On Joint Order and Bandwidth Selection for Identification of Nonstationary Autoregressive Processes

Maciej Niedźwiecki and Marcin Ciołek

Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control
Gdańsk University of Technology, ul. Narutowicza 11/12, Gdańsk, Poland

maciekn@eti.pg.gda.pl, marcin.ciolek@pg.gda.pl

Abstract—**When identifying a nonstationary autoregressive process, e.g. for the purpose of signal prediction or parametric spectrum estimation, two important decisions must be taken. First, one should choose the appropriate order of the autoregressive model, i.e., the number of autoregressive coefficients that will be estimated. Second, if identification is carried out using the local estimation technique, such as the localized version of the method of least squares, one should select the most appropriate estimation bandwidth, i.e., the effective width of the local data window used for the purpose of parameter tracking. The paper presents the first unified treatment of the problem of joint order and bandwidth selection. Two solutions to this problem are examined, first based on the predictive least squares principle, and second exploiting the suitably modified Akaike's final prediction error statistic. It is shown that the best results are obtained if the two approaches mentioned above are combined.**

I. INTRODUCTION

Autoregressive signal modeling is a basis for practical applications in different research areas such as biology and biomedicine, telecommunications, seismology etc. [1] - [9]. Since in majority of cases the analyzed signals are nonstationary, the time-varying autoregressive representation has to be used. When signal characteristics vary slowly with time, the time-varying autoregressive models can be obtained using the local estimation techniques. In this approach the measure of fit, minimized by the identification routine, is modified in such a way that estimation results depend only or mainly on the most recently collected data samples.

Two important decisions that must be taken when identifying the time-varying autoregressive model are the choice of the number of estimated autoregressive coefficients, i.e., the model order, and selection of the size of the local analysis interval, i.e., the estimation bandwidth (frequency range in which parameter changes can be tracked "successfully"). Both decisions may have important quantitative (estimation accuracy) and quantitative (estimation adequacy) implications.

For stationary signals order estimation is a well-explored statistical problem, which can be solved in many different ways. The most popular solutions are those based on the Akaike information criterion (AIC) [10], Schwarz criterion, frequently referred to as the Bayesian information criterion (BIC) [11], and Rissanen's minimum description length (MDL) criterion [12]. Generalized versions of the AIC and BIC criteria, applicable to local estimation schemes, were proposed in [13] and [14], respectively.

Selection of the estimation bandwidth for the purpose of analysis of nonstationary signals is a less investigated topic. The solution based on prequential analysis [15] was proposed in [16], [17]. Another solution, based on the analysis of the intersection of the confidence intervals (ICI), was suggested in [18] and further developed in [19] and [20]. Finally, for noncausal estimation schemes, the problem of bandwidth selection was solved using the cross-validation approach [21], [22].

To the authors best knowledge the problem of joint order and bandwidth selection has not been considered so far.

II. STATIONARY AUTOREGRESSIVE PROCESSES

Stationary autoregressive (AR) process of order n is governed by the equation

$$y(t) = \sum_{i=1}^{n} a_i y(t-i) + e(t), \quad \text{var}[e(t)] = \rho_n \qquad (1)$$

where t denotes the discrete (normalized) time, $\{e(t)\}$ is a sequence of zero-mean independent random variables and all zeros of the characteristic polynomial $A_n(z^{-1}) = 1 - \sum_{i=1}^{n} a_i z^{-i}$ lie inside the unit circle in the complex plane. Note that the AR equation (1) can be also put down in the form

$$y(t) = \boldsymbol{\varphi}_n^{\mathrm{T}}(t)\boldsymbol{\theta}_n + e(t) \qquad (2)$$

where $\boldsymbol{\varphi}_n(t) = [y(t-1), \ldots, y(t-n)]^{\mathrm{T}}$ denotes regression vector and $\boldsymbol{\theta}_n = [a_1, \ldots, a_n]^{\mathrm{T}}$ is the vector of AR coefficients.

The spectral density function of the AR process (1) can be expressed in the form

$$S(\omega) = \frac{\rho_n}{|A_n(e^{-j\omega})|^2} \qquad (3)$$

where $\omega \in [0, \pi]$ denotes normalized angular frequency and $j = \sqrt{-1}$.

III. NONSTATIONARY AUTOREGRESSIVE PROCESSES AND THEIR IDENTIFICATION

In this paper we will consider nonstationary AR process governed by

$$y(t) = \boldsymbol{\varphi}_n^{\mathrm{T}}(t)\boldsymbol{\theta}_n(t) + e(t), \quad \mathrm{var}[e(t)] = \rho_n(t) \qquad (4)$$

where $\boldsymbol{\theta}_n(t) = [a_1(t), \ldots, a_n(t)]^{\mathrm{T}}$ is the vector of time-varying AR coefficients.

When the coefficients vary slowly with time, the AR process can be regarded as locally stationary and can be identified using local estimation techniques. An elegant theory of locally stationary processes (under agreeable technical assumptions – such as uniform stability and bounded parameter variation – the time-varying AR process fits this framework) was recently presented by Dahlhaus [23]. One of the key elements of this theory is the concept of instantaneous spectrum, which for (4) is defined as

$$S(\omega, t) = \frac{\rho_n(t)}{|A_n(e^{-j\omega}, t)|^2} \qquad (5)$$

where

$$A_n(z^{-1}, t) = 1 - \sum_{i=1}^{n} a_i(t) z^{-i}.$$

Identification of the model (4) will be carried out using the technique known as exponentially weighted least squares (EWLS). Denote by λ_k, $0 < \lambda_k < 1$, the so-called forgetting constant. The EWLS estimates of $\boldsymbol{\theta}_n(t)$ and $\rho(t)$ can be obtained from

$$\widehat{\boldsymbol{\theta}}_{n|k}(t) = [\widehat{a}_{1,n|k}(t), \ldots, \widehat{a}_{n,n|k}(t)]^{\mathrm{T}}$$
$$= \arg\min_{\boldsymbol{\theta}_n} \sum_{i=0}^{t-1} \lambda_k^i [y(t-i) - \boldsymbol{\varphi}_n^{\mathrm{T}}(t-i)\boldsymbol{\theta}_n]^2$$
$$\widehat{\rho}_{n|k}(t) = \frac{\beta_{n|k}(t)}{L_k(t)} \qquad (6)$$
$$\beta_{n|k}(t) = \sum_{i=0}^{t-1} \lambda_k^i [y(t-i) - \boldsymbol{\varphi}_n^{\mathrm{T}}(t-i)\widehat{\boldsymbol{\theta}}_{n|k}(t)]^2$$

where $L_k(t) = \sum_{i=0}^{t-1} \lambda_k^i$ denotes the effective width of the exponential window (which converges to $L_k(\infty) = 1/(1-\lambda_k)$ as t becomes large). It is well-known that the EWLS estimates can be computed recursively [24]

$$\widehat{\boldsymbol{\theta}}_{n|k}(t) = \widehat{\boldsymbol{\theta}}_{n|k}(t-1) + \mathbf{R}_{n|k}^{-1}(t)\boldsymbol{\varphi}_n(t)\varepsilon_{n|k}(t)$$
$$\mathbf{R}_{n|k}(t) = \lambda_k \mathbf{R}_{n|k}(t-1) + \boldsymbol{\varphi}_n(t)\boldsymbol{\varphi}_n^{\mathrm{T}}(t)$$
$$\beta_{n|k}(t) =$$
$$\lambda_k \left[\beta_{n|k}(t-1) + \frac{\varepsilon_{n|k}^2(t)}{\lambda_k + \boldsymbol{\varphi}_n^{\mathrm{T}}(t)\mathbf{R}_{n|k}^{-1}(t-1)\boldsymbol{\varphi}_n(t)} \right]$$
$$L_k(t) = \lambda_k L_k(t-1) + 1$$

where

$$\varepsilon_{n|k}(t) = y(t) - \boldsymbol{\varphi}_n^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_{n|k}(t-1)$$

denotes the one-step-ahead prediction error. Moreover, the inverse of the exponentially weighted regression matrix

$$\mathbf{P}_{n|k}(t) = \mathbf{R}_{n|k}^{-1}(t) = \left[\sum_{i=0}^{t-1} \lambda_k^i \boldsymbol{\varphi}_n(t-i)\boldsymbol{\varphi}_n^{\mathrm{T}}(t-i) \right]^{-1}$$

can be also evaluated in a recursive manner.

$$\mathbf{P}_{n|k}(t)$$
$$= \frac{1}{\lambda_k} \left[\mathbf{P}_{n|k}(t-1) - \frac{\mathbf{P}_{n|k}(t-1)\boldsymbol{\varphi}_n(t)\boldsymbol{\varphi}_n^{\mathrm{T}}(t)\mathbf{P}_{n|k}(t-1)}{\lambda_k + \boldsymbol{\varphi}_n^{\mathrm{T}}(t)\mathbf{P}_{n|k}(t-1)\boldsymbol{\varphi}_n(t)} \right]$$

Based on the model (6), one arrives at the following estimate of the instantaneous spectrum

$$\widehat{S}_{n|k}(\omega, t) = \frac{\widehat{\rho}_{n|k}(t)}{|\widehat{A}_{n|k}(e^{-j\omega}, t)|^2} \qquad (7)$$

where

$$\widehat{A}_{n|k}(z^{-1}, t) = 1 - \sum_{i=1}^{n} \widehat{a}_{i,n|k}(t) z^{-i}.$$

Properties of the spectral estimate (7), both quantitative and qualitative, strongly depend on the choice of forgetting constant λ_k and the adopted order of autoregression n. Forgetting constant decides upon the estimation bandwidth which should be chosen in accordance with the speed of parameter variation. When parameters vary at a slow rate, λ_k should be set close to 1, which guarantees large estimation memory $L_k(\infty)$, i.e., small variance of parameter estimation errors. When parameter variation is (relatively) fast, estimation memory should be shortened (by moving λ_k away from 1) to reduce bias errors. If the speed of parameter variation is not known *a priori*, or if it also changes over time, resolution of the bias-variance trade-off may be a difficult task. The choice of the model order n has important qualitative implications. If the order of autoregression is underestimated, the corresponding spectral estimates may not reveal some important frequency components of the analyzed signal. When it is overestimated, the corresponding spectrum estimate may show spurious peaks suggesting the presence of nonexistent frequency modes.

When the signal is nonstationary decision upon the model order and the most appropriate estimation bandwidth should be made in an adaptive, data-dependent fashion. This task can be accomplished using parallel estimation schemes, i.e., by running simultaneously several estimation algorithms with different bandwidth and order settings, and choosing the one which provides the best estimates according to some local performance measure.

Remark

For a given value of λ_k the estimates $\widehat{\boldsymbol{\theta}}_{n|k}(t)$, $\widehat{\rho}_{n|k}(t)$, $n = 1, \ldots, N$ can be evaluated in a computationally efficient way using time- and order-recursive algorithms based on Hausholder transformation [25], or using the least squares lattice (ladder) algorithms [26], [27].

IV. Adaptive order and bandwidth selection

Consider K EWLS algorithms, equipped with forgetting constants $\lambda_1, \ldots, \lambda_K$, working in parallel, each providing N structural variants of the local AR model corresponding to different hypothetical model orders $n = 1, \ldots, N$. The selection task amounts to finding the most appropriate local estimates of n and k. Based on these estimates, the instantaneous spectrum estimate can be obtained in the form

$$\widehat{S}_{\widehat{n}(t)|\widehat{k}(t)}(\omega, t) = \frac{\widehat{\rho}_{\widehat{n}(t)|\widehat{k}(t)}(t)}{|\widehat{A}_{\widehat{n}(t)|\widehat{k}(t)}(e^{-j\omega}, t)|^2}. \tag{8}$$

A. Predictive least squares

Predictive least squares (PLS) statistic was originally proposed by Rissanen [28], [29] as a tool for model order estimation in the case where the identified process is stationary. Later on, the localized version of this statistic was successfully used to select estimation bandwidth of competing finite-memory adaptive filters [16], [17]. The local PLS measure of fit is given by

$$\text{PLS}_{n,k}(t) = \sum_{i=0}^{l-1} \varepsilon_{n|k}^2(t - i) \tag{9}$$

where $l \in [20, 50]$ is the width of the local decision window $T(t) = [t - l + 1, t]$.

The estimates of n and k can be obtained by minimizing (9) over $\mathcal{K} = \{1, \ldots, K\}$ and $\mathcal{N} = \{1, \ldots, N\}$

$$\{\widehat{n}(t), \widehat{k}(t)\} = \arg \min_{\substack{n \in \mathcal{N} \\ k \in \mathcal{K}}} \text{PLS}_{n,k}(t). \tag{10}$$

According to (10), from NK competing models we choose the one with the "recently the best" prediction record.

B. Final prediction error

Denote by $\widetilde{y}(t)$ another realization of the analyzed nonstationary process, independent of $y(t)$ used for process identification. As a measure of predictive capability of the model characterized by the vector of parameters $\widehat{\boldsymbol{\theta}}_{n|k}(t)$, one can use the following quantity

$$\delta_{n|k}(t) = \text{E}\{[\widetilde{y}(t) - \widetilde{\boldsymbol{\varphi}}_n^{\text{T}}(t)\widehat{\boldsymbol{\theta}}_{n|k}(t)]^2\} \tag{11}$$

where the expectation is taken over $\widetilde{y}(t)$ *and* $y(t)$. Note that according to (11) the quality of the model is checked on an independent data set, different from the one used to estimate model parameters.

Assume that n is not smaller than the true order of the analyzed AR process. We will work out a "stationary" approximation of $\delta_{n|k}(t)$. Suppose that $y(t)$ is governed by (2). Then it holds that $\text{E}[\widehat{\boldsymbol{\theta}}_{n|k}(t)] \cong \boldsymbol{\theta}_n$ and [24]

$$\text{cov}[\widehat{\boldsymbol{\theta}}_{n|k}(t)] \cong \frac{\rho_n \boldsymbol{\Phi}_0^{-1}}{M_k(t)} \tag{12}$$

where $M_k(t) = (\sum_{i=0}^{t-1} \lambda_k^i)^2 / (\sum_{i=0}^{t-1} \lambda_k^{2i})$ denotes the so-called equivalent window width and $\boldsymbol{\Phi}_0 = \text{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\text{T}}(t)]$.

Table I
DESCRIPTION OF FOUR VARIANTS OF MIXED SELECTION

Variant	A	B	C	D
Step 1	FPE(n)	PLS(k)	PLS(n)	FPE(k)
Step 2	PLS(k)	FPE(n)	FPE(k)	PLS(n)

Let $\Delta\widehat{\boldsymbol{\theta}}_{n|k}(t) = \widehat{\boldsymbol{\theta}}_{n|k}(t) - \boldsymbol{\theta}_n$. Note that

$$\begin{aligned}
\delta_{n|k}(t) &= \text{E}\{[\widetilde{e}(t) - \widetilde{\boldsymbol{\varphi}}_n^{\text{T}}(t)\Delta\widehat{\boldsymbol{\theta}}_{n|k}(t)]^2\} \\
&= \rho_n + \text{E}\{\widetilde{\boldsymbol{\varphi}}_n^{\text{T}}(t)\text{cov}[\widehat{\boldsymbol{\theta}}_{n|k}(t)]\widetilde{\boldsymbol{\varphi}}_n(t)\} \\
&= \rho_n + \text{tr}\{\text{cov}[\widehat{\boldsymbol{\theta}}_{n|k}(t)]\boldsymbol{\Phi}_0\} \cong \rho_n \left[1 + \frac{n}{M_k(t)}\right].
\end{aligned} \tag{13}$$

On the other hand, it holds that [24]

$$\text{E}[\widehat{\rho}_{n|k}(t)] \cong \rho_n \left[1 - \frac{n}{M_k(t)}\right] \tag{14}$$

leading to the following estimate of $\delta_{n|k}(t)$

$$\widehat{\delta}_{n|k}(t) = \widehat{\rho}_{n|k}(t) \left[\frac{1 + \frac{n}{M_k(t)}}{1 - \frac{n}{M_k(t)}}\right] = \text{FPE}_{n,k}(t) \tag{15}$$

which is identical with the generalized version of Akaike's final prediction error (FPE) statistic proposed in [30] as a tool for model order selection. The analysis carried out above shows that the same statistic can be used for the purpose of choosing the estimation bandwidth. The joint order and bandwidth selection rule takes the form

$$\{\widehat{n}(t), \widehat{k}(t)\} = \arg \min_{\substack{n \in \mathcal{N} \\ k \in \mathcal{K}}} \text{FPE}_{n,k}(t). \tag{16}$$

Note that for large values of t the quantity $M_k(t)$ in (16) can be replaced with its steady state value $M_k(\infty) = (1+\lambda_k)/(1-\lambda_k)$.

C. Mixed solutions

Table 1 specifies 4 mixed variants obtained when one criterion is used for model order selection and the other one - for bandwidth selection. In the first case (A), initially the best value of n is searched for each value of k, then the best value of k is selected

$$\begin{aligned}
\widehat{n}_k(t) &= \arg \min_{n \in \mathcal{N}} \text{FPE}_{n,k}(t), \quad k \in \mathcal{K} \\
\widehat{k}(t) &= \arg \min_{k \in \mathcal{K}} \text{PLS}_{\widehat{n}_k(t),k}(t).
\end{aligned} \tag{17}$$

In the second case (B) the same operations are performed in the reverse order

$$\begin{aligned}
\widehat{k}_n(t) &= \arg \min_{k \in \mathcal{K}} \text{PLS}_{n,k}(t), \quad n \in \mathcal{N} \\
\widehat{n}(t) &= \arg \min_{n \in \mathcal{N}} \text{FPE}_{n,\widehat{k}_n(t)}(t).
\end{aligned} \tag{18}$$

The remaining two variants (C, D) use PLS for order selection and FPE for bandwidth selection.

Table II
VALUES OF THE COEFFICIENTS r_i AND ϕ_i USED IN SIMULATION

i	1	2	3	4	5
r_i	0.9852	0.8558	0.9480	0.9168	0.8554
ϕ_i	0.5197	0.9709	1.4047	1.8977	2.6865

Figure 1. Pole trajectories of a nonstationary AR model.

V. SIMULATION RESULTS

To verify the proposed methods, a time-varying variable-order AR process was generated using the pole location technique.

During the first phase of simulation five pairs of complex conjugate poles of the forming filter $1/A(z^{-1}, t)$, i.e., the roots of the characteristic polynomial $A(z^{-1}, t)$, were gradually moved (one at a time) from their initial zero positions to terminal locations $z_i^\pm = r_i e^{\pm j\phi_i}$, $i = 1, \dots, 5$, close to the unit circle - see Figure 1 and Table 2. The pole locus changed according to the following rule

$$\gamma_1(t) r_1 e^{\pm j\phi_1} \quad \text{for} \quad t \in T_1$$
$$r_1 e^{\pm j\phi_1}, \dots, r_{i-1} e^{\pm j\phi_{i-1}}, \gamma_i(t) r_i e^{\pm j\phi_i} \quad \text{for} \quad t \in T_i, i > 1 \tag{19}$$

where $T_i = \{800(i-1) + l, l = 1, \dots, 800\}$ and $\gamma_i(t) = [t - 800(i-1)]/800$. According to (19), in the i-th period of time T_i the i-th pair of complex poles was moved (with a constant speed) towards the unit circle, while the first $i - 1$ pairs were held at fixed positions. Note that the order of the resulting AR process gradually increased from $n = 2$ (for $t \in T_1$) to $n = 10$ (for $t \in T_5$).

During the second phase of simulation the poles were moved back to the origin in an analogous way as described above and in the same order, i.e., starting from $z_1^\pm$ and ending with $z_5^\pm$. Evolution of the spectral density function of the process generated in this way and an exemplary process realization are shown in Figure 2.

Table 3 shows the mean values of the Itakura-Saito spectral distortion measure (averaging was carried out over time and 20 realizations of a nonstationary AR process) yielded by 5 EWLS algorithms ($\lambda_1 = 0.9775$, $\lambda_2 = 0.985$, $\lambda_3 = 0.99$, $\lambda_4 = 0.9933$, $\lambda_5 = 0.9955$) of different orders ($n = 1, \dots, 20$) and by 4 variants of the parallel estimation scheme: the variant (10) based entirely on the predictive least squares statistic, the variant (16) based entirely on the final prediction error statistic, and two mixed variants (17) and (18) (the other mixed variants gave worse results).

The best results were obtained for the mixed variants, which is a straightforward consequence of the fact that, as observed, the FPE criterion is more successful in selecting the model order, and the PLS criterion - in choosing the estimation bandwidth. The same conclusions can be reached after examining the mean squared parameter estimation errors and mean squared prediction errors (not shown here due to the lack of space).
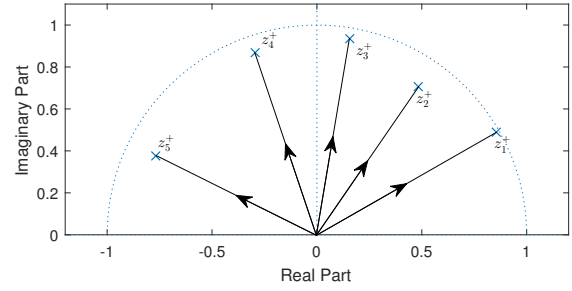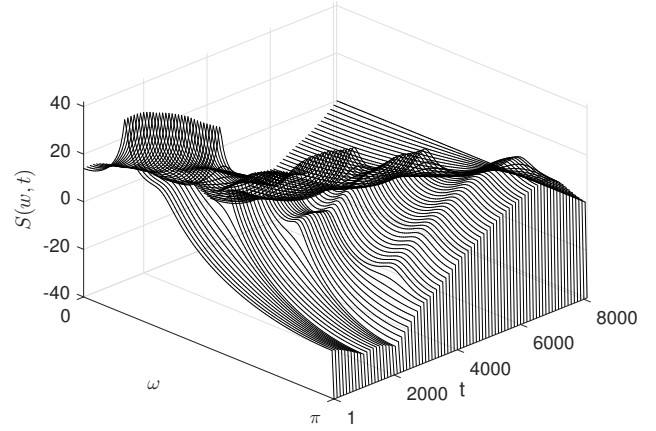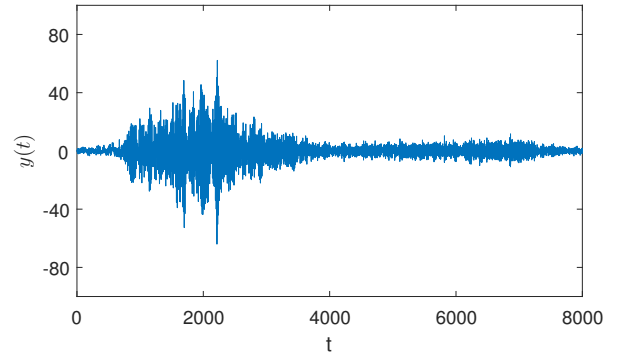

(a)

(b)

Figure 2. Evolution of the true spectral density function of a simulated process (a) and a typical process realization (b).

VI. CONCLUSION

Selection of the order of a nonstationary autoregressive process and the choice of the most appropriate estimation bandwidth are two important problems that so far have been considered and solved separately. The paper presented the first unified treatment of both problems. Two methods were proposed that can be used for joint order and bandwidth selection - the predictive least squares approach and the

Table III

COMPARISON OF ESTIMATION RESULTS (MEAN SPECTRAL DISTORTION MEASURES) OBTAINED FOR 5 FIXED-ORDER ($n = 1, \ldots, 20$) EWLS ALGORITHMS WITH DIFFERENT FORGETTING CONSTANTS $\lambda_1, \ldots, \lambda_5$, WITH THE RESULTS YIELDED BY 4 ORDER-AND-BANDWIDTH-ADAPTIVE PARALLEL ESTIMATION SCHEMES: THE ONE BASED EXCLUSIVELY ON THE PLS STATISTIC, THE ONE BASED EXCLUSIVELY ON THE FPE STATISTICS AND TWO MIXED SOLUTIONS A AND B.

n/N	λ_1	λ_2	λ_3	λ_4	λ_5	PLS	FPE	A	B
1	2.147	2.132	2.131	2.148	2.183	2.160	2.176	2.160	2.160
2	0.967	0.948	0.948	0.974	1.037	0.967	0.968	0.965	0.965
3	0.764	0.743	0.743	0.771	0.844	0.767	0.765	0.760	0.762
4	0.332	0.313	0.318	0.353	0.442	0.333	0.331	0.325	0.325
5	0.284	0.261	0.264	0.301	0.399	0.282	0.277	0.271	0.272
6	0.165	0.140	0.143	0.182	0.290	0.160	0.153	0.148	0.148
7	0.168	0.138	0.137	0.173	0.274	0.156	0.149	0.143	0.143
8	0.153	0.118	0.115	0.149	0.243	0.133	0.123	0.118	0.118
9	0.157	0.118	0.112	0.146	0.238	0.128	0.116	0.112	0.111
10	0.164	0.118	0.108	0.136	0.221	0.124	0.111	0.107	0.106
11	0.179	0.127	0.113	0.138	0.218	0.126	0.112	0.108	0.107
12	0.196	0.137	0.119	0.141	0.217	0.129	0.114	0.110	0.109
13	0.212	0.147	0.125	0.145	0.219	0.132	0.115	0.111	0.110
14	0.230	0.157	0.132	0.149	0.222	0.134	0.116	0.112	0.111
15	0.247	0.168	0.138	0.154	0.225	0.136	0.116	0.113	0.111
16	0.264	0.177	0.144	0.157	0.227	0.137	0.116	0.113	0.112
17	0.282	0.188	0.151	0.161	0.229	0.139	0.117	0.114	0.112
18	0.303	0.199	0.158	0.166	0.232	0.141	0.117	0.114	0.113
19	0.324	0.212	0.166	0.171	0.235	0.142	0.117	0.115	0.113
20	0.344	0.223	0.173	0.175	0.238	0.144	0.118	0.115	0.114

approach based on the modified Akaike's final prediction error criterion. It was shown that the best results can be obtained when both approaches are appropriately combined.

REFERENCES

[1] T. Wada, M. Jinnouchi, and Y. Matsumura, "Application of autoregressive modelling for the analysis of clinical and other biological data," *Ann. Inst. Statist. Math.*, vol. 40, pp. 211-227, 1998.

[2] V.K. Jirsa and A.R. McIntosh, Eds., *Handbook of Brain Connectivity*, Springer-Verlag, 2007.

[3] N. Channouf, P. L'Ecuyer, A. Ingolfsson, and A.N. Avramidis, "The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta," Health Care Manag. Sci., vol. 10, pp. 25–45, 2007.

[4] R. Takalo, H. Hytti, H. Ihalainen, and A. Sohlberg, *Adaptive autoregressive model for reduction of noise in SPECT*, Comp. Math. Methods in Medicine, vol. 2015, 9 p., 2015.

[5] K.E. Baddour and N.C. Beaulieu, "Autoregressive models for fading channel simulation," IEEE Trans. Wireless Comm., vol. 4, pp. 1650–1662, 2005.

[6] J.F. Hayes and T.V.J. Ganesh Babu, *Modeling and Analysis of Telecommunication Networks,* Wiley, 2004.

[7] C. Li and R.L. Nowack, "Application of autoregressive extrapolation to seismic tomography," Bull. Seism. Soc. Amer., vol. 94, pp. 1456?-1466, 2004.

[8] P. Lesage, F. Glangeaud, and J. Mars, "Applications of autoregressive models and time-frequency analysis to the study of volcanic tremor and long-period events," *J. Volc. Geotherm. Res.*, vol. 114, pp. 391?-417, 2002.

[9] D. Brillinger, E.A. Robinson, and F.P. Schoenberg, Eds., *Time Series Analysis and Applications to Geophysical Systems*, Springer, 2012.

[10] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. 19, pp. 716–723, 1974.

[11] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, pp. 461–464, 1978.

[12] J. Rissanen, "Modeling by shortest data descriptiona," *Automatica*, vol. 14, pp. 465-658, 1978.

[13] M. Niedźwiecki, "On the localized estimators and generalized Akaike's criteria," *IEEE Trans. Automat. Contr.*, vol. 29, pp. 970–983, 1984.

[14] M. Niedźwiecki, "Bayesian-like autoregressive spectrum estimation in the case of unknown process order," *IEEE Trans. Automat. Contr.*, vol. 30, pp. 950–961, 1985.

[15] A.P. Dawid, "Present position and potential developments: some personal view, statistical theory, the prequential approach," *J. Roy. Statist. Soc. A*, vol. 147, pp. 278–292, 1984.

[16] M. Niedźwiecki, "Identification of nonstationary stochastic systems using parallel estimation schemes," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 329–334, 1990.

[17] M. Niedźwiecki, "Multiple model approach to adaptive filtering,", *IEEE Trans. Signal Process.*, vol. 40, pp. 470–474, 1992.

[18] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Math. Meth. Stat.*, vol. 6, pp. 135–170, 1997.

[19] V. Katkovnik, "A new method for varying adaptive bandwidth selection," *IEEE Trans. Signal Process.*, vol. 47, pp. 2567–2571, 1999.

[20] L. Stanković, "Performance analysis of the adaptive algorithm for bias-to-variance tradeoff," *IEEE Trans. Signal Process.*, vol. 52, pp. 1228–1234, 2004.

[21] M. Niedźwiecki, "Locally adaptive cooperative Kalman smoothing and its application to identification of nonstationary stochastic systems," *IEEE Trans. Signal Process.*, vol. 60, pp. 48–59, 2012.

[22] M. Niedźwiecki, M. Ciołek, and Y. Kajikawa, "On adaptive selection of estimation bandwidth for analysis of locally stationary multivariate processes," *ICASSP 2016 – Proc. 2016 IEEE Int. Conf. Acoust. Speech Sign. Process.*, Shanghai, China, pp. 4860–4864, 2016.

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

[24] M. Niedźwiecki, *Identification of Time-varying Processes*, Wiley, 2000.

[25] V. Peterka, "A square root filter for real time multivariate regression," *Kybernetika*, vol. 11, 53–67, 1975.

[26] D.T.L. Lee, M. Morf, and B. Friedlander, "Recursive least-squares ladder estimation algorithms," *IEEE Trans. Circuits Syst.*, vol. 28, pp. 467–481, 1981.

[27] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, 2003.

[28] J. Rissanen and V. Wertz, "Structure estimation by accumulated prediction error criterion," *7th IFAC Symposium on Identification and System Parameter Estimation*, York, U.K., pp. 757–759, 1985.

[29] J. Rissanen, "A predictive least squares principle," *IMA J. Math. Control Inform.*, Vol. 3, pp. 211-222, 1986.

[30] H. Akaike, "Statistical predictor identification," *Ann. Inst. Statist. Math.*, vol. 32, pp. 203–217, 1970.