CLASSIFYING AUTOREGRESSIVE MODELS USING DISSIMILARITY MEASURES: A COMPARATIVE STUDY

Clément Magnant [1],[3] *Eric Grivel* [3] *Audrey Giremus* [3] *Laurent Ratton* [2] *Bernard Joseph* [1]

[1] [2] THALES Systèmes Aéroportés S.A.,[1] Pessac, [2] Elancourt, FRANCE
[3]University of Bordeaux - Bordeaux INP - IMS - UMR CNRS 5218, Talence, FRANCE

ABSTRACT

Autoregressive (AR) models are used in various applications, from speech processing to radar signal analysis. In this paper, our purpose is to extract different model subsets from a set of two or more AR models. The approach operates with the following steps: firstly the matrix composed of dissimilarity measures between AR-model pairs are created. This can be based on the symmetric Itakura divergence, the symmetric Itakura-Saito divergence, the log-spectral distance or Jeffrey's divergence (JD), which corresponds to the symmetric version of the Kullback-Leibler divergence. These matrices are then transformed to get the same properties as correlation matrices. Eigenvalue decompositions are performed to get the number of AR-model subsets and the estimations of their cardinals. Finally, K-means are used for classification. A comparative study points out the relevance of the JD-based method. Illustrations with sea radar clutter are also provided.

Index Terms— Autoregressive model, Jeffrey's divergence, Itakura divergence, Itakura-Saito divergence, log-spectral distance, K-means, classification.

1. INTRODUCTION

In many applications, autoregressive (AR) models are very popular. For instance, in radar processing, the properties of the Gaussian clutter can be analyzed by modeling it by an AR process [1]. In biomedical applications [2], AR models can be used to classify signals of patients with a specific pathology from signals recorded with healthy people. In speech processing such as speech analysis and Kalman-filter based enhancement, M sets of p AR parameters $\{a_l^i\}_{l=1,..,p}^{i=1,...,M}$, estimated by different methods, are often compared one another [3]. Most of the time, when $M = 2$, the authors use the Itakura divergence (ID) or its symmetric version (SID). As an alternative, since the i^{th} AR process can be seen as an infinite-impulse-response filtering of a white noise process, it can be defined by the poles $\{p_l^i\}_{l=1,..,p}$ of the corresponding transfer function, namely $H_i(z) = \frac{1}{\prod_{l=1}^{p}(1-p_l^i z^{-1})}$. Comparing two AR models amounts to compare the corresponding poles. Therefore, a visual comparison is usually done by looking at a figure which shows the poles in the unit disc in the z-plane. Another method consists in computing the symmetric version of the Itakura-Saito divergence (SISD) or the log-spectral distance (LSD). It is based on the power spectral densities (PSDs) of both AR processes that can be expressed from the AR parameters and the variances of the driving processes. As the SISD and the LSD are based on a continuous sum in the frequency domain, they are approximated by discrete sums.
Nevertheless, to our knowledge, there is no method that makes it possible to jointly compare more than two AR models at the same time. An idea would be to compare the set of the AR parameters, the set of poles or the corresponding PSDs. However it is not necessary an easy task, especially when the number of the AR processes becomes large.
In [4], we recently proposed to compare various state-space representations (SSRs), which are 1st-order Markov models, in order to select the SSRs that could be used in multiple-model based estimation approaches [5]. It was then applied in the field of maneuvering-target tracking. In our approach, the symmetric version of the Kullback-Leibler divergence (KL), i.e. Jeffrey's divergence (JD), between the joint distributions of the successive state-vector values associated to two SSRs is first recursively computed. Then, at each instant, the matrix composed of the JDs between the SSRs is built. The matrices are then modified to have the same properties as correlation matrices and hence to take advantage of correlation-matrix analysis techniques. Therefore, the eigenvalue decompositions of the resulting matrices allow the number of model subsets to be deduced. Finally, K-means are used for classification [6].
This paper deals with new ways to classify a set of various p^{th}-order AR models into several model subsets. Therefore, we first suggest extending the above approach [4] to p^{th}-order Markov models. Then, by replacing the JD by the SID, the SISD or the LSD, three other methods can be derived. Finally a comparative study between these four methods is carried out. We will see that some of them seem to be *a priori* relevant but their use is not straightforward in practice.
The rest of this paper is organized as follows: section 2 introduces the notations and gives information about the SID,

This work is part of the common research activities between Bordeaux campus and Thales (GIS Albatros).

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

the SISD and the LSD. The recursive expression of the JD applied to AR models is also presented. As it corresponds to a specific case of what we proposed in [7], the main results are recalled. In section 3, the way to classify p^{th}-order AR-models is presented. In section 4, a comparative study is carried out with synthetic data. Sea radar clutter data are also used. Finally, conclusions and perspectives are drawn.
In the following, ^{T} and ^{H} denote the transpose and the hermitian while * is the conjugate. $\mathbf{0}_{i\times j}$ and $\mathbf{1}_{i\times j}$ are matrices of $i \times j$ zeros and ones respectively. $\boldsymbol{I}_i$ is the identity matrix of size $i \times i$. Moreover, $\odot$ is the Hadamard product.

2. PRELIMINARIES

2.1. Notations

Let us consider M p^{th}-order AR models:

$$\begin{aligned} x_k &= -\sum_{l=1}^{p} a_l^i x_{k-l} + u_k^i \quad i = 1, ..., M \\ &= -\boldsymbol{X}_{k-1}^{(p)} (\bar{\theta}^i)^T + u_k^i \quad i = 1, ..., M \end{aligned} \tag{1}$$

where u_k^i is a zero-mean white Gaussian process with variance $\sigma_{u_i}^2$. In addition, $\boldsymbol{X}_{k-1}^{(p)} = [x_{k-1}, \ ... \ , x_{k-p}]$, $\bar{\theta}^i = [a_1^i, \ ... \ , \ a_p^i]$, $\theta^i = [1 \ \bar{\theta}^i]$ and $\Delta\bar{\theta} = \bar{\theta}^2 - \bar{\theta}^1$.

2.2. Definitions of the Itakura divergence (ID), the Itakura-Saito divergence (ISD), their symmetric versions (SID and SISD) and the Log-spectral distance (LSD)

When dealing with $M = 2$ AR models, the ID is defined by:

$$ID_{12} = \ln\left(\frac{\theta^2 R_1^{(p+1)} (\theta^2)^H}{\theta^1 R_1^{(p+1)} (\theta^1)^H}\right) \tag{2}$$

with $R_i^{(p+1)} = \mathrm{E}_{p_i(\boldsymbol{X}_{k-1}^{(p+1)})}\left[(\boldsymbol{X}_{k-1}^{(p+1)})^H \boldsymbol{X}_{k-1}^{(p+1)}\right]$ the autocorrelation matrix of $\boldsymbol{X}_{k-1}^{(p+1)}$, for the i^{th} AR process.
The SID, the symmetric version of the ID, is defined as:

$$SID_{12} = ID_{12} + ID_{21} \geq 0 \tag{3}$$

Concerning the ISD, it can be approximated by a discrete sum based on N angular-frequency values:

$$ISD_{ij} \approx \frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} \left[\frac{P_i(2\pi\frac{n}{N})}{P_j(2\pi\frac{n}{N})} - \ln\frac{P_i(2\pi\frac{n}{N})}{P_j(2\pi\frac{n}{N})} - 1\right] \tag{4}$$

where $P_i(\omega)$ is the PSD of the i^{th} AR model, which can be expressed from $\sigma_{u_i}^2$ and $H_i(z)$. Given the above, the SISD is deduced as follows:

$$\begin{aligned} SISD_{12} &= ISD_{12} + ISD_{21} \\ &\approx \frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} \left[\frac{P_i(2\pi\frac{n}{N})}{P_j(2\pi\frac{n}{N})} + \frac{P_j(2\pi\frac{n}{N})}{P_i(2\pi\frac{n}{N})} - 2\right] \geq 0 \end{aligned} \tag{5}$$

Concerning the LSD, it can be approximated by:

$$LSD_{12} \approx \sqrt{\frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} \left[10 \ln \frac{P_1(2\pi\frac{n}{N})}{P_2(2\pi\frac{n}{N})}\right]^2} \geq 0 \tag{6}$$

2.3. Recursive computation of the Jeffrey's divergence

To deduce the JD between the joint distributions of the successive values of two AR models, denoted $p_1(x_{0:k})$ and $p_2(x_{0:k})$, the KL must be preliminary defined:

$$KL_{12}(k) = \int_{x_{0:k}} p_1(x_{0:k}) \ln \frac{p_1(x_{0:k})}{p_2(x_{0:k})} dx_{0:k} \tag{7}$$

In [7], we showed that the corresponding JD can be recursively computed as follows:

$$\begin{aligned} JD_{21}(k) = JD_{12}(k) &\stackrel{\Delta}{=} KL_{12}(k) + KL_{21}(k) \\ &= JD_{12}(k-1) + A_{12} + B_{12} \end{aligned} \tag{8}$$

where:

$$\begin{aligned} A_{12} &= -1 + \frac{1}{2}\left[\frac{\sigma_{u_2}^2}{\sigma_{u_1}^2} + \frac{\sigma_{u_1}^2}{\sigma_{u_2}^2}\right] = A_{21} \geq 0 \\ B_{12} &= \frac{1}{2}(\mathrm{Tr}\,[R_1^{(p)} \frac{\Delta\bar{\theta}^H \Delta\bar{\theta}}{\sigma_{u_2}^2} \\ &\quad + R_2^{(p)} \frac{\Delta\bar{\theta}^H \Delta\bar{\theta}}{\sigma_{u_1}^2}]) = B_{2,1} \geq 0 \end{aligned} \tag{9}$$

Since $\boldsymbol{X}_{k-2}^{(p)}$ and u_{k-1}^i are uncorrelated, $R_i^{(p)}$ satisfies the following recursive equation:

$$R_i^{(p)} = C(\bar{\theta}^i) R_i^{(p)} C(\bar{\theta}^i)^H + \sigma_{u_i}^2 [1 \ \mathbf{0}_{1\times p-1}]^T [1 \ \mathbf{0}_{1\times p-1}] \tag{10}$$

with $C(\bar{\theta}^i) = \begin{bmatrix} -a_1^i & -a_2^i & \cdots & -a_p^i \\ & \boldsymbol{I}_{p-1} & & \mathbf{0}_{(p-1)\times 1} \end{bmatrix}$.
It should be noted that for any value of p, $R_i^{(p)}$ can be iteratively estimated with (10) by using a fixed-point method.
In addition, it is possible to compare AR models the orders of which are different. Indeed, in this case, the end of the vector that is the smallest among $\bar{\theta}^1$ or $\bar{\theta}^2$ is padded with zeros so that both can have the same size.
Finally, two properties are of interest:
1/ if $\Delta\bar{\theta} = 0$, then $B_{12} = 0$.
2/ if the driving-process variances are equal, i.e. $\sigma_{u_1}^2 = \sigma_{u_2}^2$, then $A_{12} = 0$. In this latter case, one has:

$$JD_{12}(k) = kJD_{12}(1) = kB_{12} \tag{11}$$

2.4. Remarks

Unlike SID_{12}, $SISD_{12}$ and LSD_{12}, JD_{12} depends on time. All are symmetric, always positive and null when the AR models are the same.
In the following, let us address the comparison of several AR models by using the above dissimilarity measures.

3. FOUR APPROACHES FOR AR-MODEL CLASSIFICATION

The four proposed methods operate with the same steps, but differ by the dissimilarity measure that is used.

Step 1: When more than two AR models must be compared, there are several dissimilarity measures to be studied simultaneously. For this reason, the dissimilarity measures, namely $\{SID_{ij}\}_{i=1,..,M}^{j=1,..,M}$, $\{SISD_{ij}\}_{i=1,..,M}^{j=1,..,M}$, $\{LSD_{ij}\}_{i=1,..,M}^{j=1,..,M}$ or $\{JD_{ij}(k)\}_{i=1,..,M}^{j=1,..,M}$ are structured in a matrix form.

For SID, SISD and LSD, this leads to a dissimilarity matrix D of size $M \times M$ and whose main diagonal is zero.

For the specific case of the JD-based method, there is one matrix $D(k)$ at each time instant $k = 1, ..., S$, where S is the sample number:

$$D(k) = \begin{bmatrix} JD_{11}(k) = 0 & ... & JD_{1M}(k) \\ JD_{21}(k) & ... & JD_{2M}(k) \\ ... & ... & ... \\ JD_{M1}(k) & ... & JD_{MM}(k) = 0 \end{bmatrix} \quad (12)$$

If the initial conditions are the same, $D(0) = \mathbf{0}_{M\times M}$. Given (11), $D(k)$ can be recursively computed as follows:

$$D(k) = kD(1) \ \forall\, k \geq 1 \quad (13)$$

Step 2: In order the triangular inequality to be respected, a shortest-path algorithm such as the Dijkstra algorithm must be applied on the dissimilarity matrix, as proposed by [8].

For SID, SISD and LSD, this leads to a new matrix, denoted W, with elements $\{W_{ij}\}_{i=1,..,M}^{j=1,..,M}$.

For the specific case of the JD-based method, Dijkstra algorithm is applied on $D(k)$ to get $W(k)$, the elements of which are $\{W_{ij}(k)\}_{i=1,..,M}^{j=1,..,M}$.

Step 3: To deduce the number of AR-model subsets, we suggest referring to different results dealing with correlation-matrix analysis such as subspace methods, principal component analysis, etc. [9].

For SID, SISD and LSD, the matrix W is transformed into a matrix E that has the same properties as a correlation matrix. Its elements $\{E_{ij}\}_{i=1,..,M}^{j=1,..,M}$ can be defined as follows:

$$E_{ij} = \exp\left(-\lambda W_{ij}\right) \quad (14)$$

where λ is a scale parameter to be set by the practitioner. Note that this setting will be discussed in the simulation part. In this case, if $SID_{ij} = 0$, $SISD_{ij} = 0$ or $LSD_{ij} = 0$, $W_{ij} = 0$. E_{ij} is hence equal to the maximum value that can be obtained, namely one. The higher the dissimilarity measure is, the smaller E_{ij} is. In addition, the transformation (14) emphasizes differences between small dissimilarity-measure values and higher values.

For the specific case of the JD-based method, $W(k)$ is transformed into a matrix $E(k)$ by following the same non-linear transformation as in (14). The choice of the scale parameter λ will be discussed at the end of the next step.

Step 4: Eigenvalue decompositions are computed.

For SID, SISD and LSD, as the number of model subsets should be *a priori* deduced from the number of predominant eigenvalue analysis, a K-means algorithm based classification could be performed.

For the specific case of the JD-based method, the eigenvalues of $E(k)$ are analyzed over time. According to the Gershgorin circle theorem [10] and as $E(k)$ is semi-definite positive, the eigenvalues of $E(k)$ are located in at least one of the intervals $[max(0, 1 - R_i(k)),\ 1 + R_i(k)]$ for $i = 1, ..., M$, where $R_i(k) = \sum_{j=1;j\neq i}^{M} E_{ij}(k)$.

At time $k = 0$, the eigenvalues lie in the interval $[0,\ M]$. More particularly, one eigenvalue of $E(0) = \mathbf{1}_{M\times M}$ is equal to M whereas the others are all equal to zero. Therefore, there is a set of M similar AR models.

When k increases, according to (8)-(9), $JD_{ij}(k+1) \geq JD_{ij}(k) \geq 0$ for $i \neq j$. $E_{ij}(k+1) \leq E_{ij}(k)$ for $i \neq j$ and $R_i(k)$ becomes smaller and smaller. All the eigenvalues hence tend to 1, leading to M subsets of one AR model.

Therefore, the number of predominant eigenvalues, namely those which are greater than one, provide information about the subset number. Then, the K-means algorithm can be used.

Remark about the scale-parameter setting: For the non-linear transformation in the JD-based method, we suggest considering that AR models are "equally dissimilar" at time S so that the eigenvalues are all close to one at this instant. The non-diagonal elements of $E(S)$ must be hence close to zero and under a threshold $\mathcal{T}$. If $W_{min}(S)$ is the minimal value of the elements $\{W_{ij}(S)\}_{i\neq j}$, $\lambda \geq -\frac{\ln(\mathcal{T})}{W_{min}(S)}$.

4. SIMULATION RESULTS

Various sets of AR models have been tested. In this simulation part, a comparative study is carried out between the four proposed methods with a toy example. This helps the reader to see the relevance of the JD-based method and the difficulty to adjust the scale parameter λ when using the three other methods. Then, real radar sea clutter data are used.

4.1. Comparing $M = 5$ AR models

A set of five 2nd-order AR models $\{AR_i\}_{i=1,..,M}$ is studied. Their AR parameters and corresponding poles are given in Table 1. Note that the poles are deliberately chosen so that two subsets of cardinals equal to 3 and 2 could be *a priori* expected.

The values taken by the different dissimilarity measures are not necessarily in the same ranges. When applying the approaches based on SID, SISD and LSD, arbitrarily choosing the scale parameter λ leads to poor results, as depicted

Model	a_1^i	a_2^i	Poles
AR_1	-1.178	0.462	$p_1^1 = 0.68\ exp(i\pi/6), p_1^{1*}$
AR_2	-1.230	0.504	$p_1^2 = 0.71\ exp(i\pi/6), p_1^{2*}$
AR_3	-1.282	0.548	$p_1^3 = 0.74\ exp(i\pi/6), p_1^{3*}$
AR_4	-0.530	0.281	$p_1^4 = 0.53\ exp(2i\pi/3), p_1^{4*}$
AR_5	-0.570	0.325	$p_1^5 = 0.57\ exp(2i\pi/3), p_1^{5*}$

Table 1. AR parameters of the five 2nd-order AR models and their corresponding poles

by Table 2. Indeed, when λ is very small, the components of E are all close or equal to 1, leading to a predominant eigenvalue close to 5 while the others are close to 0. When λ is very high, E tends to the identity matrix, leading to 5 eigenvalues equal to 1. In between, different values for the eigenvalues $\{\beta_i\}_{i=1,..,M=5}$ can be obtained. As an alternative, the selection of λ could take into account the distribution of the components of the dissimilarity matrix D. If D_{ij} is distributed according to the probability density function (pdf) $p_D(d)$, then the pdf of E_{ij} is $p_E(e) = \frac{1}{\lambda e} p_D(\frac{-1}{\lambda e} ln(e))$ if $e > 0$ and zero otherwise. By setting the inverse of the scale parameter, $\frac{1}{\lambda}$, at the maximum value, the mediane value or the standard deviation (std) of the components of W, the pdf of E_{ij} is modified. The maximum value is not a good choice. Indeed, the components of W can be very close or very far one another, the components of the matrix E necessarily lie in the interval $[exp(-1),\ 1]$. This hence leads to poor results when interpreting the eigenvalues of E. In this toy example, the choice of std seems convenient according to Table 2, but when using it in other cases, this setting is not necessarily appropriate because the std alone cannot characterize the multimodal distribution $p_D(d)$.
Depending on the choice made on λ, the conclusions that can be drawn on the eigenvalues $\{\beta_i\}_{i=1,..,M}$ may be erroneous. Therefore, these three methods cannot be applied easily in any case.

	Methods	β_1	β_2	β_3	β_4	β_5
$\frac{1}{\lambda} = 1$	ISD	3.52	1.45	0.02	0.01	0.00
	SISD	3.29	1.69	0.01	0.01	0.00
	LSD	2.24	1.72	0.52	0.28	0.24
$\frac{1}{\lambda} = 0.3$	ISD	2.94	1.98	0.05	0.02	0.01
	SISD	2.93	1.98	0.05	0.02	0.02
	LSD	1.46	1.34	0.92	0.66	0.62
$\frac{1}{\lambda} = std$	ISD	3.07	1.88	0.02	0.02	0.01
	SISD	3.08	1.88	0.02	0.01	0.01
	LSD	2.75	1.76	0.26	0.12	0.11

Table 2. Eigenvalues of the matrix E for different scale parameters

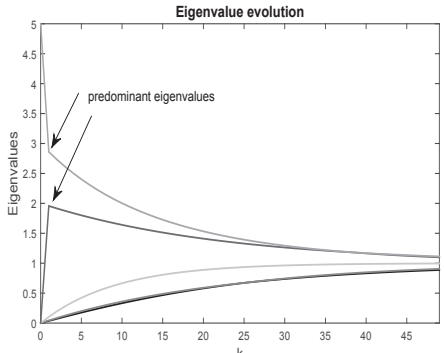

Fig. 1. Evolution over time of the eigenvalues of $E(k)$

Unlike the other approaches, thanks to the time dimension which is added in the JD-based method, two extreme cases are introduced and serve as references to give a meaning to the eigenvalues: one at $k = 0$ corresponding to a set of M models and the other at $k = S$ corresponding to M sets of one model. When using the approach based on the JD, the key feature is hence the time evolutions of the eigenvalues. This is illustrated by Fig. 1 for the specific case of Table 1. Even if some AR models belong to the same subset, they necessarily differ one another over time. Therefore, all eigenvalues converge to one. By modifying λ, the convergence can be more or less fast. In this toy example, there are two predominant eigenvalues that are clearly higher than 1. There are hence two AR-model subsets whose cardinals are equal to 3 and 2. By using the K-means algorithm on $W(k)$, with $K = 2$ subsets, the three following subsets are obtained $\{AR_1, AR_2, AR_3\}$; $\{AR_4, AR_5\}$. This confirms what could be initially expected from the model set design.
According to our tests including the one presented in this paper, the approach based on the JD provides significant results even if its computational cost is high (as there are S eigenvalue decompositions to be done).

4.2. Application with real data in radar signal processing

In this section, the relevance of the JD-based approach is analyzed with real radar data in order to characterize the similarity of real clutter data in terms of correlation properties.
In the framework of maritime surveillance, airborne radars are used to detect targets at various ranges. The radar antenna transmits a coherent burst of pulses with a given period. Then, the received signal can be disturbed by thermal noise as well as ground or sea returns, commonly called the "clutter". These disturbances can be problematic to detect targets, specifically those with low speeds. To adress this problem, approaches such as space time adaptive processing (STAP) [11] require the inversion of the clutter correlation matrix. This latter can be estimated in the maximum likelihood sense by using the "secondary data", i.e. the data corresponding to the area around the range cell of interest. However, there is no

guarantee that these data have *a priori* the same statistical and correlation properties. Statistical tests exist and are usually used to charaterize the clutter homogeneity [12].
In this paper, our purpose is to compare in a "simple" way various sets of clutter data that have been recorded in flight and to check whether they have the same kinds of correlation properties. Each set of sea clutter data is composed of the successive values of a range cell taken at different periods of time corresponding to the burst period. For these data sets, the AR parameters are estimated by using the Yule-Walker equations [9] and the JD based-algorithm is used.

Example 1:
$M = 6$ sea clutter data sets are compared. Among them, we deliberately choose four data sets characterizing only Gaussian clutter and two data sets which are composed of Gaussian clutter disturbed by the antenna secondary lobe effects related to a target detection. The scale parameter is set at $\lambda = 10^{-6}$. The result of our approach is given by Fig. 2.

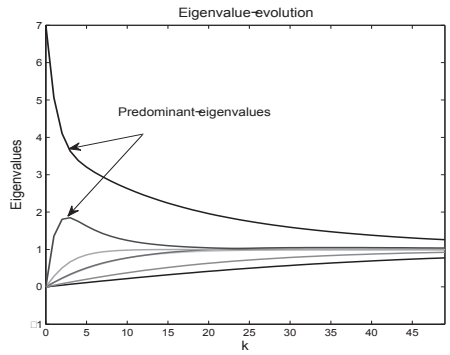

Fig. 2. Evolution over time of the eigenvalues of $E(k)$ for non-homogeneous clutter data

The eigenvalue evolution characterizes two data subsets. Therefore, they do not have the same correlation properties.

Example 2:
$M = 6$ sea clutter data sets are now selected without any *a priori*. In this case, $\lambda = 10^{-3}$. The evolution of the eigenvalues is given by Fig. 3.
Among the eigenvalues, there is only one predominant eigenvalue. Hence, only a single subset of the six data sets can be deduced. In other words, the data sets have the same correlation properties and the set of clutter data can be considered as homogeneous in terms of correlation.

5. CONCLUSIONS

This paper deals with the comparison of a set of AR models. Four methods have been proposed. Three are based on the Itakura divergence, the Itakura-Saito divergence and the log spectral distance. The last one uses Jeffrey's divergence between the joint distributions associated to the successive

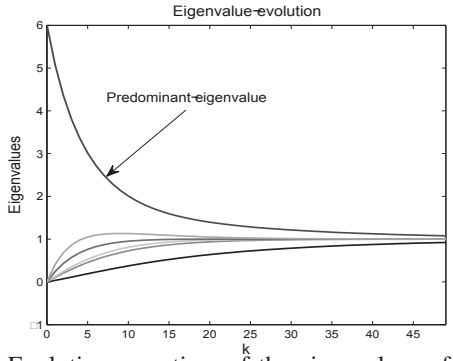

Fig. 3. Evolution over time of the eigenvalues of $E(k)$ for homogeneous clutter data with $\lambda = 10^{-3}$

values of AR models. According to our simulations on synthetic data, the JD-based method provides significant results whereas the others are too sensitive to a scale parameter to be tuned. Illustrations with radar sea clutter data confirm the relevance of the JD-based method. We are currently investigating an alternative approach in which the set of JD matrices is seen as a tensor and a high-order singular value decomposition is used.

6. REFERENCES

[1] Y. Abramovich, N. Spencer, and M. Turley, "Time-varying Autoregressive (TVAR) Models for Multiple Radar Observations," *IEEE Trans. on Signal Processing*, vol. 55, pp. 1298–1311, 2007.

[2] L. Sornmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, Elsevier Academic Press, 2005.

[3] R. Diversi, R. Guidorzi, and U. Soverini, "Identification of Autoregressive Models in the presence of Additive Noise," *International Journal of Adaptive Control and Signal Processing*, vol. 22, pp. 465–481, June 2008.

[4] C. Magnant, E. Grivel, A. Giremus, B. Joseph, and L. Ratton, "Jeffrey's Divergence for State-Space Model Comparison," *Signal Processing*, doi:10.1016/j.sigpro.2015.02.006, 2015.

[5] Y. Bar-Shalom and X. R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, Storrs, 1995.

[6] J. B. MacQueen, "Some Methods for Classification and Analysis of Multivariate Obervations," in *5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297.

[7] C. Magnant, A. Giremus, and E. Grivel, "On Computing Jeffreys Divergence Between Time-Varying Autoregressive Models," *IEEE Signal Processing Letters*, vol. 22, pp. 915–919, 2015.

[8] A. Schutz, L. Bombrun, and Y. Berthoumieu, "Centroid-Based Texture Classification Using the SIRV Representation," in *IEEE International Conf. Image Processing, ICIP 2013*, 2013, pp. 3810–3814.

[9] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Prentice Hall, 1992.

[10] R. S. Varga, *Gershgorin and his Circles*, Springer Science, 2004.

[11] S. Beau and S. Marcos, "Range Dependent Clutter Rejection Using Range-Recursive Space-Time Adaptive Processing (STAP) Algorithms," *Signal Processing*, vol. 90, pp. 57–68, January 2010.

[12] E. Conte, A. De Maio, and C. Galdi, "Statistical Analysis of Real Clutter at Different Range Resolutions," *IEEE Trans. on Aerospace and Electronics Sytems*, pp. 903–918, 2004.