14th IFAC Symposium on System Identification, Newcastle, Australia, 2006

DISTANCES BETWEEN DYNAMICAL MODELS FOR CLUSTERING TIME SERIES

Jeroen Boets* Katrien De Cock* Bart De Moor*

* *K.U.Leuven, Dept. of Electrical Engineering, SCD*
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

Abstract: In this paper we consider the clustering of time series arising from the class of scalar linear stochastic models. The properties and performance of several so-called model-free and model-based distances for these time series are compared on both artificial and real data sets. In particular, the inappropriateness of model-free distances to distinguish between time series of this class is shown, as well as several important differences between the model-based distances themselves.

Keywords: Time-series analysis, ARMA models, SISO, Classification, Distance

1. INTRODUCTION

Time series arise in many important areas, e.g. economics (evolution of share prices), bioinformatics (microarray measurements of gene expression), geology, etc. This paper is concerned with detecting similarities in the *dynamics* of different time series, i.e. the way in which consecutive measurements are related to each other. We apply for this a three-step procedure which has been applied previously in different application areas such as speech processing (Gray *et al.*, 1980) and biomedical signal processing (Gersch, 1981). Firstly, and this is the key step, each time series is represented by a dynamical model estimated from the given data. Secondly, a distance between the models is defined and computed over all models estimated in the first stage. Finally, a clustering or classification is performed based on this distance.

In this paper we mainly focus on the second step, the definition and computation of the distance. A lot of research has been done on defining appropriate distances for dynamical models (Gray *et al.*, 1980; Basseville, 1989; Martin, 2000; De Cock and De Moor, 2002; Georgiou and Lindquist, 2003), also in the domain of model reduction (Antoulas, 1999). However, several challenges remain to be tackled. For instance, while some distances are defined and computable for both single-input single-output (SISO) and multiple-input multiple-output (MIMO) models (Kazakos and Papantoni-Kazakos, 1980), others are only applicable to SISO models, e.g. the ones directly based on the cepstrum (Gray *et al.*, 1980), opening a research challenge for their multivariable extension.

We will, however, only consider SISO (linear) autoregressive moving average (ARMA) processes in this paper. We will compare the properties and performance of a (limited) set of distances for this class of time series. One of the distances we investigate is based on the concept of subspace angles within and between models. This distance will be shown, both theoretically [1] (Section 4.2) and experimentally (Section 5.3), to have a cepstral (i.e. a log-spectral) character. The extension of the concept of subspace angles for MIMO models could therefore result in a very natural extension of a cepstral norm for multivariable processes.

The paper is organized as follows. Section 2 describes the model class we work with. Several distances for time series of this class are defined in Sections 3 and 4. In Section 5 these distances are compared using artificial examples, in Section 6 using real data sets of time series. Finally, Section 7 concludes the paper.

[1] These results are not a contribution of the present paper, but can be found in (De Cock, 2002).

2. CLASS OF TIME SERIES

2.1 Model class

The model class of time series we work with are SISO linear time-invariant stochastic models (ARMA models). We consider the following state space description of such an nth order model:

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) , \\ y(k) = Cx(k) + Du(k) , \end{cases} \quad (1)$$

where the output $\{y(k)\}_{k\in\mathbb{Z}} \in \mathbb{R}$ of the model is the stochastic process that is being modelled, $\{x(k)\}_{k\in\mathbb{Z}} \in \mathbb{R}^n$ is the state process and $\{u(k)\}_{k\in\mathbb{Z}} \in \mathbb{R}$ is a zero-mean white noise process with variance 1 (henceforth, u will also be called the input process). The corresponding transfer function from u to y is

$$H(z) = C(zI_n - A)^{-1}B + D .$$

A stochastic model is fully specified by the foursome $\{A, B, C, D\}$ or the singleton $\{H(z)\}$, and will therefore be denoted by its transfer function $H(z)$ or shortly H. The model is assumed to be strictly stable and minimum phase, meaning that its poles and zeros lie strictly inside the unit circle.

The infinite *observability matrix* of the model is

$$\Gamma = \begin{pmatrix} C^\top & (CA)^\top & (CA^2)^\top & \cdots \end{pmatrix}^\top .$$

The *inverse model*, denoted by $H^{-1}(z)$, is immediately derived from the equations in (1):

$$\begin{cases} x(k+1) = (A - BD^{-1}C)x(k) + BD^{-1}y(k) , \\ u(k) = -D^{-1}Cx(k) + D^{-1}y(k) . \end{cases}$$

Its infinite observability matrix is denoted by Γ_z.

We will compute principal angles between the row spaces of *input and output Hankel matrices*. If N output observations $y(0), y(1), \ldots, y(N-1)$ are given, then the output Hankel matrix is equal to

$$Y = \begin{pmatrix} y(0) & y(1) & \cdots & y(j-1) \\ y(1) & y(2) & \cdots & y(j) \\ \vdots & \vdots & & \vdots \\ y(i-1) & y(i) & \cdots & y(i+j-2) \end{pmatrix} \quad (2)$$

where $i + j - 1 = N$ and i and j are user-defined parameters. $U \in \mathbb{R}^{i\times j}$ is defined in a similar way.

2.2 The cepstrum of an ARMA process

The cepstrum of a stochastic process is defined as the inverse Fourier transform of the logarithm of the spectrum of the process. It can be easily estimated from data using the fast Fourier transform (FFT) algorithm and the periodogram estimate of the spectrum. The cepstrum can also be expressed in terms of the model parameters. For a model with poles $\alpha_1, \ldots, \alpha_n$ and zeros $\beta_1, \ldots, \beta_n$, the cepstrum of the output process y is equal to:

$$c(k) = \begin{cases} \log D^2 & k = 0 , \\ \sum_{i=1}^{n} \frac{\alpha_i^{|k|}}{|k|} - \sum_{i=1}^{n} \frac{\beta_i^{|k|}}{|k|} & k \neq 0 . \end{cases} \quad (3)$$

3. MODEL-FREE DISTANCES BETWEEN ARMA TIME SERIES

In the first class of distances, the Euclidean distance is measured directly between the observed time series or a simple transformation of them. Six types of vector representations will be considered, each denoted by the code in parentheses: the raw data (`eu`), the normalized data ($\texttt{eu}_\texttt{n}$), the first 10 discrete Fourier transform coefficients of the data (`df`) and of the normalized data ($\texttt{df}_\texttt{n}$), the first 10 principal component scores of the data (`pc`) and of the normalized data ($\texttt{pc}_\texttt{n}$). For the normalization, we subtracted from each time series its mean and divided by its standard deviation.

4. MODEL-BASED DISTANCES BETWEEN ARMA TIME SERIES

The second class of distances are measured between dynamical models. The estimation of a dynamical model from a time series is done using the MATLAB System Identification Toolbox, namely the forward-backward least squares approach for autoregressive (AR) modelling and the N4SID subspace identification algorithm (Van Overschee and De Moor, 1996) for ARMA estimation.

4.1 The $\mathbf{H_2}$ distance ($\mathrm{h_2}$)

The $\mathbf{H_2}$ norm of a model $H(z)$ is defined as the root-mean-square (rms) gain of the output signal with respect to the input signal, when a white noise input is applied to the model. This leads to the expression $\|H(z)\|_{\mathrm{h_2}}^2 = \frac{1}{2\pi}\int_0^{2\pi} |H(e^{j\omega})|^2 d\omega$. The corresponding distance is $\mathrm{d_{h_2}}(H^{(1)}, H^{(2)}) = \|H^{(1)} - H^{(2)}\|_{\mathrm{h_2}}$, which is equal to the rms gain of $y^{(1)} - y^{(2)}$ with respect to u when the same white noise input signal u is applied to both models.

4.2 A weighted cepstral distance (`wcep`)

In this section, as in the previous, we define a model norm by applying a white noise signal to the model, but instead of measuring the input-output rms gain, we now follow a *geometric* approach by computing the *principal angles*[2] between subspaces of input and output data. This

[2] The principal angles between the row spaces of the matrices $A \in \mathbb{R}^{p\times m}$ of rank r_a and $B \in \mathbb{R}^{q\times m}$ of rank r_b, with $r_a < r_b$, are denoted by $(\theta_1, \ldots, \theta_{r_a}) = [A \lhd B]$ and are a generalization of the angle between two vectors. The

approach will be elaborated in a first subsection and a connection with the cepstrum of the process (hence the code wcep) will be shown in a second. All statements were proven in (De Cock, 2002).

The angles within and between models. Consider an nth order model $H(z)$, to which a white noise input signal u is applied.

Theorem 1. The largest n principal angles between the row spaces of the input and output Hankel matrices U and Y (see (2)), of the model $H(z)$ are equal to the principal angles between the column spaces of the observability matrices of $H(z)$ and of the inverse model $H^{-1}(z)$, provided the number of rows and columns of U and Y goes to ∞. The other principal angles are equal to 0:

$$[U \sphericalangle Y] = \left[\Gamma^\top \sphericalangle \Gamma_z^\top\right], 0, 0, \ldots$$

The n non-zero angles are called the *subspace angles within the model* $H(z)$.

In a similar way, we can define angles *between* two models $H^{(1)}(z)$ and $H^{(2)}(z)$, of orders $n^{(1)}$ and $n^{(2)}$ respectively, by looking at the principal angles between their output spaces when they are both driven by the same white noise input signal.

Theorem 2. The largest $n^{(1)} + n^{(2)}$ principal angles between the row spaces of the output Hankel matrices of $H^{(1)}(z)$ and $H^{(2)}(z)$ are equal to the principal angles between the column spaces of $\left(\Gamma^{(1)}\ \Gamma_z^{(2)}\right)$ and $\left(\Gamma^{(2)}\ \Gamma_z^{(1)}\right)$, provided the number of rows and columns of $Y^{(1)}$ and $Y^{(2)}$ goes to ∞:

$$\left[Y^{(1)} \sphericalangle Y^{(2)}\right] = \left[\left(\Gamma^{(1)}\ \Gamma_z^{(2)}\right)^\top \sphericalangle \left(\Gamma^{(2)}\ \Gamma_z^{(1)}\right)^\top\right], 0, 0, \ldots$$

The $n^{(1)} + n^{(2)}$ non-zero angles are called the *subspace angles between the models* $H^{(1)}$ *and* $H^{(2)}$.

From both theorems it can be easily shown that the subspace angles between $H^{(1)}$ and $H^{(2)}$ are equal to the subspace angles within $\frac{H^{(1)}}{H^{(2)}}$ or $\frac{H^{(2)}}{H^{(1)}}$. Analogously, the subspace angles within the model $H(z) = \frac{b(z)}{a(z)}$ are equal to the subspace angles between the AR models $\frac{z^n}{a(z)}$ and $\frac{z^n}{b(z)}$. They consequently indicate a certain 'distance' between the pole- and zero-part of the transfer function.

Relation with a weighted cepstral norm and distance. We consider a particular *weighted* cepstral distance, defined in (Martin, 2000):

$$\mathrm{d}^2_{\mathrm{wcep}}(H^{(1)}, H^{(2)}) = \sum_{k=0}^{\infty} k(c^{(1)}(k) - c^{(2)}(k))^2 , \quad (4)$$

with $c^{(1)}$ and $c^{(2)}$ the cepstra of $H^{(1)}$ and $H^{(2)}$. Denoting the subspace angles within the model $H(z)$ by $\psi_1, \ldots, \psi_n$, and the subspace angles between $H^{(1)}(z)$ and $H^{(2)}(z)$ by $\psi_1^{(12)}, \ldots, \psi_{n^{(1)}+n^{(2)}}^{(12)}$,

$$\|H\|^2_{\mathrm{wcep}} = \sum_{k=0}^{\infty} kc(k)^2 = -\log \prod_{i=1}^{n} \cos^2 \psi_i , \quad (5)$$

$$\mathrm{d}^2_{\mathrm{wcep}}(H^{(1)}, H^{(2)}) = -\log \prod_{i=1}^{n^{(1)}+n^{(2)}} \cos^2 \psi_i^{(12)} .$$

So the angles within a model are related to its wcep norm, the angles between two models to their wcep distance. It is now also clear that

$$\mathrm{d}_{\mathrm{wcep}}(H^{(1)}, H^{(2)}) = \left\| \frac{H^{(1)}}{H^{(2)}} \right\|_{\mathrm{wcep}} = \left\| \frac{H^{(2)}}{H^{(1)}} \right\|_{\mathrm{wcep}} . \quad (6)$$

Using (3), we can express (5) as (Martin, 2000):

$$\|H\|^2_{\mathrm{wcep}} = \log \frac{\prod_{i=1}^{n} \prod_{j=1}^{n} |1 - \alpha_i \bar{\beta}_j|^2}{\prod_{i,j=1}^{n} (1 - \alpha_i \bar{\alpha}_j) \prod_{i,j=1}^{n} (1 - \beta_i \bar{\beta}_j)} , \quad (7)$$

where $\bar{a}$ is the complex conjugate of a. Using the equality (6), this formula can also be used to compute the wcep distance between AR or ARMA models as a function of their poles and zeros.

*4.3 The unweighted cepstral distance (*cep*)*

The unweighted cepstral distance $\mathrm{d}_{\mathrm{cep}}$ is defined as the Euclidean distance between the cepstra $c^{(1)}$ and $c^{(2)}$ of $H^{(1)}$ and $H^{(2)}$:

$$\mathrm{d}^2_{\mathrm{cep}}(H^{(1)}, H^{(2)}) = \sum_{k=1}^{L} (c^{(1)}(k) - c^{(2)}(k))^2 .$$

Using Parseval's relation this distance can be seen to be equivalent to the L_2 distance between the log-spectra of the (normalized) models $H^{(1)}$ and $H^{(2)}$ for infinite L. As in (Kalpakis *et al.*, 2001), $L = 10$ was chosen for the experiments in this paper. Note that in the case of the weighted cepstral distance (4) there is no need to cut off the cepstral sequences, since the distance can be computed exactly through (7).

*4.4 Distance between spectra (*sp*)*

The L_2 distance between the spectra $\Phi^{(1)}$ and $\Phi^{(2)}$ of two models $H^{(1)}$ and $H^{(2)}$ is defined as

$$\mathrm{d}^2_{\mathrm{sp}}(H^{(1)}, H^{(2)}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\Phi^{(1)}(\omega) - \Phi^{(2)}(\omega))^2 d\omega .$$

In our experiments we calculated the Euclidean distance between the vectors consisting of the autocovariances $R(-10), \ldots, R(0), \ldots, R(10)$, which are expressible in terms of the model parameters.

4.5 The H_∞ *distance (*h_∞*)*

The H_∞ norm of a model $H(z)$ is defined as the maximum energy of the output signal over all possible input signals with energy one applied to the model. It is equal to $\|H(z)\|_{\mathrm{h}_\infty} =$

number of angles equal to zero, is equal to the dimension of the intersection between the row spaces. More information and a formal definition can be found in (De Cock, 2002).

$\sup_{\omega\in[0,2\pi]} |H(e^{j\omega})|$. The distance is equal to $d_{h_\infty}(H^{(1)}, H^{(2)}) = \|H^{(1)} - H^{(2)}\|_{h_\infty}$.

The Hankel and Hilbert-Schmidt-Hankel (HSH) norm were also included in this paper's experiments. We will however not further report on these norms since their results were very similar to those with the H_∞ norm. For strictly proper systems these three norms are in fact equivalent (see e.g. (Schelfhout, 1996)).

*4.6 A distance between ARMA coefficients (*coef*)*

As a last distance we consider the Euclidean distance between the vectors of AR or ARMA coefficients in $H(z) = \frac{b(z)}{a(z)}$, as proposed in (Bagnall and Janacek, 2004). The vectors of polynomial coefficients are first scaled so that the highest order coefficient in z is equal to one.

5. COMPARISON OF THE DISTANCES USING ARTIFICIAL DATA

In this section we will compare the properties of the defined distances on simulated data. All experiments were performed using all distances mentioned in Sections 3 and 4, but we will for each experiment only report on the ones that allow the most important conclusions. We also computed the distances `cep`, `wcep` and `sp` in a non-parametric way using the FFT algorithm and the periodogram estimate of the spectrum. However, for the same amount of data (and for various choices of cut-off length), they always performed worse than their model-based variants (using (3), (7) and parametric expressions for the autocovariances, respectively). So we will not further report on these non-parametric estimates.

5.1 Clustering methodology

The cluster algorithm we use in this paper is a *partitioning around medoids* algorithm (van der Laan *et al.*, 2003) which aims at finding a clustering with maximum *average silhouette* over all objects. The silhouette of an object measures how well the object belongs to its cluster and always lies between -1 (very badly) and $+1$ (very well). The input of the algorithm is a matrix of pairwise distances and a number of clusters.

All cluster experiments in this paper involve time series of which the *true* cluster partition is known. Therefore, we evaluate the experimentally obtained cluster result by measuring its similarity with the true partition, using the *adjusted Rand index* R_{adj} (Hubert and Arabie, 1985). The maximum (and best) value of R_{adj} is 1, denoting equality of both partitions, while its expected value in the case of random clusters is equal to 0. It is a meaningful quantity irrespective of the number of clusters and the cluster sizes of both partitions.

5.2 Model-free versus model-based distances

As a first experiment we used the same set-up as in (Xiong and Yeung, 2004). We sampled 15 models from each of the following 5 model ranges: $\mathbf{M}^{(1)}$ consisting of AR models of order 1 (AR(1) models) with their only pole uniformly distributed in the range 0.30 ± 0.01, $\mathbf{M}^{(2)}$ to $\mathbf{M}^{(5)}$ with their pole distributed in the ranges 0.40 ± 0.01 up to 0.55 ± 0.01 with steps of 0.05. From each of the 75 sampled models, a time series of length 250 was generated. We then performed a clustering with the 15 time series from $\mathbf{M}^{(1)}$ and the 15 time series from one of the other model ranges $\mathbf{M}^{(i)}$, thus trying to detect two clusters. The R_{adj} index of this cluster result was computed, using the known labels of these 30 time series. This was done for $i = 2$ to 5, resulting in 4 Rand indices. The correct model structure was always assumed to be known for the identification.

The whole procedure (sampling 75 models, generating 75 time series, performing 4 clusterings) was then repeated 9 times, thus resulting in 10 R_{adj} indices for each of the 4 model ranges $\mathbf{M}^{(2)}$ to $\mathbf{M}^{(5)}$. The minimum, average and maximum value of these 10 indices can be found in Table 1. In this table the results of `wcep` are representative of the results of all model-based distances (`cep`, `wcep`, `sp`, h_2, h_∞ and `coef`), while `mfree` contains in each row the best value of all model-free distances (see Section 3). There was not much difference among the model-based distances. As expected, their Rand index rises as the poles of the two sets of models are farther apart. The model-free distances, however, all appeared unable to deal with the stochastic nature of the time series. Therefore, we conclude that they are clearly unsuited for detecting similarity between these kinds of objects.

5.3 Spectral versus cepstral distances

In this section we compare the spectral distances `sp`, h_2 and h_∞ with the cepstral distances `cep` and `wcep`. One of the main differences between spectral and cepstral distances is the logarithm which appears in the cepstrum. This explains why we investigated the following two characteristics.

Distinguishing between zeros. Consider the experiment of Section 5.2 with the 5 model ranges, and the results in Table 1. When replacing for the

Table 1. Clustering results for time series generated from two AR(1) model ranges, one of which always has its pole in the range 0.30±0.01, while the other has its pole in the range given in the table.

Range of AR pole	Adjusted Rand index (min/avg/max)	
	mbased (wcep)	mfree
0.40±0.01	(0.04/0.32/0.63)	(-0.03/0.01/0.08)
0.45±0.01	(0.19/0.54/0.74)	(-0.03/0.01/0.14)
0.50±0.01	(0.43/0.82/1.00)	(-0.03/0.04/0.34)
0.55±0.01	(0.63/0.92/1.00)	(-0.03/0.01/0.13)

Table 2. Clustering results for time series (now of length 500) generated from two ARMA(1) model ranges, both always having their pole at −0.70, and one of which always has a zero in the range 0.30±0.01, while the other has a zero in the range given in the table.

Range of ARMA zero	Adjusted Rand index (min/avg/max)		
	wcep	h_2	h_∞
0.40±0.01	(0.01/0.29/0.63)	(0.00/0.02/0.08)	(-0.04/0.01/0.13)
0.45±0.01	(0.43/0.68/0.87)	(0.00/0.07/0.34)	(-0.02/0.00/0.08)
0.50±0.01	(0.74/0.90/1.00)	(0.01/0.16/0.52)	(-0.03/0.01/0.05)
0.55±0.01	(0.74/0.97/1.00)	(0.01/0.33/1.00)	(-0.02/0.00/0.04)

models in each of the 5 model ranges their (non-random) zero at 0 by a zero at −0.70, thereby obtaining ARMA(1) sampled models, and redoing the experiment with these adjusted $\mathbf{M}^{(1)}$ to $\mathbf{M}^{(5)}$, all model-based distances performed again roughly similarly well.

However, when we redid this experiment with the roles of pole and zero in the ARMA(1) model ranges switched, thereby obtaining ARMA(1) sampled models with a (non-random) pole at −0.70 and a random zero in one of the 5 given ranges, we obtained the results in Table 2, where wcep is now representative of cep, wcep and coef; h_∞ of both sp and h_∞. Clearly, h_2 and especially h_∞ and sp have much more difficulty discriminating between these sets of models. When redoing this last experiment with the (non-random) pole at −0.70 replaced by a pole at 0, all model-based distances performed again approximately similarly well, so that the reason for deterioration of h_2, h_∞ and sp lies in the combination of slightly different ranges of zeros and a pole which is strong enough to conceal these differences.

We conclude that the spectral distances, especially h_∞ and sp, are less sensitive to differences in zeros than the cepstral, when an additional and sufficiently strong pole is present in the spectrum.

Distinguishing between secondary peaks. Now consider two AR models: $H^{(1)}$ of order 2 having a pole pair at radius 0.96 and angles ±150 degrees, and $H^{(2)}$ of order 4 having the same pole pair and an additional pair at radius 0.80 and angles ±109 degrees (Fig. 1). We did a 10 times repeated clustering experiment, each time with only two sets of 15 time series, each set of time series generated from just one model, $H^{(1)}$ or $H^{(2)}$. The goal being to distinguish between both sets, the (min/avg/max) adjusted Rand index was (0.01/0.30/0.63) for sp, (0.19/0.40/0.74) for h_∞, (0.42/0.68/0.87) for h_2 and (1.00/1.00/1.00) for cep and wcep.

We can conclude that the spectral distances, especially h_∞ and sp, are less sensitive to differences in small peaks than the cepstral, when an additional and larger peak is present in the spectrum.

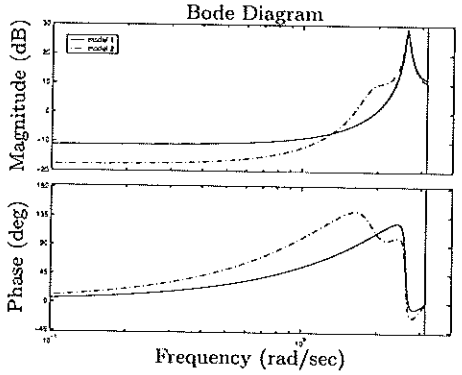

Fig. 1. Bode diagram of the AR(2) model $H^{(1)}$ (full line) and the AR(4) model $H^{(2)}$ (dash-dotted line), which have a common pole pair at frequency 2.62 rad/sec.

6. APPLICATION TO REAL DATA

In this section we use the distances of Sections 3 and 4 on a real data set used in (Xiong and Yeung, 2004). We applied the same preprocessing steps and model structure choices as they did.

The EEG (electroencephalogram) data set was obtained from the UCI KDD Archive (http://kdd.ics.uci.edu/) and contains 10 trials of the 1 second multi-channel EEG recordings of a control subject and an alcoholic subject for three different conditions, denoted by '1', 'm' and 'n'. We included the recordings from two channels (F4, P8) and thereby obtained six experiments, each consisting of 20 time series (of length 256) where we wish to discriminate between the control subject and the alcoholic subject time series.

Following the approach in (Keirn and Aunon, 1990; Xiong and Yeung, 2004), AR(6) models were estimated for the time series after a differencing step to make the time series stationary. The clustering results are in Table 3. The best results are obtained with wcep, followed by cep, sp, h_2 and h_∞. The model-free distances generally perform worse than these. coef however performs the worst, probably because of the relatively high model order. The coef distance appeared also

Table 3. Clustering results for real data.

Data Set	Adjusted Rand index						
	cep	wcep	sp	h_2	h_∞	coef	mfree
$EEG_{1,F4}$	0.62	1.00	1.00	0.62	0.80	0.04	0.04
$EEG_{1,P8}$	0.80	1.00	0.80	0.80	0.80	0.12	0.62
$EEG_{m,F4}$	1.00	1.00	1.00	1.00	0.80	0.11	0.21
$EEG_{m,P8}$	0.80	0.80	0.62	0.62	0.46	0.22	0.04
$EEG_{n,P4}$	0.62	0.21	0.22	0.33	0.21	0.21	0.62
$EEG_{n,P8}$	0.80	0.80	0.80	0.62	0.80	-0.03	0.12

with other data sets and artificial examples to perform worse with complexer model structures.

7. CONCLUSIONS AND DISCUSSION

In this paper we considered distance-based clustering of scalar ARMA time series. The cluster examples investigated in Sections 5.2 and 6 showed that model-free distances are not appropriate for clustering AR(MA) time series, and that the model-based distances (except `coef`) had a much better performance than model-free distances on the considered data sets of EEG recordings.

It was also shown in Section 5.3 that the spectral distances (h_2 and especially h_∞ and `sp`) are less sensitive to differences in zeros or small spectral peaks than the cepstral (`cep` and `wcep`), when there is an additional and sufficiently large peak present in the spectrum. It also appeared from experiments that the `coef` distance degrades in performance for complexer model structures, such as ARMA (not shown) or high order AR models.

An important remark to be made with respect to the experiments in Section 5, is that for model estimation the correct model structure was always assumed to be known. We did some experiments with wrong model structures for identification (too simple or too complex) too. This decreased the performance of the model-based distances, although they remained substantially better performing than the model-free distances. In order, however, to state general conclusions about the effects of mismodelling on the different model-based distances, more research will be needed.

Another important future topic is the extension of the `wcep` distance for MIMO models. The SISO distance was shown in this paper to have a cepstral character. Since the distance can be defined in terms of angles between certain subspaces, an extension of these angles could result in a very natural extension of the distance to a MIMO cepstral distance, thereby avoiding the difficulty of taking the logarithm of the spectrum or transfer function matrix. It could also give ideas for a definition of the cepstrum for multivariable models.

ACKNOWLEDGMENTS

Research supported by Research Council KUL: GOA AMBioRICS, CoE EF/05/006 Optimization in Engineering, several PhD/postdoc & fellow grants; Flemish Government: FWO: PhD/postdoc grants, projects, G.0407.02 (SVM), G.0197.02 (power islands), G.0141.03 (Identification and cryptography), G.0491.03 (control for intensive care glycemia), G.0120.03 (QIT), G.0452.04 (new quantum algorithms), G.0499.04 (Statistics), G.0211.05 (Nonlinear), G.0226.06 (cooperative systems and optimization), G.0321.06 (Tensors), G.0553.06 (VitaminD), research communities (ICCoS, ANMMM, MLDM); IWT: PhD Grants, GBOU (McKnow), Eureka-Flite2; Belgian Federal Science Policy Office: IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling', 2002-2006); PODO-II (CP/40: TMS and Sustainability); EU: FP5-Quprodis; ERNSI; Contract Research/agreements: ISMC/IPCOS, Data4s, TML, Elia, LMS, Mastercard. Jeroen Boets is a research assistant with the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). Dr. Katrien De Cock is a postdoctoral researcher and Prof. Dr. Bart De Moor is a full professor at the K.U.Leuven.

REFERENCES

Antoulas, A. C. (1999). Approximation of linear dynamical systems. In: *Wiley Encyclopedia of Electrical and Electronics Engineering* (J. G. Webster, Ed.). Vol. 11. pp. 403–422.

Bagnall, A. J. and G. J. Janacek (2004). Clustering time series from ARMA models with clipped data. In: *Proc. 10th ACM SIGKDD Intl. Conf. on KDD*. ACM Press. Seattle, WA, USA. pp. 49–58.

Basseville, M. (1989). Distance measures for signal processing and pattern recognition. *Signal Processing* **18**, 349–369.

De Cock, K. (2002). Principal Angles in System Theory, Information theory and Signal Processing. PhD thesis. K.U.Leuven, Faculty of Applied Sciences, Leuven, Belgium. Available on `http://homes.esat.kuleuven.be/~decock/publications.html`.

De Cock, K. and B. De Moor (2002). Subspace angles between ARMA models. *Systems & Control Letters* **46**, 265–270.

Georgiou, T. T. and A. Lindquist (2003). Kullback-Leibler approximation of spectral density functions. *IEEE Trans. Information Theory* **49**, 2910–2917.

Gersch, W. (1981). Nearest neighbor rule in classification of stationary and nonstationary time series. In: *Applied Time Series Analysis II* (D. F. Findley, Ed.). pp. 221–270. Academic Press. New York.

Gray, R. M., A. Buzo, A. H. Gray, Jr. and Y. Matsuyama (1980). Distortion measures for speech processing. *IEEE Trans. Acoustics, Speech, and Signal Processing* **28**, 367–376.

Hubert, L. and P. Arabie (1985). Comparing partitions. *Journal of Classification* **2**, 193–218.

Kalpakis, K., D. Gada and V. Puttagunta (2001). Distance measures for effective clustering of ARIMA time-series. In: *Proc. IEEE ICDM*. San Jose, CA. pp. 273–280.

Kazakos, D. and P. Papantoni-Kazakos (1980). Spectral distance measures between Gaussian processes. *IEEE Trans. Automatic Control* **25**, 950–959.

Keirn, Z. A. and J. I. Aunon (1990). A new mode of communication between man and his surroundings. *IEEE Trans. Biomedical Engineering* **37**, 1209–1214.

Martin, R. J. (2000). A metric for ARMA processes. *IEEE Trans. Signal Processing* **48**, 1164–1170.

Schelfhout, G. (1996). Model Reduction for Control Design. PhD thesis. Faculty of Engineering, K.U.Leuven. Leuven (Belgium).

van der Laan, M. J., K. S. Pollard and J. Bryan (2003). A new partitioning around medoids algorithm. *J. of Statistical Computation and Simulation* **73**, 575–584.

Van Overschee, P. and B. De Moor (1996). *Subspace Identification for Linear Systems: Theory – Implementation – Applications*. Kluwer Academic Publishers. Boston.

Xiong, Y. and D. Y. Yeung (2004). Time series clustering with ARMA mixtures. *Pattern Recognition* **37**, 1675–1689.