlhaus [Dahlhaus (2012)], the instantaneous spectral density function $S(\omega, t)$ of a time-varying AR process (1) is a well and uniquely defined characteristic.

Table 1 shows identification results – the measures $d_{\mathrm{PAR}}(t)$ and $d_{\mathrm{IS}}(t)$ averaged over $t \in [1, T_s]$ and 100 independent realizations of $\{y(t)\}$ – obtained for 4 fixed-bandwidth NWLS/NWYW algorithms ($k_1, \ldots, k_4$), and 3 variable-bandwidth algorithms based on minimization of the FPE statistics, the PPE statistics ($M = 35$), and the symmetrized PPE statistic, respectively. Both FPE-based and PPE-based algorithms with adaptive bandwidth scheduling work satisfactorily and yield comparable results, usually better than the results provided by the fixed-bandwidth algorithms. The symmetrized PPE rule works slightly better than the one-sided rule.

Finally, Fig. 3 shows the locally time averaged histograms of the results of bandwidth selection for the NWYW algorithm (each time bin covers 500 consecutive time instants) obtained for 100 process realizations. The analogous histograms for the NWLS algorithm look almost identically. On the qualitative level, the PPE-based selection seems to work better than the FPE-based one.

The important advantage of the PPE-based approach is its wider range of applicability. While the generalized FPE criterion (21) can be used for evaluation of performance of the NWLS/NWYW algorithms only, the PPE criterion (18) can be extended to other classes of adaptive identification algorithms, such as the Kalman filter based parameter smoothers described in [Niedźwiecki (2012)]. Hence, using this approach, one can control banks of adaptive filters made up of noncausal algorithms based on different principles.

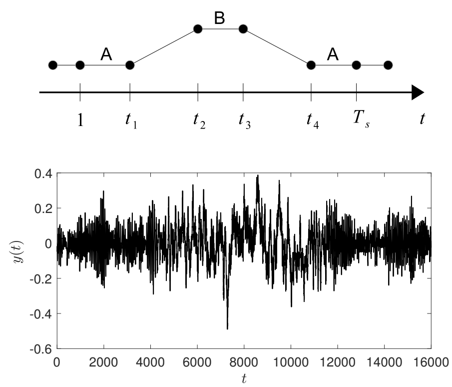

Fig. 2. Simulation scenario (top figure) and a typical realization of the corresponding nonstationary AR process (bottom figure).

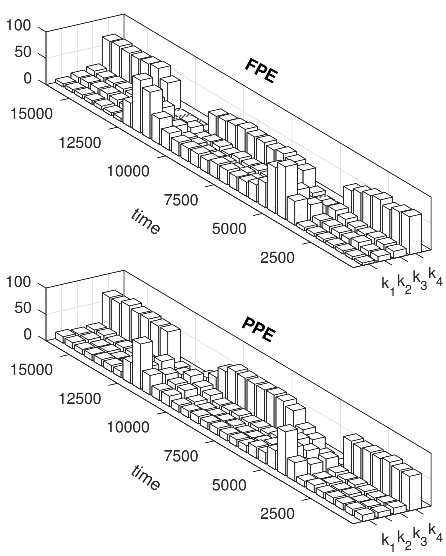

Fig. 3. Locally time averaged histograms of the results of bandwidth selection obtained for the NWYW algorithm ($T_s = 16000$).

Table 1. Identification results obtained for 4 fixed-bandwidth NWLS/NWYW algorithms ($k_1, \ldots, k_4$) and 3 variable-bandwidth algorithms based on minimization of the FPE statistics, the PPE statistics ($M = 35$), and the symmetrized PPE statistic, respectively. The results were obtained for 2 quality measures (mean squared parameter estimation error, mean Itakura-Saito spectral distortion measure) and 3 speeds of parameter variation SoV (S_1, S_2, S_3).

parameter estimation errors

	SoV	k_1	k_2	k_3	k_4	FPE	PPE	PPE*
NWLS	S_1	0.366	0.251	0.188	0.190	0.236	0.207	0.205
	S_2	0.369	0.268	0.245	0.343	0.247	0.225	0.223
	S_3	0.434	0.427	0.629	1.260	0.323	0.319	0.316

	SoV	k_1	k_2	k_3	k_4	FPE	PPE	PPE*
NWYW	S_1	0.423	0.256	0.189	0.190	0.199	0.211	0.207
	S_2	0.431	0.275	0.246	0.344	0.212	0.231	0.227
	S_3	0.490	0.431	0.629	1.260	0.289	0.323	0.319

spectrum estimation errors

	SoV	k_1	k_2	k_3	k_4	FPE	PPE	PPE*
NWLS	S_1	0.042	0.028	0.021	0.023	0.028	0.023	0.022
	S_2	0.043	0.031	0.029	0.041	0.029	0.025	0.025
	S_3	0.052	0.051	0.070	0.119	0.040	0.037	0.036

	SoV	k_1	k_2	k_3	k_4	FPE	PPE	PPE*
NWYW	S_1	0.042	0.027	0.021	0.023	0.023	0.022	0.022
	S_2	0.043	0.030	0.029	0.041	0.024	0.025	0.024
	S_3	0.053	0.050	0.070	0.119	0.035	0.037	0.036

REFERENCES

Baddour, K.E. & Beaulieu, N.C. (2005). Autoregressive models for fading channel simulation. *IEEE Trans. Wireless Comm.*, (4), 1650–1662.

Brillinger, D., Robinson, E.A. & Schoenberg, F.P. Eds. (2012). *Time Series Analysis and Applications to Geophysical Systems*. Springer.

Dahlhaus, R. & Giraitis, L. (1998). On the optimal segment length for parameter estimates for locally stationary time series. *J. Time Series Anal.*, (19), 629–655.

Dahlhaus, R. (2012). Locally stationary processes. *Handbook Statist.*, (25), 1–37.

Fabri, S.G., Camilleri, K.P. & Cassar, T. (2011). Parametric modelling of EEG data for the identification of mental tasks. *Biomed. Eng. Trends in Electron., Commun. Software*, (A. Laskovski Ed.), 367-386.

Hayes, J.F. & Ganesh Babu, T.V.J. (2004). *Modeling and Analysis of Telecommunication Networks*. Wiley.

Lesage, P., Glangeaud, F. & Mars, J. (2002). Applications of autoregressive models and time-frequency analysis to the study of volcanic tremor and long-period events. *J. Volc. Geotherm. Res.*, (114), 391417.

Li, C. & Nowack, R.L. (2004). Application of autoregressive extrapolation to seismic tomography. *Bull. Seism. Soc. Amer.*, 1456-1466.

Niedźwiecki, M. (1990). Identification of nonstationary stochastic systems using parallel estimation schemes. *IEEE Trans. Automat. Contr.*, (35), 329–334.

Niedźwiecki, M. (1992). Multiple model approach to adaptive filtering. *IEEE Trans. Signal Process.*, (40), 470–474.

Niedźwiecki, M. (2000). *Identification of Time-varying Processes*. Wiley.

Niedźwiecki, M. & Gackowski, S. (2011). On noncausal weighted least squares identification of nonstationary stochastic systems. *Automatica*, (47), 2239–2245.

Niedźwiecki, M. (2012). Locally adaptive cooperative Kalman smoothing and its application to identification of nonstationary stochastic systems. *IEEE Trans. Signal Process.*, (60), 48–59.

Niedźwiecki, M., Ciołek, M. & Kajikawa, Y. (2017). On adaptive covariance and spectrum estimation of locally stationary multivariate processes. *Automatica*, (82), 1–12.

Schlögl, A. (2000). *The Electroencephalogram and the Adaptive Autoregressive Model: Theory and Applications*. Aachen, Germany: Shaker Verlag.

Söderström, T. & Stoica, P. (1988) *System Identification*, Englewood Cliffs NJ: Prentice-Hall.

Wada, T., Jinnouchi, M. & Matsumura, Y. (1998). Application of autoregressive modelling for the analysis of clinical and other biological data. *Ann. Inst. Statist. Math.*, (40), 211-227.