This is a post-peer-review, pre-copyedit version of an article published in 2017 IEEE 56th Annual Conference on Decision and Control (CDC).
The final authenticated version is available online at: https://doi.org/10.1109/CDC.2017.8264188

On autoregressive spectrum estimation using the model averaging technique

Michał Meller, Maciej Niedźwiecki, Damian Chojnacki and Adam Lasota

***Abstract*— The problem of estimating spectral density of a nonstationary process satisfying local stationarity conditions is considered. The proposed solution is a two step procedure based on local autoregressive (AR) modeling. In the first step Bayesian-like averaging of AR models, differing in order, is performed. The main contribution of the paper is development of a new final-prediction-error-like statistic, which can be used to select optimal estimation bandwidth in the second step of the procedure. Simulation experiments demonstrate that the combined cooperative-competitive approach outperforms the previously introduced fully competitive scheme.**

I. Introduction

Spectral estimation of stationary processes can be performed using multitude of methods, among which autoregressive (AR) modeling takes an important place. This is due to its ease of use, accuracy, high resolution and direct links with the Burg's maximum entropy analysis [1].

Recent advances in statistics show that AR modeling is an effective tool which can be used in spectral estimation of nonstationary processes [2]. Under such circumstances, the ever-changing spectral content of a nonstationary signal requires one to use local estimation techniques. This however, requires solution of two problems. First, the structure of the AR model should be parsimonious, yet offer enough capacity to accommodate the local spectral content. Second, the estimation bandwidth (related to the estimation memory of the parameter tracking algorithm) should be locally adjusted so as to match the rate of nonstationarity of the process.

In a recent paper [3] it was shown that the two problems mentioned above can be solved jointly using a competitive estimation scheme. The method is based on parallel estimation of multiple AR models, differing in order and bandwidth settings. At each time instant the best model is selected using the generalized Akaike's final prediction error criterion.

In this study we present a different solution, which combines the competitive approach with the cooperative one. Cooperation is achieved using the model averaging technique, introduced by Akaike [4]. Since model averaging using the Akaike's framework can be performed in a meaningful way only for models estimated using the same bandwidth settings, the optimal estimation bandwidth selection must be carried out separately. To this end we extend the final prediction error statistic to averaged AR models, which is the main contribution of this work.

The paper is organized as follows. Section II states the problem of interest and summarizes the proposed solution. The core part of the paper is Section III, where the final prediction error statistic is developed for averaged models. Section IV presents results of computer simulations and Section V concludes.

II. Problem statement and the main result

Consider the problem of estimation of a spectral density function of a nonstationary autoregressive (AR) signal $\{y(t)\}$ governed by

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

where $t = \ldots, -1, 0, 1, \ldots$ denotes normalized (dimensionless) time, $a_i(t), i = 1, \ldots, n$ denote time-varying autoregressive coefficients, and $\varepsilon_n(t)$ denotes white noise with variance $\rho_n(t)$. Note that (1) can be rewritten in a more compact form

$$y(t) = \boldsymbol{\varphi}_n^{\mathrm{T}}(t)\boldsymbol{\alpha}_n(t) + \varepsilon_n(t)$$

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

According to Dahlhaus [2], when $\{y(t)\}$ obeys local stationarity assumptions (i.e., when the AR model is uniformly stable and parameter trajectories are of bounded variation), its time-varying spectral density function given by

$$S_n(\omega, t) = \frac{\rho_n(t)}{\left|1 - \sum_{i=1}^{n} a_i(t)e^{-j\omega i}\right|^2} \quad (2)$$

where $\omega \in (-\pi, \pi]$ denotes normalized angular frequency and $j = \sqrt{-1}$, is a well and uniquely defined characteristic in the rescaled time domain (in this framework a fixed-length time interval is sampled over a finer and finer grid of points as the sample size increases).

When parameters of the AR model (1) are not known, one can replace them in the formula (2) with local estimates, e.g. with the estimates obtained using the weighted least squares

(WLS) approach

$$\widehat{\boldsymbol{\alpha}}_{n|k}(t) = [\widehat{a}_{1,n|k}(t), \ldots, \widehat{a}_{n,n|k}(t)]^{\mathrm{T}} = \arg\min_{\boldsymbol{\alpha}_n} \sum_{l=-k}^{k} w_k(l)[y(t+l) - \boldsymbol{\varphi}_n^{\mathrm{T}}(t+l)\boldsymbol{\alpha}_n]^2$$

$$\widehat{\rho}_{n|k}(t) = \frac{1}{L_k} \sum_{l=-k}^{k} w_k(l)[y(t+l) - \boldsymbol{\varphi}_n^{\mathrm{T}}(t+l)\widehat{\boldsymbol{\alpha}}_{n|k}(t)]^2 \tag{3}$$

where $w_k(l) = g(l/k)$, $l = -k, \ldots, k$, $g : [-1, 1] \rightarrow [0, 1], g(0) = 1$, denotes a symmetric bell-shaped window of effective width

$$L_k = \sum_{l=-k}^{k} w_k(l).$$

Such a two-sided (noncausal) estimation allows one to significantly reduce bias errors caused by the effect known as estimation delay [5] (in causal estimation schemes the expected trajectory of parameter estimates can be regarded, to a certain extent, as a delayed version of the true trajectory).

A. Competitive estimation

The value of k should be chosen in accordance with the rate of parameter variation, and the value of n – in accordance with the spectral richness (resonant structure) of the analyzed signal. When the most appropriate values of k and n are not known, one can simultaneously run several WLS algorithms with different window width and order settings, and select at each time instant t the best configuration.

Denote by $\mathcal{K} = \{k_1, \ldots, k_K\}$ the set of competing window widths, and by $\mathcal{N} = \{1, \ldots, N\}$ – the set of considered model orders. As shown in [3], the local estimates of k and n can be obtained using the suitably modified Akaike's final prediction error (FPE) criterion, namely

$$\{\widehat{k}(t), \widehat{n}(t)\} = \arg\min_{\substack{k \in \mathcal{K} \\ n \in \mathcal{N}}} \mathrm{FPE}_{n|k}(t) \tag{4}$$

where

$$\mathrm{FPE}_{n|k}(t) = \frac{1 + \frac{n}{M_k}}{1 - \frac{n}{M_k}} \widehat{\rho}_{n|k}(t) \tag{5}$$

and

$$M_k = \frac{\left(\sum_{l=-k}^{k} w_k(l)\right)^2}{\sum_{l=-k}^{k} w_k^2(l)}$$

is the so-called equivalent window width.

Based on (4), the instantaneous spectrum estimate can be expressed in the form

$$\widehat{S}_{\widehat{n}(t)|\widehat{k}(t)}(\omega, t) = \frac{\widehat{\rho}_{\widehat{n}(t)|\widehat{k}(t)}(t)}{\left|1 - \sum_{i=1}^{\widehat{n}(t)} \widehat{a}_{i,\widehat{n}(t)|\widehat{k}(t)}(t) e^{-j\omega i}\right|^2}. \tag{6}$$

B. Cooperative estimation

In the competitive approach one looks for the best-local values of k and n, ignoring the uncertainty embedded in the underlying decision process – the point estimates $\widehat{k}(t)$ and $\widehat{n}(t)$ are in some sense the most likely the best, but not certainly the best, choices of k and n, respectively. The uncertainty factor can be accounted for when estimation is carried out within the Bayesian framework. In this framework the estimated quantities, such as model parameters, are regarded as realizations of random variables with assigned prior distributions. Consider, for example, the problem of one-step-ahead prediction of a stationary AR signal based on its available observation history $\mathcal{Y}(t) = \{y(s), s \leq t\}$. The optimal, in the Bayesian sense, predictor of $y(t+1)$ takes the form [6]

$$\widehat{y}(t+1|t) = \sum_{n=1}^{N} \mu_n(t) \widehat{y}_n(t+1|t) \tag{7}$$

where

$$\widehat{y}_n(t+1|t) = \boldsymbol{\varphi}_n^{\mathrm{T}}(t+1)\widehat{\boldsymbol{\alpha}}_n(t)$$

denotes predictor based on the AR model of order n, and $\mu_n(t)$ is the posterior probability of n given $\mathcal{Y}(t)$. Since posterior probabilities are nonnegative and sum up to 1

$$\mu_n(t) \geq 0, n \in \mathcal{N}, \quad \sum_{n=1}^{N} \mu_n(t) = 1 \tag{8}$$

the obtained solution is a convex combination of predictors obtained for different hypothetical values of n. Note that the formula (7) can be rewritten in the following equivalent form

$$\widehat{y}(t+1|t) = \boldsymbol{\varphi}_N^{\mathrm{T}}(t+1)\bar{\boldsymbol{\alpha}}_N(t) \tag{9}$$

where $\bar{\boldsymbol{\alpha}}_N(t)$ denotes the vector of averaged parameters

$$\bar{\boldsymbol{\alpha}}_N(t) = \sum_{n=1}^{N} \mu_n(t) \widehat{\boldsymbol{\alpha}}_n^u(t) \tag{10}$$

and $\widehat{\boldsymbol{\alpha}}_n^u(t) = [\widehat{\boldsymbol{\alpha}}_n^{\mathrm{T}}(t), \mathbf{0}_{N-n}^{\mathrm{T}}]^{\mathrm{T}}$ denotes the vector of parameters corresponding to the AR model of order n, extended with $N - n$ zeros. In a similar way, one can obtain the Bayesian estimate of the driving noise variance

$$\bar{\rho}_N(t) = \sum_{n=1}^{N} \mu_n(t) \widehat{\rho}_n(t). \tag{11}$$

The estimates $\bar{\boldsymbol{\alpha}}_N(t)$ and $\bar{\rho}_N(t)$ correspond to the "averaged" signal model (of order N) – introduced by Akaike [4]. As shown in [4], under uniform, i.e., noninformative prior distribution of n, the posterior probabilities of n, referred to by Akaike as model likelihoods, can be obtained from

$$\mu_n(t) \propto \exp\left\{-\frac{1}{2}\mathrm{AIC}_n(t)\right\} \tag{12}$$

where

$$\mathrm{AIC}_n(t) = t \log \widehat{\rho}_n(t) + 2n \tag{13}$$

denotes the Akaike's information statistic [7].

In [8] the concept of model averaging was extended to WLS estimators. For a fixed value of k, parameters of the averaged model are given in the form analogous to (10)-(11)

$$\begin{aligned}\bar{\boldsymbol{\alpha}}_{N|k}(t) &= [\bar{a}_{1,N|k}(t),\ldots,\bar{a}_{N,N|k}(t)]^{\mathrm{T}} \\ &= \sum_{n=1}^{N}\mu_{n|k}(t)\widehat{\boldsymbol{\alpha}}^{a}_{n|k}(t) \\ \bar{\rho}_{N|k}(t) &= \sum_{n=1}^{N}\mu_{n|k}(t)\widehat{\rho}_{n|k}(t)\end{aligned} \tag{14}$$

where

$$\mu_{n|k}(t) \propto \exp\left\{-\frac{1}{2}\mathrm{AIC}_{n|k}(t)\right\} \tag{15}$$

and

$$\mathrm{AIC}_{n|k}(t) = L_k \log\widehat{\rho}_{n|k}(t) + \frac{2nL_k}{M_k} \tag{16}$$

denotes the generalized AIC statistic.

Our interest in the model averaging technique stems from the fact that spectral estimates based on (14)

$$\bar{S}_{N|k}(\omega,t) = \frac{\bar{\rho}_{N|k}(t)}{\left|1-\sum_{i=1}^{N}\bar{a}_{i,N|k}(t)e^{-j\omega i}\right|^2} \tag{17}$$

are usually more accurate than the "point" estimates

$$\widehat{S}_{\widehat{n}_k(t)|k}(\omega,t) = \frac{\widehat{\rho}_{\widehat{n}_k(t)|k}(t)}{\left|1-\sum_{i=1}^{\widehat{n}(t)}\widehat{a}_{i,\widehat{n}_k(t)|k}(t)e^{-j\omega i}\right|^2} \tag{18}$$

obtained when the model order n is fixed at its "most likely" value $\widehat{n}_k(t)$ [8]

$$\widehat{n}_k(t) = \arg\min_{n\in\mathcal{N}}\mathrm{AIC}_{n|k}(t) \cong \arg\min_{n\in\mathcal{N}}\mathrm{FPE}_{n|k}(t) \tag{19}$$

Remark: Note that when $n \ll M_k$, it holds that

$$\begin{aligned}&\log\mathrm{FPE}_{n|k}(t) \\ &\quad= \log\widehat{\rho}_{n|k}(t) + \log(1+n/M_k) - \log(1-n/M_k) \\ &\quad\cong \log\widehat{\rho}_{n|k}(t) + \frac{2n}{M_k} = \frac{1}{L_k}\mathrm{AIC}_{n|k}(t)\end{aligned}$$

which means, that selection of the model order based on minimization of the AIC statistic yields approximately the same results as that based on minimizing the FPE statistic.

The spectral estimate (17) corresponds to a particular (fixed) value of k. In the next section we will show how the model averaging technique can be used in the case where, due to signal nonstationarity, the most appropriate value of k is not known and possibly time-dependent. The proposed spectrum estimation formula, which constitutes the main contribution of the paper, has the form

$$\bar{S}_{N|\widehat{k}(t)}(\omega,t) = \frac{\bar{\rho}_{N|\widehat{k}(t)}(t)}{\left|1-\sum_{i=1}^{N}\bar{a}_{i,N|\widehat{k}(t)}(t)e^{-j\omega i}\right|^2} \tag{20}$$

where

$$\widehat{k}(t) = \arg\min_{k\in\mathcal{K}}\mathrm{FPE}^{*}_{N|k}(t) \tag{21}$$

and

$$\mathrm{FPE}^{*}_{N|k}(t) = \frac{1+\frac{1}{M_k}\sum_{m=1}^{N}\sum_{n=1}^{N}\mu_{m|k}(t)\mu_{n|k}(t)\min(m,n)}{1-\frac{\widehat{n}_k(t)}{M_k}}\widehat{\rho}_{\widehat{n}_k(t)|k}(t) \tag{22}$$

denotes the FPE-like statistic.

III. Derivation of the Modified Final Prediction Error Statistic

In our quest for the locally the best value of k we will use the Akaike's concept of the final prediction error. Denote by $\widetilde{\mathcal{Y}}_k(t) = \{\widetilde{y}(t-k-n),\ldots,\widetilde{y}(t+k)\}$ another realization of the analyzed data sequence, independent of the sequence $\mathcal{Y}_k(t) = \{y(t-k-n),\ldots,y(t+k)\}$ used for identification purposes, and let $\widetilde{\boldsymbol{\varphi}}_N(t) = [\widetilde{y}(t-1),\ldots,\widetilde{y}(t-N)]^{\mathrm{T}}$. Final prediction error for the model (14) is defined as

$$\delta_k(t) = \mathrm{E}\left\{[\widetilde{y}(t) - \widetilde{\boldsymbol{\varphi}}^{\mathrm{T}}_N(t)\bar{\boldsymbol{\alpha}}_{N|k}(t)]^2\right\} \tag{23}$$

where expectation is carried out with respect to to $\widetilde{\mathcal{Y}}_k(t)$ and $\mathcal{Y}_k(t)$, i.e., it is the mean squared prediction error observed when the model is verified using an independent data set.

Our approximation of $\delta_k(t)$ will take advantage of some well-known properties of the AIC/FPE based order selection. It is known that AIC shows some tendency to overestimate the true model order n_0 (under stationary conditions the probability of selecting the order $n > n_0$ does not tend to zero when the number of observations tends to infinity) but it is efficient in eliminating underestimated models (under stationary conditions the probability of selecting $n < n_0$ quickly decays to zero for growing sample size) [7]. This means that in the case considered one can assume that

$$(\mathbf{A}_1) \qquad \mu_{n|k}(t) \cong 0, \quad \forall n < n_0.$$

The second assumption concerns local signal stationarity

$(\mathbf{A}_2)$ The signal $\{y(t)\}$ can be regarded as stationary in the interval $[t-k-N, t+k]$ with "true" parameters $\boldsymbol{\alpha}_{n_0}$ and ρ_{n_0}.

Denote by $\boldsymbol{\alpha}^{a}_{n_0} = [\boldsymbol{\alpha}^{\mathrm{T}}_{n_0}, \mathbf{0}^{\mathrm{T}}_{N-n_0}]^{\mathrm{T}}$ the vector $\boldsymbol{\alpha}_{n_0}(t)$ extended with $N-n_0$ zeros to length N, and define $\widehat{\boldsymbol{\alpha}}^{a}_{n|k}(t)$ in an analogous way. Then the following result holds true

Lemma 1: For $n_0 \le n \le N$ and $n_0 \le m \le N$ it holds that

$$\begin{aligned}\mathrm{E}[\widehat{\boldsymbol{\alpha}}^{a}_{n|k}(t)] &\cong \boldsymbol{\alpha}^{a}_{n_0} \\ \mathrm{E}\{\Delta\widehat{\boldsymbol{\alpha}}^{a}_{n|k}(t)[\Delta\widehat{\boldsymbol{\alpha}}^{a}_{m|k}(t)]^{\mathrm{T}}\} &\cong \frac{\rho_{n_0}}{M_k}\mathbf{P}_{\min(m,n)}\end{aligned} \tag{24}$$

where

$$\Delta\widehat{\boldsymbol{\alpha}}^{a}_{n|k}(t) = \widehat{\boldsymbol{\alpha}}^{a}_{n|k}(t) - \boldsymbol{\alpha}^{a}_{n_0}$$

denotes estimation error and

$$\begin{aligned}\mathbf{P}_n &= \begin{bmatrix}\mathbf{R}_n^{-1} & \mathbf{0}_{n\times(N-n)} \\ \mathbf{0}_{(N-n)\times n} & \mathbf{0}_{(N-n)\times(N-n)}\end{bmatrix} \\ \mathbf{R}_n &= \mathrm{E}[\boldsymbol{\varphi}_n(t)\boldsymbol{\varphi}^{\mathrm{T}}_n(t)]\,.\end{aligned}$$

Proof: See Appendix I.

The next lemma is a straightforward consequence of the former one.

Lemma 2: The averaged parameter estimator (14) is unbiased with covariance matrix of estimation errors given by

$$\mathrm{E}\{\Delta\bar{\boldsymbol{\alpha}}^a_{n|k}(t)[\Delta\bar{\boldsymbol{\alpha}}^a_{m|k}(t)]^{\mathrm{T}}\} \cong \frac{\rho_{n_0}}{M_k}\sum_{m=1}^{N}\sum_{n=1}^{N}\mu_{m|k}(t)\mu_{n|k}(t)\mathbf{P}_{\min(m,n)} \cdot \qquad (25)$$

Proof: The proof is elementary. Since $\mu_{n|k}(t) \cong 0$ for $n < n_0$, and each estimator $\widehat{\boldsymbol{\alpha}}^a_{n|k}(t)$ is unbiased for $n_0 \le n \le N$, so is their convex combination $\bar{\boldsymbol{\alpha}}^a_{n|k}(t)$. Substituting (24) into (14) one gets, after some elementary manipulations, (25).

We are now ready to state the following proposition regarding predictive properties of the averaged estimate.

Proposition 1: Final prediction error of the averaged estimate (14) can be expressed in the form

$$\delta_k(t) \cong \left[1 + \frac{1}{M_k}\sum_{m=1}^{N}\sum_{n=1}^{N}\mu_{m|k}(t)\mu_{n|k}(t)\min(m,n)\right]\rho_{n_0} \cdot \qquad (26)$$

Proof: See Appendix II.

It is well known that for $n \ge n_0$ it holds that [5], [3],

$$\mathrm{E}[\widehat{\rho}_{n|k}(t)] \cong \left(1 - \frac{n}{M_k}\right)\rho_{n_0} \cdot$$

i.e., $\widehat{\rho}_{n|k}(t)$ is a biased estimator of the true variance ρ_{n_0}. Replacing ρ_{n_0} in (26) with the debiased variant of $\widehat{\rho}_{\widehat{n}_k(t)|k}(t)$ yields the estimate which can be immediately recognized as our modified prediction error statistic (22).

IV. Computer simulations

As shown in [3], the WLS estimator equipped with the window $w_k(i)$ yields approximately the same results as the Yule-Walker (YW) estimator equipped with data taper $v_k(i) = \sqrt{w_k(i)}$. The YW estimates can be obtained by solving the set of YW equations, after replacing the true autocorrelation coefficients r_n with their estimates

$$\widehat{r}_{n|k}(t) = \frac{1}{L_k}\sum_{i=n-k}^{k} y_k(t+i|t)y_k(t+i-n|t) \qquad (27)$$
$$n = 0, \ldots, N$$

where $y_k(t+i|t) = v_k(i)y(t+i), i = -k, \ldots, k$ is the tapered data sequence.

In our simulation experiments we used YW estimators in lieu of WLS estimators for two important reasons. First, the YW estimates $\widehat{\boldsymbol{\alpha}}_{n|k}(t)$, $\widehat{\rho}_{n|k}(t)$, $n = 1, \ldots, N$ can be evaluated using the well-known order-recursive Levinson-Durbin algorithm, which is computationally attractive and guarantees that the resulting AR models of different orders are at all times stable. Model stability is an important property (which may not hold if the original WLS scheme is used) since it is a prerequisite for well-posed parametric spectrum estimation.

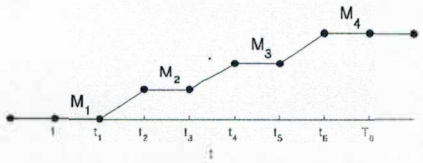

Fig. 1: Model morphing scenario.

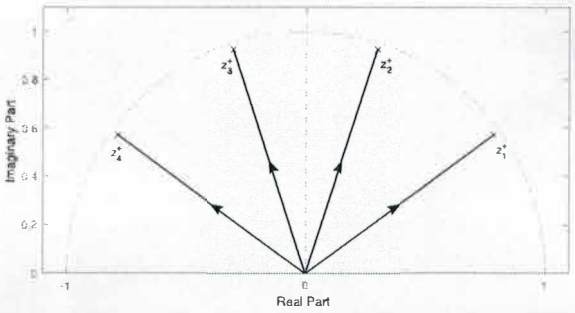

Fig. 2: Trajectories of zeros of the characteristic polynomial.

Second, when the Hann (raised cosine) window $w_k(i) = [1 + \cos(\pi i/(k+1))]/2 = [\cos(\pi i/(2(k+1)))]^2$ is used, i.e., when $v_k(i) = \cos(\pi i/(2(k+1)))$, the estimates (27) are recursively computable [3], [9], which further reduces the computational load.

To show advantages of the proposed method, a nonstationary AR process was generated, defined in terms of 4 time-invariant "anchor" AR models M_i, of orders $2i$, $i = 1, \ldots, 4$. The forming filter $1/A_i(z^{-1}, t)$ corresponding to model M_i had i resonant modes, determined by i pairs of complex-conjugate zeros of its characteristic polynomial $A_i(z^{-1}, t)$: $z_k^{\pm} = 0.995e^{\pm jk\pi/5}$, $k = 1, \ldots, i$.

The analyzed signal $\{y(t), t = 1 \ldots, T_0\}$, $T_0 = 5500$, consisted of segments generated by time-invariant models $\mathrm{M}_1, \ldots, \mathrm{M}_4$ and segments governed by time-varying models obtained by means of applying the "morphing" technique – see Fig. 1. A smooth transition from model M_{i-1} to M_i was realized by relocating progressively the i-th pair of zeros from their initial positions at the origin to terminal positions close to the unit circle. The corresponding zero trajectories are depicted in Fig. 2. The breakpoints shown in Fig. 1 had the following coordinates: $t_1 = 1000$, $t_2 = 1500$, $t_3 = 2500$, $t_4 = 3000$, $t_5 = 4000$ and $t_6 = 4500$.

The parallel estimation scheme was made up of 5 Yule-Walker estimators equipped with cosinusoidal windows of widths $k_1 = 100$, $k_2 = 150$, $k_3 = 225$, $k_4 = 337$ and $k_5 = 500$. Data generation was started 500 instants prior to $t = 1$ and was continued for 500 instants after $T_0 = 5500$. Such an approach allows one to start the estimation process at the instant 1 and end it at the instant T_0 for all bandwidths considered.

The mean Itakura-Saito spectral distortion measure [10]

TABLE I: Comparison of estimation results obtained for 5 fixed-order ($n = 1, \ldots, 20$) Yule-Walker algorithms with different bandwidths k_1, k_2, k_3, k_4, k_5, with the results yielded by 2 order-and-bandwidth-adaptive parallel estimation schemes based on the FPE statistic and the FPE* statistic, respectivety.

n/N	k_1	k_2	k_3	k_4	k_5	FPE	FPE*
1	3,6122	3,5804	3,5575	3,5425	3,5342	3,5606	3,5606
2	2,1265	2,0874	2,0649	2,0588	2,0713	2,0928	2,0928
3	2,0251	1,9815	1,9556	1,9480	1,9608	1,9980	1,9979
4	1,1665	1,1247	1,1023	1,0997	1,1210	1,1446	1,1443
5	1,1769	1,1276	1,0997	1,0943	1,1154	1,1582	1,1568
6	0,5484	0,5105	0,4922	0,4944	0,5260	0,5218	0,5209
7	0,5483	0,5024	0,4789	0,4773	0,5039	0,5196	0,5170
8	0,0760	0,0462	0,0331	0,0332	0,0512	0,0291	0,0278
9	0,0835	0,0515	0,0365	0,0348	0,0493	0,0304	0,0290
10	0,0919	0,0569	0,0401	0,0369	0,0493	0,0314	0,0299
11	0,0998	0,0621	0,0434	0,0387	0,0492	0,0322	0,0307
12	0,1082	0,0676	0,0469	0,0407	0,0499	0,0329	0,0315
13	0,1162	0,0729	0,0503	0,0426	0,0505	0,0335	0,0321
14	0,1245	0,0781	0,0534	0,0442	0,0505	0,0337	0,0324
15	0,1329	0,0835	0,0569	0,0463	0,0513	0,0340	0,0328
16	0,1413	0,0890	0,0605	0,0485	0,0524	0,0342	0,0331
17	0,1499	0,0946	0,0640	0,0507	0,0537	0,0343	0,0333
18	0,1601	0,1009	0,0677	0,0528	0,0549	0,0345	0,0336
19	0,1687	0,1065	0,0712	0,0549	0,0560	0,0347	0,0337
20	0,1778	0,1123	0,0749	0,0572	0,0573	0,0347	0,0339

was used to evaluate spectral estimation results

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

Table 1 presents the mean IS scores, obtained for 100 independent realizations of $\{y(t)\}$ and averaged over time ($t \in [1, T_0]$). The first five columns show results yielded by the Yule-Walker estimators corresponding to different choices of estimation bandwidth (k) and model order (n). The next two columns present results yielded by the parallel estimation scheme with FPE based and FPE* based joint bandwidth and order selection, for different values of the maximum model order N. The results presented in Table 1 clearly demonstrate the advantage of the FPE* based scheme over the FPE based one. Note also that both adaptive order-and-bandwidth selection schemes provide better results than the best fixed-order fixed-bandwidth ones.

The locally time averaged histogram (obtained for 100 independent realizations of $\{y(t)\}$) of the results of bandwidth selection is depicted in Fig. 3. Each time bin incorporates 250 consecutive time instants. Note that during transition phases, when AR parameters change, the smaller bandwidths are selected more frequently than the larger ones. On the other hand, during time-invariant phases (constant AR parameters) the larger bandwidths are preferred.

Fig. 4 shows the locally time and ensemble averaged histogram of the likelihood coefficients $\mu_{n|\widehat{k}(t)}(t)$, evaluated according to (15) and (21). Note good order adaptivity of the proposed order scheduling method.

V. Conclusions

We considered the problem of local autoregressive modeling of nonstationary random processes. The proposed solution combines the "soft" Bayesian-like model averaging with "hard" selection of optimal estimation bandwidth, carried out

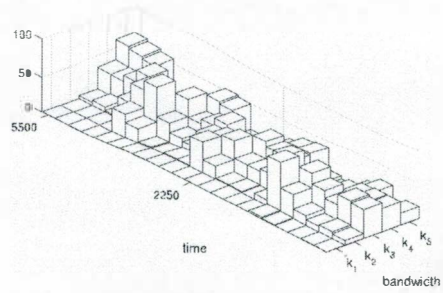

Fig. 3: Locally time averaged histogram of the results of bandwidth selection.

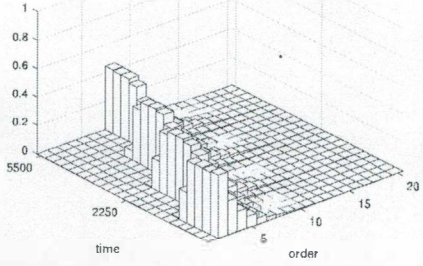

Fig. 4: Locally time averaged histogram of the likelihood coefficients.

using the newly developed extension of the Akaike's final prediction error criterion. Simulation experiments show improved performance of the new adaptive estimation method.

References

[1] S.M. Kay, *Modern Spectral Estimation: Theory and Applications*, Englewoods Cliffs NJ: Prentice-Hall, 1999.

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

[3] M. Niedźwiecki, M. Ciołek, and Y. Kajikawa, "On adaptive covariance and spectrum estimation of locally stationary multivariate processes," *Automatica*, vol. 82, pp. 1-12, 2017.

[4] H. Akaike, "On entropy maximization principle," in Applications of Statistics, P.R. Krishnaiah, Ed., Amsterdam. The Netherlands: North-Holland, pp. 27-41, 1977.

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

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

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

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

[9] M. Ciołek and M. Niedźwiecki, "Detection of impulsive disturbances in archive audio signals," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2017, pp. 671-675.

[10] R.M. Gray, A. Buzo, A.H. Gray, Jr., and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Trans. Acoust. Speech Sign. Process.*, vol. 28, pp. 367–376, 1980.

[11] R.L. Taylor, "Stochastic convergence of weighted sums of random elements in linear spaces," *Lecture Notes in Math.*, vol. 672, 1978.

APPENDIX I
PROOF OF LEMMA 1

We will first show that $\widehat{\boldsymbol{\alpha}}^{a}_{n|k}(t)$ is an unbiased estimate of $\boldsymbol{\alpha}^{a,n}_{n_0}$. Recall (from e.g. [5]) that the WLS estimator (3) admits the following closed form solution

$$\widehat{\boldsymbol{\alpha}}_{n|k}(t) = \mathbf{Q}^{-1}_{n|k}(t)\mathbf{q}_{n|k}(t) \,, \tag{28}$$

where

$$\mathbf{Q}_{n|k}(t) = \sum_{l=-k}^{k} w_k(l)\boldsymbol{\varphi}_n(t+l)\boldsymbol{\varphi}^{\mathrm{T}}_n(t+l)$$

$$\mathbf{q}_{n|k}(t) = \sum_{l=-k}^{k} w_k(l)\boldsymbol{\varphi}_n(t+l)y(t+l) \,.$$

Also note that, for $n \geq n_0$, it holds that

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}_n(t)\boldsymbol{\alpha}^{a,n}_{n_0} + \varepsilon_{n_0}(t) \,, \tag{29}$$

where

$$\boldsymbol{\alpha}^{a,n}_{n_0} = [\boldsymbol{\alpha}^{\mathrm{T}}_{n_0} \; \mathbf{0}^{\mathrm{T}}_{n-n_0}]^{\mathrm{T}} = \mathbf{X}_{n_0\to n}\boldsymbol{\alpha}_{n_0}$$

and

$$\mathbf{X}_{n_0\to n} = [\mathbf{I}_{n_0} \; \mathbf{0}_{n_0\times(n-n_0)}]^{\mathrm{T}}$$

denotes the expansion matrix.

Substituting (29) into (28) leads to

$$\widehat{\boldsymbol{\alpha}}_{n|k}(t) = \boldsymbol{\alpha}^{a,n}_{n_0} + \mathbf{Q}^{-1}_{n|k}(t)\mathbf{p}_{n|k}(t) \tag{30}$$

where

$$\mathbf{p}_{n|k}(t) = \sum_{l=-k}^{k} w_k(l)\boldsymbol{\varphi}_n(t+l)\varepsilon_{n_0}(t+l) \,.$$

Using the generalized law of large numbers for weighted sums of random variables [11], one arrives at the following approximation

$$\mathbf{Q}^{-1}_{n|k}(t) \cong \mathbf{V}^{-1}_{n|k}$$

where

$$\mathbf{V}_{n|k} = \mathrm{E}[\mathbf{Q}_{n|k}(t)] = \sum_{l=-k}^{k} w_k(l)\mathrm{E}[\boldsymbol{\varphi}_n(t+l)\boldsymbol{\varphi}^{\mathrm{T}}_n(t+l)]$$

$$= \mathbf{R}_n \sum_{l=-k}^{k} w_k(l)$$

leading to

$$\Delta\widehat{\boldsymbol{\alpha}}_{n|k}(t) \cong \mathbf{V}^{-1}_{n|k}\mathbf{p}_{n|k}(t)$$

and $\mathrm{E}[\Delta\widehat{\boldsymbol{\alpha}}_{n|k}(t)] \cong \mathbf{V}^{-1}_{n|k}\mathrm{E}[\mathbf{p}_{n|k}(t)] = \mathbf{0}_n$, i.e. $\mathrm{E}[\widehat{\boldsymbol{\alpha}}_{n|k}(t)] \cong \boldsymbol{\alpha}^{a,n}_{n_0}$. Taking into account the fact that $\boldsymbol{\alpha}^{a}_{n_0} = \mathbf{X}_{n\to N}\boldsymbol{\alpha}^{a,n}_{n_0}$ and $\widehat{\boldsymbol{\alpha}}^{\bullet}_{n|k}(t) = \mathbf{X}_{n\to N}\widehat{\boldsymbol{\alpha}}_{n|k}(t)$, one immediately completes the first part of the proof.

To evaluate crosscorrelation of estimation errors assume, without loss of generality, that $m \leq n$. Note that

$$\mathrm{E}[\Delta\widehat{\boldsymbol{\alpha}}_{n|k}(t)\Delta\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}_{m|k}(t)] \cong \mathrm{E}[\mathbf{V}^{-1}_{n|k}\mathbf{p}_{n|k}(t)\mathbf{p}^{\mathrm{T}}_{m|k}(t)\mathbf{V}^{-1}_{m|k}]$$

and

$$\mathrm{E}[\mathbf{p}_{n|k}(t)\mathbf{p}^{\mathrm{T}}_{m|k}(t)] \cong \mathbf{R}_{n,m}\rho_{n_0}\sum_{l=-k}^{k} w^2_k(l)$$

where $\mathbf{R}_{n,m} = \mathrm{E}[\boldsymbol{\varphi}_n(t)\boldsymbol{\varphi}^{\mathrm{T}}_m(t)]$. Since for $m \leq n$ it holds that

$$\boldsymbol{\varphi}_m(t) = [\mathbf{I}_m \; \mathbf{0}_{n-m}]\boldsymbol{\varphi}_n(t) = \mathbf{X}^{\mathrm{T}}_{m\to n}\boldsymbol{\varphi}_n(t) \,.$$

one can express $\mathbf{R}_{n,m}$ in terms of $\mathbf{R}_n$

$$\mathbf{R}_{n,m} = \mathbf{R}_n\mathbf{X}_{m\to n}.$$

Combining all earlier results, one arrives at

$$\mathrm{E}[\Delta\widehat{\boldsymbol{\alpha}}_{n|k}(t)\Delta\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}_{m|k}(t)]$$

$$\cong \rho_{n_0}\frac{\sum_{l=-k}^{k} w^2_k(l)}{\left[\sum_{l=-k}^{k} w_k(l)\right]^2}\mathbf{R}^{-1}_n\mathbf{R}_n\mathbf{X}_{m\to n}\mathbf{R}^{-1}_m$$

$$= \frac{\rho_{n_0}}{M_k}\mathbf{X}_{m\to n}\mathbf{R}^{-1}_m \,.$$

Employing the fact that $\widehat{\boldsymbol{\alpha}}^{a}_{n|k}(t) = \mathbf{X}_{n\to N}\widehat{\boldsymbol{\alpha}}_{n|k}(t)$ and $m \leq n$, one obtains

$$\mathrm{E}\{\Delta\widehat{\boldsymbol{\alpha}}^{\bullet}_{n|k}(t)[\Delta\widehat{\boldsymbol{\alpha}}^{a}_{m|k}(t)]^{\mathrm{T}}\}$$

$$\cong \mathbf{X}_{n\to N}\mathrm{E}[\Delta\widehat{\boldsymbol{\alpha}}_{n|k}(t)\Delta\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}_{m|k}(t)]\mathbf{X}^{\mathrm{T}}_{m\to N} = \frac{\rho_{n_0}}{M_k}\mathbf{P}_m \,.$$

Using similar arguments for $n \leq m$, one gets

$$\mathrm{E}\{\Delta\widehat{\boldsymbol{\alpha}}^{a}_{n|k}(t)[\Delta\widehat{\boldsymbol{\alpha}}^{a}_{m|k}(t)]^{\mathrm{T}}\} \cong \frac{\rho_{n_0}}{M_k}\mathbf{P}_n \,.$$

which finally leads to (24). □

APPENDIX II
PROOF OF PROPOSITION 1

Let

$$\widetilde{e}(t) = \widetilde{y}(t) - \widetilde{\boldsymbol{\varphi}}^{\mathrm{T}}_N(t)\bar{\boldsymbol{\alpha}}_{N|k}(t) \,.$$

Using $\widetilde{y}(t) = \widetilde{\boldsymbol{\varphi}}^{\mathrm{T}}_N(t)\boldsymbol{\alpha}^{\bullet}_{n_0} + \widetilde{\varepsilon}_{n_0}(t)$ and $\bar{\boldsymbol{\alpha}}_{N|k}(t) = \boldsymbol{\alpha}^{a}_{n_0} + \Delta\bar{\boldsymbol{\alpha}}_{N|k}(t)$, one arrives at

$$\widetilde{e}(t) = \widetilde{\varepsilon}_{n_0}(t) - \widetilde{\boldsymbol{\varphi}}^{\mathrm{T}}_N(t)\Delta\bar{\boldsymbol{\alpha}}_{N|k}(t) \,.$$

Since $\Delta\bar{\boldsymbol{\alpha}}_{N|k}(t)$ is, by definition, independent of $\widetilde{\varepsilon}_{n_0}(t)$ and $\widetilde{\boldsymbol{\varphi}}_N(t)$, one obtains

$$\delta_k(t) = \rho_{n_0} + \mathrm{E}[\widetilde{\boldsymbol{\varphi}}^{\mathrm{T}}_N(t)\bar{\mathbf{C}}_k\widetilde{\boldsymbol{\varphi}}_N(t)] \,, \tag{31}$$

where

$$\bar{\mathbf{C}}_k = \mathrm{E}[\Delta\bar{\boldsymbol{\alpha}}_{N|k}(t)\Delta\bar{\boldsymbol{\alpha}}^{\mathrm{T}}_{N|k}(t)] \,.$$

According to Lemma 2 it holds that

$$\mathrm{E}[\widetilde{\boldsymbol{\varphi}}^{\mathrm{T}}_N(t)\bar{\mathbf{C}}_k\widetilde{\boldsymbol{\varphi}}_N(t)] \cong \frac{\rho_{n_0}}{M_k}\sum_{m=1}^{N}\sum_{n=1}^{N}\mu_{m|k}(t)\mu_{n|k}(t)$$

$$\times \mathrm{tr}\left\{\mathbf{P}_{\min(m,n)}\mathrm{E}[\widetilde{\boldsymbol{\varphi}}_N(t)\widetilde{\boldsymbol{\varphi}}^{\mathrm{T}}_N(t)]\right\} \,. \tag{32}$$

Note that $\mathrm{E}[\widetilde{\boldsymbol{\varphi}}_N(t)\widetilde{\boldsymbol{\varphi}}^{\mathrm{T}}_N(t)] = \mathbf{R}_N$ and

$$\mathrm{tr}\left[\mathbf{P}_{\min(m,n)}\mathbf{R}_N\right]$$

$$= \mathrm{tr}\left\{\begin{bmatrix}\mathbf{R}^{-1}_{\min(m,n)} & \mathbf{O}\\ \mathbf{O} & \mathbf{O}\end{bmatrix}\begin{bmatrix}\mathbf{R}_{\min(m,n)} & *\\ * & *\end{bmatrix}\right\}$$

$$= \mathrm{tr}\left\{\begin{bmatrix}\mathbf{I}_{\min(m,n)} & *\\ \mathbf{O} & \mathbf{O}\end{bmatrix}\right\} = \min(m,n) \,.$$

where the symbol $*$ is used to denote matrices that have no effect on the final result. Combining this formula with (31) and (32) leads to (26) and completes the proof. □