Quantile-based fuzzy clustering of multivariate time series in the frequency domain

Ángel López-Oriona [a,*], José A. Vilar [a,b], Pierpaolo D'Urso [c]

[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*
[c] *Department of Social Sciences and Economics, Sapienza University of Rome, P. le Aldo Moro 5, Roma, Italy*

Abstract

A novel procedure to perform fuzzy clustering of multivariate time series generated from different dependence models is proposed. Different amounts of dissimilarity between the generating models or changes on the dynamic behaviours over time are some arguments justifying a fuzzy approach, where each series is associated to all the clusters with specific membership levels. Our procedure considers quantile-based cross-spectral features and consists of three stages: (i) each element is characterized by a vector of proper estimates of the quantile cross-spectral densities, (ii) principal component analysis is carried out to capture the main differences reducing the effects of the noise, and (iii) the squared Euclidean distance between the first retained principal components is used to perform clustering through the standard fuzzy C-means and fuzzy C-medoids algorithms. The performance of the proposed approach is evaluated in a broad simulation study where several types of generating processes are considered, including linear, nonlinear and dynamic conditional correlation models. Assessment is done in two different ways: by directly measuring the quality of the resulting fuzzy partition and by taking into account the ability of the technique to determine the overlapping nature of series located equidistant from well-defined clusters. The procedure is compared with the few alternatives suggested in the literature, substantially outperforming all of them whatever the underlying process and the evaluation scheme. Two specific applications involving air quality and financial databases illustrate the usefulness of our approach.
© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords: Multivariate time series; Clustering; Quantile cross-spectral density; Fuzzy C-means; Fuzzy C-medoids; Principal component analysis

1. Introduction

Time series clustering is a pivotal problem in data mining with applications in a wide variety of fields, including machine learning, economics, finance, physics, computer science, medicine, biology, geology, among others. The goal

is to split a large set of unlabeled 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 detect 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 during the last two decades. Comprehensive overviews including current advances, future prospects, significant references and specific application areas are provided by [1–4], and more recently in the monograph by [5]. 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 grouping 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 dimensions might be high in some clusters but non-significant in others. Additionally, 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.

There exist two important decisions to be made in any clustering problem, namely the notion of dissimilarity between the elements subject to the grouping problem and the clustering algorithm determining how the assignment of objects to the groups is done. A proper dissimilarity measure must be totally dependent on the nature and specific purpose of the clustering task, thus providing a meaningful clustering solution according to the grouping target. If the goal 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 dependence models. In the latter case, the clustering performance may be seriously affected by noise, change in the conditional variance or heavy-tailed errors, and hence distance 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 [6] and many of them are implemented in the R package **TSclust** [7].

Regarding how the assignment of the elements to the clusters is made, there are two classical paradigms which are usually referred to as "hard" and "soft" clustering. The partition provided by hard clustering procedures locates each data object in exactly one cluster, thus being constituted by disjoint subsets. This approach does not allow for overlapping clusters and could become too inflexible in some scenarios. For instance, hard clustering is incapable of giving insights into which elements are equidistant from two or more clusters and also to account for a closer alignment to patterns of other clusters due to changes in the dynamic of the series over time. On the other hand, fuzzy clustering strategies [8,9] provide a more versatile approach to address the clustering task. They rely on the notion of membership of an element in a given cluster, which indicates the degree of confidence in that particular assignment. Therefore, the output of these methods is a soft partition where the objects can belong to several groups with specific membership degrees.

Several works have considered fuzzy clustering of time series, specially in the univariate setting. [10] proposed a fuzzy clustering approach based on estimates of the autocorrelation function of the time series up to a given lag. The corresponding quantities are used as input to the traditional fuzzy C-means algorithm. [11] provided three different methods relying on different characteristics computed from the UTS, namely the periodogram, the normalized periodogram and the logarithm of the normalized periodogram. A fuzzy approach to the clustering of UTS based on estimated wavelet variances is presented in [12], proving itself capable of identifying time series with switching patterns in terms of variability. [13] developed an approach focused on grouping together time series with similar seasonal structures using extreme value analysis. The input features in the fuzzy clustering algorithm are parameter estimates of time varying location, scale and shape obtained by means of a fitting of the generalised extreme value distribution. [14] developed two fuzzy clustering strategies aimed at clustering financial time series. The first approach employs the autoregressive representation of GARCH models whereas the second one is based on estimates of GARCH parameters. A method for grouping heteroskedastic time series was designed by [15]. The approach assumes that the series follow a GARCH model. Estimates of the so-called unconditional volatility and time-varying volatility are obtained and then used to feed the classical fuzzy C-medoids technique. Robust alternatives of the method are also provided. [16] extended their work in [17] by introducing a procedure which employs estimates of the quantile autocovariance function in order to perform fuzzy clustering of UTS by means of the fuzzy C-medoids algorithm.

The method takes advantage of the nice properties of the quantile autocovariance function as robustness to heavy tails or no requirements about the existence of moments. Robust approaches derived from this technique are developed in [18]. [19] introduced a copula-based procedure for finding groups in spatial time series. The methodology obtains the empirical copula for a given MTS and computes a dissimilarity between this copula end the Frechet copula by taking into account also spatial information. Note that all the previous mentioned approaches pertain to the so-called feature-based clustering, which exploits specific features extracted from a time series, or to the termed model-based clustering, which groups the time series based on estimates of the parameters or of the residuals of a given model.

Approaches based on direct computation of a given distance have also been suggested for fuzzy clustering of UTS. [20] designed an approach employing the dynamic time warping distance (DTW) in the standard fuzzy C-means algorithm. A specific procedure to obtain the average of a set of UTS according to DTW is utilised. A trimmed fuzzy clustering technique for financial time series relying on DTW and the fuzzy C-medoids algorithm is developed in [21]. The method is designed so as to attain robustness against outlying elements by trimming away the series which are more distant from the bulk of the data.

By contrast, a fewer number of papers have dealt with fuzzy clustering of MTS. [22] analysed different dissimilarity measures between multivariate time trajectories, which are classified in three classes according to their characteristics. [23] constructed different unsupervised clustering models for multivariate data time arrays which take into account geometric and trend patterns of the multivariate time arrays, among others. [24] introduced three fuzzy C-means clustering approaches of multivariate time trajectories considering the so-called positional dissimilarity, velocity dissimilarity and a mixture of both. These models are particularly beneficial when anomalous trajectories are present in the dataset. [25] proposed to perform unsupervised classification of MTS by means of a novel objective function containing two terms, one taking into account longitudinal features of the MTS and the other considering the Shannon entropy measure concerning fuzzy partitions. A procedure based on the maximum overlap discrete wavelet transform is provided in [26]. Estimates of wavelet variances and correlations are computed from each MTS and used to feed the fuzzy C-means, fuzzy C-medoids and fuzzy relational clustering algorithms. [27] devised an approach where each MTS sample is treated as matrix data. First, a dimensionality reduction technique is applied over the original MTS dataset, and then a fuzzy clustering algorithm based on spatial weighted distance matrix is performed on the reduced dataset. Four different robust clustering models for MTS aimed at neutralizing the negative effects of outliers are developed in [28]. All of them take into account the exponential transformation, which assigns "small weights" to outliers in the clustering process, hence achieving the desired robustness. [29] designed a sophisticated feature-weighted clustering method based on two dissimilarity measures, DTW and the named shape-based distance. The approach consists of several steps in which the contribution of each independent dimension to the overall clustering process is considered. Note that, except for [26], all the previously cited works on fuzzy clustering of MTS are not suitable for making the grouping in terms of underlying dependence structures, since they are aimed to measure dissimilarity in shape. Thus, it is clear that there is a need for developing fuzzy clustering approaches capable of addressing the task from the former perspective.

This paper is aimed at evaluating the performance of a distance measure based on the quantile cross-spectral density (QCD) in fuzzy clustering of MTS. Our assumption is that the goal consists in grouping the series in terms of their generating processes, that is, we consider two MTS to be similar if their underlying dependence patterns are alike. Note that this premise is quite realistic if, for instance, one intends to detect the temporal pattern driving a time series which is observed in practice under large amounts of noise. Under this criterion, a metric capable of effectively discriminating between different generating mechanisms while displaying a large degree of robustness to the type of underlying processes is undoubtedly needed in order to attain a meaningful clustering solution. Indeed, QCD always exists under the assumption of strictly stationarity and allows to account for complex forms of dependence that other time series features as traditional autocovariances are unable to uncover [30]. In addition, QCD takes advantage of automatically considering all the lags. These nice properties insinuate that a dissimilarity measure based on QCD can achieve great results in grouping MTS according to the stated goal. In our previous work [31], we decided to answer this question by analysing the performance of a distance measure termed d_{QCD} in different scenarios of MTS clustering through a comprehensive simulation study. The results showed that d_{QCD} is highly competitive when grouping linear processes and significantly outperforms alternative dissimilarities suggested in the literature when dealing with processes showing a high amount of heteroskedascity or complicated types of dependence. In addition, d_{QCD} exhibited a substantial degree of robustness against changes in the distributional form of the error terms. The distance was also evaluated in [31] with real datasets contained in the well-known UEA multivariate time series

classification archive [32], which consists of 30 MTS databases covering a wide range of cases, dimensions and series lengths. The results indicated that d_{QCD} is one of the best performing metrics also in a practical framework.

The main contribution of this work consists of proposing a novel fuzzy procedure for MTS clustering taking advantage of the high capability of the quantile cross-spectral density to characterize any type of serial dependence structure. Specifically, we take into consideration a modified version of the distance d_{QCD} given in [31]. While d_{QCD} is directly constructed from estimates of QCD, the alternative metric we propose here considers the principal component analysis (PCA) transformation over the mentioned estimates. This way, a lot of the noise contained in these (correlated) estimates is removed and the most important information is retained, thus getting improved performance in comparison with the original distance while inheriting all its advantageous characteristics. The new distance is utilised in both the standard fuzzy C-means and the fuzzy C-medoids algorithms. Assessment of the proposed approaches is carried out by means of an extensive simulation study including linear, nonlinear and conditional heteroskedastic processes. The dissimilarity proposed in [26], a natural extension of that introduced in [10] and the distance constructed by [33] are also analysed for comparison purposes. Two evaluation schemes are considered. The first one is aimed at examining the capability of the procedures in assigning high (low) membership values if a given series pertains (not pertains) to a specific cluster defined in advance. The second scheme also analyses the ability of the approaches to handle outlying series. Lastly, two specific applications to multivariate financial and environmental datasets are presented to highlight the usefulness of the proposed clustering technique.

The remainder of this paper is structured as follows. Section 2 presents d_{QCD}, a dissimilarity measure between a pair of MTS that takes into consideration proper estimates of QCD. The estimation procedure is detailed and some powerful properties of the metric holding under very general conditions are highlighted. In Section 3, the dissimilarity is considered to develop two novel fuzzy clustering approaches relying on the traditional fuzzy C-means and fuzzy C-medoids algorithms. A direct modification of the distance d_{QCD} based on the PCA transformation is proposed. The new measure considers the transformed QCD-based features concerning the principal components space. Advantages of this metric in comparison with the original dissimilarity are shown by means of a toy example. The dissimilarity based on QCD and PCA is used to perform fuzzy clustering in Section 4. Three scenarios characterised by the kind of generating process are considered, namely linear, nonlinear and dynamic conditional correlation. The assessment task is performed in a fair and general manner, and the results are compared with those obtained using alternative dissimilarity measures. Section 5 contains applications to real datasets and some concluding remarks are summarized in Section 6.

2. A distance measure based on the quantile cross-spectral density

Consider a set of n multivariate time series $\mathcal{S} = \left\{ \boldsymbol{X}_t^{(1)}, \ldots, \boldsymbol{X}_t^{(n)} \right\}$, where the j-th element $\boldsymbol{X}_t^{(j)} = \left\{ \boldsymbol{X}_1^{(j)}, \ldots, \boldsymbol{X}_{T_j}^{(j)} \right\}$ 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 clustering algorithm 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 $\{X_t, t \in \mathbb{Z}\} = \{(X_{t,1}, \ldots, X_{t,d}), t \in \mathbb{Z}\}$ be a d-variate real-valued strictly stationary stochastic process. Denote by F_j the marginal distribution function of $X_{t,j}$, $j = 1, \ldots, d$, 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

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

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 *quantile autocovariance function* (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 joint distribution 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 [17] to perform UTS clustering with very satisfactory results.

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

$$\mathbf{\Gamma}(l, \tau, \tau') = \big(\gamma_{j_1,j_2}(l, \tau, \tau')\big)_{1 \le j_1, j_2 \le d}, \tag{2}$$

which jointly provides information about both the cross-dependence (when $j_1 \ne j_2$) and the serial dependence (because the lag l is considered). To obtain a much richer picture of the underlying dependence structure, $\mathbf{\Gamma}(l, \tau, \tau')$ 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

$$\mathbf{\Gamma}_{X_t}(\mathcal{L}, \mathcal{T}) = \big\{\mathbf{\Gamma}(l, \tau, \tau'), \ l \in \mathcal{L}, \ \tau, \tau' \in \mathcal{T}\big\}. \tag{3}$$

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* (QCD) 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 $1 \le j_1, j_2 \le d, \omega \in \mathbb{R}$ 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

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

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

$$\mathfrak{f}_{X_t}(\Omega, \mathcal{T}) = \big\{\mathfrak{f}(\omega, \tau, \tau'), \ \omega \in \Omega, \ \tau, \tau' \in \mathcal{T}\big\}, \tag{6}$$

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

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

Representing X_t through $\mathfrak{f}_{X_t}$, a 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 [34], [35] and [30], 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 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

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

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

Let $\{X_1, \ldots, X_T\}$ be a realization from the process $(X_t)_{t \in \mathbb{Z}}$ so that $X_t = (X_{t,1}, \ldots, X_{t,d})$, $t = 1, \ldots, T$. For arbitrary $j_1, j_2 \in \{1, \ldots, d\}$ and $(\tau, \tau') \in [0, 1]^2$, [30] 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 [36] 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 $d_{T,R}^{j}(\omega, \tau) = \sum_{t=1}^{T} I\{\hat{F}_{T,j}(X_{t,j}) \leq \tau\} e^{-i\omega t}$.

The asymptotic properties of the CCR-periodogram are established in Proposition S4.1 of [30]. Likewise the standard cross-periodogram, the CCR-periodogram is not a consistent estimator of $\mathfrak{f}_{j_1, j_2}(\omega, \tau, \tau')$ [30]. 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

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

where $W_T(u) = \sum_{v=-\infty}^{\infty} \frac{1}{h_T} W\left(\frac{u + 2\pi v}{h_T}\right)$, with $h_T > 0$ a sequence of bandwidths such that $h_T \to 0$ and $T h_T \to \infty$ 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 $\hat{G}_{T,R}^{j_1, j_2}(\omega, \tau, \tau')$ are established in Theorem S4.1 of [30].

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

$$\hat{\mathfrak{f}}_{X_t}(\Omega, \mathcal{T}) = \left\{ \hat{\mathfrak{f}}(\omega, \tau, \tau'), \ \omega \in \Omega, \ \tau, \tau' \in \mathcal{T} \right\},$$

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

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

Throughout this article, the smoothed CCR-periodograms were obtained by using the R-package **quantspec** [37].

2.2. A spectral dissimilarity measure between MTS

A simple dissimilarity criterion between a pair of d-variate time series $X_t^{(1)}$ and $X_t^{(2)}$ can be obtained by comparing their estimated sets of complex-valued matrices $\hat{\mathfrak{f}}_{X_t^{(1)}}(\Omega, \mathcal{T})$ and $\hat{\mathfrak{f}}_{X_t^{(2)}}(\Omega, \mathcal{T})$ evaluated on a common range of frequencies and quantile levels. Specifically, each time series $X_t^{(u)}$, $u = 1, 2$, is characterized by means of a set of d^2 vectors $\{\boldsymbol{\Psi}_{j_1, j_2}^{(u)}, 1 \leq j_1, j_2 \leq d\}$ 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 $\boldsymbol{\Psi}_{j_1, j_2}^{(u)}$ is given by

$$\boldsymbol{\Psi}_{j_1, j_2}^{(u)} = (\boldsymbol{\Psi}_{1, j_1, j_2}^{(u)}, \ldots, \boldsymbol{\Psi}_{K, j_1, j_2}^{(u)}), \tag{11}$$

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

$$\left(\hat{G}_{T,R}^{j_1, j_2}(\omega_k, \tau_i, \tau_{i'}) \right)_{1 \leq i, i' \leq r}. \tag{12}$$

Once the set of d^2 vectors $\boldsymbol{\Psi}_{j_1, j_2}^{(u)}$ is obtained, they are all concatenated in a vector $\boldsymbol{\Psi}^{(u)}$ in the same way as vectors $\boldsymbol{\Psi}_{k, j_1, j_2}^{(u)}$ constitute $\boldsymbol{\Psi}_{j_1, j_2}^{(u)}$ in (11). In this manner, the dissimilarity between $X_t^{(1)}$ and $X_t^{(2)}$ is obtained by means of the Euclidean distance between the complex vectors $\boldsymbol{\Psi}^{(1)}$ and $\boldsymbol{\Psi}^{(2)}$

$$d_{QCD}(X_t^{(1)}, X_t^{(2)}) = \left[||\Re_v(\mathbf{\Psi}^{(1)}) - \Re_v(\mathbf{\Psi}^{(2)})||^2 + ||\Im_v(\mathbf{\Psi}^{(1)}) - \Im_v(\mathbf{\Psi}^{(2)})||^2\right]^{1/2} =$$

$$\left[\sum_{j_1=1}^{d}\sum_{j_2=1}^{d}\sum_{i=1}^{r}\sum_{i'=1}^{r}\sum_{k=1}^{K}\left(\Re(\hat{G}_{T,R}^{j_1,j_2}(\omega_k, \tau_i, \tau_{i'})^{(1)}) - \Re(\hat{G}_{T,R}^{j_1,j_2}(\omega_k, \tau_i, \tau_{i'})^{(2)})\right)^2 + \right.$$

$$\left.\sum_{j_1=1}^{d}\sum_{j_2=1}^{d}\sum_{i=1}^{r}\sum_{i'=1}^{r}\sum_{k=1}^{K}\left(\Im(\hat{G}_{T,R}^{j_1,j_2}(\omega_k, \tau_i, \tau_{i'})^{(1)}) - \Im(\hat{G}_{T,R}^{j_1,j_2}(\omega_k, \tau_i, \tau_{i'})^{(2)})\right)^2\right]^{1/2},$$

(13)

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

Computation of vectors $\mathbf{\Psi}^{(1)}, \dots, \mathbf{\Psi}^{(n)}$ for every MTS in the set S could be used to perform fuzzy clustering in S by means of an algorithm as fuzzy C-means or fuzzy C-medoids considering the distance d_{QCD}. This distance has been successfully applied to perform clustering on MTS in a crisp framework [31], and the corresponding QCD-based features, to develop classification [38] and outlier detection [39] procedures.

2.3. Properties of d_{QCD}

Now we present some important properties of the distance d_{QCD} which make it a very powerful dissimilarity to perform clustering of MTS.

In the following, we assume that X_t^i is a d-variate, real-valued, strictly stationary process and $X_t^{(i)}$ is a realization of length T from the process X_t^i. The j-th component of X_t^i, $j = 1, \dots, d$, is denoted by $X_{t,j}^i$. The notation F_j^i stands for the marginal cumulative distribution function of $X_{t,j}^i$. Given a lag $l \in \mathbb{Z}$ and a couple of components $j_1, j_2 = 1, \dots, d$, the joint cumulative distribution function of the pair $(X_{t,j_1}^i, X_{t+l,j_2}^i)$ is denoted by $F_{j_1,j_2,l}^i$. We suppose that all the mentioned cumulative distribution functions are continuous functions. Now we state the following properties.

Property 1. If $X_t^1 = X_t^2$. Then $d_{QCD}(X_t^{(1)}, X_t^{(2)}) \underset{p}{\to} 0$ as $T \to \infty$, where the notation $\underset{p}{\to}$ stands for convergence in probability.

Property 2. Assume that there exists some $l \in \mathbb{Z}$ and a couple of dimensions $j_1, j_2 = 1, \dots, d$ such that $F_{j_1,j_2,l}^1 \neq F_{j_1,j_2,l}^2$ and that $F_j^1 = F_j^2$, $j = 1, \dots, d$. Then there exist an infinite number of probability levels and an infinite number of frequencies such that $d_{QCD}(X_t^{(1)}, X_t^{(2)}) \underset{p}{\to} a$, $a \neq 0$ as $T \to \infty$.

The proofs of the previous properties are deferred to the Appendix A so as not to impair the flow of the paper. Some remarks about the results are given below.

Remark 1. Property 1 is a desirable characteristic of any dissimilarity measure aimed at performing clustering of MTS based on underlying dependence patterns. Indeed, it tells us that the distance between two MTS generated from the same process is expected to be negligible for a sufficiently large value of the series length. The majority of the dissimilarity measures suggested in the literature for UTS or MTS clustering according to the stated goal have this property.

Remark 2. Property 2 assumes that the marginal distribution function of the j-th component of X_t^1 is equal to the marginal distribution function of the j-th component of X_t^2. This can be supposed without loss of generality. Indeed, if $F_{j'}^1 \neq F_{j'}^2$ for some $j' \in \{1, \dots, d\}$, then the quantile cross-covariance in (1) regarding $l = 0$ and $j_1 = j_2 = j'$ is going to be different between both processes for some pair of probability levels $(\tau, \tau') \in [0, 1]^2$. It is not difficult to derive (see (4)) that this discrepancy is transmitted to the corresponding smoothed CCR-periodograms, resulting in the convergence in probability of the respective distance between two realizations of the processes to a quantity distinct from zero.

Remark 3. Property 2 can be directly extended to more than two processes. In fact, given a collection of N processes X_t^b and the corresponding realizations $X_t^{(b)}$, $b = 1, \ldots, N$, if we assume that (i) for every pair (i, i'), $i, i' \in \{1, \ldots, N\}$, there exist j_i, $j_{i'}$ and $l_{i,i'} \in \mathbb{Z}$ such that $F_{j_i,j_{i'},l_{i,i'}}^i \neq F_{j_i,j_{i'},l_{i,i'}}^{i'}$, (ii) every couple of functions of the form $F_{j_k,j_{k'},l_{k,k'}}^k$, $k, k' \in \{1, \ldots, N\}$ are different from one another, and (iii) $F_j^1 = F_j^2 = \ldots = F_j^N$, $j = 1, \ldots, d$, then there exist an infinite number of probability levels, an infinite number of frequencies, and a set of different real numbers $\{a_{q,q'}\}$, $q, q' = 1, \ldots, N, q \neq q'$ such that $d_{QCD}(X_t^{(q)}, X_t^{(q')}) \underset{p}{\rightarrow} a_{q,q'} \neq 0$ as $T \to \infty$. Note that this is often the case in practice when we apply cluster analysis through d_{QCD} to a set of MTS coming from more than two different generating processes.

Remark 4. Property 2 is perhaps the most important characteristic of the distance d_{QCD}. Broadly speaking, it tells us that, under appropriate conditions, the metric is able to capture even the slightest change in the dependence structure between two generating processes. This trait is not shared, to the best of our knowledge, by any of the metrics proposed in the literature for clustering of UTS or MTS based on dependence patterns. Generally, it is easy to find a counterexample where two MTS are produced from distinct generating processes but the corresponding dissimilarities fail to detect any difference between the corresponding realizations even for very large values of the series lengths. In fact, our previous work [31] provides an insightful example of this happening. There we simulated bivariate MTS from three types of the so-called QVAR processes. This kind of processes are capable of generating rich forms of quantile dependence while keeping uncorrelatedness within and between components. The results shown in that work determined that most of the alternative metrics were totally unable to unmask the underlying patterns (see the bottom panel in Figure 3 of [31]), whereas the distance d_{QCD} perfectly discriminated between processes. This property is due to the relationship between QCD, the copula and the marginal distributions of a given pair $(X_{t,j_1}^i, X_{t+l,j_2}^i)$, which gives d_{QCD} the ability to capture any kind of deviation in the dependence structure of the stochastic processes.

3. Fuzzy clustering methods based on the quantile cross-spectral density and PCA

In this section we introduce two fuzzy clustering procedures based on the distance d_{QCD}, namely the QCD-based fuzzy C-means clustering model and the QCD-based fuzzy C-medoids clustering model, and show how the effectiveness of these models can be significantly improved by applying the PCA transformation over the corresponding QCD-based features and performing clustering in the transformed space.

3.1. QCD-based fuzzy C-means clustering model (QCD-FCMn)

As in previous sections, consider a set S of n realizations of multivariate time series $\{X_t^{(1)}, \ldots, X_t^{(n)}\}$ and denote by $\boldsymbol{\Psi} = \{\boldsymbol{\Psi}^{(1)}, \ldots, \boldsymbol{\Psi}^{(n)}\}$ the corresponding vectors of estimated quantile cross-spectral densities obtained as indicated in Section 2.2. Assume that all vectors $\boldsymbol{\Psi}^{(i)}$ have the same length, $2d^2r^2(\lfloor T/2 \rfloor + 1)$, being d the number of dimensions, r the number of probability levels, T the series length and $\lfloor \cdot \rfloor$ the floor function. In this framework, we propose to perform partitional fuzzy clustering on S by using the QCD-based fuzzy C-means clustering model (QCD-FCMn), whose aim is to find a set of centroids $\overline{\boldsymbol{\Psi}} = \{\overline{\boldsymbol{\Psi}}^{(1)}, \ldots, \overline{\boldsymbol{\Psi}}^{(C)}\}$, and the $n \times C$ matrix of fuzzy coefficients, $\boldsymbol{U} = (u_{ic})$, $i = 1, \ldots, n, c = 1, \ldots, C$, which define the solution of the minimization problem

$$
\begin{cases}
\min_{\overline{\boldsymbol{\Psi}}, \boldsymbol{U}} \sum_{i=1}^n \sum_{c=1}^C u_{ic}^m \left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c)} \right\|^2 \\
\text{with respect to the constraints:} \\
\sum_{c=1}^C u_{ic} = 1, \, u_{ic} \geq 0 \qquad \text{(constraints on the membership degree)}, \\
\overline{\boldsymbol{\Psi}}_L \leq \overline{\boldsymbol{\Psi}}_k^{(c)} \leq \overline{\boldsymbol{\Psi}}_U \qquad \text{(possible constraints on } \overline{\boldsymbol{\Psi}}_k^{(c)}),
\end{cases}
\tag{14}
$$

where $u_{ic} \in [0, 1]$ represents the membership degree of the i-th series in the c-th cluster, $\overline{\boldsymbol{\Psi}}^{(c)}$ is the vector of estimated quantile cross-spectral densities with regards to the centroid series for the cluster c, $m > 1$ is a parameter controlling

---

Algorithm 1 The QCD-based fuzzy C-means clustering algorithm (QCD-FCMn).

1: Fix C, m, *max.iter*, *tol*, a matrix norm $\|\cdot\|_M$
2: Set $iter = 0$
3: Initialize the membership matrix $\boldsymbol{U} = \boldsymbol{U}^{(0)}$
4: **repeat**
5:     Set $\boldsymbol{U}_{\text{OLD}} = \boldsymbol{U}$ ▷ Store the current membership matrix
6:     Compute the centroids $\overline{\boldsymbol{\Psi}}^{(c)}$, $c = 1, \ldots, C$, by means of (16)
7:     Compute u_{ic}, $i = 1, \ldots, n$, $c = 1, \ldots, C$, using (15)   ▷ Update the membership matrix
8:     $iter \leftarrow iter + 1$
9: **until** $\|\boldsymbol{U} - \boldsymbol{U}_{\text{OLD}}\|_M < tol$ or $iter = max.iter$

---

the fuzziness of the partition, usually referred to as fuzziness parameter, and $\overline{\boldsymbol{\Psi}}_k^{(c)}$ is the k-th component of centroid $\overline{\boldsymbol{\Psi}}^{(c)}$. Constraints on u_{ic} are standard requirements in fuzzy clustering. Specifically, that the sum of the membership degrees for each series equals one implies that all of them contribute with the same weight to the clustering process. The fuzziness parameter controls the level of fuzziness considered in the grouping procedure. In the naive case, when $m = 1$, we have $u_{ic} = 1$ if $c = \underset{c' \in \{1, \ldots, C\}}{\arg\min} \, d_{QCD}^2(\boldsymbol{\Psi}^{(i)}, \overline{\boldsymbol{\Psi}}^{(c')})$ and 0 otherwise so that the crisp version of the procedure is obtained. As the value of m increases, the boundaries between clusters become softer and therefore the grouping is fuzzier. Note that the centroid of a cluster is the mean of all points (in this case, the quantile cross-spectral features describing the MTS), weighted by the degree of belonging to the cluster. Hence, we can think of the centroids as the prototypes of each cluster, i.e., a time series or a feature vector artificially representing the characteristics of the time series belonging to the corresponding cluster with a high membership degree. In (14), $\overline{\boldsymbol{\Psi}}_L$ an $\overline{\boldsymbol{\Psi}}_U$ stand for the possible lower and upper bound of $\overline{\boldsymbol{\Psi}}_k^{(c)}$, respectively.

The goal of QCD-FCMn is to find a fuzzy partition into C clusters such that the squared QCD-distance between the clusters and their prototypes is minimized.

The quality of the clustering solution strongly depends on the capability of the distance d_{QCD} to identify different dependence structures. Note that, unlike in a crisp clustering procedure, here the non-stochastic uncertainty inherent to the assignment of series to clusters is incorporated to the procedure by means of the membership degrees.

By taking into consideration only the membership degree constraints, the constrained optimization problem in (14) can be solved by means of the Lagrangian multipliers method, given rise to a two-step iterative process. The first step consists of the minimization of the objective function with respect to u_{ic}, being $\overline{\boldsymbol{\Psi}}$ fixed.

$$u_{ic} = \left[\sum_{c'=1}^{C} \left(\frac{\left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c)} \right\|^2}{\left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c')} \right\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}, \text{ for } i = 1, \ldots, n \text{ and } c = 1, \ldots, C. \tag{15}$$

The second step is based on the minimization of the objective function regarding $\overline{\boldsymbol{\Psi}}$, being u_{ic} fixed

$$\overline{\boldsymbol{\Psi}}_k^{(c)} = \frac{\sum_{i=1}^{n} u_{ic}^m \boldsymbol{\Psi}_k^{(i)}}{\sum_{i=1}^{n} u_{ic}^m}, \text{ for } k = 1, \ldots, d^2 r^2 (\lfloor T/2 \rfloor + 1) \text{ and } c = 1, \ldots, C. \tag{16}$$

Note that, in the previous iterative solutions, $\overline{\boldsymbol{\Psi}}_k^{(c)}$ already verifies that $\overline{\boldsymbol{\Psi}}_L \leq \overline{\boldsymbol{\Psi}}_k^{(c)} \leq \overline{\boldsymbol{\Psi}}_U$, since it inherits the possible constraints of the estimated quantile cross-spectral density features from the observed series, $\boldsymbol{\Psi}^{(i)}$ ($i = 1, \ldots, n$), i.e., $\overline{\boldsymbol{\Psi}}_L \leq \boldsymbol{\Psi}_k^{(i)} \leq \overline{\boldsymbol{\Psi}}_U$. Indeed, from the previous inequality we have $\sum_{i=1}^{n} u_{ic}^m \overline{\boldsymbol{\Psi}}_L \leq \sum_{i=1}^{n} u_{ic}^m \boldsymbol{\Psi}_k^{(i)} \leq \sum_{i=1}^{n} u_{ic}^m \overline{\boldsymbol{\Psi}}_U$, which, by dividing by $\sum_{i=1}^{n} u_{ic}^m$, yields $\overline{\boldsymbol{\Psi}}_L \leq \overline{\boldsymbol{\Psi}}_k^{(c)} \leq \overline{\boldsymbol{\Psi}}_U$. An outline of the QCD-FCMn clustering algorithm is shown in Algorithm 1.

3.2. QCD-based fuzzy C-medoids clustering model (QCD-FCMd)

One natural alternative to QCD-FCMn is the QCD-based fuzzy C-medoids clustering model (QCD-FCMd). Following the context of Section 3.1, the goal is now to find the subset of $\boldsymbol{\Psi}$ of size C, $\widetilde{\boldsymbol{\Psi}} = \{\widetilde{\boldsymbol{\Psi}}^{(1)}, \ldots, \widetilde{\boldsymbol{\Psi}}^{(C)}\}$, and the $n \times C$ matrix of fuzzy coefficients, $\boldsymbol{U}$, solving the minimization problem

---

Algorithm 2 The QCD-based fuzzy C-medoids clustering algorithm (QCD-FCMd).

1: Fix C, m and *max.iter*
2: Set $iter = 0$
3: Pick the initial medoids $\widetilde{\mathbf{\Psi}} = \{\widetilde{\mathbf{\Psi}}^{(1)}, \ldots, \widetilde{\mathbf{\Psi}}^{(C)}\}$
4: **repeat**
5:     Set $\widetilde{\mathbf{\Psi}}_{\text{OLD}} = \widetilde{\mathbf{\Psi}}$ ▷ Store the current medoids
6:     Compute u_{ic}, $i = 1, \ldots, n$, $c = 1, \ldots, C$, using (18)
7:     For each $c \in \{1, \ldots, C\}$, determine the index $j_c \in \{1, \ldots, n\}$ satisfying:

$$j_c = \underset{1 \le j \le n}{\arg\min} \sum_{i=1}^{n} u_{ic}^m \left\| \mathbf{\Psi}^{(i)} - \mathbf{\Psi}^{(j)} \right\|^2$$

8:     **return** $\widetilde{\mathbf{\Psi}}^{(c)} = \mathbf{\Psi}^{(j_c)}$, for $c = 1, \ldots, C$ ▷ Update the medoids
9:       $iter \leftarrow iter + 1$
10: **until** $\widetilde{\mathbf{\Psi}}_{\text{OLD}} = \widetilde{\mathbf{\Psi}}$ or $iter = max.iter$

$$\min_{\widetilde{\mathbf{\Psi}}, U} \sum_{i=1}^{n} \sum_{c=1}^{C} u_{ic}^m \left\| \mathbf{\Psi}^{(i)} - \widetilde{\mathbf{\Psi}}^{(c)} \right\|^2 \text{ with respect to } \sum_{c=1}^{C} u_{ic} = 1 \text{ and } u_{ic} \ge 0, \tag{17}$$

where $\widetilde{\mathbf{\Psi}}^{(c)}$ is the vector of estimated quantile cross-spectral density with regards to the medoid series for the cluster c and the remaining elements are the same as in (14).

By solving the constrained optimization problem in (17) with the Lagrangian multipliers method, we can obtain an iterative algorithm that alternately optimizes the membership degrees and the medoids. The iterative solution for the membership degrees takes the form ([40])

$$u_{ic} = \left[\sum_{c'=1}^{C} \left(\frac{\left\| \mathbf{\Psi}^{(i)} - \widetilde{\mathbf{\Psi}}^{(c)} \right\|^2}{\left\| \mathbf{\Psi}^{(i)} - \widetilde{\mathbf{\Psi}}^{(c')} \right\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}, \text{ for } i = 1, \ldots, n \text{ and } c = 1, \ldots, C. \tag{18}$$

Once the membership degrees are obtained through (18), the C series minimizing (17) are selected as new medoids. This two-step procedure is iterated until there is no change in the medoids or a maximum number of iterations is achieved. An outline of the QCD-FCMd clustering algorithm is given in Algorithm 2.

3.3. Effectiveness of combining the quantile cross-spectral density and PCA

This section illustrates the advantages of applying the PCA transformation over the features obtained through QCD in order to perform fuzzy cluster analysis.

Consider two bivariate VAR processes with matrices of coefficients given by $\begin{pmatrix} 0.2 & 0.2 \\ 0.2 & 0.2 \end{pmatrix}$ and $\begin{pmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{pmatrix}$, which we will refer to as VAR 1 and VAR 2, respectively, and a bivariate white noise process. In the three cases, we assume that the error vector is formed by independent observations following a standard bivariate normal distribution. We simulated 5 MTS of length $T = 500$ from each one the processes, giving rise to a set of 15 MTS. The QCD-based features, that is, the vectors in (11), were extracted from each series in order to perform fuzzy clustering. The set of probability levels $\mathcal{T} = \{0.1, 0.5, 0.9\}$ was considered. This set if often enough for the quantile-based features to give a meaningful description of the underlying process [16–18,31,38,39]. Additionally, the PCA transformation was applied over the matrix containing the QCD-based feature vectors as rows. The so-called matrix of scores (giving the position of each observation in the new coordinate system) was obtained and the first $\lfloor 0.12p \rfloor$ principal components were retained, being p the total number of principal components, 15 in this example. The rate 0.12 was chosen due to the fact that it proved very effective in retaining enough information about the original dataset while removing most of the noise for classification tasks (see [38]). In fact, we have also tried other values for this parameter, getting worse results in terms of clustering accuracy. The QCD-FCMn algorithm (Algorithm 1) was applied over the original features and over the matrix of scores by setting $C = 3$. Four values for the fuzziness parameter m were considered, namely $m = 1.5, 1.8, 2, 2.2$. This mechanism was repeated 100 times.

The assessment of both approaches was carried out by means of the fuzzy extension of the Adjusted Rand Index (see Section 4.2.1), denoted by FARI. This performance metric is specifically designed to evaluate a fuzzy clustering partition, is bounded between -1 and 1, and the closer to one its value, the better the clustering partition. The partition

Table 1
Average values of FARI for the fuzzy C-means procedures
QCD and QCD-PCA according to the fuzziness parameter. For
each value of m, the best result is shown in bold.

Fuzziness parameter	QCD	QCD-PCA
$m = 1.5$	0.747	**0.960**
$m = 1.8$	0.511	**0.900**
$m = 2.0$	0.402	**0.848**
$m = 2.2$	0.322	**0.792**

Table 2
Average maximum membership degrees for fuzzy C-means
procedures QCD and QCD-PCA according to the fuzziness pa-
rameter. For each value of m, the best result is shown in bold.

		QCD	QCD-PCA
$m = 1.5$	Cluster 1	0.915	**1**
	Cluster 2	0.916	**1**
	Cluster 3	0.916	**1**
$m = 1.8$	Cluster 1	0.770	**0.996**
	Cluster 2	0.779	**0.995**
	Cluster 3	0.771	**0.996**
$m = 2.0$	Cluster 1	0.693	**0.987**
	Cluster 2	0.687	**0.987**
	Cluster 3	0.692	**0.991**
$m = 2.2$	Cluster 1	0.620	**0.975**
	Cluster 2	0.624	**0.978**
	Cluster 3	0.630	**0.978**

defined by the generating processes was assumed to be the true partition. Table 1 contains the average FARI for each of the values of m with both approaches, which are referred to as QCD and QCD-PCA. The results achieved by QCD-PCA were better than the ones attained by using the raw features, specially as the value of m got larger.

To gain a deeper insight into the behaviour of both approaches, we examined the resulting membership degrees. First, we obtained the average value of the maximum membership degree concerning each of the three clusters (see Table 2). The membership degrees generated by QCD-PCA are far larger than those from QCD, thus concluding that QCD-PCA best resolves the level of uncertainty by providing a fuzzy partition with membership values close to one or zero.

Then, we recorded the proportion of times that the fuzzy partition produced by both approaches resulted in the true crisp partition. To this end, we decided to assign the i-th MTS to the c-th cluster according to two criteria: $u_{ic} > 0.8$ and $u_{ic} > 0.6$, thus evaluating both approaches under different degrees of requirement in the assignment rule. The corresponding rates of correct identification are shown in Table 3. We can see that, in both cases, QCD-PCA outperformed QCD by a large extent. The latter was unable to get the true partition when the value of m was 2 or 2.2. These results suggest that the solutions attained by QCD-PCA are far closer to the true partition. Note that it is reasonable that both procedures decreased their performance when increasing the value of m, since higher values of m imply softer partitions.

Thus, one can conclude that, generally, when PCA transformation is not considered, the resulting partitions show a high degree of overlap between clusters, thus giving little informative solutions in the case of underlying crisp partitions. On the other hand, by applying PCA, large membership degrees are usually attained, which provides more reliable solutions. This holds true for the most common values of m used in practice.

It is worth highlighting that the results achieved by QCD-PCA in this toy example are likely to be attributable to the noise reduction and higher stability resulting from the PCA transformation. PCA is often used as a preprocessing step to unsupervised classification [41], usually providing a robustification to the clustering technique [42]. Indeed, several works show its usefulness in different application domains [43–46]. Note that, due to the definition of the smoothed

Table 3
Rates of correctly identified crisp partitions for fuzzy C-means procedures QCD and QCD-PCA according to the fuzziness parameter and two different cutoff values. For each value of m and the cutoff, the best result is shown in bold.

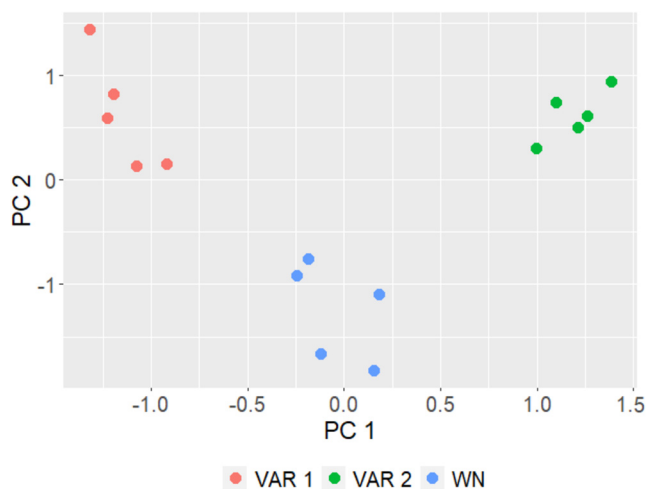| m	$u_{ic} > 0.8$		$u_{ic} > 0.6$	
	QCD	QCD-PCA	QCD	QCD-PCA
$m = 1.5$	0.17	**0.92**	0.87	**0.92**
$m = 1.8$	0	**0.54**	0.47	**0.85**
$m = 2.0$	0	**0.30**	0.04	**0.82**
$m = 2.2$	0	**0.14**	0	**0.73**

Fig. 1. PCA score plot of the first two PCs for one simulation trial considering the QCD-PCA procedure. The points were coloured according to the generating process. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

CCR-periodogram in (10), the QCD-based features are highly correlated. Therefore, by considering the raw features, some variables could get a higher weight than others in the distance computation, thus creating a bias in the clustering algorithm. The PCA transformation avoids this problem by removing the underlying correlation between features, thus making the grouping process easier.

We also took advantage from this toy example to illustrate the high discriminatory power of QCD-PCA. Specifically, we selected a random a trial from the set of simulations and recorded the matrix of scores over the first two principal components, PC 1 and PC 2. The scatter plot based on these scores is displayed in Fig. 1, where the points were coloured according to their generating structures. It is clearly observed that, by projecting on the PC 1–PC 2 plane, one can properly discriminate between the three underlying processes, thus showing the effectiveness of QCD-PCA to discriminate between the generating processes of a given collection of MTS.

The better performance of QCD-PCA over QCD holds generally true whatever the generating processes. Specifically, in the simulated scenarios considered in Section 4, QCD-PCA achieved substantially better results than QCD. For this reason, from now on, the distance d_{QCD} and the clustering procedures QCD-FCMn and QCD-FCMd are going to refer to the PCA-transformed features rather than to the original features, although we maintain the notation for the sake of simplicity and readability. More precisely, given a set of n MTS, $\{X_t^{(1)}, \ldots, X_t^{(n)}\}$, the original QCD-based features are extracted from each series, thus providing the set $\boldsymbol{\Psi} = \{\boldsymbol{\Psi}^{(1)}, \ldots, \boldsymbol{\Psi}^{(n)}\}$. These vectors are transformed by means of PCA, obtaining the set of score vectors $\boldsymbol{\Psi}_{PCA} = \{\boldsymbol{\Psi}_{PCA}^{(1)}, \ldots, \boldsymbol{\Psi}_{PCA}^{(n)}\}$. This is the set subject to clustering by means of Algorithms 1 and 2, and the corresponding distance is $d_{QCD_{PCA}}(X_t^{(1)}, X_t^{(2)}) = \left\| \boldsymbol{\Psi}_{PCA}^{(1)} - \boldsymbol{\Psi}_{PCA}^{(2)} \right\|^2$. The subscript PCA is removed from now on.

4. Simulation study

In this section, we carry out a set of simulations with the aim of evaluating the performance of d_{QCD} in different scenarios of fuzzy clustering of MTS. Firstly we describe some alternative metrics that we have considered for comparison purposes. Then we explain the two ways in which the assessment task was performed, along with the corresponding simulation mechanism and results.

4.1. Alternative metrics

To shed light on the performance of d_{QCD} in a fuzzy clustering context, it was compared with some other clustering models based on alternative dissimilarities. Note that, according to the fuzzy approach based on features extracted from a MTS, a fuzzy C-medoids model and a fuzzy C-means model can be formalized as the minimization problems in (17) and (14), respectively, only by replacing $\widetilde{\boldsymbol{\Psi}}$, $\widetilde{\boldsymbol{\Psi}}^{(c)}$, $\boldsymbol{\Psi}^{(i)}$, $\overline{\boldsymbol{\Psi}}$, $\overline{\boldsymbol{\Psi}}^{(c)}$, $\overline{\boldsymbol{\Psi}}_L$, $\overline{\boldsymbol{\Psi}}_U$ by $\widetilde{\boldsymbol{\varphi}}$, $\widetilde{\boldsymbol{\varphi}}^{(c)}$, $\boldsymbol{\varphi}^{(i)}$, $\overline{\boldsymbol{\varphi}}$, $\overline{\boldsymbol{\varphi}}^{(c)}$, $\overline{\boldsymbol{\varphi}}_L$, $\overline{\boldsymbol{\varphi}}_U$, where $\boldsymbol{\varphi}^{(i)}$ represents the vector of estimated features for the i-th series, $i = 1 \ldots, n$, and the remaining terms are defined analogously. In the same way, the iterative solutions are obtained through (18), (15) and (16) by considering the corresponding features.

The alternative approaches are described below:

- *Wavelet-based features*. [26] introduced a squared Euclidean distance between wavelet features of two MTS, specifically between estimates of wavelet variances and wavelet correlations. 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. Thus, in this case the vector $\boldsymbol{\varphi}^{(i)}$ contains estimates of wavelet variances and wavelet correlations of a given MTS. The corresponding methods are referred to as Wavelet-based fuzzy C-medoids clustering model (W-FCMd) and Wavelet-based fuzzy C-means clustering model (W-FCMn). After performing some brief preliminary analyses, we reached the conclusion that the wavelet filter of length 4 of the Daubechies family, DB4, along with the maximum allowable number of scales, were the choices that led to the best average results in the considered simulation scenarios (see Sections 4.2.1 and 4.2.2). Hence, they were the hyperparameters chosen for the simulation study.

- *Correlation-based features*. In the univariate framework, [10] proposed a fuzzy procedure in which the extracted features are the estimated autocorrelations of a UTS for lags between 1 and fixed l. Here we propose to generalize this approach to a multivariate context. This generalization is straightforward. Given a MTS, we fix a lag l and compute estimates of the autocorrelations up to lag l for each component (UTS) and of the cross-correlations up to lag l between each pair of different components. This set of features defines the vector $\boldsymbol{\varphi}^{(i)}$ describing the i-th MTS and is used to perform clustering. We call the corresponding approaches Correlation-based fuzzy C-medoids clustering model (C-FCMd) and Correlation-based fuzzy C-means clustering model (C-FCMn). The hyperparameter l was set to $l = 1$ throughout the simulation study, as the majority of the considered generating processes contain only one significant lag (see Sections 4.2.1 and 4.2.2).

- *A versatile approach based on features of different nature*. [33] provided a two-step procedure for MTS clustering. First, a set of features of different nature (skewness, kurtosis, nonlinear structure...) are extracted from each UTS within the MTS and stored in a vector. Second, these vectors are concatenated with each other in order to construct a unique vector describing the i-th MTS, $\boldsymbol{\varphi}^{(i)}$. We refer to the corresponding approaches as Features-based fuzzy C-medoids clustering model (F-FCMd) and Features-based fuzzy C-means clustering model (F-FCMn).

- *VPCA metric*. It was proposed by [27]. The procedure is based on PCA and a spatial weighted distance matrix. The grouping is carried out over a set of matrices which are constructed applying dimensionality reduction techniques to the original set of MTS. Using a distance between matrices called SWMD, a fuzzy C-means approach is proposed (see Section III-B in [27]). Its extension to the case of a fuzzy-C medoids is straightforward. Note that this method does not pertain to the fuzzy framework based on extracted features, so here there is no vector $\boldsymbol{\varphi}^{(i)}$. The approaches are called the VPCA-based fuzzy C-medoids clustering model (VPCA-FCMd) and the VPCA-based fuzzy C-means clustering model (VPCA-FCMn).

It is worth remarking that we did not consider the well-known DTW distance as an alternative procedure. This was due to the fact that, in our previous work [31], this distance proved totally useless for grouping MTS according to the

generating processes in a crisp clustering framework. Indeed, we have made some proofs with this metric in a fuzzy context and it attained very poor results.

4.2. Experimental design and results

A broad simulation study was carried out to evaluate the performance of the proposed methods, QCD-FCMn and QCD-FCMd. We intended to construct the evaluation process in a way that general conclusions about the performance of both approaches can be reached. To this end, we designed two different assessment schemes. The first one contains scenarios with three different groups of MTS, and is aimed at evaluating the capability of the procedures in assigning high (low) membership if a given MTS pertains (not pertains) to a given cluster. The second one consists of scenarios with two different groups of MTS and a MTS which does not belong to any of the groups. It assesses again the membership degrees of the series in the two groups but also that the switching series is not assigned to any of the clusters with high membership. The latter scheme utilises a cutoff value determining whether or not a membership degree in a given group is enough to assign the corresponding MTS to that cluster.

4.2.1. First assessment scheme

In this section, the performance of QCD-FCMn and QCD-FCMd is analysed by means of three different simulated scenarios. We attempted to recreate scenarios with a wide variety of generating processes (linear, nonlinear and conditional heteroskedastic), number of dimensions, and series lengths. Each scenario consisted of three clusters (i.e., generating models) with five MTS each, defining the true clustering partition. The generating models concerning each class of processes are given below.

Scenario 1. Fuzzy clustering of VARMA processes.

(a) VAR(1)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} = \begin{pmatrix} 0.4 & -0.3 & 0.9 \\ 0.4 & 0.3 & -0.1 \\ 0.5 & 0.3 & -0.2 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \\ X_{t-1,3} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \\ \epsilon_{t,3} \end{pmatrix},$$

(b) VMA(1)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} = \begin{pmatrix} 0.3 & -0.7 & -0.9 \\ 0.2 & 0.3 & 0.1 \\ 0.2 & 0.1 & -0.3 \end{pmatrix} \begin{pmatrix} \epsilon_{t-1,1} \\ \epsilon_{t-1,2} \\ \epsilon_{t-1,3} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \\ \epsilon_{t,3} \end{pmatrix},$$

(c) VARMA(1, 1)

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

Scenario 2. Fuzzy clustering of nonlinear processes.

(a) NAR (nonlinear autoregressive process)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0.7|X_{t-1,1}|/(|X_{t-1,2}| + 1) \\ 0.7|X_{t-1,2}|/(|X_{t-1,1}| + 1) \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix},$$

(b) TAR (threshold autoregressive process)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0.9X_{t-1,2}I_{\{|X_{t-1,1}|\leq 1\}} - 0.3X_{t-1,1}I_{\{|X_{t-1,1}|>1\}} \\ 0.9X_{t-1,1}I_{\{|X_{t-1,2}|\leq 1\}} - 0.3X_{t-1,2}I_{\{|X_{t-1,2}|>1\}} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix},$$

(c) BL (bilinear process)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0.7X_{t-1,1}\epsilon_{t-2,2} \\ 0.7X_{t-1,2}\epsilon_{t-2,1} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix}.$$

Scenario 3. Fuzzy clustering of dynamic conditional correlation processes. Consider $(X_{t,1}, X_{t,2})^{\mathsf{T}} = (a_{t,1}, a_{t,2})^{\mathsf{T}} = (\sigma_{t,1}\epsilon_{t,1}, \sigma_{t,2}\epsilon_{t,2})^{\mathsf{T}}$, denoting T the transpose vector. The data generating process consists of two GARCH models. Specifically,

$$\sigma_{t,1}^2 = 0.01 + 0.05a_{t-1,1}^2 + 0.94\sigma_{t-1,1}^2,$$
$$\sigma_{t,2}^2 = 0.5 + 0.2a_{t-1,2}^2 + 0.5\sigma_{t-1,2}^2,$$
$$\begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix} \right].$$

The correlation between the standardized shocks, ρ_t, is given by the following expressions:

(a) Piecewise constant correlation

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

(b) Constant correlation

$$\rho_t = 0.5,$$

(c) Piecewise constant correlation

$$\rho_t = 0.9I_{\{t \leq (T/2)\}} - 0.2I_{\{t > (T/2)\}}.$$

The error vector in Scenarios 1 and 2 is formed by independent observations following a multivariate standard Gaussian distribution.

VARMA models are broadly used in many fields but the determination of the models order is quite complex since fixing orders too small leads to inconsistent estimators whereas too large orders produce less accurate predictions. Note that our approach does not require prior modelling. Scenario 2 consists of multivariate extensions of univariate NAR, TAR, and BL processes proposed in [47]. Nonlinear UTS arise in several application fields [48–50]. Although nonlinear MTS have received much less attention than linear ones, there exist some fields as neurophysiology [51] and economy [52] in which nonlinear analysis of MTS has proven to be critical. Thus, a good fuzzy clustering method should be able to specify proper membership degrees between different nonlinear generating processes. Scenario 3 is based on Scenario 2 in [31], which is in turn motivated by a simulation study in the landmark work [53], where dynamic conditional correlation models are introduced. Multivariate GARCH models have been comprehensively investigated over the last decades (an extensive survey is offered in [54]). Specifically, estimation of dynamic conditional correlation models has been widely applied to financial series of different nature [55–57]. Furthermore, we have decided to include in Scenario 3 both positive and negative correlations, since it has been shown that some financial quantities are either positive or negative correlated depending on the period [58]. Some of the generating processes in Scenarios 1, 2 and 3 have already been used either for clustering [31] or classification [38] purposes.

We considered different values for the series length, namely $T \in \{100, 150, 200\}$ in Scenario 1, $T \in \{300, 400, 500\}$ in Scenario 2 and $T \in \{1000, 1500, 2000\}$ in Scenario 3 in order to study its effect in the proposed approaches. Note that, as the three scenarios contain very distinct types of processes, it is logical that very different values of T are needed in order to make an appropriate evaluation. Particularly, we considered specially large values of T in Scenario 3. However, this is not necessarily a drawback, as these sample sizes are offered encountered in real MTS fitted by means of dynamic conditional correlation models [55,57]. Indeed multivariate series of stock returns and other related financial quantities, which consist of measures of daily or even intra-daily data, are one common example of series fitted through this class of models.

The fuzziness parameter m plays a crucial role in the obtained clustering solution. When $m = 1$, the crisp version of either fuzzy C-means or fuzzy C-medoids is obtained. On the other hand, excessively large values of m result in a partition with all memberships close to $1/C$, thus having a large degree of overlap between clusters. As a consequence,

choosing these values for m is not recommended [59]. There is a broad range of literature on determining proper values for the fuzziness parameter appropriately. For instance, [8] showed that values of m between 1.5 and 2.5 are typically a good choice for the fuzzy-C means clustering algorithm. This is also confirmed by [60,61]. However, there seems to be no consensus about the optimal value for m (see discussion in Section 3.1.6 of [11]). In the context of time series clustering, the majority of works consider values of m between 1.3 and 2.6 when performing simulation studies [16,26,10,11,18]. Based on the previous considerations, we have decided to take into consideration the values $m = 1.5$, 1.8, 2 and 2.2.

Given a scenario, a value for m and a value for T, 200 simulations were performed. In each trial, we applied the QCD-FCMn method, as well as the fuzzy C-means versions of the competitors in Section 4.1. The same procedure was repeated regarding the QCD-FCMd technique and the corresponding fuzzy-C-medoids variants. The set of probability levels $\mathcal{T} = \{0.1, 0.5, 0.9\}$ and the first $\lfloor 0.12p \rfloor$ principal components, being p the total number of principal components, were used to compute the QCD-based features. The number of clusters was set to $C = 3$. The effectiveness of each clustering procedure was measured by means of the fuzzy extension of the Adjusted Rand Index (FARI) devised by [62]. This index compares the true partition with the membership matrix resulting from a clustering algorithm. It is based on the original definition of the Adjusted Rand Index and some notions of the fuzzy set theory. The resulting index is also bounded between -1 and 1, as the original Adjusted Rand Index, indicating values close to 1 a more accurate clustering solution. One of the biggest advantages of the FARI is that it allows to elude the loss of information which is unavoidable when using the original formulation of the index to assess the quality of a fuzzy partition; e.g., by transforming the membership matrix into a crisp partition. This way, two membership matrices giving rise to the same crisp partition can be compared with one another, thus making the comparison process fairer. Computation of FARI requires the selection of a triangular norm (see Section 3.2 in [62]) and the corresponding triangular conorm (the dual norm of the former). As the triangular norm, we have considered the minimum norm, which is a common choice in fuzzy logic [8]. The respective triangular conorm is the maximum norm.

Results

Table 4 shows the average values of the FARI for the fuzzy C-means procedures. We can see that all the methods decreased their performance as the value of the parameter m increased. This is reasonable and expected since we are considering scenarios with three well-defined clusters. As it was already mentioned, small values of m lead to near-crisp partitions. On the contrary, larger values of m imply smoothing the boundary between clusters and make the classification fuzzier, thus decreasing the value of the FARI.

The algorithm QCD-FCMn achieved the best scores in all the considered setups except for Scenario 1 with $T = 100$ and $m = 1.5$, where it was slightly outperformed by the correlation-based approach C-FCMn. In Scenario 1, these two methods obtained very similar scores for all values of T and $m = 1.5, 1.8$. However, for $m = 2, 2.2$, the difference in favour of QCD-FCMn got significant. The wavelet-based technique also obtained acceptable scores in this scenario, particularly for the largest value of T and the smallest values of m.

As for Scenario 2, the wavelet and the correlation-based approaches attained similar results, the former outperforming the latter to a small extent as the series length increased. Both techniques were clearly defeated by QCD-FCMn, which attained, in all the setups, an average FARI at least 0.08 points better than either. With regards to Scenario 3, the proposed method also beat W-FCMn and C-FCMn by a considerable degree. The difference with respect to the correlation-based procedure, the second best-performing method, was smaller for $T = 1000$, but substantial for $T = 1500$ and 2000.

The remaining methods F-FCMn and VPCA-FCMn attained in general poor results. The former achieved an average FARI above 0.30 in almost all the settings of Scenarios 1 and 2, thus indicating that it was able to distinguish between generating processes to some extent. However, it did not show a significant improvement when larger series were considered. This is probably due to the fact that some of the features taken into consideration by this technique are useless to differentiate between underlying dependence structures, thus making the clustering process noisy. Indeed, in Scenario 3, F-FCMn did not perform better than choosing a membership matrix at random. The procedure based on the spatial weighted distance matrix VPCA-FCMn, got by far the worst results among the five considered algorithms. It was unable to reach meaningful conclusions in all the situations, always obtaining an average FARI near to zero. It is clear from the results that this approach is not appropriate to perform fuzzy clustering based on generating processes. For this reason, we decided to not take this method into account for further analysis.

Table 4
Average FARI obtained by the fuzzy C-means clustering procedures.

		QCD-FCMn	W-FCMn	C-FCMn	F-FCMn	VPCA-FCMn
Scenario 1						
$T = 100$	$m = 1.5$	0.952	0.536	**0.975**	0.309	0.027
	$m = 1.8$	**0.910**	0.426	0.886	0.296	0.002
	$m = 2.0$	**0.867**	0.364	0.810	0.286	0
	$m = 2.2$	**0.818**	0.312	0.734	0.274	0
$T = 150$	$m = 1.5$	**0.991**	0.865	0.990	0.369	0.026
	$m = 1.8$	**0.961**	0.728	0.932	0.354	0.001
	$m = 2.0$	**0.927**	0.641	0.872	0.342	0
	$m = 2.2$	**0.886**	0.564	0.806	0.327	0
$T = 200$	$m = 1.5$	**0.995**	0.931	0.994	0.376	0.026
	$m = 1.8$	**0.973**	0.813	0.950	0.365	0.001
	$m = 2.0$	**0.946**	0.729	0.898	0.355	0
	$m = 2.2$	**0.910**	0.652	0.839	0.342	0
Scenario 2						
$T = 300$	$m = 1.5$	**0.852**	0.703	0.694	0.398	0.043
	$m = 1.8$	**0.811**	0.626	0.638	0.380	0.002
	$m = 2.0$	**0.772**	0.573	0.594	0.364	0
	$m = 2.2$	**0.733**	0.521	0.549	0.348	0
$T = 400$	$m = 1.5$	**0.931**	0.839	0.764	0.405	0.046
	$m = 1.8$	**0.892**	0.744	0.700	0.388	0.002
	$m = 2.0$	**0.855**	0.681	0.654	0.374	0
	$m = 2.2$	**0.812**	0.622	0.607	0.359	0
$T = 500$	$m = 1.5$	**0.962**	0.876	0.790	0.407	0.058
	$m = 1.8$	**0.929**	0.784	0.726	0.393	0.002
	$m = 2.0$	**0.896**	0.723	0.681	0.380	0
	$m = 2.2$	**0.856**	0.664	0.635	0.365	0
Scenario 3						
$T = 1000$	$m = 1.5$	**0.631**	0.402	0.568	0.003	0.002
	$m = 1.8$	**0.569**	0.340	0.499	0.002	0
	$m = 2.0$	**0.522**	0.298	0.451	0.002	0
	$m = 2.2$	**0.475**	0.261	0.405	0.001	0
$T = 1500$	$m = 1.5$	**0.783**	0.482	0.622	-0.010	0.002
	$m = 1.8$	**0.701**	0.414	0.556	-0.009	0
	$m = 2.0$	**0.643**	0.369	0.509	-0.008	0
	$m = 2.2$	**0.590**	0.326	0.463	-0.007	0
$T = 2000$	$m = 1.5$	**0.863**	0.575	0.718	0.009	0
	$m = 1.8$	**0.782**	0.493	0.643	0.006	0
	$m = 2.0$	**0.721**	0.438	0.589	0.005	0
	$m = 2.2$	**0.660**	0.391	0.535	0.005	0

It is worth remarking that, according to the results in Table 4, QCD-FCMn was the less affected approach by increasing the value of m among the best-performing ones. Note that, whereas some methods as the wavelet-based procedure usually decreased the average FARI by far more than 0.10 when a given value of m is replaced by the next, the maximum variation of QCD-FCMn was 0.082 in Scenario 3 with $T = 1500$, when moving from $m = 1.5$ to $m = 1.8$ (in the remaining settings, the variation was far less than that). The stability exhibited by QCD-FCMn against the modification of the values of m is a very beneficial property of the proposed approach. As already stated, the fuzziness parameter plays a pivotal role in the quality of the clustering solution, and has to be set in advance in practical applications, usually without guarantees about the rightness of the choice. The devised algorithm QCD-FCMn gets around this limitation, ensuring a high probability of meaningful results whatever the value of m. Thus, it could be seen as a sure bet to be used in real clustering problems.

In order to gain illustrative insights into the previous results, Fig. 2 displays the boxplots based on the FARI according to the 200 simulation trials for intermediate values of T (150, 400 and 1500 for Scenarios 1, 2 and 3,
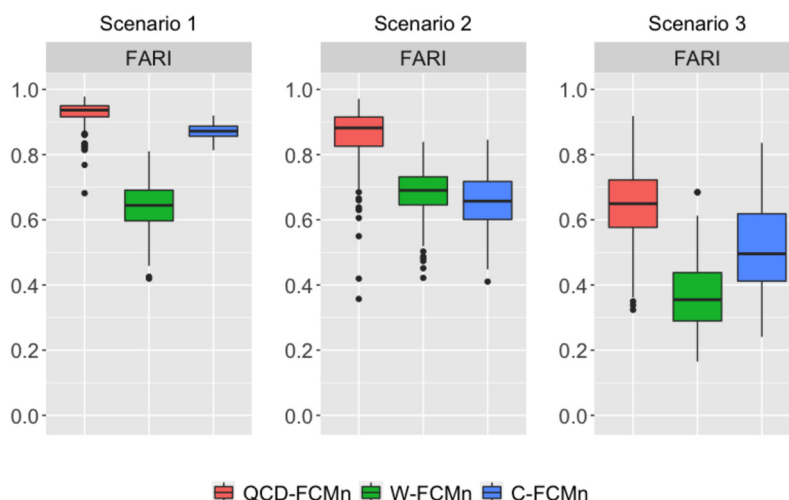
Fig. 2. Boxplots of FARI index according to the 200 trials of the simulation procedure. The considered length was $T = 150$ for Scenario 1, $T = 400$ for Scenario 2 and $T = 1500$ for Scenario 3. The fuzziness parameter was set to $m = 2$.

respectively) and $m = 2$. We depicted the results only for the three best performing approaches according to Table 4, QCD-FCMn, W-FCMn and C-FCMn. The superiority of QCD-FCMn over the remaining strategies is obvious from Fig. 2. Additionally, the plots give us an idea about the variability of the results associated with each procedure. In Scenario 1, the results of QCD-FCMn and C-FCMn showed very little dispersion in comparison with those of W-FCMn. On the contrary, in Scenarios 2 and 3, the approaches exhibiting less variability were QCD-FCMn and W-FCMn, whereas the correlation-based technique C-FCMn displayed the most. A similar situation arises when considering the rest of the values for T and m. Hence, the proposed algorithm also has the desirable property of giving less variable results than the considered competitors.

The results involving the fuzzy C-medoids versions of the approaches are given in Table 5. Generally, the fuzzy C-medoids algorithms showed worse effectiveness than the fuzzy C-means algorithms. Aside from this fact, some interesting conclusions can be reached from Table 5. On the one hand, all the approaches substantially decreased their performance in Scenario 3 as compared with their fuzzy C-means counterparts. A different story went for Scenario 1 and 2. In these scenarios, whereas W-FCMn and C-FCMn significantly worsened their scores in comparison with those of Table 4, QCD-FCMn barely suffered. Indeed, whereas the former approaches decreased in some setups their average FARI by more than 0.15 when changing the algorithm, the greatest decrease for QCD-FCMn, occurring in Scenario 2 when $T = 500$ and $m = 2.2$, was 0.043. Consequently, the proposed procedure is the one showing the most robustness against a change in the considered clustering algorithm.

We have redone the simulations by considering heavy tails in the error distribution. Note that this property often arises in real time series, specially within the field of finance [63–66]. Specifically, we have simulated the (independent) innovations in Scenarios 1, 2 and 3 from a multivariate t distribution with 3 degrees of freedom. For the sake of simplicity, we give the results only for the fuzzy C-means-based approaches. They are depicted in Table 6.

By comparing Table 4 with Table 6, one can state that QCD-FCMn did not suffer when some amount of fat tailedness was introduced in the error distribution. It inherits the robustness of quantile methods, which gives this technique a desirable stability against the distributional form of the error. On the contrary, W-FCMn and C-FCMn significantly decreased their efficacy, specially in Scenarios 2 and 3. These procedures rely mainly on measures of traditional correlation and cross-correlation, and these quantities get less accurate under tail behaviour.

It is worth noting that some other indexes aside from the FARI have been computed for the simulations carried out throughout this section. Particularly, the fuzzy versions of the Rand Index, the Jaccard Index and the Fowlkes-Mallows Index. All of them lead to similar conclusions as the ones stated above. Thus, the results concerning the mentioned alternative indexes are not presented in this work, although they are available upon request.

Table 5
Average FARI obtained by the fuzzy C-medoids clustering procedures.

		QCD-FCMd	W-FCMd	C-FCMd	F-FCMd
Scenario 1					
$T = 100$	$m = 1.5$	**0.957**	0.453	0.861	0.327
	$m = 1.8$	**0.900**	0.342	0.706	0.301
	$m = 2.0$	**0.852**	0.289	0.645	0.279
	$m = 2.2$	**0.801**	0.251	0.585	0.262
$T = 150$	$m = 1.5$	**0.988**	0.774	0.910	0.386
	$m = 1.8$	**0.952**	0.609	0.761	0.356
	$m = 2.0$	**0.912**	0.522	0.723	0.335
	$m = 2.2$	**0.872**	0.446	0.684	0.317
$T = 200$	$m = 1.5$	**0.991**	0.862	0.890	0.394
	$m = 1.8$	**0.951**	0.717	0.779	0.373
	$m = 2.0$	**0.927**	0.628	0.726	0.350
	$m = 2.2$	**0.894**	0.543	0.706	0.330
Scenario 2					
$T = 300$	$m = 1.5$	**0.837**	0.624	0.560	0.437
	$m = 1.8$	**0.799**	0.524	0.518	0.386
	$m = 2.0$	**0.747**	0.461	0.507	0.356
	$m = 2.2$	**0.705**	0.409	0.466	0.330
$T = 400$	$m = 1.5$	**0.913**	0.692	0.593	0.491
	$m = 1.8$	**0.863**	0.574	0.559	0.409
	$m = 2.0$	**0.818**	0.528	0.540	0.378
	$m = 2.2$	**0.781**	0.458	0.496	0.351
$T = 500$	$m = 1.5$	**0.936**	0.671	0.625	0.558
	$m = 1.8$	**0.917**	0.594	0.601	0.473
	$m = 2.0$	**0.868**	0.504	0.571	0.421
	$m = 2.2$	**0.813**	0.477	0.517	0.380
Scenario 3					
$T = 1000$	$m = 1.5$	**0.537**	0.341	0.520	-0.002
	$m = 1.8$	**0.448**	0.246	0.411	0.002
	$m = 2.0$	**0.382**	0.204	0.328	0.003
	$m = 2.2$	**0.353**	0.170	0.271	0.003
$T = 1500$	$m = 1.5$	**0.656**	0.408	0.591	-0.010
	$m = 1.8$	**0.512**	0.306	0.469	-0.005
	$m = 2.0$	**0.451**	0.241	0.379	-0.002
	$m = 2.2$	**0.414**	0.204	0.331	-0.004
$T = 2000$	$m = 1.5$	0.648	0.487	**0.687**	0.011
	$m = 1.8$	0.527	0.357	**0.548**	0.010
	$m = 2.0$	**0.477**	0.282	0.445	0.008
	$m = 2.2$	**0.453**	0.251	0.378	0.007

4.2.2. Second assessment scheme

In this section the proposed approaches and their competitors are evaluated in a different manner than that of Section 4.2.1. Here, we considered scenarios with two well-separated clusters consisting of five time series each and a single switching series arising from a different generating process. The corresponding scenarios and generating processes are described below.

Scenario 4. Fuzzy clustering of VARMA processes with switching series.

Cluster 1: VAR(1)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0 & 0.2 \\ 0.2 & 0.2 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix},$$

Table 6

Average FARI obtained by the fuzzy C-means clustering procedures. Innovations were drawn from a multivariate t distribution with 3 degrees of freedom.

		QCD-FCMn	W-FCMn	C-FCMn	F-FCMn
Scenario 1					
$T = 100$	$m = 1.5$	**0.969**	0.381	0.922	0.315
	$m = 1.8$	**0.927**	0.321	0.796	0.299
	$m = 2.0$	**0.886**	0.282	0.709	0.286
	$m = 2.2$	**0.839**	0.243	0.629	0.270
$T = 150$	$m = 1.5$	**0.990**	0.458	0.954	0.356
	$m = 1.8$	**0.962**	0.407	0.852	0.339
	$m = 2.0$	**0.929**	0.367	0.773	0.325
	$m = 2.2$	**0.889**	0.330	0.697	0.310
$T = 200$	$m = 1.5$	**0.994**	0.544	0.965	0.373
	$m = 1.8$	**0.971**	0.488	0.876	0.357
	$m = 2.0$	**0.943**	0.443	0.802	0.344
	$m = 2.2$	**0.907**	0.402	0.729	0.329
Scenario 2					
$T = 300$	$m = 1.5$	**0.899**	0.163	0.471	0.336
	$m = 1.8$	**0.851**	0.160	0.413	0.310
	$m = 2.0$	**0.807**	0.156	0.369	0.292
	$m = 2.2$	**0.759**	0.156	0.328	0.267
$T = 400$	$m = 1.5$	**0.957**	0.151	0.484	0.347
	$m = 1.8$	**0.918**	0.150	0.433	0.337
	$m = 2.0$	**0.879**	0.149	0.393	0.316
	$m = 2.2$	**0.833**	0.149	0.350	0.294
$T = 500$	$m = 1.5$	**0.986**	0.152	0.505	0.397
	$m = 1.8$	**0.956**	0.151	0.455	0.383
	$m = 2.0$	**0.922**	0.149	0.415	0.360
	$m = 2.2$	**0.881**	0.150	0.377	0.334
Scenario 3					
$T = 1000$	$m = 1.5$	**0.723**	0.002	0.233	-0.001
	$m = 1.8$	**0.644**	0.001	0.189	0
	$m = 2.0$	**0.590**	0.001	0.165	-0.001
	$m = 2.2$	**0.540**	0.001	0.143	-0.001
$T = 1500$	$m = 1.5$	**0.828**	0.001	0.163	0.009
	$m = 1.8$	**0.735**	0.001	0.138	0.007
	$m = 2.0$	**0.668**	0.001	0.120	0.006
	$m = 2.2$	**0.610**	0.001	0.104	0.005
$T = 2000$	$m = 1.5$	**0.881**	0.002	0.111	0
	$m = 1.8$	**0.783**	0.002	0.090	0
	$m = 2.0$	**0.716**	0.002	0.085	0
	$m = 2.2$	**0.653**	0.002	0.074	0.001

Cluster 2: VMA(1)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} -0.4 & -0.4 \\ -0.4 & -0.2 \end{pmatrix} \begin{pmatrix} \epsilon_{t-1,1} \\ \epsilon_{t-1,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix},$$

Switching series: VARMA(1, 1)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0 & 0.2 \\ 0.2 & 0.2 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} -0.4 & -0.4 \\ -0.4 & -0.2 \end{pmatrix} \begin{pmatrix} \epsilon_{t-1,1} \\ \epsilon_{t-1,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix}.$$

Scenario 5. Fuzzy clustering of nonlinear processes with switching series.

Cluster 1: NLVMA (nonlinear VMA process)

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0.1\epsilon_{t-1,1} + 0.6\epsilon_{t-1,2}^2 + \epsilon_{t,1} \\ 0.1\epsilon_{t-1,2} + 0.6\epsilon_{t-1,1}^2 + \epsilon_{t,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix},$$

Cluster 2: NLVMA

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} -0.1\epsilon_{t-1,1} - 0.6\epsilon_{t-1,2}^2 + \epsilon_{t,1} \\ -0.1\epsilon_{t-1,2} - 0.6\epsilon_{t-1,1}^2 + \epsilon_{t,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix},$$

Switching series: a bivariate white noise process.

Scenario 6. Fuzzy clustering of dynamic conditional correlation processes with switching series. Consider the GARCH models and the correlation between the standardized shocks, ρ_t, in Scenario 3.

Cluster 1: $\rho_t = 0.9I_{\{t \le (T/2)\}} - 0.3I_{\{t > (T/2)\}}$,
Cluster 2: $\rho_t = -0.9I_{\{t \le (T/2)\}} + 0.3I_{\{t > (T/2)\}}$,
Switching series: $\rho_t = 0$.

The error vector in Scenarios 4 and 5 is formed by independent observations following a bivariate standard Gaussian distribution.

Nonlinear ARMA processes have attracted a great deal of attention in the univariate framework [67–71], proving themselves useful in several application fields. Particularly, nonlinear MA models have given rise to specific works due to the difficulty in their estimation [72,73]. In a multivariate context, some works have also dealt with the topic of nonlinear VARMA models [74,75].

Note that Scenarios 4, 5 and 6 have been designed in a way that the switching series is expected to lay "in the middle" of both clusters. In other words, a distance measure aimed to discriminate among generating processes should be able to produce very similar distance values from the switching series to a series from Cluster 1 and Cluster 2 indistinctly. In order to see if the proposed metric d_{QCD} verifies this fact, we performed a metric two-dimensional scaling (2DS) based on the pairwise QCD-based dissimilarity matrix. The 2DS performs in the following way. Given a distance matrix $\boldsymbol{D} = (D_{ij}), i, j = 1, \ldots, n$, it finds the set of points $\{(a_i, b_i), i = 1, \ldots, n\}$ such that the stress function

$$\sqrt{\frac{\sum_{i \ne j = 1}^{n} (\| (a_i, b_i) - (a_j, b_j) \| - D_{ij})^2}{\sum_{i \ne j = 1}^{n} D_{ij}^2}} \tag{19}$$

is minimized. Thus, the goal is to represent the distances D_{ij} in terms of Euclidean distances into a 2-dimensional space so that the smaller the value of the stress function, the better the 2DS configuration. The corresponding 2-D plot gives usually a good visual representation of how the elements are located with respect to each other according to the considered distance.

In order to visualize the corresponding 2-D graphs for each scenario, we simulated 50 MTS of length $T = 500$ and $T = 2000$ from the series defining the clusters and the switching series. The 2DS was carried out for each set of 150 MTS and value of the series length. The corresponding points in the new coordinate space are depicted in Fig. 3, where each set of points has been coloured according to the underlying generating process. The top panels show the 2DS for $T = 500$ and the bottom panels, for $T = 2000$.

With the aim of assessing the quality of the embedding, we computed the R^2 value to determine what proportion of variance of the scaled data can be accounted for the 2DS procedure. We obtained the values 0.667 (Scenario 4), 0.507 (Scenario 5) and 0.604 (Scenario 6) for the small sample sizes and 0.866 (Scenario 4), 0.788 (Scenario 5) and 0.852 (Scenario 6) for the large sample sizes. It is worth remarking that values above 0.6 are considered to provide an acceptable scaling procedure, whereas values above 0.8 mean a very good fit [76]. Although the values for the nonlinear case are lower than those for the remaining scenarios, they are quite close to the mentioned thresholds. Thus, generally, the plots in Fig. 3 provide an accurate picture about the distance d_{QCD} between the different underlying processes.

The reduced bivariate spaces in Fig. 3 show three compact, well-separated groups defined by the series of Cluster 1, Cluster 2, and the set of switching series. In all cases, the group of switching series (blue colour) is prone to position
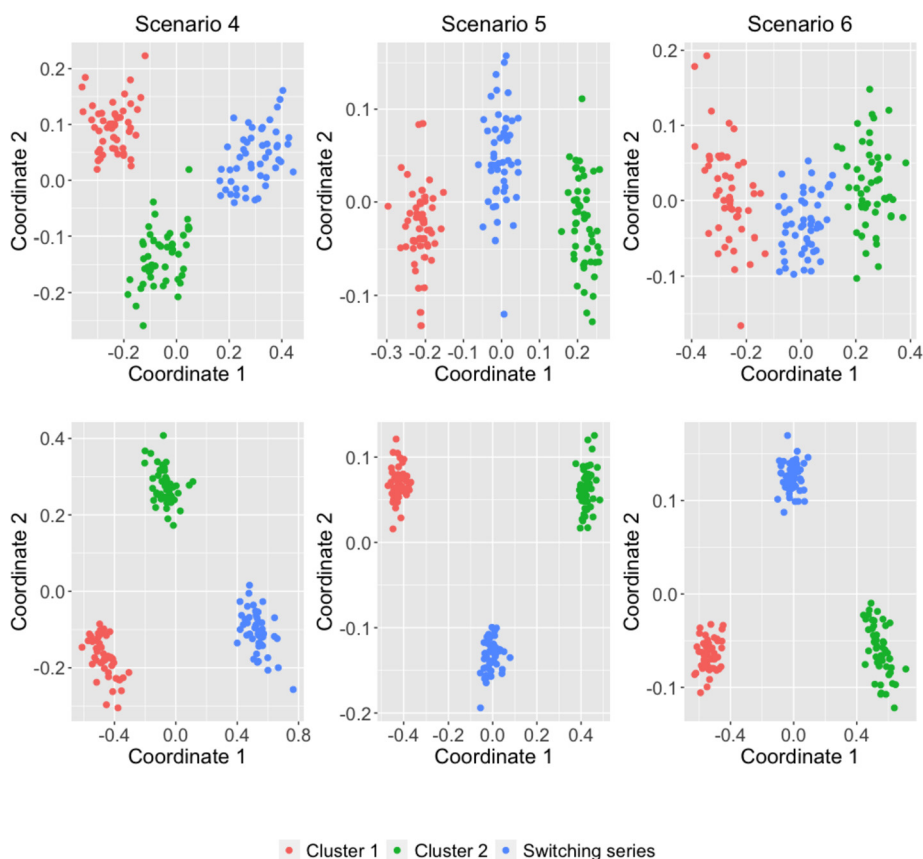
Fig. 3. Two-dimensional scaling planes based on QCD distances between simulated series in Scenarios 4, 5 and 6. The series length is $T = 500$ in the top panels and $T = 2000$ in the bottom panels.

at an intermediate place between both clusters. This is a great trait exhibited by the distance d_{QCD}, since, as remarked previously, Scenarios 4, 5 and 6 were chosen so that the switching series is in the middle of Clusters 1 and 2 in terms of dependence structures. Note that, for a given value of the series length, the most challenging scenario for the proposed clustering algorithms seems to be Scenario 6, since the 2DS plot produces near overlapping clusters when $T = 500$. As expected, when increasing the series length, the QCD-based features are more accurately estimated and the groups become more distant from one another. In summary, Fig. 3 shows that the QCD-based distance should be capable of discriminating between the original clusters and the switching series in Scenarios 4, 5 and 6, thus being a good candidate for the second assessment design.

We considered again different values for the series length, namely, $T \in \{200, 400, 600\}$ in Scenario 4, $T \in \{300, 600, 900\}$ in Scenario 5 and $T \in \{500, 1000, 1500\}$ in Scenario 6. The values taken into account for the fuzziness parameter m were the same as in Section 4.2.1. Again, 200 simulations were performed, and the fuzzy C-means and fuzzy C-medoids versions of the procedures were applied. The same hyperparameters as in Section 4.2.1 were taken into consideration to obtain the QCD-based features and those concerning the alternative procedures. This time, the number of clusters was set to $C = 2$. In this second evaluation scheme, we assessed the clustering methods by means of the frequency with which the five series from Cluster 1 grouped together in one group, the five series from Cluster 2 clustered together in another group, and the switching series had a relatively high membership degree with regards to both groups. To that end, we had to define a cutoff point in order to decide when a given realization was assigned to a specific cluster. We decided to use the cutoff value of 0.7, so that the i-th MTS was placed into the c-th cluster if $u_{ic} > 0.7$. This cutoff value has already been considered in some works [26,10,11]. A discussion about the reasoning for choosing this membership degree constraint can be seen in [11]. In this way, the switching series was considered to concurrently pertain to both clusters if its membership degrees were both below 0.7.

Results

The average success rates attained by the C-means techniques according to the previous criteria are presented in Table 7. The best approach was QCD-FCMn, achieving the best average scores in all the considered configurations excluding Scenario 4 with $m = 1.5$ and $T = 600$, where it was outplayed by F-FCMn. As expected, its results improved when increasing T. It also got better for larger values of m in the three scenarios. This is owing to the fact that, for the smaller values of m, QCD-FCMn returns a quite crisp partition so that the switching series is located a lot of times in one cluster with high membership. Indeed this is the cause that QCD-FCMn attained poor classification rates for $m = 1.5$ whatever the value of the series length in Scenario 4. However, when m becomes larger, the boundaries between clusters become blurrier and the switching series is simultaneously located in both clusters. Nevertheless, this gives rise to a different type of classification error, since, for larger values of m, frequently some of the non switching series display similar membership values in both clusters (e.g., 0.55 and 0.45). This trade-off situation concerning Scenario 4 for QCD-FCMn gets fixed as the value of T increases.

The approaches W-FCMn and C-FCMn only reached satisfactory success rates in Scenario 6, whereas they completely failed in Scenarios 4 and 5. This seems surprising, since at least C-FCMn was expected to perform well when dealing with Scenario 4, as the corresponding metric should be able to distinguish between linear processes. We ran this procedure for values of m greater than 2.2 in Scenario 4 and found out that the corresponding distance was able to get higher success rates for some of those values (but less than those associated with QCD-FCMn). The reason was that, for C-FCMn, the value $m = 2.2$ still means a very crisp partition so its failures are attributable to the switching series. This fact highlights the paramount importance of the fuzziness parameter m when evaluating the strategies through a cutoff point.

Table 8 contains the average frequencies of correct classification for the fuzzy C-medoids variants of the methods. Generally speaking, QCD-FCMd showed approximately the same performance as QCD-FCMn in the three scenarios. W-FCMd and C-FCMd somewhat improved their scores in Scenario 4 but worsened it in Scenarios 5 and 6. F-FCMd achieved better scores than F-FCMn in Scenario 4. As in Table 7, the results appear to be highly influenced by the values of the fuzziness coefficient m.

In order to complement the evaluation results of Tables 7 and 8, we considered the largest values of the series length (600, 900 and 1500) and a grid of values for m equispaced between 1 and a value large enough so that all the methods achieve near-zero rates of correct classification. Simulations were performed in the same way as before. The corresponding curves of frequencies of correct classification as a function of m for the four fuzzy C-means methods are shown in Fig. 4. These curves give a much more complete picture about the discriminatory capability of the approaches over the whole range of values for the fuzziness parameter. As stated before, although values of m above 2.5 or 3 are not commonly used in practice, they can be taken into consideration in the assessment mechanism in order to get a fair comparison between the considered approaches. Given the graphs in Fig. 4, a reasonable measure of performance is the area under the fuzziness curve (AUFC). The corresponding quantities are given in the first part of Table 9. QCD-FCMn was the approach associated with the greatest value of AUFC in the three scenarios. It clearly outmatched the remaining strategies in Scenarios 1 and 2, and got a slightly better value than C-FCMn in Scenario 3. This latter approach was by far the second best-performing one.

Fig. 5 shows the fuzziness curves for the fuzzy C-medoids strategies. The situation is rather similar to that of Fig. 4. On the whole, there is a moderate decline in the performance of all techniques concerning most values of m. The respective values of the AUFC are presented in the second part of Table 9. Indeed, there is a decrease in AUFC for all the methods in comparison with the fuzzy C-means setting except for F-FCMn in Scenario 1, which slightly improved its score. Again, QCD-FCMd reached the best overall results among the four analysed algorithms.

As in Section 4.2.1, the simulations concerning Scenarios 4, 5 and 6 were repeated by taking into account a multivariate t distribution with 3 degrees of freedom for the innovations. The respective fuzziness curves for the fuzzy C-means methods are displayed in Fig. 6, whereas the corresponding values for the AUFC are shown in the last part of Table 9. Clearly, QCD-FCMn is the only approach capable of performing an effective classification under these circumstances, exhibiting again a substantial robustness against the departure from normality in the error distribution. The remaining procedures got a poor rate of correct classification in Scenario 1 and zero in Scenarios 2 and 3.

It is worth highlighting that, in all the previous analyses, the fuzzy C-means and the fuzzy C-medoids algorithms were executed over many random assignments in order to avoid the issues of local optima.

Table 7
Average success rates obtained by the fuzzy C-means clustering procedures.

		QCD-FCMn	W-FCMn	C-FCMn	F-FCMn
Scenario 4					
$T = 200$	$m = 1.5$	**0.070**	0.005	0	0.015
	$m = 1.8$	**0.130**	0.010	0.005	0.015
	$m = 2.0$	**0.115**	0.015	0.025	0.015
	$m = 2.2$	**0.065**	0.015	0.030	0.010
$T = 400$	$m = 1.5$	**0.070**	0	0	0.065
	$m = 1.8$	**0.315**	0	0	0.110
	$m = 2.0$	**0.495**	0.025	0.010	0.125
	$m = 2.2$	**0.455**	0.075	0.070	0.085
$T = 600$	$m = 1.5$	0.055	0	0	**0.245**
	$m = 1.8$	**0.440**	0	0	0.390
	$m = 2.0$	**0.610**	0.005	0.015	0.395
	$m = 2.2$	**0.725**	0.045	0.115	0.410
Scenario 5					
$T = 300$	$m = 1.5$	**0.435**	0	0.025	0
	$m = 1.8$	**0.610**	0	0.005	0
	$m = 2.0$	**0.700**	0	0.005	0
	$m = 2.2$	**0.650**	0	0.005	0
$T = 600$	$m = 1.5$	**0.520**	0	0.090	0
	$m = 1.8$	**0.810**	0	0.085	0
	$m = 2.0$	**0.895**	0	0.065	0
	$m = 2.2$	**0.945**	0	0.040	0
$T = 900$	$m = 1.5$	**0.690**	0	0.150	0
	$m = 1.8$	**0.920**	0.010	0.220	0
	$m = 2.0$	**0.980**	0.005	0.170	0
	$m = 2.2$	**0.990**	0.005	0.105	0
Scenario 6					
$T = 500$	$m = 1.5$	**0.370**	0.280	0.330	0
	$m = 1.8$	**0.645**	0.450	0.560	0
	$m = 2.0$	**0.735**	0.515	0.670	0
	$m = 2.2$	**0.820**	0.500	0.695	0
$T = 1000$	$m = 1.5$	**0.500**	0.335	0.395	0
	$m = 1.8$	**0.790**	0.595	0.595	0
	$m = 2.0$	**0.890**	0.675	0.725	0
	$m = 2.2$	**0.955**	0.735	0.840	0
$T = 1500$	$m = 1.5$	**0.550**	0.415	0.425	0
	$m = 1.8$	**0.860**	0.680	0.730	0
	$m = 2.0$	**0.940**	0.785	0.850	0
	$m = 2.2$	**0.980**	0.840	0.920	0

5. Applications

This section is devoted to show two applications of the proposed clustering procedures.

5.1. Fuzzy clustering of the top 20 companies in the S&P 500 index

Time series clustering has proven very useful when dealing with financial time series. More often than not, the aim is to group different assets or companies according to how they behave over a certain period of time. This way, the clustering solution can represent groups sharing similar risk profiles, management behaviour, or even future expected returns. There are a broad variety of works coping with clustering of financial time series [77–81,14]. Here we present how one of the proposed fuzzy clustering approaches can be applied to group the most important companies in the US. We want to remark that the following analyses are not aimed at giving financial advice nor deriving economical

Table 8
Average success rates obtained by the fuzzy C-medoids clustering procedures.

		QCD-FCMd	W-FCMd	C-FCMd	F-FCMd
Scenario 4					
$T = 200$	$m = 1.5$	**0.080**	0.010	0.010	0.020
	$m = 1.8$	**0.085**	0.025	0.020	0.015
	$m = 2.0$	**0.075**	0.030	**0.075**	0.020
	$m = 2.2$	0.045	0.020	**0.060**	0.005
$T = 400$	$m = 1.5$	0.135	0	0.005	**0.180**
	$m = 1.8$	**0.365**	0.025	0.035	0.180
	$m = 2.0$	**0.395**	0.065	0.085	0.130
	$m = 2.2$	**0.360**	0.140	0.240	0.110
$T = 600$	$m = 1.5$	0.130	0	0	**0.415**
	$m = 1.8$	0.415	0.005	0.020	**0.450**
	$m = 2.0$	**0.570**	0.070	0.120	0.440
	$m = 2.2$	**0.655**	0.215	0.285	0.355
Scenario 5					
$T = 300$	$m = 1.5$	**0.425**	0	0.005	0
	$m = 1.8$	**0.540**	0	0	0
	$m = 2.0$	**0.620**	0	0	0
	$m = 2.2$	**0.540**	0	0	0
$T = 600$	$m = 1.5$	**0.600**	0	0.040	0
	$m = 1.8$	**0.820**	0	0.025	0
	$m = 2.0$	**0.895**	0	0.015	0
	$m = 2.2$	**0.930**	0	0.005	0
$T = 900$	$m = 1.5$	**0.700**	0	0.145	0
	$m = 1.8$	**0.910**	0.005	0.095	0
	$m = 2.0$	**0.960**	0	0.035	0
	$m = 2.2$	**0.975**	0	0.015	0
Scenario 6					
$T = 500$	$m = 1.5$	**0.400**	0.315	0.360	0
	$m = 1.8$	**0.650**	0.440	0.535	0
	$m = 2.0$	**0.735**	0.455	0.580	0
	$m = 2.2$	**0.800**	0.455	0.635	0
$T = 1000$	$m = 1.5$	**0.545**	0.375	0.405	0
	$m = 1.8$	**0.795**	0.510	0.605	0
	$m = 2.0$	**0.880**	0.585	0.685	0
	$m = 2.2$	**0.950**	0.655	0.760	0
$T = 1500$	$m = 1.5$	**0.610**	0.425	0.430	0
	$m = 1.8$	**0.840**	0.660	0.685	0
	$m = 2.0$	**0.925**	0.775	0.800	0
	$m = 2.2$	**0.980**	0.820	0.860	0

implications, but at illustrating the suitability of the designed technique to recognize homogeneous groups with similar stochastic dependence patterns and to what extent each company pertains to each one of the identified groups.

The data we are going to use were taken from the finance section of the Yahoo website[1]. It contains daily stock returns and trading volume of the current top 20 companies of the S&P 500 index according to market capitalization. The sample period spans from 6th July 2015 to 7th February 2018, thus resulting serial realizations of length $T = 655$. The S&P 500 is a stock market index that tracks the stocks of 500 large-cap US companies. The top 20 contains some of the most important companies in the world, as Apple, Google, Facebook or Berkshire Hathaway.

It is important to highlight that the relationship between price and volume has been extensively analyzed in the literature [82–84] and constitutes itself a topic of great financial interest. Prices and trading volume are known to
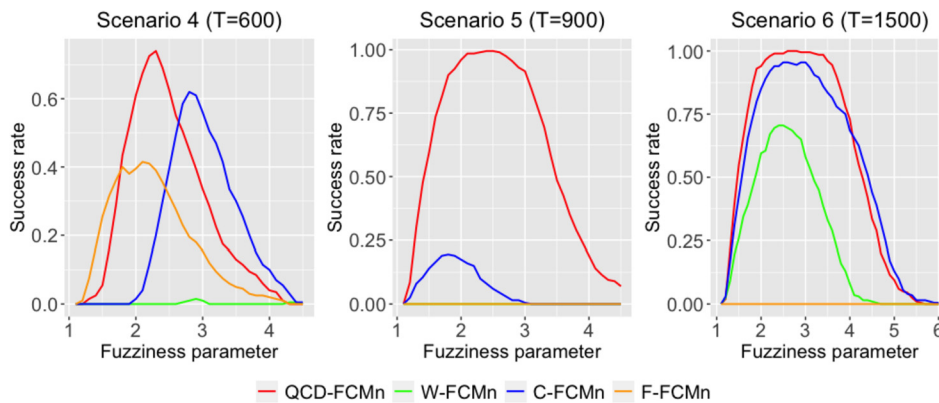
---

[1] https://es.finance.yahoo.com.

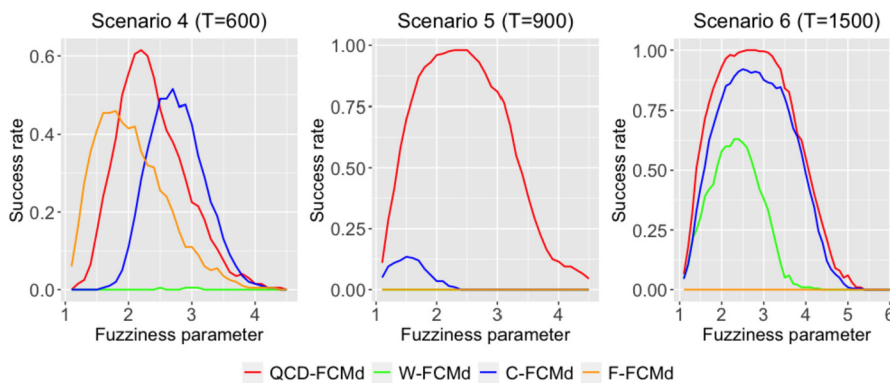Fig. 4. Average rates of correct classification for the fuzzy C-means procedures as a function of m.

Fig. 5. Average rates of correct classification for the fuzzy C-medoids procedures as a function of m.

Table 9
Area under the curves of Figs. 4, 5 and 6.

		QCD	W	C	F
Fuzzy C-means	Scenario 1	**0.9205**	0.0050	0.6943	0.5845
	Scenario 2	**2.0855**	0	0.1925	0
	Scenario 3	**2.7825**	1.2710	2.6483	0
Fuzzy C-medoids	Scenario 1	**0.7620**	0.0002	0.5740	0.6290
	Scenario 2	**1.9908**	0	0.0975	0
	Scenario 3	**2.5923**	0.9825	2.2283	0
Fuzzy C-means	Scenario 1	**1.1568**	0.0215	0.0495	0.1635
(heavy tails)	Scenario 2	**3.7380**	0	0	0
	Scenario 3	**2.2058**	0	0	0

exhibit some empirical linkages over the fluctuations of stock markets. Thus, it is reasonable to describe each of the considered companies by means of these two quantities. Our goal is to analyze the joint behaviour of prices and volume in order to perform fuzzy clustering. Thus, we assume that two companies behave similarly if the corresponding bivariate time series exhibit similar dependence structures.

It can be observed that both the UTS of prices and trading volume are non-stationary in mean. Thus, 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. Finally, all UTS are normalized to have zero mean and unit variance. The resulting MTS are shown in Fig. 7, where the returns and the change in volume are displayed through the red and blue colour, respectively. Overall, plots in Fig. 7 exhibit common traits of financial
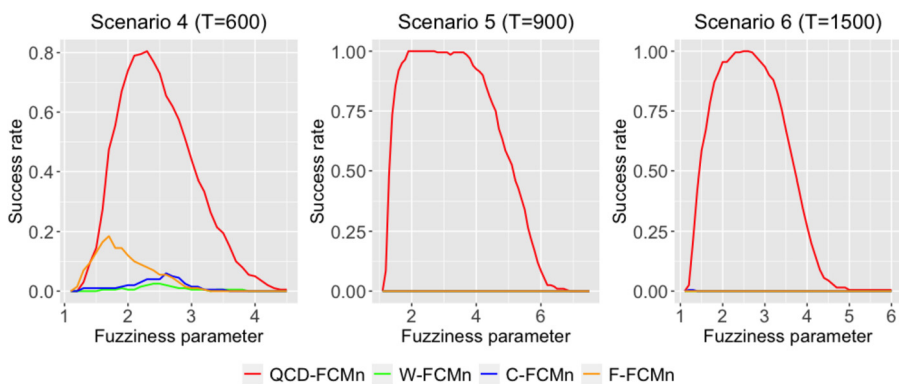
Fig. 6. Average rates of correct classification for the fuzzy *C*-means procedures as a function of *m*. Innovations were drawn from a multivariate *t* distribution with 3 degrees of freedom.

time series. There is a substantial degree of heteroskedasticity in both prices and change in volume. In addition, both quantities exhibit the so-called phenomenon of volatility clustering: large values (positive or negative) tend to group together, resulting in a marked persistence. These special properties of financial time series, usually referred to as stylized facts, are generally accounted for by modelling the series by way of multivariate GARCH-type models, for instance the dynamic conditional correlation models of Scenarios 3 and 6. It is worth remembering that the devised fuzzy clustering algorithms demonstrated their efficacy when coping with this type of models, specially when the error distribution possesses some amount of fat-tailedness. This property also relates to the stylized facts [63,85,86]. Thus, given the high capability of QCD-FCMn and QCD-FCMd to discriminate between conditional heteroskedastic models, it is expected that both methods can provide a meaningful fuzzy partition determining groups of companies following a similar behavioural pattern.

As a preliminary exploratory step, we performed a 2DS based on the pairwise QCD-dissimilarity matrix. That way, a projection of the companies on a two-dimensional plane preserving the original distances as well as possible is available. The location of the top 20 companies in the transformed space is displayed in Fig. 8. The R^2 value is 0.7251. Thus, the scatter plot in Fig. 8 can be considered an acceptable representation of the underlying distance configuration [76].

Overall, Fig. 8 suggests that the grouping of the top 20 companies calls for a fuzzy partition, since the points in the new coordinate space do not appear to be split in clear, well-separated, nonoverlapping groups. Hence, a fuzzy clustering algorithm is expected to give more meaningful insights into the distribution of the corporations than a hard clustering one, where each company would be allowed to pertain to only one cluster. By observing Fig. 8, one could conclude the existence of 5 or 6 overlapping clusters. There is a cluster including the three technological giants Google Alphabet Class A (GOOGL), Google Alphabet Class C (GOOG) and Amazon (AMZN), along with the company PayPal (PYPL). The remaining tech giants Apple (AAPL), Facebook (FB) and Microsoft (MSFT) appear to form another group with some other businesses as Tesla (TSLA) and Walt Disney (DIS). The rest of the firms are more spread out. Berkshire Hathaway (BRK.B) seems to be isolated, well-separated from the remaining companies, and it could be thought of as an atypical corporation. Johnson & Johnson (JNJ) and Visa (V) are placed close to each other, so they could constitute another cluster. The five remaining organizations Procter & Gamble (PG), JPMorgan Chase (JPM), Bank of America (BAC), Walmart (WMT) and Home Depot (HD) could comprise one or two different groups (PG-JPM-BAC and WMT-HD).

As the QCD-FCMn procedure achieved overall slightly better results than the QCD-FCMd algorithm in the simulation study carried out in Section 4, the former was the method chosen for the application. Just as in the simulations, the metric d_{QCD} was constructed by using the probability levels 0.1, 0.5 and 0.9. As it was already pointed out, and as we saw in Section 4, the fuzziness parameter *m* highly influences the quality of the obtained clustering partition. Thus, we decided to choose this parameter as well as the number of clusters, *C*, by way of a data driven approach. In order to do so, we took into consideration four different cluster validity indexes, the Xie-Beni Index (XBI) [87], the Kwon Index (KI) [88] and the indexes proposed in [89] (TI) and in [90] (BI). Now we define those indexes according to the fuzzy approach based on features extracted from an MTS. Let d_{ic} be the Euclidean distance between the element $\boldsymbol{\varphi}^{(i)}$ and the centroid of the *c*-th cluster, $\overline{\boldsymbol{\varphi}}^{(c)}$. The XBI is defined as
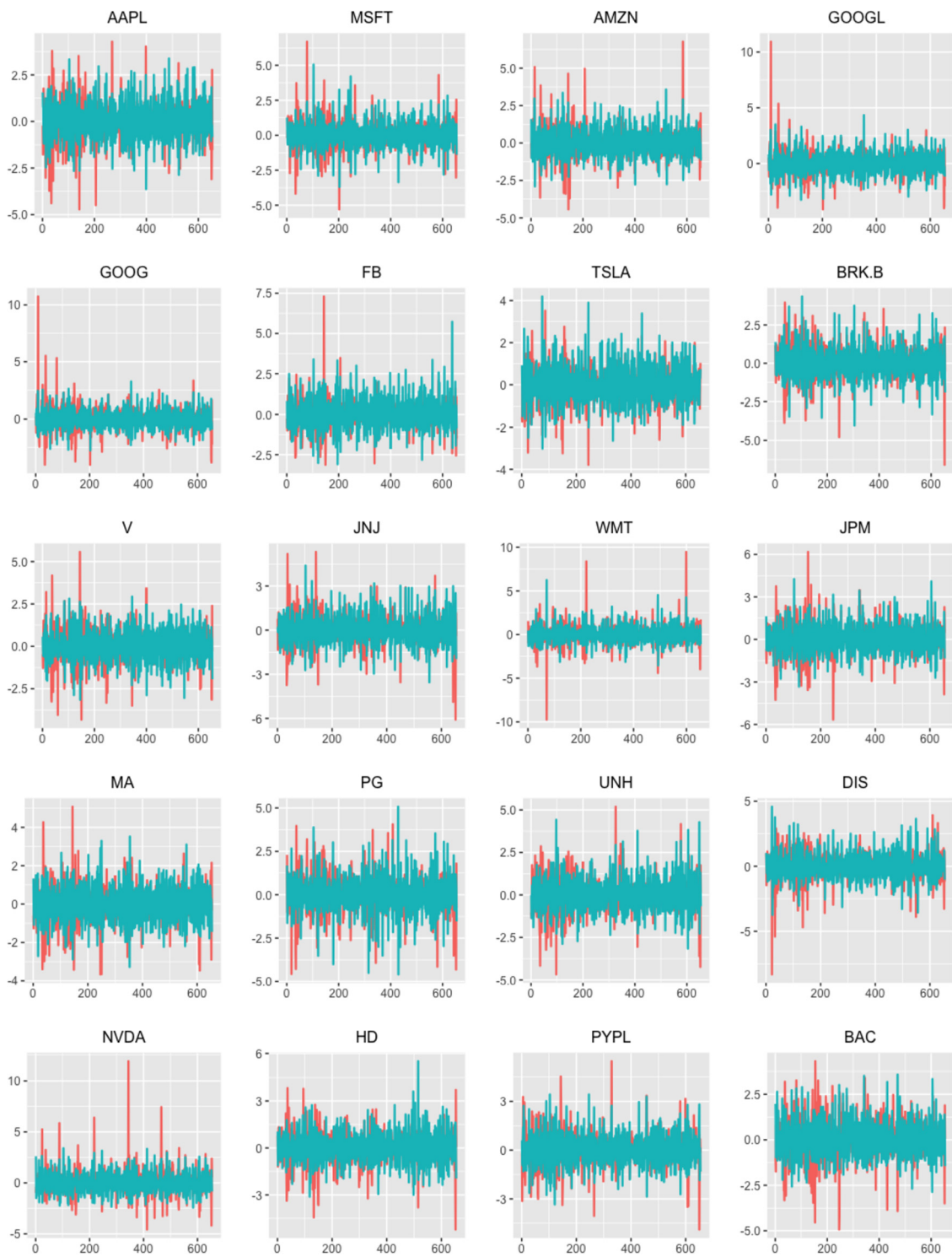
Fig. 7. Daily returns (red colour) and change in volume (blue colour) of the top 20 companies in the S&P 500 index.
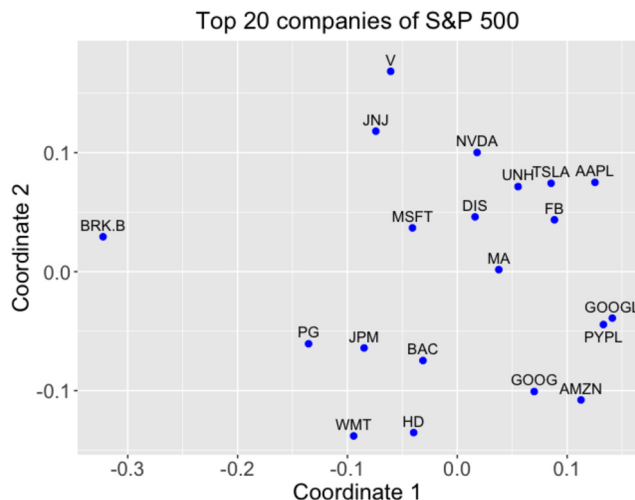
Fig. 8. Two-dimensional scaling plane based on the QCD distances for the daily returns and change in volume of the top 20 companies in the S&P 500 index.

$$\text{XBI}(C, m) = \frac{\sum_{i=1}^{n} \sum_{c=1}^{C} u_{ic}^2 d_{ic}^2}{n \min_{c \neq c'} \left\| \overline{\boldsymbol{\varphi}}^{(c)} - \overline{\boldsymbol{\varphi}}^{(c')} \right\|^2}. \tag{20}$$

Note that minimizing the numerator of (20), which measures the compactness of the fuzzy partition, is precisely the goal of QCD-FCMn when $m = 2$. The denominator in (20) measures the degree of separation between clusters. Thus, the index decreases with separation between clusters. Note that XBI is a function of both the number of clusters C and the fuzziness coefficient m, since both parameters must be used as input to a fuzzy C-means-type clustering algorithm to produce the membership matrix $\boldsymbol{U}$ and the set of centroids that are required to compute the index.

The KI is an extension of the XBI which is aimed at penalizing the monotonically decreasing trend exhibited by the latter when C becomes large. The KI is given by

$$\text{KI}(C, m) = \frac{\sum_{i=1}^{n} \sum_{c=1}^{C} u_{ic}^2 d_{ic}^2 + (1/C) \sum_{c=1}^{C} \left\| \overline{\boldsymbol{\varphi}}^{(c)} - \overline{\boldsymbol{\varphi}} \right\|^2}{\min_{c \neq c'} \left\| \overline{\boldsymbol{\varphi}}^{(c)} - \overline{\boldsymbol{\varphi}}^{(c')} \right\|^2}, \tag{21}$$

where $\overline{\boldsymbol{\varphi}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\varphi}^{(i)}$. Note that the punishing term $(1/C) \sum_{c=1}^{C} \left\| \overline{\boldsymbol{\varphi}}^{(c)} - \overline{\boldsymbol{\varphi}} \right\|^2$ increases substantially as the number of clusters is close to n.

Although KI allows to evaluate the quality of a fuzzy C-means procedure when $C \to n$, it becomes unstable or unpredictable as $m \to \infty$. The TI tries to solve this problem by adding a penalty function both in the numerator and the denominator. This index is defined as

$$\text{TI}(C, m) = \frac{\sum_{i=1}^{n} \sum_{c=1}^{C} u_{ic}^2 d_{ic}^2 + \frac{1}{C(C-1)} \sum_{c=1}^{C} \sum_{c'=1; c' \neq c}^{C} \left\| \overline{\boldsymbol{\varphi}}^{(c)} - \overline{\boldsymbol{\varphi}}^{(c')} \right\|^2}{\min_{c \neq c'} \left\| \overline{\boldsymbol{\varphi}}^{(c)} - \overline{\boldsymbol{\varphi}}^{(c')} \right\|^2 + 1/C}. \tag{22}$$

The BI is given by

$$\text{BI}(C, m) = \sum_{c=1}^{C} \frac{\sum_{i=1}^{n} u_{ic}^2 d_{ic}^2}{\sum_{i=1}^{n} u_{ic} \sum_{c'=1}^{C} \left\| \overline{\boldsymbol{\varphi}}^{(c)} - \overline{\boldsymbol{\varphi}}^{(c')} \right\|^2}. \tag{23}$$

The numerator in (23) measures the compactness and the denominator accounts for the separation between clusters.

Note that, concerning all the previous defined indexes, smaller values mean better fuzzy partitions. Indeed, the minimum value of some of those indexes is frequently utilised to choose the optimal number of clusters [18,10] and

Table 10

Membership degrees for top 20 companies in the S&P 500 index by considering the QCD-FCMn model and a 6-cluster partition.

Company	C_1	C_2	C_3	C_4	C_5	C_6
AAPL	0.115	0.213	0.016	**0.516**	0.090	0.050
MSFT	0.147	0.068	0.060	**0.494**	0.137	0.095
AMZN	**0.891**	0.022	0.006	0.041	0.012	0.028
GOOGL	**0.757**	0.039	0.013	0.127	0.028	0.036
GOOG	**0.923**	0.012	0.005	0.032	0.009	0.019
FB	0.003	**0.983**	0.001	0.005	0.005	0.004
TSLA	0.068	0.040	0.009	**0.824**	0.038	0.021
BRK.B	0.000	0.000	**1**	0.000	0.000	0.000
V	0.006	0.019	0.006	0.025	**0.933**	0.010
JNJ	0.004	0.013	0.004	0.013	**0.959**	0.008
WMT	0.032	0.045	0.022	0.032	0.035	**0.834**
JPM	0.096	0.067	0.055	0.122	0.089	**0.572**
MA	0.129	0.160	0.017	**0.499**	0.087	0.109
PG	0.087	0.071	0.120	0.117	0.118	**0.486**
UNH	0.010	**0.905**	0.004	0.026	0.037	0.018
DIS	0.057	0.120	0.014	**0.619**	0.128	0.061
NVDA	0.044	0.038	0.013	**0.802**	0.077	0.025
HD	0.035	0.052	0.015	0.032	0.030	**0.836**
PYPL	0.189	**0.420**	0.019	0.172	0.071	0.129
BAC	0.035	0.042	0.011	0.041	0.029	**0.841**

the optimal value for the fuzziness parameter [91]. Here, our concern is to determine the combination of these two values which must be considered to optimize the clustering process. The goal is simple: starting from a grid of input parameters (several values for C and m), we solve the minimization problem in (14) and select the values of C and m that lead to the minimum average value of the indexes XBI, KI, TI and BI. The process is carried out as follows.

Step 1. Select the initial grid of values for the number of clusters, C, and the fuzziness parameter, m; and the number of random initializations for the fuzzy C-means clustering procedure, R.

Step 2. Solve the minimization problem in (14) for each pair (C, m) in the grid by considering R random initializations for the membership matrix. Store the resulting membership matrices and centroids.

Step 3. For each pair (C, m) in the grid, obtain the R values of (20), (21), (22) and (23) and select the corresponding minimum values. The results of this step are 4 vectors of length Cm storing the corresponding optimal values of the indexes, $\boldsymbol{v}_{\mathrm{XBI}}$, $\boldsymbol{v}_{\mathrm{KI}}$, $\boldsymbol{v}_{\mathrm{TI}}$ and $\boldsymbol{v}_{\mathrm{BI}}$.

Step 4. Scale the vectors resulting from *Step 3* so that they have zero mean and unit variance to get $\boldsymbol{v}^s_{\mathrm{XBI}}$, $\boldsymbol{v}^s_{\mathrm{KI}}$, $\boldsymbol{v}^s_{\mathrm{TI}}$ and $\boldsymbol{v}^s_{\mathrm{BI}}$.

Step 5. Average the vectors computed in *Step 4*, obtaining the vector

$$\boldsymbol{v}_{\mathrm{avg}} = \frac{\boldsymbol{v}^s_{\mathrm{XBI}} + \boldsymbol{v}^s_{\mathrm{KI}} + \boldsymbol{v}^s_{\mathrm{TI}} + \boldsymbol{v}^s_{\mathrm{BI}}}{4}.$$

Step 6. Return the pair (C, m) associated with the minimum value in the vector $\boldsymbol{v}_{\mathrm{avg}}$.

Note that standardization in *Step 4* is necessary to bring the four indexes to the same scale.

We selected a grid between 1 and 10 (step size of 1) for the number of clusters C and a grid between 1.1 and 3 (step size of 0.1) for the fuzziness parameter m and applied the previous selection procedure by taking into account the QCD-FCMn algorithm. The optimal (C, m) pair was $(6, 1.9)$. The value $C = 6$ seems consistent with the plot in Fig. 8.

Table 10 shows the membership degrees obtained by taken into consideration the previous values of C and m. For each single corporation, the entries in bold enhance the highest membership degree, i.e., the cluster assignment from a crisp perspective.

On balance, the clustering partition obtained with the QCD-FCMn model is quite consistent with the distances between the points in Fig. 8. Cluster C_4 contains the companies AAPL, MSFT, TSLA, MasterCard (MA), DIS and

NVIDIA (NVDA) with high membership. Cluster C_1 consists of AMZN and the two branches of Google (GOOGL, GOOG). It is worth noting that AAPL and MSFT also possess a non negligible membership degree in this cluster. The remaining one of the biggest five companies, FB, presents a large membership degree in cluster C_2 (0.983). It is noticeable that FB is barely present in clusters C_1 and C_4. This fact suggests that its business model greatly differs from that of its tech giants counterparts. Cluster C_2 is also constituted by the health insurance company UnitedHelath (UNH) with high membership and by the online payments company PYPL. Interestingly enough, this latter company is substantially spread out among the six clusters. Indeed, most of the companies in Table 10 offer PYPL as a form of online payment in their corresponding websites. Thus, it is expected that the financial behaviour of PYPL is somehow related to that of the remaining corporations. The multinational conglomerate holding company BRK.B forms itself an isolated cluster (C_3), which is congruent with Fig. 8 where BRK.B is represented as the most outlying point. An economical explanation of this fact can be easily obtained: BRK.B is not the typical corporation. This business model consists of investing in and holding dozens of major public and private companies. Thus, it is not surprising that it shares no similarity with the rest of the firms. Cluster C_5 is formed by V and JNJ presenting large membership degrees. Finally, cluster C_6 includes the three consumer-based corporations WMT, PG and HD, and the two banking companies JPM and BAC.

Note that a great deal of information can be extracted from the resulting partition by virtue of the fuzzy nature of the clustering solution. For instance, by looking at Table 10, one investor could determine that, whereas the financial behaviour of BRK.B is not shared by any of the remaining firms, there are another companies as PYPL and PG which show a changing behaviour. This type of insights could be invaluable in order to make informed investing decisions.

5.2. Fuzzy clustering of air pollution data

Now we develop a study case related to the non-supervised classification of geographical zones in terms of their temporal records of air pollutants. In the Spanish autonomous community of Galicia the air quality is analysed by a monitoring network consisting of 14 public and 33 private stations situated at different locations. These 47 stations provide hourly data on air pollutant concentration. The following pollutant concentrations: SO_2, NO, NO_2, NO_X, CO, O_3, PM_{10} (particulate matter 10 micrometers or less in diameter) and $PM_{2.5}$ (particulate matter 2.5 micrometers or less in diameter) are recorded in at least one station of the network.

We considered trivariate time series of hourly concentrations of nitrogen dioxide (NO_2), ozone (O_3) and nitrogen monoxide (NO) during the whole year 2018 in 20 different stations. The choice of this subset of pollutants was based on (i) several studies have uncovered serious health effects associated with the continuous exposure to high levels of NO_2, O_3 and NO, [92–94] and (ii) they were the most monitored gases throughout the network. It is important to highlight that our intention is only to show the usefulness of the proposed clustering algorithms without seeking to give any type of environmental implications, although this study together with other analyses could lead to the taking of some steps in order to reduce pollution. The 20 corresponding stations are Ferrol-Parque RS (FE), Coruña-Torre de Hércules (CO-T), Coruña-Riazor (CO-R), Lugo-Fingoy (LU), Santiago-Campus (SDC-C), Santiago-San Caetano (SDC-SC), Sur (SU), Pontevedra-Campolongo (PO-CL), Vigo-Coia (VGO-CO), Vigo-Lope de Vega (VGO-L), Ponteareas (PT), Ourense-Gómez Franqueira (OR), Pontevedra-Campelo (PO-CP), Fraga Redonda (FR), Xove (XO), Vigo-Citroën (VGO-CT), Paiosaco-Laracha (PA), Magdalena (MA), Louseiras (LO) and Mourence (MO). All data were sourced from the website of Ministry for the Ecological Transition and the Demographic challenge[2]. Table 11 contains some general information about the location of the stations. The categorization of the location of each station as "urban", "suburban", "rural" and "near power plant" was made based on information provided in the above website. Thus, from an environmental point of view, it is reasonable to think that the joint behaviour of the concentration of the three considered gases is different depending on where the station is situated.

The 20 MTS available are formed by $T = 8760$ hourly records and are non-stationary in mean. For this reason, the former series were transformed by taking the first differences of the natural logarithm of the original values. The new series are depicted in Fig. 9. It can be observed that the behaviour of the trivariate series is substantially different among the considered stations. In addition, it is reasonable to think that a fuzzy behaviour might be present, with MTS sharing features of distinct and well-defined patterns of hourly changes of concentrations of NO_2, O_3 and NO.
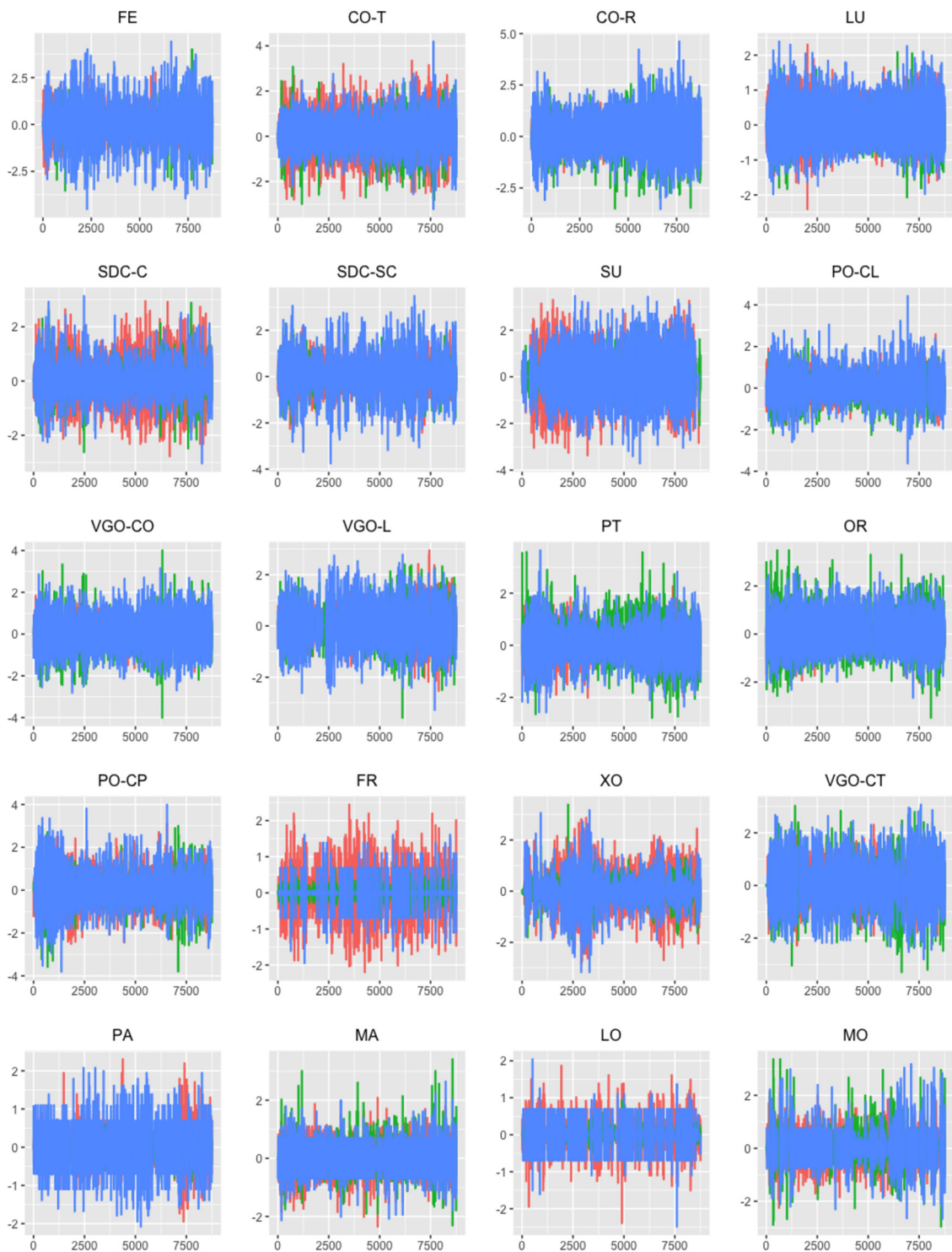

Fig. 9. Transformed levels of NO_2 (red colour) O_3 (green colour) and NO (blue colour) in the 20 monitoring stations of Galicia.

Table 11
Location information of the 20 stations monitoring the air quality in the community of Galicia.

Station	Abbreviation	Location area
Ferrol-Parque RS	FE	Suburban
Coruña-Torre de Hércules	CO-T	Urban
Coruña-Riazor	CO-R	Urban
Lugo-Fingoy	LU	Urban
Santiago-Campus	SDC-C	Suburban
Santiago-San Caetano	SDC-SC	Urban
Sur	SU	Rural
Pontevedra-Campolongo	PO-CL	Urban
Vigo-Coia	VGO-CO	Urban
Vigo-Lope de Vega	VGO-L	Urban
Ponteareas	PT	Suburban
Ourense-Gómez Franqueira	OR	Urban
Pontevedra-Campelo	PO-CP	Rural
Fraga Redonda	FR	Near power plant
Xove	XO	Rural
Vigo-Citroën	VGO-CT	Urban
Paiosaco-Laracha	PA	Near power plant
Magdalena	MA	Near power plant
Louseiras	LO	Near power plant
Mourence	MO	Near power plant
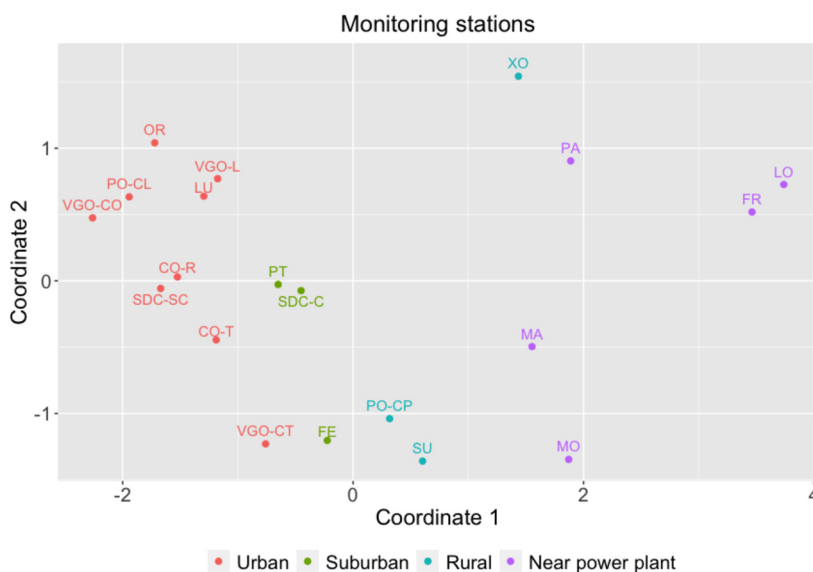
Fig. 10. Two-dimensional scaling plane based on the QCD distances for the hourly levels of NO_2, O_3 and NO in the 20 monitoring stations of Galicia.

First, as in the above case study, we carried out a 2DS based on the QCD-dissimilarity matrix. The resulting 2DS plane in Fig. 10, whose respective R^2 value is 0.8921, gives illustrative insights into the proximity of the time series according to the QCD-based distance. The points have been coloured according to the categories introduced in Table 11 concerning the location of the stations.

Fig. 10 shows that the QCD-based distance effectively captures the underlying categories of the stations according to Table 11. In particular the first coordinate clearly separates the stations with regards to the specific location. Stations situated in urban areas are located in the left side, whereas those positioned near a power plant are placed in the right side. The remaining categories, which correspond to suburban and rural locations, lie somewhere in the middle. It is

Table 12
Membership degrees for the 20 monitoring stations in Galicia by considering the QCD-FCMn model and a 3-cluster partition.

Station	Area	C_1	C_2	C_3
FE	Suburban	0.113	0.035	**0.853**
CO-T	Urban	**0.701**	0.041	0.258
CO-R	Urban	**0.799**	0.042	0.159
LU	Urban	**0.952**	0.010	0.039
SDC-C	Suburban	*0.452*	*0.044*	*0.504*
SDC-SC	Urban	**0.928**	0.012	0.060
SU	Rural	0.089	0.076	**0.835**
PO-CL	Urban	**0.963**	0.008	0.029
VGO-CO	Urban	**0.938**	0.014	0.048
VGO-L	Urban	**0.903**	0.020	0.076
PT	Suburban	*0.469*	*0.069*	*0.462*
OR	Urban	**0.944**	0.014	0.042
PO-CP	Rural	0.022	0.012	**0.966**
FR	Near power plant	0.018	**0.947**	0.036
XO	Rural	*0.191*	*0.538*	*0.271*
VGO-CT	Urban	0.227	0.043	**0.730**
PA	Near power plant	0.049	**0.850**	0.101
MA	Near power plant	*0.088*	*0.345*	*0.567*
LO	Near power plant	0.023	**0.933**	0.044
MO	Near power plant	*0.115*	*0.294*	*0.590*

worth pointing out that the five stations situated near power plants also pertain to rural zones. Note that, whereas the distinction between urban stations and those located in rural regions or near a power plant is evident, the classification of the suburban locations seems vaguer, implying that the dependence relationship of the three pollutants could exhibit traits related to both urban and rural areas. Therefore, a fuzzy clustering approach seems far more suitable to tackle the grouping task of the 20 stations than a hard clustering one.

By observing Fig. 10 and ignoring the underlying categories, one could hypothesize the existence of three or four underlying clusters. There is a first cluster formed by all the urban stations except for VGO-CT together with PT and SDC-C. VGO-CT, FE, PO-CP and SU constitute another cluster, as well as the two stations near a thermal power plant, MA and MO. Finally, XO, PA, FR and LO could be considered to compose one or two different clusters.

As in the preceding application, the QCD-FCMn approach was applied to the series in Fig. 9. Selection of the optimal values for m and C was effectuated by using the same procedure as in the above analysis. After performing the corresponding steps, the existence of three major groups ($C = 3$) was concluded, which is consistent with the 2DS plot in Fig. 10. The optimal value for the fuzziness coefficient was $m = 1.9$.

The resulting 3-cluster fuzzy partition is displayed in Table 12. For a given element, the highest membership degree is shown in bold provided that its value is larger than 0.6. As 3 clusters are being considered, this cutoff seems a sensible choice. When the three membership values are below this cutoff, the corresponding quantities are written in italic font. Fundamentally, the model QCD-FCMn produces the expected classification by grouping the transformed series according to the kind of location where the respective stations are placed. Regarding the urban locations, they are grouped together in cluster C_1 with a high membership aside from VGO-CT, which mainly pertains to cluster C_3. This cluster brings together the mentioned VGO-CT, two rural stations (SU and PO-CP) and one suburban (FE) station. The misclassification of VGO-CT with respect to the underlying location can be explained because, despite being in the same city as VGO-CO and VGO-L, it is located in a car factory, probably exposing himself to special types of emissions avoided by VGO-CO and VGO-L. The last cluster, C_2, includes three of the stations close to a thermal power plant.

The stations concerning the fuzziest allocations were SDC-C, PT, XO, MA and MO. SDC-C and PT correspond to suburban areas, and are placed in C_1 (urban cluster) with membership values of 0.452 and 0.469, respectively, and in C_3 (rural cluster) with membership values of 0.504 and 0.462, respectively. XO is the station displaying the most spread between clusters, corresponding its largest membership degree to cluster C_2. Finally, both MA and MO are mainly assigned to clusters C_2 and C_3, showing a higher membership in the later. Note that all the previous

Table 13
FARI obtained by the fuzzy C-means procedures in grouping
the 20 monitoring stations in Galicia. The ground truth is given
by the location of the stations.

Method	QCD-FCMn	W-FCMn	C-FCMn	F-FCMn
FARI	**0.3649**	0.2253	0.2083	0.0040

assignments are coherent according to the particularities of each cluster. The stations placed in suburban areas are expected to share traits of both urban and rural regions. In the same way, as the five stations located close to a power plant belong to rural zones, they are presumed to show features typical of rural locations. These results stress the power of a fuzzy allocation when there exist overlapping classes.

For the sake of illustration and comparison purposes, we have obtained the clustering solutions according to the approaches W-FCMn, C-FCMn and F-FCMn. The selection of C and m was carried out similarly than in the above analyses. The corresponding fuzzy partitions were obtained for the optimal values of m and C. Assuming that the true partition is given by the categories standing for the location of each station, the FARI was obtained for the three mentioned procedures as well as QCD-FCMn. Note that this quantity is a good indicator of to what extent each procedure is capable of deriving the underlying geographical distribution. The results are displayed in Table 13. QCD-FCMn significantly outperformed the remaining approaches. Whereas W-FCMn and C-FCMn were able to figure out some insights into the region of each station, F-FCMn produced a random partition according to this criterion. Lastly, it is worth enhancing that all the procedures have determined some series showing a fuzzy nature, which supports the helpfulness of the fuzzy approach.

6. Concluding remarks and future work

In this work we have proposed two novel approaches for fuzzy clustering of MTS based on the quantile cross-spectral density (QCD) and principal component analysis (PCA), the so-called QCD-FCMn and QCD-FCMd. The former builds on the traditional fuzzy C-means algorithm while the latter employs the fuzzy C-medoids. The methods utilise a slight modification of the distance d_{QCD} in our previous work [31] regarding the projection of the QCD-based features onto the principal components space. The advantages of performing dimensionality reduction via PCA in terms of clustering effectiveness have been shown through a motivating example.

To evaluate the performance of QCD-FCMn and QCD-FCMd, we have carried out numerical experiments including scenarios formed by MTS pertaining to well-defined clusters and scenarios involving series equidistant from two clusters. Several types of generating processes were considered. The assessment task was executed in two different fashions. Concerning scenarios lacking a switching series, a fuzzy extension of the Adjusted Rand Index was considered. This way, the quality of the resulting fuzzy partition, in the sense of assigning high membership values to the correct clusters, was directly evaluated. On the other hand, the capability of the techniques in scenarios incorporating an equidistant series was measured also by taking into account their ability to determine the fuzzy nature of this series. The methods were compared with other alternative dissimilarities suggested in the literature. Regardless of the considered models and assessment schemes, QCD-FCMn and QCD-FCMd produced the best results, the former slightly outperforming the latter overall. Both methods inherit the powerful characteristics of QCD, as no requirements about the existence of moments, robustness to changes in the error distribution, and computational efficiency. They also preserve the properties of the former distance d_{QCD}, as being able to uncover any type of disparity in the dependence structure of two MTS. Two specific case studies involving environmental and financial databases have illustrated the usefulness of the proposed techniques.

It is worth pointing out that this paper represents an original extension of [31] in four different ways. First, this manuscript presents explicitly some useful properties of the distance d_{QCD} indicating that this dissimilarity is able to detect any type of discrepancy in the dependence structure of two generating processes whatever their complexity. Second, this work highlights the power of applying dimensionality reduction techniques to QCD-based features in relation to clustering performance. Third, whereas the simulated scenarios in [31] did not include nonlinear processes, we have considered them here in order to examine the approaches under a vast assortment of generating patterns. Lastly, our study in [31] was limited to crisp clustering procedures. In this article we introduced fuzzy clustering strategies, thus combining the versatility of the fuzzy logic by permitting overlapping clusters with the high ability

of the QCD-based metric to differentiate between underlying mechanisms. In fact, the superiority showed by the proposed dissimilarity over the alternative metrics in a fuzzy context is substantially greater than that in a crisp framework. In short, this paper contributes to the few works on fuzzy clustering of MTS based on generating processes.

There are indeed some appealing issues for further research in relation to the use of QCD in soft clustering of MTS. Specifically, this work can be extended in two different ways. On the one hand, it would be interesting to obtain robust versions of QCD-FCMn and QCD-FCMd capable of properly neutralizing the effect of outlying MTS. For instance, we could consider the techniques used in [95] and [18], namely the metric approach (by smoothing the distance), the trimmed approach (by trimming away a small proportion of the series) and the noise approach (by considering a noise cluster expected to contain the outlying series). On the other hand, note that, by using the smoothed CCR-periodograms to describe an MTS dataset, each MTS is characterized by a set of curves of the form

$$\left\{ W\left(\hat{G}_{T,R}^{j_1,j_2}(\omega, \tau, \tau') \right), 1 \le j_1, j_2 \le d, \tau, \tau' \in \mathcal{T} \right\}, \tag{24}$$

where $W(\cdot)$ is used interchangeably to denote the real part and the imaginary part operator. Our numerical studies have revealed that some of these curves contain far more information than others in terms of the generating process of each MTS. Thus, it would be reasonable to create a fuzzy clustering algorithm giving more importance to the functions with more discriminative power. This could be naturally accomplished by introducing weights in the objective functions (14) and (17). Even an approach considering only two weights, for real and imaginary parts, respectively, could be devised. The mentioned topics for further research will be properly addressed in the upcoming months.

CRediT authorship contribution statement

Not applicable.

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.

Availability of data and materials

Not applicable.

Code availability

The code and the data used to perform the analysis carried out in Section 5.1 is available in https://github.com/anloor7/PhD_degree/tree/master/r_code/papers/paper_fuzzy_qcd. The computation of the PCA-feature matrix shown there can be used to replicate the results provided throughout the whole paper. Please contact the corresponding author for further information.

Appendix A

We now derive Properties 1 and 2 in Section 2.3. Property 1 follows directly from the definition of QCD, the fact that the smoothed CCR-periodogram is a consistent estimator of this quantity and the definition of d_{QCD}.

To show Property 2, note that the assumption on the continuity of the cumulative probability distribution functions implies that there exists a neighbourhood V of $\mathbb{R}^2$ in which $F^1_{j_1,j_2,l}$ and $F^2_{j_1,j_2,l}$ differ from one another. Let $(c,d) \in V$, and therefore $F^1_{j_1,l}(c,d) \neq F^2_{j_1,j_2,l}(c,d)$. By virtue of Sklar's theorem, every cumulative distribution function can be expressed by means of its marginals and a unique copula. Hence, if $C^1_{j_1,j_2,l}$ and $C^2_{j_1,j_2,l}$ denote the copulas associated with the pairs $(X^1_{t,j_1}, X^1_{t+l,j_2})$ and $(X^2_{t,j_1}, X^2_{t+l,j_2})$, respectively, we can write

$$
\begin{aligned}
F^1_{j_1,j_2,l}(x_1,x_2) &= C^1_{j_1,j_2,l}(F^1_{j_1}(x_1), F^1_{j_2}(x_2)), \\
F^2_{j_1,j_2,l}(y_1,y_2) &= C^2_{j_1,j_2,l}(F^2_{j_1}(y_1), F^2_{j_2}(y_2)),
\end{aligned}
\tag{25}
$$

for all $x_1, x_2, y_1, y_2 \in \mathbb{R}$. Take now $x_1 = y_1 = c$, $x_2 = y_2 = d$. From the fact that $F^1_{j_1,j_2,l}(c,d) \neq F^2_{j_1,j_2,l}(c,d)$ and the equality of the marginal distributions, we have

$$
C^1_{j_1,j_2,l}(F^1_{j_1}(c), F^1_{j_2}(d)) \neq C^2_{j_1,j_2,l}(F^1_{j_1}(c), F^1_{j_2}(d)).
\tag{26}
$$

Now, consider the following relationships

$$
\begin{aligned}
\int_{-\pi}^{\pi} \mathfrak{f}^1_{j_1,j_2}(\omega, \tau_1, \tau_2) e^{il\omega} d\omega &= C^1_{j_1,j_2,l}(\tau_1,\tau_2) - \tau_1\tau_2, \\
\int_{-\pi}^{\pi} \mathfrak{f}^2_{j_1,j_2}(\omega, \tau_3, \tau_4) e^{il\omega} d\omega &= C^2_{j_1,j_2,l}(\tau_3,\tau_4) - \tau_3\tau_4,
\end{aligned}
\tag{27}
$$

where $\mathfrak{f}^i_{j_1,j_2}$, is the quantile cross-spectral density for the pair of processes X^i_{t,j_1} and X^i_{t,j_2}, $i = 1,2$, $\omega \in \mathbb{R}$ is an arbitrary frequency, and (τ_1, τ_2) and (τ_3, τ_4) are arbitrary couples of probability levels. Then, denoting by $\tau = F^1_{j_1}(c)$ and $\tau' = F^1_{j_2}(d)$, (26) and (27) allow to conclude that

$$
\mathfrak{f}^1_{j_1,j_2}(\omega', \tau, \tau') \neq \mathfrak{f}^2_{j_1,j_2}(\omega', \tau, \tau')
\tag{28}
$$

for some $\omega' \in V_\omega \subset \mathbb{R}$, with V_ω a neighbourhood where the functions differ. From (28) we know that at least one of the facts $\Re(\mathfrak{f}^1_{j_1,j_2}(\omega', \tau, \tau')) \neq \Re(\mathfrak{f}^2_{j_1,j_2}(\omega', \tau, \tau'))$ or $\Im(\mathfrak{f}^1_{j_1,j_2}(\omega', \tau, \tau')) \neq \Im(\mathfrak{f}^2_{j_1,j_2}(\omega', \tau, \tau'))$ is true. Assume without loss of generality that $\Re(\mathfrak{f}^1_{j_1,j_2}(\omega', \tau, \tau')) \neq \Re(\mathfrak{f}^2_{j_1,j_2}(\omega', \tau, \tau'))$. Now, in order to compute the distance d_{QCD}, select $\mathcal{T}$ and Ω such that $\tau, \tau' \in \mathcal{T}$ and $\omega' \in \Omega$. Let $\mathbf{\Psi}^{(1)}$ and $\mathbf{\Psi}^{(2)}$ be the feature vectors computed from the realizations $X^{(1)}_t$ and $X^{(2)}_t$, respectively. From the definition of d_{QCD}, it necessarily exists an integer k such that the k-th components of vectors $\mathbf{\Psi}^{(1)}$ and $\mathbf{\Psi}^{(2)}$ are given by $\Re(\hat{G}^{1,j_1,j_2}_{T,R}(\omega', \tau, \tau'))$ and $\Re(\hat{G}^{2,j_1,j_2}_{T,R}(\omega', \tau, \tau'))$, respectively, being $\hat{G}^{1,j_1,j_2}_{T,R}$ and $\hat{G}^{2,j_1,j_2}_{T,R}$ the smoothed CCR-periodograms computed from the realizations $X^{(1)}_t$ and $X^{(2)}_t$, respectively. As the smoothed CCR-periodogram is a consistent estimate of the quantile cross-spectral density, consistency also holds for real and imaginary parts so we have

$$
\begin{aligned}
\Re(\hat{G}^{1,j_1,j_2}_{T,R}(\omega', \tau, \tau')) &\xrightarrow{p} \Re(\mathfrak{f}^1_{j_1,j_2}(\omega', \tau, \tau')), \\
\Re(\hat{G}^{2,j_1,j_2}_{T,R}(\omega', \tau, \tau')) &\xrightarrow{p} \Re(\mathfrak{f}^2_{j_1,j_2}(\omega', \tau, \tau')),
\end{aligned}
\tag{29}
$$

from which the convergence in probability of the distance d_{QCD} to some $a \neq 0$ is trivially derived. Note that by choosing another pair (c',d'), with $c' \neq c$ and $d' \neq d$ in the neighbourhood V, a different pair of probability levels such that the result holds could be extracted. The same is true for another $\omega'' \in V_\omega$ This process could be repeated infinitely. The proof is completed. □

References

[1] T.W. Liao, Clustering of time series data: a survey, Pattern Recognit. 38 (11) (2005) 1857–1874.

[2] T.-c. Fu, A review on time series data mining, Eng. Appl. Artif. Intell. 24 (1) (2011) 164–181.

[3] S. Rani, G. Sikka, Recent techniques of clustering of time series data: a survey, Int. J. Comput. Appl. 52 (15) (2012) 1–9.

[4] S. Aghabozorgi, A.S. Shirkhorshidi, T.Y. Wah, Time-series clustering—a decade review, Inf. Sci. 53 (2015) 16–38.
[5] E. Maharaj, P. D'Urso, J. Caiado, Time Series Clustering and Classification, Chapman & Hall/CRC Computer Science and Data Analysis Series, CRC Press, 2019.
[6] P. Montero, J.A. Vilar, TSclust: an R package for time series clustering, J. Stat. Softw. 62 (1) (2014) 1–43.
[7] P. Montero, J.A. Vilar, TSclust: time series clustering utilities, R package version 1.2.1, http://CRAN.R-project.org/package=TSclust, 2014.
[8] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Springer Science & Business Media, 2013.
[9] S. Miyamoto, H. Ichihashi, K. Honda, H. Ichihashi, Algorithms for Fuzzy Clustering, Springer, 2008.
[10] P. D'Urso, E.A. Maharaj, Autocorrelation-based fuzzy clustering of time series, Fuzzy Sets Syst. 160 (24) (2009) 3565–3589.
[11] E.A. Maharaj, P. D'Urso, Fuzzy clustering of time series in the frequency domain, Inf. Sci. 181 (7) (2011) 1187–1211.
[12] E.A. Maharaj, P. D'Urso, D.U. Galagedera, Wavelet-based fuzzy clustering of time series, J. Classif. 27 (2) (2010) 231–275.
[13] P. D'Urso, E.A. Maharaj, A.M. Alonso, Fuzzy clustering of time series using extremes, Fuzzy Sets Syst. 318 (2017) 56–79.
[14] P. D'Urso, C. Cappelli, D. Di Lallo, R. Massari, Clustering of financial time series, Physica A 392 (9) (2013) 2114–2129.
[15] P. D'Urso, L. De Giovanni, R. Massari, Garch-based robust clustering of time series, Fuzzy Sets Syst. 305 (2016) 1–28.
[16] J.A. Vilar, B. Lafuente-Rego, P. D'Urso, Quantile autocovariances: a powerful tool for hard and soft partitional clustering of time series, Fuzzy Sets Syst. 340 (2018) 38–72.
[17] B. Lafuente-Rego, J.A. Vilar, Clustering of time series using quantile autocovariances, Adv. Data Anal. Classif. 10 (3) (2016) 391–415.
[18] B. Lafuente-Rego, P. D'Urso, J. Vilar, Robust fuzzy clustering based on quantile autocovariances, Stat. Pap. (2018) 1–56.
[19] M. Disegna, P. D'Urso, F. Durante, Copula-based fuzzy clustering of spatial time series, Spat. Stat. 21 (2017) 209–225.
[20] H. Izakian, W. Pedrycz, I. Jamal, Fuzzy clustering of time series data using dynamic time warping distance, Eng. Appl. Artif. Intell. 39 (2015) 235–244.
[21] P. D'Urso, L. De Giovanni, R. Massari, Trimmed fuzzy clustering of financial time series based on dynamic time warping, Ann. Oper. Res. (2019) 1–17.
[22] P. D'Urso, Dissimilarity measures for time trajectories, Stat. Methods Appl. 9 (1) (2000) 53–83.
[23] P. D'Urso, Fuzzy c-means clustering models for multivariate time-varying data: different approaches, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 12 (03) (2004) 287–326.
[24] P. D'Urso, Fuzzy clustering for data time arrays with inlier and outlier time trajectories, IEEE Trans. Fuzzy Syst. 13 (5) (2005) 583–604.
[25] R. Coppi, P. D'Urso, Fuzzy unsupervised classification of multivariate time trajectories with the Shannon entropy regularization, Comput. Stat. Data Anal. 50 (6) (2006) 1452–1477.
[26] P. D'Urso, E.A. Maharaj, Wavelets-based clustering of multivariate time series, Fuzzy Sets Syst. 193 (2012) 33–61.
[27] H. He, Y. Tan, Unsupervised classification of multivariate time series using VPCA and fuzzy clustering with spatial weighted matrix distance, IEEE Trans. Cybern. 50 (3) (2018) 1096–1105.
[28] P. D'Urso, L. De Giovanni, R. Massari, Robust fuzzy clustering of multivariate time trajectories, Int. J. Approx. Reason. 99 (2018) 12–38.
[29] H. Li, M. Wei, Fuzzy clustering based on feature weights for multivariate time series, Knowl.-Based Syst. 197 (2020) 105907.
[30] J. Baruník, T. Kley, Quantile coherency: a general measure for dependence between cyclical economic variables, Econom. J. 22 (2) (2019) 131–152.
[31] Á. López-Oriona, J.A. Vilar, Quantile cross-spectral density: A novel and effective tool for clustering multivariate time series, Expert Syst. Appl. 185 (2021) 115677.
[32] A. Bagnall, H.A. Dau, J. Lines, M. Flynn, J. Large, A. Bostrom, P. Southam, E. Keogh, The UEA multivariate time series classification archive, arXiv preprint, arXiv:1811.00075, 2018.
[33] X. Wang, A. Wirth, L. Wang, Structure-based statistical features and multivariate time series clustering, in: Seventh IEEE International Conference on Data Mining, ICDM 2007, IEEE, 2007, pp. 351–360.
[34] J. Lee, S.S. Rao, The quantile spectral density and comparison based tests for nonlinear time series, arXiv:1112.2759, 2012.
[35] H. Dette, M. Hallin, T. Kley, S. Volgushev, Of copulas, quantiles, ranks and spectra: an l_1-approach to spectral analysis, Bernoulli 21 (2) (2015) 781–831.
[36] T. Kley, S. Volgushev, H. Dette, M. Hallin, Quantile spectral processes: asymptotic analysis and inference, Bernoulli 22 (3) (2016) 1770–1807.
[37] T. Kley, Quantile-based spectral analysis in an object-oriented framework and a reference implementation in R: the quantspec package, J. Stat. Softw. 70 (3) (2016) 1–27, https://doi.org/10.18637/jss.v070.i03.
[38] Á. López-Oriona, J.A. Vilar, F4: an all-purpose tool for multivariate time series classification, Mathematics 9 (23) (2021) 3051.
[39] Á. López-Oriona, J.A. Vilar, Outlier detection for multivariate time series: a functional data approach, Knowl.-Based Syst. 233 (2021) 107527.
[40] F. Höppner, F. Klawonn, R. Kruse, T. Runkler, Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition, John Wiley & Sons, 1999.
[41] B.S. Everitt, S. Landau, M. Leese, D. Stahl, Cluster Analysis, 5th ed., 2011.
[42] A. Ben-Hur, I. Guyon, Detecting stable clusters using principal component analysis, in: Functional Genomics, Springer, 2003, pp. 159–182.
[43] S. Raychaudhuri, J.M. Stuart, R.B. Altman, Principal components analysis to summarize microarray experiments: application to sporulation time series, in: Biocomputing 2000, World Scientific, 1999, pp. 455–466.
[44] J. Xue, C. Lee, S.G. Wakeham, R.A. Armstrong, Using principal components analysis (PCA) with cluster analysis to study the organic geochemistry of sinking particles in the ocean, Org. Geochem. 42 (4) (2011) 356–367.
[45] C. Ding, X. He, K-means clustering via principal component analysis, in: Proceedings of the Twenty-First International Conference on Machine Learning, 2004, p. 29.
[46] N. Gaitani, C. Lehmann, M. Santamouris, G. Mihalakakou, P. Patargias, Using principal component and cluster analysis in the heating evaluation of the school building sector, Appl. Energy 87 (6) (2010) 2079–2086.
[47] G.P. Zhang, B.E. Patuwo, M.Y. Hu, A simulation study of artificial neural networks for nonlinear time-series forecasting, Comput. Oper. Res. 28 (4) (2001) 381–396.

[48] C.W. Granger, T. Terasvirta, et al., Modelling Non-linear Economic Relationships, OUP Catalogue, 1993.

[49] C.W.J. Granger, A. Andersen, An Introduction to Bilinear Time Series Models, 1978.

[50] H. Tong, K.S. Lim, Threshold autoregression, limit cycles and cyclical data, in: Exploration of a Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics, World Scientific, 2009, pp. 9–56.

[51] E. Pereda, R.Q. Quiroga, J. Bhattacharya, Nonlinear multivariate analysis of neurophysiological signals, Prog. Neurobiol. 77 (1–2) (2005) 1–37.

[52] G. Koop, M.H. Pesaran, S.M. Potter, Impulse response analysis in nonlinear multivariate models, J. Econom. 74 (1) (1996) 119–147.

[53] R. Engle, Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models, J. Bus. Econ. Stat. 20 (3) (2002) 339–350.

[54] L. Bauwens, S. Laurent, J.V. Rombouts, Multivariate GARCH models: a survey, J. Appl. Econom. 21 (1) (2006) 79–109.

[55] Y.-H.H. Ku, H.-C. Chen, K.-h. Chen, On the application of the dynamic conditional correlation model in estimating optimal time-varying hedge ratios, Appl. Econ. Lett. 14 (7) (2007) 503–509.

[56] K. Naoui, N. Liouane, S. Brahim, A dynamic conditional correlation analysis of financial contagion: the case of the subprime credit crisis, Int. J. Econ. Finance 2 (3) (2010) 85–96.

[57] G.H. Kuper, et al., Dynamic conditional correlation analysis of financial market interdependence: an application to Thailand and Indonesia, J. Asian Econ. 18 (4) (2007) 670–684.

[58] M. Andersson, E. Krylova, S. Vähämaa, Why does the correlation between stock and bond returns vary over time?, Appl. Financ. Econ. 18 (2) (2008) 139–151.

[59] P. Arabie, J.D. Carroll, W. DeSarbo, J. Wind, Overlapping clustering: a new method for product positioning, J. Mark. Res. 18 (3) (1981) 310–317.

[60] R.L. Cannon, J.V. Dave, J.C. Bezdek, Efficient implementation of the fuzzy c-means clustering algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 2 (1986) 248–255.

[61] L.O. Hall, A.M. Bensaid, L.P. Clarke, R.P. Velthuizen, M.S. Silbiger, J.C. Bezdek, A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain, IEEE Trans. Neural Netw. 3 (5) (1992) 672–682.

[62] R.J. Campello, A fuzzy extension of the rand index and other related indexes for clustering and classification assessment, Pattern Recognit. Lett. 28 (7) (2007) 833–841.

[63] A.C. Harvey, Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series, vol. 52, Cambridge University Press, 2013.

[64] M. Bernardi, A. Maruotti, L. Petrella, Multiple risk measures for multivariate dynamic heavy–tailed models, J. Empir. Finance 43 (2017) 1–32.

[65] S.T. Rachev, Handbook of Heavy Tailed Distributions in Finance: Handbooks in Finance, Book 1, Elsevier, 2003.

[66] T. Mikosch, Modeling dependence and tails of financial time series, in: Extreme Values in Finance, Telecommunications, and the Environment, 2003, pp. 185–286.

[67] H.T. Pham, B.-S. Yang, et al., A hybrid of nonlinear autoregressive model with exogenous input and autoregressive moving average model for long-term machine state forecasting, Expert Syst. Appl. 37 (4) (2010) 3310–3317.

[68] S. Lu, K.H. Chon, Nonlinear autoregressive and nonlinear autoregressive moving average model parameter estimation by minimizing hyper-surface distance, IEEE Trans. Signal Process. 51 (12) (2003) 3020–3026.

[69] I.M. Yassin, A. Zabidi, M.S. Amin Megat Ali, N. Md Tahir, H. Zainol Abidin, Z.I. Rizman, Binary particle swarm optimization structure selection of nonlinear autoregressive moving average with exogenous inputs (NARMAX) model of a flexible robot arm, Int. J. Adv. Sci. Eng. Inf. Technol. 6 (5) (2016) 630–637.

[70] A. Awwad, H. Abu-Rub, H. Toliyat, Nonlinear autoregressive moving average (NARMA-L2) controller for advanced ac motor control, in: 2008 34th Annual Conference of IEEE Industrial Electronics, IEEE, 2008, pp. 1287–1292.

[71] R. Boynton, M. Balikhin, S. Billings, O. Amariutei, Application of nonlinear autoregressive moving average exogenous input models to geospace: advances in understanding and space weather forecasts, Ann. Geophys. 31 (2013) 1579–1589, Copernicus GmbH.

[72] O. Karakuş, E.E. Kuruoğlu, M.A. Altınkaya, Bayesian estimation of polynomial moving average models with unknown degree of nonlinearity, in: 2016 24th European Signal Processing Conference, EUSIPCO, IEEE, 2016, pp. 1543–1547.

[73] E.M. Delgado, M.R. Cogollo, J.D. Velásquez, Parameters estimation for nonlinear moving average model using supernova metaheuristic, in: 2012 7th Colombian Computing Congress, CCC, IEEE, 2012, pp. 1–5.

[74] M.Y. Khan, Advances in applied nonlinear time series modeling, Ph.D. thesis, LMU, 2015.

[75] C.L. Weise, The asymmetric effects of monetary policy: a nonlinear vector autoregression approach, J. Money Credit Bank. (1999) 85–108.

[76] J.F. Hair, Multivariate Data Analysis, 2009.

[77] G. Kou, Y. Peng, G. Wang, Evaluation of clustering algorithms for financial risk analysis using MCDM methods, Inf. Sci. 275 (2014) 1–12.

[78] F. Durante, R. Pappadà, N. Torelli, Clustering of financial time series in risky scenarios, Adv. Data Anal. Classif. 8 (4) (2014) 359–376.

[79] F. Pattarin, S. Paterlini, T. Minerva, Clustering financial time series: an application to mutual funds style analysis, Comput. Stat. Data Anal. 47 (2) (2004) 353–372.

[80] J.G. Dias, J.K. Vermunt, S. Ramos, Clustering financial time series: new insights from an extended hidden Markov model, Eur. J. Oper. Res. 243 (3) (2015) 852–864.

[81] J.A. Bastos, J. Caiado, Clustering financial time series with variance ratio statistics, Quant. Finance 14 (12) (2014) 2121–2133.

[82] J.M. Karpoff, The relation between price changes and trading volume: a survey, J. Financ. Quant. Anal. (1987) 109–126.

[83] J.Y. Campbell, S.J. Grossman, J. Wang, Trading volume and serial correlation in stock returns, Q. J. Econ. 108 (4) (1993) 905–939.

[84] B. Gebka, M.E. Wohar, Causality between trading volume and returns: evidence from quantile regressions, Int. Rev. Econ. Finance 27 (2013) 144–159.

[85] T.A. Schmitt, D. Chetalova, R. Schäfer, T. Guhr, Non-stationarity in financial time series: generic features and tail behavior, Europhys. Lett. 103 (5) (2013) 58003.

[86] B.O. Bradley, M.S. Taqqu, Financial risk and heavy tails, in: Handbook of Heavy Tailed Distributions in Finance, Elsevier, 2003, pp. 35–103.

[87] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 13 (8) (1991) 841–847.

[88] S.H. Kwon, Cluster validity index for fuzzy clustering, Electron. Lett. 34 (22) (1998) 2176–2177.

[89] Y. Tang, F. Sun, Z. Sun, Improved validation index for fuzzy clustering, in: Proceedings of the 2005, American Control Conference, IEEE, 2005, pp. 1120–1125.

[90] A.M. Bensaid, L.O. Hall, J.C. Bezdek, L.P. Clarke, M.L. Silbiger, J.A. Arrington, R.F. Murtagh, Validity-guided (re)clustering with applications to image segmentation, IEEE Trans. Fuzzy Syst. 4 (2) (1996) 112–123.

[91] K. Zhou, C. Fu, S. Yang, Fuzziness parameter selection in fuzzy c-means: the perspective of cluster validation, Sci. China Inf. Sci. 57 (11) (2014) 1–8.

[92] N.M. Elsayed, Toxicity of nitrogen dioxide: an introduction, Toxicology 89 (3) (1994) 161–174.

[93] M.A. Mehlman, C. Borek, Toxicity and biochemical mechanisms of ozone, Environ. Res. 42 (1) (1987) 36–53.

[94] S. Depayras, T. Kondakova, H.J. Heipieper, M.G. Feuilloley, N. Orange, C. Duclairoir-Poc, The hidden face of nitrogen oxides species: from toxic effects to potential cure?, in: Emerging Pollutants-Some Strategies for the Quality Preservation of Our Environment, IntechOpen, 2018.

[95] P. D'Urso, L. De Giovanni, Robust clustering of imprecise data, Chemom. Intell. Lab. Syst. 136 (2014) 58–80.