Time series clustering and classification by autoregressive metric. Comput Stat Data Anal 52: 1860–1872

ARTICLE *in* COMPUTATIONAL STATISTICS & DATA ANALYSIS · FEBRUARY 2008

Impact Factor: 1.4 · DOI: 10.1016/j.csda.2007.06.001 · Source: RePEc

2 AUTHORS:

Marcella Corduas
University of Naples Federico II
31 PUBLICATIONS **135** CITATIONS
SEE PROFILE

Domenico Piccolo
University of Naples Federico II
59 PUBLICATIONS **438** CITATIONS
SEE PROFILE

Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2008) 1860–1872
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Time series clustering and classification by the autoregressive metric

Marcella Corduas*, Domenico Piccolo

Dipartimento di Scienze Statistiche, Università di Napoli Federico II, Via L. Rodinò 22, 80138 Napoli, Italy

Received 16 February 2006; received in revised form 31 May 2007; accepted 1 June 2007
Available online 3 June 2007

Abstract

The statistical properties of the autoregressive (*AR*) distance between *ARIMA* processes are investigated. In particular, the asymptotic distribution of the squared *AR* distance and an approximation which is computationally efficient are derived. Moreover, the problem of time series clustering and classification is discussed and the performance of the *AR* distance is illustrated by means of some empirical applications.

Keywords: Time series; Classification; Discrimination; Distances; Autoregressive Metric; ARIMA models

1. Introduction

Statistical techniques for time series clustering and classification are often necessary to provide useful information for the solution of real problems arising from different domains. For this reason, the study of distance measures and discriminating rules for time series has represented an important area of research in several scientific fields. In economics, for instance, the investigation of the economic cycle sometimes requires the seasonal adjustment of a consistent number of time series. In such a situation time series under scrutiny can be classified into groups with similar dynamic behavior so that they can be treated by applying the same seasonal adjustment filter (Corduas and Piccolo, 1999). In speech recognition, synthesized voice signals need to be attributed to specific word patterns (Gray and Markel, 1976). In astronomy, clustering helps to classify time series of star brightness in massive archives (Ng and Huang, 1999). In medicine, the study of biological signals requires to discriminate between signals caused by particular illness with respect to those of healthy people. This is the case of ECG (Kalpakis et al., 2001; Ge et al., 2002), EEG (Gersh et al., 1979; Alagón, 1989), EMG data (Kang et al., 1995; Kosěc, 2000). In seismology, it is relevant to discriminate the nature of seismic waves (Shumway and Unger, 1974; Dargahi-Noubary and Laycock, 1981; Kakizawa et al., 1998). Finally, in recent times the search for data mining techniques for the management of large data archives has renewed interest in time series clustering and discrimination (Agrawal et al., 1993; Ananthanarayana et al., 2001; Keogh and Kasetty, 2003).

Several approaches to time series comparison have been proposed in literature. The first approach is merely descriptive and relies on dissimilarity measures which directly compare observations or some features extracted from raw data. To

* Corresponding author. Tel.: +39 0812537461; fax: +39 0812537466.
E-mail address: marcella.corduas@unina.it (M. Corduas).

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.06.001

this end, in time domain, Bohte et al. (1980) and Kovačić (1996) used the autocorrelation and cross-correlation functions to summarize the temporal structure whereas Agrawal et al. (1994) considered the discrete Fourier transform of data and Struzik and Siebes (1999) the wavelets to map time series in frequency domain and extract dominant dynamic features. In a recent article, Caiado et al. (2006) proposed a metric based on periodogram as an aid to discriminate between stationary and non-stationary time series.

The second approach, instead, moves within an inferential framework. It reduces the comparison of time series to the definition of a dissimilarity measure between the underlying generating processes which are generally assumed to be linear and Gaussian. Consequently, dissimilarity is evaluated exploiting known results for multivariate Normal vectors leading, amongst the others, to measures such as the Kullback–Liebler information (Shumway and Unger, 1974; Kazakos and Papantoni-Kazakos, 1980) and the Bhattacharyya distance (Chaudury et al., 1991; Kailath, 1967). These results were extended to include vector stationary time series (Kakizawa et al., 1998; Taniguchi and Kakizawa, 2000) and non-stationary series (Shumway, 2003). In the same framework, other dissimilarity criteria based on the difference of estimated parameters of linear models have been defined. Thomson and De Souza (1985) introduced the Mahalanobis distance between autoregressive (*AR*) models and derived its distributional properties. This criterion was widely applied to speech recognition. Later, Maharaj (1999, 2000) extended an analogous testing procedure to the case of correlated univariate and multivariate stationary time series. The discrimination problem was also investigated as a model selection problem, see for instance Galeano and Peña (2000). Moreover, some contributions attempted at constructing a test statistic in order to verify whether two sets of data can be considered as coming from a common generating process using autocovariance functions (Mélard and Roy, 1984), spectral distributions (Anderson, 1993) and likelihood functions (Basawa et al., 1984).

More recently, the interest has focussed on composite procedures which combine different statistical techniques to obtain more reliable classification, such as the algorithm for clustering financial time series proposed by Pattarin et al. (2004), the method based on the use of functional analysis explored by Ingrassia et al. (2003), and the clustering technique developed by Alonso et al. (2006) based on the full probability density of forecasts.

An extensive review of the topic was illustrated by Liao (2005) who described the past research-work, the areas that time series clustering had been applied to and the source of data used.

In this article we investigate the statistical properties of the *AR* distance between *ARIMA* processes which measures the dissimilarity of two time series through the corresponding forecasting functions (Piccolo, 1990 and references included therein). This criterion has proved to be an effective tool in providing useful information for various aims: to produce a preliminary clustering of time series for seasonal adjustment, and to compare filters and results from different adjustment procedures (Corduas and Piccolo, 1999; Otranto and Triacca, 2002); to improve the design of a pollution monitoring network system (Sarno, 2005); to measure diversity in a genetic algorithm for clustering (Baragona et al., 2001); to develop a clustering algorithm (Maharaj, 1996); to classify price index series (Sarno and Zazzaro, 2002), and to daily hydrological time series (Grimaldi, 2004); and, finally, to compare the null and alternative hypotheses in testing problems involving *ARMA* parameters (Gonzalo and Lee, 1996). In addition, the *AR* metric is well defined also for *ARFIMA* processes and it has been successfully applied to derive a minimum distance estimator for the fractional parameter (Corduas, 2000).

The rest of this article is organized as follows: Section 2 introduces the *AR* distance on the class of *ARIMA* invertible models and derives its asymptotic distribution giving an approximation which is easily computable; Section 3 discusses the clustering and classification problem and provides a new discriminant rule. Finally, in Section 4 the resulting methodology is illustrated by some empirical examples concerning the identification of similarities among Industrial Production Index series in Italy and the clustering of ECG data for cardiac diseases classification.

2. The AR distance

According to the standard notation (Box and Jenkins, 1976; Brockwell and Davies, 1991), let Z_t be a zero mean invertible $ARIMA(p, d, q)(P, D, Q)_s$ process defined as

$$\varphi(B)\nabla^d\nabla_s^D Z_t = \vartheta(B)a_t, \tag{1}$$

where a_t is a white noise (WN) process with constant variance σ^2, B is the backshift operator such that $B^k Z_t = Z_{t-k}$, $\forall k = 0, \pm 1, \ldots$, the polynomials $\varphi(B) = \phi(B)\Phi(B^s) = (1 - \phi_1 B - \cdots - \phi_p B^p)(1 - \Phi_1 B^s - \cdots - \Phi_P B^{sP})$ and $\vartheta(B) = \theta(B)\Theta(B^s) = (1 - \theta_1 B - \cdots - \theta_q B^q)(1 - \Theta_1 B^s - \cdots - \Theta_Q B^{sQ})$, for any $s \geqslant 0$, have no common factors,

and all the roots of $\varphi(B)\vartheta(B)=0$ lie outside the unit circle. Moreover, we will assume that any outlier or deterministic component (such as trading days, calendar effects, mean level) has been previously removed from the series.

The invertibility assumption ensures that Z_t can be represented in terms of its past values according to the $AR(\infty)$ formulation

$$\pi(B)Z_t = a_t \tag{2}$$

with $\pi(B)=(1-B)^d(1-B^s)^D\varphi(B)\vartheta^{-1}(B)=1-\sum_{j=1}^{\infty}\pi_j B^j$ and $\sum_{j=1}^{\infty}|\pi_j|<\infty$. Assuming that a_t is a Gaussian process, given the initial values, the operators $\varphi(B)$, $\vartheta(B)$ and the WN variance characterize the probabilistic structure of the process Z_t. Moving from this consideration, Piccolo (1984, 1990) introduced the Euclidean distance between the π-weights of the $AR(\infty)$ formulations:

$$d=\sqrt{\sum_{j=1}^{\infty}\left(\pi_{xj}-\pi_{yj}\right)^2} \tag{3}$$

as a measure of *structural dissimilarity* between two *ARIMA* processes, X_t and Y_t, with given orders. The distance d is a well defined measure because of the absolute convergence of the π-sequences of processes belonging to the admissible *ARIMA* class, and satisfies the properties of a metric. Moreover, it has an interesting interpretation in terms of the forecast function of a linear process which is simply determined by the past lagged values of Z_t and the π-sequence. Therefore, distance between two *ARIMA* processes, with given orders, is zero if, provided the same set of initial values, the corresponding models produce the same forecasts.

The *AR* distance does not take the WN variance into account. This is considered as a mere scale factor depending on the measurement unit and such that it does not affect the temporal structure of a process. Nevertheless, the role of the WN variance is recovered in order to derive the distribution of the distance criterion.

Different approaches which explicitly introduce this factor have been already discussed in literature. For instance, in order to detect influential observations, Peña (1990) has considered the (squared) Mahalanobis distance to assess how the parameters of a model change when a single observation is iteratively removed from the time series and the estimation is performed in presence of a missing value. Such a measure is not a metric, it directly depends on the WN variance, and results in the squared Euclidean distance between the observed series and the one step ahead forecast series. Consequently, the Mahalanobis distance is strongly affected by the scale unit and cannot be used for clustering purposes since in real applications the comparison may involve phenomena observed on different scale measurements. The same remark applies to the similarity and dissimilarity measures that Tong and Dabas (1990) suggested for clustering the residuals from various statistical models fitted to the same time series. These criteria are effective for the purpose that the authors considered, but in general they cannot be used to compare two time series.

2.1. The asymptotic distribution

In this Section we derive the asymptotic distribution of the squared *AR* distance in order to set the comparison of time series within the hypotheses testing framework.

We note that d^2 is a mathematical function of the *AR* and *MA* parameters of the considered models. However, deriving an exact expression for (3) may be a rather difficult task given the increasing computing complexity related to models with large orders. For this reason, a numerical approximation of d^2, denoted as d_m^2, is introduced. This approximation is defined by truncating the π-weight expansion in (3) at the first m terms so that the contribution of ignored terms to the summation is negligible.

Given two observed time series $\{x_t\}$ and $\{y_t\}$, $t=1,\ldots,n$, the maximum likelihood (ML) estimator, $\widehat{d}_m^2$, is simply the squared Euclidean distance between the ML estimators of the π-weight sequences related to each *ARIMA* model.

Therefore, we recall briefly the asymptotic properties of ML estimators of *ARMA* parameters.

Let $\boldsymbol{\beta}=\{\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q,\Phi_1,\ldots,\Phi_P,\ \Theta_1,\ldots,\Theta_Q\}'$ be the parameters' vector of an $ARIMA(p,d,q)(P,D,Q)_s$ model and let $\boldsymbol{\pi}^A=(\pi_1^A,\ldots,\pi_m^A)'$ be the vector of the first m coefficients of the $AR(\infty)$ polynomial related to the *ARMA* part of the overall model, that is: $\pi(B)=(1-B)^d(1-B^s)^D\pi^A(B)$. We set $m>p+q+Ps+Qs$ and note that each π-weight is a function of the parameter vector: $\pi_i^A=g_i(\boldsymbol{\beta})$, $i=1,2,\ldots,m$. Furthermore, let $\widehat{\boldsymbol{\beta}}$ be the ML estimator of β derived from $\{x_t, t=1,\ldots,n\}$. Under standard regularity assumptions, $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ is asymptotically

Normal with zero mean vector and variance and covariance matrix $\mathbf{V}$. Specifically, $\mathbf{V} = \sigma^2[E(\mathbf{H}'\mathbf{H})]^{-1}$, where $\mathbf{H}$ is a $(n, p+q+P+Q)$ matrix whose jth column is given by the derivatives: $-\partial a_t/\partial\beta_j$, $j = 1, \ldots, p+q+P+Q$. Note that the matrix $\mathbf{V}$ only depends on the parameter vector $\boldsymbol{\beta}$ and does not depend on σ^2 (Box and Jenkins, 1976, pp. 281–284, 325). Then, asymptotically $\sqrt{n}(\widehat{\boldsymbol{\pi}}^A - \boldsymbol{\pi}^A) \overset{a}{\sim} N(\mathbf{0}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = \mathbf{BVB}'$, being $\mathbf{B} = \{b_{ij}\}$ a matrix with elements

$$b_{ij} = \left\{ \frac{\partial g_i(\widehat{\boldsymbol{\beta}})}{\partial \widehat{\beta}_j} \right\}_{\widehat{\boldsymbol{\beta}}=\boldsymbol{\beta}} \quad \text{for } i = 1, \ldots, m; \quad j = 1, \ldots, p+q+P+Q.$$

This result will be extended to *ARIMA* model since, in general, the vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)'$ may be expressed as

$$\boldsymbol{\pi} = \mathbf{A}\boldsymbol{\pi}^A + \boldsymbol{u}, \tag{4}$$

where $\mathbf{A}$ is a (m, m) lower triangular matrix where the elements of the jth column, a_{ij} for $j \leqslant i \leqslant j + d + sD$, are the coefficients of the terms B^h in the convolution $(1-B)^d(1-B^s)^D = \sum_{h=0}^{d+Ds} \alpha_h B^h$ and $a_{ij} = 0$ elsewhere; $\boldsymbol{u}$ is a vector such that: $u_h = -\alpha_h$, $h = 1, \ldots, d + Ds$; $u_h = 0$ for $h > d + Ds$.

Then, the *ARIMA* π-weights will satisfy $\sqrt{n}(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \overset{a}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Omega}\mathbf{A}'$.

In order to assess whether two time series have been originated from generating processes with the same temporal structure, we need to test the hypothesis $\mathrm{H}_0 : d^2 = 0$, which is equivalent to: $\mathrm{H}_0 : \boldsymbol{\pi}_x - \boldsymbol{\pi}_y = \mathbf{0}$. Let us assume that the processes X_t and Y_t are independent, and suppose that we observe the time series $\{x_t, t = 1, \ldots, n_x\}$ and $\{y_t, t = 1, \ldots, n_y\}$. As mentioned above, asymptotically, the ML estimators for the π-weights will satisfy $\sqrt{n_x}\,(\widehat{\boldsymbol{\pi}}_x - \boldsymbol{\pi}_x) \overset{a}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_x)$ and $\sqrt{n_y}\,(\widehat{\boldsymbol{\pi}}_y - \boldsymbol{\pi}_y) \overset{a}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_y)$. Hence, $(\widehat{\boldsymbol{\pi}}_x - \widehat{\boldsymbol{\pi}}_y) \overset{a}{\sim} N(\boldsymbol{\pi}_x - \boldsymbol{\pi}_y, (n_x^{-1}\boldsymbol{\Sigma}_x + n_y^{-1}\boldsymbol{\Sigma}_y))$.

Under H_0, since $\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}$, we get $(\widehat{\boldsymbol{\pi}}_x - \widehat{\boldsymbol{\pi}}_y) \overset{a}{\sim} N(\mathbf{0}, \mathbf{C}_0)$ being

$$\mathbf{C}_0 = \left(\frac{1}{n_x} + \frac{1}{n_y}\right) \mathbf{ABVB}'\mathbf{A}', \tag{5}$$

a $(p+q+P+Q, p+q+P+Q)$ matrix. Defining $\boldsymbol{\delta} = \mathbf{C}_0^{-1/2}(\widehat{\boldsymbol{\pi}}_x - \widehat{\boldsymbol{\pi}}_y)$, the ML estimator of d_m^2 can be written as $\widehat{d}_m^2 = \boldsymbol{\delta}'\mathbf{C}_0\boldsymbol{\delta}$. Consequently, the asymptotic distribution of $\widehat{d}_m^2$ is derived applying known results on quadratic forms in standard normal random variables since

$$\widehat{d}_m^2 = \boldsymbol{\delta}'\mathbf{C}_0\boldsymbol{\delta} = \sum_j \lambda_j \chi^2_{g_j}, \tag{6}$$

where λ_j, $j \leqslant p+q+P+Q$ are the non-zero eigenvalues of $\mathbf{C}_0$, and $\chi^2_{g_j}$ are independent Chi-square random variables with g_j degrees of freedom given by the multiplicity of each eigenvalue (usually, $g_j \equiv 1, \forall j$). The percentiles of such distribution are evaluated by numerical techniques (see: Imhof, 1961; Farebrother, 1990; Piccolo, 1989).

To facilitate the use of the *AR* distance an approximation of the distribution of a linear combination of Chi-squared random variable is needed (see Mathai and Provost, 1992, for an extensive review). In particular, we consider the approximation: $\widehat{d}_m^2 \sim a\chi^2_v + b$ where a, b, and v (which may be a non-integer value) are determined by the method of moments:

$$a = t_3/t_2, \quad b = t_1 - (t_2^2/t_3), \quad v = t_2^3/t_3^2 \tag{7}$$

with $t_k = \sum_j \lambda_j^k = \mathrm{tr}(\mathbf{C}_0^k)$, $k = 1, 2, 3$. It is important to note that $\mathrm{tr}(\mathbf{C}_0^k)$ involves only the $p+q+P+Q$ diagonal elements of the matrix $\mathbf{C}_0$ which, in general, has rather small dimension since *ARIMA* models satisfy the principle of parsimony. In practice, the matrix $\mathbf{C}_0$ will be estimated from observed time series by replacing the *ARMA* unknown parameters with the corresponding estimates in the expression: $(n_x^{-1}\boldsymbol{\Sigma}_x + n_y^{-1}\boldsymbol{\Sigma}_y)$.

In general, the complexity of the models object of comparison, which may involve non-seasonal or seasonal *MA* polynomials of high order, requires a strategy for selecting the value of m. As mentioned previously, the proposed distance uses the forecast function as a basis for comparing two *ARIMA* processes. When a stretch of n observations from a given process is observed, one of the approaches to evaluating the one-step ahead prediction at time $n+1$ is computing a linear combination of the past n observed values by means of the estimated π-weights (Brockwell and

Davies, 1991, pp. 183–184). In this context the truncation point for the approximated predictor is simply the number of observations available. By analogy, this consideration leads to a simple rule for the selection of the value of m for the computation of d_m^2. In fact, m can be determined so to be equal to the length of the shorter time series involved in the comparison. In such a way, in computing the *AR* distance the user is behaving according to the same strategy he/she would use if he/she had to evaluate the one-step ahead forecast of the time series using the truncated predictor.

Of course, as it happens for other inferential problems, the "pile up effect" of the sampling distribution of the *MA* estimators affects also the computation and the distributional properties of the proposed criterion (see for instance, Ansley and Newbold, 1980). However, empirical experiences have shown that this extreme situation, due to models with *MA* operators with roots very close to the unit circle, is not relevant in practice since it occurs very rarely in real applications.

Moreover, the distribution of the *AR* metric is derived under the assumption that the orders characterizing the *ARIMA* processes, X_t and Y_t, are known. This implies that both uncorrect model specification and unit root problem are not considered.

3. Clustering and discriminating time series

The *AR* distance supplies a measure for clustering time series. Using (3) on a sample of time series leads to a distance matrix that can be analyzed by one of the hierarchical clustering procedures or, alternatively, by a k-means method, such as that described by Kaufman and Rousseeuw (1990) which produces partitions around medoids.

Although clustering techniques provide useful data representations which enhance similarities among series, whatever method is chosen, the interpretation of results is confined to a descriptive level.

The *AR* metric, instead, allows the comparison of time series within a testing hypotheses framework and gives a substantive way to assess their "nearness". On one hand, the *AR* metric provides, in fact, an aid to understand the results of clustering displayed by means of a dendrogram, on the other it is a tool in itself for constructing time series clustering as we will illustrate in Section 4 by means of two real applications.

In this respect, we suggest to transform the original matrix of squared *AR* distances into a binary matrix where the (i, j)th entry takes the value 1 when the squared distance between the ith and jth models is not significant (at a specified significance level) and takes the value 0 otherwise.

Hence, clusters of time series are found by reordering such a matrix into an approximate block diagonal form. The original sequence of objects are so arranged that the elements of every cluster always lie in consecutive rows and columns of the permuted matrix. In general, a clearly defined (unit) triangle immediately under the diagonal will indicate a cluster well separated from neighboring points and such that all the squared *AR* distances between elements of the cluster are not significant at a specified significance level. This fact implies that the hypothesis that all the time series corresponding to this triangle were generated by the same *ARIMA* process cannot be rejected. Moreover, if the triangle contains any zero value, the cluster may be elongated or it may contain other small clusters that are themselves separated.

Of course, the problem becomes the selection of the rows/columns permutation which achieves the better partition of data. For this purpose, one of the many algorithms proposed in literature to construct diagonal block matrices can be applied. This technique has been extensively used over the last three decades though it has been referred to with various names (structuring of matrices, data reorganization, restricted partitioning, rearrangement clustering, etc.). The problem is in fact common to a number of different applications in diverse areas such as operation research (the well known *travelling salesman problem*), marketing, imaging, data mining and clustering of genes (see for instance Tran-Luu and DeClaris, 1997; Climer and Zhang, 2006 and references reported therein). In the rest of this article we will refer to the bond energy algorithm (BEA) which has been widely studied in literature as a tool for rearrangement clustering (McCormick et al., 1972; Arabie and Hubert, 1990). The detailed description of BEA algorithm and properties goes beyond the scope of this article; we will limit ourselves to illustrate the method briefly and to show, in the following section, the results that can be achieved in practice.

The original algorithm, developed by McCormick et al. (1972), operates on an $M \times N$ matrix $\mathbf{A}$ of non-negative entries. The rows and columns of $\mathbf{A}$ are permuted so as to maximize the expression:

$$ME = \sum_{j=1}^{M} \sum_{k=1}^{N} a_{j,k}[a_{j,k-1} + a_{j,k+1} + a_{j-1,k} + a_{j+1,k}], \tag{8}$$

where the maximization is over all $N!M!$ possible arrays that can be obtained from permuting $\mathbf{A}$ (with the convention that $a_{0,k}=a_{M+1,k}=a_{j,0}=a_{j,N+1}=0$). The authors denoted ME as the sum of the "bond strengths" in the matrix where the "bond strength" between two nearest-neighbor elements is given by their product. The idea is that large values will be drawn to other large values (and vice versa small values to other small values) so as to increase the overall sum of the products.

In general, the row and column permutation problem can be separated due to the additivity property of the method so that the optimization of ME can be performed in two passes. Nevertheless, in the specific case that we are considering, since the binary matrix is symmetric, the same optimal ordering must hold for both rows and columns; hence it is only necessary to compute this ordering once.

3.1. Discriminant rule

The squared *AR* distance can be used to classify a finite observed realization $\{x_t, t=1,\ldots,n\}$ into one of several categories describing the temporal structure of the *ARIMA* generating process as specified by the corresponding operator $\pi_j(B)$. Although we focus on the case of two populations, the results can be easily extended to more general situations. In particular, we assume that the two categories are described by the hypotheses H_1 and H_2 concerning the π-sequences characterizing the process X_t, that is: $\boldsymbol{\pi}_1=\{\pi_{1,k}, k=1,\ldots,m\}$ and $\boldsymbol{\pi}_2=\{\pi_{2,k}, k=1,\ldots,m\}$.

The squared *AR* distance is used to measure how close the estimated weights $\hat{\boldsymbol{\pi}}_x=\{\hat{\pi}_{x,k}, k=1,\ldots,m\}$ are to the π-sequences implied by the two hypotheses. Then, we propose the following rule:

$$\text{assign } x_t \text{ to } H_1 \text{ (or } H_2\text{) according as } D>0 \text{ (or } D\leqslant 0),$$

where $D=d^2(\hat{\boldsymbol{\pi}}_x,\boldsymbol{\pi}_2)-d^2(\hat{\boldsymbol{\pi}}_x,\boldsymbol{\pi}_1)$.

Given the properties of the *AR* metric, the discriminant function D matches the observed time series to the category whose temporal structure is closer.

It is easy to see that $\sqrt{n}/2\{D+(-1)^j d^2(\boldsymbol{\pi}_1,\boldsymbol{\pi}_2)\}=\sqrt{n}(\boldsymbol{\pi}_1-\boldsymbol{\pi}_2)'(\hat{\boldsymbol{\pi}}_x-\boldsymbol{\pi}_j)$ for $j=1,2$. Then, recalling that, under H_j with $j=1,2$, asymptotically, $\sqrt{n}(\hat{\boldsymbol{\pi}}_x-\boldsymbol{\pi}_j)\overset{a}{\sim}N(0,\boldsymbol{\Sigma}_j)$, we have that:

$$\frac{\sqrt{n}}{2}\{D+(-1)^j d^2(\boldsymbol{\pi}_1,\boldsymbol{\pi}_2)\}\overset{a}{\sim}N(0,v_j^2), \tag{9}$$

where

$$v_j^2=(\boldsymbol{\pi}_1-\boldsymbol{\pi}_2)'\boldsymbol{\Sigma}_j(\boldsymbol{\pi}_1-\boldsymbol{\pi}_2), \quad j=1,2.$$

Assuming $d^2(\boldsymbol{\pi}_2,\boldsymbol{\pi}_1)>0$, the misclassification probabilities are

$$P(2|1)=P(D\leqslant 0|H_1)=\Phi\left(-\frac{\sqrt{n}}{2}\frac{d^2(\boldsymbol{\pi}_1,\boldsymbol{\pi}_2)}{v_1}\right)$$

and

$$P(1|2)=P(D>0|H_2)=1-\Phi\left(\frac{\sqrt{n}}{2}\frac{d^2(\boldsymbol{\pi}_2,\boldsymbol{\pi}_1)}{v_2}\right).$$

Note that $\lim_{n\to\infty}P(2|1)=\lim_{n\to\infty}P(1|2)=0$, that is the discriminant statistic is consistent in the sense that the misclassification probabilities tend to zero as $n\to\infty$. Moreover, if $d^2(\boldsymbol{\pi}_1,\boldsymbol{\pi}_2)=0$, as one would expect, $P(2|1)=P(1|2)=0.5$.

In real applications, the π-weight sequences $\{\pi_{1,k}\}$ and $\{\pi_{2,k}\}$ in H_1 and H_2 are often unknown and they have to be estimated using learning samples. Suppose that for $j=1,2$ we have s_j observed time series: $x_{i,t}^{(j)}$, $t=1,\ldots,n$, $i=1,\ldots,s_j$, correctly classified with respect to the two categories defined by H_j. These represent the learning samples. Assume, also, that *ARIMA* models are fitted to those time series and that the *medoid* model for each category is identified. In this context, we define the *medoid* model as the one which minimizes the sum of the *AR* distances with respect to all the other models belonging to the same category. Then, we suggest to use the estimated π-weight sequence of the two *medoid* models in place of the unknown $\{\pi_{1,k}\}$ and $\{\pi_{2,k}\}$.

Table 1
ARIMA models for industrial production indices series in Italy

No.	Series	d	D	ϕ_1	ϕ_2	ϕ_3	θ	Θ	$\widehat{\sigma}^2$
1	Mining: non-energetic ore	0	12	0.536 (*0.073*)	0.152 (*0.082*)	0.128 (*0.073*)		0.618 (*0.053*)	60.48
2	Mining: energetic ore	1	0				0.434 (*0.063*)	−0.175 (*0.069*)	43.72
3	Food, beverage, tobacco	0	12	0.346 (*0.072*)	0.224 (*0.072*)			0.700 (*0.053*)	7.20
4	Textile and clothing products	1	12				0.541 (*0.062*)	0.436 (*0.067*)	14.08
5	Leather and leather products	1	12				0.563 (*0.061*)	0.350 (*0.069*)	17.19
6	Wood and wood products	1	12				0.600 (*0.057*)	0.593 (*0.056*)	13.47
7	Coke, oil, nuclear fuel	0	12	0.313 (*0.071*)		0.115 (*0.075*)		0.882 (*0.043*)	24.51
8	Paper, print and publishing	0	12	0.398 (*0.073*)	0.191 (*0.077*)	0.130 (*0.073*)		0.761 (*0.056*)	10.45
9	Chemical products	1	12				0.663 (*0.055*)	0.585 (*0.062*)	8.82
10	Rubber and plastic products	1	12				0.547 (*0.067*)	0.636 (*0.063*)	12.53
11	Non-metallic mineral products	0	12	0.584 (*0.077*)		0.256 (*0.078*)		0.634 (*0.061*)	7.44
12	Basic and fabricated metal products	1	12				0.597 (*0.067*)	0.625 (*0.065*)	13.60
13	Machinery and equipment	1	12				0.540 (*0.067*)	0.427 (*0.065*)	18.82
14	Electrical and optical equipment	1	12				0.500 (*0.065*)	0.342 (*0.069*)	14.52
15	Transport equipment	1	12				0.382 (*0.071*)	0.708 (*0.053*)	20.99
16	Other manufacturing industry	1	12				0.523 (*0.063*)	0.710 (*0.057*)	27.86
17	Electrical power, gas and water supply	0	12	0.568 (*0.060*)				0.588 (*0.068*)	9.82
18	General industrial production	1	12				0.515 (*0.063*)	0.809 (*0.046*)	2.82

Table 2
Rearranged binary matrix for industrial production indices series in Italy

No.	13	14	5	4	6	9	12	10	3	16	11	15	7	8	18	1	17	2
13	1																	
14	1	1																
5	1	1	1															
4	1	1	1	1														
6	1	0	0	1	1													
9	0	0	0	1	1	1												
12	0	0	0	1	1	1	1											
10	0	0	0	1	1	1	1	1										
3	0	0	0	0	1	0	1	1	1									
16	0	0	0	0	1	1	1	1	1	1								
11	0	0	0	0	0	0	0	1	0	1	1							
15	0	0	0	0	0	0	0	1	0	1	1	1						
7	0	0	0	0	0	0	0	0	0	0	0	0	1					
8	0	0	0	0	1	1	1	1	1	1	1	1	1	1				
18	0	0	0	0	0	0	0	1	1	1	0	1	0	1	1			
1	0	0	0	0	1	0	1	1	1	1	1	1	0	1	1	1		
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

The performance of the proposed rule, assuming that learning samples are used, has been investigated by Corduas (2004) by means of a simulation study. In particular, the rule was tested, by analogy with other contributions (see for instance Zhang and Taniguchi, 1995), in the case that $H_1 : X_t = \phi_1 X_{t-1} + a_t$ and $H_2 : X_t = \phi_2 X_{t-1} + a_t$ with $\phi_2 = \phi_1 + h/\sqrt{(n)}$ for varying parameters. The asymptotic misclassification probabilities of the D-rule resulted very close with those of the approximated quadratic discriminant rule derived in the frequency domain by Shumway (1982, pp. 17–18).

4. Case studies

We illustrate the use of the *AR* metric for time series classification by means of two real examples. The first application refers to economic time series data and is aimed at identifying similarities among industrial production index series in Italy. The second example concerns the comparison of various electrocardiogram time series in order to detect a special type of arrhythmia with respect to normal sinus rhythm (NSR) of heart of healthy people.

4.1. Italian industrial production indices

We consider 17 time series consisting of monthly Italian industrial production (by branch) indices, adjusted for the working days effects, from January 1990 to August 2006 (source: http://www.istat.it, ATECO 2002 classification). The time series models are reported in Table 1. The estimated parameters standard errors are given in parentheses. We have also considered the general Industrial Production indices series (18) as a reference for the comparison of the indices by branch.

As discussed previously, a binary matrix was obtained from the squared distance matrix after having tested the hypothesis of homogeneity for each couple of models at 5% significance level. Then, the BEA algorithm was applied to produce a block diagonal matrix. Due to the symmetry of the matrix, only the lower triangle is illustrated in Table 2.

The clusters are identified along the main diagonal of the matrix by blocks of unit values. In particular, the following groups are identified: $g_1 = (13, 14, 5, 4)$, $g_2 = (6, 9, 12, 10, 3, 16)$, $g_3 = (11, 15)$, $g_4 = (8, 18, 1)$ whereas the elements 2, 7 and 17 are isolated. Note that some elements act as a bridge between groups. Specifically, 4 connects g_1 to g_2 whereas 8 and 16 links g_2 to g_3 and g_4.

The partition is similar in many aspects to the clustering produced by complete linkage method as shown by the dendrogram in Fig. 1. However, sectioning a dendrogram at a certain height is a critical issue; varying the threshold level will tend to produce groups which have a different degree of internal homogeneity, and for that reason, hierarchical

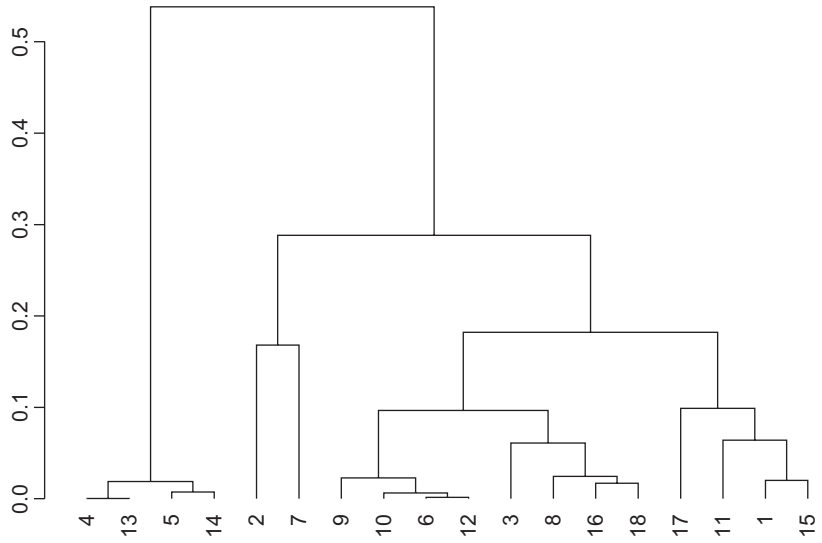

Fig. 1. Dendrogram of Industrial Production Indices series.

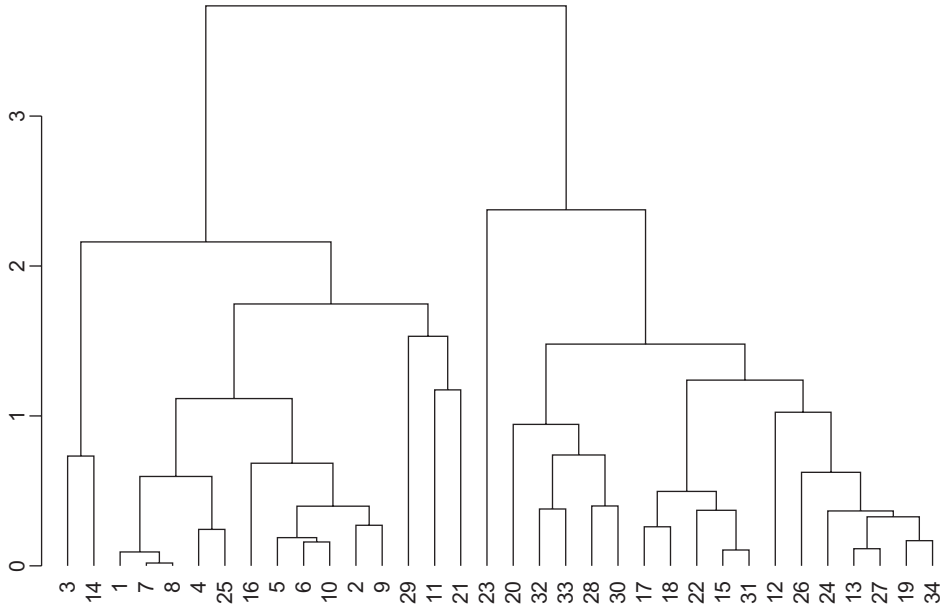

Fig. 2. Dendrogram by complete linkage method (1–10 are NSR series; 11–34 are SVT series).

clustering methods are usually regarded as a form of exploratory data analysis. In this example, for instance, if the dendrogram was sectioned at level 0.2, the series 2 and 7 would be joined together whereas, instead, according to the *AR* distance, the hypothesis of homogeneity is rejected at 5% significance level (see Table 2).

As mentioned previously, the *AR* distance assesses how close the forecasting functions of two *ARIMA* models are. The forecasting function is determined by all the operators characterizing a model. Then, as this example shows, the comparison is possible irrespective of the presence/absence of difference operators in the models. This is an advantage of the *AR* distance over other measures defined in frequency domain, such as, for instance, the Kullback–Liebler divergence (Shumway and Unger, 1974), which are well defined only for stationary series.

Table 3
Rearranged binary matrix for ECG series (1–10 are NSR series; 11–34 are SVT series)

No.	17	31	18	15	22	13	27	19	34	24	26	33	20	28	30	11	32	1	3	14	8	7	6	25	10	9	5	2	4	16	29	23	21	12
17	1																																	
31	1	1																																
18	1	1	1																															
15	1	1	1	1																														
22	1	1	1	1	1																													
13	1	0	0	0	1	1																												
27	0	0	0	0	0	1	1																											
19	1	0	0	0	1	1	1	1																										
34	0	0	0	0	1	1	1	1	1																									
24	0	0	0	0	0	1	1	1	1	1																								
26	0	0	0	0	0	1	1	1	1	1	1																							
33	0	0	0	0	0	1	1	1	1	1	1	1																						
20	0	0	0	0	0	0	0	0	0	1	0	1	1																					
28	0	0	0	0	0	0	0	0	0	0	0	1	1	1																				
30	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1																			
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1																		
32	0	0	0	0	0	0	0	1	1	0	0	1	0	1	1	1	1																	
1	0	0	0	0	0	0	0	1	1	0	0	1	0	1	1	1	1	1																
3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1															
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1														
8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1													
7	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1												
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1											
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1										
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1									
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1								
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1							
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1						
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1					
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1				
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0	1			
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

4.2. ECG data

We now apply the *AR* distance to the analysis of ECG data in order to detect a special type of arrhythmia, the supraventricular tachycardia (SVT), with respect to the NSR of heart of healthy people. The data are obtained from the MIT-BIH arrhythmia database (available from http://www.physionet.org). The development of algorithms based on ECG for accurate diagnosis of cardiac arrhythmias has originated a vast literature but the reference to it overcomes the purpose of this paper. We mention only the recent contributions of Ge et al. (2002) and Kalpakis et al. (2001) since both refer to the use of *AR* models for cardiac arrhythmia time series. The first contribution classifies the *AR* coefficients using a generalized linear model based algorithm, the second one, instead, applies the Euclidean distance between cepstral coefficients and compares it with a weighted Euclidean distance between the *AR* coefficients. The criteria are again used as exploratory tools.

For our aim, we selected 10 time series for the NSR and 24 time series for the SVT. The sampling frequency was set at 250 Hz. Each series consisted of 300 observations which corresponds to 1.2 s of recordings. This time period was considered adequate to capture at least one complete cardiac cycle. For ECG series, modelling which can be easily implemented is generally needed. For this reason, *AR* models were fitted to the data. The number of *AR* parameters was identified by minimizing the Schwarz Bayesian information criterion (Schwarz, 1978). We preferred this criterion instead of AIC criterion (Akaike, 1974) since the former tends to select simpler models than those chosen by AIC.

The problem with AIC, as noticed by a number of authors, including Schwarz, is that it tends to overfit the data and in some situations, such as in presence of small samples or when the number of parameters is moderate with respect to the sample size, the overfitting can be considerably large (see Hurvich and Tsai, 1989).

Firstly, we applied standard clustering techniques to the *AR* distance matrix. In Fig. 2 we report the dendrogram obtained by the complete linkage method. Sectioning the dendrogram at the highest level, we find two large clusters. The first one includes 18 series from SVT patients. The second one consists of 10 NSR series and 6 SVT series. Moving at a lower level the classification improves since some elements of the second groups (29,11,21) are isolated from the others so that only three SVT series are misclassified. At this stage, the dendrogram is providing only a description of data and, therefore, no conclusive deduction on the groups can be taken since the result depends on the user's selection of the threshold value.

Instead, by means of the *AR* distance the identification of groups is inferred using a testing hypotheses framework. Following the approach described earlier, we constructed the binary matrix indicating which couple of series were generated by a common *ARIMA* process according to the test on the *AR* distance at 5% significance level. In such a case, $\mathbf{C}_0$ has a very simple expression being: $\mathbf{C}_0 = n^{-1}(\boldsymbol{V}_x + \boldsymbol{V}_y)$.

In Table 3, we report the binary matrix rearranged by means of the BEA algorithm. We identify the following groups of series: $g_1 = (17, 31, 18, 15, 22)$, $g_2 = (13, 27, 19, 34, 24, 26)$, $g_3 = (20, 28, 30)$ $g_4 = (11, 32)$, $g_5 = (1, 3, 14, 8, 7, 6)$, $g_6 = (25, 10, 9, 5, 2, 4, 16)$. Three SVT series are joined with NSR series; these recordings are probably related to patients at the initial state of the illness. The isolated elements—{12}, {21}, {23} and {29}—refer only to SVT series. They are correctly separated from the NSR series, suggesting the presence of an abnormal behavior, although they are not joined to any group. This could be justified by the fact that the SVT arrhythmia can manifest itself in several forms. Finally, several elements of g_5 are linked with elements in g_6, hence, the two groups may be joined in a single group reconstructing the whole set of healthy people ECG recordings.

5. Conclusions

We discussed the statistical properties of the *AR* distance deriving the asymptotic distribution and an adequate approximation which is easily computable. We demonstrated by means of two real examples how the *AR* distance can be applied to classification problems.

The inferential properties of the proposed criterion is a meaningful result since, differently from other criteria proposed in literature, it allows clustering not to be confined within the bounds of exploratory methods. Testing the hypothesis that two time series are originated by the same generating process, in fact, contributes to give strength to the results from a clustering analysis.

Finally, it is important to note that the *AR* metric is well defined for stationary and non-stationary, seasonal and non seasonal, short and long memory processes and, in this respect, it is a wide applicable statistical tool.

Acknowledgement

The authors gratefully acknowledge the helpful comments and suggestions of the associate editor and referees on an earlier version of this paper. This research was supported by Dipartimento di Scienze Statistiche—Università di Napoli Federico II and CFEPSR (Portici, Italy).

References

Agrawal, R., Imielinski, T., Swami, A., 1993. Database mining: a performance perspective. IEEE Trans. Knowledge Data Eng. 5, 914–925.

Agrawal, R., Faloutsos, C., Swami, A., 1994. Efficient similarity search in sequence databases. Fourth Proceedings of F.O.D.O. '93, Lecture Notes in Computer Science, vol. 730, Springer, New York, pp. 69–84.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automatic Control AC 19, 203–217.

Alagón, J., 1989. Spectral discrimination of two groups of time series. J. Time Ser. Anal. 10, 203–214.

Alonso, A.M., Berrendero, J.R., Hernández, A., Justel, B., 2006. Time series clustering based on forecast densities. Comput. Statist. Data Anal. 51, 762–776.

Ananthanarayana, V.S., Murty, M.N., Subramanian, D.K., 2001. Efficient clustering of large data set. Pattern Recognition 34, 2561–2563.

Anderson, T.W., 1993. Goodness of fit tests for spectral distributions. The Ann. Statist. 21, 830–847.

Ansley, C.F., Newbold, P., 1980. Finite sample properties for Autoregressive Moving Average models. J. Econo. 13, 159–183.

Arabie, P., Hubert, L.J., 1990. The bond energy algorithm revisited. IEEE Trans. Systems Man Cybernet. 20, 268–274.

Baragona, R., Battaglia, F., Cucina, D., 2001. Clustering of time series with genetic algorithms. Metron 59, 113–130.
Basawa, I.V., Billard, L., Srnivasan, R., 1984. Large sample tests of homogeneity for time series. Biometrika 71, 203–206.
Bohte, Z., Cepar, D., Kosmelij, K., 1980. Clustering of time series. Proceedings of COMPSTAT80, pp. 587–593.
Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control (rev edition). Holden-Day, San Francisco.
Brockwell, P.J., Davies, R.A., 1991. Time Series: Theory and Methods. second ed. Springer, New York.
Caiado, J., Crato, N., Peña, D., 2006. A periodogram-based metric for time series classification. Comput. Statist. Data Anal 50, 2668–2684.
Chaudury, G., Borwarkar, J.D., Rao, P.R.K., 1991. Bhattacharyya distance based linear discriminant function for stationary time series. Comm. Statist. Theory Methods 20, 2195–2205.
Climer, S., Zhang, W., 2006. Rearrangement clustering: pitfalls, remedies and applications. J. Mach. Learn. 7, 919–943.
Corduas, M., 2000. Preliminary estimation of ARFIMA models. In: Betlehem, J.G., van der Heijden, P.G.M. (Eds.), Proceedings in Computational Statistics. Physica, Heidelberg, pp. 247–252.
Corduas, M., 2004. Time series discrimination using AR metric. Proceedings of XLII Riunione Scientifica SIS, CLEUP, Padova, pp. 143–146.
Corduas, M., Piccolo D., 1999. An application of the AR metric to seasonal adjustment. Bulletin of the International Statistical Institute, vol. LVIII, pp. 217–218.
Dargahi-Noubary, G.R., Laycock, P.J., 1981. Spectral ratio discriminants and information theory. J. Time Ser. Anal. 2, 71–86.
Farebrother, R.W., 1990. The distribution of a quadratic form in normal variables. Appl. Statist. 39, 294–309.
Galeano, P., Peña, D., 2000. Multivariate analysis in vector time series. Resenhas 4, 383–404.
Ge, D., Srinivasan, N., Krishnan, S.M., 2002. Cardiac arrhythmia classification using autoregressive modeling. Biomed. Eng. OnLine, ⟨http://www.biomedical-engineering-online.com⟩.
Gersh, W., Martinelli, F., Yonemoto, J., Low, M.D., McEwan, J.A., 1979. Automatic classification of electroencephalograms: Kullback–Liebler nearest neighbor rules. Science 205, 193–195.
Gonzalo, J., Lee, T.H., 1996. Relative power of t type tests for stationary and unit root processes. J. Time Ser. Anal. 17, 37–47.
Gray, A.H., Markel, J.D., 1976. Distance measures for speech processing. IEEE Trans. Acoust., Speech and Signal Processing ASSP-24, 380–391.
Grimaldi, S., 2004. Linear parametric models applied on daily hydrological series. J. Hydrol. Eng. 9, 383–391.
Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307.
Imhof, P.J., 1961. Computing the distribution of quadratic forms in Normal variables. Biometrika 48, 419–426.
Ingrassia, S., Cerioli, A., Corbellini, A., 2003. Some issues on clustering of functional data. In: Schader, M., Gaul, W., Vichi, M. (Eds.), Between Data Science and Applied Data Analysis. Springer, Berlin, pp. 49–56.
Kailath, T., 1967. The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans. Comm. Technol. COM-15, 52–60.
Kakizawa, Y., Shumway, R.H., Taniguchi, M., 1998. Discrimination and clustering for multivariate time series. J. Amer. Statist. Assoc. 93, 328–340.
Kalpakis, K., Gada, D., Puttagunda, V., 2001. Distance measures for effective clustering of ARIMA time series. Proc. IEEE Internat. Conf. Data Mining, 273–280.
Kang, W., Cheng, C., Lai, J., Tsao, H., 1995. The application of cepstral coefficients and maximum likelihood method in EGM pattern recognition. IEEE Trans. Biomed. Eng. 42, 777–785.
Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York.
Kazakos, D., Papantoni-Kazakos, P., 1980. Spectral distances between Gaussian processes. IEEE Trans. Automat. Control AC-25, 950–959.
Keogh, E., Kasetty, S., 2003. On the need for time series data mining benchmarks: a survey and empirical demonstration. Data Mining Knowledge Discovery 7, 349–371.
Kosĕc, D., 2000. Parametric estimation of continuous non stationary spectrum and its dynamics in surface EMG studies. Internat. J. Med. Inform. 58-59, 59–69.
Kovačić, Z.J., 1996. Classification of time series with application to the leading indicator selection. Proceedings of the Fifth Conference of IFCS, vol. 2, pp. 204–207.
Liao, T.W., 2005. Clustering time series data—a survey. Pattern Recognition 38, 1857–1874.
Maharaj, E.A., 1996. A significance test for classifying ARMA models. J. Statist. Comput. Simulation 54, 305–331.
Maharaj, E.A., 1999. Comparison and classification of stationary multivariate time series. Pattern Recognition 32, 1129–1138.
Maharaj, E.A., 2000. Clusters of time series. J. Classification 17, 297–314.
Mathai, A.M., Provost, S.B., 1992. Quadratic Forms in Random Variables. Marcell Decker, New York.
McCormick, W.T., Schweitzer, P.J., White, T.W., 1972. Problem decomposition and data reorganization by a clustering technique. Oper. Res. 20, 993–1009.
Mélard, G., Roy, R., 1984. Sur un test d'égalité des autocovariances de deux séries chronologiques. Canad. J. Statist. 12, 333–342.
Ng, M.K., Huang, Z., 1999. Data mining massive time series astronomical data: challenges, problems and solutions. Inform. Software Technol. 41, 545–556.
Otranto, E., Triacca, U., 2002. Measures to evaluate the discrepancy between direct and indirect model-based seasonal adjustment. J. Official Statist. 18, 511–530.
Pattarin, F., Paterlini, S., Minerva, T., 2004. Clustering financial time series: an application to mutual funds style analysis. Comput. Statist. Data Anal. 47, 353–372.
Peña, D., 1990. Influential observation in time series. J. Business and Econom. Statist. 8, 235–242.
Piccolo, D., 1984. Una topologia per la classe dei processi ARIMA. Statistica, XLIV, 47–59.
Piccolo, D., 1989. On the measure of dissimilarity between ARIMA models. In: Proceedings of the A.S.A. Meetings, Business and Economic Statistics Section, Washington, DC, pp. 231–236.
Piccolo, D., 1990. A distance measure for classifying ARIMA models. J. Time Ser. Anal. 11, 153–164.
Sarno, E., 2005. Testing information redundancy in environmental monitoring networks. Environmetrics 16, 71–79.

Sarno, E., Zazzaro, A., 2002. An index of dissimilarity among time series: an application to the inflation rates of the EU countries. In: Klinke, S., Ahrend, P., Richter, L. (Eds.), Proceedings of COMPSTAT 2002. Springer, Berlin.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.

Shumway, R.H., 1982. Discriminant analysis for time series. In: Krishnaiah, P.R., Kanal, L.N. (Eds.), Handbook of Statistics, vol. 2. North Holland, Amsterdam, pp. 1–46.

Shumway, R.H., 2003. Time-frequency clustering and discriminant analysis. Statist. Probab. Lett. 63, 307–314.

Shumway, R.H., Unger, A.N., 1974. Linear discriminant functions for stationary time series. J. Amer. Statist. Assoc. 65, 1527–1546.

Struzik, Z.R., Siebes, A., 1999. The Haar wavelet in the time series similarity paradigm. In: Proceedings of the Third European Conference on Principles of Data Mining and Knowledge Discovery. Springer, Prague, pp. 12–22.

Taniguchi, M., Kakizawa, Y., 2000. Asymptotic Theory of Statistical Inference for Time Series. Springer, New York.

Thomson, P.J., De Souza, P., 1985. Speech recognition using LPC distance measures. In: Hannan, E.J., Krishnaiah, P.R., Rao, M.M. (Eds.), Handbook of Statistics, vol. 5. North Holland, Amsterdam, pp. 389–412.

Tong, H., Dabas, P., 1990. Cluster of time series. J. Appl. Statist. 17, 187–198.

Tran-Luu, T.D., DeClaris, N., 1997. Visual heuristics for data clustering. IEEE Trans. Systems Man Cybernet. 1, 19–24.

Zhang, G., Taniguchi, M., 1995. Nonparametric approach for discriminant analysis in time series. Nonparametric Statist. 5, 91–101.