Outlier detection for multivariate time series: A functional data approach ®

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

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

ARTICLE INFO

ABSTRACT

A method for detecting outlier samples in a multivariate time series dataset is proposed. It is assumed that an outlying series is characterized by having been generated from a different process than those associated with the rest of the series. Each multivariate time series is described by means of an estimator of its quantile cross-spectral density, which is treated as a multivariate functional datum. Then an outlier score is assigned to each series by using functional depths. A broad simulation study shows that the proposed approach is superior to the alternatives suggested in the literature and demonstrates that the consideration of functional data constitutes a critical step. The procedure runs in linear time with respect to both the series length and the number of series, and in quadratic time with respect to the number of dimensions. Two applications concerning financial series and ECG signals highlight the usefulness of the technique.

1. Introduction

Complexity, speed and volume of data is growing at an unprecedented pace. In particular, time series data have become ubiquitous in our days, arising frequently in a broad variety of fields including medicine, computer science, economics, finance, and environmental sciences, among many others. Although univariate time series (UTS) were the norm until recently, multivariate time series (MTS) have received lately a great deal of attention due to the advance of technology and storage capabilities of everyday machines. Examples of MTS are multi-lead ECG signals of patients or temporal records of several economic indicators for a specific country.

In particular, detection and analysis of anomalous behaviors in temporal data constitutes an important problem in time series data mining and has been addressed in different works in the literature. A current review presenting a structured and comprehensive state-of-the-art on outlier detection techniques in time series data can be seen in [1]. Several methods have been proposed either for UTS [2–11] or MTS [12–15]. These techniques focus on detecting time series showing an unusual behavior over a more or less long-term period or even in a specific time point. Hence, the outlier is characterized by an unexpected performance over a time window. Nevertheless, only a few works have addressed the detection of outlier time series objects, i.e., whole time series exhibiting an anomalous behavior compared with the majority of time series in a given dataset. Examples of works handling this problem are [16–20] for UTS and [21] for MTS.

The first contribution of this paper is to propose an algorithm aimed to detect outlier MTS samples. We consider this problem to be of paramount importance. For instance, companies are frequently interested in detecting unusual consumer behaviors based on temporal patterns. In the same way, practitioners aim to discover atypical ECG signals indicating that a given patient could be at risk. Consequently, the proposal of new approaches intended to successfully identify anomalous MTS in a given collection is highly desirable.

Several strategies have been proposed in the outlier detection literature for multidimensional time series. [22] generalized four common types of outliers in UTS to the MTS framework,

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

and developed an iterative method for outlier detection based on two different test statistics. [12] designed a procedure for anomaly identification considering projections of VARMA models. [13] constructed a method that uses independent component analysis to extract possible outlier patterns from an MTS. [14] proposed a robust algorithm for detecting anomalies in noisy MTS data. The approach uses a kernel matrix alignment method capturing the dependence relationships between variables.

All the previous techniques focus on detecting series exhibiting a distorted behavior during a period of time rather than identifying entire MTS objects as anomalous. In the present paper, we are interested in the latter topic, which is concerned with extracting whole outlying MTS from a given database. This perspective has been considered only in a few works. [23] presented a procedure for identifying outlying samples in an MTS dataset based on local sparsity coefficient [24]. In a latter work, [21] designed a methodology to extract the top outlying MTS by means of the Extended Frobenius norm (*Eros*). However, these works do not address anomaly detection directly from the point of view of the underlying stochastic processes. In other words, they do not consider that an MTS is atypical when it exhibits a different dependence structure, thus showing a dynamical pattern distinct from the rest of the series in the dataset. This is particularly remarkable, since a substantial part of the literature on time series is devoted to modeling and explaining the underlying dependence structures of these objects [25,26]. In this regard, the second contribution of this work is to address the outlier identification task from the mentioned viewpoint. It is worth emphasizing the high complexity of this problem when treating with multidimensional time series due to the interdependence relation between the univariate components.

Our approach for outlier detection relies on the notion of quantile cross-spectral density. This via has been successfully applied in clustering of MTS [27]. The excellent results reached in clustering, clearly outperforming other alternative criteria, motivated its consideration for detecting anomalous MTS. Furthermore, the proposed method takes advantage of the functional nature of the quantile cross-spectral density, thus allowing to use the concept of functional depth to assign an outlier score to each series. This way, the whole dataset is ranked according to the outlying likelihood of its elements.

The rest of the paper is organized as follows. Section 2 reviews some background knowledge on the quantile cross-spectral density and functional data depths. Section 3 describes the novel approach we propose to detect outlier samples in an MTS dataset. The main results from a wide simulation study aimed to assess the proposed procedure are shown in Section 4. Section 5 analyses the computation times of the method. The usefulness of the approach is illustrated in Section 6 by means of its application to real MTS datasets. Finally, the main conclusions are summarized in Section 7.

2. Background and related works

This section reviews some background knowledge on the quantile cross-spectral density and the concept of depth for functional data.

2.1. The quantile cross-spectral density

Following [27], 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 a lag $l \in \mathbb{Z}$ and an arbitrary couple of quantile levels $(\tau, \tau') \in [0, 1]^2$, consider the cross-covariance of the

indicator functions $I\left\{X_{t,j_1} \leq q_{j_1}(\tau)\right\}$ and $I\left\{X_{t+l,j_2} \leq q_{j_2}(\tau')\right\}$ given by

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

for $1 \leq j_1, j_2 \leq d$. Taking $j_1 = j_2 = j$, the function $\gamma_{j,j}(l, \tau, \tau')$, with $(\tau, \tau') \in [0, 1]^2$, so-called quantile autocovariance function of lag l, generalizes the traditional autocovariance function.

The $d \times d$ matrix given by

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

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

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

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

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

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

Proceeding as in (3), QCD can be evaluated on a range of frequencies Ω and of quantile levels $\mathcal{T}$ for every couple of components in order to obtain a complete representation of the process. In this way, we can consider the set of matrices

$$\mathfrak{f}_{\boldsymbol{X}_t}(\Omega, \mathcal{T}) = \left\{\mathfrak{f}(\omega, \tau, \tau'), \ \omega \in \Omega, \ \tau, \tau' \in \mathcal{T}\right\}, \quad (5)$$

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

$$\mathfrak{f}(\omega, \tau, \tau') = \left(\mathfrak{f}_{j_1,j_2}(\omega, \tau, \tau')\right)_{1 \leq j_1, j_2 \leq d}. \quad (6)$$

Representing $\{\boldsymbol{X}_t, \ t \in \mathbb{Z}\}$ through $\mathfrak{f}_{\boldsymbol{X}_t}$, complete information on the general dependence structure of the process is available. However, the true QCD is unknown in practice and a proper estimator is required given a realization of length T, $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T\}$, from the process $\{\boldsymbol{X}_t, \ t \in \mathbb{Z}\}$. Following [28], a consistent estimator of $\mathfrak{f}_{j_1,j_2}(\omega, \tau, \tau')$ can be obtained by smoothing a modified version of the traditional cross-periodogram based on ranks. Specifically, the called *rank-based copula cross-periodogram* (hereafter referred as CCR-periodogram) is defined by

$$I_{T,R}^{j_1,j_2}(\omega, \tau, \tau') = \frac{1}{2\pi T} d_{T,R}^{j_1}(\omega, \tau) d_{T,R}^{j_2}(-\omega, \tau'), \quad (7)$$

where

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

with $\hat{F}_{T,j}(x) = T^{-1}\sum_{t=1}^{T} I\{X_{t,j} \leq x\}$ the empirical distribution function of $X_{t,j}$. Consistency and asymptotic performance of the smoothed CCR-periodogram, $\hat{G}_{T,R}^{j_1,j_2}(\omega, \tau, \tau')$, are established

in Theorem S4.1 of [28]. A detailed description of the smoothed CCR-periodogram is deferred to the Appendix.

Based on previous comments, the set of complex-valued matrices $\mathfrak{f}_{\boldsymbol{X}_t}(\Omega, \mathcal{T})$ in (5) characterizing the underlying process can be estimated by

$$\hat{\mathfrak{f}}_{\boldsymbol{X}_t}(\Omega, \mathcal{T}) = \left\{ \hat{\mathfrak{f}}(\omega, \tau, \tau'), \ \omega \in \Omega, \ \tau, \tau' \in \mathcal{T} \right\}, \tag{8}$$

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

$$\hat{\mathfrak{f}}(\omega, \tau, \tau') = \left(\hat{G}_{T,R}^{j_1, j_2}(\omega, \tau, \tau') \right)_{1 \leq j_1, j_2 \leq d}. \tag{9}$$

2.2. The notion of depth for functional data

A functional variable is one whose values depend on a continuous magnitude such as time. Thus, a univariate functional dataset is a set of real-valued curves $\{\mathcal{X}_1(t), \ldots, \mathcal{X}_n(t)\}, t \in [a, b]$, whereas a multivariate functional dataset is a set of vector-valued curves $\{\boldsymbol{\mathcal{X}}_1(t), \ldots, \boldsymbol{\mathcal{X}}_n(t)\}, t \in [a, b]$. Without loss of generality, it is often assumed that $a = 0$ and $b = 1$. In practice, each curve is observed only in a finite set of points $t_0, \ldots, t_m \in [0, 1] = \mathcal{D}$.

Given a set of curves, it is often desirable to know which of them lay "in the middle of the set" for most of the time. This idea corresponds in a functional setting to the intuitive notion of median. In the same way, the curves laying "outward the set" for most of the time could be considered to exhibit an anomalous behavior. The concept of depth for functional data arises to address this problem by providing a centrality measure for the elements of the dataset. Although many depth notions for functional data have been established [29–32], here we focus our attention on the Fraiman–Muniz depth [30]. This depth frequently leads to the best performance of the approach proposed in this manuscript (see Section 4).

Let D be a depth measure in $\mathbb{R}$. The Fraiman–Muniz depth for a univariate functional datum $\mathcal{X}_i(t)$ concerning the set $\{\mathcal{X}_1(t), \ldots, \mathcal{X}_n(t)\}$ is defined as

$$FMD(\mathcal{X}_i) = \int_{\mathcal{D}} Z_i(t) dt, \tag{10}$$

where, for every $t_j \in \mathcal{D}$, $Z_i(t_j) = D(\mathcal{X}_i(t_j))$ is the univariate depth of the quantity $\mathcal{X}_i(t_j)$ with respect to the set $\{\mathcal{X}_k(t_j)\}_{k=1}^{n}$. Several notions of depth for univariate data have been also proposed. In this paper, we consider the Tukey depth, also referred to as halfspace depth. The Tukey depth of a point y with respect to a set of univariate points $\{y_1, \ldots, y_s\}$ is defined as

$$TD(y) = \frac{1}{s} \min\{\#\{y_i \leq y\}, \#\{y_i \geq y\}\}, \tag{11}$$

where # denotes the cardinal of a set. It is straightforward to see that the median is the point with maximal Tukey depth.

The notion of depth for multivariate functional data can be easily generalized from the concept of depth for univariate functional data just introduced. Let $\boldsymbol{\mathcal{X}}_i(t) = \{\mathcal{X}_{1,i}(t), \ldots, \mathcal{X}_{d,i}(t)\}$ be a d-variate functional datum and D_{UF} any depth measure for univariate functional data. The corresponding depth for the multivariate functional datum $\boldsymbol{\mathcal{X}}_i$ is given by

$$\boldsymbol{D}_{MF}(\boldsymbol{\mathcal{X}}_i) = \sum_{k=1}^{d} D_{UF}(\mathcal{X}_{k,i}). \tag{12}$$

2.3. The quantile cross-spectral density as multivariate functional data

Note that, for fixed j_1, j_2, τ and τ', the quantile cross-spectral density in (4) is a complex-valued function of the frequency ω. Consider now the set of complex-valued functions

$$\{\mathfrak{f}_{j_1, j_2}(\omega, \tau, \tau'), 1 \leq j_1, j_2 \leq d, \tau, \tau' \in \mathcal{T}\}. \tag{13}$$

Taking separately the real and imaginary parts of the functions in (13), we obtain the set of curves

$$\{W[\mathfrak{f}_{j_1, j_2}(\omega, \tau, \tau')], 1 \leq j_1, j_2 \leq d, \tau, \tau' \in \mathcal{T}\}, \tag{14}$$

where $W(\cdot)$ is used interchangeably to denote the real part ($\Re(\cdot)$) and imaginary part ($\Im(\cdot)$) operators. Hence, each multivariate process $\{\boldsymbol{X}_t, t \in \mathbb{Z}\}$ can be characterized by means of a set of $2d^2r^2$ real-valued functions, with $r = \#\mathcal{T}$ being the number of selected probability levels. In the same way, we can use the smoothed CCR-periodogram in (9) to describe each MTS through a $2d^2r^2$-dimensional functional datum.

The functional representation of QCD is useful for different reasons. First, it is well known that, in practice, considering a functional data representation is often advantageous in comparison with the use of large finite-dimensional vectors [33]. On the other hand, our goal is to rank the sample data to detect the most extreme elements, which are associated with the lowest depths. However, the effective calculus of multivariate depths in high dimensional spaces is extremely complex for computational reasons [30].

3. Algorithms

In this section, the proposed approach for outlier detection is described and structured throughout two algorithms. The first one generates a multivariate functional dataset from an MTS dataset. The second algorithm extracts the top outlier samples from the functional dataset.

3.1. From time series to functional data

The first stage in our method involves obtaining a collection of multivariate functional curves from the original MTS dataset. Note that each MTS is stored in a matrix of the form: number of time observations × number of dimensions. For each time series, the smoothed CCR-periodogram is computed according to all pairs of variables and quantile levels as a function of the Fourier frequencies. Two real curves are extracted from the smoothed CCR-periodograms, one corresponding to the real part and the other to the imaginary part, as indicated in Section 2.3. Then, all the curves associated with a given MTS are stored together and transformed into functional data. Algorithm 1, entitled *From time series to functional data* (FTSTFD), describes the steps.

In Algorithm 1, the ith MTS is described by means of the matrix **AL**[[i]] before the consideration of functional data. The first half of the rows in this matrix are associated with real parts of the smoothed CCR-periodograms, whereas the second half of the rows are associated with imaginary parts. Specifically, the auxiliary matrices **ALR**[[i]] (real parts) and **ALI**[[i]] (imaginary parts) are stacked together by rows to constitute **AL**[[i]]. Each row vector in **AL**[[i]] can be seen as a discretization of a real curve in the interval $(0, 2\pi)$. Note that the length of each row in **AL**[[i]] is the same as the length of the vector Ω, as the smoothed CCR-periodograms are computed at the Fourier frequencies. The function fdata() performs a row-wise transformation of **AL**[[i]] into a functional space, returning the multivariate functional element **DSMF**[[i]]. The dimension of this functional object (i.e., its number of curves) is $2 \times \text{cols}^2 \times \text{levels}^2$ as indicated in Section 2.3. Note that the factor 2 is due to the fact that real and imaginary parts are treated independently.

Each CCR-periodogram is obtained by means of the Fast Fourier Transform (FFT). The FFT of a series of length T has a time complexity of $O(T \log T)$ [34], and the algorithm performs a number of $r^2 d^2$ FFT for each MTS in the initial set. Therefore, it is concluded that the complexity of Algorithm 1 is $O(nr^2d^2T \log T)$.

Algorithm 1. [FTSTFD(**DSTS**, $\mathcal{T}$)]

Input: An MTS list with n samples of length T, called *Dataset of time series* (**DSTS**), a set $\mathcal{T}$ of r probability levels.

Output: A list of n multivariate functional data, called *Dataset of multivariate functional* (**DSMF**). Each element in this set is formed of $2d^2r^2$ curves.

1:    $n \leftarrow$ length(**DSTS**)

2:    $cols \leftarrow$ number of columns of **DSTS**[[1]]

3:    $levels \leftarrow$ cardinality of the set $\mathcal{T}$, r

4:    $T \leftarrow$ number of rows of **DSTS**[[1]]

5:    $\Omega \leftarrow$ the vector of Fourier frequencies (function of T)

6:    Initialize the array of smoothed CCR-periodograms, called *Smoothed CCR-periodograms* (**SCCRP**)

7:    Initialize the array with the real parts of the smoothed CCR-periodograms, called *Real smoothed CCR-periodograms* (**RSCCRP**)

8:    Initialize the array with the imaginary parts of the smoothed CCR-periodograms, called *Imaginary smoothed CCR-periodograms* (**ISCCRP**)

9:    Initialize an auxiliary list of n matrices for the real parts of the smoothed CCR-periodograms, called *Auxiliary list real* (**ALR**)

10:   Initialize an auxiliary list of n matrices for the imaginary parts of the smoothed CCR-periodograms, called *Auxiliary list imaginary* (**ALI**)

11:   Initialize an auxiliary list of n matrices called *Auxiliary list* (**AL**)

12:   Initialize the list **DSMF**

13:   **for** (i in $1:n$)

     for (j_1 in $1:cols$)

        for (j_2 in $1:cols$)

           for (k_1 in $1:levels$)

              for (k_2 in $1:levels$)

                 SCCRP$[i,j_1,j_2,k1,k2,:] \leftarrow$ Compute $\{\hat{G}_{T,R}^{j_1,j_2}(\omega,k_1,k_2), \omega \in \Omega\}$, i.e., evaluated at the Fourier frequencies

                 RSCCRP$[i,j_1,j_2,k1,k2,:] \leftarrow \{\Re(\hat{G}_{T,R}^{j_1,j_2}(\omega,k_1,k_2)), \omega \in \Omega\}$

                 ISCCRP$[i,j_1,j_2,k1,k2,:] \leftarrow \{\Im(\hat{G}_{T,R}^{j_1,j_2}(\omega,k_1,k_2)), \omega \in \Omega\}$

              end for

           end for

        end for

     end for

     ALR$[[i]] \leftarrow$ matrix whose rows are formed by the vectors **RSCCRP**$[i,j_1,j_2,k1,k2,:]$

     ALI$[[i]] \leftarrow$ matrix whose rows are formed by the vectors **ISCCRP**$[i,j_1,j_2,k1,k2,:]$

     AL$[[i]] \leftarrow$ rbind(**ALR**$[[i]]$, **ALI**$[[i]]$)

     DSMF$[[i]] \leftarrow$ fdata(**AL**$[[i]]$)

   end for

14: Return the list of multivariate functional data, **DSMF**

3.2. Outlier detection

The second algorithm corresponds to the outlier detection stage. Given a multivariate functional dataset and a univariate functional depth, D_{UF}, the multivariate depth of each element in the set is computed and stored in a vector. The quantities of this vector are sorted in increasing order and their corresponding indexes are stored. This way, the original samples are ranked from the most outlying series to the less outlying one. A rate α has to be set as to decide which is the desired proportion of outliers to detect. Algorithm 2, entitled *Outlier detection* (OD), gives a description of the steps.

Algorithm 2. [OD(**DSMF**, α, D_{UF})]

Input: A list of n multivariate functional data, called *Dataset of multivariate functional* (**DSMF**), the desired proportion of outliers, α, a depth for univariate functional data, D_{UF}.

Output: The top $\lceil \alpha n \rceil$ outlier samples in **DSMF**, where $\lceil \cdot \rceil$ denotes the ceiling function.

1:   Initialize the vector of depths, **mdepth**

2:   For each object **DSMF**[[i]], obtain the multivariate functional depth according to (12), **mdepth**[i]

3:   **mdepth_order** $\leftarrow$ Order(**mdepth**, decreasing = FALSE)

4:   n_outliers $\leftarrow$ Ceiling($n\alpha$)

5:   **outliers** $\leftarrow$ **mdepth_order**[1 : n_outliers]

6:   Return the top $\lceil \alpha n \rceil$ outlying samples in **DSMF**, **outliers**

In Algorithm 2, the vector **mdepth** contains the depth of each element in the multivariate functional collection **DSMF**. The indexes associated with the lowest values in this vector are returned in the vector **outliers**.

The time complexity of Algorithm 2 depends on the selected functional depths. In our numerical experiments, we have considered the Fraiman–Muniz and the Tukey depths (see Section 4). The input of Algorithm 2 is a multivariate functional dataset with n elements, each one containing $2d^2r^2$ curves evaluated at the Fourier frequencies. Let ω be one of these frequencies and fix one of the $2d^2r^2$ curves. According to (10), we have to compute the univariate Tukey depths (i.e., the elements Z_i) for the corresponding n values. Note that, by virtue of (11), computing the depth of n points with respect to themselves can be done by sorting these points. The best algorithm for sorting a set of points is **Quicksort** [35], having an average case complexity of $O(p \log p)$ for a set of p points. Thus, we assume a complexity of $O(n \log n)$ for this stage. As we need to repeat the process for all the $\lfloor T/2 \rfloor + 1$ Fourier frequencies ($\lfloor \cdot \rfloor$ denoting the floor function), this yields a complexity of $O(Tn \log n)$. This is the complexity to compute the Tukey depths for one of the curves. Replicating the methodology for all the curves gives a complexity of $O(Tnd^2r^2 \log n)$ (Step 1). Once this is done, the integral in (10) needs to be computed for each one of the curves. Note that, at this point, the curves are described by means of $\lfloor T/2 \rfloor + 1$ values (univariate Tukey depths) calculated in Step 1. By using Simpson's rule [36], the complexity of each integral is $O(T)$, thus being the complexity of the integration step $O(Tnd^2r^2)$ (Step 2). Finally, once we have the final depths for each multivariate functional element, sorting it through **Quicksort** requires $O(n \log n)$ time (Step 3).

As the time complexity of Steps 2 and 3 is lower than that of Step 1, the total computational complexity of Algorithm 2 is $O(Tnd^2r^2 \log n)$.

As the number r of probability levels is fixed and small, we can conclude that the complexities of Algorithms 1 and 2 are $O(Tnd^2 \log T)$ and $O(Tnd^2 \log n)$, respectively. Thus, the complexity of the proposed outlier detection method is $O(Tnd^2(\log T + \log n))$. Note that the difference between T, n and $\log T$, $\log n$, respectively, becomes larger as T and n increase. Therefore, for moderately large values of n and T (which is often the case in practice), the approach follows approximately linear time in T and n, and quadratic time in d.

Fig. 1 displays a flowchart of the outlier detection procedure.

4. Assessing the outlier detection approach: A simulation study

In this section, we carry out a set of simulations with the aim of assessing the performance of the outlier detection procedure.
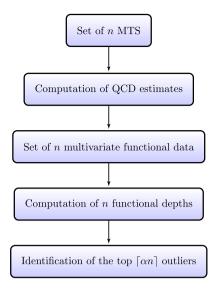
Fig. 1. Flowchart of the outlier detection method.

Different scenarios are considered. Firstly we describe the simulation mechanism, and then we explain how the assessment of the proposed approach was done. Finally, we show the results of the simulation study.

4.1. Experimental design

The simulated scenarios cover a wide variety of generating processes. Specifically, three bivariate setups were taken into account, namely outlier detection of (1) VARMA processes, (2) nonlinear processes, and (3) dynamic conditional correlation processes. The selection of such kind of processes was made with the goal of performing the assessment task in a fair and general manner. Indeed, the three chosen setups are pivotal in several application domains.

Each of the considered settings is formed by twenty realizations from a base generating process plus one or two outliers series drawn out from different generating models. The considered scenarios and the specific generation schemes are given below.

Outlier detection of linear models

BASE GENERATING PROCESS: A VAR(1) process given by

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

SCENARIO 1.1: Twenty series simulated from the base generating process plus one outlier time series simulated from the VAR(1) process

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

SCENARIO 1.2: The twenty one series in Scenario 1.1 plus a second outlier generated from a bivariate white noise process (WN).

Outlier detection of nonlinear models

BASE GENERATING PROCESS: A nonlinear autoregressive (NAR) process given by

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

SCENARIO 2.1: Twenty series simulated from the base generating process plus one outlier time series simulated from the exponential autoregressive (EXPAR) process

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

SCENARIO 2.2: The twenty one series in Scenario 2.1 plus a second outlier consisting of a realization from the threshold autoregressive (TAR) process

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

where I stands for the indicator function.

Outlier detection of dynamic conditional correlation models

BASE GENERATING PROCESS: Consider

$$(X_{t,1}, X_{t,2})^\intercal = (a_{t,1}, a_{t,2})^\intercal = (\sigma_{t,1}\epsilon_{t,1}, \sigma_{t,2}\epsilon_{t,2})^\intercal,$$

denoting $\intercal$ the transpose operator. The data-generating process consists of two Gaussian GARCH models [37], one which is highly persistent and the other which is not.

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

The correlation between the standardized shocks, ρ_t, is given by 0.5.

SCENARIO 3.1: Twenty series simulated from the base generating process plus one outlier time series simulated from an analogous process with $\rho_t = 0.2$.

SCENARIO 3.2: The twenty one series in Scenario 3.1 plus a second outlier which is given by an analogous process with $\rho_t = \frac{0.99}{\log(t+2)}I_{\{t \text{ odd}\}} - \frac{0.99}{\log(t+2)}I_{\{t \text{ even}\}}$.

In Scenarios 1.1, 1.2, 2.1 and 2.2, the vector error process $\{(\epsilon_{t,1}, \epsilon_{t,2})^\intercal, t \in \mathbb{Z}\}$ consists of iid vectors following a standard bivariate Gaussian distribution.

Scenarios 1.1 and 1.2 deal with classical VARMA models, which are broadly used in many fields. Note that the choice of the coefficient matrices is driven by the requirements of stationarity. Scenarios 2.1 and 2.2 consist of multivariate extensions of univariate NAR and TAR processes proposed in [38] and the univariate EXPAR process given in [39]. Nonlinear UTS arise in several application fields [40–42]. Hence, it is natural to consider their multivariate extensions. Scenarios 3.1 and 3.2 are motivated by the landmark work [43], where the dynamic conditional correlation models are introduced. Some of the processes in these scenarios have been partially considered in [27].

To bring insight into the usefulness of the proposed outlier detection approach, we generated large sample size realizations ($T = 2000$) from two processes concerning the linear models setting. Specifically, we selected the VAR(1) model in the base generating process and the WN process. Then, we simulated 50 MTS from each process. For each one of the total 100 realizations, we obtained the imaginary part of the smoothed CCR-periodogram for $j_1 = 1, j_2 = 2, \tau = 0.1$ and $\tau' = 0.5$, evaluated over the set of Fourier frequencies, Ω. The corresponding collection, $\{\Im(\hat{G}_{T,R}^{1,2}(\omega, 0.1, 0.5)), \omega \in \Omega\}$ was transformed into a function. Note that the considered curve is one of the components of the multivariate functional datum used for describing each MTS as explained in Section 2.3.

Plots of the 100 curves are given in Fig. 2. A different color was used according to the underlying generating process. The curves
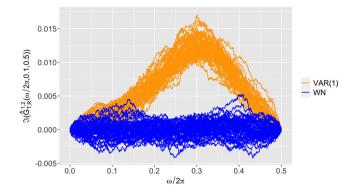
Fig. 2. Imaginary part of smoothed CCR-periodogram for $j_1 = 1$, $j_2 = 2$, $\tau = 0.1$, and $\tau' = 0.5$, regarding 50 large sample size ($T = 2000$) realizations from each one of the processes VAR(1) (orange color) and white noise (blue color).
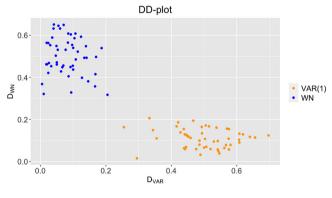

Fig. 3. DD-plot of the curves in Fig. 2.

from the VAR(1) model (orange color) are clearly different from the ones corresponding to the WN process (blue color). Indeed, the latter functions oscillate around zero, as the QCD of a WN process is zero across all frequencies. The differences between both groups of curves can be explained by means of functional depths. For instance, the average Fraiman–Muniz depth of the group of WN curves with respect to the group of VAR(1) curves is 0.0822 (standard deviation of 0.0497). Conversely, the average depth of the group of VAR(1) functions with regard to itself is 0.4997 (standard deviation of 0.0907). Both quantities clearly differ and give understanding about the fact that a WN series could be seen as an anomalous element concerning the group of VAR(1) series.

Another interesting graphical tool to visualize the suitability of functional depths to differentiate between underlying structures is the DD-plot [44]. In the case of the previous example, the DD-plot displays, for each functional datum $\mathcal{X}$, the pair of points

$$\left(D_{VAR}(\mathcal{X}), D_{WN}(\mathcal{X}) \right), \tag{15}$$

where $D_{VAR}(\mathcal{X})$ and $D_{WN}(\mathcal{X})$ are the depths of $\mathcal{X}$ with respect to the data in the VAR(1) and WN groups, respectively. Fig. 3 shows the DD-plot, where we have used the same color as in Fig. 2 to represent the processes. Each group of points is located at a different region of the plane. The depths of the VAR(1) (WN) series with regards to the WN (VAR(1)) group are smaller, whereas the depths of the WN (VAR(1)) series in relation to its group are larger. Thus, in this context, the use of depths seems convenient to differentiate between the generating processes of a group of MTS.

The previous toy example highlights the usefulness of the proposed approach for outlier detection of MTS samples based on QCD and functional data depths.

The simulation study was carried out as follows. For each one of the six scenarios, 200 simulations were performed for different values of the series length. We considered $T \in \{200, 400, 600\}$ in Scenarios 1.1, 1.2, 2.1 and 2.2, and $T \in \{400, 800, 1200\}$ in Scenarios 3.1 and 3.2.

With regards to the implementation of the proposed approach, the set of probability levels in Algorithm 1 was chosen to be $\mathcal{T} = \{0.1, 0.5, 0.9\}$. This set of levels has been shown to perform well in practice when working with quantile-based quantities [27, 39,45,46]. Therefore, we have decided not to increase the computational time of the algorithm by looking for a proper set. The hyperparameter α was chosen as to detect the top 1 outlier in Scenarios 1.1, 2.1 and 3.1, and the top 2 outliers in Scenarios 1.2, 2.2 and 3.2. The depth D_{UF} in Algorithm 2 was chosen as the Fraiman–Muniz depth in (10). Note that this depth has an associated depth for univariate data (represented by Z_i in (10)). The choice for this univariate depth was the Tukey depth provided in expression (11). We have also analyzed other types of depths, but the mentioned choices gave the best overall results in terms of outlier detection success.

4.2. Alternative approaches and assessment criteria

To shed light on the performance of the designed method, which we will refer to as *quantile cross-spectral density functional for outlier detection* (QCD-F-OD), we decided to compare this approach with a simpler version of itself that does not consider functional data and with two well-known approaches. The alternative techniques are described below.

- *The quantile cross-spectral density as multivariate data for outlier detection* (QCD-M-OD). This strategy performs in the following way. First, for each MTS, the smoothed CCR-periodograms in (9) are computed for a fixed set of probability levels, $\mathcal{T}$, and evaluated at the Fourier frequencies. Next, real and imaginary parts of all the elements are obtained separately and concatenated in a vector describing the MTS. Once this vector has been obtained for all the series, outlier detection is carried out by considering the Tukey depth for multivariate data. Several other depths have been analyzed, but the Tukey depth attained the best results in terms of outlier identification (see Section 4.3). Note that this approach can be seen as a benchmark regarding whether or not the treatment of QCD as functional data is advantageous for outlier detection. Concerning the implementation, the probability levels and values for α employed for QCD-F-OD were also considered for QCD-M-OD.

- *Feature-based approach for outlier detection* (FB-OD). We considered the approach for anomaly detection proposed in [16]. This strategy is originally designed to deal with UTS. It consists of three stages: (i) extracting a variety of features from each UTS, (ii) performing principal component analysis (PCA) on the feature-space and (iii) retaining the first two principal components and applying outlier detection in the reduced space. The third stage can be performed either by means of a density-based method or a procedure considering α-hulls. The numerical experiments carried out in [16] showed that the latter approach is more effective. Here, we extend the technique of [16] to the multidimensional setting by applying the algorithm to each UTS conforming the MTS. α-hulls are considered for the third stage.

Table 1

Averages rates of correct outlier identification for scenarios with one outlying series. For each scenario and value of the series length, the best result is shown in bold.

	Length	QCD-M-OD	QCD-F-OD	FB-OD	EB-OD
Scenario 1.1	$T = 200$	0.335	**0.670**	0.175	0.270
	$T = 400$	0.440	**0.955**	0.300	0.435
	$T = 600$	0.540	**0.990**	0.305	0.600
Scenario 2.1	$T = 200$	0.550	**0.985**	0.210	0.225
	$T = 400$	0.755	**1**	0.275	0.280
	$T = 600$	0.830	**1**	0.335	0.335
Scenario 3.1	$T = 400$	0.350	**0.550**	0.100	0.460
	$T = 800$	0.450	**0.920**	0.095	0.520
	$T = 1200$	0.605	**0.990**	0.125	0.590

- *Eros-based procedure for outlier detection* (EB-OD). A methodology for discovering outlying elements within an MTS database is provided in [21]. The procedure is based on the *Eros* distance measure [47]. First, the pairwise distance matrix of the MTS dataset is computed. Next, an outlier score is assigned to each element by taking into account the distance to its k nearest neighbors.

Fixed an scenario, a value for T, and a simulation trial, we computed the top 1 outlier (Scenarios 1.1, 2.1 and 3.1) or the top 2 outliers (Scenarios 1.2, 2.2 and 3.2) according to the four analyzed procedures. In the former scenarios, we recorded the proportion of times over the 200 trials that the outlier detection task was successful. In other words, we obtained the proportion of times that the series identified as outlying by the methods was the true outlier series. In the case of Scenarios 1.2, 2.2 and 3.2, we computed the proportion of times that the procedures: (i) successfully detected both true anomalous series, (ii) successfully identified one true anomalous series but failed to detect the remaining one, and (iii) identified as anomalous series two series generated from the base generating process. Note that the three previous metrics provide a comprehensive summary about the effectiveness of the different algorithms in scenarios containing two outliers.

Since the feature-based approach FB-OD performs anomaly identification in each component of the MTS, we considered this method successful if it succeeded at least in one of the components. Concerning EB-OD, we analyzed all the possible values for the number of neighbors, k, being enough that the procedure detected the outlier MTS for one of the values of k. Note that the previous remarks mean a considerable advantage of FB-OD and EB-OD over QCD-F-OD and QCD-M-OD in the assessment task.

4.3. Results and discussion

Performance metrics concerning the 200 trials of the simulation procedure are given in Table 1 for scenarios with one outlying series and in Table 2 for scenarios with two outlying series. The abbreviation CDO in the third column of Table 2 stands for *Correctly Detected Outliers*. As indicated in Section 4.2, results for Scenarios 1.2, 2.2 and 3.2 are given separately depending on the number of true outliers correctly detected. Note that, for a given scenario, method, and value for the series length, the three corresponding quantities in Table 2 add up to one.

When there was only one outlier, we can see from Table 1 that QCD-F-OD outperformed the remaining methods by a large degree. The feature-based approach FB-OD obtained the worst average scores for all scenarios and series lengths. This is probably due to the fact that a lot of the extracted features are not useful to distinguish the outlier series from the rest, thus incorporating a lot of noise in the anomaly classification task. The method

based on the *Eros* distance achieved acceptable results, specially in Scenarios 1.1 and 3.1. It is clear that this method is able to detect MTS coming from a distinct generating process, as its performance significantly improves with the series length. However, this procedure needs very long series to get perfect results, a condition which is not always guaranteed in practice. The results concerning QCD-M-OD show that this technique is useful for outlier detection purposes, but significantly worse than QCD-F-OD. The latter strategy attained average rates of correct outlier identification substantially greater than the former. In fact, QCD-F-OD exhibited success rates doubling those obtained by QCD-M-OD in several settings. Table 1 also shows that both QCD-based methods and the feature-based approach attained the best results in the nonlinear setting, just where EB-OD struggled the most.

Results in Table 2 also reveal the superiority of QCD-F-OD in the scenarios with two anomalous series. QCD-M-OD and QCD-F-OD correctly detected the two outliers (CDO = 2) with similar rates to those obtained in scenarios with one outlying series. The scores attained in the former case are slightly worse than in the latter, which is expected since identifying two outliers is more challenging than detecting only one. It is worth noting that these methods barely failed to discover both atypical series (CDO = 0). In fact, the functional-based approach QCD-F-OD always identified correctly at least one outlying series in all the considered settings and simulation trials. The worst results of QCD-F-OD arose from Scenarios 1.2 and 3.2 with the shortest series ($T = 200$ and $T = 400$, respectively), where this technique failed to detect one outlier (CDO = 1) a considerable number of times. However, its performance substantially improved when increasing the series length.

Regardless of the value for T, the procedure QCD-M-OD was unable to identify at least one outlier in Scenarios 1.2 and 3.2 in more than 50% of the trials. Furthermore, the rates associated with CDO = 1 for QCD-M-OD in Scenario 3.2 do not decrease when increasing the series length. This is unexpected, since it indicates that QCD-M-OD struggled to accurately identify both anomalous MTS even for large sample sizes. It is interesting to remark that, most of the times, the method successfully detected the series associated with the time-varying correlation. By contrast, it was not able to spot the series associated with $\rho_t = 0.2$. Thus, We can conclude that the use of functional data through QCD-F-OD clearly provides a useful tool for the detection of this series.

With regards to FB-OD and EB-OD, both procedures substantially decreased their performance in comparison to the simpler scenarios. In particular, FB-OD completely failed to detect both outliers (CDO = 2) in all scenarios whatever the value of T. Actually, this method identified as outliers two non-anomalous series (CDO = 0) most of the times. On the other hand, the *Eros*-based method behaved poorly in Scenario 2.2 (nonlinear processes), where it generally misidentified to regular series as outliers. This technique attained the best results in Scenario 3.2 (dynamic conditional correlation processes) where it either identified both outliers (half of the time) or at least one. Finally, in Scenario 1.2, EB-OD clearly improved its behavior when increasing the series length, although it still frequently missed at least one outlier for $T = 600$.

In summary, the proposed method QCD-F-OD is clearly superior to the alternative approaches for outlier detection of MTS provided that the outlyingness is characterized by the generating process. Additionally, a big part of its success is undoubtedly attributed to the treatment of QCD as functional data.

Table 2
Proportion of times that each method correctly identified 2, 1 or 0 outliers. The column CDO stands for the number of *Correctly Detected Outliers*. For each scenario and value of the series length, the best result is shown in bold provided that CDO = 2.

	Length	CDO	QCD-M-OD	QCD-F-OD	FB-OD	EB-OD
Scenario 1.2	$T = 200$	2	0.245	**0.615**	0	0.070
		1	0.610	0.385	0.315	0.465
		0	0.145	0	0.685	0.465
	$T = 400$	2	0.325	**0.945**	0	0.200
		1	0.650	0.055	0.340	0.575
		0	0.025	0	0.660	0.226
	$T = 600$	2	0.460	**0.990**	0	0.375
		1	0.535	0.010	0.305	0.545
		0	0.005	0	0.695	0.085
Scenario 2.2	$T = 200$	2	0.545	**0.940**	0	0.015
		1	0.450	0.060	0.270	0.295
		0	0.005	0	0.730	0.690
	$T = 400$	2	0.680	**1**	0	0.020
		1	0.320	0	0.370	0.285
		0	0	0	0.630	0.695
	$T = 600$	2	0.750	**1**	0	0.045
		1	0.250	0	0.385	0.455
		0	0	0	0.615	0.500
Scenario 3.2	$T = 400$	2	0.330	**0.515**	0	0.425
		1	0.500	0.485	0.150	0.560
		0	0.170	0	0.850	0.015
	$T = 800$	2	0.415	**0.805**	0	0.525
		1	0.545	0.195	0.175	0.465
		0	0.040	0	0.825	0.010
	$T = 1200$	2	0.445	**0.920**	0	0.550
		1	0.550	0.080	0.180	0.450
		0	0.005	0	0.820	0

5. Time consumption analysis

In this section, we analyze the time consumption of QCD-F-OD as a function of the series length, T, the number of MTS in the collection, n, and the number of components (UTS), d. We recorded the runtime of the corresponding programs until the depths of all the elements were obtained. In order to measure the computation time as a function of one parameter, we maintained the rest of the parameters constant. The corresponding fixed values were $T = 500$, $d = 2$, $n = 40$. All the experiments were carried out in a MacBook Pro with processor Quad-Core Intel Core i7, a speed of 2.9 GHz and a RAM memory of 16 GB. The programs were coded and executed in RStudio. The R version was 3.6.1.

Figs. 4 show the CPU runtime according to T, n and d, respectively. As expected, in the three cases, there is an increase in time consumption of the method as the value of the corresponding parameter increases. With regards to the series length and the number of MTS, there is a linear increase, whereas for the number of components, there is a quadratic trend. This is reasonable, as the smoothed CCR-periodogram is computed for each pair of dimensions, and the number of obtained smoothed CCR-periodograms is in direct relationship with the dimension of the considered multivariate functional data. These empirical results are totally in accordance with the time complexity of the procedure derived in Section 3.

f5

6. Applications

In this section, we illustrate the usefulness of the proposed approach in applications with real MTS datasets. It is important to note that, in practice, some MTS are non-stationary. Although our method relies on QCD, which is well-defined only for stationary processes, the results from a comprehensive set of numerical experiments provided in [27] reveal that the use of the QCD-based features as descriptive quantities can give also valuable information about non-stationary series.
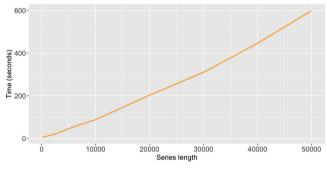

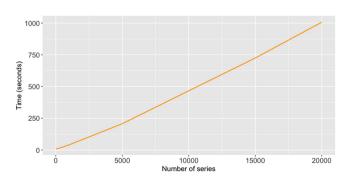**Fig. 4.** The CPU runtime versus the length of each MTS.

Fig. 5. The CPU runtime versus the number of components.

First, we apply the proposed methodology to two databases containing financial series and ECG signals, respectively. Next we provide some comments about the validity of the identified outliers in real applications as the ones presented here.

Table 3
The top 5 outlier samples in the financial MTS dataset.

Symbol	Company	Depth value
TSLA	Tesla Inc.	0.2555
NEE	NextEra Energy	0.2584
CVX	Chevron Corp	0.2628
NFLX	Netflix Inc.	0.2643
MA	Mastercard Inc.	0.2644

6.1. Outlier detection of financial time series

The first dataset was taken from the finance section of the Yahoo website[1]. It contains daily stock returns and trading volume of the top 50 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 50 contains some of the most important companies in the world, as Apple, Google, Facebook or Berkshire Hathaway.

It should be noted that the relationship between price and volume has been extensively analyzed in the literature [48–50] 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 valuable to characterize each of the considered companies in terms of the time evolution of these two quantities. Our goal is to analyze the joint behavior of prices and volume in order to perform outlier detection. Thus, we assume that a company shows an anomalous behavior with respect to the rest if the corresponding bivariate time series is an outlier series.

It can be observed that both the UTS of prices and trading volume are non-stationary in mean. Thus, all UTS are transformed by taking the first differences of the natural logarithm of the original values. This way, prices give rise to stock returns, and volume to what we call change in volume. It has been already remarked that the proposed method does not require stationary MTS to be successful. However, this transformation is common when dealing with this kind of series [51]. Finally, all UTS are normalized to have zero mean and unit variance.

We applied the outlier detection approach proposed in Section 3 to the set of transformed bivariate MTS. Just as in the simulations, we chose $\mathcal{T} = \{0.1, 0.5, 0.9\}$ and the Fraiman–Muniz and Tukey depths. The hyperparameter α was set to $\alpha = 0.10$ so that the top 5 outlying MTS were detected. The results of the anomaly detection procedure are given in Table 3. We can see that the most outlying MTS sample corresponds to Elon Musk's company Tesla Inc (TSLA). This fact is not surprising since TSLA is the only company in the sector of Automobile Manufacturers among the considered top 50. The remaining anomalous samples correspond to two companies from the Energy sector (NextEra and Chevron), and the well-known companies Netflix and Mastercard.

Fig. 6 depicts the bivariate series of Tesla, whereas Fig. 7 displays the series corresponding to the deepest element in the multivariate functional dataset (depth value of 0.3026), which represents the company Danaher Corp (DHR). This company could be considered as the prototype company among the top 50. By comparing Figs. 6 and 7, one can see that the outlier and the prototype series show patterns substantially different. For instance, it is clear from the plots that TSLA has been suffering from a greater degree of volatility than DHR over the considered years. This feature is observed for both returns and change in volume.
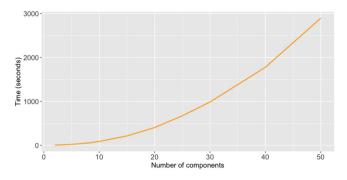

Fig. 6. Returns and change in volume of the company Tesla Inc.

Fig. 7. Returns and change in volume of the company Danaher Corp.

Note that the previous analysis could be valuable in many different financial contexts. For instance, it is common that an investor looks for companies deviating from the regular behavior of the index to invest in.

6.2. Outlier detection of ECG signals

The second dataset was extracted from the PTB Diagnostic ECG Database [52] and has been broadly used for ECG classification [53–57]. We have downloaded this database from the Kaggle repository[2]. The collection contains 15-lead ECG signals from both healthy patients and people with myocardial infarction condition. The minimum series length in this dataset is $T = 32000$. It is worth remarking that ECG signals are known to be nonstationary.

In this application we focused on the subset of healthy volunteers. There are 80 instances coming from 52 healthy volunteers. For the sake of simplicity, the first 6 leads and 500 time observations of each ECG signal were selected. The choice of certain subsets of dimensions and time observations is common in ECG data mining [53,58].

The proposed procedure was applied to the reduced set of 80 MTS. The considered parameters were the same as in the previous application. Table 4 contains the top 8 outlier ECG signals along with the corresponding depth values. The ECG signal No. 75 is the most outlying one.

Figs. 8 and 9 show the ECG signals No. 75 and No. 74, respectively. The signal No. 74 corresponds to the maximum depth (depth value of 0.3533), so it can be considered as the dataset prototype. It is clear from the figures that both signals differ substantially over the considered time periods. Whereas signal

---

Table 4
The top 8 outlier samples in the ECG MTS dataset.

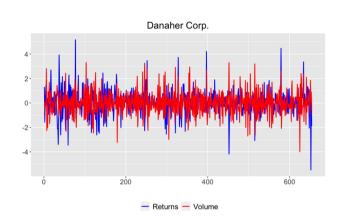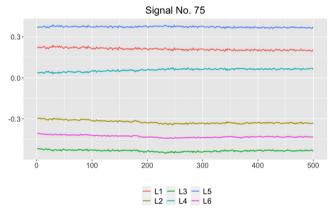| ECG signal	Depth value
No. 75	0.1666
No. 15	0.1670
No. 31	0.1731
No. 62	0.1951
No. 67	0.1969
No. 46	0.1973
No. 47	0.2114
No. 6	0.2249

Fig. 8. The ECG signal No. 75.

Fig. 9. The ECG signal No. 74.

No. 74 shows some marked fluctuations, signal No. 75 remains almost flat with regards to all its components and along all the time span.

In this context, the use of the proposed outlier detection technique could be clearly useful. In fact, the detection of outlier ECG signals in a group of healthy patients could assist the physician in detecting some patients at risk of suffering a given condition.

6.3. Assessing the accuracy of outlier identification in practical applications

A common problem faced by Machine Learning practitioners concerns the validity of the results provided by a given algorithm. For instance, it is often unfeasible to determine whether the labels assigned by a classification technique for new observations are right or wrong. The same issue arises when applying outlier detection algorithms as the one proposed in this manuscript. However, it is worth highlighting that the underlying uncertainty is precisely what makes this class of algorithms valuable. Having
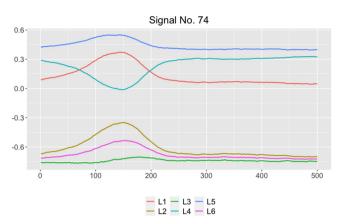

Fig. 10. Boxplots indicating the distribution of the depths concerning the S&P 500 companies (left panel) and the ECG signals (right panel).

said that, there exist some heuristic procedures which try to validate the output of a particular algorithm.

An interesting way to assess the accuracy of the outliers detected by QCD-F-OD in practical applications is by means of the distribution of the computed depths. The idea is to look for extreme values in the lower tail of this distribution. The observations associated with these extreme depths hold a high likelihood of outlyingness. For this purpose, a classical boxplot displaying the distribution of the depths can be depicted. Indeed, the existence of isolated points in the lower part of this graph could indicate the presence of anomalous observations. Note that this strategy is not only valid for the method QCD-F-OD, but for every anomaly identification approach relying on depths.

This criterion was applied to the two case studies analyzed in Sections 6.1 and 6.2 . Fig. 10 contains the corresponding boxplots regarding the financial series (left panel) and the ECG signals (right panel). Concerning the former boxplot, only one point can be observed in the lower end of the graph. This point corresponds to the company TSLA, which is associated with the lowest value of the depth as indicated in Section 6.1. On the other hand, three extreme observations can be noticed in the boxplot of the right panel. These observations are associated with signals No. 75, No. 15 and No. 31 as seen in Table 4. Therefore, either TSLA in the first case study or the three mentioned signals in the second one could be regarded as potential outliers, clearly deserving a careful investigation.

Note that, in the analyses carried out in Sections 6.1 and 6.2, the number of outlier series to be detected was set in advance by choosing $\alpha = 0.10$. It is worth remarking that, in some applications, the researcher decides beforehand the number of anomalous observations to be trimmed away. In many other cases, however, this quantity is not known, entailing a challenge for the practitioner. In this regard, the approach proposed in this section can be seen as a heuristic way of selecting the parameter α. This tool could be also valuable in exploratory analysis stages.

Finally, it is important to highlight that it is always challenging to guarantee the accuracy of the identified outliers in practical cases. A wise roadmap would be to run first the outlier detection procedure and then make a specific analysis of each potentially outlier series. This way, the strength of the proposed algorithm can be combined with the domain knowledge of the practitioner in order to make an informed decision about the outlying nature of a given element.

7. Conclusions

This work has developed a method for detecting outlying MTS samples in a given set according to the generating process, a topic

barely addressed in the literature. The procedure relies on the quantile cross-spectral density, functional data, and functional depths. These elements make a novel combination that, to the best of our knowledge, has not been considered before. The notion of depth allows to assign to each MTS an outlier score, hence making straightforward the identification of the top m outliers. A broad simulation study has shown that the treatment of the quantile cross-spectral density as functional data is advantageous for anomaly detection. In fact, the approach shows a higher efficacy than its non-functional counterpart. The proposed methodology also outperforms the few other existing approaches aimed to identify anomalous MTS. The technique has been applied to perform outlier detection in two real datasets containing financial time series and ECG signals. A graphical tool for assessing the quality of the detected outliers in real applications has also been introduced.

Besides its effectiveness, the proposed method is computationally efficient. The parameter which increases the computational time the most is the number of dimensions, d. With respect to this parameter, the approach has a computational complexity of $O(d^2)$. Still, it beats the second best performed method, Weng's method, by an order of magnitude, being the latter $O(d^3)$.

This work opens three new research directions. First, given that the utilization of the functional smoothed CCR-periodograms has proven useful for anomaly recognition of MTS, it would be interesting to study if these advantages transfer to a framework of clustering or classification. Second, additional time series features as the autocorrelations or cross-correlations could be also considered as functional data. In this way, classical algorithms for time series data mining could be adapted to the functional context. Third, it is worth mentioning that the pervasiveness of MTS with a huge number of dimensions is becoming increasingly common (e.g., sensor data). Therefore, it is highly desirable to reduce the computational complexity of the proposed procedure to one which is linear, or almost linear, with respect to d. The three directions will be properly addressed in further works.

CRediT authorship contribution statement

Ángel López-Oriona: Conceptualisacion, Writing – review & editing, Methodology, Software, Data processing, Visualisation. **José A. Vilar:** Investigation, Supervision, Writing – review & editing, Project administration.

Declaration of competing interest

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

Acknowledgments

Appendix. Description and implementation of the smoothed CCR-periodogram

A detailed description of the smoothed CCR-periodogram mentioned in Section 2 is provided below.

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 $(\tau, \tau') \in [0, 1]^2$, [28] propose to estimate $\mathfrak{f}_{j_1, j_2}(\omega, \tau, \tau')$ by considering a smoother of the cross-periodograms based on the indicator functions $I\{\hat{F}_{T,j}(X_{t,j})\}$, where $\hat{F}_{T,j}(x) = T^{-1} \sum_{t=1}^{T} I\{X_{t,j} \leq x\}$ denotes the empirical distribution function of $X_{t,j}$. This approach extends to the multivariate case the estimator proposed by [59] in the univariate setting. More specifically, the called rank-based copula cross periodogram (CCR-periodogram) is defined by

$$I_{T,R}^{j_1, j_2}(\omega, \tau, \tau') = \frac{1}{2\pi T} d_{T,R}^{j_1}(\omega, \tau) d_{T,R}^{j_2}(-\omega, \tau'), \tag{16}$$

where

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

The asymptotic properties of the CCR-periodogram are established in Proposition S4.1 of [28]. Likewise the standard cross-periodogram, the CCR-periodogram is not a consistent estimator of $\mathfrak{f}_{j_1, j_2}(\omega, \tau, \tau')$. To achieve consistency, the CCR-periodogram ordinates (evaluated on the Fourier frequencies) are convolved with weighting functions $W_T(\cdot)$. The *smoothed CCR-periodogram* takes the form

$$\hat{G}_{T,R}^{j_1, j_2}(\omega, \tau, \tau') = (2\pi / T) \sum_{s=1}^{T-1} W_T\left(\omega - \frac{2\pi s}{T}\right) I_{T,R}^{j_1, j_2}\left(\frac{2\pi s}{T}, \tau, \tau'\right), \tag{17}$$

where

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

with $h_T > 0$ being 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 $[-\pi, \pi]$, frequently called kernel function. Consistency and asymptotic performance of the smoothed CCR-periodogram, $\hat{G}_{T,R}^{j_1, j_2}(\omega, \tau, \tau')$, are established in Theorem S4.1 of [28].

The CCR-periodograms were computed throughout this manuscript by considering the so-called Epanechnikov kernel, which is defined as

$$W(u) = \frac{3}{4\pi}\left(1 - \frac{u}{\pi}\right)^2, \quad \text{for } u \in [-\pi, \pi]. \tag{18}$$

The associated bandwidth h_T was chosen as $h_T = T^{-1/5}/10$, which fulfills the assumptions required in Theorem S4.1 of [28] to ensure that the smoothed CCR-periodogram is a consistent estimate of QCD.

Computations of the smoothed CCR-periodogram were carried out by using the R-package **quantspec** [60].

References

[1] A. Blázquez-García, A. Conde, U. Mori, J.A. Lozano, A review on outlier/anomaly detection in time series data, 2020, arXiv preprint arXiv: 2002.04236.

[2] B. Abraham, A. Chuang, Outlier detection and time series modeling, Technometrics 31 (2) (1989) 241–248.

[3] G.M. Ljung, On outlier detection in time series, J. R. Stat. Soc. Ser. B Stat. Methodol. 55 (2) (1993) 559–567.

[4] R.S. Tsay, Outliers, level shifts, and variance changes in time series, J. Forecast. 7 (1) (1988) 1–20.

[5] S. Basu, M. Meckesheimer, Automatic outlier detection for time series: an application to sensor data, Knowl. Inf. Syst. 11 (2) (2007) 137–154.

[6] F. Battaglia, L. Orfei, Outlier detection and estimation in nonlinear time series, J. Time Series Anal. 26 (1) (2005) 107–121.

[7] J.-i. Takeuchi, K. Yamanishi, A unifying framework for detecting outliers and change points from time series, IEEE Trans. Knowl. Data Eng. 18 (4) (2006) 482–492.

[8] C. Caroni, V. Karioti, Detecting an innovative outlier in a set of time series, Comput. Statist. Data Anal. 46 (3) (2004) 561–570.

[9] P. Perron, G. Rodríguez, Searching for additive outliers in nonstationary time series, J. Time Series Anal. 24 (2) (2003) 193–220.

[10] K. Yamanishi, J.-i. Takeuchi, A unifying framework for detecting outliers and change points from non-stationary time series data. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002, pp. 676–681.

[11] Y. Cai, N. Davies, A simple diagnostic method of outlier detection for stationary Gaussian time series, J. Appl. Stat. 30 (2) (2003) 205–223.

[12] P. Galeano, D. Peña, R.S. Tsay, Outlier detection in multivariate time series by projection pursuit, J. Amer. Statist. Assoc. 101 (474) (2006) 654–669.

[13] R. Baragona, F. Battaglia, Outliers detection in multivariate time series by independent component analysis, Neural Comput. 19 (7) (2007) 1962–1984.

[14] H. Cheng, P.-N. Tan, C. Potter, S. Klooster, Detection and characterization of anomalies in multivariate time series, in: Proceedings of the 2009 SIAM International Conference on Data Mining, SIAM, 2009, pp. 413–424.

[15] D. Cucina, A. Di Salvatore, M.K. Protopapas, Outliers detection in multivariate time series using genetic algorithms, Chemometr. Intell. Lab. Syst. 132 (2014) 103–110.

[16] R.J. Hyndman, E. Wang, N. Laptev, Large-scale unusual time series detection, in: 2015 IEEE International Conference on Data Mining Workshop (ICDMW), IEEE, 2015, pp. 1616–1619.

[17] L. Beggel, B.X. Kausler, M. Schiegg, M. Pfeiffer, B. Bischl, Time series anomaly detection based on shapelet learning, Comput. Statist. 34 (3) (2019) 945–976.

[18] S.-E. Benkabou, K. Benabdeslem, B. Canitia, Unsupervised outlier detection for time series by entropy and dynamic time warping, Knowl. Inf. Syst. 54 (2) (2018) 463–486.

[19] U. Rebbapragada, P. Protopapas, C.E. Brodley, C. Alcock, Finding anomalous periodic time series, Mach. Learn. 74 (3) (2009) 281–313.

[20] V. Karioti, C. Caroni, Detecting outlying series in sets of short time series, Comput. Statist. Data Anal. 39 (2) (2002) 351–364.

[21] X. Weng, J. Shen, Detecting outlier samples in multivariate time series dataset, Knowl.-Based Syst. 21 (8) (2008) 807–812.

[22] R.S. Tsay, D. Pena, A.E. Pankratz, Outliers in multivariate time series, Biometrika 87 (4) (2000) 789–804.

[23] X. Weng, J. Shen, Outlier mining for multivariate time series based on local sparsity coefficient, in: 2006 6th World Congress on Intelligent Control and Automation, 2, IEEE, 2006, pp. 5957–5961.

[24] M. Agyemang, C.I. Ezeife, Lsc-mine: Algorithm for mining local outliers, in: Proceedings of the 15th Information Resource Management Association (IRMA) International Conference, New Orleans, Vol. 1, Citeseer, 2004, pp. 5–8.

[25] R.H. Shumway, D.S. Stoffer, D.S. Stoffer, Time Series Analysis and Its Applications, vol. 3, Springer, 2000.

[26] G.E. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time series analysis: forecasting and control, John Wiley & Sons, 2015.

[27] A. López-Oriona, J.A. Vilar, Quantile cross-spectral density: A novel and effective tool for clustering multivariate time series, Expert Syst. Appl. (2021) 115677, http://dx.doi.org/10.1016/j.eswa.2021.115677, URL https://www.sciencedirect.com/science/article/pii/S0957417421010630.

[28] J. Baruník, T. Kley, Quantile coherency: A general measure for dependence between cyclical economic variables, Econom. J. 22 (2) (2019) 131–152.

[29] S. López-Pintado, J. Romo, On the concept of depth for functional data, J. Amer. Statist. Assoc. 104 (486) (2009) 718–734.

[30] R. Fraiman, G. Muniz, Trimmed means for functional data, Test 10 (2) (2001) 419–440.

[31] A. Cuevas, M. Febrero, R. Fraiman, Robust estimation and classification for functional data via projection-based depth notions, Comput. Statist. 22 (3) (2007) 481–496.

[32] T. Harris, J.D. Tucker, B. Li, L. Shand, Elastic depths for detecting shape anomalies in functional data, Technometrics (2020) 1–11.

[33] J. Ramsay, B.W. Silverman, Functional data analysis (springer series in statistics), 1997.

[34] E.O. Brigham, R. Morrow, The fast Fourier transform, IEEE Spectr. 4 (12) (1967) 63–70.

[35] C.A. Hoare, Quicksort, Comput. J. 5 (1) (1962) 10–16.

[36] P.J. Davis, P. Rabinowitz, Methods of Numerical Integration, Courier Corporation, 2007.

[37] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Econometrics 31 (3) (1986) 307–327.

[38] 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.

[39] B. Lafuente-Rego, J.A. Vilar, Clustering of time series using quantile autocovariances, Adv. Data Anal. Classif. 10 (3) (2016) 391–415.

[40] C.W. Granger, T. Terasvirta, et al., Modelling non-linear economic relationships, OUP Categ. (1993).

[41] C.W.J. Granger, G. CWJ, A. AP, An introduction to bilinear time series models, 1978.

[42] 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, 2009, pp. 9–56.

[43] R. Engle, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, J. Bus. Econom. Statist. 20 (3) (2002) 339–350.

[44] R.Y. Liu, J.M. Parelius, K. Singh, et al., Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh), Ann. Statist. 27 (3) (1999) 783–858.

[45] 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 and Systems 340 (2018) 38–72.

[46] B. Lafuente-Rego, P. D'Urso, J. Vilar, Robust fuzzy clustering based on quantile autocovariances, Statist. Papers 61 (6) (2020) 2393–2448.

[47] K. Yang, C. Shahabi, A PCA-based similarity measure for multivariate time series, in: Proceedings of the 2nd ACM international workshop on Multimedia databases, 2004, pp. 65–74.

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

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

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

[51] S.-S. Chen, Revisiting the empirical linkages between stock returns and trading volume, J. Bank. Financ. 36 (6) (2012) 1781–1788.

[52] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, Physiobank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circ. 101 (23) (2000) e215–e220.

[53] E.A. Maharaj, A.M. Alonso, Discriminant analysis of multivariate time series: Application to diagnosis based on ECG signals, Comput. Statist. Data Anal. 70 (2014) 67–87.

[54] M. Kachuee, S. Fazeli, M. Sarrafzadeh, Ecg heartbeat classification: A deep transferable representation, in: 2018 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, 2018, pp. 443–444.

[55] R. Remya, K. Indiradevi, K.A. Babu, Classification of myocardial infarction using multi resolution wavelet analysis of ECG, Proc. Technol. 24 (2016) 949–956.

[56] S. Banerjee, M. Mitra, Ecg feature extraction and classification of anteroseptal myocardial infarction and normal subjects using discrete wavelet transform, in: 2010 International Conference on Systems in Medicine and Biology, IEEE, 2010, pp. 55–60.

[57] S. Banerjee, M. Mitra, Application of cross wavelet transform for ECG pattern analysis and classification, IEEE Trans. Instrum. Meas. 63 (2) (2013) 326–333.

[58] D. Sadhukhan, S. Pal, M. Mitra, Automated identification of myocardial infarction using harmonic phase distribution pattern of ECG data, IEEE Trans. Instrum. Meas. 67 (10) (2018) 2303–2313.

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

[60] T. Kley, Quantile-based spectral analysis in an object-oriented framework and a reference implementation in R: The quantspec package, J. Stat. Softw. 70 (3) (2016) 1–27, http://dx.doi.org/10.18637/jss.v070.i03.