International Journal of Approximate Reasoning 150 (2022) 55–82

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Quantile-based fuzzy *C*-means clustering of multivariate time series: Robust techniques

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

[a] *Department of Mathematics, Research Group MODES, Research Center for Information and Communication Technologies (CITIC), University of A Coruña, Spain*
[b] *Department of Social Sciences and Economics, Sapienza University of Rome, Italy*

ARTICLE INFO

Article history:
Received 29 October 2021
Received in revised form 10 June 2022
Accepted 30 July 2022
Available online 6 August 2022

Keywords:
Multivariate time series
Robust fuzzy *C*-means
Quantile cross-spectral density
Exponential distance
Noise cluster
Trimming

ABSTRACT

Robust fuzzy clustering of multivariate time series is addressed when the clustering purpose is grouping together series generated from similar stochastic processes. Robustness to the presence of anomalous series is attained by considering three well-known robust versions of a fuzzy *C*-means model based on a spectral dissimilarity measure with high discriminatory power. The dissimilarity measure compares principal component scores obtained from estimates of quantile cross-spectral densities, and the robust techniques follow the so-called metric, noise and trimmed approaches. The metric approach incorporates in the objective function a distance aimed at neutralizing the effect of the outliers, the noise approach builds an artificial cluster expected to contain the outlying series, and the trimmed approach removes the most atypical series in the dataset. As result, the proposed clustering methods take advantage of both the robust nature of these techniques and the capability of the quantile cross-spectral density to identify complex dependence structures. An extensive simulation study including multivariate linear, nonlinear and GARCH processes shows that the algorithms are substantially effective in coping with the presence of outlying series, clearly outperforming other alternative procedures. Two specific applications regarding financial and environmental series illustrate the usefulness of the presented methods.

1. Introduction

In the last years there has been an unprecedented growth in complexity, speed and volume of data. In particular, time series data have become ubiquitous in our days, arising frequently in a broad variety of fields including medicine, computer science, finance, environmental sciences, machine learning, marketing and neuroscience, among many others. Typically, time series involve a huge number of records, present dynamic behaviour patterns which might change over time, and one frequently has to deal with realizations of different length. Due to this complex nature, standard techniques to perform data mining tasks as classification, clustering or anomaly detection often produce unsatisfactory results. Complexity is still greater by treating with multivariate time series, where the interdependence structure and high dimensionality are serious obstacles to develop efficient procedures. Univariate time series (UTS) were the main focus of intensive research until recently, but multivariate time series (MTS) have received lately a great deal of attention due to the advance of technology and storage

* Corresponding author.
E-mail addresses: oriona38@hotmail.com, a.oriona@udc.es (Á. López-Oriona), pierpaolo.durso@uniroma1.it (P. D'Urso), jose.vilarf@udc.es (J.A. Vilar), borja.lafuente@udc.es (B. Lafuente-Rego).

https://doi.org/10.1016/j.ijar.2022.07.010

capabilities of everyday devices. Well-known examples of MTS are multi-lead ECG signals of patients or temporal records of economic indicators of a given country, but many other examples can be easily obtained from different fields.

Among time series data mining tasks, clustering is a central problem. In fact, identifying groups of similar series is basic for many applications in order to detect a few representative patterns, forecast future performances, quantify affinity, recognize dynamic changes and structural breaks, … However, unlike traditional databases, similarity search in time series data is a complex issue that cannot be addressed with conventional methods. Notice that a suitable distance between time series objects should take into account the underlying dynamic patterns. This problem accentuates when coping with MTS data, since the interdependence relationship between the different dimensions makes more complex to characterize the dependence structure. Moreover, it is not uncommon for MTS databases to contain some outlying MTS, i.e., series exhibiting a very different behaviour from the remaining MTS in the dataset. This way, a potentially suitable distance is likely to fail in the clustering process if it lacks robustness to anomalous series. Computational complexity is also a challenging issue when performing MTS clustering. The high dimension of MTS samples implies a fast increase of the dataset size, thus requiring more storage space and computing resources. In fact, clustering procedures specifically designed to deal with UTS data can be inefficient or even infeasible with MTS data. In sum, the complex nature of the dependence structure and large dimensionality of MTS make particularly challenging to develop effective and robust clustering procedures.

The aim of this paper is to perform robust clustering of MTS so that the resulting partitions are not disrupted by the presence of anomalous series. In addition, the proposed algorithms are able to detect potential outlying elements, thus providing a way of simultaneously performing robust clustering and outlier identification. Note that this is a more challenging task than simply carrying out outlier detection in an MTS dataset, since the latter approach does not involve the detection of a hidden clustering structure.

It is assumed that the similarity principle relies on the underlying dependence structures, i.e., the target is grouping MTS generated by the same multivariate stochastic process. This interesting problem often arises in a natural way when dealing with sets of MTS consisting of environmental, financial, EEG and fMRI data, which may exhibit complex serial dependence structures. According to this clustering principle, a given MTS is considered to be an outlier if it has been generated from a stochastic process different from those of the majority of the series in the dataset. This definition of outlier has already been considered in some works [1,2], and an interpretation of why this notion of outlyingness naturally arises in practice is given in [3].

Our proposal consists in considering three robust versions of a fuzzy C-means model based on a spectral dissimilarity measure, which compares principal component scores of nonparametric estimates of quantile cross-spectral densities. In essence, our approach intends to take advantage of: (i) the versatility of the fuzzy paradigm, (ii) the high capability of the spectral dissimilarity to discriminate between complex dependence structures, and (iii) the robustness inherent to the variants considered for the fuzzy C-means algorithm.

The behaviour of the constructed robust extensions is analyzed by means of an extensive simulation study involving linear, nonlinear and the so-called BEKK models, one well-known specification of multivariate GARCH processes. The simulated scenarios consist of two well-defined clusters contaminated with one or two outlying series. This structure allows us to evaluate the ability of the methods to correctly locate the nonabnormal MTS and properly handle the anomalous series.

The remainder of this paper is structured as follows. In Section 2, we provide a brief overview of related work, which motivates our approach. Section 3 is devoted to introduce the concept of quantile cross-spectral density (QCD) and construct a dissimilarity measure between MTS (d_{QCD}) based on proper QCD-estimates. Three robust extensions of the fuzzy C-means clustering model considering d_{QCD} introduced by [4] are presented in Section 4. The behaviour of the proposed clustering approaches with respect to different types of outliers is examined via simulations in Sections 5 and 6. The usefulness of the robust fuzzy models is illustrated through two study cases with real data in Section 7, and some concluding remarks are summarized in Section 8.

2. Related work and motivation

A range of procedures have been proposed for clustering of time series during the last two decades, either considering the "hard" or the "soft" paradigm. The former techniques locate each data object in exactly one cluster, whereas the latter strategies provide a more versatile approach, constructing a partition where each object can belong to several groups with specific membership degrees. While a lot of works have addressed the topic of clustering of UTS [1,5–22], the literature on MTS clustering is considerably more scarce [4,23–29]. Comprehensive reviews on the topic can be seen in [30,31].

Robust clustering procedures for time series have also arisen in recent years. In the univariate framework, [32] introduced a robust procedure based on the traditional fuzzy C-medoids algorithm, where the effect of outlying series is neutralized by introducing the so-called noise cluster, expected to contain the outlier elements. This work is extended in [10] by considering a suitable robust metric for time series, the exponential distance measure, and further in [9], where a trimming-based rule is considered for trimming away the most anomalous series. [14] proposed different fuzzy robust clustering approaches for UTS based on GARCH models. [1] developed three robust fuzzy clustering strategies relying on quantile autocovariances. In [33], a robust clustering for stationary series is introduced by considering estimates of the spectral densities as functional data and then applying a robust clustering algorithm.

Some papers have also proposed robust clustering strategies for MTS. [28] introduced three fuzzy C-means clustering approaches for multivariate time trajectories considering the so-called positional dissimilarity, velocity dissimilarity and a

generalization of both. [29] proposed four fuzzy methods for grouping MTS based on the fuzzy C-medoids approach and making use of the exponential transformation for dissimilarity measures. The procedures differentiate from one another by the kind of discrepancy they take into account. Three methods consider instantaneous-based, longitudinal-based, and both types of features, respectively, whereas the fourth algorithm employs the dynamic time warping (DTW) distance for MTS. DTW is also used in [34] to construct a novel trimmed fuzzy approach to cluster multivariate financial time series, which is successfully applied to 40 MTS representing companies in the Italian Stock Exchange. Four techniques to perform robust fuzzy clustering of time series employing B-splines fitted coefficients are proposed in [35] along with feasible algorithms for implementing the corresponding methodologies.

It is worth highlighting that the previous works regarding the multivariate framework are mainly aimed at capturing dissimilarity in shape, that is, they intend to discriminate between different geometric profiles. To the best of our knowledge, no robust clustering method for MTS has been suggested in the literature when outlyingness is characterized by atypical dependence patterns.

Based on previous references, one can deduce that the majority of robust clustering strategies for time series consider a fuzzy approach and rely on at least one of the following three methodological strategies:

1) **Metric approach**. It is based on incorporating in the objective function of the clustering algorithm distances with high robustness against outliers.
2) **Noise approach**. A term representing an artificial cluster is included in the objective function. This cluster is expected to contain the outlying series in the dataset with a high membership.
3) **Trimmed approach**. The clustering technique is applied to the time series remaining after a fixed fraction of the most distant series is removed.

These three strategies are indeed useful to construct robust algorithms. However, we also need a dissimilarity measure capable of separating distinct generating mechanisms since this is our clustering principle. Motivated by these arguments, our proposal combines the mentioned robust approaches with a metric exhibiting a high discriminatory power. Specifically, we propose to use the distance based on estimates of the quantile cross-spectral densities and principal component analysis (PCA). This metric has been used in our previous work [4] to perform fuzzy clustering of MTS through the standard fuzzy C-means and fuzzy C-medoids clustering algorithms. Both procedures showed very good results in grouping MTS simulated from a wide range of generating processes, clearly outperforming other alternative metrics suggested in the literature. However, they are prone to errors in presence of atypical series due to their lack of robustness. To address this issue, we consider the metric, noise and trimmed versions of the standard fuzzy C-means clustering model (see [36], [37] and [38], respectively) in combination with the spectral metric given in [4]. Thus, the presented algorithms can be seen as robust extensions of the fuzzy C-means clustering model introduced by [4], so-called QCD-FCMn. We use the fuzzy C-means algorithm because it produced slightly better results than the C-medoids algorithm in [4]. Each of the introduced methods achieves robustness in different manner and all of them take advantage of the high capability of the QCD-based metric to discriminate between independent realizations from a broad range of stationary processes.

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

Let $\mathcal{S} = \left\{ \boldsymbol{X}_t^{(1)}, \ldots, \boldsymbol{X}_t^{(s)} \right\}$ be a set of s multivariate time series, 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. We wish to perform clustering on $\mathcal{S}$ in such a way that the series generated from the same stochastic process are grouped together. We propose to use a partitional algorithm starting from a pairwise dissimilarity matrix based on comparing estimated quantile cross-spectral densities. In this section, the quantile cross-spectral density notion is presented and then used to define a distance between MTS.

3.1. The quantile cross-spectral density

Let $\{\boldsymbol{X}_t,\, t \in \mathbb{Z}\} = \{(X_{t,1}, \ldots, X_{t,d}),\, t \in \mathbb{Z}\}$ 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

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

for $1 \leq j_1, j_2 \leq d$. Taking $j_1 = j_2 = j$, the function $\gamma_{j,j}(l,\tau,\tau')$, 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

dependence structure of $(X_{t,j}, X_{t+l,j})$, enabling us to capture serial features that standard autocovariances cannot detect. Note that $\gamma_{j_1,j_2}(l,\tau,\tau')$ always exists since no assumptions about moments are required. Furthermore, QAF also takes advantage of the local distributional properties inherent to the quantile methods, including robustness against heavy tails, dependence in the extremes and changes in the conditional shapes (skewness, kurtosis). Motivated by these nice properties, a dissimilarity between UTS based on comparing estimated quantile autocovariances over a common range of quantiles was proposed by [5] to perform UTS clustering with very satisfactory results.

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

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

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

$$\boldsymbol{\Gamma}_{\boldsymbol{X}_t}(\mathcal{L},\mathcal{T}) = \left\{\boldsymbol{\Gamma}(l,\tau,\tau'),\ l\in\mathcal{L},\ \tau,\tau'\in\mathcal{T}\right\}. \tag{2}$$

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

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

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.

For fixed quantile levels (τ,τ'), the quantile cross-spectral density is the cross-spectral density of the bivariate process $(I\{X_{t,j_1}\le q_{j_1}(\tau)\}, I\{X_{t,j_2}\le q_{j_2}(\tau')\})$. 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 (2), the quantile cross-spectral density can be evaluated for every couple of components on a range of frequencies Ω and of quantile levels $\mathcal{T}$ in order to obtain a complete representation of the process, i.e., consider the set of matrices

$$\mathfrak{f}_{\boldsymbol{X}_t}(\Omega,\mathcal{T}) = \left\{\mathfrak{f}(\omega,\tau,\tau'),\ \omega\in\Omega,\ \tau,\tau'\in\mathcal{T}\right\}, \tag{3}$$

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

$$\mathfrak{f}(\omega,\tau,\tau') = \left(\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau')\right)_{1\le j_1,j_2\le d}. \tag{4}$$

Representing $\boldsymbol{X}_t$ through $\mathfrak{f}_{\boldsymbol{X}_t}$, complete information on the general dependence structure of the process is available. Comprehensive discussions about the nice properties of the quantile cross-spectral density are given in [39], [40] and [41], including invariance to monotone transformations, robustness and capability to detect nonlinear dependence. It is also worth enhancing that the quantile cross-spectral density provides a full description of all copulas of pairs of components in $\boldsymbol{X}_t$, since the difference between the copula of an arbitrary couple (X_{t,j_1}, X_{t+l,j_2}) evaluated in (τ,τ') and the independence copula at (τ,τ') can be written as

$$\mathbb{P}\left(X_{t,j_1}\le q_{j_1}(\tau), X_{t+l,j_2}\le q_{j_2}(\tau')\right) - \tau\tau' = \int_{-\pi}^{\pi}\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau')e^{il\omega}\,d\omega. \tag{5}$$

According to the prior arguments, a dissimilarity measure between realizations of two multivariate processes, $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$, could be established by comparing their representations in terms of the quantile cross-spectral density matrices, $\mathfrak{f}_{\boldsymbol{X}_t}$ and $\mathfrak{f}_{\boldsymbol{Y}_t}$, respectively. To this aim, the quantile cross-spectral densities must be previously estimated.

Let $\{\boldsymbol{X}_1,\ldots,\boldsymbol{X}_T\}$ be a realization from the process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ so that $\boldsymbol{X}_t=(X_{t,1},\ldots,X_{t,d})$, $t=1,\ldots,T$. For arbitrary $j_1,j_2\in\{1,\ldots,d\}$ and $(\tau,\tau')\in[0,1]^2$, [41] propose to estimate $\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau')$ considering a smooth estimator of the cross-periodograms based on the indicator functions $I\{\hat{F}_{T,j}(X_{t,j})\}$, where $\hat{F}_{T,j}(x)=T^{-1}\sum_{t=1}^{T}I\{X_{t,j}\le x\}$ denotes the empirical distribution function of $X_{t,j}$. This approach extends to the multivariate case the estimator proposed by [42] in the univariate setting. More specifically, the called *rank-based copula cross periodogram* (CCR-periodogram) is defined by

$$I_{T,R}^{j_1,j_2}(\omega,\tau,\tau') = \frac{1}{2\pi T}d_{T,R}^{j_1}(\omega,\tau)d_{T,R}^{j_2}(-\omega,\tau'), \tag{6}$$

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

The asymptotic properties of the CCR-periodogram are established in Proposition S4.1 of [41]. Likewise the standard cross-periodogram, the CCR-periodogram is not a consistent estimator of $\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau')$ [41]. To achieve consistency, the CCR-periodogram ordinates (evaluated on the Fourier frequencies) are convolved with weighting functions $W_T(\cdot)$. The *smoothed CCR-periodogram* takes the form

$$\hat{G}^{j_1,j_2}_{T,R}(\omega,\tau,\tau')=\frac{2\pi}{T}\sum_{s=1}^{T-1} W_T\left(\omega-\frac{2\pi s}{T}\right) I^{j_1,j_2}_{T,R}\left(\frac{2\pi s}{T},\tau,\tau'\right), \tag{7}$$

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 $Th_T\to\infty$ as $T\to\infty$, and W is a real-valued, even weight function with support $[-\pi,\pi]$. Consistency and asymptotic performance of the smoothed CCR-periodogram $\hat{G}^{j_1,j_2}_{T,R}(\omega,\tau,\tau')$ are established in Theorem S4.1 of [41].

The set of complex-valued matrices $\mathfrak{f}_{\boldsymbol{X}_t}(\Omega,\mathcal{T})$ in (3) characterizing the underlying process can be estimated by

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

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

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

3.2. A spectral dissimilarity measure between MTS

A simple dissimilarity criterion between two d-variate time series $\boldsymbol{X}^{(1)}_t$ and $\boldsymbol{X}^{(2)}_t$ can be obtained by comparing their estimated sets of complex-valued matrices, $\hat{\mathfrak{f}}_{\boldsymbol{X}^{(1)}_t}(\Omega,\mathcal{T})$ and $\hat{\mathfrak{f}}_{\boldsymbol{X}^{(2)}_t}(\Omega,\mathcal{T})$, respectively, evaluated on a common range of frequencies and quantile levels. Specifically, each time series $\boldsymbol{X}^{(u)}_t$, $u=1,2$, is characterized by means of a set of d^2 vectors $\{\boldsymbol{\Psi}^{(u)}_{j_1,j_2}, 1\leq j_1,j_2\leq d\}$ constructed as follows. For a given set of K frequencies $\Omega=\{\omega_1,\ldots,\omega_K\}$, and r quantile levels $\mathcal{T}=\{\tau_1,\ldots,\tau_r\}$, each vector $\boldsymbol{\Psi}^{(u)}_{j_1,j_2}$ is given by

$$\boldsymbol{\Psi}^{(u)}_{j_1,j_2}=(\boldsymbol{\Psi}^{(u)}_{1,j_1,j_2},\ldots,\boldsymbol{\Psi}^{(u)}_{K,j_1,j_2}), \tag{8}$$

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

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

All the d^2 vectors $\boldsymbol{\Psi}^{(u)}_{j_1,j_2}$, $1\leq j_1,j_2\leq d$, are then concatenated in a vector $\boldsymbol{\Psi}^{(u)}$ in the same way as vectors $\boldsymbol{\Psi}^{(u)}_{k,j_1,j_2}$ constitute $\boldsymbol{\Psi}^{(u)}_{j_1,j_2}$ in (8). Based on previous considerations, the dissimilarity between $\boldsymbol{X}^{(1)}_t$ and $\boldsymbol{X}^{(2)}_t$ is defined by means of the Euclidean distance between $\boldsymbol{\Psi}^{(1)}$ and $\boldsymbol{\Psi}^{(2)}$, that is

$$\begin{aligned} d_{QCD}(\boldsymbol{X}^{(1)}_t,\boldsymbol{X}^{(2)}_t)=&\Big[\left\|\Re_v(\boldsymbol{\Psi}^{(1)})-\Re_v(\boldsymbol{\Psi}^{(2)})\right\|^2 \\ &+\left\|\Im_v(\boldsymbol{\Psi}^{(1)})-\Im_v(\boldsymbol{\Psi}^{(2)})\right\|^2\Big]^{1/2}, \end{aligned} \tag{9}$$

where $\Re_v$ and $\Im_v$ denote the element-wise real and imaginary part operators, respectively.

If $\boldsymbol{\Psi}^{(1)},\ldots,\boldsymbol{\Psi}^{(n)}$ are obtained from the series in S, we could perform fuzzy clustering using an algorithm as fuzzy C-means or fuzzy C-medoids based on the distance d_{QCD}. This distance has been successfully applied to perform crisp MTS clustering [25], and the QCD-based features used to develop classification [43] and outlier detection [3] procedures. Notice that calculation of vectors $\boldsymbol{\Psi}^{(1)},\ldots,\boldsymbol{\Psi}^{(n)}$ entails a low computational cost. In [25], the time consumption of several dissimilarity measures was evaluated in the context of partitional clustering, and the results indicated that d_{QCD} is computationally efficient and outperforms several metrics employing alternative representations for the series.

If we focus our attention on the clustering task, the vectors $\left(\Re_v(\boldsymbol{\Psi}^{(1)}),\Im_v(\boldsymbol{\Psi}^{(1)})\right),\ldots,\left(\Re_v(\boldsymbol{\Psi}^{(n)}),\Im_v(\boldsymbol{\Psi}^{(n)})\right)$ can be transformed by PCA to obtain the score vectors $\boldsymbol{\Psi}^{(1)}_{PCA},\ldots,\boldsymbol{\Psi}^{(n)}_{PCA}$, and then apply the clustering algorithm to this set of score vectors, i.e., considering the distance between an arbitrary pair of MTS, $\boldsymbol{X}^{(i)}_t$ and $\boldsymbol{X}^{(j)}_t$, defined by $d_{QCD_{PCA}}(\boldsymbol{X}^{(i)}_t,\boldsymbol{X}^{(j)}_t)=\left\|\boldsymbol{\Psi}^{(i)}_{PCA}-\boldsymbol{\Psi}^{(j)}_{PCA}\right\|^2$.

In [4], we showed that, by proceeding this way, the clustering algorithm generally increases its effectiveness because a lot of noise gets removed. When there are well-defined clusters, the use of PCA provides a partition with less overlap between groups, thus giving more informative solutions. This behaviour is not surprising, since the QCD-based features are

highly correlated due to the definition of the smoothed CCR-periodogram in (7). Hence, by considering the raw features, some variables could get a higher weight than others in the distance computation, thus introducing bias in the clustering algorithm. The PCA transformation overcomes this drawback by removing the underlying correlation between features and making the grouping process easier.

Indeed, only a subset with the first p principal components is considered, although the optimal selection of p is not addressed here. Actually, the choice of the optimal subset of principal components can be seen as a classical feature-selection problem and multiple procedures are available in a clustering context [44,45]. In simulation experiments, a reasonable criterion consists in considering internal clustering validity indexes. Anyway, we simply proceeded in line with the empirical rule provided in [4], which is based on retaining the first $\lceil 0.12p^* \rceil$ principal components, being p^* the total number of principal components. This guideline led generally to the best results in the simulation studies of Sections 5 and 6.

Based on previous remarks, from now on, the distance d_{QCD} and the QCD-based quantities are going to refer to the PCA-transformed features rather than the original features, although we maintain the former notation for the sake of simplicity, that is, the subscript PCA is removed.

4. Robust fuzzy clustering based on the QCD-based distance

In this section, we extend the model QCD-FCMn proposed by [4] (see Section 4.1) by considering the metric, noise and trimmed approaches (see [36], [37] and [38], respectively). The direct application of these robust strategies to the QCD time series representation leads to different models for fuzzy clustering of MTS. To make the work more self-contained and easier to read, the construction of the three models is described in Sections 4.2, 4.3 and 4.4, respectively. Also, for the reader's convenience, outlines of the respective resulting algorithms are provided in Supplementary material (see Algorithms 1, 2, 3 and 4).

4.1. QCD-based fuzzy C-means clustering model

As in previous sections, consider a set S of n realizations of multivariate time series $\{\boldsymbol{X}_t^{(1)}, \ldots, \boldsymbol{X}_t^{(n)}\}$ 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 3.2. 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, $\boldsymbol{U} = (u_{ic})$, $i = 1, \ldots, n$, $c = 1, \ldots, C$, defining the solution of the minimization problem

$$\begin{cases} \min\limits_{\overline{\boldsymbol{\Psi}}, \boldsymbol{U}} \sum\limits_{i=1}^{n} \sum\limits_{c=1}^{C} u_{ic}^m \left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c)} \right\|^2, \\ \text{subject to } \sum\limits_{c=1}^{C} u_{ic} = 1 \text{ and } u_{ic} \geq 0, \end{cases} \tag{10}$$

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 for the c-th cluster, and $m > 1$ is a parameter controlling the fuzziness of the partition, usually referred to as fuzziness parameter. Constraints on u_{ic} are standard requirements in fuzzy clustering.

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 d_{QCD} of identifying different dependence structures. Unlike crisp clustering, the nonstochastic uncertainty inherent to the assignment of series to clusters is here incorporated to the procedure by means of the membership degrees.

The optimization problem in (10) is solved by the Lagragian multipliers method in a two-step iterative process. In our setting, the updates of the classical solutions take the form

$$u_{ic} = \left[\sum_{c'=1}^{C} \left(\frac{\left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c)} \right\|^2}{\left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c')} \right\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}, \tag{11}$$

for $i = 1, \ldots, n$ and $c = 1, \ldots, C$, and

$$\overline{\boldsymbol{\Psi}}^{(c)} = \frac{\sum_{i=1}^{n} u_{ic}^m \boldsymbol{\Psi}^{(i)}}{\sum_{i=1}^{n} u_{ic}^m}, \quad c = 1, \ldots, C. \tag{12}$$

4.2. QCD-based exponential fuzzy C-means clustering model

More robust metrics than the Euclidean distance can be employed in the conventional objective function of the fuzzy algorithm. In particular, [36] consider an exponential-type distance, whose direct application to the QCD time series representation leads to the QCD-based exponential fuzzy C-means clustering model (QCD-FCMn-E). To remind readers this approach, the objective function takes the form

$$\begin{cases} \min\limits_{\overline{\boldsymbol{\Psi}},\boldsymbol{U}} \sum\limits_{i=1}^{n}\sum\limits_{c=1}^{C} u_{ic}^{m}\left[1-\exp\left\{-\beta\left\|\boldsymbol{\Psi}^{(i)}-\overline{\boldsymbol{\Psi}}^{(c)}\right\|^{2}\right\}\right] \\ \text{subject to } \sum\limits_{c=1}^{C} u_{ic}=1 \text{ and } u_{ic}\geq 0, \end{cases} \tag{13}$$

where $\beta>0$ is a constant.

The local optimal solution for (13) is given by (see [36])

$$u_{ic}=\left[\sum_{c'=1}^{C}\left(\frac{1-\exp\left\{-\beta\left\|\boldsymbol{\Psi}^{(i)}-\overline{\boldsymbol{\Psi}}^{(c)}\right\|^{2}\right\}}{1-\exp\left\{-\beta\left\|\boldsymbol{\Psi}^{(i)}-\overline{\boldsymbol{\Psi}}^{(c')}\right\|^{2}\right\}}\right)^{\frac{1}{m-1}}\right]^{-1}. \tag{14}$$

The fuzzy clustering based on the exponential distance is more robust than the one based on the Euclidean distance [36]. By definition, exponential distance gives different weights in accordance to whether an element is anomalous or not, namely small weights to outliers and large weights to compact points in the dataset ([10,36]).

An appropriate choice of the hyperparameter β is totally crucial for a good performance of QCD-FCMn-E (see Section 4 in [36] for further details). When β tends to zero, the QCD-FCMn-E algorithm tends to the QCD-FCMn algorithm, which gives equal weight to all elements in the dataset regardless of their outlying nature. The value of β is usually determined as the inverse of the variability in the data (the more variability in the data, the less the value of β). This quantity has an impact on the membership degrees (14) in terms of robustness to outliers. The following choice for β has proven to be suitable for different types of datasets:

$$\beta=\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Psi}^{(i)}-\overline{\boldsymbol{\Psi}}^{(k)}\right\|^{2}\right)^{-1}, \tag{15}$$

where $\overline{\boldsymbol{\Psi}}^{(k)}$ corresponds to the index k satisfying $k=\underset{1\leq i'\leq n}{\arg\min}\sum\limits_{i''=1}^{n}\left\|\boldsymbol{\Psi}^{(i'')}-\boldsymbol{\Psi}^{(i')}\right\|^{2}$ (see [10] for more details).

In essence, QCD-FCMn-E adjusts the impact of anomalous series by smoothing their effect through suitable weights. This way, the membership degrees of the outliers are similarly distributed across the clusters but the true clustering structure is not seriously perverted because of their presence.

4.3. QCD-based fuzzy C-means clustering with noise cluster

The QCD-based fuzzy C-means clustering model with noise cluster (QCD-FCMn-NC) arises from applying to the PCA-transformed features the algorithm proposed by [37]. The corresponding minimization problem is given by

$$\begin{cases} \min\limits_{\tilde{\boldsymbol{\Psi}},\boldsymbol{U}} \sum\limits_{i=1}^{n}\sum\limits_{c=1}^{C-1} u_{ic}^{m}\left\|\boldsymbol{\Psi}^{(i)}-\overline{\boldsymbol{\Psi}}^{(c)}\right\|^{2}+\sum\limits_{i=1}^{n}\delta^{2}\left(1-\sum\limits_{c=1}^{C-1}u_{ic}\right)^{m} \\ \text{subject to } \sum\limits_{c=1}^{C} u_{ic}=1 \text{ and } u_{ic}\geq 0, \end{cases} \tag{16}$$

where $\delta>0$ is the so-called noise distance.

Note that the approach proposed by [37] involves C clusters, but only $(C-1)$ are "valid" clusters. The additional cluster, the noise cluster, is artificially created for outlier identification purposes. The aim is to locate the outliers and place them in the noise cluster, which is represented by a fictitious prototype time series with a constant distance δ from every MTS (the noise distance). If the distance from a given MTS to a centroid series is smaller than δ, then the MTS is assigned to the real cluster. Otherwise, it is located into the noise cluster.

The value $u_{i*} = 1 - \sum_{c=1}^{C-1} u_{ic}$ in (16) expresses the membership degree of the i-th time series to the noise cluster, $i = 1, \ldots, n$, and it is expected to be high for series showing an outlying nature. This way, the usual constraint on the membership degrees for the real clusters is here relaxed to $\sum_{c=1}^{C-1} u_{ic} < 1$, which allows outlying time series to have small membership values in real clusters.

Minimization of (16) with regards to the membership degrees (see [37]) yields

$$u_{ic} = \left[\sum_{c'=1}^{C} \left(\frac{\left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c)} \right\|^2}{\left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c')} \right\|^2} \right)^{\frac{1}{m-1}} + \left(\frac{\left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c)} \right\|^2}{\delta^2} \right)^{\frac{1}{m-1}} \right]^{-1}. \tag{17}$$

The hyperparameter δ determines the boundary of the noise cluster so that a proper choice of δ is crucial to make the most of the QCD-FCMn-NC approach. When δ is too large, QCD-FCMn-NC degenerates to the nonrobust version of the model and outliers are forced to belong to real clusters. On the other hand, if δ is too small, many objects can be considered as noise and misclassified into the noise cluster [46]. Although there are some heuristic approaches to estimate δ, its optimal choice under general conditions is still an open issue. [47] suggest to adequate the computation of δ to the concept of "scale" in robust statistics [48]. However, as stated by [46], "unfortunately, the proper estimation of [the] scale is not a trivial task [49] and requires some knowledge of the data, which cannot always be supposed in real clustering applications". In the original noise clustering algorithm [37], the value of δ was set to

$$\delta^2 = \frac{\lambda}{n(C-1)} \sum_{i=1}^{n} \sum_{c=1}^{C-1} \left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c)} \right\|^2, \tag{18}$$

where λ is a scale multiplier to be chosen in accordance with the nature of the data.

To select the most suitable value of λ, [46] propose to execute the fuzzy clustering model with noise cluster with decreasing values of λ and study the distribution of the percentage of objects assigned to the noise cluster. This distribution has a sudden change of slope (elbow) when the value of the noise distance is so small that elements belonging to real clusters are grouped into the noise cluster. According to the elbow it is possible to figure out the optimal noise distance. The authors approximate the distribution of percentages with a Pareto distribution.

4.4. QCD-based trimmed fuzzy C-means clustering

The Least Trimmed Squares approach proposed by [38] leads to the QCD-based trimmed fuzzy C-means clustering model (QCD-FCMn-T). Here, robustness is attained by removing a certain proportion of the series and requires the specification of the fraction α of the data to be trimmed. Then, all nontrimmed series are classified using the QCD-FCMn model. For a fixed trimming ratio α, the minimization problem is given by

$$\begin{cases} \min_{\boldsymbol{Y}, \boldsymbol{U}} \sum_{i=1}^{H(\alpha)} \sum_{c=1}^{C} u_{ic}^m \left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c)} \right\|^2 \\ \text{subject to } \sum_{c=1}^{C} u_{ic} = 1 \text{ and } u_{ic} \geq 0, \end{cases} \tag{19}$$

where $\boldsymbol{Y}$ ranges on all the subsets of $\boldsymbol{\Psi}$ of size $H(\alpha) = \lfloor n(1-\alpha) \rfloor$. If α is set to 0, then none of the series is removed and the nonrobust QCD-FCMn model is obtained. According to [38], the local optimal solution for the estimation of u_{ic} is provided by (11), with i ranging in the subset of the nontrimmed series and $c = 1, \ldots, C$. Then, replacing the optimal expression for u_{ic} in (19), one has

$$\sum_{i=1}^{H(\alpha)} \left[\sum_{c=1}^{C} \left(\left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c)} \right\|^2 \right)^{1/(1-m)} \right]^{1-m} = \sum_{i=1}^{H(\alpha)} h_i,$$

with $h_i = \left[\sum_{c=1}^{C} \left(\left\| \boldsymbol{\Psi}^{(i)} - \overline{\boldsymbol{\Psi}}^{(c)} \right\|^2 \right)^{1/(1-m)} \right]^{1-m}$.

The objective function in (19) takes the form

$$\sum_{i=1}^{H(\alpha)} h_{i:n}, \tag{20}$$

where $h_{i:n}$ represents the i-th item when h_i, $i = 1, \ldots, n$, are arranged in ascending order. Specification (20) is very advantageous to implement the procedure.

The value of $H(\alpha) < n$ is chosen depending on how many series we would like to eliminate in the clustering process. For instance, when $H(\alpha) = \lfloor n/2 \rfloor$, 50% of the time series are not involved in the clustering process, and the objective function is minimized when the centroids are computed in such a way that the sum of harmonic mean (squared) Euclidean distance of 50% of the MTS to the corresponding centroids is as small as possible (further details in [11] and [9]).

5. Evaluating robustness with respect to global outliers

In this section, we analyze the results from an extensive simulation study designed to assess the effectiveness of the proposed robust methods by considering the so-called global outliers, i.e., elements substantially deviating from the rest in a general manner. Since our goal is to discriminate between generating processes, we focus on anomalies characterized by atypical dynamic behaviours. Specifically, the considered scenarios are formed by two regular groups of MTS plus one or two outlying series generated from processes different from the ones defining the regular clusters.

First, some alternative metrics considered for comparison purposes are described. Then, the simulation and assessment mechanisms are detailed and the main results are discussed.

5.1. Alternative metrics

To shed light on the clustering accuracy of QCD-FCMn-E, QCD-FCMn-NC and QCD-FCMn-T, a comparison with other robust clustering models based on alternative metrics was carried out. It is worthy to remark the lack in the literature of robust procedures to perform MTS clustering. However, the QCD-based fuzzy approach developed in Section 4.1 can be easily adjusted to involve other kind of extracted features. Thus, an alternative fuzzy C-means model can be formalized as the minimization problem in (10) by replacing $\boldsymbol{\Psi}^{(i)}$, $\overline{\boldsymbol{\Psi}}$ and $\overline{\boldsymbol{\Psi}}^{(c)}$ by $\boldsymbol{\varphi}^{(i)}$, $\overline{\boldsymbol{\varphi}}$ and $\overline{\boldsymbol{\varphi}}^{(c)}$, respectively, with $\boldsymbol{\varphi}^{(i)}$ denoting the estimated vector for the i-th series of selected features, and the remaining terms defined analogously. The iterative solutions are obtained through (11) and (12) by considering the new features. In the same way, exponential, noise, and trimmed versions are constructed starting from (13), (16) and (19), respectively.

For comparison purposes, we introduced the robust fuzzy models based on the MTS features described below.

- *Wavelet-based features*. [7] introduced a squared Euclidean distance between wavelet features of two MTS considering estimates of wavelet variances and 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. Therefore, in this case the vector $\boldsymbol{\varphi}^{(i)}$ contains estimates of wavelet variances and wavelet correlations of a given MTS. The corresponding robust techniques are referred to as wavelet-based exponential fuzzy C-means clustering model (W-FCMn-E), wavelet-based fuzzy C-means clustering model with noise cluster (W-FCMn-NC) and wavelet-based trimmed fuzzy C-means clustering (W-FCMn-T). After performing some preliminary analyses, we concluded that the wavelet filter of length 4 of the Daubechies family, DB4, along with the maximum allowable number of scales, were the choices producing the best average results in the simulation scenarios considered in Sections 5 and 6.

- *Correlation-based features*. In the univariate setting, [6] proposed a fuzzy model based on estimated autocorrelations of a UTS for a range of lags, which was extended to the multivariate context in [4]. Given an MTS, estimates of both autocorrelations for each component (UTS) and cross-correlations between each pair of components are calculated up to a fixed lag l. This set of features defines the vector $\boldsymbol{\varphi}^{(i)}$ characterizing the i-th MTS and used to perform clustering. We call the corresponding robust approaches correlation-based exponential fuzzy C-means clustering model (C-FCMn-E), correlation-based fuzzy C-means clustering model with noise cluster (C-FCMn-NC) and correlation-based trimmed fuzzy C-means clustering (C-FCMn-T). The hyperparameter l was set to $l = 1$ throughout our simulation study because most of the considered generating processes contain only one significant lag.

To get insight into the robustness of the proposed approaches, a wide range of generating processes were considered to build our simulation scenarios, including linear, nonlinear and conditionally heteroskedastic models.

5.2. Experimental design and exploratory analyses

Each of the considered setups consisted of two well-established clusters (base scenario), with five realizations each from the same generating process, successively contaminated by adding one and two outlier series. The specific scenarios and the generation schemes are given below.

Robust fuzzy clustering of linear models

Base Scenario 1: A VAR(1) process given by

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

and a VMA(1) process given by

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

Scenario 1.1: The ten series simulated from the Base Scenario 1 plus one outlier generated from a VARMA process with matrices of coefficients given in the Base Scenario 1.

Scenario 1.2: The eleven series from Scenario 1.1 plus one additional outlier time series simulated from the NAR process introduced in Scenario 2.1 below.

Robust fuzzy clustering of nonlinear models

Base Scenario 2: A EXPAR (exponential autoregressive) process given by

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0.3 - 10\exp(-X_{t-1,1}^2 - X_{t-1,2}^2)X_{t-1,2} \\ 0.3 - 10\exp(-X_{t-1,1}^2 - X_{t-1,2}^2)X_{t-1,1} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix},$$

and a BL (bilinear) process given by

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

Scenario 2.1: The ten series simulated from the Base Scenario 2 plus one outlier time series simulated from the NAR (nonlinear autoregressive) process

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

Scenario 2.2: The series from Scenario 2.1 plus an additional outlier time series simulated from the VAR process

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

Robust fuzzy clustering of conditional heteroskedastic models

Base Scenario 3: Consider $\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \Sigma_t^{1/2} \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix}$.

The data generating processes consist of two bivariate BEKK (Baba-Engle-Kraft-Kroner) models given by

$$\Sigma_t = C^\intercal C + A^\intercal \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} (X_{t-1,1}, X_{t-1,2}) A + G^\intercal \Sigma_{t-1} G,$$

where C is a lower triangular matrix and A and G are 2×2 matrices. In the first generating process, $C = \begin{pmatrix} 0.1 & 0 \\ 0.1 & 0.1 \end{pmatrix}$, $A = \begin{pmatrix} 0.2 & 1.2 \\ 0.4 & 0.5 \end{pmatrix}$ and $G = \begin{pmatrix} 0.2 & -0.1 \\ -0.1 & -0.1 \end{pmatrix}$, whereas in the second process, $C = \begin{pmatrix} 0.1 & 0 \\ 0.1 & 0.1 \end{pmatrix}$, $A = \begin{pmatrix} 0.5 & 0.4 \\ 0.7 & -0.2 \end{pmatrix}$ and $G = \begin{pmatrix} -0.5 & -0.4 \\ -0.1 & -0.4 \end{pmatrix}$.

Scenario 3.1: The ten series simulated from the Base Scenario 3 plus one outlier time series simulated from a bivariate white noise process (WN).

Scenario 3.2: The eleven series from Scenario 3.1 plus one outlier time series simulated from the BL process in Base Scenario 2.

In all cases, the error process $(\epsilon_{t,1}, \epsilon_{t,2})^\intercal$ consisted of iid realizations from a standard bivariate Gaussian distribution.

VARMA models are broadly used in many fields but determining the model order is a complex task. 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. Scenarios 2.1 and 2.2 consist of a multivariate extension of the univariate NAR process proposed in [50] and EXPAR and BL processes proposed in [1]. Nonlinear UTS arise in several application fields [51–53]. Although nonlinear MTS have received much less attention, there exist some fields as neurophysiology [54] and economy [55] 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. Scenarios 3.1 and 3.2 are motivated by the BEKK models introduced in [56]. BEKK models are a formulation of multivariate GARCH models frequently used to

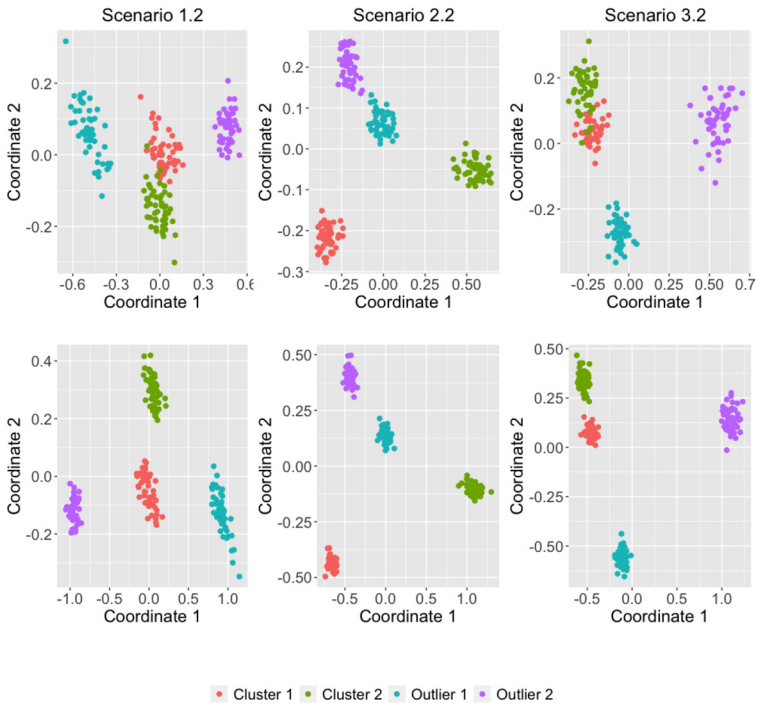

Fig. 1. Two-dimensional scaling based on QCD distances between simulated series in Scenarios 1.2, 2.2 and 3.2. The series length is $T = 500$ (top panel) and $T = 2000$ (bottom panel). (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

model MTS of financial variables as stock returns, commodity prices, inflation and exchange rates, among others [57–59]. Financial time series are known to exhibit empirical statistical regularities, so-called "stylized-facts". The most common stylized facts include: heavy tails and a peak centre compared to the normal distribution, volatility clustering (periods of low volatility mingle with periods of high volatility), leverage effects (the quantities are negatively correlated with volatility), and autocorrelation at much longer horizons than expected. In our previous works [3,4,25,43], the distance d_{QCD} was tested either for clustering, classification and outlier detection purposes by considering a different extension of multivariate GARCH models, the so-called dynamic conditional correlation models (DCC) [60], achieving great results. BEKK and DCC models are the most utilized formulations of multivariate GARCH models. It is worth remarking that each formulation is useful in its own right, and the choice of the most suitable one often depends on the empirical application [61]. In this work we have decided to evaluate the distance d_{QCD} with BEKK models. Note that in Scenarios 1.2, 2.2 and 3.2, the second anomalous series is generated from a kind of process different from the ones concerning the remaining series. For instance, three types of nonlinear processes along with a linear one are involved in Scenario 2.2. Heterogeneity in terms of generating processes was introduced in these scenarios with the aim of making more demanding the clustering task.

As a preliminary step, a metric two-dimensional scaling (2DS) based on QCD distances was carried out to examine if d_{QCD} is expected to perform robust fuzzy clustering correctly in the considered scenarios. 2DS represents the pairwise QCD distances in terms of Euclidean distances into a 2-dimensional space preserving the original distances as well as possible (by minimizing a loss function). For Scenarios 1.2, 2.2 and 3.2, we simulated 50 MTS of equal length T from each involved process (including both regular and outlying models). The resulting scatter plots for $T = 500$ and $T = 2000$ are depicted in Fig. 1, where points coming from the same process have the same colour.

The quality of the embedding can be measured by the R^2 value, reporting the proportion of variance accounted for the 2DS procedure. We obtained the values 0.767 (Scenario 1.2), 0.751 (Scenario 2.2) and 0.746 (Scenario 3.2), when $T = 500$, and 0.8905 (Scenario 1.2), 0.900 (Scenario 2.2) and 0.898 (Scenario 3.2), when $T = 2000$. Notice that values above 0.6 are considered to provide an acceptable scaling procedure, whereas values above 0.8 mean a very good fit [62]. In the three scenarios, we have very similar values and all of them above 0.6 for $T = 500$ and above 0.8 for $T = 2000$, thus concluding that the plots in Fig. 1 provide an accurate picture on the behaviour of d_{QCD} between the different underlying processes.

Indeed, Fig. 1 shows different configurations according to the scenario and the series length. For Scenarios 1.2 and 3.2 with $T = 500$, the 100 series generated from the base scenario exhibit a slight overlapping, whereas the groups of outlying series are well-separated. In the opposite way, for Scenario 2.2 with $T = 500$, the two regular clusters are far away from one another and from the groups of anomalous series (which are very close to each other). When increasing the series length ($T = 2000$), the same configuration is observed, but there are no longer overlapping clusters in any scenario. This is expected since the QCD-based features are more accurately estimated and hence groups become more distant. In summary, Fig. 1 uncovers that the robust approaches based on d_{QCD} should be capable of discriminating between the regular clusters

and the atypical series in Scenarios 1.2, 2.2 and 3.2 provided that the series length is large enough. Among all of them, Scenario 3.2 seems the more challenging scenario for a fixed T.

Every proposed clustering procedure, QCD-FCMn-E, QCD-FCMn-NC and QCD-FCMn-T, and the robust versions of the alternative metrics were applied to cluster the series from Scenarios 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2. In addition, the nonrobust QCD-FCMn model was also run to get a clear idea about to what extent the robust techniques are successful. Accuracy is measured by the proportion of times in which the series generated from the same process belong to the same cluster with a high membership degree. Robustness is examined by studying to what extent the outlying series affect the resulting fuzzy partition. Concerning QCD-FCMn-T and QCD-FCMn-NC, we regard the proportion of times that the anomalous series are trimmed or located in the noise cluster, respectively.

Since the three scenarios contain two base generating processes, the number of clusters was set to $C = 2$. As we are handling fuzzy partitions, it is necessary to specify a cutoff value in order to assign an MTS to a given cluster. For the nonanomalous series, we decided to assign the i-th element to the c-th cluster if $u_{ic} > 0.7$. As for the outlying series, the assignment criterion was different among the three considered robust algorithms. By using the exponential cluster models, we assumed that an atypical series was correctly handled when $\max\{u_{i1}, u_{i2}\} < 0.7$, that is, when the outlying series exhibits similar membership degrees in both clusters. Regarding the noise cluster methods, an anomalous series was correctly identified if $u_{ic_{NC}} > 0.5$, with c_{NC} indicating the index of the noise cluster. Note that, in this case, the minimization problem in (16) involves three underlying clusters, so the value 0.5 is a stringent threshold. Finally, concerning the trimmed approaches, we examined if the anomalous series were correctly trimmed. It is worthy mentioning that these criteria and thresholds are in line with the suggestions recommended in the literature [6,10,12,32,63,64] and have already been considered in some works [1]. A given method is assumed to provide a correct clustering solution if it accurately handles both sets of series anomalous and nonanomalous.

Note that the accuracy of the algorithms is not evaluated by means of external clustering quality indexes, i.e., indexes comparing the experimental partition with the true clustering structure (usually referred to as ground truth). There are two main reasons for it. First, an extensive simulation study assessing the performance of QCD-FCMn through the fuzzy extension of the Adjusted Rand Index (FARI) was carried out in [4], and the results indicated that QCD-FCMn is substantially superior to wavelet-based and correlation-based procedures. Therefore, it is expected that the robust methods QCD-FCMn-E, QCD-FCMn-NC and QCD-FCMn-T inherit the excellent behaviour of QCD-FCMn in this regard. Second, the presence of anomalous series makes it hard to define the ground truth.

The QCD-based features were obtained by using three quantile levels, namely $\mathcal{T} = \{0.1, 0.5, 0.9\}$. This choice has proven to be enough to reach suitable results in an extensive range of data mining tasks involving QCD and similar quantities, as hard and soft clustering [4,5,17,25], classification [43] and outlier detection [3], among others. The number p of retained principal components, i.e., the length of the vectors $\mathbf{\Psi}^{(i)}$ representing the QCD-based features, was calculated by applying the empirical rule presented in Section 3.2, resulting $p = 2$ in all cases.

Since the fuzziness parameter m plays a crucial role, a suitable choice of m is a key issue broadly treated in the literature. For instance, [65] showed that values of m between 1.5 and 2.5 are typically a good choice for the fuzzy-C means algorithm, which is also confirmed by [66] and [67]. However, there seems to be no consensus about the optimal value for m (see discussion in Section 3.1.6 of [12]). Treating with time series, the majority of works consider values of m between 1.3 and 2.6 when performing simulation studies [1,6,7,12,17]. Based on the previous considerations, we decided to use $m = 1.8$, 2, 2.2 and 2.5.

The three robust procedures involve hyperparameters, namely β, δ and α for exponential, noise and trimmed approaches, respectively. Our numerical experiments have revealed the following: (i) strong influence of these parameters on the clustering results and (ii) their optimal values heavily rely on the selected value for m. Thus, we decided to proceed as follows. Regarding the exponential and noise techniques, for a given method and value of m, the correct classification rate was recorded for a grid of equispaced values for the hyperparameter ranging from 0 to L, L being large enough to achieve near-zero rates. By doing so, we intend to assess not only the maximum rate of correct classification a method is able to achieve, but also its sensitivity against the choice of the hyperparameter. Concerning the trimmed procedures, the trimming ratio α was chosen as to detect the real number of outliers, one in Scenarios 1.1, 2.1 and 3.1 and two in Scenarios 1.2, 2.2 and 3.2.

For each one of the six scenarios, 100 trials were carried out with all the described techniques and the average percentages of correct classification were computed as a measure of clustering effectiveness. In each scenario, two values for the series length were considered: $T \in \{750, 1500\}$ for Scenarios 1.1 and 1.2, $T \in \{600, 900\}$ for Scenarios 2.1 and 2.2, and $T \in \{1500, 3000\}$ for Scenarios 3.1 and 3.2. Since each scenario contains very distinct types of processes, it is reasonable that very different values of T are needed to make an appropriate evaluation. In particular, the large values of T in Scenarios 3.1 and 3.2 are due to the high variability inherent to the estimation process associated with heteroskedastic models, which requires to employ large realizations to obtain accurate results, i.e., to enable d_{QCD} for discriminating among clusters. The 2DS plots in Fig. 1 clearly support this argument by showing that, for a fixed T, Scenario 3.2 is expected to be the most demanding scenario. It is worth enhancing that this requirement is not necessarily a drawback since these sample sizes are often encountered in real MTS fitted by BEKK models [68,69]. Indeed, multivariate series of stock returns and other related financial quantities, usually formed by daily or even intra-daily data, are one common example of series fitted through this class of models.

Table 1 presents a summary of relevant dimensions involved at each scenario, including series length, number of series, number of entries in the corresponding MTS, and the length of the vectors $\mathbf{\Psi}^{(i)}$ before and after applying the PCA

Table 1
Summary of relevant dimensions involved at each of the simulated scenarios.

Scenario	T	Series	$D(\boldsymbol{X}_t^{(i)})$	$l(\boldsymbol{\Psi}^{(i)})$ (pre PCA)	$l(\boldsymbol{\Psi}^{(i)})$ (post PCA)
1.1, 1.2	750	12	1500	121338	2
	1500	12	3000	242838	2
2.1, 22	600	12	1200	97038	2
	900	12	1800	145638	2
3.1, 3.2	1500	12	3000	242838	2
	3000	12	6000	485838	2

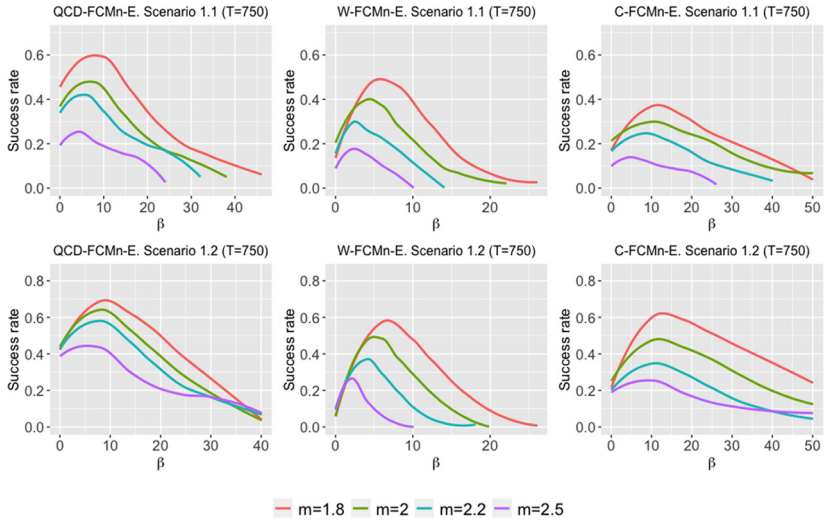

Fig. 2. Average rates of correct classification for QCD-FCMn-E, W-FCMn-E and C-FCMn-E as a function of β in Scenarios 1.1 and 1.2 for series of length $T = 750$ and four fuzziness levels m.

transformation. The notation $D(\boldsymbol{X}_t^{(i)})$ stands for the number of elements in MTS $\boldsymbol{X}_t^{(i)}$ and $l(\boldsymbol{v})$ denotes the length of vector $\boldsymbol{v}$.

5.3. Results for Scenarios 1.1 and 1.2

The curves of average rates of correct classification for the exponential-based approaches concerning Scenarios 1.1 and 1.2 are given in Fig. 2, for $T = 750$, and in Fig. 8 in Supplementary material, for $T = 1500$.

Figs. 2 and 8 illustrate that the three procedures performed pretty well when dealing with linear processes. For $T = 750$, QCD-FCMn-E achieved a maximum classification rate considerably above the one for W-FCMn-E and C-FCMn-E in both scenarios. For $T = 1500$, the three techniques showed a maximum average score close to one for $m = 1.8$, 2 and 2.2 and a bit worse for $m = 2.5$. The only exception was C-FCMn-E in Scenario 1.1, displaying a maximum average accuracy quite below that of its competitors for all values of m. C-FCMn-E was the least affected by the choice of β, usually attaining acceptable classification rates for large values of β, for which the remaining two techniques obtained poor results. This fact is not surprising since the sample autocorrelations are known to exhibit high capability for discriminating among linear models.

The results of Figs. 2 and 8 are summarized in Table 2, where the maximum accuracy values and the areas under the curves (AUC) are given. To make comparisons easier, the AUC quantities were normalized with respect to the maximum value of AUC attained by a method for fixed T and m. This way, the displayed AUC scores are bounded between 0 and 1. The results in Table 2 corroborate our prior remarks. In short, the three methods seem fairly capable of performing robust clustering of linear models, QCD-FCMn-E drawing better results for shorter lengths and C-FCMn-E exhibiting the highest robustness to the selection of β. Furthermore, the maximum rates of correct classification were considered to carry out statistical tests comparing pairs of proportions. Specifically, McNemar tests for paired proportions were performed for each scenario and values of m and T. An asterisk was incorporated in Table 2 if the corresponding method resulted significantly more effective than its two competitors for a significance level 0.05. Thus, method QCD-FCMn-E was significantly better than its competitors for $T = 750$. However, when $T = 1500$, the great behaviour of the wavelet-based technique, W-FCMn-E, stopped QCD-FCMn-E from taking the lead. Similar comparisons have been carried out in subsequent analyses.

Results for the noise approaches regarding Scenarios 1.1 and 1.2 are shown in Table 3, Fig. 3 for $T = 750$, and Fig. 9 in Supplementary material for $T = 1500$. Again, the three strategies performed well, although the results are slightly distinct

Table 2
Maximum correct classification rate and normalized AUC for QCD-FCMn-E, W-FCMn-E and C-FCMn-E. Scenarios 1.1 and 1.2. An asterisk indicates that a given method is significantly more effective than its competitors at level 0.05.

Exponential-based approach		Maximum			AUC		
		QCD	W	C	QCD	W	C
Scenario 1.1							
$T = 750$	$m = 1.8$	**0.60**	0.50	0.38	0.94	0.45	**1**
	$m = 2.0$	*__0.49__	0.41	0.29	0.97	0.48	**1**
	$m = 2.2$	*__0.43__	0.27	0.23	**1**	0.41	0.89
	$m = 2.5$	*__0.25__	0.14	0.12	**1**	0.32	0.41
$T = 1500$	$m = 1.8$	0.97	**0.98**	0.84	0.54	0.54	**1**
	$m = 2.0$	0.95	**0.96**	0.73	0.51	0.52	**1**
	$m = 2.2$	0.90	**0.92**	0.65	0.53	0.53	**1**
	$m = 2.5$	**0.83**	0.79	0.50	0.66	0.58	**1**
Scenario 1.2							
$T = 750$	$m = 1.8$	**0.71**	0.59	0.63	0.77	0.30	**1**
	$m = 2.0$	*__0.63__	0.49	0.48	0.81	0.29	**1**
	$m = 2.2$	*__0.57__	0.39	0.37	0.87	0.26	**1**
	$m = 2.5$	*__0.44__	0.26	0.25	**1**	0.23	0.94
$T = 1500$	$m = 1.8$	**0.99**	0.96	0.96	0.64	0.53	**1**
	$m = 2.0$	**0.98**	0.95	0.89	0.55	0.45	**1**
	$m = 2.2$	**0.95**	0.90	0.83	0.48	0.40	**1**
	$m = 2.5$	*__0.88__	0.78	0.67	0.54	0.41	**1**

Table 3
Maximum correct classification rate and normalized AUC for QCD-FCMn-NC, W-FCMn-NC and C-FCMn-NC. Scenarios 1.1 and 1.2. An asterisk indicates that a given method is significantly more effective than its competitors at level 0.05.

Noise cluster approach		Maximum			AUC		
		QCD	W	C	QCD	W	C
Scenario 1.1							
$T = 750$	$m = 1.8$	**0.63**	0.55	0.58	0.98	**1**	0.70
	$m = 2.0$	*__0.43__	0.29	0.36	**1**	0.86	0.72
	$m = 2.2$	*__0.26__	0.12	0.17	**1**	0.44	0.52
	$m = 2.5$	*__0.07__	0.01	0.01	**1**	0.11	0
$T = 1500$	$m = 1.8$	**0.95**	0.93	0.94	**1**	0.83	0.59
	$m = 2.0$	0.88	0.84	**0.91**	**1**	0.75	0.59
	$m = 2.2$	**0.73**	0.60	0.67	**1**	0.63	0.52
	$m = 2.5$	*__0.31__	0.15	0.20	**1**	0.31	0.35
Scenario 1.2							
$T = 750$	$m = 1.8$	**0.61**	0.34	0.58	**0.14**	0.06	0.12
	$m = 2.0$	*__0.43__	0.15	0.34	**1**	0.25	0.75
	$m = 2.2$	**0.20**	0.01	0.18	**1**	0	0.70
	$m = 2.5$	**0.05**	0	0.02	**1**	0	0.50
$T = 1500$	$m = 1.8$	**0.95**	0.89	0.93	**1**	0.48	0.62
	$m = 2.0$	**0.85**	0.68	0.81	**1**	0.39	0.57
	$m = 2.2$	*__0.74__	0.40	0.54	**1**	0.28	0.48
	$m = 2.5$	*__0.32__	0.05	0.16	**1**	0.08	0.32

from the ones obtained with the exponential approach. For $T = 1500$ and $m = 1.8$, we obtained excellent rates of correct classification close to one. When the value of m got larger, the overall performance substantially decreased, specially for $m = 2.5$ with very poor scores. When $m = 2.5$, all the procedures produced very blurry partitions in which the outlying series were generally allocated in the noise cluster with a membership degree less than 0.5, thus causing failed trials. Table 3 shows that the scores from QCD-FCMn-NC are clearly above the ones concerning W-FCMn-NC and C-FCMn-NC, thus concluding that QCD-FCMn-NC was the best performing model in terms of both maximum accuracy and robustness to the choice of δ.

The three trimming-based procedures accomplished very similar proportions of successful trials in the linear scenarios (Table 4). Overall, QCD-FCMn-T outperformed W-FCMn-T and C-FCMn-T, particularly in Scenario 1.2 with $T = 750$. By comparing the results in Tables 2, 3 and 4, we conclude that the trimmed approaches are slightly more effective than their counterparts based on the exponential and noise techniques in Scenarios 1.1 and 1.2.

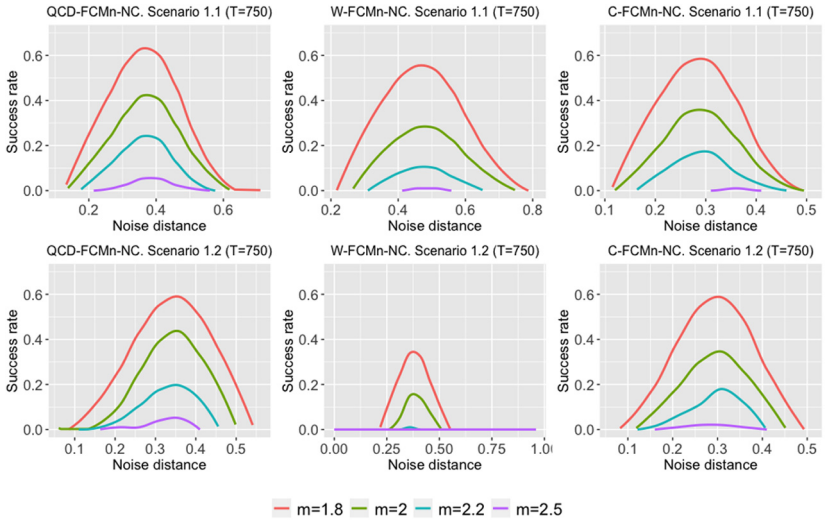

Fig. 3. Average rates of correct classification for QCD-FCMn-NC, W-FCMn-NC and C-FCMn-NC as a function of δ in Scenarios 1.1 and 1.2 for series of length $T = 750$ and four fuzziness levels m.

Table 4
Average rates of correct classification for QCD-FCMn-T, W-FCMn-T and C-FCMn-T. Scenarios 1.1 and 1.2. An asterisk indicates that a given method is significantly more effective than its competitors at level 0.05.

Trimmed approach		Scenario 1.1			Scenario 1.2		
		QCD	W	C	QCD	W	C
$T = 750$	$m = 1.8$	0.67	0.65	**0.71**	**0.73**	0.67	0.69
	$m = 2.0$	**0.59**	0.55	0.56	**0.66**	0.59	0.61
	$m = 2.2$	**0.52**	0.44	0.44	***0.60**	0.48	0.49
	$m = 2.5$	**0.33**	0.30	0.23	***0.47**	0.31	0.30
$T = 1500$	$m = 1.8$	**1**	0.98	0.95	0.98	**0.99**	0.97
	$m = 2.0$	0.97	**0.98**	0.93	**0.95**	0.93	0.91
	$m = 2.2$	0.95	**0.98**	0.87	**0.93**	0.91	0.87
	$m = 2.5$	0.95	**0.98**	0.87	**0.93**	0.91	0.87

5.4. Results for Scenarios 2.1 and 2.2

Results of QCD-FCMn-E, W-FCMn-E and C-FCMn-E in Scenarios 2.1 and 2.2, with nonlinear models, are given in Table 5, Fig. 4 for $T = 600$ and Fig. 10 in Supplementary material for $T = 900$. In all cases, QCD-FCMn-E substantially outmatched the alternative approaches. The correlation-based strategy C-FCMn-E was able to make some correct classifications for a broader range of values for β, but attaining an overall accuracy much less than QCD-FCMn-E. W-FCMn-E accomplished very poor results for all values of m, β and T. It is worth noting that the average rates of successful trials attained by our proposal were always close to one.

In Scenarios 2.1 and 2.2, the noise-based approaches led to very similar results (Table 6, see also Figs. 11 and 12 in Supplementary material). Again, QCD-FCMn-NC clearly obtained the best average scores by far. The wavelet-based method W-FCMn-NC was incapable of making correct classifications, and C-FCMn-NC achieved very low success rates in Scenario 2.1, specially for larger values of m, and was totally ineffective in Scenario 2.2.

From Table 6 clearly follows that W-FCMn-NC and C-FCMn-NC are not suitable techniques for treating with the nonlinear processes defining Scenarios 2.1 and 2.2. To unravel the reasons behind of this poor behaviour, we carried out a detailed examination of the clustering solutions. It was observed that W-FCMn-NC often allocated the single or both outlying series in the noise cluster with large membership degrees, but in addition, some of the series generated from the BL process were also placed in the noise cluster, thus provoking unsuccessful trials. When increasing the value of noise distance δ, the anomalous MTS get assigned to regular clusters, and at least one of the BL series remains located in the noise cluster until δ is large enough. As a result, W-FCMn-NC is not able to perform successful classifications. A similar situation occurs in Scenario 2.2 for C-FCMn-NC. The partitions returned by C-FCMn-NC in this scenario are characterized by the allocation of at least one of the BL series in the noise cluster, hence producing failure trials. In Scenario 2.1, although some BL series are also frequently placed in the noise cluster, there exist some trials in which all these series are correctly classified whereas the single outlying series is usually located into the noise cluster. This accounts for the low success rate achieved by C-FCMn-NC in Scenario 2.1. In short, the procedures W-FCMn-NC and C-FCMn-NC find it hard to classify the series generated from the

Table 5
Maximum correct classification rate and normalized AUC for QCD-FCMn-E, W-FCMn-E and C-FCMn-E. Scenarios 2.1 and 2.2. An asterisk indicates that a given method is significantly more effective than its competitors at level 0.05.

Exponential-based approach		Maximum			AUC		
		QCD	W	C	QCD	W	C
Scenario 2.1							
$T = 600$	$m = 1.8$	*__0.93__	0.30	0.35	*__1__	0.02	0.32
	$m = 2.0$	*__0.94__	0.27	0.40	*__1__	0.01	0.43
	$m = 2.2$	*__0.96__	0.16	0.44	*__1__	0.01	0.54
	$m = 2.5$	*__0.98__	0.09	0.46	*__1__	0.01	0.74
$T = 900$	$m = 1.8$	*__0.98__	0.35	0.32	*__1__	0.03	0.38
	$m = 2.0$	*__0.99__	0.36	0.43	*__1__	0.02	0.55
	$m = 2.2$	*__0.99__	0.31	0.52	__1__	0.02	0.74
	$m = 2.5$	*__1__	0.17	0.58	0.98	0.01	__1__
Scenario 2.2							
$T = 600$	$m = 1.8$	*__0.99__	0.27	0.27	*__1__	0.02	0.32
	$m = 2.0$	*__0.99__	0.28	0.35	*__1__	0.01	0.42
	$m = 2.2$	*__0.99__	0.21	0.37	*__1__	0.01	0.47
	$m = 2.5$	*__0.99__	0.07	0.45	*__1__	0.01	0.53
$T = 900$	$m = 1.8$	*__1__	0.28	0.22	*__1__	0.03	0.31
	$m = 2.0$	*__1__	0.26	0.29	*__1__	0.02	0.46
	$m = 2.2$	*__1__	0.24	0.35	*__1__	0.01	0.58
	$m = 2.2$	*__1__	0.12	0.46	*__1__	0.01	0.70

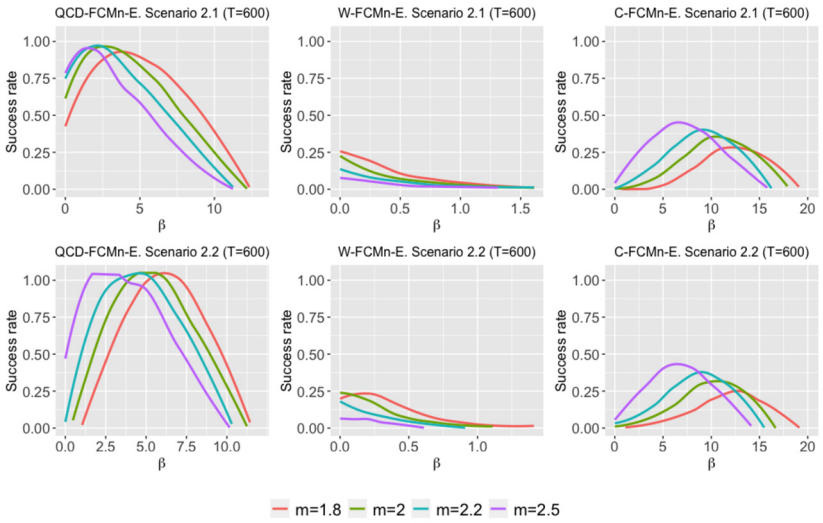

Fig. 4. Average rates of correct classification for QCD-FCMn-E, W-FCMn-E and C-FCMn-E as a function of β in Scenarios 2.1 and 2.2 for series of length $T = 600$ and four fuzziness levels m.

BL process in Scenarios 2.1 and 2.2, which in turn leads to these methods being incapable of counterbalancing the effects of the abnormal MTS.

Interestingly enough, the QCD-based approach substantially ameliorated its efficacy when dealing with two outliers. To gain insights into this fact, we fixed $T = 600$, $m = 1.8$, and took the membership matrices produced by QCD-FCMn-NC in each one of the trials with the optimal values of δ (see Fig. 11 in Supplementary material), namely $\delta = 0.38$ for Scenario 2.1 and $\delta = 0.3$ for Scenario 2.2. In each trial, we first recorded the maximum membership value for each MTS, and then group these values separately by kind of series (EXPAR and BL) and selected the minimum at each group. These minimum values are crucial to know if the method made a correct classification. In addition, the membership values of the outlying series in relation to the noise cluster were also recorded. The boxplots of these minimum membership degrees over all the trials are shown in Fig. 5.

The left panel of Fig. 5 illustrates that the failure trials of QCD-FCMn-NC in Scenario 2.1 are mainly due to the series generated from the BL process (Cluster 2). Almost half of the times at least one BL series was assigned to Cluster 1 with a membership value less than 0.7. Although some outlier series were located in the noise cluster with a membership degree less than 0.5, the number of misclassifications due to this fact was negligible. The success rate of QCD-FCMn-NC in this setting was 0.53. When the second atypical series from the VAR process was included (right panel of Fig. 5), the minimum

Table 6
Maximum correct classification rate and normalized AUC for QCD-FCMn-NC, W-FCMn-NC and C-FCMn-NC. Scenarios 2.1 and 2.2. An asterisk indicates that a given method is significantly more effective than its competitors at level 0.05.

Noise cluster approach		Maximum			AUC		
		QCD	W	C	QCD	W	C
Scenario 2.1							
$T = 600$	$m = 1.8$	***0.53**	0	0.14	***1**	0	0.14
	$m = 2.0$	***0.34**	0	0.09	***1**	0	0.10
	$m = 2.2$	***0.29**	0	0.03	***1**	0	0.07
	$m = 2.5$	***0.17**	0	0	***1**	0	0
$T = 900$	$m = 1.8$	***0.88**	0	0.23	***1**	0	0.09
	$m = 2.0$	***0.81**	0	0.14	***1**	0	0.07
	$m = 2.2$	***0.70**	0	0.07	***1**	0	0.03
	$m = 2.5$	***0.58**	0	0.01	***1**	0	0
Scenario 2.2							
$T = 600$	$m = 1.8$	***0.97**	0	0	***1**	0	0
	$m = 2.0$	***0.94**	0	0	***1**	0	0
	$m = 2.2$	***0.87**	0	0	***1**	0	0
	$m = 2.5$	***0.66**	0	0	***1**	0	0
$T = 900$	$m = 1.8$	***1**	0	0	***1**	0	0
	$m = 2.0$	***1**	0	0	***1**	0	0
	$m = 2.2$	***0.99**	0	0	***1**	0	0
	$m = 2.5$	***0.91**	0	0	***1**	0	0

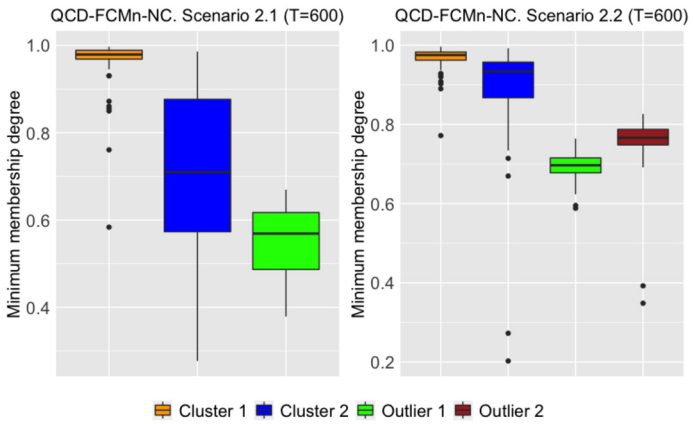

Fig. 5. Minimum membership degrees of the series in cluster 1, the series in cluster 2 and the outlying series in the noise cluster.

membership degrees of the BL series considerably increased and both anomalous series were generally allocated in the noise cluster with membership values far beyond 0.5. Thus, including the VAR series made QCD-FCMn-NC better capable of both neutralizing the effects of the outliers and detecting the true clustering structure, thus attaining a greater success rate (0.97). The corresponding graph for Scenario 2.2 in the 2DS scatter plots of Fig. 1 help us to understand the improvement exhibited by QCD-FCMn-NC when dealing with two outliers. We observe that the first anomalous series is closer to the regular clusters than the second one. Thus, when the second outlying series, which is close to the first one, is introduced, a small "cluster" of two outliers is formed, making it easier for the technique to distinguish between the true clusters and the abnormal MTS.

The scores generated by the trimmed techniques in Scenarios 2.1 and 2.2 are given in Table 7. Not surprisingly, QCD-FCMn-T obtained the best scores by a large degree. The method C-FCMn-T attained markedly greater results than C-FCMn-E and C-FCMn-NC, but they are not comparable with the ones of QCD-FCMn-T. It is worth noting that, in the context of nonlinear processes, the exponential approach QCD-FCMn-E displayed a better performance than QCD-FCMn-NC and QCD-FCMn-T.

5.5. Results for Scenarios 3.1 and 3.2

With regards to the results for Scenarios 3.1 and 3.2, it is worth pointing out that both the wavelet-based and the correlation-based methods achieved very poor scores. In fact, the exponential and the noise approaches led to success rates close to zero. The four strategies W-FCMn-E, C-FCMn-E, W-FCMn-NC and C-FCMn-NC are prone to the same types of error.

Table 7
Average rates of correct classification for QCD-FCMn-T, W-FCMn-T and C-FCMn-T. Scenarios 2.1 and 2.2. An asterisk indicates that a given method is significantly more effective than its competitors at level 0.05.

Trimmed approach		Scenario 2.1			Scenario 2.2		
		QCD	W	C	QCD	W	C
$T = 600$	$m = 1.8$	*__0.78__	0.13	0.54	*__0.98__	0.21	0.53
	$m = 2.0$	*__0.76__	0.15	0.54	*__0.98__	0.17	0.54
	$m = 2.2$	*__0.74__	0.10	0.53	*__0.98__	0.14	0.52
	$m = 2.5$	*__0.69__	0.08	0.53	*__0.98__	0.12	0.51
$T = 900$	$m = 1.8$	*__0.96__	0.22	0.68	*__1__	0.17	0.63
	$m = 2.0$	*__0.96__	0.18	0.68	*__1__	0.17	0.62
	$m = 2.2$	*__0.95__	0.18	0.68	*__1__	0.14	0.63
	$m = 2.5$	*__0.95__	0.18	0.68	*__1__	0.14	0.63

Table 8
Maximum correct classification rate and AUC for QCD-FCMn-E and QCD-FCMn-NC. Scenarios 3.1 and 3.2.

		QCD-FCMn-E		QCD-FCMn-NC	
		Max	AUC	Max	AUC
Scenario 3.1					
$T = 1500$	$m = 1.8$	0.99	30.83	0.99	0.419
	$m = 2.0$	0.99	27.18	0.97	0.349
	$m = 2.2$	0.99	22.88	0.90	0.309
	$m = 2.5$	0.99	18.03	0.71	0.157
$T = 3000$	$m = 1.8$	1	21.45	1	0.659
	$m = 2.0$	1	19.15	1	0.582
	$m = 2.2$	0.99	17.08	1	0.513
	$m = 2.5$	1	13.66	0.95	0.396
Scenario 3.2					
$T = 1500$	$m = 1.8$	0.56	4.87	0.92	0.294
	$m = 2.0$	0.53	4.26	0.85	0.242
	$m = 2.2$	0.47	3.18	0.75	0.191
	$m = 2.5$	0.42	1.99	0.48	0.103
$T = 3000$	$m = 1.8$	0.59	3.10	1	0.494
	$m = 2.0$	0.62	2.61	1	0.440
	$m = 2.2$	0.57	2.01	0.97	0.374
	$m = 2.5$	0.47	1.16	0.88	0.282

Depending on the value of β or δ, they usually (i) are not capable of clearly differentiating between both BEKK processes, since the ten MTS in the base scenario obtained membership degrees close to 0.5 in both clusters (the same occurred for the atypical series), or (ii) locate all the BEKK series in one cluster with high memberships and the one or two outlying series in another cluster, also with high memberships. In addition, the noise cluster techniques W-FCMn-NC and C-FCMn-NC were sometimes capable of assigning correctly the outlying series into the noise cluster, but in this case the BEKK series are mixed. In conclusion, it is clear that these approaches are not able to properly distinguish between BEKK models nor neutralize the effect of the abnormal series. Therefore, in Scenarios 3.1 and 3.2, we decided only to show the results based on the QCD approaches. In particular, the results for QCD-FCMn-E and QCD-FCMn-NC are jointly shown in Table 8 (see also Figs. 13, 14, 15 and 16 in Supplementary material). Note that in this case the AUC values were not normalized because only the results based on QCD are considered.

The results in Table 8 indicate that both models performed very well when only one outlier was present (Scenario 3.1), reaching perfect results regardless of the value of m. The scores worsened when the additional outlier given by the BL process was added (Scenario 3.2). It seems that this second outlying series made much more challenging the clustering task, particularly for QCD-FCMn-E. In fact, in Scenario 3.2, the noise cluster approach performed clearly better than the exponential one for both values of T and all values of m. For $T = 3000$, QCD-FCMn-NC attained perfect scores when $m = 1.8$, 2, 2.2 for some values of δ.

In these scenarios involving conditional heteroskedastic processes, the trimmed method QCD-FCMn-T exhibited the same patterns as QCD-FCMn-E and QCD-FCMn-NC (see Table 18 in Supplementary Material): high scores in Scenario 3.1 and worse behaviour in Scenario 3.2. Notice that, in the case of BEKK processes, the exponential-based strategy defeated both QCD-FCMn-NC and QCD-FCMn-T.

It can be concluded that each one of the approaches, exponential, noise and trimmed, has shown to be the most appropriate choice for coping with a different type of generating processes. This insight is illuminating, as it indicates that

Table 9
Rates of correct classification reached by QCD-FCMn and by the best performing robust QCD approach (in brackets).

Scenario	T	$m = 1.8$	$m = 2$	$m = 2.2$	$m = 2.5$
1.1	750	0.37 (**0.67**)	0.39 (**0.59**)	0.38 (**0.52**)	0.24 (**0.33**)
	1500	0.33 (**1**)	0.47 (**0.97**)	0.53 (**0.95**)	0.63 (**0.95**)
1.2	750	0.48 (**0.73**)	0.56 (**0.66**)	0.55 (**0.60**)	0.43 (**0.47**)
	1500	0.87 (**0.99**)	0.95 (**0.98**)	**0.95** (**0.95**)	0.88 (**0.93**)
2.1	600	0.55 (**0.93**)	0.79 (**0.94**)	0.89 (**0.96**)	**0.98** (**0.98**)
	900	0.67 (**0.98**)	0.87 (**0.99**)	0.96 (**0.99**)	**1** (**1**)
2.2	600	0.29 (**0.99**)	0.53 (**0.99**)	0.78 (**0.99**)	0.94 (**0.99**)
	900	0.32 (**1**)	0.56 (**1**)	0.81 (**1**)	0.99 (**1**)
3.1	1500	0.76 (**1**)	0.97 (**0.99**)	**1** (0.99)	**1** (0.99)
	3000	0.70 (**1**)	**1** (**1**)	**1** (**1**)	**1** (**1**)
3.2	1500	0 (**0.92**)	0 (**0.85**)	0 (**0.75**)	0 (**0.48**)
	3000	0 (**1**)	0 (**1**)	0 (**0.97**)	0 (**0.88**)

no method dominates the remaining ones under all circumstances. Thus, practitioners should always take into account the three analyzed techniques when performing robust clustering of MTS.

5.6. Additional results. Comparison between QCD-FCMn and its robust extensions

To better understand to what extent the QCD-based robust approaches are useful, Table 9 simultaneously provides the rates of correct classification reached by the nonrobust QCD-FCMn method and the highest one obtained among the three robust approaches (in brackets). On balance, the robust techniques are clearly beneficial when there exist outlier series in the dataset. Although the differences are more pronounced for the smallest values of m, they hold for all m, thus concluding that the robust strategies are far less dependent on the suitable choice of m. This is a very desirable property given the difficulty of properly selecting m in practice. The results for Scenario 3.2 are particularly noteworthy since, while the method QCD-FCMn was incapable of performing correct classifications, the robust versions still took advantage of their powerful neutralizing capacity and produced high scores.

So far, we have evaluated the quality of the cluster assignment, that is, the capability of each model to correctly allocate each series to its cluster, separating the outlying series. When considering robust clustering models, it is also important to assess to what extent the centroids are affected by the presence of outliers. Indeed, robust techniques are expected to produce more accurate prototypes than their nonrobust counterparts. Given a scenario and fixed values for m and T, we proceeded to measure the discrepancy between empirical and theoretical centroids by computing the quantity

$$\frac{1}{100}\sum_{i=1}^{100}\frac{1}{2}\left(\left\|\overline{\boldsymbol{\Psi}}_i^{(1)}-\overline{\boldsymbol{\Psi}}^{*(1)}\right\|+\left\|\overline{\boldsymbol{\Psi}}_i^{(2)}-\overline{\boldsymbol{\Psi}}^{*(2)}\right\|\right),$$

where, for $i = 1, \ldots, 100$ and $j = 1, 2$, $\overline{\boldsymbol{\Psi}}_i^{(j)}$ is the centroid of j-th cluster returned by the corresponding algorithm in the i-th trial, and $\overline{\boldsymbol{\Psi}}^{*(j)}$ is the theoretical centroid of the j-th cluster approximated through Monte Carlo simulations. Thus, the average deviation over the 100 simulated trials is considered in our assessment. The results are given in Table 10, where only scenarios with two outlying series (Scenarios 1.2, 2.2 and 3.2), and the largest values of T ($T = 1500$, $T = 900$ and $T = 3000$, respectively), are considered for the sake of simplicity.

According to Table 10, the three robust techniques always provide more reliable centroids than the standard method QCD-FCMn, which is negatively affected by the anomalous series. The metric approach performs the worst in the comparison, attaining only marginal gains in Scenarios 1.2 and 2.2. On the contrary, the noise and trimmed approaches produce excellent approximations to the theoretical centroids, particularly in Scenarios 2.2 and 3.2, where they exhibit a dramatic advantage over QCD-FCMn.

In short, the results presented in this section clearly corroborate the usefulness of the considered robust extensions.

5.7. Results for a different error distribution

All simulations were replicated by considering heavy tails in the error distribution, since this feature often arises in real time series, specially in finance [70–73]. Specifically, the innovations in all scenarios were generated from a multivariate t distribution with 3 degrees of freedom. Results for exponential, noise, and trimmed techniques are provided in Supplementary material, in Tables 19, 20 and 21, respectively. It is clearly observed that the quantile-based approaches exhibited the most robustness against the effect of heavy tails. For instance, whereas the three robust approaches concerning W-FCMn and

Table 10
Average deviations exhibited by the empirical centroids with respect to the theoretical ones.

Scenario 1.2	$m = 1.8$	$m = 2$	$m = 2.2$	$m = 2.5$
QCD-FCMn	0.090	0.082	0.078	0.074
QCD-FCMn-E	0.073	0.073	0.073	0.073
QCD-FCMn-NC	**0.059**	**0.059**	0.059	0.059
QCD-FCMn-T	**0.059**	**0.059**	**0.058**	**0.058**
Scenario 2.2	$m = 1.8$	$m = 2$	$m = 2.2$	$m = 2.5$
QCD-FCMn	0.154	0.140	0.128	0.112
QCD-FCMn-E	0.148	0.135	0.124	0.110
QCD-FCMn-NC	0.034	0.035	0.036	0.036
QCD-FCMn-T	**0.029**	**0.029**	**0.029**	**0.029**
Scenario 3.2	$m = 1.8$	$m = 2$	$m = 2.2$	$m = 2.5$
QCD-FCMn	1.301	1.281	1.318	1.356
QCD-FCMn-E	0.741	0.749	0.778	0.797
QCD-FCMn-NC	0.053	0.053	0.053	0.054
QCD-FCMn-T	**0.052**	**0.052**	**0.052**	**0.052**

C-FCMn performed quite accurately when handling linear processes with normal innovations (Tables 2-4), they totally failed when some amount of fat-tailedness was introduced. Only the correlation-based technique was capable of attaining some successful trials, particularly the trimmed approach (C-FCMn-T). Similar behaviour is observed for the rest of scenarios. Interestingly, the trimmed-based approaches W-FCMn-T and C-FCMn-T, although outperformed by QCD-FCMn-T, obtained high scores in Scenario 3.1 involving BEKK processes, even improving their results with Gaussian innovations (Table 18). However, for Scenario 3.2, QCD-FCMn-T was the only approach attaining high success rates.

Overall, the change in the error distribution did not affect the performance of the proposed robust approaches in Scenarios 1.1 and 1.2, and slightly decreased their scores in Scenarios 2.1, 2.2, 3.1 and 3.2, specially for the shortest value of the series length. In any case, the robust QCD-based versions have proven to be the best ones in terms of robustness not only against outlying series, but also against the distributional form of the error terms.

6. Evaluating robustness with respect to contextual outliers

This section is devoted to assess the effectiveness of the analyzed robust methods with respect to contextual outliers, that is, data objects differing from the remaining according to a specific context, such as local deviations over a range of neighbouring time points or changes in the distribution of the innovations, among others.

First, we describe the experimental design and then we show the results from the simulation study.

6.1. Experimental design

We consider two classes of outliers which are pervasive in the MTS literature: the multivariate innovational outliers (MIO) and the multivariate temporary (or transitory) changes (MTC) (see e.g., [74,75]). Note that, although these types of outliers are usually defined in the context of VARMA processes, their extension to general types of processes is straightforward. Generally speaking, MIO appear when the noise distribution of the original series is perverted and a MTC occurs when the series receives a sudden impact that disappears gradually with time. Both outliers are assumed to emerge at some particular time point, say t_0. Specifically, if $\boldsymbol{X}_t$ denotes an arbitrary d-dimensional realization generated from a given cluster with innovations following a distribution F, then a MIO series $\boldsymbol{X}'_t$ is generated identically as $\boldsymbol{X}_t$ for $t = 1, \ldots, t_0 - 1$, but with innovations following a distribution F' for $t = t_0, \ldots, T$. On the other hand, a MTC series $\boldsymbol{X}''_t$ is constructed as $\boldsymbol{X}''_t = \boldsymbol{X}_t$ if $t < t_0$ and $\boldsymbol{X}''_{t_0+k} = \boldsymbol{X}_{t_0} + \eta^k \boldsymbol{w}$ for $k = 0, 1, \ldots T - t_0$, where $\boldsymbol{w} = (w_1, \ldots, w_d)^\intercal$ is the size of the outlier and $0 < \eta < 1$ is a constant regulating the propagation of the anomalous effect in subsequent observations.

Six new simulation scenarios, referred to as MTC 1, MTC 2, MTC 3, MIO 1, MIO 2 and MIO 3, were designed to investigate how the robust techniques deal with this class of outliers. For $i \in \{1, 2, 3\}$, Scenario MTC (MIO) i consisted of 5 realizations of each generating process in Base Scenario i plus an MTC (MIO) outlier constructed from the first generating process in Base Scenario i. The outlying effect was introduced at $t_0 = T/2$ in all cases. The outlier sizes used in the definition of the MTC outliers were $\boldsymbol{w} = (5, -5)^\intercal$ in Scenarios MTC 1 and MTC 2, and $\boldsymbol{w} = (1, -1)^\intercal$ in Scenario MTC 3. The parameter η was always set to 0.99. As for the MIO outliers, the distribution F' was chosen to be the χ^2 distribution with 3 degrees of freedom in Scenarios MIO 1 and MIO 2, and the χ^2 with 3/10 degrees of freedom in Scenario MIO 3. Note that the selected parameters concerning Scenarios MTC 3 and MIO 3 produce less pronounced changes in the corresponding outlying series. The reason is that the marginal variance of both components in the first generating process of Base Scenario 3 is far less than that of their counterparts in Base Scenarios 1 and 2. Hence, a modification in both the size of the outliers and the mean and variance of the innovations needed to be made to maintain the difficulty in the outlier detection task in

Table 11
Summary of relevant dimensions involved at each of the new simulated scenarios.

Scenario	T	Series	$D(\boldsymbol{X}_t^{(i)})$	$l(\boldsymbol{\Psi}^{(i)})$ (pre PCA)	$l(\boldsymbol{\Psi}^{(i)})$ (post PCA)
MTC 1, MIO 1	1500	11	3000	242838	2
MTC 2, MIO 2	900	11	1800	145638	2
MTC 3, MIO 3	3000	11	6000	485838	2

Table 12
Averages rates of correct classification for robust approaches in Scenarios MTC 1, MTC 2, MTC 3, MIO 1, MIO 2 and MIO 3. An asterisk indicates that a given method is more effective than the remaining ones at a significance level of 0.05.

Scenario	T	m	Exponential (E)			Noise (NC)			Trimmed (T)		
			QCD	W	C	QCD	W	C	QCD	W	C
MTC 1	1500	1.8	**0.61**	0.55	0.13	*1	0	0.91	***0.99**	0.50	0.97
		2.2	0.48	***0.58**	0.14	***0.97**	0	0.67	***0.99**	0.33	0.87
MTC 2	900	1.8	**1**	0	**1**	*1	0	0.89	**1**	0	0.99
		2.2	**1**	0	0.99	***0.98**	0	0.42	**1**	0	0.99
MTC 3	3000	1.8	*1	0.3	0	*1	0	0	*1	0.53	0.02
		2.2	*1	0.25	0	*1	0	0	*1	0.34	0.01
MIO 1	1500	1.8	***0.83**	0.01	0.72	**0.99**	0.95	0.91	**0.99**	0.97	0.96
		2.2	**0.73**	0	0.65	**0.91**	0.70	0.45	**0.98**	0.93	0.93
MIO 2	900	1.8	**1**	0.14	0.99	**1**	0	0.99	**1**	0.23	0.99
		2.2	**1**	0.09	0.98	**0.99**	0	0.97	**1**	0.19	0.99
MIO 3	3000	1.8	*1	0.01	0	*1	0.41	0.02	*1	0.48	0.04
		2.2	*1	0.01	0	*1	0.16	0	*1	0.35	0

Scenarios MTC 3 and MIO 3. The remaining simulation features (cut-off values for fuzzy membership degrees, sample sizes, input parameters,...) and the alternative procedures are exactly the same as in Section 5.

Note that, unlike the outlying series handled in Section 5, MIO and MTC outliers can be detected by visual inspection of the realizations (see Fig. 17 in Supplementary material to elucidate this assertion). However, in practice, by dealing with a dataset with hundreds or thousands of MTS, it is often unfeasible to perform outlier detection by visual examination. Thus, it is desirable to have available a proper robust clustering algorithm to be able to detect and neutralize the effect of these types of atypical series.

A summary of the dimensions concerning each one of the new scenarios is provided in Table 11.

6.2. Results and discussion

Table 12 contains the average correct classification rates concerning the six new scenarios. For the sake of simplicity, we have included the results only for the largest value of the series length T and $m = 1.8$ and 2.2. It is evident from Table 12 that, overall, the QCD-based robust approaches outperform the other two methods in terms of outlier neutralization. Only in one occasion a QCD-based method did not attain the best rate of correct classification. This was QCD-FCMn-E in Scenario MTC 1 with $m = 2.2$, being outmatched by the wavelet-based procedure. Unlike the alternative techniques, the proposed methods achieved perfect classification rates in most of the settings. The correlation-based methods acquired very similar successful rates on some scenarios, but totally failed in several others. Specifically, Scenario MTC 1 along with the scenarios concerning conditional heteroskedastic processes were particularly challenging for this metric. The wavelet-based strategies reached very poor results.

From scores in Table 12 follows that the three QCD-based robust procedures are highly capable of dealing with the types of anomalous series presented in this section. Overall, the trimmed-based method yielded the best classification rates, closely followed by the noise approach. The results reached by the exponential procedure, although with quite high success rates, were worse by working with linear processes (Scenarios MTC 1 and MIO 1). Of course, the scores could be increased by using a less stringent cut-off value (e.g., 0.6 instead of 0.7). It is worth revealing that, although not shown in the manuscript, we have repeated the simulation study for different parameters concerning the definition of MTC and MIO type outliers. Particularly, we considered $\boldsymbol{w} = (2.5, -2.5)^\intercal$ ($\boldsymbol{w} = (0.5, -0.5)^\intercal$ in Scenario MTC 3) and $\eta = 0.9$ in the case of MTC type outliers, and 1.5 degrees of freedom for the χ^2 distribution (1.5/10 in Scenario MIO 3) in the case of MIO type outliers. As expected, the rates of correct classification were worse than those provided in Table 12 for all methods, but the QCD-based models were again the best performing ones. Simulations were also carried out introducing the outlying effect in $t_0 = 3T/4$. As the anomalous period is shorter, the results were again worse than those in Table 12, but one more time the quantile-based metric generally defeated its competitors. In short, the exponential, noise and trimmed approaches based on d_{QCD} have proven to be successful in counterbalancing the effects of distorted individual MTS.

7. Applications

The QCD-FCMn algorithm was used in [4] to perform clustering on two real data sets involving financial and environmental multidimensional time series, respectively. In both cases, meaningful solutions were obtained. However, in this section, the goal is to take a step further by applying the three robust versions of QCD-FCMn to make a proper treatment of potentially outlying series, which may have affected the solutions attained in [4]. It is worth pointing out that, in both cases, the goal is to show the usefulness of the proposed algorithms without intending to give financial or environmental advice.

7.1. Robust fuzzy clustering of the top 20 companies of S&P 500 index

The first dataset was taken from the finance section of the Yahoo website https://es.finance.yahoo.com. 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 U.S. companies. The top 20 involves some of the most important companies in the world, as Apple, Google, Facebook or Berkshire Hathaway.

It is worth noting that the relationship between price and volume has been extensively analyzed in the literature ([76–78]) and constitutes itself a topic of great financial interest. Prices and trading volume are known to exhibit some empirical linkages over the fluctuations of stock markets. Thus, it is interesting to identify groups of companies showing similar joint behaviour of prices and volume. Given the stochastic nature of both variables, we assume that two companies are close if their corresponding bivariate time series exhibit similar dependence structures.

Since the UTS of prices and trading volume are nonstationary in mean, all of them were 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. Next, all UTS were normalized to have zero mean and unit variance. The resulting MTS are depicted in Fig. 18 in Supplementary material. Overall, plots in Fig. 18 exhibit common traits of financial time series. There is a substantial degree of heteroskedasticity and both components exhibit the so-called phenomenon of volatility clustering: large values (positive or negative) tend to group together, resulting in a marked persistence. These particular 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 BEKK models considered in Scenarios 3.1 and 3.2. Recall that the robust fuzzy clustering algorithms showed high efficacy to cope with this type of models, specially when the error distribution possesses some amount of fat-tailedness (other property related to the stylized facts [70,79,80]). Therefore, QCD-FCMn-E, QCD-FCMn-NC and QCD-FCMn-T are expected to provide meaningful fuzzy partitions with regular groups of companies following a similar behavioural pattern and isolated MTS exhibiting atypical dependence structures.

In [4], the number of clusters C and the fuzziness parameter m were simultaneously selected according to the minimization of four internal clustering validity indexes. The optimal values were $C = 6$ and $m = 1.9$. In fact, the 6-cluster solution provided by QCD-FCMn (see Table 22 in Supplementary material) gave rise to meaningful groups, putting together companies sharing characteristics as the market capitalization or the company nature, among others. For instance, a cluster included the two giants of information technology, Apple (AAPL) and Microsoft (MSFT), and another cluster grouped together the two branches of Google (GOOGL and GOOG) and Amazon (AMZN). Fig. 6 shows a metric 2-dimensional scaling plot of the companies according to the pairwise QCD-based distance matrix. Since the associated R^2 is 0.7251, the scatter plot can be considered an acceptable representation of the underlying distance configuration [62].

From Fig. 6, it could be concluded that Berkshire Hathaway (BRK.B) is by far the most outlying company. In fact, Table 22 in Supplementary material shows that BRK.B constitutes an isolated cluster (C_3) in the resulting partition, with negligible membership values for the remaining companies. Thus, BRK.B is expected to be detected as an outlier by the robust approaches. Fig. 6 also shows any other potentially anomalous elements, located partly far from the bulk of the data points, as Visa (V) or Walmart (WMT). Therefore, it is interesting to analyze the results obtained with the three QCD-based robust procedures, particularly the resulting membership vectors for these companies.

Concerning the hyperparameters of each procedure, the coefficient β in QCD-FCMn-E was chosen by means of (15) in Section 4.2, resulting $\beta = 174.85$. The trimming rate α associated with QCD-FCMn-T was selected by considering a grid of values for α and choosing the one giving rise to the least average value of the indexes considered in [4], resulting $\alpha = 0.15$. As for the noise distance δ for QCD-FCMn-NC, we proceeded as follows. The algorithm was run several times for decreasing values of δ and the proportion of series placed in the noise cluster was recorded. Then, we chose the value of δ associated with a sudden change in this proportion. The underlying rationale here is that, by gradually decreasing the value of δ, an appropriate threshold will be figured out since substantially low values of δ lead to partitions where nonoutlying elements are located in the noise cluster. Lastly, three principal components were retained to set the score vectors defining the QCD distance according to the empirical rule presented in Section 3.2.

A summary of the relevant dimensions involved in the financial application is presented in Table 13.

Table 14 provides the fuzzy partition obtained by QCD-FCMn-E. Compared to the one generated by QCD-FCMn (Table 22), the main difference is that QCF-FCMn-E establishes that BRK.B is an outlier company since its membership degrees are split across the clusters almost uniformly. By dealing here with a 6-cluster solution, we decided to determine the i-th MTS as

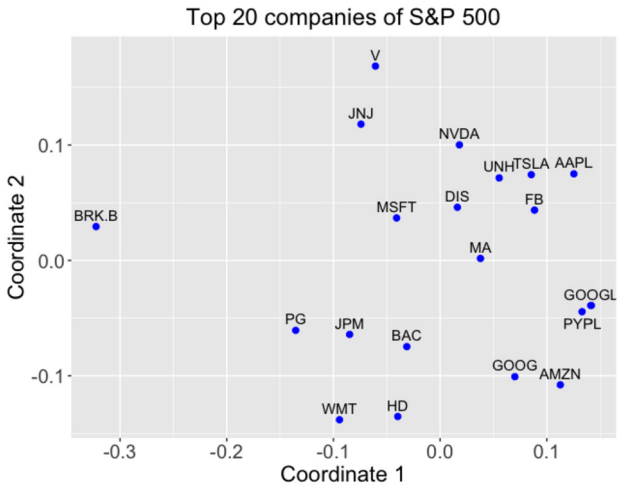

Fig. 6. Two-dimensional scaling plane based on the QCD-based distance for the daily returns and change in volume of the top 20 companies in the S&P 500 index.

Table 13
Summary of relevant dimensions involved in the financial application.

T	Series	$D(\boldsymbol{X}_t^{(i)})$	$l(\boldsymbol{\Psi}^{(i)})$ (pre PCA)	$l(\boldsymbol{\Psi}^{(i)})$ (post PCA)
655	20	1310	105948	3

Table 14
Membership degrees for top 20 companies in the S&P 500 index by considering the QCD-FCMn-E.

Company	C_1	C_2	C_3	C_4	C_5	C_6
AAPL	0.118	0.205	0.061	**0.469**	0.094	0.053
MSFT	0.134	0.072	0.203	**0.397**	0.125	0.069
AMZN	**0.854**	0.027	0.029	0.044	0.018	0.028
GOOGL	**0.666**	0.053	0.055	0.137	0.044	0.044
GOOG	**0.905**	0.014	0.023	0.030	0.011	0.016
FB	0.005	**0.966**	0.005	0.009	0.009	0.007
TSLA	0.064	0.041	0.029	**0.808**	0.038	0.020
BRK.B	0.126	0.121	**0.273**	0.147	0.183	0.151
V	0.010	0.024	0.017	0.029	**0.907**	0.012
JNJ	0.006	0.016	0.012	0.015	**0.944**	0.008
WMT	0.014	0.017	0.042	0.013	0.015	**0.899**
JPM	0.013	0.010	**0.918**	0.015	0.012	0.031
MA	0.137	0.164	0.110	**0.415**	0.092	0.081
PG	0.008	0.007	**0.944**	0.010	0.010	0.021
UNH	0.012	**0.902**	0.014	0.023	0.032	0.017
DIS	0.067	0.133	0.083	**0.533**	0.133	0.052
NVDA	0.044	0.039	0.040	**0.786**	0.070	0.022
HD	0.009	0.012	0.020	0.008	0.008	**0.942**
PYPL	0.185	**0.359**	0.089	0.162	0.082	0.123
BAC	0.074	0.084	0.257	0.077	0.061	**0.446**

an anomalous one if $u_{ic} \in [0.05, 0.35]$, $c = 1, \ldots, 6$. According to this criterion, no additional outliers were found by QCD-FCMn-E. Notice that PayPal (PYPL) could be seen as a potential outlier requiring individual analysis since it presented a membership degree slightly above 0.35 in cluster C_2. This is not surprising, since PayPal is also the MTS displaying the most spread out membership values in the partition reported by QCD-FCMn (Table 22).

The partition obtained with QCD-FCMn-NC is given in Table 15. Four outlying companies were identified, namely BRK.B, V, Johnson & Johnson (JNJ) and PYPL. As QCD-FCMn-NC is here handling a 7-cluster solution, we resolved to consider a series as anomalous when its maximum membership corresponded to the noise cluster and was above 0.25. BRK.B displayed the highest membership value in the noise cluster, followed by V and JNJ, and lastly by PYPL. It is worth remarking that three of the top five companies, AAPL, MSFT and GOOGL, exhibited membership values above 0.20 in the noise cluster, thus suggesting that these companies, specially AAPL (whose highest membership value is 0.267 in C_1) could be seen as plausible outliers. Compared to the partition provided by QCD-FCMn-E, the main difference is that the cluster C_4, formed by AAPL, MSFT, Tesla (TSLA), Mastercard (MA), Walt Disney (DIS) and Nvidia (NVDA), is split by QCD-FCMn-NC into two

Table 15
Membership degrees for top 20 companies in the S&P 500 index by considering the QCD-FCMn-NC.

Company	C_1	C_2	C_3	C_4	C_5	C_6	NC
AAPL	**0.267**	0.247	0.076	0.132	0.036	0.028	0.214
MSFT	0.171	**0.335**	0.079	0.036	0.130	0.034	0.215
AMZN	0.032	0.022	**0.822**	0.014	0.018	0.016	0.076
GOOGL	0.072	0.093	**0.516**	0.03	0.035	0.023	0.231
GOOG	0.021	0.019	**0.872**	0.008	0.016	0.010	0.054
FB	0.015	0.006	0.004	**0.941**	0.004	0.005	0.025
TSLA	0.127	**0.657**	0.050	0.030	0.022	0.012	0.102
BRK.B	0.054	0.055	0.036	0.033	0.127	0.053	**0.642**
V	0.176	0.167	0.038	0.127	0.080	0.046	**0.366**
JNJ	0.203	0.129	0.037	0.147	0.106	0.061	**0.317**
WMT	0.011	0.006	0.007	0.009	0.030	**0.890**	0.047
JPM	0.013	0.007	0.007	0.004	**0.932**	0.015	0.022
MA	**0.812**	0.050	0.027	0.029	0.024	0.015	0.043
PG	0.015	0.009	0.008	0.006	**0.900**	0.022	0.040
UNH	0.015	0.006	0.003	**0.942**	0.004	0.005	0.025
DIS	**0.936**	0.026	0.005	0.01	0.007	0.003	0.013
NVDA	0.030	**0.924**	0.007	0.006	0.007	0.003	0.023
HD	0.008	0.004	0.005	0.007	0.014	**0.935**	0.027
PYPL	0.184	0.078	0.124	0.210	0.053	0.073	**0.278**
BAC	0.136	0.045	0.055	0.058	0.253	**0.298**	0.155

clusters, C_1 (AAPL-MA-DIS) and C_2 (MSFT-TSLA-NVDA). In short, the fuzzy partition determined by QCD-FCMn-NC is also consistent with the 2D plot in Fig. 6, where BRK.B constitutes the most isolated point and V and JNJ are located at the top of the graph, moderately distant from the rest of the points.

Concerning QCD-FCMn-T, three series, BRK.B, WMT and The Home Depot (HD), were trimmed away (see Table 23 in Supplementary material) in accordance with the optimal trimmed ratio $\alpha = 0.15$. Note that the latter two companies were not determined as anomalous by QCD-FCMn-E and QCD-FCMn-NC. However, Fig. 6 suggests that WMT and HD could exhibit a distinct behaviour than that of the rest of companies, as the corresponding points are located in the lower part of the graph, far from the bulk.

In summary, each of the approaches QCD-FCMn-E, QCD-FCMn-NC and QCD-FCMn-T provides a different conclusion in terms of identification of outlying series, but all of them are consistent with the plot in Fig. 6. Moreover, there was consensus on establishing BRK.B as an outlier, which is quite reasonable from a financial point of view. BRK.B is the only company among the top 20 whose main task consists of investing in the remaining companies. Consequently, it is unsurprising that this firm exhibits a particular behaviour in terms of returns and trading volume.

7.2. Robust fuzzy clustering of air pollution data

The second study case concerns clustering of geographical zones in terms of their temporal records of air pollutants. 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 located in Galicia, an autonomous community of Spain. The choice of these pollutants was mainly based on the fact that several studies have uncovered serious health effects associated with the continuous exposure to high levels of NO_2 and O_3. The data were sourced from the website of Ministry for the Ecological Transition and the Demographic challenge https://www.miteco.gob.es/, which also contains information about the type of location, namely "urban", "suburban", "rural" and "near power plant". Thus, from an environmental point of view, it is reasonable to think that the joint behaviour of the concentration of these pollutants is different depending on where the station is situated.

The 20 MTS available are formed by $T = 8760$ hourly records and, given their nonstationarity in mean, we proceeded to transform them by taking first differences. The new set of series, which was subject to clustering in [4], is depicted in Fig. 19 in Supplementary material. The resulting 2DS plane based on the QCD distance is given in Fig. 7. In this case study, the empirical rule presented in Section 3.2 led to the selection of three principal components to construct d_{QCD}.

Proceeding as in Section 7.1, the optimal values for C and m found in [4] were $C = 3$ and $m = 1.9$. The 3-cluster solution attained by QCD-FCMn by considering 0.6 as cutoff (Table 24 in Supplementary material) identified groups of stations sharing the same location category, as it was expected from Fig. 7. Specifically, a cluster (C_1) grouped all the urban stations except for VGO-CT, other cluster (C_2) was formed by VGO-CT, one suburban (FE) and two rural (SU and PO-CP) stations, and the cluster C_3 involved three stations located near a power plant. The remaining 5 stations (SDC-C, PT, XO, MA and MO) displayed a relative high membership value in two clusters. These results reveal that the standard fuzzy model QCD-FCMn produced meaningful groups from a point of view of geographical location, but it was not able to identify anomalous series since no station displays membership degrees uniformly distributed among the three clusters. However, the 2D plot in Fig. 7 clearly indicates the existence of some MTS whose behaviour deviates significantly from the majority, as LO, FR

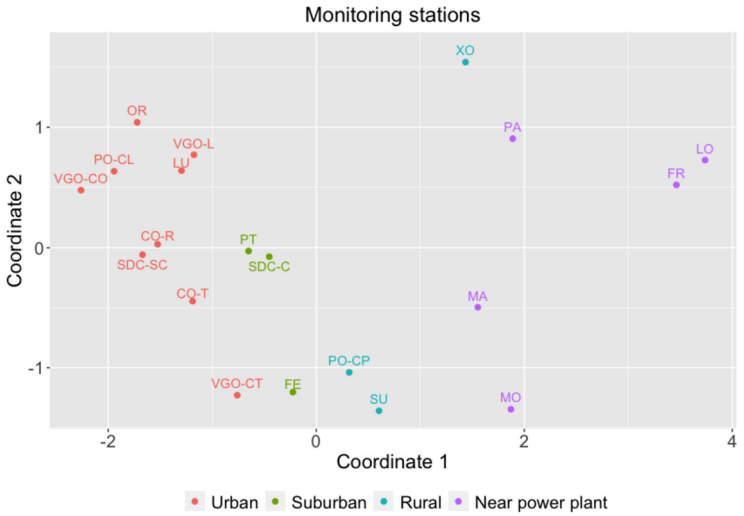

Fig. 7. 2D scaling plane based on the QCD distances for the hourly levels of NO_2, O_3 and NO in the 20 monitoring stations of Galicia.

Table 16
Summary of relevant dimensions involved in the environmental application.

T	Series	$D(\boldsymbol{X}_t^{(i)})$	$l(\boldsymbol{\Psi}^{(i)})$ (pre PCA)	$l(\boldsymbol{\Psi}^{(i)})$ (post PCA)
8760	20	26280	1418958	3

Table 17
Membership degrees for the 20 monitoring stations in Galicia by considering QCD-FCMn-E, QCD-FCMn-NC and QCD-FCMn-T models.

Station	QCD-FCMn-E			QCD-FCMn-NC				QCD-FCMn-T		
	C_1	C_2	C_3	C_1	C_2	C_3	NC	C_1	C_2	C_3
FE	0.114	0.046	**0.840**	0.043	0.022	**0.813**	0.122	0.064	0.042	**0.894**
CO-T	**0.656**	0.065	0.279	**0.357**	0.040	0.274	0.329	**0.604**	0.071	0.325
CO-R	**0.745**	0.071	0.184	**0.431**	0.040	0.145	0.384	**0.748**	0.068	0.184
LU	**0.945**	0.014	0.041	**0.893**	0.009	0.030	0.068	**0.931**	0.022	0.047
SDC-C	0.431	0.057	**0.513**	0.250	0.048	**0.493**	0.209	0.340	0.101	**0.559**
SDC-SC	**0.900**	0.024	0.076	**0.684**	0.020	0.098	0.198	**0.899**	0.023	0.078
SU	0.107	0.100	**0.793**	0.059	0.077	**0.554**	0.310	0.089	0.144	**0.767**
PO-CL	**0.947**	0.016	0.037	**0.927**	0.005	0.015	0.053	**0.961**	0.012	0.027
VGO-CO	**0.906**	0.029	0.065	**0.779**	0.013	0.042	0.166	**0.932**	0.021	0.047
VGO-L	**0.887**	0.030	0.083	**0.783**	0.020	0.059	0.138	**0.866**	0.047	0.087
PT	**0.454**	0.094	0.452	0.273	0.060	0.297	**0.370**	0.387	0.164	**0.449**
OR	**0.923**	0.025	0.052	**0.871**	0.010	0.025	0.094	**0.937**	0.023	0.040
PO-CP	0.030	0.020	**0.950**	0.039	0.036	**0.778**	0.147	0.037	0.063	**0.900**
FR	0.041	**0.900**	0.059	0.027	0.449	0.049	**0.475**	-	-	-
XO	0.213	**0.514**	0.273	0.071	0.349	0.087	**0.493**	0.091	**0.792**	0.11
VGO-CT	0.226	0.063	**0.711**	0.108	0.031	**0.610**	0.251	0.148	0.065	**0.787**
PA	0.056	**0.847**	0.097	0.006	**0.934**	0.011	0.049	0.029	**0.918**	0.053
MA	0.109	0.383	**0.508**	0.051	**0.381**	0.228	0.340	0.052	**0.711**	0.237
LO	0.055	**0.869**	0.076	0.027	0.387	0.047	**0.539**	-	-	-
MO	0.155	0.321	**0.524**	0.053	0.176	0.204	**0.567**	0.100	**0.500**	0.400

or XO. Thus, it should be desirable to use a robust method capable of simultaneously preserving the true cluster structure while identifying outlier series.

A summary of the relevant dimensions involved in the environmental application is presented in Table 16.

The membership matrices associated with the proposed robust models, QCD-FCMn-E, QCD-FCMn-NC and QCD-FCMn-T, are jointly given in Table 17.

QCD-FCMn-E produced a very similar partition to the one given by the standard algorithm, although all the series have membership degrees slightly more spread out. As we are dealing with 3 clusters, the i-th station was deemed anomalous if $u_{ic} > 0.2$, $c = 1, 2, 3$, which implies a reasonable equidistribution of its membership degrees among the three clusters, and hence its consideration as an outlier. In accordance with this benchmark, only XO was identified as outlier, which corresponds to the highest point in Fig. 7.

The models QCD-FCMn-NC and QCD-FCMn-T led to different results. The former identified 4 outlying stations, FR, XO, LO and MO, according to the assignment rule of locating a station in the noise cluster if the corresponding membership is above 0.4. These locations correspond to a rural station and three stations located near a power plant. Cluster C_1 showed the same composition as in the solutions given by QCD-FCMn and QCD-FCMn-E, and C_3 was also similar except for PT and MO (now outliers). There is also a new cluster C_2 formed by MA and PA, two stations situated near a power plant. Finally, the partition provided by QCD-FCMn-T shares traits with those of both QCD-FCMn and QCD-FCMn-E but is slightly different. The cluster C_1 is the standard urban cluster, but clusters C_2 and C_3 are constituted differently from clusters C_2 and C_3 in the mentioned models. The optimal α was determined to be $\alpha = 0.1$ so that two MTS corresponding to the most peripheral points in Fig. 7, FR and LO, were trimmed away.

8. Concluding remarks

In this work we have analyzed three robust approaches for MTS clustering by extending the QCD-based fuzzy C-means clustering model (QCD-FCMn) defined in [4]. All of them are aimed at counterbalancing the negative effects that outlying series (i.e., series generated by a different stochastic process than those of the regular clusters) provoke in the clustering solution. The considered strategies are based on three robust clustering methodologies suggested in the literature, namely the metric, noise and trimmed approaches (see [36], [37] and [38], respectively). The metric approach (QCD-FCMn-E) acquires its robustness with respect to outliers by considering a exponential distance measure. The noise approach (QCD-FCMn-NC) obtains its robustness against outlying series by introducing an artificial cluster represented by a noise prototype. The trimmed approach (QCD-FCMn-T) removes a certain proportion of anomalous time series data. Fundamentally, we intended to take advantage of three pivotal elements:

- The versatility of the fuzzy logic by allowing overlapping clusters. This permits to appropriately handling situations with inherent uncertainty, as it is often the case in MTS clustering.
- The powerful properties of the QCD-based distance to discriminate between general dependence structures (see [4]). By construction, this metric inherits the nice properties of the quantiles, hence exhibiting robustness against outliers and heavy tails.
- The high capability of the metric, noise and trimmed approaches to neutralize the disruptive impact of atypical series.

A broad simulation study involving different types of generating processes, namely linear, nonlinear and BEKK models, was carried out to evaluate the performance of the analyzed procedures. In order to make the assessment task fairly general, two types of anomalous series were considered: (i) series entirely generated from a different process and (ii) series suffering an abrupt change at a particular time point. For comparison purposes, two alternative dissimilarity measures were also taken into account. Overall, the suggested techniques outperformed the remaining metrics in terms of classification accuracy for the two classes of outlying series. Furthermore, they were the least sensitive to the selection of the corresponding hyperparameters. None of the three robust approaches proved to be better than the remaining ones in all the considered scenarios. Thus, each one of the techniques is useful in its own right. The three of them outperformed the nonrobust version QCD-FCMn in terms of outlier neutralization by a large degree, which highlights the usefulness of the robust algorithms. For illustrative purposes, we applied the methods to two MTS datasets containing financial and environmental series. Our analyses showed that the robust techniques were able to identify some series showing anomalous dynamic patterns, leading to interesting conclusions.

Although we have presented three accurate methods for robust fuzzy clustering of MTS based on generating processes, there is still room for further research in this topic. First, introducing suitable ways for automatically selecting the input parameters of these techniques would be highly desirable, as the performance of the methods is substantially dependent on an appropriate choice. Second, some extensions of the introduced strategies regarding a combination of the QCD-based metric and a shape-based dissimilarity could be appropriate to perform robust clustering in some specific scenarios. For instance, by dealing with nonstationary series or datasets in which the different patterns are characterized by both generating processes and geometric profiles. Third, note that, in our approach, each MTS is characterized by a set of curves of the form $\left\{W\left(\hat{G}_{T,R}^{j_1,j_2}(\omega,\tau,\tau')\right), 1 \leq j_1, j_2 \leq d, \tau, \tau' \in \mathcal{T}\right\}$, 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 about the dependence patterns. Thus, it would be reasonable to create a robust fuzzy clustering algorithm giving more importance to the functions with more discriminatory power. This could be naturally accomplished by introducing weights in the objective function (10). Finally, spatial extensions and possibilistic versions of the robust models introduced here could be constructed. The former techniques could be useful when dealing with series containing geographical information, as the ones in Section 7.2. The mentioned topics for further research will be properly addressed in the future.

Declaration of competing interest

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

Acknowledgements

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

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijar.2022.07.010.

References

[1] B. Lafuente-Rego, P. D'Urso, J. Vilar, Robust fuzzy clustering based on quantile autocovariances, Stat. Pap. (2018) 1–56.
[2] A. Blázquez-García, A. Conde, U. Mori, J.A. Lozano, A review on outlier/anomaly detection in time series data, arXiv preprint, arXiv:2002.04236, 2020.
[3] Á. López-Oriona, J.A. Vilar, Outlier detection for multivariate time series: a functional data approach, Knowl.-Based Syst. 233 (2021) 107527.
[4] Á. López-Oriona, J. Vilar, P. D'Urso, et al., Quantile-based fuzzy clustering of multivariate time series in the frequency domain, Fuzzy Sets Syst. 443 (B) (2022) 115–154.
[5] B. Lafuente-Rego, J.A. Vilar, Clustering of time series using quantile autocovariances, Adv. Data Anal. Classif. 10 (3) (2016) 391–415.
[6] P. D'Urso, E.A. Maharaj, Autocorrelation-based fuzzy clustering of time series, Fuzzy Sets Syst. 160 (24) (2009) 3565–3589.
[7] P. D'Urso, E.A. Maharaj, Wavelets-based clustering of multivariate time series, Fuzzy Sets Syst. 193 (2012) 33–61.
[8] P. D'Urso, C. Cappelli, D. Di Lallo, R. Massari, Clustering of financial time series, Physica A 392 (9) (2013) 2114–2129.
[9] P. D'Urso, R. Massari, C. Cappelli, L. De Giovanni, Autoregressive metric-based trimmed fuzzy clustering with an application to pm10 time series, Chemom. Intell. Lab. Syst. 161 (2017) 15–26.
[10] P. D'Urso, L. De Giovanni, R. Massari, Time series clustering by a robust autoregressive metric with application to air pollution, Chemom. Intell. Lab. Syst. 141 (2015) 107–124.
[11] P. D'Urso, L. De Giovanni, Robust clustering of imprecise data, Chemom. Intell. Lab. Syst. 136 (2014) 58–80.
[12] E.A. Maharaj, P. D'Urso, Fuzzy clustering of time series in the frequency domain, Inf. Sci. 181 (7) (2011) 1187–1211.
[13] D. Piccolo, A distance measure for classifying arima models, J. Time Ser. Anal. 11 (2) (1990) 153–164.
[14] P. D'Urso, L. De Giovanni, R. Massari, Garch-based robust clustering of time series, Fuzzy Sets Syst. 305 (2016) 1–28.
[15] A.M. Alonso, D. Peña, Clustering time series by linear dependency, Stat. Comput. 29 (4) (2019) 655–676.
[16] E.A. Maharaj, A significance test for classifying arma models, J. Stat. Comput. Simul. 54 (4) (1996) 305–331.
[17] 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.
[18] 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.
[19] J. Caiado, N. Crato, D. Peña, A periodogram-based metric for time series classification, Comput. Stat. Data Anal. 50 (10) (2006) 2668–2684.
[20] G. De Luca, P. Zuccolotto, Hierarchical time series clustering on tail dependence with linkage based on a multivariate copula approach, Int. J. Approx. Reason. 139 (2021) 88–103.
[21] M. La Rocca, F. Giordano, C. Perna, Clustering nonlinear time series with neural network bootstrap forecast distributions, Int. J. Approx. Reason. 137 (2021) 1–15.
[22] R. Cerqueti, M. Giacalone, R. Mattera, Model-based fuzzy time series clustering of conditional higher moments, Int. J. Approx. Reason. 134 (2021) 34–52.
[23] Y. Kakizawa, R.H. Shumway, M. Taniguchi, Discrimination and clustering for multivariate time series, J. Am. Stat. Assoc. 93 (441) (1998) 328–340.
[24] P. D'Urso, L. De Giovanni, E.A. Maharaj, R. Massari, Wavelet-based self-organizing maps for classifying multivariate time series, J. Chemom. 28 (1) (2014) 28–51.
[25] 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.
[26] 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.
[27] E.A. Maharaj, Comparison and classification of stationary multivariate time series, Pattern Recognit. 32 (7) (1999) 1129–1138.
[28] P. D'Urso, Fuzzy clustering for data time arrays with inlier and outlier time trajectories, IEEE Trans. Fuzzy Syst. 13 (5) (2005) 583–604.
[29] P. D'Urso, L. De Giovanni, R. Massari, Robust fuzzy clustering of multivariate time trajectories, Int. J. Approx. Reason. 99 (2018) 12–38.
[30] S. Aghabozorgi, A.S. Shirkhorshidi, T.Y. Wah, Time-series clustering–a decade review, Inf. Sci. 53 (2015) 16–38.
[31] E.A. Maharaj, P. D'Urso, J. Caiado, Time Series Clustering and Classification, CRC Press, 2019.
[32] P. D'Urso, L. De Giovanni, R. Massari, D. Di Lallo, Noise fuzzy clustering of time series by autoregressive metric, Metron 71 (3) (2013) 217–243.
[33] D. Rivera-García, L.A. García-Escudero, A. Mayo-Iscar, J. Ortega, Robust clustering for time series using spectral densities and functional data analysis, in: International Work-Conference on Artificial Neural Networks, Springer, 2017, pp. 142–153.
[34] 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.
[35] P. D'Urso, L.A. García-Escudero, L. De Giovanni, V. Vitale, A. Mayo-Iscar, Robust fuzzy clustering of time series based on b-splines, Int. J. Approx. Reason. 136 (2021) 223–246.
[36] K.-L. Wu, M.-S. Yang, Alternative c-means clustering algorithms, Pattern Recognit. 35 (10) (2002) 2267–2278.
[37] R.N. Dave, Characterization and detection of noise in clustering, Pattern Recognit. Lett. 12 (11) (1991) 657–664.
[38] R. Krishnapuram, A. Joshi, O. Nasraoui, L. Yi, Low-complexity fuzzy relational clustering algorithms for web mining, IEEE Trans. Fuzzy Syst. 9 (4) (2001) 595–607.
[39] J. Lee, S.S. Rao, The quantile spectral density and comparison based tests for nonlinear time series, arXiv:1112.2759, 2012.
[40] 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.
[41] J. Baruník, T. Kley, Quantile coherency: a general measure for dependence between cyclical economic variables, Econom. J. 22 (2) (2019) 131–152.
[42] T. Kley, S. Volgushev, H. Dette, M. Hallin, Quantile spectral processes: asymptotic analysis and inference, Bernoulli 22 (3) (2016) 1770–1807.
[43] Á. López-Oriona, J.A. Vilar, F4: an all-purpose tool for multivariate time series classification, Mathematics 9 (23) (2021) 3051.
[44] D.M. Witten, R. Tibshirani, A framework for feature selection in clustering, J. Am. Stat. Assoc. 105 (490) (2010) 713–726.
[45] S. Alelyani, J. Tang, H. Liu, Feature selection for clustering: a review, in: Data Clustering, 2018, pp. 29–60.

[46] M.G. Cimino, G. Frosini, B. Lazzerini, F. Marcelloni, On the noise distance in robust fuzzy c-means, in: International Conference on Computational Intelligence, 2004, pp. 361–364.
[47] R.N. Davé, S. Sen, Robust fuzzy clustering of relational data, IEEE Trans. Fuzzy Syst. 10 (6) (2002) 713–727.
[48] R.N. Davé, R. Krishnapuram, Robust clustering methods: a unified view, IEEE Trans. Fuzzy Syst. 5 (2) (1997) 270–293.
[49] R. Davé, S. Sen, Noise clustering algorithm revisited, in: 1997 Annual Meeting of the North American Fuzzy Information Processing Society-NAFIPS (Cat. No. 97TH8297), IEEE, 1997, pp. 199–204.
[50] 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.
[51] C.W. Granger, T. Teräsvirta, Modelling Non-linear Economic Relationships, Oxford University Press, 1993.
[52] C. Granger, A. Andersen, An Introduction to Bilinear Time Series Models, Angewandte Statistik und Ökonometrie, Vandenhoeck und Rupreckt, Gottingen, 1978.
[53] 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.
[54] E. Pereda, R.Q. Quiroga, J. Bhattacharya, Nonlinear multivariate analysis of neurophysiological signals, Prog. Neurobiol. 77 (1–2) (2005) 1–37.
[55] G. Koop, M.H. Pesaran, S.M. Potter, Impulse response analysis in nonlinear multivariate models, J. Econom. 74 (1) (1996) 119–147.
[56] R.F. Engle, K.F. Kroner, Multivariate simultaneous generalized arch, Econom. Theory (1995) 122–150.
[57] J. Chevallier, Time-varying correlations in oil, gas and co2 prices: an application using bekk, ccc and dcc-mgarch models, Appl. Econ. 44 (32) (2012) 4257–4274.
[58] S. Rahman, A. Serletis, Oil price uncertainty and the Canadian economy: evidence from a varma, garch-in-mean, asymmetric bekk model, Energy Econ. 34 (2) (2012) 603–610.
[59] H. Heidari, S.T. Katircioglu, S. Bashiri, Inflation, inflation uncertainty and growth in the Iranian economy: an application of bgarch-m model with bekk approach, J. Bus. Econ. Manag. 14 (5) (2013) 819–832.
[60] R. Engle, Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models, J. Bus. Econ. Stat. 20 (3) (2002) 339–350.
[61] M. Caporin, M. McAleer, Do we really need both bekk and dcc? A tale of two multivariate garch models, J. Econ. Surv. 26 (4) (2012) 736–751.
[62] J. Hair, R. Anderson, B. Babin, Multivariate Data Analysis, Prentice Hall, 2009.
[63] E.A. Maharaj, P. D'Urso, D.U. Galagedera, Wavelet-based fuzzy clustering of time series, J. Classif. 27 (2) (2010) 231–275.
[64] D. Dembele, P. Kastner, Fuzzy c-means method for clustering microarray data, Bioinformatics 19 (8) (2003) 973–980.
[65] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Springer Science & Business Media, 2013.
[66] 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.
[67] 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.
[68] Y. Wu, J.M. Hernández-Lobato, G. Zoubin, Dynamic covariance models for multivariate financial time series, in: International Conference on Machine Learning, PMLR, 2013, pp. 558–566.
[69] O. Efimova, A. Serletis, Energy markets volatility modelling using garch, Energy Econ. 43 (2014) 264–273.
[70] A.C. Harvey, Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series, vol. 52, Cambridge University Press, 2013.
[71] M. Bernardi, A. Maruotti, L. Petrella, Multiple risk measures for multivariate dynamic heavy–tailed models, J. Empir. Finance 43 (2017) 1–32.
[72] S.T. Rachev, Handbook of Heavy Tailed Distributions in Finance, Handbooks in Finance, Book 1, Elsevier, 2003.
[73] T. Mikosch, Modeling dependence and tails of financial time series, in: Extreme Values in Finance, Telecommunications, and the Environment, 2003, pp. 185–286.
[74] R.S. Tsay, D. Pena, A.E. Pankratz, Outliers in multivariate time series, Biometrika 87 (4) (2000) 789–804.
[75] P. Galeano, D. Peña, R.S. Tsay, Outlier detection in multivariate time series by projection pursuit, J. Am. Stat. Assoc. 101 (474) (2006) 654–669.
[76] J.M. Karpoff, The relation between price changes and trading volume: a survey, J. Financ. Quant. Anal. (1987) 109–126.
[77] J.Y. Campbell, S.J. Grossman, J. Wang, Trading volume and serial correlation in stock returns, Q. J. Econ. 108 (4) (1993) 905–939.
[78] B. Gebka, M.E. Wohar, Causality between trading volume and returns: evidence from quantile regressions, Int. Rev. Econ. Finance 27 (2013) 144–159.
[79] 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.
[80] B.O. Bradley, M.S. Taqqu, Financial risk and heavy tails, in: Handbook of Heavy Tailed Distributions in Finance, Elsevier, 2003, pp. 35–103.