Expert Systems With Applications 185 (2021) 115677

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Quantile cross-spectral density: A novel and effective tool for clustering multivariate time series

Ángel López-Oriona [a,*], José A. Vilar [a,b]

[a] Research Group MODES, Research Center for Information and Communication Technologies (CITIC), University of A Coruña, 15071 A Coruña, Spain
[b] Technological Institute for Industrial Mathematics (ITMATI), Spain

ARTICLE INFO

Keywords:
Multivariate time series
Clustering
Dissimilarity measure
Quantile cross-spectral density
S&P 500
UEA archive

ABSTRACT

Clustering of multivariate time series is a central problem in data mining with applications in many fields. Frequently, the clustering target is to identify groups of series generated by the same multivariate stochastic process. Most of the approaches to address this problem include a prior step of dimensionality reduction which may result in a loss of information or consider dissimilarity measures based on correlations and cross-correlations but ignoring the serial dependence structure. We propose a novel approach to measure dissimilarity between multivariate time series aimed at jointly capturing both cross dependence and serial dependence. Specifically, each series is characterized by a set of matrices of estimated quantile cross-spectral densities, where each matrix corresponds to a pair of quantile levels. Then the dissimilarity between every couple of series is evaluated by comparing their estimated quantile cross-spectral densities, and the pairwise dissimilarity matrix is taken as starting point to develop a partitioning around medoids algorithm. Since the quantile-based cross-spectra capture dependence in quantiles of the joint distribution, the proposed metric has a high capability to discriminate between high-level dependence structures. An extensive simulation study shows that our clustering procedure outperforms a wide range of alternative methods and exhibits robustness to noise distribution besides being computationally efficient. A real data application involving bivariate financial time series illustrates the usefulness of the proposed approach. The procedure is also applied to cluster nonstationary series from the UEA multivariate time series classification archive.

1. Introduction

Time series clustering is a central problem in data mining with applications in many fields. The objective is to split a large set of unlabelled time series realizations into homogeneous groups so that similar series are placed together in the same group and dissimilar series are located in different groups. This unsupervised classification process is useful to characterize different dynamic patterns without the need to analyse and model each single time series, which is computationally intensive and often far from being the real target. Many methods to cluster time series have been proposed in the literature. Comprehensive overviews including current advances, future prospects, significant references and specific application areas are provided by Aghabozorgi, Shirkhorshidi, and Wah (2015), Fu (2011), Liao (2005) and Rani, and Sikka (2012), and more recently in the monograph by Maharaj, D'Urso, and Caiado (2019). However, most of the proposed approaches concern univariate time series (UTS) while clustering of multivariate time series (MTS) has received much less attention. Unlike UTS, MTS involve a number of variables which must be jointly considered to characterize the underlying dynamic pattern. From the clustering point of view, this is a challenging issue because a dissimilarity measure between MTS should take into account the interdependence relationship between variables. For example, the cross-correlation between some specific variables might be high in some clusters but non-significant in others. In addition, MTS are two-dimensional objects, which increases the computational complexity, making inefficient or even infeasible some of the clustering procedures proposed to deal with UTS. In short, high dimensionality and complexity to assess dissimilarity make particularly challenging the MTS clustering task.

A key point in cluster analysis is to establish the dissimilarity notion since different dissimilarity criteria can lead to different groupings. A proper dissimilarity measure must be based on the nature and specific purpose of the clustering task, thus allowing to interpret the clustering solution in terms of the grouping target. If the target is to discriminate between geometric profiles of the time series, then a shape-based dissimilarity criterion is suitable. In contrast, a structure-based dissimilarity is desirable if the intention is to compare underlying

[*] Corresponding author.
E-mail addresses: a.oriona@udc.es, oriona38@hotmail.com (Á. López-Oriona), jose.vilarf@udc.es (J.A. Vilar).

https://doi.org/10.1016/j.eswa.2021.115677
Received 4 November 2020; Received in revised form 15 July 2021; Accepted 25 July 2021
Available online 29 July 2021

dependence models. In the latter case, the clustering performance may be seriously affected by noise, heteroskedasticity, heavy tails…and hence dissimilarity measures capable of capturing high level dynamic structures are particularly helpful. Many criteria to assess dissimilarity between UTS are available in the literature, including measures based on raw data, extracted features, generating models, complexity levels, and forecast performances, among others. A survey of measures can be seen in Montero, and Vilar (2014) and many of them are implemented in the R package **TSclust** (Montero, & Vilar, 2014b).

The focus of this paper is on clustering of MTS according to the underlying dependence structures, i.e., on identifying groups of MTS generated by the same multivariate stochastic process. This problem often arises when dealing with sets of MTS consisting of environmental, financial, EEG and fMRI data, which may exhibit complex serial dependence structures. Due to the high dimensionality of MTS, most of the proposed dissimilarity criteria combine a previous step regarding dimensionality reduction methods with the application of a feature-based dissimilarity into the new coordinate space, with lower dimensionality. Principal component analysis (PCA) is one of the most widely used reduction techniques and metrics such as Euclidean distance, dynamic time warping (DTW) distance, or an extended Frobenius norm (so-called *Eros*) have been employed in different studies to measure dissimilarity between principal components. Approaches considering PCA and some variants such as weighted principal component analysis (WPCA), two-dimensional principal component analysis (2dPCA), common principal component analysis (CPCA) and variable-based principal component analysis (VPCA) were proposed by He, and Tan (2020), Karamitopoulos, Evangelidis, and Dervos (2010), Li (2016, 2019), Singhal, and Seborg (2005) and Yang, and Shahabi (2004), among others. However, dimensionality reduction may result in a loss of information on the structural relationships of the MTS objects, thus hindering the detection of the underlying cluster structure. In fact, our numerical experiments have shown that some of these procedures behave poorly in clustering of stationary MTS models with different cross-correlation between their components (see Section 3). Wang, Wirth, and Wang (2007) directly consider a feature-based approach where each MTS is replaced by a vector of statistical features extracted from its univariate components and then a standard K-means clustering is performed. By construction, this method ignores the interdependence structure of the MTS data.

A more reduced number of procedures addressed the evaluation of MTS dissimilarity without dimensionality reduction and taking into account the dependence relationship between components. In an early work, Kakizawa, Shumway, and Taniguchi (1998) considered the Kullback–Leibler and Chernoff information measures as particular cases of a general disparity measure between sample spectral matrices, showing its robustness to departures from Gaussianity. In the time domain, Maharaj (1999) proposed a model-based measure designed to compare vector autoregressive parameter estimates of the series. For each pair of series, the fitted parameter vectors are used to check the equality of generating models and then a clustering algorithm based on the p-values is carried out. A procedure based on finite mixture models where the group-specific model parameters are simultaneously estimated by using Bayesian Markov chain Monte Carlo simulation methods was proposed by Fröhwirth-Schnatter, and Kaufmann (2008). In D'Urso, and Maharaj (2012), the dissimilarity between a pair of MTS is evaluated in terms of the Euclidean distance between their corresponding representations by uni and multidimensional wavelet features. Specifically, the maximum overlap discrete wavelet transform of each MTS component is obtained on a number of scales and then the wavelet variances of all components and the wavelet correlations of all pairs of components are concatenated to construct a vector representing the MTS. Later, D'Urso, De Giovanni, Maharaj, and Massari (2014) introduced a weighted version of this wavelet-based dissimilarity where the weights are computed by training self-organizing maps. The wavelet-based approach is non-parametric, no model assumptions are required and the relationship between components is taken into account.

Apart from the model-based approaches, which require modelling the time series and rely obviously on prior model requirements, the rest of procedures considering the dependence structure of MTS are connected to the analysis of means, correlations and cross-correlations but they ignore the serial dependence. Indeed, these clustering procedures yield good results as the series are jointly Gaussian, but they are expected to behave poorly when the MTS under consideration exhibit serial features related to the joint distribution of their components, such as changes in the conditional shape, heavy tails and dependence in the extremes (Dette, Hallin, Kley, & Volgushev, 2015; Hagemann, 2013; Kley, Volgushev, Dette, & Hallin, 2016). For UTS clustering, this fact was highlighted in Lafuente-Rego, and Vilar (2016), where a dissimilarity measure comparing estimated quantile autocovariance functions (QAF) was introduced. Quantile autocovariances provide information about the serial dependence structure at different pairs of quantile levels allowing to identify dependence features that covariance-based methods are unable to detect. The QAF metric was used with the partitioning around medoids (PAM) algorithm (Kaufman, & Rousseeuw, 2009) and an extensive simulation study showed its high capability to cluster UTS generated from a broad range of dependence models, particularly to discriminate between conditionally heteroskedastic models, and its robustness to the distributional form of the errors. In a further work, Vilar, Lafuente-Rego, and D'Urso (2018) obtained asymptotic properties and developed a novel fuzzy C-medoids approach based on the QAF metric.

Motivated by the good behaviour of the QAF metric in UTS clustering, the aim of this paper is to extend this principle to MTS clustering by introducing a metric addressing jointly both cross dependence and serial dependence. In this case, we propose to work in the frequency domain. A small set of quantile levels is fixed and every MTS is characterized by estimates of the quantile cross-spectral densities associated with each pair of components and each pair of quantile levels (see e.g., Baruník, & Kley, 2019). Real and imaginary parts of all these estimates are concatenated in a vector and the Euclidean distance between any two vectors provides the dissimilarity between the corresponding MTS. Since the quantile-based cross-spectra capture dependence in arbitrary quantiles of the joint distribution, the proposed metric is able to discriminate between any type of dependence structures. The proposed approach has several advantages over its counterpart in the time domain. For instance, the former only requires specification of the quantile levels whereas the latter also need to set the lags to be considered. Additionally, from a theoretical point of view, consistency and asymptotic behaviour of the utilized estimates of the quantile cross-spectral density have been properly established in the literature under specific conditions on the underlying multivariate process (Baruník & Kley, 2019).

Our experimental analyses show that the proposed metric produces excellent results in clustering of MTS using the PAM algorithm. Compared to other alternative dissimilarities, our metric is clearly more effective in scenarios involving complex dependence models and highly competitive in conventional setups where the cross-spectral density fully characterizes the underlying dependence. Our approach also exhibits robustness to the presence of heavy-tailed noise distributions and high computational efficiency.

The rest of the paper is organized as follows. The dissimilarity measure between MTS based on estimated quantile cross-spectral densities is presented in Section 2. First, a precise definition of the quantile cross-spectral density is provided, the estimation procedure is detailed and the capability of the dissimilarity to discriminate between multivariate processes is motivated by means of a simple example. Its behaviour in MTS clustering is analysed in Section 3 throughout an extensive simulation study where different clustering scenarios featured by the kind of generating processes are considered. Effects of different distributional forms for the errors and different lengths of the series are also examined and the results are compared with the ones obtained using other dissimilarity measures. A time consumption comparison between the analysed approaches is shown in Section 4. In Section 5, we apply the proposed tool to cluster real bivariate time series belonging to the field of Finance. Furthermore, the new clustering algorithm is tested on some additional sets of time series from various fields extracted from the UEA multivariate time series classification archive in Section 6. Some concluding remarks and future work are given in Section 7. All the simulation experiments and the real data analysis have been carried out using self-programmed code implemented in the R language (R Core Team, 2020), which is available upon request.

2. A novel structure-based approach for multivariate time series clustering

Consider a set of s multivariate time series $\mathcal{S} = \left\{\boldsymbol{X}_t^{(1)}, \dots, \boldsymbol{X}_t^{(s)}\right\}$, where the j-th element $\boldsymbol{X}_t^{(j)} = \left\{\boldsymbol{X}_1^{(j)}, \dots, \boldsymbol{X}_{T_j}^{(j)}\right\}$ is a T_j-length partial realization from any d-variate real-valued strictly stationary stochastic process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$. We wish to perform clustering on the elements of $\mathcal{S}$ in such a way that the series generated from the same stochastic process are grouped together. We propose to use a partitional algorithm starting from a pairwise dissimilarity matrix based on comparing estimated quantile cross-spectral densities. In this section, the quantile cross-spectral density notion is presented and then used to define a distance between MTS.

2.1. The quantile cross-spectral density

Let $\{\boldsymbol{X}_t, t \in \mathbb{Z}\} = \{(X_{t,1}, \dots, X_{t,d}), t \in \mathbb{Z}\}$ be a d-variate real-valued strictly stationary stochastic process. Denote by F_j the marginal distribution function of $X_{t,j}$, $j = 1, \dots, d$, and by $q_j(\tau) = F_j^{-1}(\tau)$, $\tau \in [0,1]$, the corresponding quantile function. Fixed $l \in \mathbb{Z}$ and an arbitrary couple of quantile levels $(\tau, \tau') \in [0,1]^2$, consider the cross-covariance of the indicator functions $I\left\{X_{t,j_1} \leq q_{j_1}(\tau)\right\}$ and $I\left\{X_{t+l,j_2} \leq q_{j_2}(\tau')\right\}$ given by

$$\gamma_{j_1,j_2}(l,\tau,\tau') = \mathrm{Cov}\left(I\left\{X_{t,j_1} \leq q_{j_1}(\tau)\right\}, I\left\{X_{t+l,j_2} \leq q_{j_2}(\tau')\right\}\right), \tag{1}$$

for $1 \leq j_1, j_2 \leq d$. Taking $j_1 = j_2 = j$, the function $\gamma_{j,j}(l,\tau,\tau')$, with $(\tau,\tau') \in [0,1]^2$, so-called QAF of lag l, generalizes the traditional autocovariance function. While autocovariances measure linear dependence between different lags evaluating covariability with respect to the average, quantile autocovariances examine how a part of the range of variation of X_j helps to predict whether the series will be below quantiles in a future time. This way, QAF entirely describes the dependence structure of $(X_{t,j}, X_{t+l,j})$, enabling us to capture serial features that standard autocovariances cannot detect. Note that $\gamma_{j_1,j_2}(l,\tau,\tau')$ always exists since no assumptions about moments are required. Furthermore, QAF also takes advantage of the local distributional properties inherent to the quantile methods, including robustness against heavy tails, dependence in the extremes and changes in the conditional shapes (skewness, kurtosis). Motivated by these nice properties, a dissimilarity between UTS based on comparing estimated quantile autocovariances over a common range of quantiles was proposed by Lafuente-Rego and Vilar (2016) to perform UTS clustering with very satisfactory results.

In the case of the multivariate process $\{\boldsymbol{X}_t, t \in \mathbb{Z}\}$, we can consider the $d \times d$ matrix

$$\boldsymbol{\Gamma}(l,\tau,\tau') = \left(\gamma_{j_1,j_2}(l,\tau,\tau')\right)_{1\leq j_1,j_2\leq d}, \tag{2}$$

which jointly provides information about both the cross-dependence (when $j_1 \neq j_2$) and the serial dependence (because the lag l is considered). To obtain a much richer picture of the underlying dependence structure, $\boldsymbol{\Gamma}(l,\tau,\tau')$ can be computed over a range of prefixed values of L lags, $\mathcal{L} = \{l_1, \dots, l_L\}$, and r quantile levels, $\mathcal{T} = \{\tau_1, \dots, \tau_r\}$, thus having available the set of matrices

$$\boldsymbol{\Gamma}_{\boldsymbol{X}_t}(\mathcal{L},\mathcal{T}) = \left\{\boldsymbol{\Gamma}(l,\tau,\tau'),\ l \in \mathcal{L},\ \tau,\tau' \in \mathcal{T}\right\}. \tag{3}$$

In the same way as the spectral density is the representation in the frequency domain of the autocovariance function, the spectral counterpart for the cross-covariances $\gamma_{j_1,j_2}(l,\tau,\tau')$ can be introduced. Under suitable summability conditions (mixing conditions), the Fourier transform of the cross-covariances is well-defined and the *quantile cross-spectral density* is given by

$$\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau') = (1/2\pi)\sum_{l=-\infty}^{\infty}\gamma_{j_1,j_2}(l,\tau,\tau')e^{-il\omega}, \tag{4}$$

for $1 \leq j_1, j_2 \leq d$, $\omega \in \mathbb{R}$ and $\tau,\tau' \in [0,1]$. Note that $\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau')$ is complex-valued so that it can be represented in terms of its real and imaginary parts, which will be denoted by $\Re(\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau'))$ and $\Im(\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau'))$, respectively. The quantity $\Re(\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau'))$ is known as quantile cospectrum of $(X_{t,j_1})_{t\in\mathbb{Z}}$ and $(X_{t,j_2})_{t\in\mathbb{Z}}$, whereas the quantity -$\Im(\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau'))$ is called quantile quadrature spectrum of $(X_{t,j_1})_{t\in\mathbb{Z}}$ and $(X_{t,j_2})_{t\in\mathbb{Z}}$.

For fixed quantile levels (τ,τ'), the quantile cross-spectral density is the cross-spectral density of the bivariate process

$$(I\{X_{t,j_1} \leq q_{j_1}(\tau)\}, I\{X_{t,j_2} \leq q_{j_2}(\tau')\}). \tag{5}$$

Therefore, the quantile cross-spectral density measures dependence between two components of the multivariate process in different ranges of their joint distribution and across frequencies. Proceeding as in (3), the quantile cross-spectral density can be evaluated on a range of frequencies Ω and of quantile levels $\mathcal{T}$ for every couple of components in order to obtain a complete representation of the process, i.e., consider the set of matrices

$$\mathfrak{f}_{\boldsymbol{X}_t}(\Omega,\mathcal{T}) = \left\{\mathfrak{f}(\omega,\tau,\tau'),\ \omega \in \Omega,\ \tau,\tau' \in \mathcal{T}\right\}, \tag{6}$$

where $\mathfrak{f}(\omega,\tau,\tau')$ denotes the $d \times d$ matrix in $\mathbb{C}$

$$\mathfrak{f}(\omega,\tau,\tau') = \left(\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau')\right)_{1\leq j_1,j_2\leq d}. \tag{7}$$

Representing $\boldsymbol{X}_t$ through $\mathfrak{f}_{\boldsymbol{X}_t}$, complete information on the general dependence structure of the process is available. Comprehensive discussions about the nice properties of the quantile cross-spectral density are given in Baruník and Kley (2019), Dette et al. (2015) and Lee, and Rao (2012), including invariance to monotone transformations, robustness and capability to detect nonlinear dependence. It is also worth enhancing that the quantile cross-spectral density provides a full description of all copulas of pairs of components in $\boldsymbol{X}_t$, since the difference between the copula of an arbitrary couple (X_{t,j_1}, X_{t+l,j_2}) evaluated in (τ,τ') and the independence copula at (τ,τ') can be written as

$$\mathbb{P}\left(X_{t,j_1} \leq q_{j_1}(\tau), X_{t+l,j_2} \leq q_{j_2}(\tau')\right) - \tau\tau' = \int_{-\pi}^{\pi}\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau')e^{il\omega}\,d\omega. \tag{8}$$

According to the prior arguments, a dissimilarity measure between realizations of two multivariate processes $\{\boldsymbol{X}_t, t \in \mathbb{Z}\}$ and $\{\boldsymbol{Y}_t, t \in \mathbb{Z}\}$, could be established by comparing their representations in terms of the quantile cross-spectral density matrices, $\mathfrak{f}_{\boldsymbol{X}_t}$ and $\mathfrak{f}_{\boldsymbol{Y}_t}$, respectively. For it, estimates of the quantile cross-spectral densities must be obtained.

Let $\left\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_T\right\}$ be a realization from the process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ so that $\boldsymbol{X}_t = (X_{t,1}, \dots, X_{t,d})$, $t = 1, \dots, T$. For arbitrary $j_1, j_2 \in \{1, \dots, d\}$ and $(\tau,\tau') \in [0,1]^2$, Baruník and Kley (2019) propose to estimate $\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau')$ considering a smoother of the cross-periodograms based on the indicator functions $I\{\hat{F}_{T,j}(X_{t,j})\}$, where $\hat{F}_{T,j}(x) = T^{-1}\sum_{t=1}^{T} I\{X_{t,j} \leq x\}$ denotes the empirical distribution function of $X_{t,j}$. This approach extends to the multivariate case the estimator proposed by Kley et al. (2016) in the univariate setting. More specifically, the called *rank-based copula cross-periodogram* (CCR-periodogram) is defined by

$$I_{T,R}^{j_1,j_2}(\omega,\tau,\tau') = \frac{1}{2\pi T}d_{T,R}^{j_1}(\omega,\tau)d_{T,R}^{j_2}(-\omega,\tau'), \tag{9}$$

where

$$d^{j}_{T,R}(\omega,\tau) = \sum_{t=1}^{T} I\{\hat{F}_{T,j}(X_{t,j}) \leq \tau\} e^{-i\omega t}.$$

The asymptotic properties of the CCR-periodogram are established in Proposition S4.1 of Baruník and Kley (2019). Likewise the standard cross-periodogram, the CCR-periodogram is not a consistent estimator of $\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau')$ (Baruník & Kley, 2019). To achieve consistency, the CCR-periodogram ordinates (evaluated on the Fourier frequencies) are convolved with weighting functions $W_T(\cdot)$. The *smoothed CCR-periodogram* takes the form

$$\hat{G}^{j_1,j_2}_{T,R}(\omega,\tau,\tau') = (2\pi/T)\sum_{s=1}^{T-1} W_T\left(\omega - \frac{2\pi s}{T}\right) I^{j_1,j_2}_{T,R}\left(\frac{2\pi s}{T},\tau,\tau'\right), \tag{10}$$

where

$$W_T(u) = \sum_{v=-\infty}^{\infty} (1/h_T) W\left(\frac{u+2\pi v}{h_T}\right),$$

with $h_T > 0$ a sequence of bandwidths such that $h_T \to 0$ and $Th_T \to \infty$ as $T \to \infty$, and W is a real-valued, even weight function with support $[-\pi,\pi]$. Consistency and asymptotic performance of the smoothed CCR-periodogram $\hat{G}^{j_1,j_2}_{T,R}(\omega,\tau,\tau')$ are established in Theorem S4.1 of Baruník and Kley (2019).

This way, the set of complex-valued matrices $\mathfrak{f}_{\boldsymbol{X}_t}(\Omega,\mathcal{T})$ in (6) characterizing the underlying process can be estimated by

$$\hat{\mathfrak{f}}_{\boldsymbol{X}_t}(\Omega,\mathcal{T}) = \left\{\hat{\mathfrak{f}}(\omega,\tau,\tau'),\ \omega \in \Omega,\ \tau,\tau' \in \mathcal{T}\right\}, \tag{11}$$

where $\hat{\mathfrak{f}}(\omega,\tau,\tau')$ is the matrix

$$\hat{\mathfrak{f}}(\omega,\tau,\tau') = \left(\hat{G}^{j_1,j_2}_{T,R}(\omega,\tau,\tau')\right)_{1\leq j_1,j_2 \leq d}. \tag{12}$$

Throughout this article, the smoothed CCR-periodograms were obtained by using the R-package **quantspec** (Kley, 2016).

2.2. Motivating example

In order to illustrate the high capability of the quantile cross-spectral density to distinguish between generating processes, we have considered realizations of three different bivariate processes:

P1

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix},$$

P2

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0.8 & 0.7 \\ -0.4 & 0.6 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix},$$

P3

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0 & 1.5(U_{t,1}-0.5) \\ 1.5(U_{t,2}-0.5) & 0 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} \Phi^{-1}(U_{t,1}) \\ \Phi^{-1}(U_{t,2}) \end{pmatrix}.$$

The process $(\epsilon_{t,1},\epsilon_{t,2})^\intercal$ in P1 and P2, denoting $\intercal$ the transpose vector, follows a multivariate standard normal distribution. With regards to P3, $U_{t,1}$ and $U_{t,2}$ are assumed to be independent and uniformly distributed on $[0,1]$, and Φ^{-1} stands for the quantile function of the standard normal distribution. Additionally, the error process in the three cases is an i.i.d. vector process. Processes P1, P2 and P3 are, respectively, a bivariate white noise process, a VAR(1) process, and a QVAR(1) process (Baruník & Kley, 2019). Note that in both P1 and P3 the univariate processes $(X_{t,1})_{t\in\mathbb{Z}}$ and $(X_{t,2})_{t\in\mathbb{Z}}$ are uncorrelated. However, whereas in process P1 they are independent, this is not the case for process P3.

After generating a large sample size ($T = 20000$) realization from each one of the processes above, we have depicted estimates of both the quantile cospectrum and the quantile quadrature spectrum of $(X_{t,1})_{t\in\mathbb{Z}}$ and $(X_{t,2})_{t\in\mathbb{Z}}$ as a function of ω, according to (9). The considered quantile levels were $\tau_1 = 0.5$ and $\tau_2 \in \{0.1, 0.5, 0.9\}$. The corresponding plots are given in Fig. 1.

It is clear from Fig. 1 that both the quantile cospectrum and the quantile quadrature spectrum show a different behaviour depending on the generating process. For instance, in the top panels ($\tau_1 = 0.5$ and $\tau_2 = 0.1$), the red lines, which stand for the white noise process P1, follow closely zero with deviations due to random variations, which is expected since P1 is formed by independent components. On the other hand, the blue and green lines show distinctive features for processes P2 and P3, exhibiting different levels of cross-dependence across the frequencies between the 0.5 and 0.1 quantiles of the joint distribution. Therefore, the top panels of Fig. 1 clearly reveal that each of the simulated time series comes from a different process, thus proving themselves as a useful graphical aid to distinguish between realizations of different generating processes.

A different phenomenon is observed for levels $\tau_1 = \tau_2 = 0.5$ (middle panels). Whereas the blue lines show the existence of a clear dependence structure between components of the VAR(1) process (P2), the green line representing the QVAR(1) process (P3) is now virtually indistinguishable from the red line because the two components in P3 are uncorrelated. Under the normality assumption, the use of $\tau_1 = \tau_2 = 0.5$ to detect dependence is equivalent to assess the correlation between $(X_{t,1})_{t\in\mathbb{Z}}$ and $(X_{t,2})_{t\in\mathbb{Z}}$. Thus, when both quantile levels are set to 0.5, the quantile cross-spectral-based quantities fail to detect the dependence relationship in P3, which remains completely hidden. Hence, the middle panels of Fig. 1 do not allow to deduce the existence of three different dependence structures. As in the top panels, the lines in both plots of the bottom panels exhibit fairly distinct behaviours, thus unravelling the existence of three different generating processes.

To highlight the huge discriminative power of the quantile cross-spectral density in comparison to its classical counterpart, estimates of traditional cospectrum and quadrature spectrum for processes P1, P2 and P3 are displayed in Fig. 2. The estimates were obtained by smoothing the classical periodogram with a series of modified Daniell smoothers (Bloomfield, 2004; Daniell, 1946; Priestley, 1981). The top and bottom panels, which correspond to the white noise process (P1) and the QVAR(1) process (P3), respectively, are almost identical except for random variations, as the traditional cross-spectral density between uncorrelated variables is zero across all frequencies. On the other hand, the middle panels of Fig. 2 illustrate that the estimates are able to differentiate the VAR(1) process (P2) from processes P1 and P3. Note that the situation in Fig. 2 is rather similar to that in the middle panels of Fig. 1, in the sense that one can only distinguish two classes of dependence structures. Hence, when $\tau_1 = \tau_2 = 0.5$, the quantile cross-spectral density resembles the traditional cross-spectral density, failing to tell apart different underlying processes in some situations as the one shown here. However, when one of the quantile levels is different from 0.5, the dependence scheme becomes visible.

In sum, we can conclude that, in a potential scenario with series coming from processes P1, P2 and P3, a clustering algorithm taking into account the quantities displayed in Fig. 1 would probably succeed, as the series coming from different generating processes are represented by means of clearly distinct features. On the contrary, the use of the quantities in Fig. 2 would likely prove useless, since series from processes P1 and P3 would be treated as identical. This toy example highlights the usefulness of the quantile cross-spectral density as a powerful tool to distinguish between unequal dependence structures under some circumstances in which the only use of the traditional cross-spectral density can lead to erroneous conclusions.

2.3. An innovative spectral dissimilarity measure between MTS

A simple dissimilarity criterion between a pair of d-variate time series $\boldsymbol{X}_t^{(1)}$ and $\boldsymbol{X}_t^{(2)}$ can be obtained by comparing their estimated sets of complex-valued matrices $\hat{\mathfrak{f}}_{\boldsymbol{X}_t^{(1)}}(\Omega,\mathcal{T})$ and $\hat{\mathfrak{f}}_{\boldsymbol{X}_t^{(2)}}(\Omega,\mathcal{T})$ evaluated on a common range of frequencies and quantile levels. Specifically,

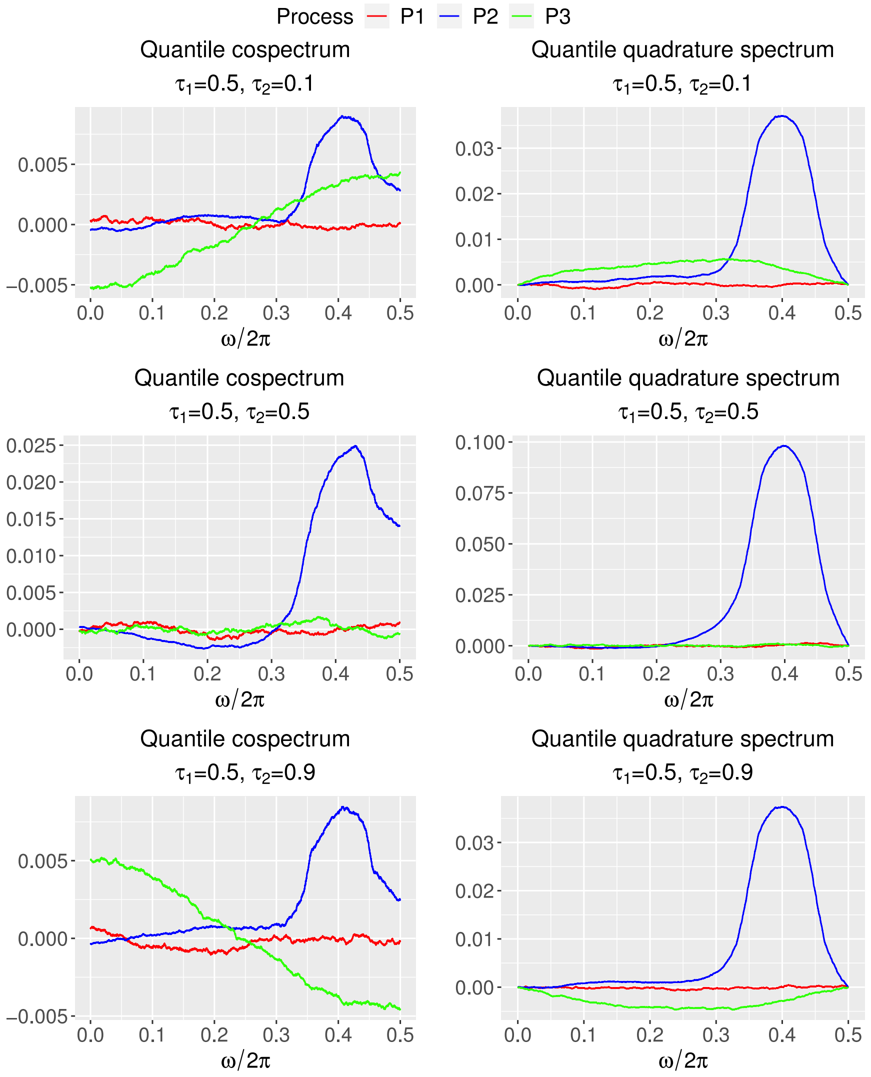

Fig. 1. Estimates of quantile cospectrum and quantile quadrature spectrum for large sample size realizations of processes P1, P2 and P3 according to different quantile levels.

each time series $\boldsymbol{X}_t^{(u)}$, $u = 1, 2$, is characterized by means of a set of d^2 vectors $\{\boldsymbol{\Psi}^{(u)}_{j_1,j_2}, 1 \leq j_1, j_2 \leq d\}$ constructed as follows. For a given set of K different frequencies $\Omega = \{\omega_1, \ldots, \omega_K\}$, and r quantile levels $\mathcal{T} = \{\tau_1, \ldots, \tau_r\}$, each vector $\boldsymbol{\Psi}^{(u)}_{j_1,j_2}$ is given by

$$\boldsymbol{\Psi}^{(u)}_{j_1,j_2} = (\boldsymbol{\Psi}^{(u)}_{1,j_1,j_2}, \ldots, \boldsymbol{\Psi}^{(u)}_{K,j_1,j_2}), \tag{13}$$

where each $\boldsymbol{\Psi}^{(u)}_{k,j_1,j_2}$, $k = 1, \ldots, K$ consists of a vector of length r^2 formed by rearranging by rows the elements of the matrix

$$\left(\hat{G}^{j_1,j_2}_{T,R}(\omega_k, \tau_i, \tau_{i'})\right)_{1 \leq i,i' \leq r}. \tag{14}$$

Once the set of d^2 vectors, $\{\boldsymbol{\Psi}^{(u)}_{j_1,j_2}, 1 \leq j_1, j_2 \leq d\}$, is obtained, its elements are all concatenated in a vector $\boldsymbol{\Psi}^{(u)}$ in the same way as vectors of the form $\boldsymbol{\Psi}^{(u)}_{k,j_1,j_2}$ constitute $\boldsymbol{\Psi}^{(u)}_{j_1,j_2}$ in (13). In this manner, the dissimilarity between $\boldsymbol{X}_t^{(1)}$ and $\boldsymbol{X}_t^{(2)}$ is obtained by means of the Euclidean distance between the complex vectors $\boldsymbol{\Psi}^{(1)}$ and $\boldsymbol{\Psi}^{(2)}$

$$\begin{aligned} d_{QCD}(\boldsymbol{X}_t^{(1)}, \boldsymbol{X}_t^{(2)}) &= \left[\|\Re_v(\boldsymbol{\Psi}^{(1)}) - \Re_v(\boldsymbol{\Psi}^{(2)})\|^2 + \|\Im_v(\boldsymbol{\Psi}^{(1)}) - \Im_v(\boldsymbol{\Psi}^{(2)})\|^2\right]^{1/2} = \\ &\left[\sum_{j_1=1}^{d}\sum_{j_2=1}^{d}\sum_{i=1}^{r}\sum_{i'=1}^{r}\sum_{k=1}^{K}\left(\Re\left(\hat{G}^{j_1,j_2}_{T,R}(\omega_k, \tau_i, \tau_{i'})^{(1)}\right)\right.\right. \\ &\left. - \Re\left(\hat{G}^{j_1,j_2}_{T,R}(\omega_k, \tau_i, \tau_{i'})^{(2)}\right)\right)^2 + \\ &\sum_{j_1=1}^{d}\sum_{j_2=1}^{d}\sum_{i=1}^{r}\sum_{i'=1}^{r}\sum_{k=1}^{K}\left(\Im\left(\hat{G}^{j_1,j_2}_{T,R}(\omega_k, \tau_i, \tau_{i'})^{(1)}\right)\right. \\ &\left.\left. - \Im\left(\hat{G}^{j_1,j_2}_{T,R}(\omega_k, \tau_i, \tau_{i'})^{(2)}\right)\right)^2\right]^{1/2}, \end{aligned} \tag{15}$$

where $\Re_v$ and $\Im_v$ denote the element-wise real and imaginary part operations, respectively, and $\hat{G}^{j_1,j_2}_{T,R}(\omega_k, \tau_i, \tau_{i'})^{(u)}$ is the corresponding smoothed CCR-periodogram for the series $\boldsymbol{X}_t^{(u)}$, $u = 1, 2$.

Computation of d_{QCD} for every couple of MTS subjected to cluster produces a pairwise dissimilarity matrix which can be used as input to the PAM algorithm. Of course, developing another partitional clustering method, for instance, using the K-means algorithm, is also possible by averaging the vectors $(\Re_v(\boldsymbol{\Psi}^{(u)}), \Im_v(\boldsymbol{\Psi}^{(u)}))$, $1 \leq u \leq s$ in order to compute the centroids. Then, the dissimilarity d_{QCD} would be used to obtain the required distances in the iterative process. However, for the sake of simplicity and interpretation, our analyses have been limited to PAM.

3. Experimental evaluation of the proposed clustering procedure

In this section we carry out a set of simulations with the aim of assessing the performance of d_{QCD} in different scenarios of MTS clustering. Firstly we describe the simulation mechanism, then we explain how the assessment of the proposed approach was done and finally we show the results of the simulation study.

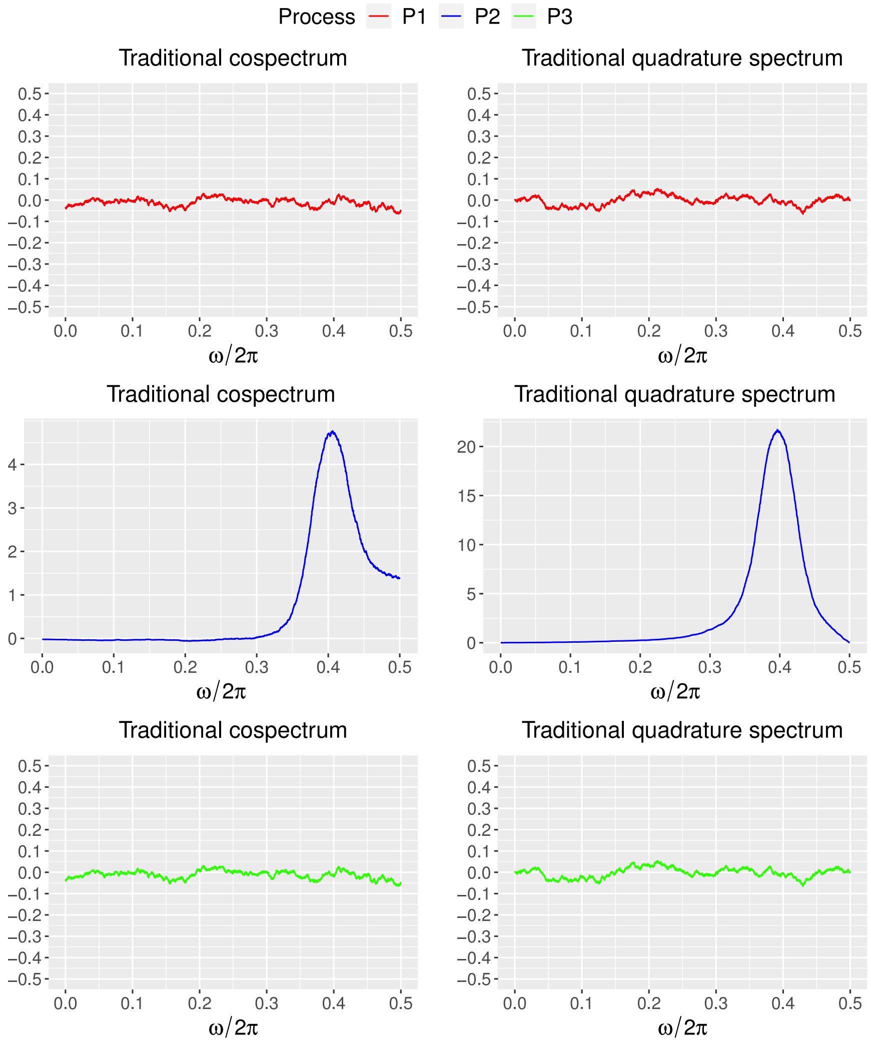

Fig. 2. Estimates of traditional cospectrum and traditional quadrature spectrum for large sample size realizations of processes P1, P2 and P3.

3.1. Experimental design

The simulated scenarios cover a wide variety of generating processes. Specifically, three unsupervised classification setups were considered, namely clustering of (1) VARMA processes, (2) dynamic conditional correlation processes, and (3) processes exhibiting different types of quantile dependence. The selection of such kind of processes was made with the goal of performing the assessment task in a fair and general manner. Indeed, the three chosen setups are pivotal in several application domains. The generating models concerning each class of processes are given below.

Scenario 1. VARMA processes clustering.

(a) VAR(1)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} = \begin{pmatrix} 0.6 & 0.5 & 0 \\ -0.4 & 0.5 & 0.3 \\ 0 & -0.5 & 0.7 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \\ X_{t-1,3} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \\ \epsilon_{t,3} \end{pmatrix},$$

(b) VAR(1)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} = \begin{pmatrix} 0.4 & 0.4 & 0 \\ -0.4 & 0.5 & 0.4 \\ 0 & -0.5 & 0.7 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \\ X_{t-1,3} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \\ \epsilon_{t,3} \end{pmatrix},$$

(c) VMA(1)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} = \begin{pmatrix} 0.6 & 0.5 & 0 \\ -0.4 & 0.5 & 0.3 \\ 0 & -0.5 & 0.7 \end{pmatrix} \begin{pmatrix} \epsilon_{t-1,1} \\ \epsilon_{t-1,2} \\ \epsilon_{t-1,3} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \\ \epsilon_{t,3} \end{pmatrix},$$

(d) VMA(1)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} = \begin{pmatrix} 0.4 & 0.4 & 0 \\ -0.4 & 0.5 & 0.4 \\ 0 & -0.5 & 0.7 \end{pmatrix} \begin{pmatrix} \epsilon_{t-1,1} \\ \epsilon_{t-1,2} \\ \epsilon_{t-1,3} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \\ \epsilon_{t,3} \end{pmatrix},$$

(e) VARMA(1,1)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} = \begin{pmatrix} 0.6 & 0.5 & 0 \\ -0.4 & 0.5 & 0.3 \\ 0 & -0.5 & 0.7 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \\ X_{t-1,3} \end{pmatrix} + \begin{pmatrix} 0.6 & 0.5 & 0 \\ -0.4 & 0.5 & 0.3 \\ 0 & -0.5 & 0.7 \end{pmatrix} \begin{pmatrix} \epsilon_{t-1,1} \\ \epsilon_{t-1,2} \\ \epsilon_{t-1,3} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \\ \epsilon_{t,3} \end{pmatrix},$$

where, in all cases, $(\epsilon_{t,1}, \epsilon_{t,2}, \epsilon_{t,3})^{\intercal}$ is an i.i.d. vector error process following the trivariate normal distribution with zero mean and covariance matrix equals the identity matrix.

Scenario 2. Dynamic conditional correlation processes clustering. Consider $(X_{t,1}, X_{t,2})^{\intercal} = (a_{t,1}, a_{t,2})^{\intercal} = (\sigma_{t,1}\epsilon_{t,1}, \sigma_{t,2}\epsilon_{t,2})^{\intercal}$. The data-generating process consists of two Gaussian GARCH models (Bollerslev, 1986), one which is highly persistent and the other which is not.

$$\sigma^2_{t,1} = 0.01 + 0.05a^2_{t-1,1} + 0.94\sigma^2_{t-1,1},$$

$$\sigma^2_{t,2} = 0.5 + 0.2a^2_{t-1,2} + 0.5\sigma^2_{t-1,2},$$

$$\begin{pmatrix}\epsilon_{t,1}\\ \epsilon_{t,2}\end{pmatrix} \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho_t\\ \rho_t & 1\end{pmatrix}\right].$$

The correlation between the standardized shocks, ρ_t, is given by the following expressions:
(a) Constant correlation

$$\rho_t = 0.5,$$

(b) Piecewise constant correlation

$$\rho_t = 0.7I_{\{t\leq(T/2)\}} - 0.9I_{\{t>(T/2)\}},$$

(c) Piecewise constant correlation

$$\rho_t = 0.9I_{\{t\leq(T/2)\}} - 0.7I_{\{t>(T/2)\}},$$

(d) Piecewise varying correlation

$$\rho_t = \frac{0.99}{\log(t+2)}I_{\{t\ \text{odd}\}} - \frac{0.99}{\log(t+2)}I_{\{t\ \text{even}\}},$$

where I stands for the indicator function.

Scenario 3. QVAR processes clustering. Consider $= (U_{t,1}, U_{t,2})^\intercal$ a sequence of independent random vectors with independent components $U_{t,k}$ which are uniformly distributed on $[0,1]$, and $\Phi^{-1}(u)$, $u \in (0,1)$, the quantile function of the standard normal distribution.
(a) QVAR(1)

$$\begin{pmatrix}X_{t,1}\\ X_{t,2}\end{pmatrix} = \begin{pmatrix}0 & -0.5(U_{t,1}-0.5)\\ -0.5(U_{t,2}-0.5) & 0\end{pmatrix}\begin{pmatrix}X_{t-1,1}\\ X_{t-1,2}\end{pmatrix} + \begin{pmatrix}\Phi^{-1}(U_{t,1})\\ \Phi^{-1}(U_{t,2})\end{pmatrix},$$

(b) QVAR(1)

$$\begin{pmatrix}X_{t,1}\\ X_{t,2}\end{pmatrix} = \begin{pmatrix}0 & 0.5(U_{t,1}-0.5)\\ 0.5(U_{t,2}-0.5) & 0\end{pmatrix}\begin{pmatrix}X_{t-1,1}\\ X_{t-1,2}\end{pmatrix} + \begin{pmatrix}\Phi^{-1}(U_{t,1})\\ \Phi^{-1}(U_{t,2})\end{pmatrix},$$

(c) QVAR(1)

$$\begin{pmatrix}X_{t,1}\\ X_{t,2}\end{pmatrix} = \begin{pmatrix}0 & 1.5(U_{t,1}-0.5)\\ 1.5(U_{t,2}-0.5) & 0\end{pmatrix}\begin{pmatrix}X_{t-1,1}\\ X_{t-1,2}\end{pmatrix} + \begin{pmatrix}\Phi^{-1}(U_{t,1})\\ \Phi^{-1}(U_{t,2})\end{pmatrix}.$$

Scenario 1 was partially considered before in Bandyopadhyay, Maulik, and Baragona (2010) with the purpose of checking whether or not their proposed clustering algorithm based on genetic multiobjective optimization is able to distinguish series generated from different VARMA processes. Here, our goal is pretty much the same, showing to what extent the dissimilarity measure based on the quantile cross-spectral density is capable of dealing with this kind of processes. Note that the choice of the coefficient matrices in Scenario 1 is driven by the requirements of stationarity. Scenario 2 was motivated by the landmark work of Engle (2002), where the dynamic conditional correlation models are introduced. A simulation study is performed there in order to check the quality of several correlation estimators. The models presented here are very similar to the models offered in that analysis, except for the fact that we have decided to include also negative correlation to cover a wider range of situations. Indeed, negative correlation arises in many real life scenarios in which the use of dynamic conditional correlation models is common. For instance, Andersson, Krylova, and Vähämaa (2008) analysed data from US, UK and Germany and stated that stock and bonds are either positive or negative correlated depending on the period. Scenario 3 is based on QVAR models introduced in Baruník and Kley (2019), which provide a natural way of generating rich dependence structures between two random variables and over different frequencies. QVAR models extend the quantile autoregression models defined in Koenker, and Xiao (2006), who showed their usefulness when dealing with series as unemployment rates or gasoline prices. This last scenario is aimed to demonstrate how the quantile cross-spectral density is capable of detecting some forms of dependence that remain otherwise invisible.

The simulation study was carried out as follows. For each scenario, five time series of length $T \in \{250, 1000\}$ were generated from each model in order to perform clustering twice, thus allowing to assess the effect of the series length. The distance d_{QCD} between each pair of MTS was calculated using $r = 3$ quantiles of levels 0.1, 0.5 and 0.9 along with the set of Fourier frequencies $\{\omega_k = 2\pi k/T, 0 \leq k \leq T/2\}$. The obtained pairwise dissimilarity matrix was then processed by the PAM algorithm to reach the clustering solution. The simulation procedure was repeated 100 times for each scenario and each value of T.

3.2. Alternative metrics and assessment criteria

To shed light on the performance of d_{QCD}, clustering solutions based on some state of the art approaches measuring dissimilarity between MTS were also obtained. The considered dissimilarities are summarized below.

- *Dynamic time warping-based distances*. Particularly, the two multivariate extensions of dynamic time warping discussed in Shokoohi-Yekta, Hu, Jin, Wang, and Keogh (2017). The "independent" warping version (d_{DTWI}) computes the classical dynamic time warping between each pair of univariate time series, whereas the "dependent" version (d_{DTWD}) forces all dimensions to warp identically, in a single warping matrix.
- *Model-based distance*. Specifically, the distance proposed by Maharaj (1999) assessing the difference between vector autoregressive parameter estimates of the series. The original approach implies three main steps. First, a finite order VAR model is fitted to each series via a given criterion. Second, a p-value is obtained for each pair of series, regarding the null hypothesis stating that there is no significant difference between both underlying generating processes. Finally, hierarchical clustering is applied to the set of MTS via the p-values, but only those series whose associated p-value is greater than some predetermined number (e.g 0.05 or 0.01) are grouped together. This implies that the number of clusters is determined by the outcome of the tests of hypotheses and therefore the desired number of groups cannot be set in advance. In this case, to make homogeneous comparisons, we simply used the Euclidean distance between the vectors of coefficient estimates of the VAR models (d_{VAR}) in the first step to construct the initial dissimilarity matrix. The number of considered lags was determined by the maximum fitted order amongst the MTS as given by the Akaike Information Criterion. This way, when computing the distance between two vectors of coefficients of unequal length, the shortest vector was padded with zeros until it reached the length of the longest vector.
- *PCA-based distance*. Singhal and Seborg (2005) consider a weighted similarity factor based on principal components and the angles between the principal components subspaces. Then, a dissimilarity measure (d_{PCA}) is obtained by subtracting the similarity factor from one. In order to perform PCA, we applied the singular value decomposition to the correlation matrices and considered a number of principal components, r, as the minimum value such that at least 95% of the variability of all MTS was explained by means of the first r principal components. This criterion led always to the retention of all principal components.
- *Wavelet-based distance*. D'Urso and Maharaj (2012) introduce a Euclidean distance between wavelet features of MTS, specifically, between estimates of wavelet variances and wavelet correlations (d_W). The estimates are obtained through the maximum overlap discrete wavelet transform, which requires choosing a wavelet filter of a given length and a number of scales. After performing some brief preliminary analyses, we reached the conclusion that the wavelet filter of length 4 of the Daubechies family, DB4,

Table 1
Summary of the dissimilarities used in the simulation study.

Dissimilarity	Principle
d_{QCD}	Quantile cross-spectral density
d_{DTWI}	Independent dynamic time warping
d_{DTWD}	Dependent dynamic time warping
d_{VAR}	Estimated VAR coefficients
d_{PCA}	PCA
d_W	Wavelets
d_{GCC}	Generalized cross-correlation
d_J	Nonparametric spectral distance

along with the maximum allowable number of scales (which depends on the series length), were the choices that led to the best average results in Scenarios 1, 2 and 3. Hence, they were the hyperparameters chosen for the simulation study.

– *Generalized cross-correlation-based distance*. Alonso, and Peña (2019) propose to measure similarity between two UTS, X_t and Y_t, via the generalized cross correlation ($GCC(X_t, Y_t)$), which compares the determinant of the correlation matrix until some lag r of the bivariate vector with those of the two univariate time series. Conceptually, $GCC(X_t, Y_t)$ evaluates the level of linear dependency among both series. Considering the sample correlation matrices, a matrix of pairwise distances of the form $1-\widehat{GCC}(X_t, Y_t)$ can be directly constructed from every couple of series subjected to the clustering procedure. We propose to extend their approach to a multivariate framework as follows. Each d-variate series is first represented through a matrix $\mathbf{X} = \left(X_{t,j}\right)_{1\leq t\leq T,\ 1\leq j\leq d}$, and then described by means of a vector of length $d(d-1)$ whose components are given by

$$1-\widehat{GCC}(X_{:,j_1}, X_{:,j_2}),\ j_1, j_2 \in \{1,\dots,d\},\ j_1 \neq j_2, \tag{16}$$

where $X_{:,i}$ is the i-th column of the matrix $\mathbf{X}$. Then we construct a distance matrix by considering the Euclidean distance between the vectors of features (d_{GCC}) given by (16). It is important to remark that the primary goal of Alonso and Peña (2019) is to cluster UTS by linear dependence, i.e., two series are grouped together if they present a high degree of dependence, but the purpose of our extension is different. In our case, the degree of dependence between each pair of UTS within the MTS is evaluated in order to characterize the MTS. For the simulation study, the hyperparameter r was set to $r = 1$, a reasonable choice since all models in Scenarios 1, 2 and 3 present one significant lag.

– *Nonparametric dissimilarity in the frequency domain* (Kakizawa et al., 1998). A dissimilarity between estimates of the spectral density matrices via the smoothed periodogram. The J-divergence was used to compute the distance between the estimated matrices (d_J).

A summary of the dissimilarities considered in the simulation study is given in Table 1.

The quality of the clustering procedure was assessed by comparing the clustering partition given by the algorithm, P_k, with the true cluster solution, which is usually referred to as ground truth, G_k. The ground truth consisted of $k = 5$ groups in Scenario 1, $k = 4$ groups in Scenario 2, and $k = 3$ groups in Scenario 3, each one of them involving five time series with the same generating process. The value of k was given as an input parameter to the PAM algorithm. Partitions P_k and G_k were then compared by using three well-known external clustering validity indexes: the Larsen–Aone index (LA) (Larsen, & Aone, 1999), the adjusted Rand index (ARI) (Hubert, & Arabie, 1985), and the Jaccard index (JI). It is worth remarking that the expected value of ARI is zero for two partitions picked at random according to the generalized hypergeometric distribution. Hence, the value of zero can be associated with a noninformative clustering solution. Additionally, we have also computed a fourth index by considering the 1-nearest neighbour classifier evaluated by leave-one-out cross-validation (LOO1NN). Specifically, LOO1NN index returns the proportion of series which, according to G_k, are in the same cluster that their nearest series, based on the given dissimilarity measure. Notice that LOO1NN does not evaluate the clustering algorithm, but gives insights into the quality of the dissimilarity measure. This evaluation criterion has been extensively used in a broad range of pattern recognition applications, including time series clustering (see e.g., Keogh, & Kasetty, 2003 and Lafuente-Rego & Vilar, 2016). The indexes LA, JI and LOO1NN take values between 0 and 1. As for ARI index, it takes values between -1 and 1. In all cases, the closer to one the index, the better the clustering solution.

3.3. Results and discussion

Averages and standard deviations of the quality indexes over the 100 trials for the best performing metrics, namely d_{QCD}, d_W, d_{GCC}, d_J, d_{VAR} and d_{PCA}, are given in Tables 2 ($T = 250$) and 3 ($T = 1000$). The results concerning d_{DTWI} and d_{DTWD} for Scenario 1 and $T = 250$ are separately shown in Table 4.

One can work out from Table 4 that the dynamic time warping-based measures showed a very similar performance and, most importantly, they were not able to distinguish between the different processes appropriately. By comparing the results of Table 4 with those of Table 2 for Scenario 1, it is easy to see that the dissimilarity measures in Table 2 outperformed d_{DTWI} and d_{DTWD} by far. Regarding the ARI, both d_{DTWI} and d_{DTWD} got average values close to zero, 0.046 and 0.027, respectively, whereas the worst dissimilarity in Table 2 for Scenario 1, d_{PCA}, led to an average value of 0.290. Something similar occurred when d_{DTWI} and d_{DTWD} dealt with series from the remaining settings. Consequently, we have decided to omit the corresponding results. The poor performance of the dynamic time warping-based distances was expected since they are aimed to compare shape patterns being dominated by local comparisons. In contrast, our simulation scenarios are characterized by different latent dependence structures which describe the global behaviour of the series.

According to results in Table 2, the dissimilarity based on the quantile cross-spectral density d_{QCD} produced the highest average scores in Scenarios 2 and 3, and presented worse behaviour in Scenario 1. With VARMA models, d_{QCD} outperformed the wavelet-based metric d_W and the cross-correlation-based metric d_{GCC}, although presenting clustering quality indexes lower than the ones reached by the dissimilarities d_J and d_{VAR}. The metric d_{QCD} showed the best results in Scenario 3, where it indisputably defeated the remaining measures. By means of the quantile cross-spectral density, the dissimilarity d_{QCD} was able to unmask complex types of dependence as those generated by QVAR models, which remained completely hidden when other dissimilarities were considered.

As expected, the metric based on VAR models d_{VAR} was affected by model misspecification and hence it performed fine in Scenario 1 but produced unsatisfactory results in Scenarios 2 and 3. In fact, according to ARI, d_{VAR} did not a better job than picking a clustering solution at random in the latter scenarios. The nonparametric dissimilarity d_J achieved the best results in Scenario 1, even better than the ones reached by d_{VAR}, but its results worsened substantially when dealing with heteroskedastic models. Despite its poor results in Scenario 3, it was the only measure other than d_{QCD} that seemed to detect some clustering structure in the data in that complex scenario.

With regards to dissimilarities d_W and d_{GCC}, they achieved in Scenarios 1 and 2 results lower than but close to the ones reached by d_{QCD}. It is worth mentioning here than even though the dissimilarity d_{GCC} is based on cross-correlations, which are clearly different between the generated groups of series in Scenarios 1 and 2, its results are still worse than the ones achieved by both d_{QCD} and d_W, indicating that even in

Table 2
Averages and standard deviations (in brackets) of four clustering validity indexes for measures d_{QCD}, d_W, d_{GCC}, d_J, d_{VAR} and d_{PCA}, according to the 100 trials of the simulation procedure. For each scenario and index, the best result is shown in bold. The length of each series was $T = 250$.

	Index	d_{QCD}	d_W	d_{GCC}	d_J	d_{VAR}	d_{PCA}
Scenario 1	ARI	0.592	0.507	0.588	**0.801**	0.730	0.290
		(0.087)	(0.110)	(0.076)	(0.122)	(0.132)	(0.132)
	LA	0.777	0.739	0.777	**0.902**	0.856	0.622
		(0.060)	(0.069)	(0.053)	(0.069)	(0.081)	(0.083)
	LOO1NN	0.746	0.707	0.711	**0.914**	0.901	0.534
		(0.094)	(0.102)	(0.105)	(0.067)	(0.062)	(0.132)
	JI	0.504	0.432	0.502	**0.732**	0.653	0.278
		(0.080)	(0.092)	(0.072)	(0.151)	(0.149)	(0.089)
Scenario 2	ARI	**0.384**	0.303	0.285	0.256	−0.006	0.260
		(0.121)	(0.090)	(0.082)	(0.106)	(0.061)	(0.112)
	LA	**0.700**	0.634	0.627	0.612	0.443	0.613
		(0.081)	(0.061)	(0.053)	(0.069)	(0.054)	(0.079)
	LOO1NN	**0.656**	0.519	0.489	0.516	0.226	0.447
		(0.109)	(0.105)	(0.129)	(0.126)	(0.114)	(0.141)
	JI	**0.364**	0.306	0.307	0.281	0.144	0.291
		(0.013)	(0.059)	(0.049)	(0.067)	(0.033)	(0.078)
Scenario 3	ARI	**0.436**	0.024	0.000	0.154	0.016	0.006
		(0.172)	(0.112)	(0.077)	(0.131)	(0.087)	(0.098)
	LA	**0.746**	0.534	0.516	0.615	0.524	0.519
		(0.092)	(0.076)	(0.059)	(0.081)	(0.065)	(0.067)
	LOO1NN	**0.719**	0.370	0.294	0.467	0.302	0.295
		(0.116)	(0.150)	(0.150)	(0.140)	(0.144)	(0.148)
	JI	**0.453**	0.203	0.192	0.275	0.213	0.193
		(0.131)	(0.059)	(0.043)	(0.068)	(0.048)	(0.055)

Table 3
Averages and standard deviations (in brackets) of four clustering validity indexes for measures d_{QCD}, d_W, d_{GCC}, d_J, d_{VAR} and d_{PCA}, according to the 100 trials of the simulation procedure. For each scenario and index, the best result is shown in bold. The length of each series was $T = 1000$.

	Index	d_{QCD}	d_W	d_{GCC}	d_J	d_{VAR}	d_{PCA}
Scenario 1	ARI	0.778	0.666	0.693	**0.994**	0.894	0.612
		(0.096)	(0.103)	(0.069)	(0.030)	(0.133)	(0.174)
	LA	0.886	0.832	0.833	**0.998**	0.943	0.812
		(0.057)	(0.058)	(0.042)	(0.013)	(0.072)	(0.099)
	LOO1NN	0.887	0.826	0.848	**0.999**	0.983	0.810
		(0.064)	(0.075)	(0.083)	(0.013)	(0.025)	(0.116)
	JI	0.701	0.577	0.606	**0.991**	0.860	0.536
		(0.121)	(0.107)	(0.078)	(0.013)	(0.176)	(0.165)
Scenario 2	ARI	**0.900**	0.400	0.339	0.340	0.013	0.372
		(0.013)	(0.113)	(0.106)	(0.108)	(0.067)	(0.157)
	LA	**0.960**	0.698	0.649	0.661	0.458	0.682
		(0.013)	(0.078)	(0.062)	(0.070)	(0.053)	(0.097)
	LOO1NN	**0.979**	0.589	0.568	0.617	0.218	0.604
		(0.013)	(0.109)	(0.130)	(0.118)	(0.109)	(0.147)
	JI	**0.866**	0.374	0.344	0.336	0.152	0.368
		(0.013)	(0.085)	(0.065)	(0.072)	(0.035)	(0.113)
Scenario 3	ARI	**0.893**	0.046	0.003	0.401	−0.009	0.009
		(0.137)	(0.123)	(0.084)	(0.133)	(0.077)	(0.100)
	LA	**0.963**	0.547	0.522	0.722	0.509	0.523
		(0.013)	(0.075)	(0.065)	(0.066)	(0.057)	(0.067)
	LOO1NN	**0.973**	0.415	0.328	0.587	0.297	0.299
		(0.043)	(0.138)	(0.144)	(0.137)	(0.134)	(0.166)
	JI	**0.874**	0.213	0.193	0.425	0.195	0.192
		(0.158)	(0.067)	(0.048)	(0.088)	(0.043)	(0.054)

an "ideal situation" like this, the use of standard cross-correlations to perform clustering could not be the best choice.

Table 3 shows the results for $T = 1000$. With regards to VARMA processes, the results are pretty much the same as in Table 2. The nonparametric dissimilarity d_J achieved almost perfect results, significantly better than the ones reached by the model-based dissimilarity d_{VAR}. The dissimilarity d_{QCD} got again acceptable results. The findings concerning Scenario 2 are quite impressive. Unlike in Table 2, where d_{QCD} slightly outperformed most of the measures, here d_{QCD} showed a noteworthy superiority over the remaining dissimilarities, including d_J. Whereas increasing the length of the series from 250 to 1000 had little to moderate effect in their quality indexes, there was a different story for d_{QCD}, whose quality indexes skyrocketed. This is worth noting because it makes d_{QCD} the most versatile measure, capable of dealing either with VARMA or with conditional heteroskedastic processes. Regarding Scenario 3, only d_{QCD}, with high scores, and d_J, with moderate scores, were able to detect the underlying structure in the data.

It is important to mention that we have redone the simulations in Scenario 1 by considering correlated error terms, i.e., a nondiagonal covariance matrix for $(\epsilon_{t,1}, \epsilon_{t,2}, \epsilon_{t,3})'$. Our aim was to see whether the performance of the analysed dissimilarities, at least in relation to each other, was the same as under the assumption of independent error terms. Indeed, the results, which are available upon request, were very similar to the ones in Tables 2 and 3.

Table 4
Averages and standard deviations (in brackets) of four clustering validity indexes for measures d_{DTWI}, d_{DTWD} in Scenario 1, according to the 100 trials of the simulation procedure. The length of each series was $T = 250$.

Scenario 1	Index	d_{DTWI}	d_{DTWD}
	ARI	0.046	0.027
		(0.013)	(0.015)
	LA	0.372	0.367
		(0.004)	(0.004)
	LOO1NN	0.210	0.206
		(0.018)	(0.015)
	JI	0.179	0.169
		(0.007)	(0.008)

In order to gain illustrative insights into the previous remarks, Fig. 3 displays the boxplots based on the clustering validity indexes from the 100 simulation trials and for length $T = 1000$.

As we can deduce from the middle and the bottom panel, d_{QCD} was by far the best performing dissimilarity in Scenarios 2 and 3. In Scenario 3, according to ARI, the dissimilarities d_W, d_{GCC}, d_{VAR} and d_{PCA} were not able to detect the relationship of dependency generated by QVAR models at all. The nonparametric dissimilarity d_J lies somewhere in the middle between d_{QCD} and the remaining metrics. In Scenario 2, the dissimilarity d_{VAR} suffered from its model-based nature, thus attaining very poor results. The metrics d_W, d_{GCC}, d_J and d_{PCA} achieved similar results in terms of the four indexes, but they were no match for d_{QCD}, which dealt considerably well with dynamic conditional correlation models.

The top panel of Fig. 3 corroborates that the worst performance of d_{QCD} occurred when VARMA models came into play. In that case, the J-divergence-based measure achieved impressive results. For instance, in terms of the goodness-of-assignment index LOO1NN, it got the maximum score 98 out of 100 times. It is followed by Maharaj's distance, which is specifically designed to deal with this type of models. Among the remaining dissimilarities, though, d_{QCD} was the best performing according to the four clustering quality indexes.

To better clarify the clustering solutions achieved by each metric, we present in Fig. 4 the distribution (in percentage) of the number of correctly identified clusters over the 100 trials of the simulation mechanism. By correctly identified cluster we mean a group in the clustering solution which contains only the five series generated by the same process.

According to the top panel of Fig. 4, the J-divergence-based distance reached the authentic solution of five clusters almost 100% of the times in Scenario 1, whereas the model-based measure was capable of detecting the true solution only 60% of the times. The metric d_{QCD} got the perfect partition around 10% of the trials but 61% of the times identified correctly 3 clusters, clearly outperforming the rest of metrics, which obtained the right solution only the 1% of the times. These arguments corroborate an acceptable behaviour of d_{QCD} also under VARMA models.

The middle panel of Fig. 4 shows the distribution for Scenario 2. Again, it is clear that the Maharaj's metric is highly dependent on the underlying model, since it was not capable of correctly identifying even only a group in the 100 simulation trials. The distance d_{QCD} obtained very good results in terms of clustering identification, usually identifying 4 or 2 correct groups with percentages of 47% and 42%, respectively. The remaining dissimilarities performed worse, never discovering the true partition and generally discerning 0 or 1 correct clusters.

As for Scenario 3, the bottom panel shows the clear superiority of d_{QCD} in this setting. Its percentages of 1 and 3 correctly identified clusters were 41% and 59%, respectively. The rest of the metrics obtained almost always 0 correct clusters, aside from d_J, which unmasked 1 true group almost half of the trials.

Table 5
Percentage of times that the series of each process in Scenario 1 were grouped together in the clustering solution, according to the six measures whose performance is given in Table 3. The length of each series was $T = 1000$.

Measure	VAR(1) (a)	VAR(1) (b)	VMA(1) (c)	VMA(1) (d)	VARMA(1, 1) (e)
d_{QCD}	84	82	9	9	98
d_W	41	41	3	2	74
d_{GCC}	53	47	1	1	100
d_J	100	100	96	96	100
d_{VAR}	68	69	84	84	100
d_{PCA}	46	39	7	9	28

Table 6
Percentage of times that the series of each process in Scenario 2 were grouped together in the clustering solution, according to the six measures whose performance is given in Table 3. The length of each series was $T = 1000$.

Measure	Constant (a)	Piecewise (b)	Piecewise (c)	Piecewise (d)
d_{QCD}	100	64	68	54
d_W	80	2	0	0
d_{GCC}	47	0	0	0
d_J	63	0	1	0
d_{VAR}	0	0	0	0
d_{PCA}	28	2	4	0

As a way of clarifying which were the easiest processes to group, as well as the most difficult ones, Tables 5–7 contain the percentage of times that each process was correctly classified in Scenarios 1, 2 and 3, respectively.

As expected, Table 5 shows that the dissimilarities d_J and d_{VAR} correctly grouped each of the five processes most of the times. The main difference lies in VAR processes, whose patterns were more successfully detected by d_J than by d_{VAR}. As for the rest of the distances, they often failed to unmask the underlying structure of VMA processes, and were more successful in distinguishing VAR processes. It is worth mentioning that the metric based on the quantile cross-spectral density was able to detect uniformity in the series coming from VAR processes far more often than the wavelet-based, the cross-correlation-based and the PCA-based dissimilarities. Regarding the VARMA process, d_{QCD} was able to achieve a rate of detection close to 100%, which was reached by d_{GCC}. It is followed by d_W, which correctly grouped the series coming from the VARMA process 74% of the trials. In this regard, d_{PCA} was the worst performing dissimilarity, mixing series generated from the VARMA process with those coming from both VAR processes.

Concerning Scenario 2, we can see in Table 6 that all the measures, aside from d_{QCD}, were unable to detect homogeneity in series generated from processes that exhibit piecewise correlation. Although the average correlation between the standardized shocks is different in the processes (b), (c) and (d) presented in Scenario 2, those distances struggled to classify accurately this kind of processes, mixing the three types of series with one another. On the other hand, the quantile-based distance succeeded in discriminating the constant correlation process 100% of the times, and indisputably outperformed all of its competitors in the piecewise correlation context. It achieved a success percentage above 50% classifying processes (b), (c) and (d) in Scenario 2.

Finally, Table 7 reveals that dissimilarities other than d_{QCD} and d_J are not suitable to tackle the forms of dependence that arise in Scenario 3. The metric d_{QCD} always grouped together the series generated from process (a) in Scenario 3, sometimes mixing the series coming from processes (b) and (c). The distance d_J failed to distinguish between processes (a) and (b), but often succeeded in grouping together series from process (c).

We repeated the simulations for Scenarios 1 and 2, but this time the processes $(\epsilon_{t,1}, \epsilon_{t,2})'$ and $(\epsilon_{t,1}, \epsilon_{t,2}, \epsilon_{t,3})'$ were generated from a multivariate t distribution with 3 degrees of freedom. This allowed us to check the performance of the analysed dissimilarities under some amount of fat-tailedness in the error distribution. This feature is frequently

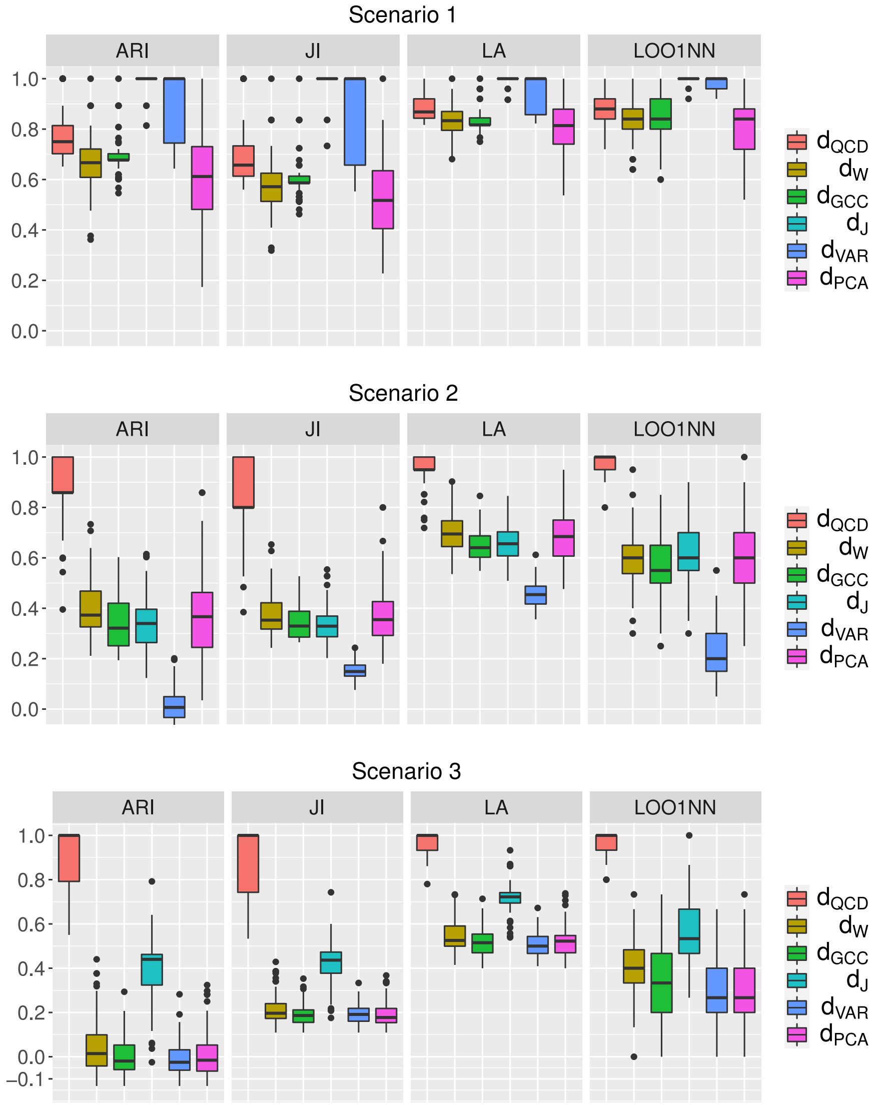

Fig. 3. Boxplots of four clustering validity indexes according to the 100 trials of the simulation procedure. The length of each series was $T = 1000$.

Table 7
Percentage of times that the series of each process in Scenario 3 were grouped together in the clustering solution, according to the six measures whose performance is given in Table 3. The length of each series was $T = 1000$.

Measure	QVAR(1) (a)	QVAR(1) (b)	QVAR(1) (c)
d_{QCD}	100	59	59
d_W	0	0	1
d_{GCC}	0	0	0
d_J	0	0	47
d_{VAR}	0	0	0
d_{PCA}	0	0	0

exhibited by some series, mainly within the field of Finance. Therefore, it is reasonable to introduce fat-tailedness in the simulations, especially in Scenario 2, since dynamic conditional correlation models originally arose to model financial time series of stock returns.

The results involving this new distribution for the error terms are given in Table 8 for $T = 250$ and in Table 9 for $T = 1000$. Observing the latter, one can reach several conclusions. In Scenario 1, the quantile cross-spectral dissimilarity d_{QCD} does not seem to be affected by the fat-tailedness of the error distribution, achieving a value of 0.772 for the ARI, versus 0.778 when the error terms follow a normal distribution. On the contrary, the remaining dissimilarities decreased their performance, especially d_J and d_{PCA}. The model-based dissimilarity d_{VAR} was the one achieving the best results, followed closely by the ones reached by d_{QCD}.

In Scenario 2, all the dissimilarities worsened their performance, but d_{QCD} was still the measure getting the highest scores. According to ARI, its average scores were at least twice as large as those reached by all the remaining dissimilarities. These results are very powerful, since they indicate that the quantile cross-spectral dissimilarity not only offers the best general performance when a wide variety of situations are taken into account, but it is also quite robust to changes in the error distribution. This makes d_{QCD} probably one of the best dissimilarities for practitioners to perform time series clustering, since it is well known that normality of the error terms cannot be guaranteed in many time series arising in several fields.

Again, in order to better understand the previous results, we have depicted in Fig. 5 the boxplots based on the clustering validity indexes for the 100 simulation trials ($T = 1000$).

The top panel of Fig. 5 shows that, when introducing heavy tails in the error distribution, the nonparametric dissimilarity d_J substantially

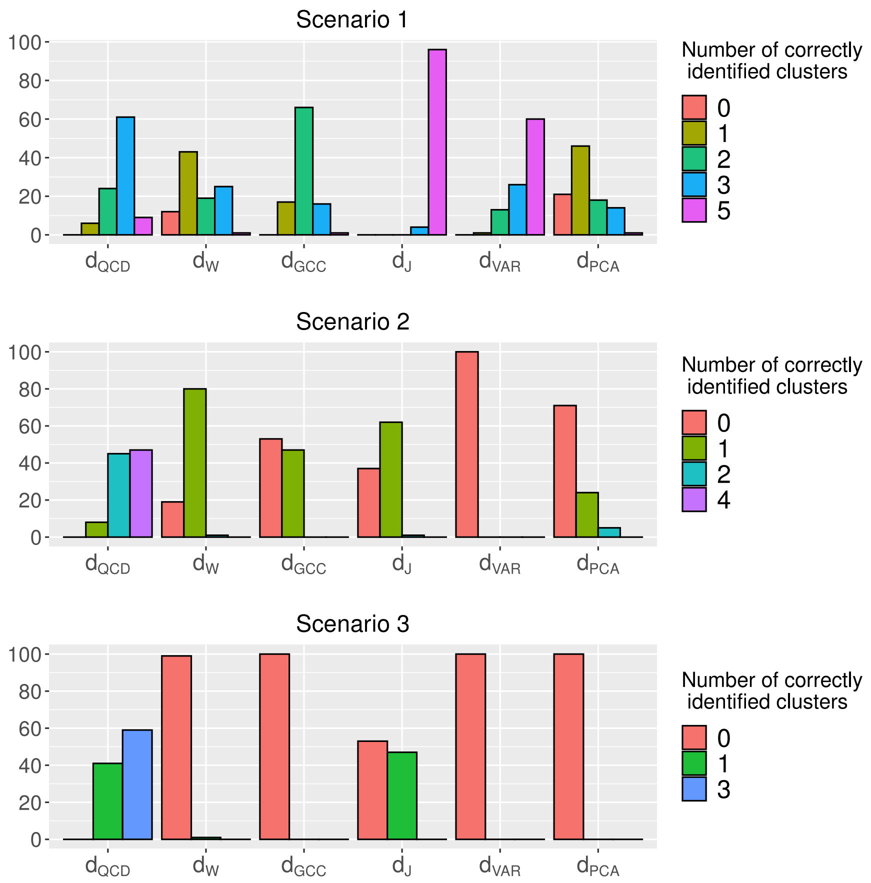

Fig. 4. Distribution (in percentage) of the number of correctly identified clusters in each trial of the simulation procedure. The length of each series was $T = 1000$.

Table 8
Averages and standard deviations (in brackets) of four clustering validity indexes for measures d_{QCD}, d_W, d_{GCC}, d_J, d_{VAR} and d_{PCA}, according to the 100 trials of the simulation procedure. Innovations were drawn from a multivariate t distribution with 3 degrees of freedom. For each scenario and index, the best result is shown in bold. The length of each series was $T = 250$.

	Index	d_{QCD}	d_W	d_{GCC}	d_J	d_{VAR}	d_{PCA}
Scenario 1	ARI	0.607	0.421	0.441	0.531	**0.645**	0.142
		(0.097)	(0.112)	(0.121)	(0.063)	(0.013)	(0.102)
	LA	0.790	0.659	0.690	0.723	**0.808**	0.513
		(0.067)	(0.079)	(0.076)	(0.052)	(0.013)	(0.072)
	LOO1NN	0.735	0.640	0.632	0.770	**0.852**	0.354
		(0.084)	(0.102)	(0.121)	(0.090)	(0.013)	(0.117)
	JI	0.519	0.374	0.384	0.455	**0.562**	0.187
		(0.093)	(0.077)	(0.094)	(0.056)	(0.013)	(0.060)
Scenario 2	ARI	**0.714**	0.411	0.255	0.440	0.012	0.333
		(0.013)	(0.123)	(0.119)	(0.150)	(0.070)	(0.112)
	LA	**0.867**	0.687	0.613	0.718	0.461	0.653
		(0.085)	(0.078)	(0.072)	(0.088)	(0.054)	(0.080)
	LOO1NN	**0.875**	0.682	0.428	0.676	0.242	0.517
		(0.013)	(0.128)	(0.141)	(0.122)	(0.112)	(0.153)
	JI	**0.653**	0.389	0.274	0.411	0.151	0.331
		(0.013)	(0.090)	(0.073)	(0.114)	(0.030)	(0.086)

worsened its performance, at least as far as the ARI, JI and LA indexes are concerned. In fact, the median values of these three indexes for d_{QCD} and d_{VAR} are very close, showing the latter measure better average performance because it sometimes got scores close to one as it can be deduced from the corresponding position of the box. On the other hand, the bottom panel of Fig. 5 displays almost the same picture as the middle panel of Fig. 3. The only difference is that all the measures, impacted by the fat-tailedness of the standardized shocks, decreased their scores regarding the four indexes by a similar degree. As a consequence, d_{QCD} retained its status as the best performing dissimilarity in the transformed scenario.

The barplots depicting the distribution of correctly identified clusters for each dissimilarity are shown in Fig. 6 for this new setting. The main difference of the top panel of Fig. 6 with respect to the top panel of Fig. 4 is that no longer the dissimilarity d_J identified the true cluster partition most of the times. In fact, under these new circumstances, d_J usually recognized one correct cluster, being able to reach the perfect partition a tiny percentage of the trials, 3%. Perhaps the most important information that the top panel of Fig. 6 contains is that, as far as d_{QCD} is concerned, the distribution of correctly identified clusters remained almost the same as in the previous framework, where

Table 9
Averages and standard deviations (in brackets) of four clustering validity indexes for measures d_{QCD}, d_W, d_{GCC}, d_J, d_{VAR} and d_{PCA}, according to the 100 trials of the simulation procedure. Innovations were drawn from a multivariate t distribution with 3 degrees of freedom. For each scenario and index, the best result is shown in bold. The length of each series was $T = 1000$.

	Index	d_{QCD}	d_W	d_{GCC}	d_J	d_{VAR}	d_{PCA}
Scenario 1	ARI	0.772	0.577	0.625	0.650	**0.845**	0.262
		(0.092)	(0.083)	(0.093)	(0.111)	(0.013)	(0.124)
	LA	0.883	0.766	0.797	0.806	**0.917**	0.589
		(0.057)	(0.058)	(0.055)	(0.074)	(0.013)	(0.088)
	LOO1NN	0.884	0.763	0.785	0.914	**0.970**	0.519
		(0.059)	(0.093)	(0.092)	(0.056)	(0.013)	(0.127)
	JI	**0.692**	0.495	0.540	0.568	0.650	0.261
		(0.013)	(0.073)	(0.085)	(0.116)	(0.111)	(0.078)
Scenario 2	ARI	**0.775**	0.359	0.282	0.352	0.000	0.313
		(0.013)	(0.116)	(0.098)	(0.115)	(0.070)	(0.121)
	LA	**0.897**	0.671	0.627	0.667	0.452	0.646
		(0.013)	(0.075)	(0.064)	(0.068)	(0.056)	(0.073)
	LOO1NN	**0.898**	0.552	0.490	0.527	0.222	0.480
		(0.013)	(0.114)	(0.135)	(0.109)	(0.115)	(0.172)
	JI	**0.713**	0.352	0.305	0.348	0.144	0.316
		(0.013)	(0.077)	(0.061)	(0.080)	(0.037)	(0.081)

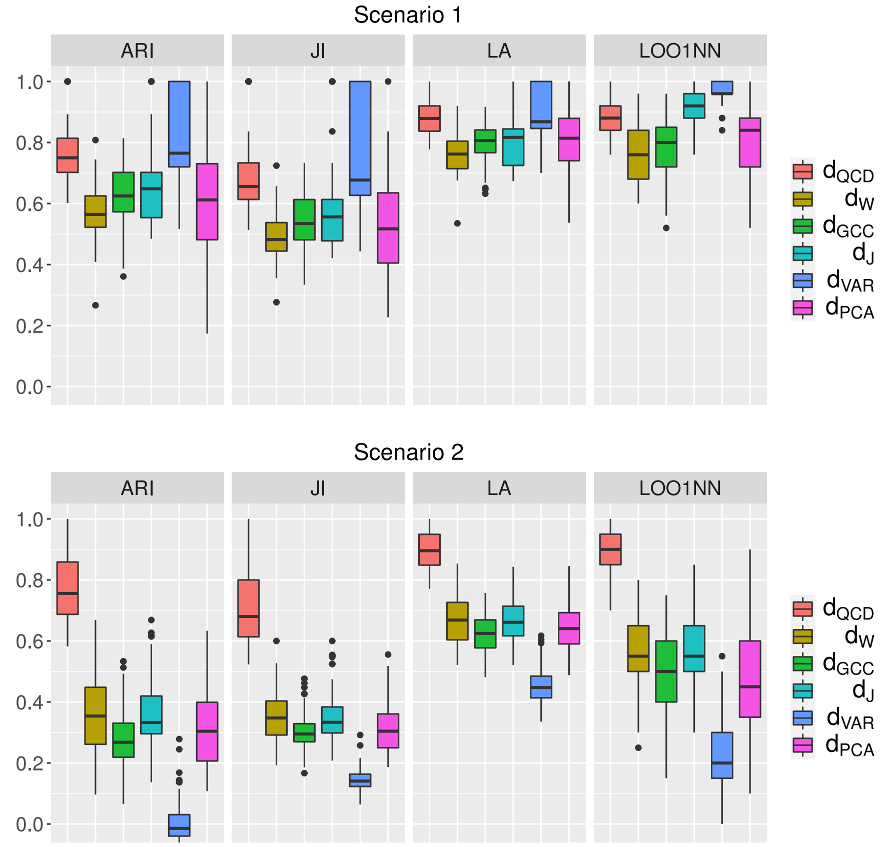

Fig. 5. Boxplots of four clustering validity indexes according to the 100 trials of the simulation procedure. Innovations were drawn from a multivariate t distribution with 3 degrees of freedom. The length of each series was $T = 1000$.

the error terms were normally distributed. Indeed, aside from the VAR-based dissimilarity, the metric d_{QCD} was the only one that identified at least one correct cluster in each one of the trials. This strongly reinforces the nature of d_{QCD} as an all-purpose distance measure.

The main difference of the bottom panel of Fig. 6 in comparison to the middle panel of Fig. 4 rests on the fact that, when heavy tails were considered in the error distribution, the dissimilarity d_{QCD} arrived at the true solution only 10% of the times, whereas this percentage was 47% under normality of errors. In the new framework, d_{QCD} almost always reached two correct clusters, being again the only metric that always identified at least one correct group.

As before, the percentage of times that each one of the processes was correctly classified is provided in Tables 10 and 11 for Scenarios 1 and 2, respectively. The most remarkable element of Table 10 are the results achieved by d_J, which totally failed to discriminate between both VMA processes when the normality assumption for the error terms was removed. Except for the model-based dissimilarity d_{VAR}, the quantile-based distance d_{QCD} was the single metric capable of detecting the VARMA process almost 100% of the times. With regards to Scenario 2, by observing Table 11, we can now better understand the distribution of correctly identified clusters associated with d_{QCD} that we previously talked about (displayed on the bottom panel of

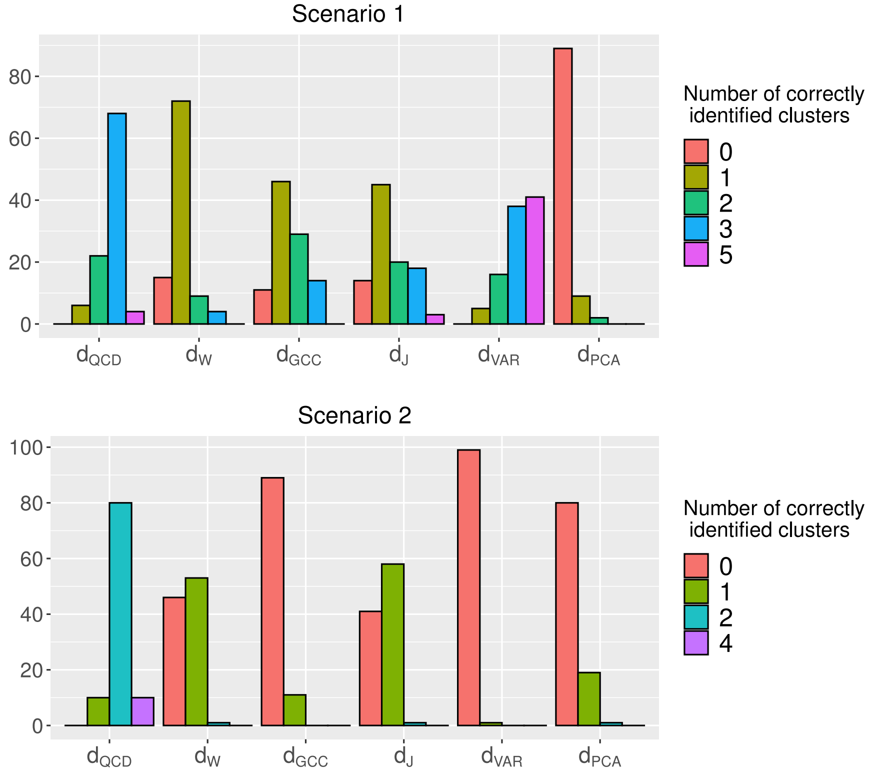

Fig. 6. Distribution (in percentage) of the number of correctly identified clusters in each trial of the simulation procedure. Innovations were drawn from a multivariate t distribution with 3 degrees of freedom. The length of each series was $T = 1000$.

Table 10
Percentage of times that the series of each process in Scenario 1 were grouped together in the clustering solution, according to the six measures whose performance is given in Table 9. Innovations were drawn from a multivariate t distribution with 3 degrees of freedom. The length of each series was $T = 1000$.

Measure	VAR(1) (a)	VAR(1) (b)	VMA(1) (c)	VMA(1) (d)	VARMA(1, 1) (e)
d_{QCD}	87	78	5	5	99
d_W	7	14	0	0	81
d_{GCC}	33	33	0	0	80
d_J	37	36	3	4	74
d_{VAR}	55	79	71	71	100
d_{PCA}	4	7	1	1	0

Table 11
Percentage of times that the series of each process in Scenario 2 were grouped together in the clustering solution, according to the six measures whose performance is given in Table 9. Innovations were drawn from a multivariate t distribution with 3 degrees of freedom. The length of each series was $T = 1000$.

Measure	Constant (a)	Piecewise (b)	Piecewise (c)	Piecewise (d)
d_{QCD}	95	95	10	10
d_W	54	1	0	0
d_{GCC}	10	1	0	0
d_J	58	2	0	0
d_{VAR}	0	0	1	0
d_{PCA}	16	5	0	0

Fig. 6). When heavy tails were assumed for the error distribution, the distance d_{QCD} generally (95% of the times) grouped correctly the series generated from processes (a) and (b) in Scenario 2. However, it was often incapable of telling apart the series coming from processes (c) and (d). This was probably due to the fact that the standardized shocks in these two processes exhibit the most similar average correlations, 0.2 and 0, respectively.

4. Time consumption assessment

In order to assess the efficiency of the main six dissimilarity measures analysed throughout Section 3, we have recorded the runtime of the corresponding programs to finish the clustering task regarding Scenario 1. Note that, for some metrics, as d_{QCD}, the clustering task consists of a feature extraction, which is the most time consuming part, followed by the computation of a distance matrix. In contrast, for other metrics, as d_J, the clustering task consists directly of the computation of a distance matrix. Of course, in both cases, the distance matrix is provided as input to the PAM algorithm, which gives the clustering solution, thus finishing the clustering task. In order to uncover the evolution of the running time as a function of the length T, it was recorded for $T = 250$ and $T = 1000$. Besides, with the aim of removing the uncertainty caused by uncontrollable factors, we have taken the running time over several trials. Thus, given a metric and a value for T, we reported the average CPU runtime spent in finishing the clustering task for one of the sets of 25 series of length T which were simulated in Scenario 1.

The runtime was recorded in exactly the same way for the six dissimilarity measures. The computer used to run the programs was a MacBook Pro with processor Quad-Core Intel Core i7, a speed of 2.9 GHz and a RAM memory of 16 GB. The programs were coded and executed in RStudio. The R version was 3.6.1.

The CPU runtime for the six dissimilarity measures is provided in Table 12. The more efficient distances were the wavelet-based dissimilarity and the PCA-based dissimilarity, both spending less than 0.17 s in completing the clustering task for $T = 1000$. The metric d_{QCD} consumed about 1.57 s when dealing with series of length 250, and about 5.3 s when coping with series of length 1000. This suggests that the time complexity of d_{QCD} scales linearly with the length of the series. The generalized cross-correlation-based dissimilarity d_{GCC} spent roughly 12 s in finishing the clustering task for $T = 250$, and

Table 12
The CPU runtime for the six measures in Table 3, regarding the 100 simulation trials in Scenario 1. Two values for the length of the series were considered, T = 250 and T = 1000.

Measure	CPU runtime (minutes) for T = 250	CPU runtime (minutes) for T = 1000
d_{QCD}	0.0262	0.0889
d_{W}	0.0017	0.0024
d_{GCC}	0.2038	0.2058
d_{J}	0.3516	1.6845
d_{VAR}	2.9431	14.0209
d_{PCA}	0.0027	0.0028

almost the same time for $T = 1000$. This implies that, whereas the autocorrelations and the cross-correlations are computed in an efficient way, the calculation of determinants and inverse matrices associated with each pair of components of a MTS (see equations (8) and (9) in Alonso & Peña, 2019) slows down the process. Finally, the model-based dissimilarity d_{VAR} was the slowest by a large degree, spending about 14 min to obtain the experimental partition when T = 1000. This was expected, since performing clustering through d_{VAR} implies fitting VAR models to each MTS, which is computationally expensive.

In summary, although the quantile-based metric is not the best dissimilarity in terms of time consumption, it has proven to be computationally efficient in performing MTS clustering. This fact, combined with the effectiveness that d_{QCD} has shown in grouping time series generated from a wide variety of different processes, makes d_{QCD} an attractive dissimilarity in terms of both performance and efficiency.

5. A case study: Clustering bivariate series of daily returns and trading volume of some S&P 500 companies

In this section, the dissimilarity based on the quantile cross-spectral density, d_{QCD}, is used to perform clustering on a real data example involving financial time series. We consider both the daily stock prices and trading volume of some companies belonging to the S&P 500 index, which comprises 505 common stocks issued by 500 large-cap companies and traded on American stock exchanges. The S&P 500 is commonly divided in eleven sectors. Only the companies belonging to financial and utilities sectors are considered. The reason of this choice rests on the fact that these two sectors are clearly different, in the sense that the activities performed by any two companies pertaining to each one of them greatly differ from one another. This is highly desirable, as our main purpose is to show to what extent d_{QCD} is able to distinguish between series belonging to highly different economic sectors, supposedly representing strongly different economic behaviours. Had we chosen somehow overlapping sectors (e.g., all the eleven sectors) and the achieved conclusions could be misguided. The financial sector consists of banks, insurance companies, credit card issues and a host of other money-centric enterprises. The utilities sector includes many local electricity and water companies, among many others. Our database contains information about 55 companies belonging to the financial sector and 25 companies belonging to the utilities sector. The sample period spans from 8th February 2013 to 7th February 2018, thus resulting serial realizations of length T = 1258. The data are sourced from the Kaggle repository of Cameron Nugent[1] (Nugent, 2017).

Each of the 80 considered companies is then described by means of a bivariate time series whose components are the company's daily stock price and trading volume. The relationship between price and volume has been extensively analysed in the literature (Campbell, Grossman, & Wang, 1993; Gebka, & Wohar, 2013; Karpoff, 1987) and constitutes itself a topic of great financial interest. Prices and trading volume are known to exhibit some empirical linkages over the fluctuations of stock markets. Here, however, our concern is not to analyse whether these empirical facts hold true in the considered series, but to assess if the study of the joint behaviour of prices and volume via d_{QCD} can give insights into the sector to which a company pertains. It seems reasonable to hypothesize that the joint behaviour of prices and volume shows some distinctive features depending on the sector. It can be observed that both the UTS of prices and trading volume are nonstationary in mean. For this reason, all UTS are transformed by taking the first differences of the natural logarithm of the original values. This way, prices give rise to stock returns, and volume, to what we call change in volume. This transformation is common when dealing with this kind of series (Chen, 2012). Finally, both series are normalized to have zero mean and unit variance. For each one of the sectors, we have depicted nine of the transformed series. The series corresponding to the financial sector are shown in Fig. 7, whereas the series corresponding to the utilities sector are displayed in Fig. 8. In all cases, the blue colour corresponds to the change in trading volume, while the red colour corresponds to the stock returns. We have included the symbol of each one of the companies as given in the S&P 500.

Table 13
Descriptive statistics of normalized returns and change in volume for the series in the top left panel of Fig. 7.

Descriptive statistics	Returns	Change in volume
Minimum	−7.4529	−3.1000
Maximum	3.5026	6.7142
Skewness	−0.7321	0.4207
Kurtosis	4.1400	2.0883

Similar to other financial time series, stock returns exhibit empirical statistical regularities, so-called "stylized-facts". It is crucial to be aware of them in order to perform a proper analysis. The most common stylized facts include: heavy tails and a peak centre compared to the normal distribution, volatility clustering (periods of low volatility mingle with periods of high volatility), leverage effects (returns are negatively correlated with volatility), and autocorrelation at much longer horizons than expected. In the same way, trading volume is known to empirically depart from normality. Table 13 provides some descriptive statistics regarding the returns and change in volume of the series in the top left panel of Fig. 7, which corresponds to the company Aflac. We can see that both the returns and the change in volume are skewed. In addition, the value of the kurtosis for the returns is 4.14, thus implying fatter tails than those of the normal distribution. Then, our proposal is to take advantage of the high capability of the quantile cross-spectral density to detect these stylized facts and performing cluster analysis based on d_{QCD}. In fact, d_{QCD} yielded by far the best average results classifying processes whose error terms exhibited some degree of fat-tailedness (see Table 9). Given its great performance in this kind of situations, we hope that, in the current scenario, d_{QCD} can distinguish two clusters, one mostly formed of the companies in the financial sector and the other mainly constituted of the companies in the utilities sector.

The 80 bivariate series of normalized returns and change in volume were subjected to PAM algorithm with the proposed dissimilarity, d_{QCD}. Just as in the simulations, $r = 3$ quantiles of levels 0.1, 0.5 and 0.9 along with the Fourier frequencies were considered to compute d_{QCD}. Once obtained the clustering solution, the performance of d_{QCD} was assessed by comparing this solution with the assumed true partition, which is given by the sectors to which each company pertains. The same four indexes considered in Section 3 were used here to quantify the performance of the proposed metric.

In order to show to what extent d_{QCD} is better able to detect the underlying structure in the data than other dissimilarities, the competitors in Table 3 were considered in the problem of grouping the financial time series. For all the dissimilarities, the same settings as in Section 3 were considered. It is worth noting that the selection of one lag for d_{GCC} seems appropriate, given that the 1-lagged returns

[1] https://www.kaggle.com/camnugent/sandp500?.

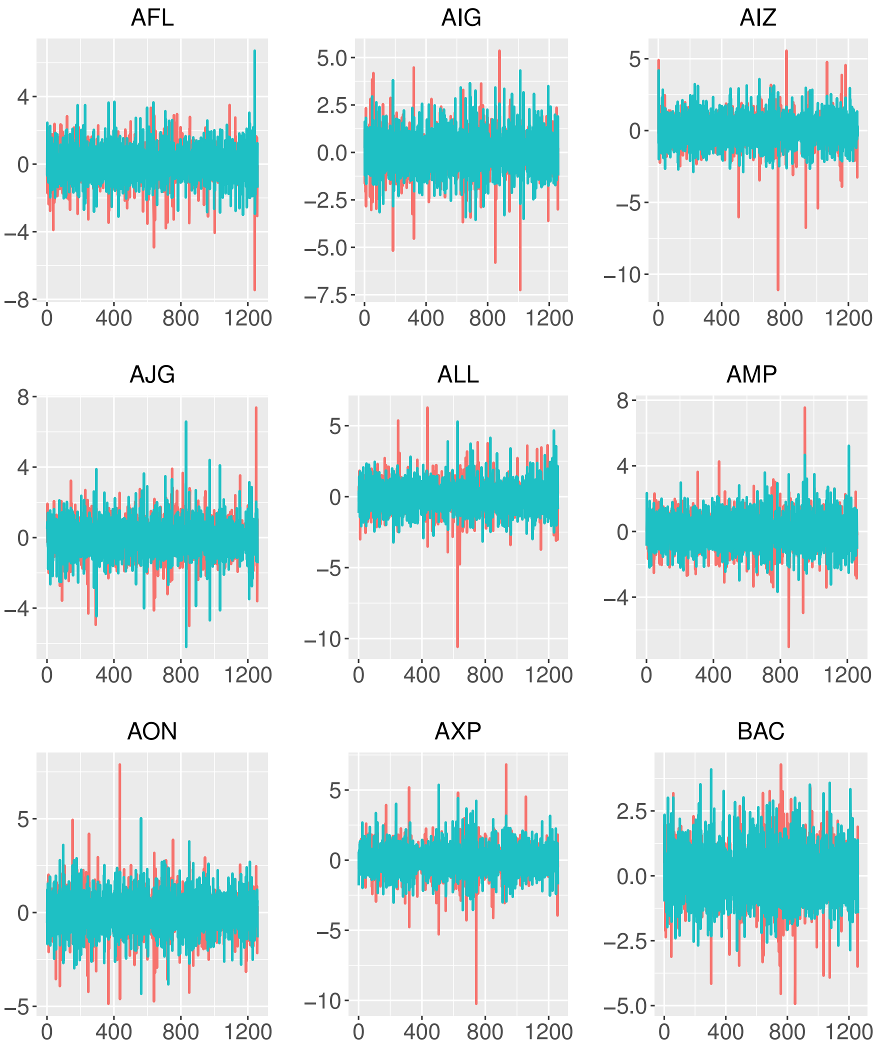

Fig. 7. Bivariate series of returns (red colour) and change in volume (blue colour) for 9 companies of the S&P 500 belonging to the financial sector.

Table 14
Four clustering validity indexes for measures d_{QCD}, d_W, d_{GCC}, d_J, d_{VAR} and d_{PCA}, with regards to the financial time series.

Measure	ARI	LI	LOO1NN	JI
d_{QCD}	**0.718**	**0.918**	**0.963**	**0.771**
d_W	0.352	0.790	0.863	0.532
d_{GCC}	0.147	0.700	0.813	0.433
d_J	0.030	0.641	0.863	0.567
d_{VAR}	0.023	0.598	0.663	0.381
d_{PCA}	0.167	0.616	0.725	0.574

Table 15
Breakdown of the number of companies from each sector located in each cluster, with regards to the clustering solution achieved by d_{QCD}.

	No. companies in the financial sector	No. companies in the utilities sector
Cluster 1	6	25
Cluster 2	49	0

have been proven to have a highly predictive ability over the trading volume (Chen, 2012).

Table 14 shows the values achieved by the six dissimilarities with regards to each one of the indexes. It can be seen that d_{QCD} got the best results, clearly outperforming the remaining dissimilarity measures in terms of all the considered indexes. Regarding the ARI, the proposed measure obtained 0.718, while its nearest competitor, d_W, only achieved 0.352. The generalized cross-correlation-based distance and the PCA-based metric slightly detected some structure in the data, whereas the Maharaj's distance and the nonparametric dissimilarity d_J were not capable of discriminating the series in both sectors at all. Similar insights can be obtained from the results with regards to the rest of the indexes. For instance, the high value of LOO1NN that d_{QCD} attained indicates that this metric is indeed the most appropriate to distinguish between underlying sectors in this kind of series.

In order to better understand the solution reached by d_{QCD}, Table 15 shows the number of companies pertaining to each sector which fell in each one of the clusters. It can be noticed that all the companies in the utilities sector were located in the first cluster, along with 6 companies belonging to the financial sector. In addition, the second cluster is a "pure" cluster, containing the remaining 49 companies in the financial sector.

The general impression that one gets from Table 15 is that the joint behaviour of price and volume is strongly related to the sector a given company pertains to, and that d_{QCD} is indeed able to distinguish between the different underlying relationships of dependence. It seems

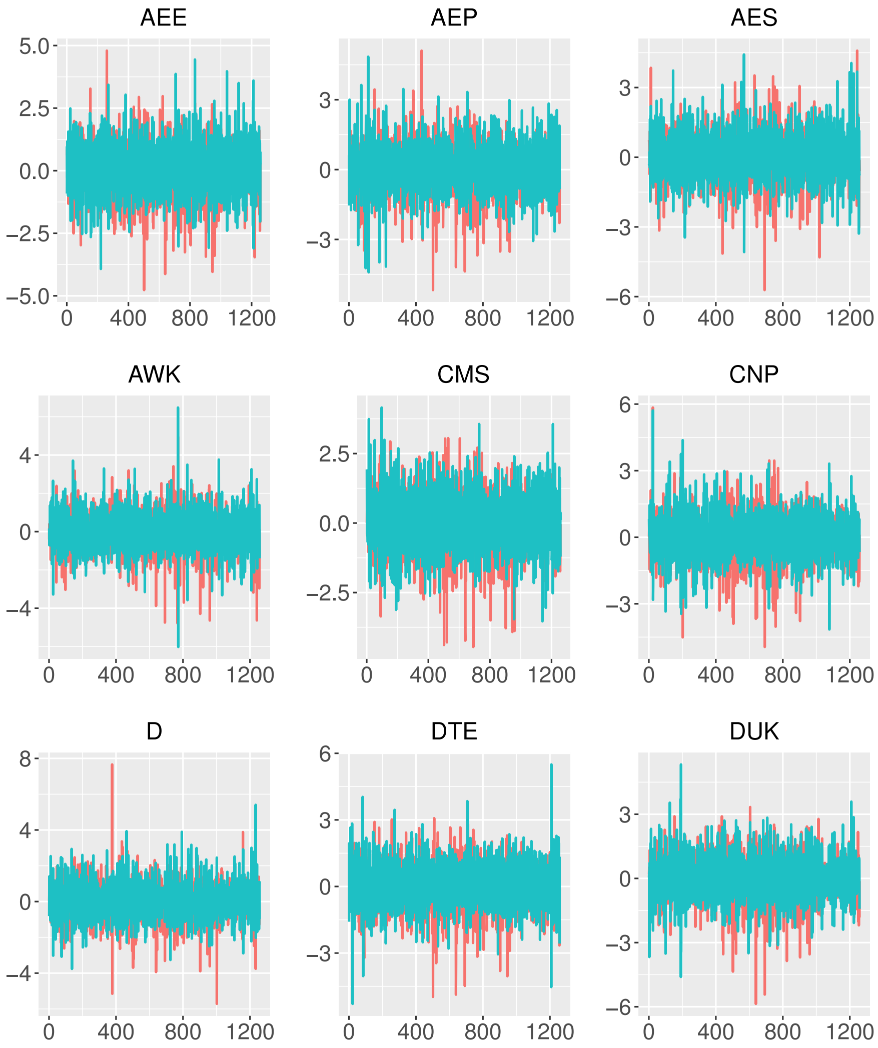

Fig. 8. Bivariate series of returns (red colour) and change in volume (blue colour) for 9 companies of the S&P 500 belonging to the utilities sector.

that, in real-life situations like this, the approach followed by this metric, aimed to capture any type of dependence, gives it a pivotal advantage over the remaining measures, which are devised to uncover particular types of dependence. By analysing this type of series via d_{QCD}, an investor could realize, for instance, that there are 6 particular companies in the financial sector whose behaviour is similar to that of the companies in the utilities sector, giving her valuable information about the market.

Perhaps one of the biggest advantages of using the PAM algorithm to perform clustering is that it produces a real prototype for each encountered group. These prototypes are usually known as medoids, and they actually pertain to the original set that was subjected to the clustering procedure, giving them high interpretation. In this example, the medoids are bivariate time series which represent all the time series belonging to each cluster. They synthesize the cluster information and represent the prototype features of the clusters, then summarizing the characteristics of the time series within each group. Given this informative power of medoids, it is undeniably interesting to know which company is playing the role of prototype in both sectors.

The medoid time series of both sectors are depicted in Fig. 9. The top panel corresponds to the medoid bivariate time series within the cluster containing the companies in the utilities sector (Cluster 1). It represents the company Dominion Energy, commonly referred to as Dominion, a power and energy company headquartered in Richmond, which supplies electricity and natural gas in different parts of the United States. The bottom panel, displaying the medoid series regarding the financial cluster (Cluster 2), corresponds to U.S. Bancorp, a bank holding company based in Minneapolis, Minnesota, which provides banking, investment, and payment service products, among others. An investor could use the companies associated with both medoids as a proxy to describe the financial situation of the corresponding sectors. Maybe, analysing these two companies gives more valuable insights into the future of both sectors than performing an extensive study over the whole groups.

It is difficult to work out from Fig. 9 that those two series correspond to two kinds of financial behaviours which are profoundly different. This is in part due to the length of both series, 1258, which makes difficult for the eye to detect any pattern. Besides, this is probably also attributable to the complex forms of dependence that exist between the price and the volume of each company. For instance, we have seen in Table 14 that d_{GCC}, a dissimilarity measure based on an intuitive quantity as the cross-correlation, barely noticed the existence of two different underlying sectors.

Given the previous results, we are compelled to emphasize that practitioners in the field of MTS clustering should take into account the dissimilarity based on the quantile cross-spectral density, capable of detecting patterns that could remain invisible otherwise.

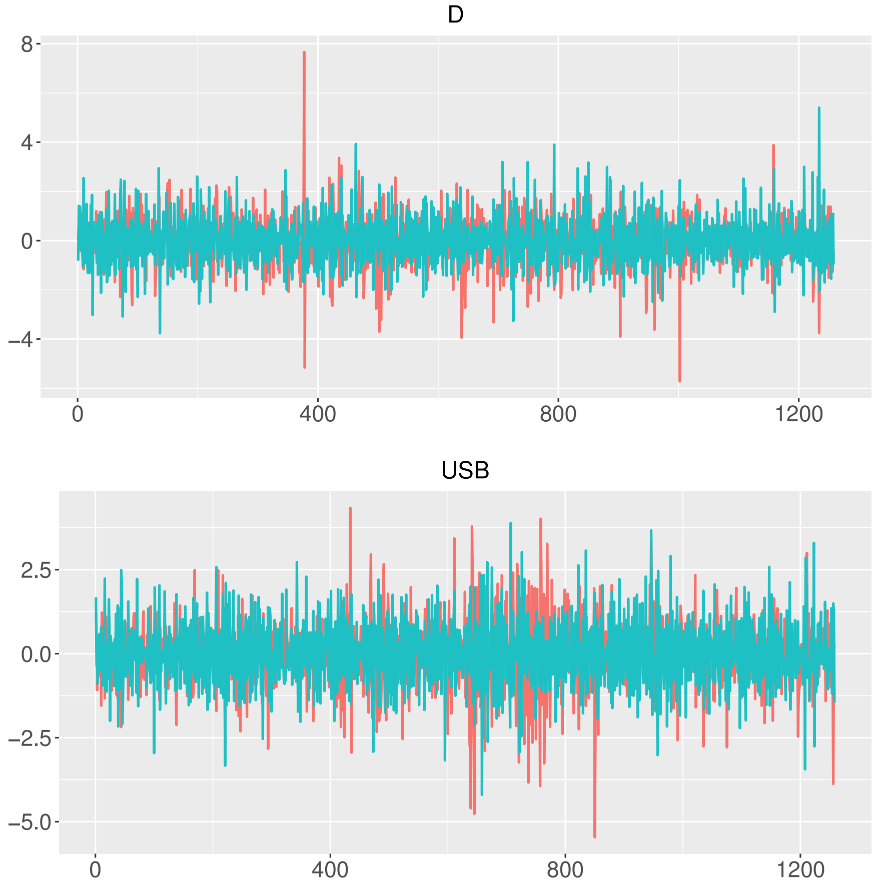

Fig. 9. Top panel: bivariate series of returns (red colour) and change in volume (blue colour) of the company Dominion Energy, which represents the medoid of the cluster of companies in the utilities sector (Cluster 1). Bottom panel: bivariate series of returns (red colour) and change in volume (blue colour) of the company U.S. Bancorp, which represents the medoid of the cluster of companies in the financial sector (Cluster 2).

6. Application to some UEA datasets

In this section, we apply the proposed approach to the MTS datasets provided in the well-known UEA multivariate time series classification archive (Bagnall et al., 2018). The archive consists of 30 MTS datasets covering a wide range of cases, dimensions and series lengths. It is worth highlighting that most of the series in this collection are nonstationary. Note that QCD is ill-defined for nonstationary series. However, even when the corresponding MTS lack stationarity, one can still compute the smoothed CCR-periodograms in (10) as descriptive features. Indeed, several works have followed a similar path, either for clustering or classification of MTS. For instance, Zagorecki (2015) proposed a generic method that can be applied to an arbitrary dataset in order to perform MTS classification. Each MTS is described by means of a set of features derived from its UTS, which are then used to feed a random forest classifier. Among these features, several autocorrelations regarding some lags are considered, although such quantities are only well-defined for stationary series. In the same way, Wang et al. (2007) introduced an approach for clustering general MTS taking into account the Box–Pierce statistic, which is based on the series autocorrelations. Both procedures are successfully applied in real MTS databases. Thus, analysing the performance of d_{QCD} when dealing with nonstationary MTS gives insights into the quality of this metric to perform clustering concerning general real situations, even when the necessary theoretical requirements are not fulfilled.

The distance d_{QCD} and the dissimilarities analysed throughout Section 3 were used to perform clustering in the 30 datasets contained in the UEA archive plus an additional dataset collected from the PTB diagnostic ECG database (Goldberger et al., 2000). Note that ECG signals are one of the most common type of nonstationary MTS. The considered databases were created for classification purposes, so training and testing sets are provided for each dataset. As we are considering a clustering algorithm, we merged both to get whole sets of MTS which are subjected to the clustering technique. The same clustering quality indexes as in Section 3 were computed. The ground truth was assumed to be given by the true labels of the MTS provided in the UEA archive and in the PTB database.

For illustrative purposes and a clear presentation of the results, we first focus on a preliminary analysis of 7 of the 31 datasets (see summary in Table 16). These databases cover a broad variety of dimensions, series lengths and numbers of classes, thus constituting, aside from the ECG dataset, a heterogeneous subset of the UEA archive. Each one of them describes a different problem. For instance, the dataset named RacketSports was created from university students playing badminton or squash while wearing a smart watch, being the target to identify the sport and the stroke of each player. A detailed summary of the corresponding problems can be seen in Bagnall et al. (2018).

The clustering quality indexes obtained for the seven databases presented in Table 16 are reported in Table 17. The symbol ∅ used for d_{VAR} with RacketSports indicates that the corresponding VAR models could not be correctly fitted to the MTS in this dataset due to numerical issues.

Overall, results in Table 17 show that the spectral metrics d_{QCD} and d_J substantially outperformed the remaining dissimilarities in relation to the four quality indexes. In fact, when the whole set of 28 scores obtained for each dissimilarity measure are ranked and compared, we observe that d_{QCD} acquired the best scores in 15 of them, the same number as d_J. The remaining distances were unable to get the best

Table 16
Summary of a subset of six datasets from the UEA multivariate time series classification archive and the ECG database.

Dataset	Number of series	Dimensions	Length	Classes
Cricket	180	6	1197	12
ArticularyWord	575	9	144	25
BasicMotions	80	6	100	4
Epilepsy	275	3	206	4
Libras	360	2	45	15
RacketSports	303	6	30	4
ECG	80	6	500	2

Table 17
Clustering validity indexes for measures d_{QCD}, d_W, d_{GCC}, d_J, d_{VAR} and d_{PCA} regarding the 7 datasets summarized in Table 16. For each dataset and index, the best result is shown in bold.

Dataset	Index	d_{QCD}	d_W	d_{GCC}	d_J	d_{VAR}	d_{PCA}
Cricket	ARI	0.848	0.496	0.720	**0.975**	0.280	0.793
	LA	0.916	0.670	0.836	**0.989**	0.493	0.876
	LOO1NN	**0.994**	0.850	0.928	**0.994**	0.706	0.972
	JI	0.754	0.369	0.592	**0.955**	0.208	0.680
ArticularyWord	ARI	0.676	0.350	0.499	**0.817**	0.123	0.575
	LA	0.811	0.534	0.678	**0.887**	0.345	0.745
	LOO1NN	0.963	0.699	0.863	**0.983**	0.580	0.899
	JI	0.526	0.232	0.351	**0.700**	0.088	0.421
RacketSports	ARI	**0.424**	0.069	0.031	0.111	∅	0.085
	LA	**0.677**	0.420	0.370	0.429	∅	0.422
	LOO1NN	**0.908**	0.710	0.548	0.832	∅	0.756
	JI	**0.430**	0.179	0.163	0.232	∅	0.196
BasicMotions	ARI	0.686	0.651	0.491	**1**	0.517	0.686
	LA	0.838	0.806	0.735	**1**	0.724	0.815
	LOO1NN	**1**	**1**	0.900	**1**	0.938	0.900
	JI	0.618	0.600	0.443	**1**	0.486	0.630
ECG	ARI	**0.520**	−0.012	0.192	0.149	0.007	0.001
	LA	**0.862**	0.488	0.724	0.698	0.649	0.608
	LOO1NN	**0.838**	0.530	0.763	0.800	0.550	0.575
	JI	**0.610**	0.326	0.422	0.403	0.459	0.389
Epilepsy	ARI	**0.693**	0.251	0.333	0.692	0.090	0.284
	LA	0.857	0.519	0.600	**0.866**	0.415	0.606
	LOO1NN	0.927	0.825	0.542	**0.967**	0.833	0.622
	JI	**0.629**	0.315	0.357	0.628	0.204	0.323
Libras	ARI	**0.323**	0.037	0.066	0.268	0.203	0.115
	LA	**0.514**	0.226	0.260	0.464	0.386	0.279
	LOO1NN	0.775	0.311	0.289	**0.820**	0.564	0.322
	JI	**0.227**	0.055	0.069	0.200	0.150	0.097

result in any instance, the only exception being d_W in the dataset BasicMotions, where it equalized d_{QCD} and d_J in terms of the LOO1NN index.

The next step was to consider the results for the 31 datasets. Here, the dissimilarity d_{VAR} was removed from the analyses due to the fact that it gave rise to numerical problems in a large number of databases. However, this is not a crucial issue since this metric usually was outperformed by the remaining ones when successfully computed (e.g., see Table 17). Additionally, due to the high computational complexity of the distance d_J (see Table 12), we selected a subset of instances, dimensions and time observations in those datasets in which d_J was computationally infeasible. The subset selection was carried out by retaining:

1. The *first* $\lceil \frac{300}{C} \rceil$ instances of each category, being C the total number of categories in the set and $\lceil \cdot \rceil$ the ceiling function. This way, a balanced subset of instances was always obtained.
2. The *first* 5 variables of each MTS.
3. The *first* 500 time observations, thus having available MTS of length $T = 500$.

Table 18
Results from the Nemenyi post-hoc tests for indexes ARI, LA, LOO1NN and JI: p-values for the pairwise comparisons, sum of ranks for each metric and homogeneous groups (different letters indicate significant differences, $p < 0.05$). 21 MTS datasets were considered. The p-values less than 0.05 are shown in bold.

Index	p-values in pairwise comparisons					Rank sum	Groups		
ARI									
	d_{QCD}	d_J	d_{PCA}	d_{GCC}	d_W				
d_{QCD}	–	0.8772	**0.0086**	**0.0035**	**5.0e−07**	86.5	A		
d_J	–	–	0.1294	0.0703	**6.7e−05**	77.0	A	B	
d_{PCA}	–	–	–	0.9991	0.1991	53.5		B	C
d_{GCC}	–	–	–	–	0.3172	51.0		B	C
d_W	–	–	–	–	–	32.0			C
LA									
	d_{QCD}	d_J	d_{PCA}	d_{GCC}	d_W				
d_{QCD}	–	1.0000	**0.0227**	**0.0086**	**3.4e−06**	83.0	A		
d_J	–	–	**0.0306**	**0.0120**	**5.7e−06**	82.0	A		
d_{PCA}	–	–	–	0.9982	0.2199	53.0		B	
d_{GCC}	–	–	–	–	0.3735	50.0		B	
d_W	–	–	–	–	–	32.0		B	
LOO1NN									
	d_{QCD}	d_J	d_{PCA}	d_{GCC}	d_W				
d_{QCD}	–	0.9148	**0.0166**	**0.0017**	**0.0017**	79.5	A		
d_J	–	–	**0.0007**	**4.2e−05**	**4.2e−05**	88.0	A		
d_{PCA}	–	–	–	0.9667	0.9667	48.5		B	
d_{GCC}	–	–	–	–	1.0000	42.0		B	
d_W	–	–	–	–	–	42.0		B	
JI									
	d_{QCD}	d_J	d_{PCA}	d_{GCC}	d_W				
d_{QCD}	–	0.9999	0.1448	**0.0141**	**7.4e−06**	80.5	A		
d_J	–	–	0.1023	**0.0086**	**3.4e−06**	82.0	A		
d_{PCA}	–	–	–	0.9148	0.0617	57.5	A	B	
d_{GCC}	–	–	–	–	0.3735	49.0		B	
d_W	–	–	–	–	–	31.0		B	

Note that the selection procedure previously described totally avoids the choice of a random seed, which could have been specifically tuned to favour the authors' interests.

In order to eliminate undesirable noise from the comparative analysis, the datasets for which the five considered dissimilarities achieved poor results were ignored. Specifically, we omitted the databases in which the values of the ARI for all the metrics were below 0.05. Note that this is a sensible choice, since the expected value of this quantity is zero for a clustering solution picked at random. By proceeding this way, 10 datasets were removed from the study, namely AtrialFibrillation, DuckDuckGeese, EthanolConcentration, FaceDetection, FingerMovements, HandMovementDirection, Heartbeat, MotorImagery, SelfRegulationSCP2 and SpokenArabicDigits. It is worth mentioning that most of these databases were subjected to the subset selection procedure described earlier.

With the aim of rigorously comparing the different dissimilarities, statistical tests were performed by taken into account the 21 resulting datasets. For each quality index, the Friedman test was first used to assess whether there are statistically significant differences between the index distributions with the considered metrics. In all cases, Friedman test was significant, with respective p-values given by $p = 1.086\text{e}{-07}$ for the ARI, $p = 7.237\text{e}{-08}$ for the LA, $p = 5.28\text{e}{-08}$ for LOO1NN and $p = 1.392\text{e}{-07}$ for the JI. Then, multiple pairwise-comparisons using the Nemenyi post-hoc test were carried out to determine which pairs of metrics are different. The corresponding p-values are reported in Table 18. Note that the Nemenyi test was developed to account for a family-wise error and is already a conservative test. Therefore, there is no need for a p-values adjustment.

Assuming a significance level of 0.05, results in Table 18 show that the proposed metric d_{QCD} attains values for ARI, LA and LOO1NN significantly better than the ones obtained with d_W, d_{GCC} and d_{PCA}. The only exception occurs with the Jaccard index, where d_{PCA} clearly exhibits lower ranges but not significant ($p > 0.05$). In all cases, no significant differences are observed between d_{QCD} and d_J, but the latter measure does not differs significantly from d_{PCA} and d_{GCC} when the ARI is considered. Overall, Table 18 suggests the existence of two homogeneous groups of metrics, $\{d_{QCD}, d_J\}$ and $\{d_{PCA}, d_{GCC}, d_W\}$, in such a way that the dissimilarities in the first group attained the best overall results in the 21 analysed datasets. This outcome could be already speculated from Table 17. Multiple pairwise comparisons using the paired Wilcoxon signed-rank test were also performed and similar results were obtained after adjusting the p-values with the Holm–Bonferroni correction (which also controls the family-wise error rate but in a more powerful approach than the Bonferroni procedure).

It should be pointed out that the different classes in the studied databases are often presumed to be characterized by means of different geometric profiles. However, even lacking shape-based information, the measures d_{QCD} and d_J are capable of attaining good results in these datasets.

The results obtained throughout this section are insightful, since they allow us to conclude that the distance d_{QCD} can be useful even when dealing with nonstationary multidimensional time series. Therefore, we encourage practitioners in the field of MTS clustering to add the smoothed CCR-periodograms to their toolbox, as these features could display a substantial discriminatory power in real MTS datasets.

7. Concluding remarks and future work

Cluster analysis of time series is a fundamental problem nowadays. We live in a society which is generating huge amounts of data every single day, a vast majority of which have a temporary nature. Whereas clustering of univariate time series (UTS) has been broadly analysed in the last twenty years, giving rise to a large amount of works, clustering of multivariate time series (MTS) has received limited attention. However, given the importance of data such as ECG, EEG, sensor data..., which can be naturally seen as MTS, it is clear that there is a need to delve deeper into the topic. Most works on MTS clustering are based on dimensionality reduction techniques as principal component analysis, thus implying some loss of information. Only a few number of studies present alternative procedures, but many of them do not take into account the interdependence relationship between the components of an MTS, which is a pivotal issue when dealing with MTS.

One of the key points in cluster analysis is deciding which type of dissimilarity to use, which highly determines the nature of the clustering solution. In the present paper, we focused on structure-based dissimilarities, i.e., dissimilarity measures aimed to compare underlying dependence structures. The challenge was to introduce an efficient dissimilarity measure with a high capability of clustering MTS generated from a broad range of dependence models. With this goal in mind, a dissimilarity based on the quantile cross-spectral density, d_{QCD}, was proposed. The great ability of the quantile cross-spectral density to distinguish between different dependence structures was shown by means of a toy example. Although an attractive tool, to the best of our knowledge, the quantile cross-spectral density has not been used before to perform time series clustering.

The proposed dissimilarity d_{QCD} was fairly tested in a broad range of simulated scenarios covering a wide variety of dependence schemes. It was compared with several state of the art dissimilarities offered in the literature. The distance d_{QCD} produced its worst results grouping linear processes. Under these circumstances, it was outperformed by two dissimilarities, one of them being specifically designed to cope with linear models. However, the results attained by d_{QCD} were not much worse than the best ones. Regarding heteroskedastic and quantile autoregression-based processes, the distance d_{QCD} obtained by far the best scores, clearly outpacing the remaining competitors by a large margin. The results have also shown a strong robustness of d_{QCD} when heavy tails are present in the error distribution, a situation which is very common in some domains. In summary, the analyses have proved that d_{QCD} is a robust and versatile dissimilarity that can be used to perform cluster of MTS in many different contexts.

Another interesting property of the quantile cross-spectral density-based metric is that it does not require the existence of any moments and it is computationally inexpensive. Besides, it can be applied to cluster series of unequal length. It must be noted that d_{QCD} requires setting a number of hyperparameters, namely the number of frequencies K, the number of quantile levels r, and the corresponding sets, $\{\omega_1, \dots, \omega_K\}$ and $\{\tau_1, \dots, \tau_r\}$, respectively. Our numerical experiments have shown that a small number of quantiles with probability levels regularly spaced in $[0, 1]$, along with the Fourier frequencies, are enough to reach satisfactory results.

In order to illustrate the usefulness of the proposed methodology in real-life scenarios, it has been applied to clustering a set of bivariate time series containing the daily stock prices and trading volume of some S&P 500 companies pertaining to two different financial sectors. The solution achieved by d_{QCD}, clearly differentiating between the two sectors and including only a small number of misclassifications, is highly interpretable from a financial point of view. We have taken advantage of the high ability of the proposed metric to distinguish between heteroskedastic processes, a task which cannot be successfully performed by other structure-based dissimilarities. In addition, the designed metric was also tested in clustering of nonstationary MTS from the UEA multivariate time series classification archive. The results indicate that d_{QCD} is effective even when the necessary requirements of stationarity are not fulfilled, suggesting that the smoothed CCR-periodograms seen as descriptive features are useful in their own right.

There are two main ways through which this work can be extended. First, the excellent properties exhibited by d_{QCD} in MTS clustering from a crisp point of view call for an extension of the distance to a soft, fuzzy context. This way, the strength of the quantile cross-spectral density and the versatility of the fuzzy logic could be combined in an undoubtedly powerful procedure, which could be probably useful in a broad range of application domains. Second, the high discriminative power demonstrated by the features extracted via the smoothed CCR-periodograms in an unsupervised learning framework suggests the possibility of a great performance also in a supervised learning setting, where these features could be used to feed a traditional classification algorithm as the random forest or the support vector machine. Both paths will be properly addressed in further work.

CRediT authorship contribution statement

Ángel López-Oriona: Conceptualization, Writing - review & editing, Methodology, Software, Data processing, Visualization. **José A. Vilar:** Conceptualizacion, Supervision, Writing and review, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research has been supported by the Ministerio de Economía y Competitividad (MINECO) grants MTM2017-82724-R and PID2020-113578RB-100, the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2020-14), and the Centro de Investigación del Sistema Universitario de Galicia “CITIC” grant ED431G 2019/01; all of them through the European Regional Development Fund (ERDF). This work has received funding for open access charge by Universidade da Coruña/CISUG.

Appendix. Supplementary material

All the code used to perform the analyses described throughout the paper is available in https://github.com/anloor7/PhD_degree/tree/master/r_code/papers/paper_qcd. Please contact the corresponding author for further information.

References

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering–a decade review. *Information Systems*, *53*, 16–38.

Alonso, A. M., & Peña, D. (2019). Clustering time series by linear dependency. *Statistics and Computing*, *29*(4), 655–676.

Andersson, M., Krylova, E., & Vähämaa, S. (2008). Why does the correlation between stock and bond returns vary over time? *Applied Financial Economics*, *18*(2), 139–151.

Bagnall, A., Dau, H. A., Lines, J., Flynn, M., Large, J., Bostrom . . ., A., & Keogh, E. (2018). The UEA multivariate time series classification archive, 2018. (pp. 1–36).

Bandyopadhyay, S., Maulik, U., & Baragona, R. (2010). Clustering multivariate time series by genetic multiobjective optimization. *Metron*, *68*(2), 161–183.

Baruník, J., & Kley, T. (2019). Quantile coherency: A general measure for dependence between cyclical economic variables. *The Econometrics Journal*, *22*(2), 131–152.

Bloomfield, P. (2004). *Fourier analysis of time series: an introduction*. John Wiley & Sons.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Campbell, J. Y., Grossman, S. J., & Wang, J. (1993). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics*, *108*(4), 905–939.

Chen, S.-S. (2012). Revisiting the empirical linkages between stock returns and trading volume. *Journal of Banking & Finance*, *36*(6), 1781–1788.

Daniell, P. J. (1946). Discussion on symposium on autocorrelation in time series. *Journal of the Royal Statistical Society*, *8*(1).

Dette, H., Hallin, M., Kley, T., & Volgushev, S. (2015). Of copulas, quantiles, ranks and spectra: An L_1-approach to spectral analysis. *Bernoulli*, *21*(2), 781–831.

D'Urso, P., De Giovanni, L., Maharaj, E. A., & Massari, R. (2014). Wavelet-based self-organizing maps for classifying multivariate time series. *Journal of Chemometrics*, *28*(1), 28–51.

D'Urso, P., & Maharaj, E. A. (2012). Wavelets-based clustering of multivariate time series. *Fuzzy Sets and Systems*, *193*, 33–61.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Fröhwirth-Schnatter, S., & Kaufmann, S. (2008). Model-based clustering of multiple time series. *Journal of Business & Economic Statistics*, *26*(1), 78–89.

Fu, T.-c. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence*, *24*(1), 164–181.

Gebka, B., & Wohar, M. E. (2013). Causality between trading volume and returns: Evidence from quantile regressions. *International Review of Economics & Finance*, *27*, 144–159.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark . . ., R. G., & Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, *101*(23), e215–e220.

Hagemann, A. (2013). Robust spectral analysis.

He, H., & Tan, Y. (2020). Unsupervised classification of multivariate time series using VPCA and fuzzy clustering with spatial weighted matrix distance. *IEEE Transactions on Cybernetics*, *50*, 1096–1105.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218.

Kakizawa, Y., Shumway, R. H., & Taniguchi, M. (1998). Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association*, *93*(441), 328–340.

Karamitopoulos, L., Evangelidis, G., & Dervos, D. (2010). PCA-based time series similarity search. In R. Stahlbock, S. F. Crone, & S. Lessmann (Eds.), *Data mining: Special issue in annals of information systems* (pp. 255–276). Boston, MA: Springer US.

Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, 109–126.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis, Vol. 344*. John Wiley & Sons.

Keogh, E., & Kasetty, S. (2003). On the need for time series data mining benchmarks: a survey and empirical demonstration. *Data Mining and Knowledge Discovery*, *7*(4), 349–371.

Kley, T. (2016). Quantile-based spectral analysis in an object-oriented framework and a reference implementation in R: The **quantspec** package. *Journal of Statistical Software*, *70*(3), 1–27. http://dx.doi.org/10.18637/jss.v070.i03.

Kley, T., Volgushev, S., Dette, H., & Hallin, M. (2016). Quantile spectral processes: Asymptotic analysis and inference. *Bernoulli*, *22*(3), 1770–1807.

Koenker, R., & Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, *101*(475), 980–990.

Lafuente-Rego, B., & Vilar, J. A. (2016). Clustering of time series using quantile autocovariances. *Advances in Data Analysis and Classification*, *10*(3), 391–415.

Larsen, B., & Aone, C. (1999). Fast and effective text mining using linear-time document clustering. In *Proceedings of the fifth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 16–22).

Lee, J., & Rao, S. S. (2012). The quantile spectral density and comparison based tests for nonlinear time series.

Li, H. (2016). Accurate and efficient classification based on common principal components analysis for multivariate time series. *Neurocomputing*, *171*, 744–753.

Li, H. (2019). Multivariate time series clustering based on common principal component analysis. *Neurocomputing*, *349*, 239–247.

Liao, T. W. (2005). Clustering of time series data: A survey. *Pattern Recognition*, *38*(11), 1857–1874.

Maharaj, E. (1999). Comparison and classification of stationary multivariate time series. *Pattern Recognition*, *32*(7), 1129–1138.

Maharaj, E., D'Urso, P., & Caiado, J. (2019). *Time series clustering and classification*. CRC Press.

Montero, P., & Vilar, J. A. (2014). **TSclust**: An R package for time series clustering. *Journal of Statistical Software*, *62*(1), 1–43.

Montero, P., & Vilar, J. A. (2014b). **TSclust**: Time series clustering utilities. [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=TSclust (R package version 1.2.1).

Nugent, C. (2017). *S&P500 stock data*. Kaggle, https://www.kaggle.com/camnugent/sandp500? (Accesed 15 July 2020).

Priestley, M. B. (1981). *Spectral analysis and time series: probability and mathematical statistics*. (Nos. 04; QA280, P7).

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, [Computer software manual]. Retrieved from https://www.R-project.org/.

Rani, S., & Sikka, G. (2012). Recent techniques of clustering of time series data: A survey. *International Journal of Computer Applications*, *52*(15), 1–9.

Shokoohi-Yekta, M., Hu, B., Jin, H., Wang, J., & Keogh, E. (2017). Generalizing DTW to the multi-dimensional case requires an adaptive approach. *Data Mining and Knowledge Discovery*, *31*(1), 1–31.

Singhal, A., & Seborg, D. E. (2005). Clustering multivariate time-series data. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *19*(8), 427–438.

Vilar, J. A., Lafuente-Rego, B., & D'Urso, P. (2018). Quantile autocovariances: A powerful tool for hard and soft partitional clustering of time series. *Fuzzy Sets and Systems*, *340*, 38–72.

Wang, X., Wirth, A., & Wang, L. (2007). Structure-based statistical features and multivariate time series clustering. In *Seventh IEEE international conference on data mining (ICDM 2007)* (pp. 351–360).

Yang, K., & Shahabi, C. (2004). A PCA-based similarity measure for multivariate time series. In *Proceedings of the 2nd ACM international workshop on multimedia databases* (pp. 65–74). New York, NY, USA: Association for Computing Machinery.

Zagorecki, A. (2015). A versatile approach to classification of multivariate time series data. In *Proceedings of the 2015 federated conference on computer science and information systems, FedCSIS 2015, Vol. 5* (pp. 407–410). Polish Information Processing Society (PIPS).