On Adaptive Spectrum Estimation of Multivariate Autoregressive Locally Stationary Processes

Michał Meller, Maciej Niedźwiecki and Damian Chojnacki
Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Computer Science,
Department of Automatic Control, ul. Narutowicza 11/12, 80-233 Gdańsk, POLAND
e-mails: michal.meller@eti.pg.edu.pl, maciekn@eti.pg.edu.pl. damianchojnacki10@wp.pl

***Abstract*—Autoregressive modeling is a widespread parametric spectrum estimation method. It is well known that, in the case of stationary processes with unknown order, its accuracy can be improved by averaging models of different complexity using suitably chosen weights. The paper proposes an extension of this technique to the case of multivariate locally stationary processes. The proposed solution is based on local autoregressive modeling, and combines model averaging with estimation bandwidth adaptation. Results of simulations demonstrate that the application of the proposed decision rules allows one to outperform the standard approach, which does not include the bandwidth adaptation.**

***Index Terms*—spectral estimation, multivariate autoregressive process, model averaging, final prediction error**

I. Introduction

Spectrum estimation plays a pivotal role in signal processing, because of its theoretical significance and importance to countless applications. All existing spectrum estimation methods can be assigned to one of the two basic classes. Nonparametric estimators, predominantly different variants of periodogram, usually rely on mild assumptions and therefore offer robust, but only mediocre, accuracy. Model-based, or parametric, methods are generally regarded superior to nonparametric ones in terms of accuracy, albeit at the cost of greater complexity and sensitivity to model mismatch.

Among parametric methods, the autoregressive (AR) modeling stands out due to its strong theoretical background, versatility, accuracy, modest computational cost and wide range of successful applications, of which many involve processing of nonstationary signals. Recent advances in the statistical theory of locally stationary processes confirm this long standing observation [1] with rigorously derived analytical results.

One of the fundamental problems associated with AR analysis of nonstationary time series is that of selecting the model order and the estimation bandwidth (related to the local ana-lysis window width). Ideally, the model should offer enough capacity to accommodate all spectral peaks and notches while avoiding any superfluous parameters. The estimation bandwidth, on the other hand, should be related to the speed at which the changes of process parameters take place. Both problems become particularly important in the nonstationary case, because in many applications the optimal order/bandwidth choices are time-dependent.

In this paper we propose a novel, partially collaborative solution to the problem outlined above. A collaborative scheme can be based on Bayesian-like model averaging using generalized Akaike's model likelihoods as weights. Unfortunately, such a solution solves only of the problem of selecting the model order. This is caused by the fact that, due to limitations of the Akaike's framework, model averaging can be carried out only across models sharing *the same* estimation bandwidth. In this paper, we propose a new final prediction error (FPE) like criterion that allows one to combine model averaging with bandwidth adaptation. When compared with the standard solution, which employs constant bandwidth/order settings, the proposed mixed collaborative-competitive estimator offers superior performance.

The paper is organized as follows: Section 2 introduces basic concepts behind modeling of nonstationary time series using autoregressive models. Section 3 outlines the proposed estimator. Section 4 presents results of statistical simulations. Finally, Section 5 concludes.

II. Multivariate autoregressive processes

A. Stationary multivariate autoregressive model

Consider m-variate discrete-time stationary vector autoregressive (VAR) process of order n

$$\boldsymbol{y}(t) = \sum_{i=1}^{n} \boldsymbol{A}_{i,n}\boldsymbol{y}(t-i) + \boldsymbol{\varepsilon}_n(t) \ , \tag{1}$$

where $t = \ldots, \ -1, \ 0, \ 1, \ \ldots$ denotes dimensionless discrete time, $\boldsymbol{y}(t) = [y_1(t) \ y_2(t) \ \ldots y_m(t)]^{\mathrm{T}}$,

$$\boldsymbol{A}_{i,n} = \begin{bmatrix} a_{11}^{i,n} & \ldots & a_{1m}^{i,n} \\ \vdots & & \vdots \\ a_{m1}^{i,n} & \ldots & a_{mm}^{i,n} \end{bmatrix}, \quad i = 1, \ 2, \ \ldots, \ n \tag{2}$$

are $m \times m$ matrices of autoregressive coefficients and $\{\boldsymbol{\varepsilon}_n(t)\}$ is m-variate zero mean white noise sequence with covariance matrix $\boldsymbol{\rho}_n$.

Let

$$\begin{aligned} \boldsymbol{\theta}_n &= \mathrm{vec}\{[\boldsymbol{A}_{1,n} \ | \boldsymbol{A}_{2,n} \ | \ \ldots \ | \ \boldsymbol{A}_{n,n}]^{\mathrm{T}}\} \\ &= \big[a_{11}^{1,n} \ \ldots \ a_{1m}^{1,n} \ \ldots \ a_{11}^{n,n} \ \ldots \ a_{1m}^{n,n} \ \ldots \\ &\quad a_{m1}^{1,n} \ \ldots \ a_{mm}^{1,n} \ \ldots \ a_{m1}^{n,n} \ \ldots \ a_{mm}^{n,n}\big]^{\mathrm{T}} \end{aligned} \tag{3}$$

978-9-0827-9703-9/19/$31.00

denote the m^2n-dimensional aggregate vector of process autoregressive coefficients. Furthermore, let

$$\begin{aligned}\boldsymbol{\varphi}_n(t) &= \left[\boldsymbol{y}^{\mathrm{T}}(t-1)\ \boldsymbol{y}^{\mathrm{T}}(t-2)\ \ldots\ \boldsymbol{y}^{\mathrm{T}}(t-n)\right]^{\mathrm{T}} \\ \boldsymbol{\Psi}_n(t) &= \boldsymbol{I}_m \otimes \boldsymbol{\varphi}_n(t)\ ,\end{aligned} \tag{4}$$

where $\otimes$ denotes the Kronecker product and $\boldsymbol{I}_m$ denotes the $m \times m$ identity matrix. Using this notation, the recursion (1) can be rewritten in the following form

$$\boldsymbol{y}(t) = \boldsymbol{\Psi}_n^{\mathrm{T}}(t)\boldsymbol{\theta}_n + \boldsymbol{\varepsilon}_n(t)\ , \tag{5}$$

which will be used later extensively.

Let

$$\boldsymbol{\mathcal{A}}_n(z, \boldsymbol{\theta}_n) = \boldsymbol{I}_m - \sum_{i=1}^{n} \boldsymbol{A}_{i,n} z^{-i}\ , \tag{6}$$

where z is a complex variable. A necessary and sufficient condition that must hold for (1) to describe an asymptotically stationary random process, is the stability of the autoregressive model. This occurs if and only if all roots of the characteristic polynomial $\det[\boldsymbol{\mathcal{A}}_n(z, \boldsymbol{\theta}_n)]$ are inside the unit circle in the $\mathcal{Z}$-plane. In such a case, the spectral density of the VAR process can be expressed in the form

$$\boldsymbol{S}_n(\omega) = \boldsymbol{\mathcal{A}}_n^{-1}(e^{j\omega}, \boldsymbol{\theta}_n)\boldsymbol{\rho}_n\boldsymbol{\mathcal{A}}_n^{-\mathrm{T}}(e^{-j\omega}, \boldsymbol{\theta}_n)\ , \tag{7}$$

where $\omega \in [-\pi, \pi)$ denotes the normalized angular frequency,

$$\boldsymbol{\mathcal{A}}_n(e^{j\omega}, \boldsymbol{\theta}_n) = \boldsymbol{\mathcal{A}}_n(z, \boldsymbol{\theta}_n)\big|_{z=e^{j\omega}}$$

and $\boldsymbol{\mathcal{A}}_n^{-\mathrm{T}}(e^{-j\omega}, \boldsymbol{\theta}_n) = [\boldsymbol{\mathcal{A}}_n^{-1}(e^{-j\omega}, \boldsymbol{\theta}_n)]^{\mathrm{T}}$.

B. Locally stationary multivariate autoregressive process

Consider a nonstationary m-variate process governed by the VAR model with time-varying coefficients

$$\boldsymbol{y}(t) = \sum_{i=1}^{n} \boldsymbol{A}_{i,n}(t)\boldsymbol{y}(t-i) + \boldsymbol{\varepsilon}_n(t)\ , \quad \mathrm{E}[\boldsymbol{\varepsilon}_n(t)\boldsymbol{\varepsilon}_n^{\mathrm{T}}(t)] = \boldsymbol{\rho}_n(t)\ . \tag{8}$$

Under regularity conditions specified in [1], the process (8) is locally stationary and its instantaneous spectral density

$$\boldsymbol{S}_n(\omega, t) = \boldsymbol{\mathcal{A}}_n^{-1}[e^{j\omega}, \boldsymbol{\theta}_n(t)]\boldsymbol{\rho}_n(t)\boldsymbol{\mathcal{A}}_n^{-\mathrm{T}}[e^{-j\omega}, \boldsymbol{\theta}_n(t)]\ , \tag{9}$$

where

$$\boldsymbol{\mathcal{A}}_n[z, \boldsymbol{\theta}_n(t)] = \boldsymbol{I}_m - \sum_{i=1}^{n} \boldsymbol{A}_{i,n}(t) z^{-i}\ , \tag{10}$$

is a well defined quantity in the rescaled time domain. In this approach, the rescaled time, which spans a fixed-length interval, is sampled uniformly on a grid that becomes finer as the number of observations increases. Under such setting, the instantaneous spectrum (9) can be interpreted as a spectrum of a stationary process $\{\boldsymbol{y}_0(t)\}$ "tangent" to $\{\boldsymbol{y}(t)\}$ at a point of interest.

Without getting into mathematical details, we note that in order to assure that the local stationarity conditions are satisfied by the time-varying VAR model, it is sufficient to assume that: 1. The parameters of the model have bounded variation. 2. The model is uniformly stable, i.e., all roots of the polynomial $\det\{\boldsymbol{\mathcal{A}}_n[z, \boldsymbol{\theta}_n(t)]\}$ remain strictly inside the unit circle at all time instants t [1].

C. Basic local estimation technique

Suppose that a prerecorded data sequence consisting of T_0 observations, $\mathcal{Y} = \{\boldsymbol{y}(1),\ \boldsymbol{y}(2),\ \ldots,\ \boldsymbol{y}(T_0)\}$, is available. Local estimates of the time-varying VAR model parameters can be obtained using the two-sided weighted least squares (WLS) method.

The two-sided WLS method can be summarized as follows. Denote by $h(x) : \mathcal{R} \to [0, 1]$ a nonnegative "prototype" weighting function with the support $[-1, 1]$. Typically, $h(x)$ is chosen to be a symmetric function with maximum at $x = 0$, $h(0) = 1$, whose values decay smoothly towards 0 as $x \to \pm 1$. These conditions are satisfied by e.g. the Hann (raised-cosine) window function

$$h(x) = \begin{cases} [1 + \cos(\pi x)]/2 & \text{for } x \in [-1, 1] \\ 0 & \text{elsewhere} \end{cases} \tag{11}$$

which is one of the standard choices in the nonparametric spectral analysis due to the fact that it offers a good variance-bias tradeoff.

Denote by $k > 0$, $k \in \mathcal{Z}$ a particular bandwidth setting. The scaled weighting function $h_k(x)$ takes the form

$$h_k(x) = h\left(\frac{x}{k}\right)\ .$$

The WLS estimate $\hat{\boldsymbol{\theta}}_{n|k}(t)$ of the parameter vector $\boldsymbol{\theta}_n(t)$ is defined as

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{n|k}(t) &= \operatorname*{argmin}_{\boldsymbol{\theta}_n} \sum_{\tau=1}^{T_0} w_{t|k}(\tau) \left\| \boldsymbol{y}(\tau) - \sum_{i=1}^{n} \boldsymbol{A}_{i,n}\boldsymbol{y}(\tau - i) \right\|^2 \\ &= \operatorname*{argmin}_{\boldsymbol{\theta}_n} \sum_{\tau=1}^{T_0} w_{t|k}(\tau) \left\| \boldsymbol{y}(\tau) - \boldsymbol{\Psi}_n^{\mathrm{T}}(\tau)\boldsymbol{\theta}_n \right\|^2\ ,\end{aligned} \tag{12}$$

where $\|\boldsymbol{v}\|^2 = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}$ denotes the squared Euclidean norm of the vector $\boldsymbol{v}$ and

$$w_{t|k}(\tau) = h_k(t - \tau)\ . \tag{13}$$

The corresponding estimate $\hat{\boldsymbol{\rho}}_{n|k}(t)$ of the driving noise covariance matrix $\boldsymbol{\rho}_n(t)$ can be obtained from

$$\begin{aligned}\hat{\boldsymbol{\rho}}_{n|k}(t) = \frac{1}{L_k(t)} \sum_{\tau=1}^{T_0} w_{t|\pi}(\tau) &\left[\boldsymbol{y}(\tau) - \boldsymbol{\Psi}_n^{\mathrm{T}}(\tau)\hat{\boldsymbol{\theta}}_{n|k}(t)\right] \times \\ &\left[\boldsymbol{y}(\tau) - \boldsymbol{\Psi}_n^{\mathrm{T}}(\tau)\hat{\boldsymbol{\theta}}_{n|k}(t)\right]^{\mathrm{T}}\ ,\end{aligned} \tag{14}$$

where

$$L_k(t) = \sum_{\tau=1}^{T_0} w_{t|k}(\tau) \tag{15}$$

denotes the so-called effective window width.

Based on (9) and (12), the estimate of the instantaneous power spectral density of the process $\{\boldsymbol{y}(t)\}$ reads

$$\hat{\boldsymbol{S}}_{n|k}(\omega, t) = \boldsymbol{\mathcal{A}}^{-1}[e^{j\omega}, \hat{\boldsymbol{\theta}}_{n|k}(t)]\hat{\boldsymbol{\rho}}_{n|k}(t)\boldsymbol{\mathcal{A}}^{-\mathrm{T}}[e^{-j\omega}, \hat{\boldsymbol{\theta}}_{n|k}(t)]\ . \tag{16}$$

The main advantage of using the two-sided (noncausal) WLS estimator, rather than its more frequently employed

single-sided (causal) counterparts, such as the recursive least squares estimator with exponential forgetting, is the reduction of the bias caused by the phenomenon known as the estimation delay – the output trajectory of a causal estimator can be approximately regarded as a delayed version of the true parameter trajectory [2]. This effect is usually negligible when the two-sided technique is used.

III. Mixed collaborative-competitive approach

The basic estimation technique will not be successful if it is not supported by a suitable mechanism for choosing the model order n and the estimation bandwidth parameter k. The former should be chosen in accordance with the local spectral richness of the analyzed signal, while the latter – in accordance with the rate of the parameter variation.

When the optimal values of n and k are not known, one can simultaneously run several WLS algorithms with different order ($\mathcal{N} = \{1, 2, \ldots, N\}$) and bandwidth ($\mathcal{K} = \{k_1, k_2, \ldots, k_K\}$) settings and, at each time instant t, select the locally best configuration $\{\hat{n}(t),\ \hat{k}(t)\}$ out of NK choices. However, such "hard" decision rules, regardless of their form and complexity, cannot avoid the issue of the uncertainty embedded in the decision process – even though the estimates $\hat{n}(t)$ and $\hat{k}(t)$ are "most likely" to be the best choices of n and k, there is always a chance that they differ from the actual optimal choice.

The uncertainty factor can be accounted for within the Bayesian framework. In this approach, the estimated quantities are regarded as realizations of random variables with assigned prior probability distributions. This leads to "soft" decision rules, based on averaging.

Consider the problem of one step ahead prediction of a stationary VAR process $\{\boldsymbol{y}(t)\}$ based on the available observation history $\mathcal{Y}(t) = \{\boldsymbol{y}(s),\ s \leq t\}$. The optimal (in the mean square sense) Bayesian predictor takes the form

$$\hat{\boldsymbol{y}}(t+1|t) = \sum_{n=1}^{N} \mu_n(t)\hat{\boldsymbol{y}}_n(t+1|t), \quad \sum_{n=1}^{N} \mu_n(t) = 1\ , \quad (17)$$

where $\hat{\boldsymbol{y}}_n(t+1|t) = \boldsymbol{\Psi}_n^{\mathrm{T}}(t+1)\hat{\boldsymbol{\theta}}_n(t)$ denotes the prediction based on the VAR model of order n, and $\mu_n(t)$ is the posterior probability of n given $\mathcal{Y}(t)$.

To show that (17) indeed corresponds to model averaging, consider augmenting (each) model of order n with $N-n$ zero autoregressive coefficients

$$\hat{\boldsymbol{A}}_{p,n}(t) = \boldsymbol{0}_{m\times m}, \quad \text{for } p > n\ .$$

Observe that, in the vectorized form,

$$\hat{\boldsymbol{\theta}}_n^a(t) = \mathrm{vec}\{[\hat{\boldsymbol{A}}_{1,n}(t)\ |\hat{\boldsymbol{A}}_{2,n}(t)\ |\ \ldots\ |\ \hat{\boldsymbol{A}}_{N,n}(t)]^{\mathrm{T}}\} \quad (18)$$

this operation can be expressed as

$$\hat{\boldsymbol{\theta}}_n^a(t) = \boldsymbol{X}_{n\to N}\hat{\boldsymbol{\theta}}_n(t)\ ,$$

where

$$\boldsymbol{X}_{n\to N} = \boldsymbol{I}_m \otimes \begin{bmatrix} \boldsymbol{I}_{nm} \\ \boldsymbol{0}_{(N-n)m\times nm} \end{bmatrix}\ . \quad (19)$$

denotes the expansion matrix.

Using (18), one can express all one-step-ahead predictors in the unified form

$$\hat{\boldsymbol{y}}_n(t+1|t) = \boldsymbol{\Psi}_N^{\mathrm{T}}(t+1)\hat{\boldsymbol{\theta}}_n^a(t)\ ,$$

which leads in a straightforward way to the following form of the Bayesian predictor

$$\hat{\boldsymbol{y}}(t+1|t) = \boldsymbol{\Psi}_N^{\mathrm{T}}(t+1)\bar{\boldsymbol{\theta}}_N(t)\ ,$$

where $\bar{\boldsymbol{\theta}}_N(t)$ is a convex combination of $\hat{\boldsymbol{\theta}}_n^a(t)$, $n = 1, 2, \ldots, N$

$$\bar{\boldsymbol{\theta}}_N(t) = \sum_{n=1}^{N} \mu_n(t)\hat{\boldsymbol{\theta}}_n^a(t)\ . \quad (20)$$

One can obtain the Bayesian estimate of the process noise covariance matrix in the same way

$$\bar{\boldsymbol{\rho}}_N(t) = \sum_{n=1}^{N} \mu_n(t)\hat{\boldsymbol{\rho}}_n(t)\ . \quad (21)$$

The estimates (20)-(21) characterize the "averaged" model, a concept introduced by Akaike [3].

Akaike has shown that – under uniform, i.e., noninformative order priors – the posterior probabilities $\mu_n(t)$, $n = 1, 2, \ldots, N$, called model likelihoods in [3], can be evaluated using the formula

$$\mu_n(t) \propto \exp\left[-\frac{1}{2}\mathrm{AIC}_n(t)\right]\ ,$$

where

$$\mathrm{AIC}_n(t) = t\log\det\hat{\boldsymbol{\rho}}_n(t) + 2m^2 n$$

denotes the Akaike's information statistic.

The concept of model averaging was extended to WLS models in [4]. For a fixed value of k, the corresponding model averaging formula takes the form

$$\bar{\boldsymbol{\theta}}_{N|k}(t) = \sum_{n=1}^{N} \mu_{n|k}(t)\hat{\boldsymbol{\theta}}_{n|k}^a(t)$$
$$\bar{\boldsymbol{\rho}}_{N|k}(t) = \sum_{n=1}^{N} \mu_{n|k}(t)\hat{\boldsymbol{\rho}}_{n|k}(t)\ , \quad (22)$$

where

$$\mu_{n|k}(t) \propto \exp\left[-\frac{1}{2}\mathrm{AIC}_{n|k}(t)\right]$$
$$\mathrm{AIC}_{n|k}(t) = L_k(t)\log\det\hat{\boldsymbol{\rho}}_{n|k}(t) + 2m^2 n\frac{L_k(t)}{M_k(t)}\ . \quad (23)$$

The resulting power spectrum estimate

$$\bar{\boldsymbol{S}}_{N|k}(\omega, t) = \boldsymbol{\mathcal{A}}^{-1}[e^{j\omega}, \bar{\boldsymbol{\theta}}_{N|k}(t)]\bar{\boldsymbol{\rho}}_{N|k}(t)\boldsymbol{\mathcal{A}}^{-\mathrm{T}}[e^{-j\omega}, \bar{\boldsymbol{\theta}}_{N|k}(t)] \quad (24)$$

is known to be more accurate than the one resulting from the classical AIC minimization procedure [4]

$$\hat{\boldsymbol{S}}_{\hat{n}(t)|k}(\omega, t) =$$
$$\boldsymbol{\mathcal{A}}^{-1}[e^{j\omega}, \hat{\boldsymbol{\theta}}_{\hat{n}(t)|k}(t)]\hat{\boldsymbol{\rho}}_{\hat{n}(t)|k}(t)\boldsymbol{\mathcal{A}}^{-\mathrm{T}}[e^{-j\omega}, \hat{\boldsymbol{\theta}}_{\hat{n}(t)|k}(t)]\ , \quad (25)$$

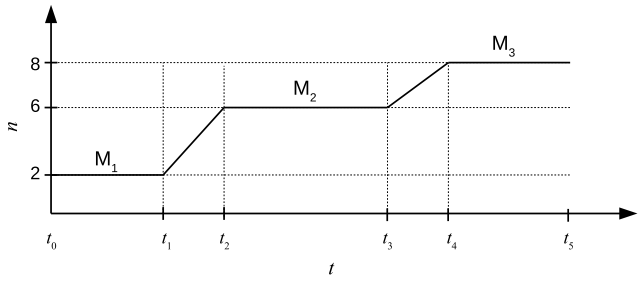

Figure 1. Model morphing scenario adopted for simulation.

where

$$\hat{n}(t) = \arg\min_{n\in\mathcal{N}} \mathrm{AIC}_{n|k}(t) \ .$$

Note that (24) was obtained for a fixed value of k. Unfortunately, it is not possible to extend the procedure outlined above to joint averaging of models obtained for different order *and* bandwidth settings (at least not within the information criterion framework) – the model likelihoods $\mu_{n|k}(t)$ can be compared *only* among models obtained for the same value of k .

To cope with this difficulty, we propose a mixed collaborative-competitive approach, based on the following spectral estimation formula

$$\bar{\boldsymbol{S}}_{N|\hat{k}(t)}(\omega,t) = \boldsymbol{\mathcal{A}}^{-1}[e^{j\omega},\bar{\boldsymbol{\theta}}_{N|\hat{k}(t)}(t)]\bar{\boldsymbol{\rho}}_{N|\hat{k}(t)}(t)\boldsymbol{\mathcal{A}}^{-\mathrm{T}}[e^{-j\omega},\bar{\boldsymbol{\theta}}_{N|\hat{k}(t)}(t)] \ , \quad (26)$$

where $\hat{k}(t)$ is chosen so as to minimize the following FPE-like statistic, whose full derivation is presented in [5]

$$\hat{k}(t) = \arg\min_{k\in\mathcal{K}} \overline{\mathrm{FPE}}_k(t)$$
$$\overline{\mathrm{FPE}}_k(t) = \mathrm{tr}\left[\hat{\boldsymbol{\rho}}_{n_k(t)|k}(t)\right] \times \left[\frac{1+\frac{m}{M_k(t)}\sum_{n=1}^{N}\sum_{n'=1}^{N}\mu_{n|k}(t)\mu_{n'|k}(t)\min(n,n')}{1-\frac{mn_k(t)}{M_k(t)}}\right] \quad (27)$$

and

$$n_k(t) = \arg\max_{n=1,2,\ldots,N} \mu_{n|k}(t) \ .$$

IV. Results of statistical simulations

To evaluate the proposed scheme, statistical simulation experiments were performed. We generated, using the model morphing technique, 100 realizations of a nonstationary process with a known power spectrum. Model morphing was realized based on three time-invariant autoregressive "anchor" models M_1, M_2 and M_3, with orders 2, 6 and 8, respectively, obtained by means of identification of three different fragments of a stereo audio recording.

The anchor models, estimated recursively using the Lee-Morf-Friedlander least squares lattice filter [6], were parametrized in terms of the matrices of the so-called normalized reflection coefficients $\boldsymbol{Q}_{i,n}, i = 1,\ldots,n$, rather than in terms of the matrices of the autoregressive coefficients $\boldsymbol{A}_{i,n}, i = 1,\ldots,n$, employed throughout the paper. The two representations of an autoregressive model are known to be equivalent [6]. However, out of these two, only reflection coefficients allow one to implement the stability-preserving model morphing.

Suppose that one intends to gradually pass from the model M_1, valid at the instant t_1 and characterized by reflection coefficients $\boldsymbol{Q}^1_{i,n}, i = 1,\ldots,n$, to the model M_2, valid at the instant t_2 and characterized by reflection coefficients $\boldsymbol{Q}^2_{i,n}, i = 1,\ldots,n$. This goal can be achieved by computing the intermediate reflection coefficients according to the following convex combination rule

$$\boldsymbol{Q}_{i,n}(t) = \alpha(t)\boldsymbol{Q}^1_{i,n} + [1-\alpha(t)]\boldsymbol{Q}^2_{i,n}$$
$$i = 1,\ldots,n, \quad t\in[t_1,t_2]$$

where $\alpha(t) : [t_1,t_2] \to [0,1]$, $\alpha(t_1) = 1$, $\alpha(t_2) = 0$, is a smoothly decreasing function of t. When the orders of the morphed models are not identical, the nonexistent reflection coefficients in the lower-order model are set to zero. The resultant time-varying model is at all times stable.

The morphing scenario adopted for the purpose of this study is depicted in Fig. 1.The following values were adopted: $t_0 = 0$, $t_1 = 1500$, $t_2 = 2000$, $t_3 = 4000$, $t_4 = 4500$ and $t_5 = 6000$.

We compared 60 baseline fixed order fixed bandwidth Hann-windowed WLS algorithms, corresponding to twenty choices of the model order ($n \in \{1,2,\ldots,20\}$) and three choices of the estimation bandwidth ($k \in \{300,500,800\}$), with three partially adaptive solutions, which employ only the model averaging technique, and the proposed mixed collaborative-competitive scheme, based on the $\overline{\mathrm{FPE}}$ statistic. For each parially adaptive / fully adaptive case, twenty subvariants, differing in the maximum model order $N \in \{1,2,\ldots,20\}$, were implemented.

The performance of each spectrum estimator was evaluated using the relative entropy rate (RER) [7], which is a multivariate extension of the Itakura-Saito spectral distortion measure. The results, obtained by combining the time averaging ($t \in [1000,5000]$) and the ensemble averaging (100 realizations), are summarized in Table I.

Observe that, in all cases the adaptive schemes yield a considerably better performance than the baseline WLS algorithms with fixed order and fixed bandwidth settings. For model order equal to 8, which is a minimum model order that can accommodate all spectral content of the process, model averaging allows one to improve RER scores by approximately 10%. When the maximum model order is overestimated the fixed order/fixed bandwidth algorithm experiences a substantial degradation, while the partially adaptive solutions exhibits a negligible penalty. The proposed approach allows one to improve on these results by another 15% for an overall accumulated improvement of 25% over the best nonadaptive variant. Furthermore, inspection of the partial results obtained for the individual realizations of the process shows that the improvement is very consistent – better performance was observed in all cases.

	Fixed order, fixed bandwidth			Adaptive order, fixed bandwidth			Fully adaptive
n/N	$k = 300$	$k = 500$	$k = 800$	$k = 300$	$k = 500$	$k = 800$	$\overline{FPE}$
1	1.6241	1.6238	1.6359	1.6241	1.6238	1.6359	1.6237
2	0.5939	0.5936	0.6050	0.5939	0.5936	0.6050	0.5921
3	0.4712	0.4695	0.4803	0.4707	0.4692	0.4801	0.4680
4	0.2970	0.2939	0.3057	0.2957	0.2932	0.3055	0.2916
5	0.0792	0.0822	0.1061	0.0772	0.0812	0.1056	0.0723
6	0.0521	0.0537	0.0775	0.0495	0.0524	0.0770	0.0432
7	0.0561	0.0555	0.0784	0.0497	0.0521	0.0767	0.0430
8	0.0556	0.0531	0.0756	0.0454	0.0476	0.0728	**0.0384**
9	0.0614	0.0562	0.0771	0.0458	0.0479	0.0730	0.0389
10	0.0673	0.0592	0.0787	0.0460	0.0480	0.0731	0.0390
11	0.0733	0.0624	0.0804	0.0461	0.0481	0.0733	0.0392
12	0.0793	0.0656	0.0822	0.0462	0.0482	0.0733	0.0392
13	0.0855	0.0688	0.0839	0.0463	0.0482	0.0734	0.0392
14	0.0919	0.0720	0.0857	0.0463	0.0483	0.0734	0.0392
15	0.0982	0.0752	0.0874	0.0463	0.0483	0.0734	0.0392
16	0.1046	0.0784	0.0891	0.0463	0.0483	0.0734	0.0392
17	0.1112	0.0817	0.0909	0.0463	0.0483	0.0735	0.0392
18	0.1179	0.0851	0.0927	0.0463	0.0483	0.0735	0.0392
19	0.1248	0.0885	0.0945	0.0463	0.0483	0.0735	0.0392
20	0.1318	0.0919	0.0964	0.0463	0.0483	0.0735	0.0392

Table I

COMPARISON OF AVERAGE RER SCORES OBTAINED USING: STANDARD WLS ESTIMATOR WITH DIFFERENT ORDER AND BANDWIDTH SETTINGS, PARTIALLY ADAPTIVE RULE BASED ON MODEL AVERAGING WITH FIXED BANDWIDTH, AND THE PROPOSED COLLABORATIVE-COMPETITIVE MECHANISM BASED ON MODEL AVERAGING AND BANDWIDTH ADAPTATION.

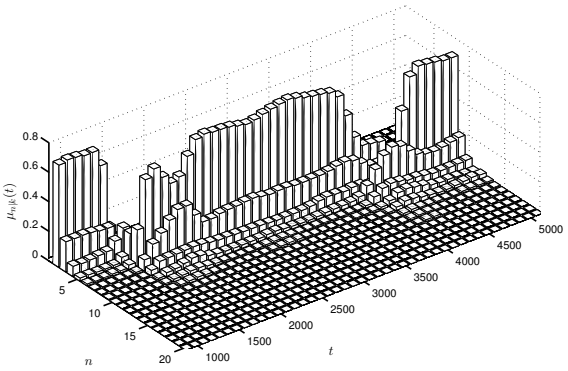

Figure 2. Ensemble and time averaged generalized Akaike likelihoods $\mu_{n|k}(t)$ obtained for autoregressive model orders $n = 1, 2, \ldots, 20$ under bandwidth $k = 500$. The data for $t < 800$ and $t > 5200$ is incomplete and was discarded.

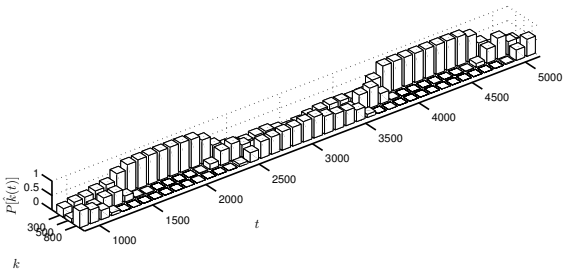

Figure 3. Ensemble and time averaged frequencies of choosing each bandwidth setting. The data for $t < 800$ and $t > 5200$ is incomplete and was discarded.

Figures 2 and 3 depict the evolution of the averaged Akaike's likelihoods for each model order ($N = 20$, $k = 500$) and the probabilities (frequencies) of choosing each bandwidth. Each time bin covers 100 consecutive time instants. Observe the good agreement of the results with the desired behavior – during the periods of process stationarity, larger bandwidth settings are selected more frequently, and model order is estimated with high confidence. On the other hand, during the periods of nonstationarity, small bandwidth settings are preferred, and a more uniform distribution of model likelihoods can be observed.

V. CONCLUSIONS

The problem of local autoregressive modeling of multivariate locally stationary random processes was considered. The proposed solution combines the "soft" Bayesian-like model averaging with the "hard" selection of optimal estimation bandwidth, carried out using the newly developed extension of the Akaike's final prediction error criterion. Simulation experiments show the improved performance of the new adaptive spectrum estimation method.

ACKNOWLEDGEMENT

This work was partially supported by the National Science Center under the agreement UMO-2015/17/B/ST7/03772.

REFERENCES

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

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

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

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

[5] M. Meller, M. Niedźwiecki, and D. Chojnacki, "Multivariate autoregressive spectrum estimation of locally stationary processes with model averaging and bandwidth adaptation," Technical report WETI/KSA/2018/1. [Online]. Available: https://pg.edu.pl/documents/1094802/f4261258-c6d1-4c3a-bb0d-70441e523893

[6] B. Friedlander, "Lattice methods for spectral estimation," *Proc. IEEE*, vol. 70, pp. 990–1017, 1982.

[7] A. Ferrante, C. Masiero, and M. Pavon, "Time and spectral domain relative entropy: A new approach to multivariate spectral estimation," *IEEE Trans. Automat. Contr.*, vol. 57, pp. 2561–2575, 2012.