

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems ●●● (●●●●) ●●●—●●●

FUZZY
sets and systemswww.elsevier.com/locate/fss

Quantile autocovariances: A powerful tool for hard and soft partitional clustering of time series

José A. Vilar^{a,*}, Borja Lafuente-Rego^a, Pierpaolo D'Urso^b^a *Universidade da Coruña, Research Group on Modeling, Optimization and Statistical Inference (MODES), Department of Mathematics, Faculty of Computer Science, Elviña, 15071 A Coruña, Spain*^b *Dipartimento di Scienze Sociali ed Economiche, Sapienza University of Rome, Pza. Aldo Moro, 5, 00185 Rome, Italy*

Received 18 September 2016; received in revised form 6 March 2017; accepted 8 March 2017

Abstract

A key issue in cluster analysis is determining a proper dissimilarity measure between two data objects, and many pairwise dissimilarities have been proposed to deal with time series. Assuming that the clustering purpose is to group series according to the underlying dependence structures, a detailed study of the behavior in clustering of a dissimilarity based on comparing estimated quantile autocovariance functions (QAF) is carried out. Quantile autocovariances provide information about the serial dependence structure that other conventional features are not able to capture, which suggests great potential to perform clustering of series. The asymptotic behavior of the sample quantile autocovariances is studied and an algorithm to determine optimal combinations of lags and pairs of quantile levels to perform clustering is introduced. The proposed metric is used to perform hard and soft partitioning-based clustering. First, a broad simulation study examines the behavior of the proposed metric in crisp clustering with the PAM procedure. A novel fuzzy *C*-medoids algorithm based on the QAF-dissimilarity is then proposed and compared with other fuzzy procedures in a new simulation study conducted to cluster fuzzy scenarios involving AR and GARCH models. In all cases, the QAF-based procedures outperform or are highly competitive with a range of dissimilarities reported in the literature, particularly exhibiting high capability to cluster conditionally heteroskedastic time series and robustness to the distributional form of the errors. Two specific applications involving air quality data and financial time series illustrate the usefulness of the proposed procedures.

© 2017 Elsevier B.V. All rights reserved.

Keywords: Time series clustering; Quantile autocovariances; Partitioning around medoids; Fuzzy *C*-medoids clustering; Heteroskedastic models

1. Introduction

Time series clustering is aimed at splitting a set of unlabeled time series into homogeneous groups, which is a central problem to many knowledge discovery tasks. Its applications span an extensive assortment of fields, including economics, finance, marketing, genetics, biology, medicine, environmental studies, machine learning and bioinfor-

* Corresponding author. Fax: +34 981 167 160.

E-mail addresses: jose.vilarf@udc.es (J.A. Vilar), borja.lafuente@udc.es (B. Lafuente-Rego), pierpaolo.durso@uniroma1.it (P. D'Urso).

<http://dx.doi.org/10.1016/j.fss.2017.03.006>

0165-0114/© 2017 Elsevier B.V. All rights reserved.

matics, among others. Determining products with similar selling patterns, identifying user profiles according to the browsing behavior on a website, classifying banks on the basis of share price series, grouping countries with similar population growth or regions with similar temperature are some typical applications where the similarity searching between time series arises in a natural way. Due to the growing interest of researchers from different disciplines, time series clustering has become a very active research area nowadays, being especially significant the huge number of contributions on the topic provided in the last two decades. A comprehensive overview of time series clustering is provided by Liao [1] and more recently by Fu [2], Rani and Sikka [3] and Caiado, Maharaj and D’Urso [4], including recent advances, significant references and specific application areas.

In the present study, the focus is on the so-called partitioning-based clustering methods, which are aimed at the iterative relocation of data objects between clusters so that the dispersion within clusters decreases at each iteration. Two important decisions must be taken to perform these methods, namely how dissimilarity between objects is measured and how the assignment of objects to the groups is carried out.

The choice of a proper dissimilarity measure mainly relies on the specific purpose of the clustering. By dealing with time series, we must be aware of their dynamic nature. If, for example, the target is to group series with similar underlying dependence structures, then a conventional dissimilarity based on static objects and overlooking temporal features will likely produce unsatisfactory results. Many approaches to measure dissimilarity between series have been proposed in the literature (see e.g. Montero and Vilar [5] and references therein). For practitioners, the R package **TSclust** [6] presents a large set of well-established peer-reviewed time series dissimilarity measures, including measures based on raw data, extracted features, underlying models, complexity levels, and forecast behaviors. Some papers reporting experimental comparisons of several dissimilarity measures are also available [7–9]. Nevertheless, many of the proposed dissimilarities require regularity conditions on the generating models (linearity, gaussianity, homoscedasticity, ...), which are not always satisfied in practice. Hence there are still important challenges to be addressed in relation to this point, and this is one of the topics considered in this work.

As far as the cluster assignment, two different paradigms are usually considered depending on whether a “hard” or “soft” partition is constructed. Traditional clustering methods assign each data object to exactly one cluster, thus producing a hard partition of the data into non-empty and disjoint subsets. This approach can result too rigid in situations with data objects equidistant from two or more groups or in presence of overlapping clusters. Fuzzy cluster techniques [10,11] provide a more versatile approach by allowing gradual membership of data objects to clusters. In the resulting soft partition, the objects can belong to several clusters with specific membership levels indicating the amount of confidence in the assignment of each data to the clusters. Adoption of the fuzzy logic in time series clustering is interestingly motivated by some authors. D’Urso and Maharaj [12] and D’Urso et al. [13,14] argue that the dynamic of a time series may change over time in such a way that it could belong to distinct clusters during different periods of time, i.e. in a fuzzy way. Aielli and Caporin [15] motivate a soft clustering based on mixture models arguing that whether similarity is based on estimated dynamic parameters, then the error estimation generates variability causing overlapping clusters. Although the hard procedures have received greater attention in the time series clustering literature, a number of recent contributions have adopted the fuzzy approach combined with different dissimilarity criteria between series, including distances based on autocorrelation functions [12], features extracted in the frequency domain (normalized periodograms and cepstral coefficients, among others) [16], autoregressive approximations [13, 17], wavelet analysis [18–20], analysis of extreme values [21] and estimated GARCH coefficients [22,23].

This paper is aimed at assessing the behavior of a distance based on estimated quantile autocovariances (QA) in partitional clustering of time series, considering both crisp and fuzzy clustering procedures. We assume that the target is to group series according to the underlying dependence structures, i.e. similarity between series is measured in terms of similarity between generating processes. This criterion makes sense if, for example, one wishes to identify a few temporal behavior patterns hidden behind long time series observed with a high amount of noise. In this framework, using a metric robust to the generating mechanism of the series is necessary to attain a suitable cluster solution. Unlike other extracted features, quantile autocovariances account for important dynamic features of time series and are well-defined for a broad class of processes, including non-linear and heteroskedastic processes [24,25], which suggests that a dissimilarity comparing quantile autocovariances could report satisfactory results. This intuition was corroborated in our previous work [26], where results from an extensive simulation study showed that the QA-based metric outperforms or is highly competitive with a range of alternative dissimilarities, also exhibiting high capability to cluster time series generated from a broad range of dependence models. Nevertheless, our study in [26] was limited

to hierarchical procedures and non-fuzzy data, and an automatic criterion to select the input parameters required by the metric was not provided, remaining as an open issue to be addressed in further research.

The contribution of this work consists of three points. First, the simulation study carried out in [26] is extended to cover a hard non-hierarchical procedure, the partitioning around medoids (PAM) algorithm [27], which returns a subset of series (medoids) representative of the identified clusters. The attained results show the good behavior of the QA-based metric compared to other commonly used dissimilarities. In particular, excellent scores are reported by classifying heteroskedastic processes, which are frequently used with economic or financial indicators [22,28–30]. Further, since heteroskedastic Gaussian models cannot often capture the asymmetry and leptokurtosis exhibited by some financial time series (e.g. log-return series of stock indices) [31,32], additional simulations based on heteroskedastic models with non-normal errors are performed attaining even better results.

Second contribution concerns the optimal selection of input parameters, i.e. establishing how many and which combinations of lags and quantile levels must be used to define the QA metric in order to optimize the clustering process. A proper adjustment of the variable selection algorithm proposed by Andrews and McNicholas [33] for clustering and classification allows us to address this problem. Nevertheless, it is worth remarking that using a small number of quantiles with probability levels regularly spaced is enough to reach satisfactory results.

Third contribution consists of introducing a novel fuzzy procedure to cluster time series. We adopt a fuzzy C -medoids approach where the QA-based metric is considered to compute distances between series and medoids. In this way, the proposed approach inherits the advantages of the fuzzy methods (flexibility to describe complex cluster structures with overlapping clusters), the partitioning around medoids technique (selection of particular series representing the underlying cluster patterns) and the QA-based metric (high capability to discriminate between a broad range of dependence structures). Here, our experiments mainly focused on the classification of heteroskedastic models, a complex scenario but frequently realistic when analyzing financial, industrial or environmental indicators, among others. We assess the capability of our proposal to clustering GARCH models, and its performance is tested against two fuzzy clustering algorithms considering distances between GARCH approximations [22], and therefore specifically designed to work in the simulated scenario. Lastly, two specific applications to environmental and financial data sets are presented to illustrate the usefulness of the proposed clustering procedure.

The rest of the paper is organized as follows. Section 2 proposes to measure dissimilarity between a pair of observed series by comparing sequences of estimated quantile autocovariances. The estimation procedure is detailed, the asymptotic behavior established, and the dissimilarity defined and motivated. Section 3 introduces an algorithm to select the optimal combinations of lags and pairs of quantile levels in order to perform clustering using the QA-based dissimilarity. In Section 4, the QA-based dissimilarity is used to perform crisp clustering with the PAM procedure in different simulation scenarios. Three classification scenarios featured by the kind of generating process are considered, namely linear, non-linear and conditional heteroskedastic models, also including different distributional forms for the errors. The results are compared with the ones obtained using other dissimilarity measures. Section 5 focuses on the fuzzy clustering approach. The fuzzy C -medoids algorithm based on the QA-dissimilarity is analyzed and compared with other alternative fuzzy procedures in a new simulation study conducted to cluster different fuzzy structures of autoregressive and generalized autoregressive conditional heteroskedasticity models. Applications to real data sets are presented in Section 6 and some concluding remarks are summarized in Section 7.

2. Distance between time series based on quantile autocovariances

2.1. The quantile autocovariance function

Let $\{X_t; t \in \mathbb{Z}\}$ be a strictly stationary process and denote by F the marginal distribution of X_t and by $q_\tau = F^{-1}(\tau)$, $\tau \in [0, 1]$, 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 $I(X_t \leq q_\tau)$ and $I(X_{t+l} \leq q_{\tau'})$ given by

$$\gamma_l(\tau, \tau') = \text{cov}\{I(X_t \leq q_\tau), I(X_{t+l} \leq q_{\tau'})\} = \mathbb{P}(X_t \leq q_\tau, X_{t+l} \leq q_{\tau'}) - \tau \tau'. \quad (1)$$

Function $\gamma_l(\tau, \tau')$, with $(\tau, \tau') \in [0, 1]^2$, is called *quantile autocovariance function (QAF) of lag l* and can be seen as a generalization of the classical autocovariance function. While the latter measures linear dependence between different lags by evaluating covariability with respect to the average, the former studies the joint variability of the events $\{X_t \leq q_\tau\}$ and $\{X_{t+l} \leq q_{\tau'}\}$, i.e. examines how a part of the range of variation of X_t helps to predict whether the series

will be below quantiles in a future time. By definition, QAF captures the sequential dependence structure of a time series, thus accounting for serial features related to the joint distribution of (X_t, X_{t+l}) that the simple autocovariances cannot detect. Unlike the usual autocovariance function, QAF is well-defined even for processes with infinite moments and takes advantage from the local distributional properties inherent to the quantile methods, in particular showing a greater robustness against heavy tails, dependence in the extremes and changes in the conditional shapes (skewness, kurtosis) [25,34–38]. Based on these nice properties, QAF and its representations in the frequency domain (quantile periodogram and quantile spectral density) have been considered in several inference problems, including evaluation of directional predictability between time series [24,39] and testing procedures for specific aspects of serial dependence such as interrelatedness, conditional homoscedasticity or conditional symmetry [40–42].

An estimator of $\gamma_l(\tau, \tau')$ can be constructed replacing the theoretical quantiles by the corresponding empirical quantiles \hat{q}_τ and $\hat{q}_{\tau'}$ based on T realizations $\{X_t, 1 \leq t \leq T\}$ of the process. This way, the estimated QAF is given by

$$\hat{\gamma}_l(\tau, \tau') = \frac{1}{T-l} \sum_{t=1}^{T-l} I(X_t \leq \hat{q}_\tau) I(X_{t+l} \leq \hat{q}_{\tau'}) - \tau \tau', \tag{2}$$

where the empirical quantiles \hat{q}_α , for $0 \leq \alpha \leq 1$, can be formally seen as the solution of the minimization problem [43, page 7] given by

$$\hat{q}_\alpha = \arg \min_{q \in \mathbb{R}} \sum_{t=1}^T \rho_\alpha(X_t - q),$$

with $\rho_\alpha(x) = x(\alpha - I(x < 0))$.

2.2. Asymptotic behavior

The asymptotic behavior of the sample quantile autocovariances defined by (2) is established in Theorem 2.1 of this section following the asymptotic analysis developed by Han et al. [39]. Specifically, consider a two-dimensional strictly stationary process $X_t = \{(X_{1t}, X_{2t}); t \in \mathbb{Z}\}$ with marginal distribution functions $F_i(\cdot)$ and quantiles $q_{i,\tau}$, for $i = 1, 2$ and $\tau \in (0, 1)$. Under general weak dependence conditions on X_t , Han et al. [39, Th. 1] obtain the asymptotic distribution of the sample cross-correlation between the events $I(X_{1t} \leq \hat{q}_{1,\tau})$ and $I(X_{2(t+l)} \leq \hat{q}_{2,\tau'})$, for an arbitrary lag l and quantile levels τ and τ' . Adapting this result to the univariate setting and considering the non-normalized version of the mentioned cross-correlations, the limiting distribution stated in Theorem 2.1 for the sample quantile autocovariances is directly derived. First, some useful notation and the required assumptions are introduced.

Given an arbitrary lag l , let $\mathcal{A} \equiv \mathcal{A}_t \times \mathcal{A}_{t+l}$ be a compact subset in $(0, 1)^2$, where \mathcal{A}_t and \mathcal{A}_{t+l} denote quantile ranges of interest for X_t and X_{t+l} , respectively. Denote by $F_l(\cdot, \cdot)$ the joint distribution of (X_t, X_{t+l}) , and for $t = 1, \dots, T$ and $(\tau, \tau') \in \mathcal{A}$, consider the vector given by

$$\xi_{t,l}(\tau, \tau') = (I(X_t \leq q_\tau, X_{t+l} \leq q_{\tau'}) - F_l(q_\tau, q_{\tau'}), I(X_t \leq q_\tau) - \tau, I(X_{t+l} \leq q_{\tau'}) - \tau')^t.$$

Now, define the three-dimensional mean-zero Gaussian process $\{\mathbb{B}_l(\tau, \tau'); (\tau, \tau') \in (0, 1)^2\}$ having covariance matrix given by

$$\Gamma_l((\tau_1, \tau'_1), (\tau_2, \tau'_2)) = \mathbb{E}(\mathbb{B}_l(\tau_1, \tau'_1) \mathbb{B}_l^t(\tau_2, \tau'_2)) = \sum_{t=-\infty}^{\infty} \text{cov}(\xi_{t,l}(\tau_1, \tau'_1), \xi_{t,l}^t(\tau_2, \tau'_2)), \tag{3}$$

for $(\tau_i, \tau'_i) \in \mathcal{A}, i = 1, 2$.

The following conditions are assumed to hold.

- A1. $\{X_t; t \in \mathbb{Z}\}$ is a strictly stationary and strongly mixing process with α -mixing coefficients satisfying $\alpha(n) = O(n^{-a})$, for $a > 1$.
- A2. The marginal distribution $F(\cdot)$ has continuous density $f(\cdot)$, which is bounded away from 0 and ∞ at $q(\tau)$ over $\tau \in \mathcal{A}_t \cup \mathcal{A}_{t+l}$.
- A3. For any $\varepsilon > 0$ there exists a $\nu(\varepsilon)$ such that $\sup_{\tau \in \mathcal{A}_t \cup \mathcal{A}_{t+l}} \sup_{|s| \leq \nu(\varepsilon)} |f(q_\tau) f(q_\tau + s)| < \varepsilon$.
- A4. The joint distribution $F_l(\cdot, \cdot)$ is continuously differentiable over the neighborhood of quantiles of interest.

Assumptions A1–A4 are mild regularity conditions and not too restrictive. While A1 entails a mixing condition for the dependence structure of X_t , A2 ensures that the quantile function is uniquely determined, and A3 and A4 impose enough smoothness and regularity for f and F_l , respectively. The weak convergence of the sample quantile autocovariance processes indexed by $(\tau, \tau') \in (0, 1)^2$ is stated in [Theorem 2.1](#) below.

Theorem 2.1. *Suppose that assumptions A1–A4 hold for a particular lag l . Then we have*

$$\sqrt{T} (\hat{\gamma}_l(\tau, \tau') - \gamma_l(\tau, \tau')) \implies \lambda_{l,(\tau,\tau')}^t \mathbb{B}_l(\tau, \tau')$$

with

$$\lambda_{l,(\tau,\tau')} = \text{diag} \left(1, \frac{1}{f(q_\tau)}, \frac{1}{f(q_{\tau'})} \right) \left(\nabla F_l(q_\tau, q_{\tau'}) \right), \tag{4}$$

where $\nabla F_l(q_\tau, q_{\tau'})$ denotes the gradient vector of $F_l(\cdot, \cdot)$ computed at $(q_\tau, q_{\tau'})$ and $\mathbb{B}_l(\tau, \tau')$ is the above-mentioned zero-mean Gaussian process with covariance matrix given by (3).

The convergence stated in [Theorem 2.1](#) is established proceeding as in the proof of Theorem 1 of Han et al. [39], but for the convenience of the reader, the main steps of the proof are outlined in [Appendix A](#).

2.3. QAF-based dissimilarity

The great sensitivity of QAF to capture complex dynamic features also suggests high capability to discriminate between generating processes, and hence an interesting potential to be applied on clustering and classification problems [26]. To illustrate this point, we have obtained the sample QAF and the sample ordinary autocovariances for series simulated from a Gaussian white noise process, a GARCH-type process and an exponential GARCH process with Gaussian innovations. Plots of the sample autocovariance function and $\hat{\gamma}_1(\tau, \tau')$, for $\tau = 0.1, 0.5$ and 0.9 , are simultaneously depicted in [Fig. 1](#) for the three series.

As the three processes are uncorrelated, the sample autocovariances in (a) are close to zero with differences simply due to the noise, and therefore the conventional autocovariances are not useful to discriminate between the generating processes. By contrast, QAF plots in panels (b)–(d) show structural differences enabling us to discriminate between the underlying processes. The graphs for the white noise are flat due to the independence, but this is not the case for the GARCH models, which are uncorrelated but not independent. For instance, the symmetry of the GARCH model produces a flat profile for $\hat{\gamma}_1(0.5, \cdot)$, indicating that if $\{X_t \leq q_{0.5}\}$ then $\{X_{t+1} \leq q_{0.5}\}$ and $\{X_{t+1} > q_{0.5}\}$ are events with equal probability. However, the asymmetry of the EGARCH model leads to a different profile for $\hat{\gamma}_1(0.5, \cdot)$ indicating that X_{t+1} likely takes values higher than X_t . On the other hand, unlike of the white noise, the heavy tails of the GARCH model are recognizable from $\hat{\gamma}_1(0.1, \cdot)$ and $\hat{\gamma}_1(0.9, \cdot)$ since large and small values at time t tend to remain that way at time $t + 1$. In short, this simple example involving GARCH processes brings insight into the potential of the quantile autocovariances to detect distinct underlying processes, providing a more comprehensive understanding on the dependence structure than the traditional autocovariances.

These considerations strongly support the idea of measuring dissimilarity between a pair of times series $X_t^{(1)}$ and $X_t^{(2)}$ by comparing estimates of their quantile autocovariances over a common range of selected quantiles, such as we propose in [26]. Specifically, each time series $X_t^{(u)}$, $u = 1, 2$, is characterized by means of the vector $\Gamma^{(u)}$ constructed as follows. For prefixed ranges of L lags, l_1, \dots, l_L , and r quantile levels, $0 < \tau_1 < \dots < \tau_r < 1$, the vector $\Gamma^{(u)}$ is given by

$$\Gamma^{(u)} = \left(\Gamma_{l_1}^{(u)}, \dots, \Gamma_{l_L}^{(u)} \right), \tag{5}$$

where each $\Gamma_{l_i}^{(u)}$, $i = 1, \dots, L$, consists of a vector of length r^2 formed by re-arranging by rows the elements of the $r \times r$ matrix

$$\left(\hat{\gamma}_{l_i}^{(u)}(\tau_j, \tau_{j'}) \right)_{j,j'=1,\dots,r}, \tag{6}$$

being $\hat{\gamma}$ the sample quantile autocovariance given in (2). This way, the dissimilarity between $X_t^{(1)}$ and $X_t^{(2)}$ is defined as the squared Euclidean distance between the corresponding representations $\Gamma^{(1)}$ and $\Gamma^{(2)}$, i.e.

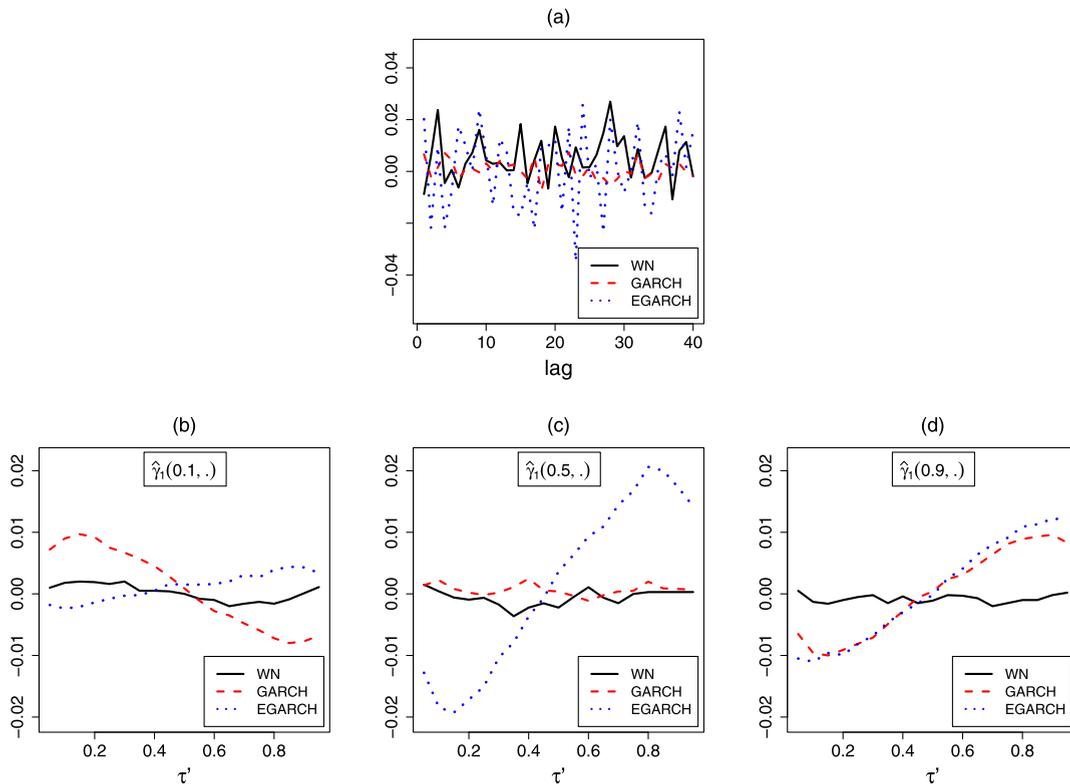


Fig. 1. Sample autocovariances (a) and sample quantile autocovariances $\hat{\gamma}_1(\tau, \tau')$ for $\tau = 0.1$ (b), 0.5 (c) and 0.9 (d), obtained from simulated realizations of a Gaussian white noise process, a GARCH-type process and an exponential GARCH with Gaussian innovations.

$$d_{QAF}(X_t^{(1)}, X_t^{(2)}) = \|\Gamma^{(1)} - \Gamma^{(2)}\|^2 = \sum_{i=1}^L \sum_{j=1}^r \sum_{j'=1}^r (\hat{\gamma}_{li}^{(1)}(\tau_j, \tau_{j'}) - \hat{\gamma}_{li}^{(2)}(\tau_j, \tau_{j'}))^2 \quad (7)$$

By definition, d_{QAF} belongs to the class of dissimilarities based on comparing features extracted of the series instead of directly comparing the observed series. Other authors have proposed feature-based dissimilarities considering distances between simple or partial autocorrelations [12,44,45], ARMA representations [46–48], periodograms or log-periodograms [49], cepstral coefficients [16] and other spectral features [8,50], among others. Obviously, all of these dissimilarities take advantage from the properties of the considered feature, and thus d_{QAF} inherits the nice properties of the quantile autocovariances. In particular, as mentioned in Subsection 2.1, the quantile autocovariance function is able to capture many types of serial dependence (including models with zero autocorrelation or presenting tail dependence) and exhibits robustness against outliers and heavy tails. From a practical point of view, it is also worthy remarking that d_{QAF} presents an efficient implementation at a very low cost in terms of computing time. Further, by construction, d_{QAF} can be evaluated on time series with unequal length. All of these interesting properties suggest that d_{QAF} has an enormous potential to perform time series clustering, and the results in this paper will corroborate this fact throughout a broad simulation study considering partitional cluster analysis, with both crisp and fuzzy clustering procedures.

3. Optimal selection of input parameters for clustering

According to definition (7), computation of d_{QAF} requires setting a number of input parameters, namely the number L of significant lags and the set of quantile levels $\{\tau_1, \dots, \tau_r\}$. Since our target is to use this metric to perform time series clustering, our concern is to determine how many and which combinations of lags and quantile levels $(l_i, \tau_j, \tau_{j'})$ must be considered to optimize the clustering process. The light computational complexity of d_{QAF} enables us to employ a reasonably large number of lags and quantiles without a significant loss of efficiency. Nev-

ertheless, working with a large set of inputs does not necessarily improve the clustering performance. In practice, introducing non-significant lags or very close quantiles means to supply noise to the classification process, thus generating worse results. Therefore, the goal is simple: starting from a preset grid of input parameters, determining a reasonably small subset that conveys the more relevant information on the underlying clustering structure.

To reach this goal, we follow a novel approach proposed by Andrews and McNicholas [33]. In a general context, Andrews and McNicholas introduce a variable selection stepwise algorithm for clustering and classification (called VSCC) based on determining the variables that simultaneously minimize the within-group variance and maximize the between-group variance. Indeed, if the variables have been standardized to have the same variance, then the minimization of the within-group variance also implies the maximization of the between-group variance. Besides this criterion in terms of “within” and “between” variances, the algorithm imposes that the correlation between the selected variables drops below a threshold. The purpose is to ignore highly correlated variables, which does not provide new information and may introduce noise. The correlation threshold is not a prefixed value but a sliding threshold allowed to be larger as the within-group variance is small. Specifically, if S denotes the subset of selected variables at a particular step of the algorithm, then a new variable s is added to S if for all $r \in S$ we have

$$|\rho_{sr}| < 1 - \mathcal{W}_s^\alpha \quad (8)$$

where ρ_{sr} is the correlation between the variables s and r , \mathcal{W}_s denotes the within-group variance for the variable s , and α is a preset parameter determining the shape of the relationship between the within-group variance and the between-variable correlation. Note that the smaller α , more stringent is the correlation threshold.

Compared to other variable selection techniques in clustering, the VSCC algorithm is intuitive, competitive and computationally efficient [33]. Based on these arguments, we decided to adapt this algorithm to address the optimal selection of lags and quantiles in order to perform clustering using d_{QAF} .

Consider a set of n realizations of time series subjected to clustering. Starting from a grid of r regularly spaced quantile levels and a number L of lags, an initial set of vectors of length Lr^2 , $\Gamma^{(u)}$ for $u = 1, \dots, n$, is computed according to (5), i.e. the observed series are replaced by vectors of estimated quantile autocovariances. These vectors are arranged by rows in a $n \times Lr^2$ matrix, \mathbf{A} , whose columns represent the variables used to perform clustering. The dissimilarity between two series $X^{(u)}$ and $X^{(v)}$ is given by the squared Euclidean distance between the rows u and v of \mathbf{A} , that is $d_{QAF}(X_t^{(u)}, X_t^{(v)}) = \|\Gamma^{(u)} - \Gamma^{(v)}\|^2$, and the clustering procedure relies on this dissimilarity criterion (details are given in Sections 4 and 5 later).

Therefore, we start with Lr^2 variables characterized by combinations of a lag and a pair of quantile levels $(l_i, \tau_j, \tau_{j'})$, and our intention is to apply the VSCC algorithm to obtain an optimal selection of these combinations. First, the columns of \mathbf{A} are standardized to have zero mean and unit variance, which allows us to concentrate our attention on minimizing the within-group variance. Then, the VSCC procedure is carried out as follows.

- Step 1.* Set the initial grid of r regularly spaced quantile levels and L lags, the number C of clusters and the value of α governing the relationship (8).
- Step 2.* Perform a partitional clustering procedure of the set of time series based on the matrix $\mathbf{A} = (\Gamma^{(u)t})_{1 \leq u \leq n}$ of estimated quantile autocovariances and d_{QAF} (see Sections 4 and 5 for hard and fuzzy algorithms, respectively).
- Step 3.* For each column of \mathbf{A} , i.e. for each combination $(l_i, \tau_j, \tau_{j'})$, with $i = 1, \dots, L$ and $j, j' = 1, \dots, r$, compute the within-group variance $\mathcal{W}_{(l_i, \tau_j, \tau_{j'})}$ defined by

$$\mathcal{W}_{(l_i, \tau_j, \tau_{j'})} = \frac{1}{n} \sum_{c=1}^C \sum_{s=1}^n z_{sc} \left(\hat{\gamma}_{l_i}^{(s)}(\tau_j, \tau_{j'}) - \bar{\gamma}_{l_i}^{(c)}(\tau_j, \tau_{j'}) \right)^2,$$

where z_{sc} is the group membership indicator function and $\bar{\gamma}_{l_i}^{(c)}(\tau_j, \tau_{j'})$ is the average of the corresponding estimated quantile autocovariances over the group c , that is $\bar{\gamma}_{l_i}^{(c)}(\tau_j, \tau_{j'}) = \frac{1}{n} \sum_{s=1}^n z_{sc} \hat{\gamma}_{l_i}^{(s)}(\tau_j, \tau_{j'})$. By dealing with a hard partition, $z_{sc} = I(\Gamma^{(s)} \in c)$ is the indicator function taking the value 1 if the series s belongs to

the cluster c and 0 otherwise. In the case of a fuzzy partition, z_{sc} takes the value 1 for the cluster c where the series s presents the highest membership degree and 0 otherwise.

- Step 4.* Sort $\mathcal{W}_{(l_i, \tau_j, \tau'_j)}$ in ascending order. Denote this sorted list by $\mathbf{W} = \{\mathcal{W}_{(1)}, \dots, \mathcal{W}_{(Lr^2)}\}$ and the combination of lag and quantile levels associated with $\mathcal{W}_{(k)}$ by $(l_{i_k}, \tau_{j_k}, \tau'_{j_k})$.
- Step 5.* $\mathcal{W}_{(1)}$ minimizes \mathbf{W} and hence $(l_{i_1}, \tau_{j_1}, \tau'_{j_1})$ is automatically placed into the subset of selected variables, denoted by S . Set count $k = 2$.
- Step 6.* If $|\rho_{sk}| < 1 - \mathcal{W}_{(2)}^\alpha$, for all $s \in S$, then the combination associated with $\mathcal{W}_{(2)}$, $(l_{i_2}, \tau_{j_2}, \tau'_{j_2})$ is added to S .
- Step 7.* While $k < Lr^2$, set $k = k + 1$ and return to Step 6. Then end algorithm.

In our experiments, we have considered up to five different values for α , namely $\alpha = 1, \dots, 5$, exactly as proposed in [33]. Overall, the most stringent choice $\alpha = 1$ led to a small number of variables and good clustering behavior. In the following section, the results from a sensitivity analysis conducted to test the efficacy of the proposed procedure against the use of different sequences of quantile levels are presented.

4. Partitioning around medoids clustering based on quantile autocovariances

This section is devoted to examine the behavior of d_{QAF} in hard clustering assuming that the grouping principle is to bring together series with the same generating process, thus the clustering task is governed by the dynamic of the series. Our previous work [26] tackled this issue by considering hierarchical clustering procedures, and d_{QAF} led to highly competitive results compared to other alternative metrics. Now, we extend this analysis to cover partitioning-based clustering methods. Assuming the existence of C clusters and starting from an initial partition, these methods proceed by iteratively relocating objects between clusters until an optimal partition is attained. At each iteration, C cluster centers (usually referred to as prototypes or centroids) are estimated and a reassignment of objects based on the updated centers is carried out. The most popular partitioning-based algorithm is the k -means procedure, where the centroids are the means of objects in the clusters and the objective is to minimize the within-cluster squared error. Nevertheless, k -means is not a proper choice in our framework because the average of quantile autocovariances does not necessarily characterize a time series model. For instance, if ARMA or GARCH models are considered, then there are no guarantees that the centroids represent one of these models. In fact, the resulting centroid could not satisfy the constraints required on the coefficients defining these models. This way the centroids may be “fictitious” time series, which leads to serious drawbacks. First, the distance d_{QAF} between observed time series objects and centroids could not be properly defined. On the other hand, time series clustering is often aimed at finding “representative” time series for each cluster, let us say a set of C patterns summarizing the different underlying dynamics, and again this is not guaranteed and the resulting centroids could fail in providing a suitable characterization of the cluster dynamics. A natural way to overcome these drawbacks is to perform a k -medoids-based algorithm where the prototypes are restricted to be chosen among the data points. The goal is to find k representative objects minimizing the average dissimilarity of all objects to their closest representative object. This way, d_{QAF} (or whatever is the selected distance) directly determines the efficacy of the clustering. In fact, the k -medoid algorithms can be run using the pairwise distances without requiring the data records. Unlike the k -means procedure, where optimization involves minimization of variance and therefore a L_2 analysis, the k -medoids-based algorithms are L_1 methods and therefore more robust to outliers and noise. Regarding these nice properties, we have carried out an extensive simulation study using the standard version of the well-known PAM algorithm (from “Partitioning Around Medoids”) [27], which is currently available in R code.

Simulations were conducted to assess the performance of d_{QAF} compared to a wide selection of model-free and model-based dissimilarity measures, and considering up to three different classification setups, namely classification of (i) ARMA models, (ii) non-linear models, and (iii) several structures of conditional heteroskedasticity. The generating models selected at each case are enumerated below.

Scenario 4.1 Classification of ARMA processes.

- (a) AR(1) $X_t = 0.9X_{t-1} + \varepsilon_t$
- (b) MA(1) $X_t = -0.7\varepsilon_{t-1} + \varepsilon_t$
- (c) AR(2) $X_t = 0.3X_{t-1} - 0.1X_{t-2} + \varepsilon_t$
- (d) MA(2) $X_t = 0.8\varepsilon_{t-1} - 0.6\varepsilon_{t-2} + \varepsilon_t$
- (e) ARMA(1,1) $X_t = 0.8X_{t-1} + 0.2\varepsilon_{t-1} + \varepsilon_t$

Scenario 4.2 Classification of non-linear processes.

- (a) NLMA $X_t = -0.5\varepsilon_{t-1} + 0.8\varepsilon_{t-1}^2 + \varepsilon_t$
- (b) EXPAR $X_t = [0.3 - 10 \exp(-X_{t-1}^2)]X_{t-1} + \varepsilon_t$
- (c) TAR $X_t = 0.5X_{t-1}I(X_{t-1} \leq 0) - 2X_{t-1}I(X_{t-1} > 0) + \varepsilon_t$

To make more complex the clustering task in this scenario, we have added series generated from the following linear model.

- (d) MA $X_t = -0.4\varepsilon_{t-1} + \varepsilon_t$

Scenario 4.3 Classification of conditional heteroskedastic processes. Consider the linear model $X_t = 0.5a_{t-1} + a_t$, with the error term satisfying $a_t = \sigma_t\varepsilon_t$, where the variance at time t , σ_t^2 , is conditional on observations at $t - 1$ by means of some of the following models.

- (a) ARCH $\sigma_t^2 = 0.2 + 0.95a_{t-1}^2$
- (b) GARCH $\sigma_t^2 = 0.2 + 0.05a_{t-1}^2 + 0.9\sigma_{t-1}^2$
- (c) GJR–GARCH $\sigma_t^2 = 0.2 + (0.05 + 1.2N_{t-1})a_{t-1}^2 + 0.1\sigma_{t-1}^2$, with $N_{t-1} = I(a_{t-1} < 0)$

Likewise the above scenario, we included a linear model MA(1) given by

- (d) MA $X_t = 0.5\varepsilon_{t-1} + \varepsilon_t$

The error process ε_t consisted of independent and identically distributed variables following different distributions, namely Gaussian innovations with unit variance, Student- t innovations with 1 degree of freedom, and exponential Exp(0.75) innovations. Using these distributions, we intend to assess the behavior of the clustering procedure also when kurtosis or skewness are present.

While clustering of linear models (Scenario 4.1) has been intensively studied and there are metrics specifically designed to deal with this kind of models, Scenario 4.2 introduces a major difficulty by including models with different conditional means that gradually depart from linearity. The models involved in Scenario 4.1 were previously considered by Maharaj [47] by performing clustering of ARMA processes, and the models in Scenario 4.2 were used in a linearity test context by Tong and Yeung [51]. Scenario 4.3 proposes a more challenging task by involving models with non-constant volatility. The autoregressive conditional heteroskedasticity models ARCH and GARCH are able to capture both time-varying volatility clustering and some amount of fat-tailedness of the distribution, features frequently exhibited for returns on assets. Unlike the GARCH models, the Glosten–Jagannathan–Runkle GARCH (GJR–GARCH) models allow to capture asymmetric effects on the conditional variance due to positive or negative past values, taking into account the leverage effect observed in many financial series.

As far as the dissimilarities to be compared, our selection must take into account the clustering purpose. We are not interested in measuring proximity between geometric profiles of series, thus shape-based dissimilarities (e.g. L_p distances) are not useful here because of clustering would be governed by local fluctuations, that is by the noise. Our aim is to bring together series generated from the same model. Hence, the selected metrics must capture differences between high level dynamic structures, which describe the global performance of the series. We decided to examine a wide range of dissimilarities, including measures comparing: estimated autocorrelations and partial autocorrelations, cross-correlations, periodograms, nonparametric spectral estimators, fitted ARMA models and cepstral coefficients, among others. All of these measures were computed using the R package **TSclust** [6]. Because of space reasons, we limit our report to the set of dissimilarities producing the best results in our numerical experiments, which are enumerated below.

- *Periodogram-based distances* [49]. Euclidean distances between periodograms, log-periodograms, normalized periodograms and log-normalized periodograms were checked, reporting in this section the results for the Euclidean distance between log-periodograms, denoted by d_{LP} .

- *Autocorrelation-based distances* [49]. Direct and weighted Euclidean distances between simple and partial autocorrelations using a number of significant lags were taken into consideration. Here we present the results considering the weighted Euclidean distance between partial autocorrelations (d_{PACFG}) based on a number of 10 lags and with weights ω_i decaying with the lag in the form $\omega_i = \pi(1 - \pi)^i$, with $\pi = 0.5$.
- *Model-based distances*. The AR metric introduced by Maharaj [47] and denoted by d_M .
- *Nonparametric dissimilarities in the frequency domain*. Although several metrics were considered within this group, we focus on the results attained with the integrated squared difference between estimated log-spectra (d_{ISD}) proposed by P ertega and Vilar [8].

Note that given the parametric models chosen to set up the simulation scenarios, it is expected that some of the selected distances work fine in Scenario 4.1. All of them were compared with the proposed metric d_{QAF} . The quantile levels were determined by means of the variable selection algorithm VSCC introduced in Section 3. As a starting point, the VSCC algorithm was implemented over a grid of regularly spaced quantile levels formed by all the combinations $(0.05j, 0.05j')$, with j and j' ranging from 1 to 19.

For each scenario, five time series of equal length T were generated from each model, thus providing a sample set of labeled series available to perform clustering. The experiments were carried out for three different series lengths, $T = 200, 500$ and 1000 . Note that all models are stationary in mean but they present differences in scale. To avoid that these differences dominate the clustering, the series were previously normalized to have unit variance.

Each set of simulated series was subjected to partitional clustering using the PAM algorithm together with each of the studied metrics. The inputs were the true number C of clusters, the pairwise dissimilarity matrix, and the initial medoids randomly selected among all the series. The output was the resulting partition, let us say $\mathcal{R} = \{R_1, \dots, R_C\}$. The quality of the clustering procedure was evaluated by means of two indexes of agreement between the true cluster partition, $\mathcal{T} = \{T_1, \dots, T_C\}$, and the experimental partition \mathcal{R} . Note that, according to the clustering target, each element T_i in \mathcal{T} is a cluster formed by all the series generated from the same model, and hence the true partition is known. The two selected criteria take into account this fact and are described below.

The first agreement index [1,52] is defined by

$$Ind_1(\mathcal{T}, \mathcal{R}) = \frac{1}{C} \sum_{i=1}^C \max_{1 \leq j \leq C} Ind_1(T_i, R_j), \quad (9)$$

where

$$Ind_1(T_i, R_j) = \frac{2|T_i \cap R_j|}{|T_i| + |R_j|},$$

and $|V|$ denotes the cardinality of a set V . Index Ind_1 accounts for the number of series sharing a same cluster in both partitions, taking exactly the value 0 if both partitions are completely dissimilar and the value 1 if they are identical. The second index, Ind_2 , is the adjusted Rand index [53], a corrected-for-chance version of the Rand index [54] which computes the proportion of pairs of series that are located together in the same or different clusters for both partitions. The adjusted Rand index modifies the Rand index in such a way that its expected value is equal to zero when the partitions are picked up at random (according to a generalized hypergeometric model) and the number of series in the clusters remain fixed. Likewise Ind_1 , the maximum value of Ind_2 is 1 and it is attained when partitions agree perfectly. Nevertheless, the adjusted Rand index typically takes values substantially lower than other agreement indexes, even occasionally negative values, and it is known to exhibit a greater sensitivity on the cluster stability than other indexes.

Besides Ind_1 and Ind_2 , we have also calculated a third index (Ind_3) using the one-nearest-neighbor (1-NN) classifier evaluated by leave-one-out cross-validation. Ind_3 returns the proportion of series correctly classified when each series has been assigned to the element of \mathcal{T} containing the nearest series according to the considered dissimilarity. Notice that Ind_3 does not evaluate the clustering algorithm, but providing insight into the efficacy of the used dissimilarity. This evaluation criterion has been intensively used in a broad range of pattern recognition applications, including time series clustering (see e.g. [7]).

The simulation procedure was replicated $N = 100$ times for each scenario and the obtained indexes were averaged over the 100 trials. The averages and standard deviations (in brackets) obtained for the different lengths of the series are reported in Tables 1, 2 and 3, including results for the considered innovation distributions.

Table 1
Indexes of clustering quality in the Monte-Carlo simulation with series of length $T = 250$.

Dissimilarity	Scenario 4.1			Scenario 4.2			Scenario 4.3		
	Ind_1	Ind_2	Ind_3	Ind_1	Ind_2	Ind_3	Ind_1	Ind_2	Ind_3
<i>Gaussian innovations</i>									
d_{LP}	0.729 (.098)	0.596 (.102)	0.736 (.092)	0.483 (.069)	0.109 (.106)	0.381 (.110)	0.414 (.046)	-0.002 (.062)	0.204 (.097)
d_{PACFG}	0.873 (.070)	0.775 (.098)	0.883 (.077)	0.714 (.088)	0.444 (.138)	0.683 (.114)	0.437 (.061)	0.017 (.073)	0.233 (.105)
d_M	0.875 (.102)	0.811 (.139)	0.937 (.068)	0.749 (.095)	0.554 (.145)	0.789 (.132)	0.428 (.051)	0.018 (.066)	0.258 (.095)
d_{ISD}	0.906 (.063)	0.821 (.095)	0.908 (.070)	0.726 (.094)	0.517 (.137)	0.722 (.117)	0.421 (.052)	0.027 (.067)	0.256 (.100)
d_{QAF}	0.767 (.090)	0.600 (.100)	0.709 (.117)	0.995 (.014)	0.986 (.039)	0.997 (.014)	0.622 (.070)	0.268 (.078)	0.496 (.109)
<i>Non-symmetric innovations</i>									
d_{LP}	0.741 (.083)	0.593 (.079)	0.742 (.107)	0.522 (.059)	0.263 (.103)	0.438 (.120)	0.509 (.101)	0.181 (.144)	0.413 (.114)
d_{PACFG}	0.876 (.073)	0.779 (.103)	0.884 (.070)	0.627 (.089)	0.401 (.091)	0.560 (.137)	0.625 (.103)	0.398 (.166)	0.619 (.120)
d_M	0.888 (.092)	0.817 (.126)	0.945 (.054)	0.631 (.093)	0.433 (.078)	0.623 (.156)	0.569 (.089)	0.286 (.145)	0.599 (.145)
d_{ISD}	0.896 (.062)	0.802 (.089)	0.911 (.058)	0.586 (.086)	0.412 (.082)	0.614 (.131)	0.598 (.079)	0.359 (.133)	0.691 (.117)
d_{QAF}	0.830 (.060)	0.728 (.071)	0.838 (.087)	1.000 (.000)	1.000 (.000)	1.000 (.000)	0.663 (.084)	0.409 (.103)	0.575 (.150)
<i>Heavy-tailed innovations</i>									
d_{LP}	0.614 (.043)	0.514 (.050)	0.714 (.104)	0.416 (.076)	0.070 (.122)	0.535 (.107)	0.483 (.051)	0.138 (.076)	0.391 (.103)
d_{PACFG}	0.872 (.097)	0.815 (.124)	0.953 (.051)	0.723 (.046)	0.556 (.077)	0.834 (.084)	0.424 (.059)	0.050 (.069)	0.351 (.087)
d_M	0.838 (.087)	0.772 (.111)	0.965 (.037)	0.738 (.073)	0.576 (.110)	0.866 (.093)	0.438 (.064)	0.062 (.083)	0.367 (.100)
d_{ISD}	0.743 (.100)	0.641 (.111)	0.905 (.060)	0.652 (.086)	0.428 (.146)	0.757 (.105)	0.481 (.055)	0.144 (.067)	0.458 (.101)
d_{QAF}	0.830 (.067)	0.735 (.084)	0.846 (.087)	0.998 (.009)	0.996 (.024)	0.999 (.007)	0.676 (.077)	0.392 (.110)	0.600 (.123)

In the case of Gaussian innovations, the dissimilarity based on quantile autocovariances d_{QAF} led to the highest average scores in clustering of non-linear and heteroskedastic models, Scenarios 4.2 and 4.3 respectively. In fact, the results in these scenarios based on d_{QAF} were substantially better than the ones obtained with the rest of metrics for the three indexes. In the linear framework (Scenario 4.1), d_{QAF} produced reasonably high indexes although a little worse than d_{PACFG} , d_M and d_{ISD} . It is worthy to point out the outstanding behavior of d_{QAF} in clustering of non-linear models, with average agreement indexes always above 0.985 for the smallest length, and exactly 1 with series of length 1000 for all the experiments (since $sd = 0$). Beyond the efficacy of the PAM algorithm, the scores of the 1-NN classifier (Ind_3) illustrate the high capability of d_{QAF} to discriminate between these processes, fairly superior to the other metrics. The worst clustering results were obtained in the Scenario 4.3, thus showing the complexity of clustering heteroskedastic structures. Only d_{QAF} is able to draw out good classification rates in this hard clustering framework, specially with large series. Note even that, except for d_{QAF} , the clustering results do not improve as the length of the series increases.

Similar conclusions derive from the results obtained with non-symmetric and heavy-tailed disturbances, although d_{QAF} reported additional nice properties. First, d_{QAF} was again the best-performed metric in Scenarios 4.2 and 4.3,

Table 2

Indexes of clustering quality in the Monte-Carlo simulation with series of length $T = 500$.

Dissimilarity	Scenario 4.1			Scenario 4.2			Scenario 4.3		
	Ind_1	Ind_2	Ind_3	Ind_1	Ind_2	Ind_3	Ind_1	Ind_2	Ind_3
<i>Gaussian innovations</i>									
d_{LP}	0.804 (.093)	0.674 (.115)	0.812 (.100)	0.530 (.070)	0.188 (.107)	0.442 (.122)	0.406 (.046)	-0.002 (.048)	0.206 (.088)
d_{PACFG}	0.949 (.066)	0.899 (.107)	0.951 (.052)	0.858 (.091)	0.694 (.152)	0.862 (.078)	0.428 (.051)	0.014 (.069)	0.249 (.103)
d_M	0.952 (.085)	0.928 (.118)	0.977 (.041)	0.905 (.103)	0.812 (.171)	0.922 (.074)	0.416 (.046)	0.020 (.067)	0.253 (.118)
d_{ISD}	0.952 (.047)	0.896 (.084)	0.958 (.044)	0.869 (.108)	0.748 (.170)	0.895 (.078)	0.405 (.039)	0.016 (.057)	0.246 (.125)
d_{QAF}	0.855 (.089)	0.730 (.131)	0.832 (.099)	0.999 (.005)	0.999 (.014)	1.000 (.000)	0.714 (.075)	0.459 (.135)	0.643 (.117)
<i>Non-symmetric innovations</i>									
d_{LP}	0.775 (.083)	0.637 (.092)	0.789 (.096)	0.556 (.079)	0.342 (.101)	0.459 (.116)	0.554 (.098)	0.263 (.135)	0.495 (.111)
d_{PACFG}	0.953 (.056)	0.903 (.092)	0.955 (.051)	0.678 (.107)	0.458 (.120)	0.660 (.130)	0.657 (.075)	0.463 (.141)	0.742 (.111)
d_M	0.956 (.081)	0.934 (.113)	0.985 (.027)	0.714 (.124)	0.544 (.135)	0.744 (.140)	0.521 (.072)	0.226 (.119)	0.681 (.120)
d_{ISD}	0.945 (.054)	0.888 (.089)	0.960 (.042)	0.664 (.099)	0.492 (.104)	0.770 (.137)	0.647 (.070)	0.436 (.119)	0.797 (.090)
d_{QAF}	0.877 (.084)	0.791 (.117)	0.916 (.062)	1.000 (.000)	1.000 (.000)	1.000 (.000)	0.777 (.109)	0.592 (.146)	0.737 (.154)
<i>Heavy-tailed errors</i>									
d_{LP}	0.620 (.042)	0.523 (.050)	0.727 (.101)	0.435 (.089)	0.100 (.140)	0.553 (.133)	0.503 (.045)	0.175 (.047)	0.419 (.103)
d_{PACFG}	0.934 (.091)	0.900 (.126)	0.984 (.030)	0.757 (.080)	0.611 (.117)	0.890 (.070)	0.415 (.052)	0.040 (.051)	0.379 (.077)
d_M	0.901 (.105)	0.860 (.141)	0.987 (.023)	0.756 (.079)	0.611 (.124)	0.607 (.062)	0.421 (.051)	0.046 (.055)	0.388 (.082)
d_{ISD}	0.784 (.128)	0.704 (.154)	0.942 (.051)	0.658 (.101)	0.445 (.161)	0.810 (.079)	0.503 (.045)	0.180 (.041)	0.507 (.099)
d_{QAF}	0.877 (.083)	0.796 (.111)	0.897 (.078)	1.000 (.000)	1.000 (.000)	1.000 (.000)	0.726 (.088)	0.523 (.125)	0.731 (.125)

increasing the average quality indexes with respect to the Gaussian setting. Particularly noteworthy was the improvement with heavy-tailed innovations where significantly high scores are now attained. The rest of metrics presented different behaviors in these two scenarios, exhibiting an improvement with heavy-tailed innovations (although fairly below d_{QAF}) but becoming substantially worse with non-symmetric innovations. Thus, we can conclude that asymmetry has an important influence over these metrics. Lastly, except for d_{QAF} , all the metrics are affected by asymmetry and kurtosis when classifying ARMA models, particularly the non-parametric dissimilarity d_{ISD} . On the contrary, d_{QAF} presents better results, being very close to the best metrics also in the linear scenario.

In short, our numerical experiments illustrate the good performance of the proposed metric in partitional clustering for a wide range of time series models. Specifically, d_{QAF} outperformed the rest of analyzed metrics in clustering of non-linear and heteroskedastic models, and exhibited clustering quality indexes reasonably high with linear models although somewhat lower than the ones obtained with metrics designed to deal with these models. Furthermore, unlike the rest of metrics, quality of the clustering results based on d_{QAF} showed robustness to the kind of disturbance distribution.

Table 3
Indexes of clustering quality in the Monte-Carlo simulation with series of length $T = 1000$.

Dissimilarity	Scenario 4.1			Scenario 4.2			Scenario 4.3		
	Ind_1	Ind_2	Ind_3	Ind_1	Ind_2	Ind_3	Ind_1	Ind_2	Ind_3
<i>Gaussian innovations</i>									
d_{LP}	0.843 (.084)	0.727 (.104)	0.864 (.084)	0.575 (.088)	0.267 (.133)	0.512 (.120)	0.418 (.052)	0.013 (.057)	0.225 (.073)
d_{PACFG}	0.995 (.018)	0.988 (.040)	0.994 (.016)	0.976 (.038)	0.936 (.091)	0.978 (.041)	0.420 (.050)	0.015 (.065)	0.252 (.104)
d_M	0.998 (.009)	0.995 (.023)	0.998 (.011)	0.989 (.031)	0.973 (.063)	0.987 (.032)	0.426 (.051)	0.044 (.060)	0.273 (.107)
d_{ISD}	0.986 (.027)	0.965 (.060)	0.991 (.019)	0.979 (.033)	0.945 (.082)	0.977 (.037)	0.416 (.046)	0.032 (.055)	0.300 (.106)
d_{QAF}	0.926 (.070)	0.845 (.118)	0.920 (.069)	1.000 (.000)	1.000 (.000)	1.000 (.000)	0.765 (.058)	0.605 (.083)	0.716 (.105)
<i>Non-symmetric innovations</i>									
d_{LP}	0.855 (.084)	0.746 (.114)	0.877 (.093)	0.388 (.057)	0.053 (.117)	0.305 (.082)	0.584 (.094)	0.315 (.144)	0.543 (.127)
d_{PACFG}	0.991 (.022)	0.978 (.050)	0.992 (.023)	0.726 (.114)	0.552 (.142)	0.812 (.094)	0.657 (.063)	0.476 (.125)	0.832 (.106)
d_M	0.997 (.023)	0.995 (.033)	0.996 (.016)	0.781 (.149)	0.674 (.175)	0.920 (.073)	0.518 (.063)	0.217 (.107)	0.750 (.091)
d_{ISD}	0.979 (.032)	0.949 (.070)	0.990 (.022)	0.692 (.120)	0.564 (.118)	0.856 (.079)	0.658 (.074)	0.448 (.124)	0.875 (.080)
d_{QAF}	0.931 (.082)	0.876 (.126)	0.953 (.050)	1.000 (.000)	1.000 (.000)	1.000 (.000)	0.890 (.084)	0.777 (.128)	0.861 (.111)
<i>Heavy-tailed innovations</i>									
d_{LP}	0.615 (.044)	0.526 (.043)	0.740 (.101)	0.426 (.084)	0.086 (.134)	0.554 (.153)	0.520 (.043)	0.199 (.039)	0.459 (.113)
d_{PACFG}	0.972 (.069)	0.959 (.096)	0.996 (.012)	0.805 (.104)	0.670 (.153)	0.928 (.059)	0.411 (.047)	0.035 (.045)	0.389 (.090)
d_M	0.941 (.093)	0.919 (.125)	0.998 (.010)	0.759 (.082)	0.612 (.123)	0.926 (.057)	0.412 (.053)	0.033 (.048)	0.402 (.088)
d_{ISD}	0.789 (.112)	0.708 (.131)	0.948 (.046)	0.665 (.098)	0.453 (.159)	0.837 (.092)	0.519 (.043)	0.201 (.039)	0.540 (.087)
d_{QAF}	0.957 (.068)	0.920 (.107)	0.963 (.056)	1.000 (.000)	1.000 (.000)	1.000 (.000)	0.702 (.083)	0.515 (.110)	0.766 (.126)

In addition, a sensitivity analysis experimenting with different sequences of regularly spaced quantile levels was conducted, including a comparison with the results based on the variable selection VSCC algorithm. This way, we intend to analyze the effect of the selection of quantile levels on the clustering results. Table 4 reports the averages of the cluster similarity indexes obtained from 100 trials of the simulation procedure for Scenarios 4.1, 4.2 and 4.3, with Gaussian innovations, series of length $T = 250$, and the metric d_{QAF} based on the following combinations of quantile levels: (i) $\tau_1 = (0.1, 0.5, 0.9)$, (ii) $\tau_2 = (0.1, 0.3, 0.5, 0.7, 0.9)$, and (iii) $\tau_3 = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$.

Table 4 reveals that the results get better as the number of quantiles is increased. Nevertheless, no large differences are observed and acceptable results are reached using only three quantile levels (τ_1). Except for Ind_3 in Scenario 4.3, the VSCC algorithm leads to the highest scores, with the additional advantage of determining a proper trade-off between number of quantile levels and clustering quality on the basis of an objective criterion. Anyway, results from Table 4 suggest that d_{QAF} should produce satisfactory results in clustering with a small number of regularly spaced quantile levels. Although this is indeed a noteworthy property, it is also worth remarking that d_{QAF} is computationally efficient because of an increase in the number quantiles does not mean a substantial cost in terms of computing time.

Table 4
Influence of the selection of quantile levels.

Vector of quantile levels	Scenario 4.1			Scenario 4.2			Scenario 4.3		
	<i>Ind</i> ₁	<i>Ind</i> ₂	<i>Ind</i> ₃	<i>Ind</i> ₁	<i>Ind</i> ₂	<i>Ind</i> ₃	<i>Ind</i> ₁	<i>Ind</i> ₂	<i>Ind</i> ₃
τ_1	0.766	0.583	0.688	0.986	0.961	0.989	0.595	0.231	0.535
τ_2	0.753	0.586	0.696	0.993	0.982	0.996	0.607	0.256	0.537
τ_3	0.767	0.600	0.716	0.993	0.982	0.997	0.610	0.251	0.560
VSCC	0.767	0.600	0.709	0.995	0.986	0.997	0.622	0.268	0.496

5. Fuzzy clustering of time series based on quantile autocovariances

So far, we have analyzed the behavior of the QA-based distance in hard partitional clustering, where the observed time series are split into *C* mutually exclusive groups. By dealing with realizations of time series, a fuzzy clustering approach might provide a solution much more attractive and interpretable. Fuzzy clustering allows the objects to belong simultaneously to several groups. More precisely, a fuzzy partition associates to each object a label vector with *C* membership degrees, ranging from zero to one and indicating the partial membership to each group. This way, an object located on the boundary between several groups may be assigned to all of them with different confidence levels, but without necessarily forcing it to be in only one of these groups. In the time series framework, besides enriching the cluster solution by recognizing the vague nature of the prototypes, the adoption of a fuzzy approach allows the time series to share different dynamic patterns. This could be perfectly supported by changes in the dynamic of the series or by the existence of time series exhibiting patterns with intermediate features respect to several clusters [12,13]. If a suitable distance is able to capture differences between the underlying dynamic patterns, then a fuzzy clustering algorithm based on such a distance should gain adaptivity to construct the prototypes, and hence to obtain a better characterization of the temporal pattern of the series (see discussion in [12]). Therefore, the versatility of the fuzzy logic combined with the high capability of *d*_{QAF} to discriminate between different generating processes motivate the study of a fuzzy clustering model based on this distance, such as we carry out throughout the present section.

5.1. QAF-based fuzzy C-medoids clustering model (QAF-FCMdC model)

As in previous sections, consider a set *S* of *n* realizations of univariate time series $\{X_t^{(1)}, \dots, X_t^{(n)}\}$ and denote by $\Gamma = \{\Gamma^{(1)}, \dots, \Gamma^{(n)}\}$ the corresponding vectors of estimated quantile autocovariances computed as defined in (5). Assume that all vectors $\Gamma^{(i)}$ have the same length *Lr*², being *L* and *r* the numbers of lags and quantile levels considered for all the series, respectively. This way, the pairwise *d*_{QAF} distances between two arbitrary series can be computed according to (7). In this framework, we propose to perform partitional fuzzy clustering on *S* by means of the QAF-based Fuzzy *C*-Medoids Clustering model (QAF-FCMdC), which aims at finding the subset of Γ of size *C*, $\tilde{\Gamma} = \{\tilde{\Gamma}^{(1)}, \dots, \tilde{\Gamma}^{(C)}\}$, and the *n* × *C* matrix of fuzzy coefficients $\Omega = (u_{ic})$, *i* = 1, ..., *n*, *c* = 1, ..., *C*, that lead to solve the minimization problem:

$$\min_{\tilde{\Gamma}, \Omega} \sum_{i=1}^n \sum_{c=1}^C u_{ic}^m \left\| \Gamma^{(i)} - \tilde{\Gamma}^{(c)} \right\|^2, \text{ subject to the constraints: } \sum_{c=1}^C u_{ic} = 1 \text{ and } u_{ic} \geq 0, \tag{10}$$

where *u*_{ic} ∈ [0, 1] represents the membership degree of the *i*-th series in the *c*-th cluster, $\tilde{\Gamma}^{(c)}$ is the vector of quantile autocovariances associated to the medoid series for the cluster *c*, and *m* > 1 is a weighting exponent that controls the fuzziness of the partition. Constraints on *u*_{ic} are standard requirements in fuzzy clustering. In particular, that the sum of the membership degrees for each series equals 1 implies that all of them contribute with the same weight to the clustering process. Parameter *m* determines the level of fuzziness introduced in the clustering procedure. In the naive case *m* = 1, we have *u*_{ic} = 1 if the *i*-th series is the medoid for the cluster *c* and 0 otherwise so that the crisp version of the procedure is obtained. As the value of *m* increases, the boundaries between clusters become softer and therefore the classification is fuzzier.

In a nutshell, the aim of QAF-FCMdC model is to determine a fuzzy partition into *C* clusters such that the QA-distance between the clusters and their prototypes is minimized. Likewise the crisp approach in Section 4, the

clustering quality strongly depends on the capability of d_{QAF} to identify different dependence structures, but now the non-stochastic uncertainty inherent to the assignment of series to clusters is incorporated to the procedure by means of the membership degrees.

An iterative algorithm that alternately optimizes the membership degrees and the medoids is used to solve the optimization problem in (10). First, the membership degrees are optimized for a set of fixed medoids. The iterative solutions for the membership degrees take the form [55]:

$$u_{ic} = \left[\sum_{c'=1}^C \left(\frac{\|\Gamma^{(i)} - \tilde{\Gamma}^{(c)}\|^2}{\|\Gamma^{(i)} - \tilde{\Gamma}^{(c')}\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}, \quad \text{for } i = 1, \dots, n \text{ and } c = 1, \dots, C. \quad (11)$$

Then, based on the membership degrees obtained from (11), the C series minimizing (10) are selected as new medoids. This two-step procedure is iterated until there is no change in the medoids or a maximum number of iterations is achieved. The QAF-based fuzzy C -medoids clustering algorithm (QAF-FCM dC) is implemented as outlined in Algorithm 1.

Algorithm 1 The QAF-based Fuzzy C -Medoids Clustering Algorithm (QAF-FCM dC).

- 1: Fix C , m and $max.iter$
 - 2: Set $iter = 0$
 - 3: Pick the initial medoids $\tilde{\Gamma} = \{\tilde{\Gamma}^{(1)}, \dots, \tilde{\Gamma}^{(C)}\}$
 - 4: **repeat**
 - 5: Set $\tilde{\Gamma}_{OLD} = \tilde{\Gamma}$ {Store the current medoids}
 - 6: Compute u_{ic} , $i = 1, \dots, n$, $c = 1, \dots, C$, using (11)
 - 7: For each $c \in \{1, \dots, C\}$, determine the index $j_c \in \{1, \dots, n\}$ satisfying:

$$j_c = \operatorname{argmin}_{1 \leq j \leq n} \sum_{i=1}^n u_{ic}^m \|\Gamma^{(i)} - \Gamma^{(j)}\|^2$$
 - 8: **return** $\tilde{\Gamma}^{(c)} = \Gamma^{(j_c)}$, for $c = 1, \dots, C$ {Update the medoids}
 - 9: $iter \leftarrow iter + 1$
 - 10: **until** $\tilde{\Gamma}_{OLD} = \tilde{\Gamma}$ or $iter = max.iter$
-

5.2. Simulation study

A new simulation study was conducted to evaluate the performance of the proposed QAF-FCM dC algorithm. We intended to recreate fuzzy scenarios with different time series models, including realizations of AR, ARCH and GARCH processes. In all cases, the base scenario consisted of two clusters with five series each, let us say \mathcal{C}_1 and \mathcal{C}_2 , and one additional time series located at equal distance from both clusters. Moreover, we add uncertainty to the classification procedure by two ways: (i) introducing variability over the parameters defining the underlying model for each cluster, and (ii) considering different levels of separation between the clusters. Variability within clusters was generated by drawing out the parameters at random according to uniform distributions with different support for each cluster. The distance between clusters is given by the distance between the means of the uniform distributions. The specific scenarios and the generation schemes for each scenario are described in detail in Table 5.

For all scenarios, innovations ε_t follow a Gaussian distribution with zero mean and unit variance. Compared to scenarios denoted by B, scenarios A exhibit greater distance between the clusters \mathcal{C}_1 and \mathcal{C}_2 , and hence less balanced memberships are expected. In other terms, the five series generated from one specific cluster should group all together with membership degrees more markedly close to one in scenarios A. As far as the time series located at an intermediate situation between \mathcal{C}_1 and \mathcal{C}_2 , it is expected that these series belong simultaneously to the two clusters showing membership degrees close to 0.5.

The QAF-FCM dC algorithm was compared with other fuzzy clustering models based on alternative dissimilarities. For Scenario 5.1, a fuzzy C -medoids algorithm considering Euclidean distances between estimated autoregressive representations was used. According to the fuzzy approach based on features extracted of the time series, a fuzzy model can be formalized as solution of the general optimization problem:

Table 5
Simulation scenarios for evaluation of the QAF-FCMdC algorithm.

Generating process	Scenario	Elements and structure
<i>Scenario 5.1: Fuzzy clustering of autoregressive processes AR(1)</i>		
$X_t = \phi X_{t-1} + \varepsilon_t$	5.1.A	Cluster \mathcal{C}_1 : 5 series with $\phi \sim U(0, 0.2)$ Cluster \mathcal{C}_2 : 5 series with $\phi \sim U(0.8, 1)$ One equidistant series with: $\phi = 0.5$
	5.1.B	Cluster \mathcal{C}_1 : 5 series with $\phi \sim U(0.2, 0.4)$ Cluster \mathcal{C}_2 : 5 series with $\phi \sim U(0.6, 0.8)$ One equidistant series with: $\phi = 0.5$
<i>Scenario 5.2: Fuzzy clustering of autoregressive conditional heteroskedastic processes ARCH(1)</i>		
$X_t = \sigma_t \varepsilon_t$, with $\sigma_t^2 = 0.1 + \alpha X_{t-1}^2$	5.2.A	Cluster \mathcal{C}_1 : 5 series with $\alpha \sim U(0, 0.1)$ Cluster \mathcal{C}_2 : 5 series with $\alpha \sim U(0.9, 1)$ One equidistant series with: $\alpha = 0.5$
	5.2.B	Cluster \mathcal{C}_1 : 5 series with $\alpha \sim U(0, 0.2)$ Cluster \mathcal{C}_2 : 5 series with $\alpha \sim U(0.8, 1)$ One equidistant series with: $\alpha = 0.5$
<i>Scenario 5.3: Fuzzy clustering of general autoregressive conditional heteroskedastic processes GARCH(1,1)</i>		
$X_t = \sigma_t \varepsilon_t$, with $\sigma_t^2 = 0.1 + \alpha X_{t-1}^2 + 0.1\sigma_{t-1}^2$	5.3.A	Cluster \mathcal{C}_1 : 5 series with $\alpha \sim U(0, 0.15)$ Cluster \mathcal{C}_2 : 5 series with $\alpha \sim U(0.85, 0.9)$ One equidistant series with: $\alpha = 0.5$
	5.3.B	Cluster \mathcal{C}_1 : 5 series with $\alpha \sim U(0.1, 0.2)$ Cluster \mathcal{C}_2 : 5 series with $\alpha \sim U(0.8, 0.9)$ One equidistant series with: $\alpha = 0.5$

$$\min : \mathcal{F}_m(\tilde{\Phi}, \Omega) = \sum_{i=1}^n \sum_{c=1}^C u_{ic}^m \left\| \Phi^{(i)} - \tilde{\Phi}^{(c)} \right\|^2, \text{ subject to the constraints: } \sum_{c=1}^C u_{ic} = 1 \text{ and } u_{ic} \geq 0, \quad (12)$$

where $\Phi^{(i)}$ represents the vector of estimated features for the i -th series, $i = 1, \dots, n$, and $\tilde{\Phi}$ denotes an arbitrary subset of C vectors $\Phi^{(i)}$ denoted by $\tilde{\Phi}^{(c)}$, $c = 1, \dots, C$. The subset $\tilde{\Phi}$ minimizing the objective function \mathcal{F}_m involves the solution with the C medoids or prototype time series. The solution is iteratively reached by optimizing alternately medoids and membership degrees. At each iteration, the membership degrees for fixed medoids are obtained by using the Lagrangian multipliers method, resulting the update formula:

$$u_{ic} = \left[\sum_{c'=1}^C \left(\frac{\left\| \Phi^{(i)} - \tilde{\Phi}^{(c)} \right\|^2}{\left\| \Phi^{(i)} - \tilde{\Phi}^{(c')} \right\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}, \text{ for } i = 1, \dots, n \text{ and } c = 1, \dots, C. \quad (13)$$

The use of different features in the class of fuzzy clustering models defined by (12) and (13) leads to distinct fuzzy algorithms. We have considered various features including single and partial autocorrelations but our best results were obtained by using autoregressive representations, which leads to the following fuzzy model.

- AR-FCMdC: Fuzzy C -medoids clustering model based on the AR-metric [56]. When the extracted features $\Phi^{(i)}$, $i = 1, \dots, n$, are the autoregressive representations of the time series, we take into consideration the distance introduced by Piccolo [46] to deal with ARIMA models [8,46,48,57,58]. Each series $X_t^{(i)}$ is identified by the AR(∞) operator approximating its ARIMA structure. In practice, the truncated AR(∞) representations are used, and thus $X_t^{(i)}$ is characterized by the vector of AR(r_i) parameter estimates, $\hat{\pi}^{(i)} = (\hat{\pi}_1^{(i)}, \dots, \hat{\pi}_{r_i}^{(i)})$, where the r_i significant lags are obtained by means of a model selection criterion such as Akaike's Information Criterion (AIC). Then $\Phi^{(i)} \equiv \pi^{(i)}$, and we have in (12) and (13):

$$\left\| \Phi^{(i)} - \tilde{\Phi}^{(c)} \right\|^2 = \sum_{u=1}^{r_{ic}} \left(\hat{\pi}_u^{(i)} - \tilde{\pi}_u^{(c)} \right)^2, \quad (14)$$

where $r_{ic} = \max(r_i, r_c)$. When $r_i \neq r_c$, the shortest AR coefficient vector is completed by adding zeros up to have two vectors with the same length.

Scenarios 5.2 and 5.3 involve conditionally heteroskedastic models and the clustering task is substantially more complex due to the peculiar features exhibited for these processes (results in experiments of Section 4 illustrate this assertion). As in Scenario 5.1, it is desirable to examine our procedure against fuzzy clustering models based on suitable distances regarding the underlying heteroskedastic structures. At this aim, we select two partitioning around medoids algorithms based on GARCH modeling recently proposed by D’Urso et al. [22]. Both models rely on distances employing the autoregressive representation of a GARCH(p, q) process. More precisely, the GARCH(p, q) model allows to model the serial dependence of the volatility by assuming that $X_t = \sigma_t \varepsilon_t$, where the innovations ε_t are independent and identically distributed variables and the squared disturbances σ_t^2 satisfy the following ARMA(p, q) representation:

$$\sigma_t^2 = \text{Var}(X_t | \mathcal{J}_{t-1}) = \gamma + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \tag{15}$$

where $\mathcal{J}_{t-1} = \sigma(X_{t-1}, X_{t-2}, \dots)$ represents the information available up to time $(t - 1)$, $\gamma > 0$, $0 \leq \alpha_i < 1$ and $0 \leq \beta_j < 1$, for $i = 1, \dots, p$ and $j = 1, \dots, q$, and $(\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j) < 1$. Based on expression (15), two distance measures between heteroskedastic processes are introduced and plugged into the fuzzy C -medoids clustering model as described below.

- GARCH-FCMdC: Fuzzy C -medoids clustering model based on the AR distance between GARCH approximations [22]. Starting from (15) and after some algebra, it can be shown that

$$X_t^2 = \gamma + \sum_{i=1}^{p^*} (\alpha_i + \beta_i) X_{t-i}^2 + \sum_{j=1}^q \beta_j \eta_{t-j} + \eta_t, \tag{16}$$

with $p^* = \max(p, q)$, $\alpha_i = 0$ for $i > p$, $\beta_i = 0$ for $i > q$, and $\eta_t = X_t^2 - \sigma_t^2$ a zero-mean error uncorrelated with the past. Equation (16) establishes an ARMA(p^*, q) representation for X_t^2 , which can be approximated by an AR(∞) structure with autoregressive coefficients π_u^G given by

$$\pi_u^G = (\alpha_u + \beta_u) + \sum_{j=1}^{\min(q, u)} \beta_j \pi_{u-j}^G$$

where $\pi_0^G = -1$, $\alpha_u = 0$ for $u > p$, and $\beta_u = 0$ for $u > q$. Then, GARCH-FCMdC model proceeds in the same line as AR-FCMdC but using estimators of these new autoregressive coefficients to compute the AR distance, i.e. replacing $(\hat{\pi}_1^{(i)}, \dots, \hat{\pi}_{r_i}^{(i)})$ by $(\hat{\pi}_1^{G, (i)}, \dots, \hat{\pi}_{r_i}^{G, (i)})$ in (14).

- GARCH-FCMdCC: Fuzzy C -medoids clustering model based on the Caiado and Crato distance between GARCH approximations [22]. Caiado and Crato [59] proposed an alternative approach to measure distance between GARCH models by taking into account the covariance between the fitted GARCH coefficients. Specifically, the distance between a pair of series $X_t^{(u)}$ and $X_t^{(v)}$ is defined by

$$d_{GARCH}(X_t^{(u)}, X_t^{(v)}) = (\mathbf{L}^{(u)} - \mathbf{L}^{(v)})^t (\mathbf{V}^{(u)} + \mathbf{V}^{(v)})^{-1} (\mathbf{L}^{(u)} - \mathbf{L}^{(v)}), \tag{17}$$

where $\mathbf{L}^{(u)} = (\hat{\boldsymbol{\alpha}}^{(u)}, \hat{\boldsymbol{\beta}}^{(u)})^t$ is the estimated vector of parameters in the GARCH representation (15) for $X_t^{(u)}$, and $\mathbf{V}^{(u)} = \mathbb{E}(\mathbf{L}^{(u)} \mathbf{L}^{(u), t})$ denotes the corresponding covariance matrix between the estimated parameters. Based on d_{GARCH} , the GARCH-FCMdCC model is derived by solving the minimization problem:

$$\min_{\tilde{\mathbf{L}}, \tilde{\boldsymbol{\Omega}}} \sum_{i=1}^n \sum_{c=1}^C u_{ic}^m (\mathbf{L}^{(i)} - \tilde{\mathbf{L}}^{(c)})^t (\mathbf{V}^{(i)} + \tilde{\mathbf{V}}^{(c)})^{-1} (\mathbf{L}^{(i)} - \tilde{\mathbf{L}}^{(c)}),$$

subject to the constraints: $\sum_{c=1}^C u_{ic} = 1$ and $u_{ic} \geq 0$,

where \tilde{L} is a subset of cardinality C of estimated GARCH vectors for the series in study.

By comparing QAF–FCMdC with AR–FCMdC in Scenario 5.1 and with GARCH–FCMdC and GARCH–FCMdCC in Scenarios 5.2 and 5.3, we choose competitors based on distances properly adjusted to the dependence structures from each simulated scenario, and therefore valuable insight into the usefulness of the proposed model and its robustness to the generating process should be obtained.

The experiments were carried out with different lengths for the time series, namely $T = 250, 500$ and 1000 for Scenario 5.1, and $T = 500, 1000$ and 2000 for Scenarios 5.2 and 5.3. The size of the series is increased for the heteroskedastic scenarios to face the high variability of the estimated GARCH parameters. Based on a controlled simulation experiment, Aielli and Caporin [15] assert that the standard quasi maximum likelihood GARCH estimates obtained from simulated realizations of a GARCH(1,1) process are characterized by higher dispersion for smaller sample sizes. These arguments account for choosing large sample sizes by treating with this kind of processes, and in fact the chosen lengths are commonly used in the literature [15,28,60]. Furthermore, large sample sizes are also usual in applications. A typical example of heteroskedastic series are the financial time series, which usually include longer sequences formed by daily or intra-daily data. Some experiments were also performed in Scenarios 5.2 and 5.3 with short series, but the results showed poor results, being particularly affected the fuzzy algorithms using GARCH-based distances due to inaccurate estimations of the GARCH parameters.

The fuzziness parameter m also has an important role, and in practice its value must be determined in advance. As already mentioned, $m = 1$ leads to a crisp partition, but very large values for m are not recommendable. Kamdar and Joshi [61] argue that very high values for m may imply to lose mobility of the medoids because all membership degrees would become very small except the one corresponding to the current medoid, which always equals 1 within its cluster. To our knowledge, there are no theoretical arguments supporting an optimal choice of m (see discussion in [62]). A popular choice is $m = 2$, although based on different heuristic arguments various authors suggest that the value of the proper level of fuzziness should be between 1.5 and 2.5 [63–65]. An interesting discussion on this point including related references is given in Section 3.1.6 of Maharaj and D’Urso [16]. We were very interested in checking the effect of moving the fuzzifier m , and based on the previous considerations we decided to take the values $m = 1.5, 2.0, 2.2$ and 2.5 , which is also a consistent choice with other recent experimental studies [16,66].

The number of clusters was set at $C = 2$, and hence the equidistant series are forced to belong simultaneously to both clusters. At all scenarios, ten sets of 100 simulations were carried out. For each set was first calculated the percentage of times in which time series were correctly classified, and then these success rates were averaged over the ten replications. At each trial, the correct classification occurs when the five series generated from the model defining C_1 are located together in one cluster, the five series coming from C_2 are grouped together in another cluster, and the single series generated from an equidistant model is simultaneously located in both clusters. Since grouping is performed in a fuzzy framework, a cut-off value for the membership degrees must be fixed to decide when a time series is assigned to a specific cluster or to both clusters simultaneously. Our assignment rule was to place the i -th series into the c -th if $u_{ic} > 0.7$. In other case, the series is simultaneously located in the two clusters because of its membership degrees are reasonably similar (both of them between 0.3 and 0.7). It is worthy remarking that the chosen cut-off point is compatible with the indications suggested in the literature [12,13,16,19,67,68].

The average percentages of correct classification were obtained with all the fuzzy models in order to be compared. In the case of the QAF–FCMdC model, the distance d_{QAF} between estimated quantile autocovariances was evaluated over a grid of regularly spaced quantile levels formed by all the combinations $(0.05j, 0.05j')$, with j and j' ranging from 1 to 19. Concerning the GARCH-based models, it is important to remark that the right number of GARCH parameters was provided as an input in the computation of d_{AR} and d_{GARCH} . Indeed, this is a substantial advantage in favor of these models since the significant number of GARCH parameters must be estimated in real scenarios. Table 6 shows the results for Scenario 5.1.

The influence of the fuzziness parameter m is evident from Table 6. The value $m = 1.5$ produced uniformly the worst percentages, and it is observed that the results seem to improve progressively when m increases. Note that using a high fuzzifier means to smooth the boundary between clusters, thus making more difficult to separate them. In particular, a reasonably high value for m implies a more uniform distribution of the membership degrees, thus

Table 6
Average percentage of correct classification in Scenario 5.1.

	Algorithm	Scenario 5.1.A			Scenario 5.1.B		
		$T = 250$	$T = 500$	$T = 1000$	$T = 250$	$T = 500$	$T = 1000$
$m = 1.5$	AR–FCMdC	31.5	45.1	52.0	17.2	26.6	31.5
	QAF–FCMdC	29.6	34.3	35.2	9.7	17.9	23.2
$m = 2.0$	AR–FCMdC	67.5	83.9	93.0	22.5	45.2	59.9
	QAF–FCMdC	69.1	76.2	77.6	28.7	44.1	58.3
$m = 2.2$	AR–FCMdC	76.9	91.8	97.5	21.1	44.3	65.2
	QAF–FCMdC	80.8	84.3	88.0	33.8	51.2	67.5
$m = 2.5$	AR–FCMdC	84.0	97.1	99.7	16.6	40.1	65.6
	QAF–FCMdC	88.8	93.7	96.1	34.1	56.0	75.7

Table 7
Average percentage of correct classification in Scenario 5.2.

	Algorithm	Scenario 5.2.A			Scenario 5.2.B		
		$T = 1000$	$T = 2000$	$T = 5000$	$T = 1000$	$T = 2000$	$T = 5000$
$m = 1.5$	GARCH–FCMdC	32.5	48.5	64.3	30.1	40.1	64.4
	GARCH–FCMdCC	20.4	17.8	10.1	20.4	21.8	20.0
	QAF–FCMdC	15.8	26.9	44.0	12.1	19.4	44.2
$m = 2.0$	GARCH–FCMdC	71.9	86.3	86.6	64.2	76.2	86.0
	GARCH–FCMdCC	47.1	51.3	53.4	48.5	51.2	53.1
	QAF–FCMdC	47.6	70.0	88.2	36.9	57.6	88.2
$m = 2.2$	GARCH–FCMdC	79.7	91.2	87.7	72.5	82.7	87.1
	GARCH–FCMdCC	56.5	65.7	69.3	56.9	62.0	70.7
	QAF–FCMdC	58.9	81.5	94.9	42.4	69.8	95.2
$m = 2.5$	GARCH–FCMdC	88.2	93.9	87.7	78.9	88.5	87.5
	GARCH–FCMdCC	70.6	82.4	85.2	67.6	75.4	84.4
	QAF–FCMdC	59.8	89.5	99.0	37.0	79.0	99.2

benefiting the correct classification of the equidistant series. As expected, the success rates substantially decreased for the Scenario 5.1.B. By increasing the level of proximity between clusters, both procedures are more sensitive to the noise and frequently the equidistant series present a membership degree $u_{ic} > 0.7$ for some c , thus producing an important number of failed trials and reducing noticeably the global success rate (even though the series of each cluster are really well-classified). Lastly, Table 6 also shows that QAF–FCMdC produces competitive results compared to AR–FCMdC. In spite of AR–FCMdC is designed to deal with ARMA series, the proposed algorithm QAF–FCMdC exhibited a similar performance, drawing out a little worse success rates in the easiest Scenario 5.1.A, but somewhat higher ones in the most difficult Scenario 5.1.B, where the clusters are closer each other.

Now, we focus on Tables 7 and 8 where the results using ARCH(1) (Scenarios 5.2) and GARCH(1,1) (Scenarios 5.3) processes are presented, respectively.

The above considerations on the effect of the fuzziness parameter m also apply in these scenarios, and it is observed that the success rates improve when m increases. In the simplest Scenario 5.2, involving ARCH(1) models, the GARCH-based algorithms outperform the QAF-based procedure when $T = 1000$. Nevertheless, by dealing with series of length $T = 2000$, QAF–FCMdC is clearly competitive, exhibiting better behavior than GARCH–FCMdCC and only a little worse than GARCH–FCMdC. As expected, again the success rates in Scenario 5.2.B are worse than the ones in Scenario 5.2.A. For the largest length of series ($T = 5000$) and regardless of the amount of separability of the clusters (Scenarios 5.2.A or 5.2.B), QAF–FCMdC is fairly the best procedure by attaining percentages of correct classification moving from 88% to 99%, while its competitors produced percentages always below 88%. Therefore, in spite of the GARCH-based algorithms take advantage of knowing the underlying parametric dependence structure (also the number of significant parameters in our simulations), QAF–FCMdC showed a similar behavior when high

Table 8
Average percentage of correct classification in Scenario 5.3.

	Algorithm	Scenario 5.3.A			Scenario 5.3.B		
		$T = 1000$	$T = 2000$	$T = 5000$	$T = 1000$	$T = 2000$	$T = 5000$
$m = 1.5$	GARCH-FCMdC	17.9	18.3	19.3	14.5	18.5	19.4
	GARCH-FCMdCC	5.0	13.5	15.0	5.3	11.9	14.1
	QAF-FCMdC	12.9	23.8	39.2	10.4	18.7	40.0
$m = 2.0$	GARCH-FCMdC	39.5	47.1	52.1	33.6	45.6	51.9
	GARCH-FCMdCC	4.7	15.9	33.5	4.5	11.5	27.3
	QAF-FCMdC	38.9	66.2	84.4	30.1	56.9	84.5
$m = 2.2$	GARCH-FCMdC	46.0	59.5	62.2	40.9	54.5	64.0
	GARCH-FCMdCC	3.6	9.3	32.2	3.3	7.4	22.4
	QAF-FCMdC	47.5	77.7	93.9	32.7	70.0	93.8
$m = 2.5$	GARCH-FCMdC	51.4	76.5	81.6	48.8	70.7	82.7
	GARCH-FCMdCC	2.7	2.8	20.2	1.9	3.4	15.7
	QAF-FCMdC	43.3	84.9	98.6	29.4	76.2	98.1

values of T were considered, and particularly excellent results (percentages of correct classification close to 100%) for series of length $T = 5000$.

This good performance of the proposed fuzzy algorithm is still more noticeable in Scenario 5.3, with GARCH(1, 1) models. In fact, Table 8 shows that QAF-FCMdC produced very similar (slightly lower) results as the ones obtained in Table 7, thus exhibiting an interesting robustness to the generating models. On the contrary, the fuzzy algorithms based on the GARCH metrics were strongly affected by the high variability in the estimation procedure of the GARCH parameters, corresponding to the GARCH-FCMdCC model the worst behavior. It must be noted that GARCH-FCMdCC requires the additional estimation of the covariance structure between GARCH parameters. The fuzzy approach based on the distance between quantile autocovariances is free of determining the underlying parametric structure and takes advantage of its enormous potential to detect complex types of dependence. These arguments account for the best results achieved by the proposed algorithm in this scenario. Only with $T = 1000$, GARCH-FCMdC seems to outperform QAF-FCMdC, but here is worthy to point out the importance of the length of the series in these heteroskedastic scenarios. In fact, the success rates for $T = 1000$ were substantially lower than using lengths 2000 and 5000, and always below 50%. Thus it is evident that larger lengths should be used. In short, although the good behavior of the QAF-based distance in clustering of heteroskedastic series was already observed by performing hard cluster analysis, the results presented in this section also illustrate how the fuzzy nature of time series presenting features intermediate between different conditionally heteroskedastic models is well-captured by a fuzzy algorithm based on d_{QAF} .

6. Applications

In this section, two study cases considering air quality data and daily returns of stocks are presented to illustrate the usefulness of the fuzzy C -medoids clustering algorithm based on quantile autocovariances. In both applications, results from different fuzzy clustering models are discussed and compared to obtain a valuable insight into the behavior of our proposal.

6.1. Application to air quality data

The first study case is related to the non-supervised classification of geographical zones in terms of their temporal records of air pollutants. Specifically, we have considered time series of daily averages of concentrations of nitrogen dioxide (NO_2) and ozone (O_3), from 1st November 2006 to 31st December 2009. All data are sourced from the official website of the Air Quality Monitoring Network of Madrid Community.¹

¹ http://gestiona.madrid.org/azul_internet/run/j/AvisosAccion.icm.

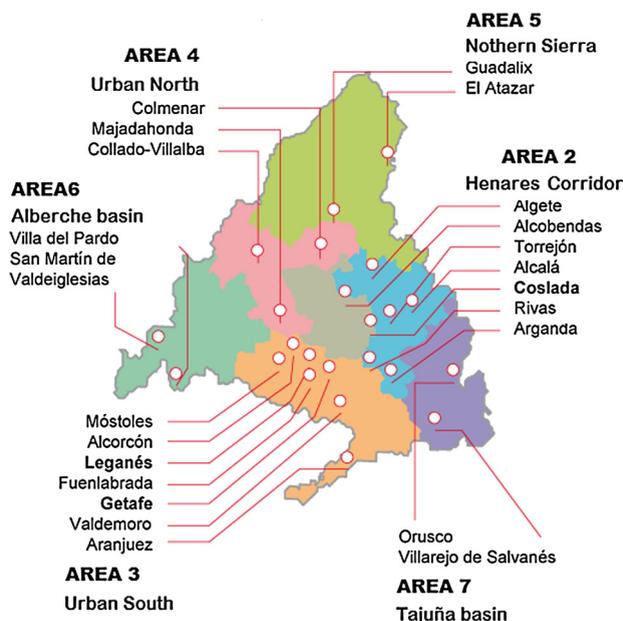


Fig. 2. Location of the stations forming the Air Quality Monitoring Network of Madrid Community.

For monitoring of emission levels, the Community of Madrid, which is an autonomous community of Spain, has a Network Control Air Quality consisting of a set of 21 fixed automatic stations and two mobile reference laboratories (a mobile unit and a bus). These 23 stations provide data on the air pollutant concentration along time and are distributed on 7 homogeneous areas that can be grouped in urban areas (Madrid, Urban North, Urban South and Henares Corridor) and rural areas (Northern Sierra, Alberche basin and Tajuña basin). We have used information extracted from 19 of the 23 stations (four stations were discarded because the database was not complete), namely Alcalá de Henares, Alcobendas, Torrejón de Ardoz, Arganda del Rey, Rivas Vaciamadrid, Leganés, Fuenlabrada, Móstoles, Aranjuez, Valdemoro, Majadahonda, Colmenar Viejo, Collado Villalba, Guadalix de la Sierra, El Atazar, S. Martín de Valdeiglesias, Villa del Prado, Villarejo de Salvanés and Orusco de Tajuña. Fig. 2 shows the geographical distribution of the stations forming the network.

Several studies have revealed serious health effects associated with the continuous exposure to high concentrations of nitrogen dioxide and ozone, and for this reason we have focused on them. While some works have considered the problem of checking by significant differences between the mean levels of these pollutants on different areas of the community (see e.g. [69]), our concern is to analyze the capability of the fuzzy clustering approach to identify locations with similar daily changes in levels of NO_2 and O_3 . Nevertheless, it is important to remark that our motivation is only to illustrate the usefulness of the proposed fuzzy algorithm, without seeking to give any type of environmental implications.

The 19 time series available are formed by $T = 1,154$ records and are non-stationary in mean, thus we proceeded to transform them taking one regular difference. Figs. 3 and 4 show plots of the transformed series for the levels of O_3 and NO_2 , respectively. It is observed that the variance is not constant over time. There are periods of time in which the series strongly fluctuate, while in others fluctuation is less marked. On the other hand, it is reasonable to think that a fuzzy behavior might be present, with time series sharing features of different and well-defined patterns of daily changes of concentrations of NO_2 and O_3 . According to our simulations results, the proposed fuzzy C -medoids clustering algorithm, QAF-FCM d_C , should lead to a proper fuzzy partition of the stations. Just as in simulations, distance d_{QAF} was computed by considering one lag ($L = 1$, with $l_1 = 1$) and a grid of quantile levels formed by all the combinations $(0.05j, 0.05j')$, with $j, j' \in \{1, \dots, 19\}$. For the purpose of comparison, the AR-FCM d_C and GARCH-FCM d_C fuzzy algorithms were also carried out. Given the underlying heteroskedasticity (particularly evident in the NO_2 case), the latter is expected to produce better results than the former.

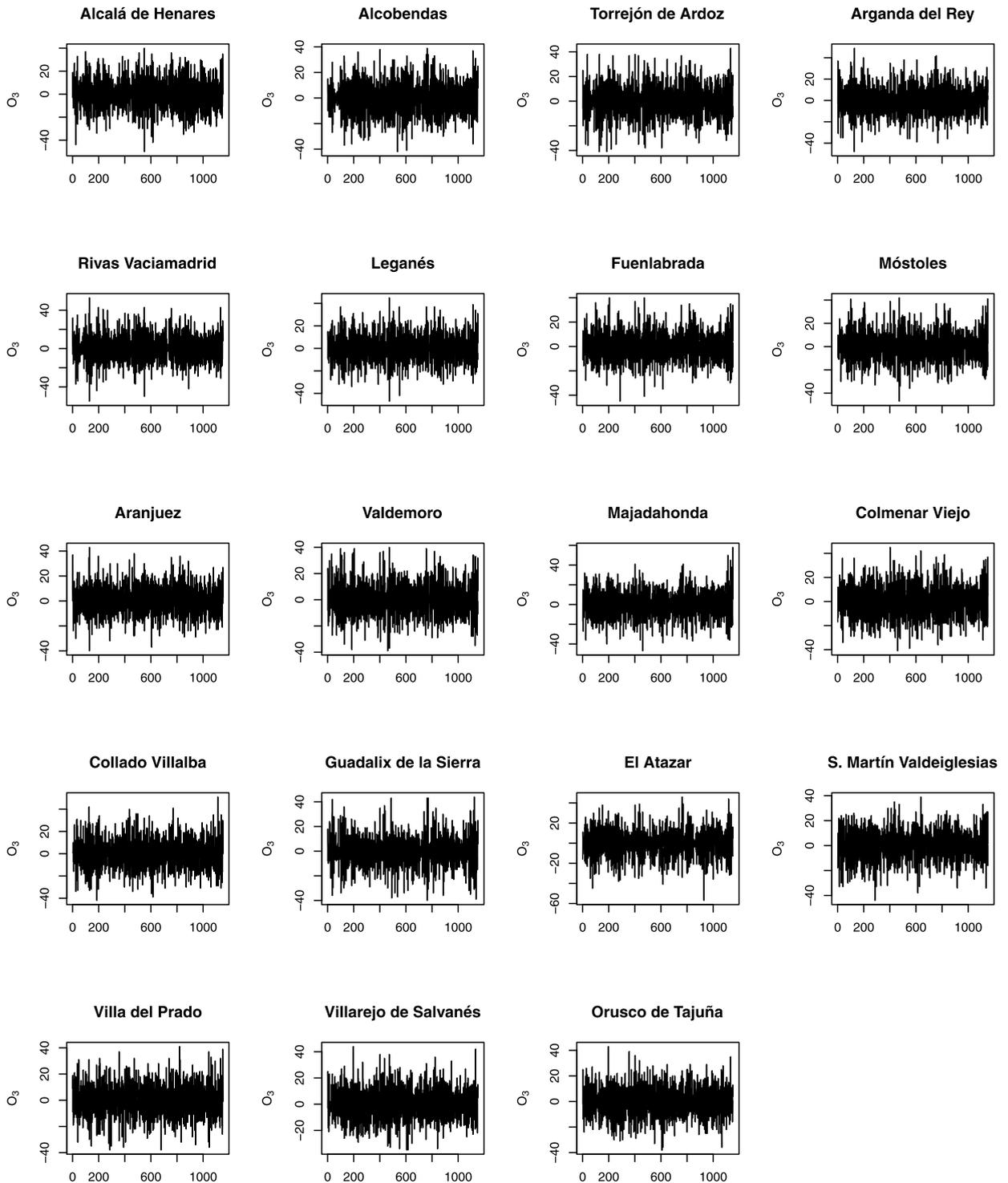


Fig. 3. Daily series of O_3 levels transformed by taking one regular difference.

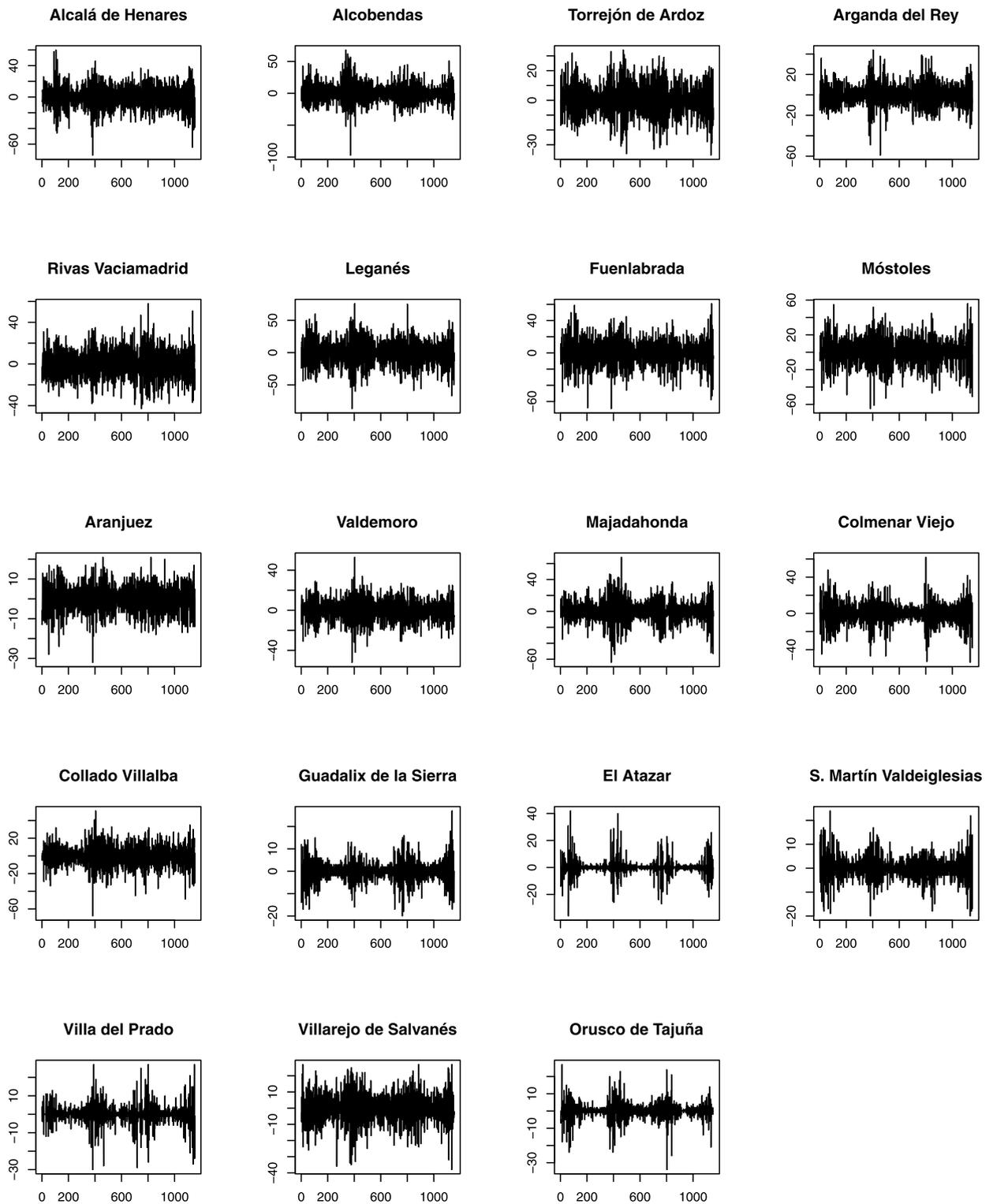


Fig. 4. Daily series of NO₂ levels transformed by taking one regular difference.

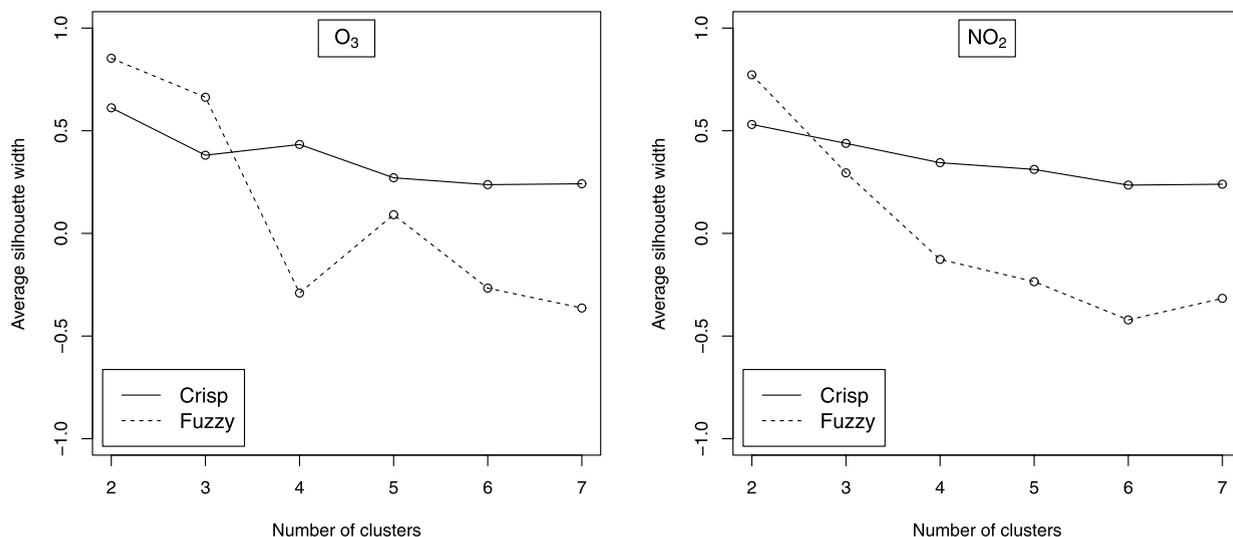


Fig. 5. Crisp and fuzzy silhouette width values for a different number partitions using QAF-FCM_dC.

A fuzzy extension of the classical silhouette width criterion was used to determine the optimal number of clusters. This fuzzy version takes into consideration the membership degrees matrix and consists in selecting the number of clusters maximizing the so-called Fuzzy Silhouette Width [70], defined by

$$FSW = \frac{\sum_{i=1}^n (u_{ir} - u_{iv})^\alpha s_i}{\sum_{i=1}^n (u_{ir} - u_{iv})^\alpha}$$

where s_i is the standard silhouette width for the i -th element, u_{ir} and u_{iv} are the first and the second largest elements of the i -th row of the fuzzy partition matrix and $\alpha \geq 0$ is a weighting coefficient. This way, FSW provides a weighted average of the individual silhouette widths, thus permitting to underweight series belonging to overlapping clusters. The value $\alpha = 1$ is commonly considered, and it was also used in our application.

Fig. 5 shows the values of the standard (crisp) and fuzzy silhouette indexes for a range of partition sizes using the QAF-FCM_dC algorithm. In all cases, the existence of two major groups is concluded. Focusing on the fuzzy approach, the highest FSW indexes were 0.854 for O₃ and 0.773 for NO₂, corresponding to partitions of two clusters in both cases. Note that these high values suggest a strong clustering structure. On the contrary, considering more than two clusters substantially reduces the values of the FSW indexes, particularly in the case of NO₂. On the other hand, a two-cluster partition is intuitively consistent with a natural grouping of the stations according to their location in urban or rural areas. Based on these arguments we decide to set the number of clusters at $C = 2$.

The 2-cluster solutions for the series of daily changes in levels of O₃ using a fuzziness parameter $m = 2$ are shown in Table 9. For each single series, the shaded cells enhance the highest membership degrees obtained with each procedure, i.e. the cluster assignments from a crisp perspective. Stations with both membership degrees within (0.3, 0.7) are fuzzy allocated between the two clusters and their memberships are shown in bold font.

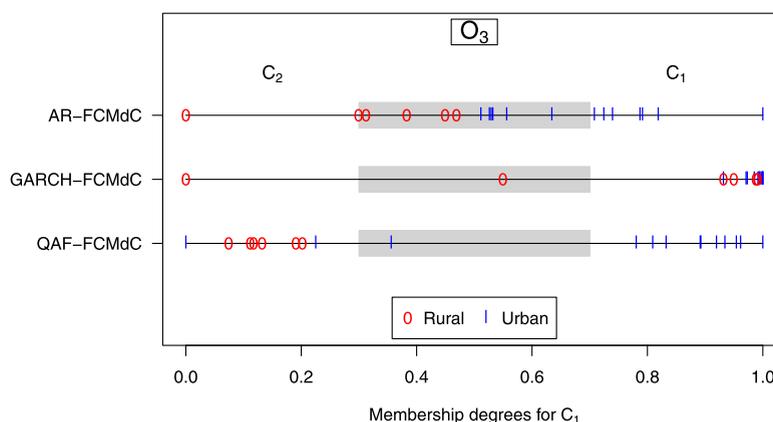
In essence, the model QAF-FCM_dC produces the expected classification by grouping the series of daily changes in O₃ according to the kind of location where they were monitored, i.e. stations placed in urban (cluster \mathcal{C}_1) and rural (cluster \mathcal{C}_2) areas. The group \mathcal{C}_2 brings together all the stations located in rural areas with memberships always above 0.798, but also including the stations of Aranjuez, Majadahonda and Colmenar Viejo. Actually, Aranjuez presents a vague location which might be explained because, despite being in an urban area, Aranjuez is located far from the rest of stations, just in the boundary of the Community. Also, in terms of ozone records, Majadahonda is set as a suburban location (see website of the Air Quality Monitoring Network of Madrid Community), which might account for its allocation in \mathcal{C}_2 . All the stations in cluster \mathcal{C}_1 belong to urban areas presenting very high membership degrees for this cluster.

The results obtained with the models AR-FCM_dC and GARCH-FCM_dC cannot be meaningfully interpreted, at least in terms of rural and urban locations. While the model GARCH-FCM_dC draws out a solution where just one cluster gathers almost all the series with memberships very close to one, the model AR-FCM_dC identifies the two

Table 9

Membership degrees in clustering of the daily change series in levels of O_3 ($C = 2$ and $m = 2.2$).

Station	Area	AR-FCMdc		GARCH-FCMdc		QAF-FCMdc	
		C_1	C_2	C_1	C_2	C_1	C_2
Alcalá de Henares	Urban	0.740	0.260	0.992	0.008	0.961	0.039
Alcobendas	Urban	0.787	0.213	1.000	0.000	0.891	0.109
Torrejón de Ardoz	Urban	0.819	0.181	0.973	0.027	0.809	0.191
Arganda del Rey	Urban	0.724	0.276	1.000	0.000	0.832	0.168
Rivas Vaciamadrid	Urban	0.708	0.292	0.971	0.029	0.893	0.107
Leganés	Urban	0.526	0.474	0.993	0.007	0.954	0.046
Fuenlabrada	Urban	1.000	0.000	0.985	0.015	1.000	0.000
Móstoles	Urban	0.511	0.489	0.998	0.002	0.920	0.080
Aranjuez	Urban	0.634	0.366	0.932	0.068	0.356	0.644
Valdemoro	Urban	0.791	0.209	1.000	0.000	0.934	0.066
Majadahonda	Urban	0.556	0.444	1.000	0.000	0.225	0.775
Colmenar Viejo	Urban	0.529	0.471	0.995	0.005	0.000	1.000
Collado Villalba	Urban	0.532	0.468	0.995	0.005	0.781	0.219
Guadalix de la Sierra	Rural	0.000	1.000	0.000	1.000	0.191	0.809
El Atazar	Rural	0.383	0.617	0.549	0.451	0.132	0.868
San Martín de Valdeiglesias	Rural	0.299	0.701	0.988	0.012	0.074	0.926
Villa del Pardo	Rural	0.312	0.688	0.932	0.068	0.202	0.798
Villarejo de Salvanes	Rural	0.449	0.551	0.991	0.009	0.117	0.883
Orusco de Tajuña	Rural	0.469	0.531	0.950	0.050	0.112	0.888

Fig. 6. Membership degrees for cluster C_1 in clustering of the daily changes in levels of O_3 .

areas but in a very fuzzy manner in most of the series. A simple way to visualize and compare the 2-cluster solutions obtained with the three models is provided by Fig. 6, where the membership degrees for cluster C_1 are depicted and the final assignment indicated.

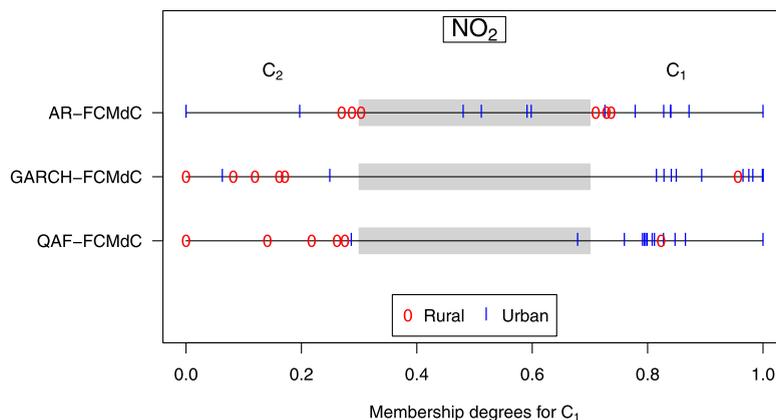
The 2-cluster solutions for the series of daily changes in levels of NO_2 are shown in Table 10 and Fig. 7. Note that GARCH-FCMdc and QAF-FCMdc lead to a very similar partition. With both models the distribution of the stations in rural and urban areas is still more evident than in the case of the ozone records. Among the rural locations, only Villarejo de Salvanes presents a pattern congruent with the urban locations, exhibiting the highest membership degree for C_1 with both algorithms. Again Majadahonda station is unexpectedly placed into C_2 and the only discrepancy is the classification of Colmenar Viejo. In contrast, AR-FCMdc again increases the fuzziness of the resulting partition, which is fairly no congruent with the grouping in urban and rural areas. It is worth noting that $C = 2$ is also the value maximizing the FSW index when AR-FCMdc is used, but in this case $FSW = 0.66$, substantially lower than the values 0.83 and 0.77 obtained with GARCH-FCMdc and QAF-FCMdc, respectively.

The main conclusions from this study case can be summarized as follows. Our fuzzy clustering approach, QAF-FCMdc, led to partitions with a meaningful interpretation for the two considered pollutants by grouping almost all

Table 10

Membership degrees in clustering of the daily change series in levels of NO₂ ($C = 2$ and $m = 2.2$).

Station	Area	AR-FCM _d C		GARCH-FCM _d C		QAF-FCM _d C	
		C_1	C_2	C_1	C_2	C_1	C_2
Alcalá de Henares	Urban	0.591	0.409	0.982	0.018	0.799	0.201
Alcobendas	Urban	0.512	0.488	0.850	0.150	0.866	0.134
Torrejón de Ardoz	Urban	0.480	0.520	0.965	0.035	1.000	0.000
Arganda del Rey	Urban	0.598	0.402	0.841	0.159	0.808	0.192
Rivas Vaciamadrid	Urban	0.197	0.803	0.828	0.172	0.791	0.209
Leganés	Urban	0.840	0.160	1.000	0.000	0.679	0.321
Fuenlabrada	Urban	0.872	0.128	1.000	0.000	0.812	0.188
Móstoles	Urban	1.000	0.000	0.815	0.185	0.794	0.206
Aranjuez	Urban	0.000	1.000	0.894	0.106	0.797	0.203
Valdemoro	Urban	0.726	0.274	0.975	0.025	0.828	0.172
Colmenar Viejo	Urban	0.778	0.222	0.249	0.751	0.760	0.240
Majadahonda	Urban	0.828	0.172	0.063	0.937	0.286	0.714
Collado Villalba	Urban	0.840	0.160	0.999	0.001	0.848	0.152
Guadalix de la Sierra	Rural	0.737	0.263	0.000	1.000	0.000	1.000
El Atazar	Rural	0.303	0.697	0.082	0.918	0.275	0.725
San Martín de Valdeiglesias	Rural	0.729	0.271	0.162	0.838	0.141	0.859
Villa del Pardo	Rural	0.710	0.290	0.120	0.880	0.218	0.782
Villarejo de Salvanés	Rural	0.287	0.713	0.957	0.043	0.823	0.177
Orusco de Tajuña	Rural	0.270	0.730	0.171	0.829	0.262	0.738

Fig. 7. Membership degrees for cluster C_1 in clustering of the daily changes in levels of O₃.

stations according to their urban or rural location. The approach based on GARCH approximations, GARCH-FCM_dC, performed in a similar way for the NO₂ series, but produced an unexpected and anomalous partition for the ozone records, thus showing less accuracy than QAF-FCM_dC. Since the observed series exhibit heteroskedasticity, the AR-FCM_dC approach relies on misspecified models, which might account for the obtained partitions, with lower quality indexes and hardly interpretable for the two air pollutants. Lastly, it is worth enhancing that all the procedures have determined some series showing fuzzy nature, which supports the usefulness of the fuzzy approach.

6.2. Application to daily stocks returns in IBEX-35 index

The second application considers daily returns of stocks included in the IBEX-35, which groups the thirty-five companies with the highest liquidity and trading volume in the Spanish stock market. Specifically, we manage a database formed by the daily returns of twenty-four stocks located in the TOP-30 ranking according to the finance

section of the Yahoo website.² The period of observation of the series spans from 1st January 2008 to 19th December 2016, thus resulting realizations of length $T = 2337$. The daily adjusted closing prices for all the stocks were sourced from the mentioned website and used to obtain the daily returns by considering the first differences of their natural logarithms. The time series are depicted in Fig. 8.

Heteroskedasticity is again observed although less pronounced for several stocks. The largest fuzzy silhouette widths are obtained with a partition in three clusters ($C = 3$) but with different strength for each model. While QAF-FCM_{dC} drawn out a value $FSW = 0.636$, the models AR-FCM_{dC} and GARCH-FCM_{dC} led to lower FSW indexes of 0.424 and 0.293, respectively. Therefore, all the models suggest the existence of three clusters but QAF-FCM_{dC} indicates in a more conclusive way that a well-defined cluster structure lies behind data. It is worthy noting that the best FSW is reached using $m = 1.5$ for AR-FCM_{dC}, while $m = 2$ is the fuzziness level producing the highest FSW for the GARCH- and QAF-based models. As in this application we lack of an intuitive idea on the underlying partition, we decided to corroborate our result by using an alternative index. Specifically, we calculated the Xie-Beni index [71] which is given by the ratio between the total variance and the minimum separation between clusters so that the optimal C is reached when this ratio is minimized. The minimum values of the Xie-Beni index corresponded to $C = 3$, with values 0.4804, 0.5537 and 0.6890 for QAF-FCM_{dC}, AR-FCM_{dC} and GARCH-FCM_{dC}, respectively, again concluding that a 3-cluster solution seems the most adequate and that QAF-FCM_{dC} produces the best-defined partition. Based on these arguments, cluster analysis using the three fuzzy models and setting $C = 3$ was carried out. The resulting membership degrees are shown in Table 11. As in previous application, the shaded cells enhance the highest membership degrees with each procedure and the stocks allocated in a fuzzy way between two or three clusters are indicated with memberships in bold font.

The 3-cluster solution generated by the QAF-FCM_{dC} model identifies a large cluster, C_1 , gathering together most of the stocks, including the ones of the sectors of Energy and Materials, Industry and Construction (except for Arcelormittal-MTE), and also the three banks with the highest capitalization level in the Financial services sector, namely BBVA, Santander-SAN and Caixabank-CABK. The cluster C_3 groups the company Arcelormittal-MTE together with the smaller banks Banco Popular-POP, Banco Sabadell-SAB and Bankinter-BKT, although SAB and BKT could be allocated in C_1 by exhibiting similar memberships for both clusters. The cluster C_2 groups together two important companies of the consumer goods industry (Viscofan-VIS and Inditex-ITX), the only insurance company (Mapfre-MAP), and a technological company related to the travel sector (Amadeus-AMS). In sum, the fuzzy partition provided by the QAF-FCM_{dC} model seems to be congruent with features like company size and business area. Nevertheless, our concern is not to obtain conclusions in financial terms such as searching proper model specifications or accurate predictions for the daily return series. These targets go beyond the scope of this work. Our motivation is to illustrate the capability of the proposed fuzzy clustering approach to identify similar dependence structures. In this sense, a relevant point by treating with daily returns is to analyze their dispersion, i.e. the underlying volatility patterns. To bring insight into this issue, nonparametric approximations of the variance between returns were obtained. The estimated volatility curves grouped according the three clusters identified with the QAF-FCM_{dC} model are depicted in Fig. 9.

Note that all the curves in Fig. 9 (a) present a very similar fluctuation pattern, with some bumps of different size in similar periods of time. The curves in Fig. 9 (b) corresponding to the cluster C_2 are characterized by a flat profile throughout the second half of the sample period. In fact, only Mapfre-MAP showed a few periods with moderate rise in the level of volatility. The cluster C_3 brings together stocks exhibiting a marked pickup in volatility in the last year, particularly Arcelormittal-MTE and Banco Popular-POP. This effect is less evident for Banco Sabadell-SAB, which could account for its vague allocation in this cluster (with a membership of 0.516). It is worthy noting that Arcelormittal-MTE presented a sharp rising of volatility during the first year and a half of the sample period, fairly above the rest of the analyzed stocks. This significant behavior might determine the atypical character of this time series. Overall, Fig. 9 allows us to describe representative volatility patterns for each of the clusters determined by the QAF-FCM_{dC} model.

The AR-FCM_{dC} model led to a reasonable cluster solution although a bit less intuitive than the one obtained with QAF-FCM_{dC}. For instance, it is unexpected that the companies Enagás-ENG, Iberdrola-IBE and Ferrovial-FER are separated from the rest of companies belonging to the same market sectors. Also, according to the volatility

² <https://finance.yahoo.com/>.

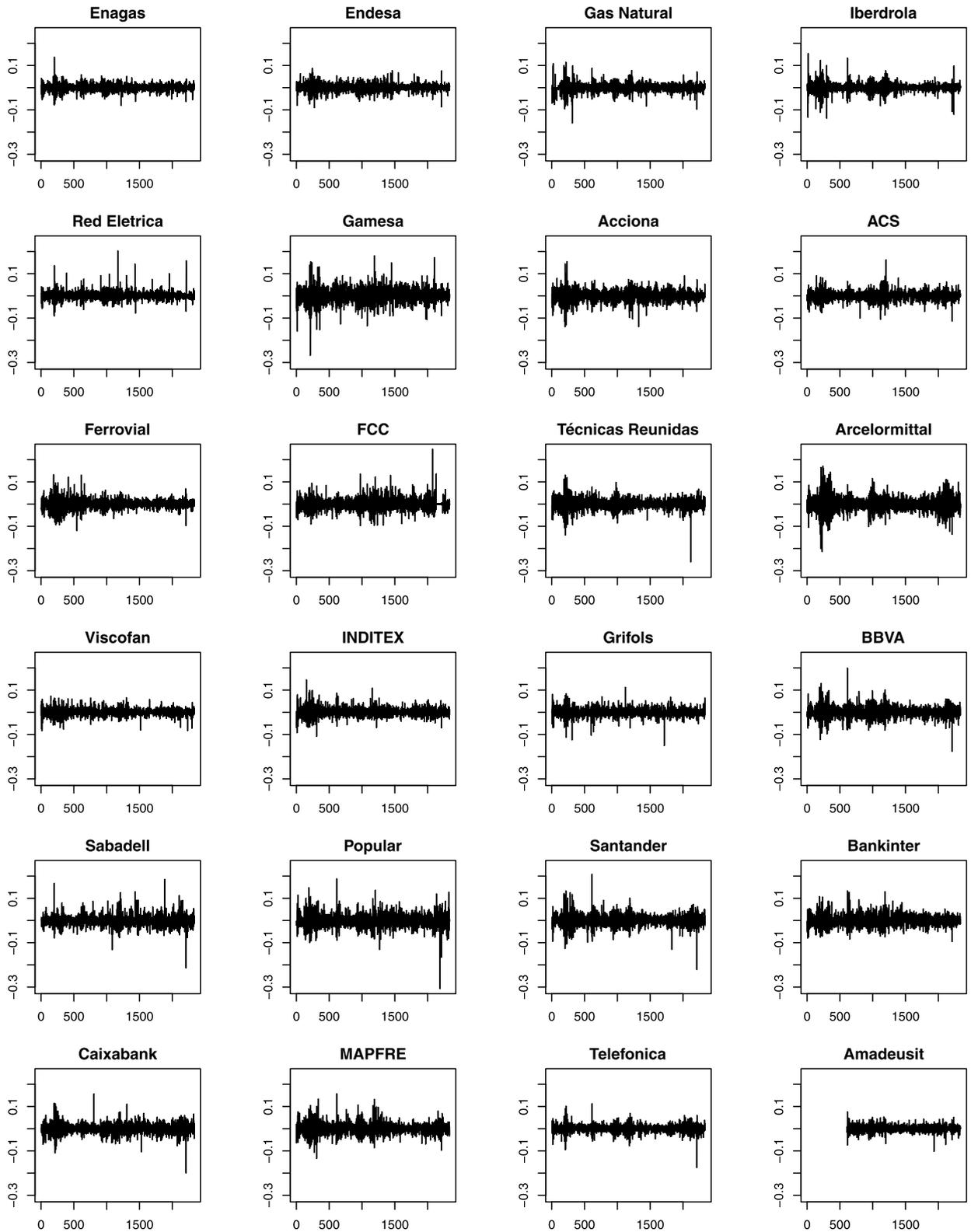


Fig. 8. Daily returns of 24 stocks included in the IBEX-35. Sample period: 1st January 2008 to 19th December 2016.

Table 11

Membership degrees in clustering of the daily returns of 24 stocks included in the IBEX-35 ($C = 3$, $m = 1.5$ for AR-FCM_{dC} and $m = 2$ for GARCH-FCM_{dC} and QAF-FCM_{dC}).

		AR-FCM _{dC}			GARCH-FCM _{dC}			QAF-FCM _{dC}		
		C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
<i>Sector: Energy</i>										
Enagás	ENG	0.142	0.002	0.856	0.482	0.447	0.071	0.558	0.275	0.167
Endesa	ELE	0.996	0.000	0.004	0.038	0.546	0.416	0.684	0.072	0.244
Gas Natural	GAS	1.000	0.000	0.000	0.031	0.943	0.026	0.946	0.013	0.041
Iberdrola	IBE	0.193	0.023	0.784	0.997	0.002	0.001	0.579	0.052	0.369
Red Eléctrica	REE	0.869	0.029	0.102	0.532	0.278	0.190	0.814	0.054	0.132
<i>Sector: Materials, Industry and Construction</i>										
Gamesa	GAM	0.616	0.020	0.364	0.001	0.003	0.996	0.703	0.079	0.218
Acciona	ANA	0.930	0.001	0.069	0.000	0.001	0.999	1.000	0.000	0.000
ACS	ACS	1.000	0.000	0.000	1.000	0.000	0.000	0.816	0.020	0.164
Ferrovial	FER	0.269	0.003	0.728	0.031	0.324	0.645	0.754	0.095	0.151
FCC	FCC	1.000	0.000	0.000	0.821	0.145	0.034	0.821	0.025	0.154
Técnicas Reunidas	TRE	0.340	0.230	0.430	0.014	0.105	0.881	0.710	0.147	0.143
Arcelormittal	MTS	0.000	1.000	0.000	0.001	0.003	0.996	0.084	0.020	0.896
<i>Sector: Consumer goods</i>										
Viscofan	VIS	0.204	0.049	0.747	0.027	0.095	0.878	0.087	0.849	0.064
Inditex	ITX	0.239	0.052	0.709	0.083	0.210	0.707	0.000	1.000	0.000
Grifols	GRF	0.429	0.022	0.549	0.008	0.054	0.938	0.539	0.070	0.391
<i>Sector: Financial services</i>										
BBVA	BBVA	0.876	0.009	0.115	0.063	0.895	0.042	0.860	0.017	0.123
Santander	SAN	0.863	0.002	0.135	0.000	1.000	0.000	0.801	0.076	0.123
Caixabank	CABK	0.052	0.002	0.946	0.000	0.000	1.000	0.860	0.037	0.103
Banco Sabadell	SAB	0.963	0.007	0.030	0.007	0.030	0.963	0.316	0.168	0.516
Banco Popular	POP	0.999	0.000	0.001	0.001	0.997	0.002	0.000	0.000	1.000
Bankinter	BKT	0.829	0.004	0.167	0.024	0.867	0.109	0.501	0.040	0.459
<i>Sector: Insurance</i>										
Mapfre	MAP	0.096	0.014	0.890	0.002	0.013	0.985	0.118	0.821	0.061
<i>Sector: Technology and Telecommunications</i>										
Telefónica	TEF	0.325	0.054	0.621	0.970	0.023	0.007	0.955	0.006	0.039
Amadeus	AMS	0.000	0.000	1.000	0.038	0.597	0.365	0.185	0.708	0.107

profiles shown in Figs. 9 (a) and (b) for the banks, it seems inappropriate to locate all of them together at the same cluster, particularly the Banco Popular-POP. These arguments support the better clustering quality indexes obtained by the QAF-FCM_{dC} model. Lastly, the GARCH-FCM_{dC} model produced a cluster partition hardly interpretable and substantially different from those obtained with the two other procedures. As in the case of the ozone records in the first application, the GARCH approximations have not been accurate enough to properly discriminate between the generating models.

7. Concluding remarks and future perspectives

In this paper we have shown that the sample quantile autocovariances are a useful tool to perform hard and soft partitional clustering of times series when the target is to group series generated from the same stochastic process. Quantile autocovariances provide valuable insight into the serial dependence and present a much richer view than other extracted features about the underlying dependence structure. Robustness to nonexistence of moments and capability to deal with heavy-tailed marginal distributions, to analyze dependence of extreme values, and to detect nonlinear features and changes in conditional shapes are appealing properties of the quantile autocovariances, which suggest their usefulness to classify a wide range of time series models. This intuition has been illustrated by means of a

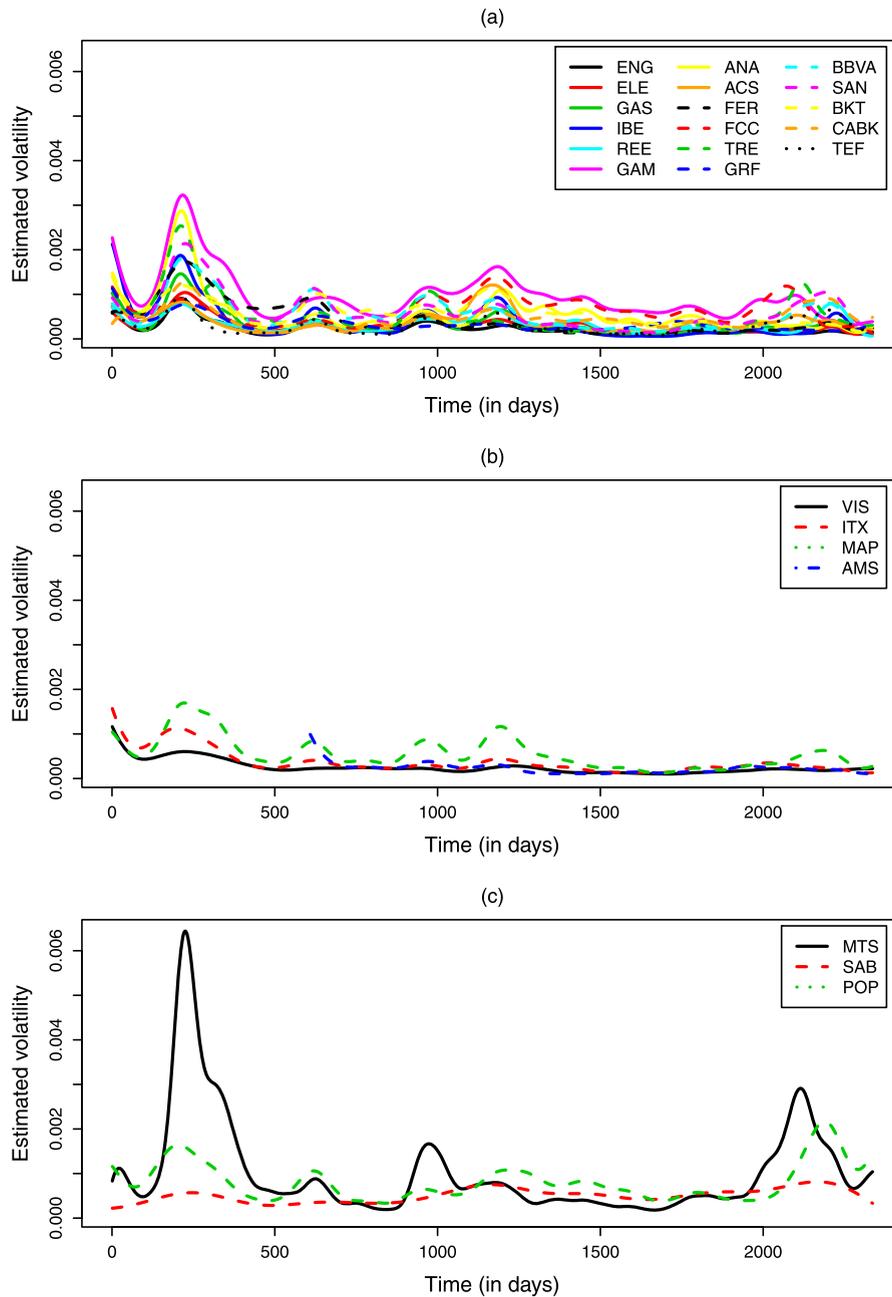


Fig. 9. Nonparametric estimators of the volatility for the daily returns of the 24 analyzed stocks grouped according to the cluster solution provided by the QAF-FCM_{dC} model: \mathcal{C}_1 (a), \mathcal{C}_2 (b) and \mathcal{C}_3 (c).

motivating example addressed to discriminate between realizations of Gaussian white noise, GARCH and exponential GARCH processes.

To perform hard clustering, we focus on the partitioning around C -medoids technique, and the results of a broad simulation study have shown the good performance of the C -crisp model based on the squared Euclidean distance between sample quantile autocovariances, d_{QAF} . Compared to other distances measuring discrepancy between generating models and other extracted features, our approach led to the best classification rates by grouping non-linear and heteroskedastic models in well-separated clusters. Clustering linear models, our proposal produced competitive success rates, in general close to the ones obtained with metrics specifically designed to deal with ARMA models. The

C -crisp model combined with d_{QAF} also exhibited a remarkable property of robustness against the kind of innovation distribution, unlike the rest of examined metrics which have been noticeably affected by skewness and kurtosis. In order to provide an automatic tool for clustering, an iterative algorithm to select the lags and quantile levels optimizing the clustering process has been introduced. Overall, a small number of quantiles is selected, and a further sensitivity analysis has illustrated that a few quantiles are enough to obtain satisfactory results. Summing up, the proposed crisp clustering procedure inherits the nice properties of the quantile autocovariances to discriminate between generating processes, and simultaneously presents the advantages of the partitioning around medoids approach, such as determining time series characterizing the clusters and robustness to noise and outliers, among others. These optimal properties and the outstanding behavior shown in simulations suggest the usefulness and interest of our proposal in hard partitional clustering of time series.

Soft partitional clustering has been considered by introducing a fuzzy C -medoids clustering model for time series based on the sample quantile autocovariances (QAF-FCMdc). Fuzzy paradigm enriches the cluster solution by permitting overlapping clusters, i.e. identifying time series with dynamics close to more than one prototype. To evaluate the QAF-FCMdc algorithm, we have carried out numerical experiments including clusters with different levels of separability and time series equidistant from several clusters. Our assessment criterion took into account the capability of the examined algorithms to detect the fuzzy nature of the equidistant series. Regardless of the considered models and compared with other fuzzy algorithms based on distances between estimates of the underlying parametric structures, the proposed fuzzy algorithm produced good results. Overall, QAF-FCMdc reported better results in the most complex scenarios, where the clusters are closer each other. The most noticeable differences in favor of QAF-FCMdc were observed by clustering GARCH(1, 1) processes, particularly for large sample sizes. In this framework, the GARCH-based algorithms were affected by the inaccurate estimation of the GARCH structure. By contrast, QAF-FCMdc is free of determining the underlying parametric structure and takes advantage of the capability of the quantile autocovariances to detect changes in conditional shapes, thus permitting to discriminate between volatility structures and identify series showing fuzzy behavior. Furthermore, QAF-FCMdc can be applied to series with different lengths, and it is simple to implement and computationally lighter than the analyzed competitors. According to these properties, the proposed fuzzy algorithm is a promising tool to be applied in many situations where it is unrealistic to assume homoscedasticity, such as we have illustrated by means of a specific case-study.

There are indeed some interesting issues for further research in relation to the use of quantile autocovariances in time series clustering. For instance, consistency of the sample quantile autocovariances has been established assuming strictly stationary processes, which can be a constraint in practice. Although stationarity is a quite common requirement in time series clustering, introducing suitable approaches to encompass non-stationary models has great interest in applications, particularly when the series in study are not easy to be transformed or such transformation does not make sense. Also notice that quantile autocovariances are well-defined for time series taking ordinal values. Therefore, it is worth analyzing the behavior of the proposed procedures in clustering of temporal sequences of ordinal data or mixed-type (metric-ordinal) data, which typically arise in social stratification and generally in social science [72]. Although we have tackled the clustering task in the time domain, an alternative approach to be addressed in future works is to consider the frequency domain by using a distance between proper estimators of the *quantile spectral densities* [25] defined by

$$G(x, y, \omega) = \frac{1}{2\pi} \sum_l \text{cov}\{I(X_0 \leq x), I(X_l \leq y)\} \exp(il\omega).$$

The quantile spectral density $G(x, y, \omega)$ specifies the frequency decomposition of the quantile autocovariances so that it can be seen as the cross spectral density of the bivariate time series $(I(X_t \leq x), I(X_t \leq y))$. Like the quantile autocovariances, $G(x, y, \omega)$ provides all the information about the serial dependence structure but now from the spectral point of view. Different approaches to estimate the quantile spectral density considering L_1 and L_2 procedures and their asymptotic properties have been provided in several works [25,36–38]. In particular, Lee and Rao [25] propose to check the equality of serial dependence of two stationary time series by using the test statistic given by

$$\mathcal{P}_T = \frac{1}{T} \sum_{k=1}^T \int |\widehat{G}_{1,T}(x, y, \omega_k) - \widehat{G}_{2,T}(x, y, \omega_k)|^2 dF(x)dF(y)$$

where ω_k denote the k -th Fourier frequency, $\widehat{G}_{1,T}$ and $\widehat{G}_{2,T}$ are the quantile spectral density estimators and F is any distribution function. This way, \mathcal{P}_T statistic might be considered as an innovative spectral dissimilarity measure between two time series.

Other additional issue deserving particular attention is to obtain robust versions of the fuzzy QAF–FCMdc algorithm to neutralize the effect of anomalous fuzzy data. Currently, we are studying generalizations of the robustness techniques considered by [73], namely the metric approach (by smoothing the distance), the noise approach (by introducing an artificial noise cluster) and the trimmed approach (by trimming away a small fraction of series).

Acknowledgements

The research carried out by the authors José A. Vilar and Borja Lafuente-Rego has been supported by the Spanish Ministerio de Economía y Competitividad (grant number MTM2014-52876-R), the Xunta de Galicia (Centro Singular de Investigación de Galicia accreditation ED431G/01 2016–2019 and Grupos de Referencia Competitiva ED431C 2016-015) and the European Union (European Regional Development Fund – ERDF). The authors wish to thank the editor and the anonymous referees for their helpful comments and suggestions to improve the quality and presentation of this manuscript.

Appendix A

Outline of the proof of Theorem 2.1. The arguments employed to establish Lemma A.1 (a), Lemma A.2 and Theorem 1 in Han et al. [39] directly lead to the asymptotic result in Theorem 2.1. Hence, a general overview of the proof is just outlined in order to provide readers insight into the key steps.

Consider an arbitrary lag l and a pair of levels of probability $(\tau, \tau') \in \mathcal{A} \equiv \mathcal{A}_\tau \times \mathcal{A}_{\tau+l}$, where \mathcal{A}_τ is a compact subset in $(0, 1)$ denoting the quantile range of interest for X_t . According to the definition of quantile autocovariance in (1) and the corresponding estimator given in (2), we have

$$\sqrt{T} (\hat{\gamma}_l(\tau, \tau') - \gamma_l(\tau, \tau')) = \sqrt{T} \left[T^{-1} \sum_{t=1}^{T-l} I(X_t \leq \hat{q}_\tau, X_{t+l} \leq \hat{q}_{\tau'}) - F_l(q_\tau, q_{\tau'}) \right]. \tag{A.1}$$

By adding and subtracting the term $T^{-1/2} F_l(\hat{q}_\tau, \hat{q}_{\tau'})$ in (A.1), we obtain

$$\sqrt{T} (\hat{\gamma}_l(\tau, \tau') - \gamma_l(\tau, \tau')) = \mathbb{V}_{T,l}(\hat{q}_\tau, \hat{q}_{\tau'}) + \sqrt{T} [F_l(\hat{q}_\tau, \hat{q}_{\tau'}) - F_l(q_\tau, q_{\tau'})], \tag{A.2}$$

where $\mathbb{V}_{T,l}$ is the empirical process indexed by $(u, u')^t \in \mathbb{R}^2$ given by

$$\mathbb{V}_{T,l}(u, u') = T^{-1/2} \sum_{t=1}^{T-l} \varphi_t(u, u', l), \text{ with } \varphi_t(u, u', l) = I(X_t \leq u, X_{t+l} \leq u') - F_l(u, u'). \tag{A.3}$$

As $F_l(\cdot, \cdot)$ is differentiable by Assumption A4, the mean value expansion leads to

$$\sqrt{T} [F_l(\hat{q}_\tau, \hat{q}_{\tau'}) - F_l(q_\tau, q_{\tau'})] = \nabla F_l(\bar{q}_\tau, \bar{q}_{\tau'})^t \sqrt{T} (\hat{q}_\tau - q_\tau, \hat{q}_{\tau'} - q_{\tau'}). \tag{A.4}$$

Also, following arguments by [39], it is shown that

$$\lim_{T \rightarrow \infty} P \left(\sup_{(\tau, \tau') \in \mathcal{A}} |\mathbb{V}_{T,l}(\hat{q}_\tau, \hat{q}_{\tau'}) - \mathbb{V}_{T,l}(q_\tau, q_{\tau'})| \right) = 0. \tag{A.5}$$

Combining (A.4) and (A.5), we can write (A.2) as follows.

$$\sqrt{T} (\hat{\gamma}_l(\tau, \tau') - \gamma_l(\tau, \tau')) = \mathbb{V}_{T,l}(q_\tau, q_{\tau'}) + \nabla F_l(q_\tau, q_{\tau'})^t \sqrt{T} (\hat{q}_\tau - q_\tau, \hat{q}_{\tau'} - q_{\tau'}) + o_p(1), \tag{A.6}$$

uniformly in $(\tau, \tau') \in \mathcal{A}$.

Now, define $\mathbb{W}_{T,l}(u_1, u_2) = (\mathbb{W}_{T,l}^{(1)}(u_1), \mathbb{W}_{T,l}^{(2)}(u_2))^t$, where $\mathbb{W}_{T,l}^{(1)}(u_1) = \lim_{u_2 \rightarrow \infty} \mathbb{V}_{T,l}(u_1, u_2)$ and $\mathbb{W}_{T,l}^{(2)}(u_2) = \lim_{u_1 \rightarrow \infty} \mathbb{V}_{T,l}(u_1, u_2)$. Based on the Bahadur representation of the sample quantiles, it holds

$$\sqrt{T}(\hat{q}_\alpha - q_\alpha) = \frac{1}{f(q_\alpha)} \mathbb{W}_{T,l}^{(l)}(q_\alpha) + o_p(1), \quad (\text{A.7})$$

for $i = 1, 2$ and uniformly in α . Expression (A.7) combined with (A.6) leads to

$$\sqrt{T}(\hat{\gamma}_l(\tau, \tau') - \gamma_l(\tau, \tau')) = \lambda_{l,(\tau, \tau')}^l \mathbf{v}_{T,l}(\tau, \tau') + o_p(1), \quad (\text{A.8})$$

uniformly in $(\tau, \tau') \in \mathcal{A}$, where $\mathbf{v}_{T,l}(\tau, \tau') = [\mathbb{V}_{T,l}(q_\tau, q_{\tau'}), \mathbf{W}_{T,l}(q_\tau, q_{\tau'})^T]^T$, and $\lambda_{l,(\tau, \tau')}$ is given in (4).

Under Assumptions A1–A4 required in Theorem 2.1, the set of constraints allowing to establish Lemma A.1 in Han et al. [39] for the particular case of a one-dimensional process are satisfied. Therefore, according to Lemma A.1 (a) in [39], the process $\mathbb{V}_{T,l}(u, u')$ defined by (A.3) converges weakly to the mean zero Gaussian process $\mathbb{V}_{\infty,l}(u, u')$ with covariance matrix given by

$$\mathbb{E}_l((u_1, u'_1), (u_2, u'_2)) = \mathbb{E}(\mathbb{V}_{\infty,l}(u_1, u'_1) \mathbb{V}_{\infty,l}(u_2, u'_2)) = \sum_{l=-\infty}^{\infty} \text{cov}[\varphi_l(u_1, u'_1, l), \varphi_0(u_2, u'_2, l)]. \quad (\text{A.9})$$

Result (A.9) together with the uniform boundedness of $\lambda_{l,(\tau, \tau')}$ (which follows from Assumption A2 and A4) ensure the finite dimensional distributions convergence of $\lambda_{l,(\tau, \tau')}^l \mathbf{v}_{T,l}(\tau, \tau')$ over $(\tau, \tau') \in \mathcal{A}$. This way, it suffices to establish the stochastic equicontinuity of $\lambda_{l,(\tau, \tau')}^l \mathbf{v}_{T,l}(\tau, \tau')$, which is derived from the stochastic equicontinuity of $\lambda_{l, \cdot}, \mathbb{V}_{T,l}(\cdot)$ and $\mathbf{W}_{T,l}(\cdot)$. \square

References

- [1] T.W. Liao, Clustering of time series data: a survey, *Pattern Recognit.* 38 (11) (2005) 1857–1874.
- [2] T.-c. Fu, A review on time series data mining, *Eng. Appl. Artif. Intell.* 24 (1) (2011) 164–181.
- [3] S. Rani, G. Sikka, Recent techniques of clustering of time series data: a survey, *Int. J. Comput. Appl.* 52 (15) (2012) 1–9.
- [4] J. Caiado, E.A. Maharaj, P. D'Urso, Time series clustering, in: C. Hennig, M. Meila, F. Murtagh, R. Rocci (Eds.), *Handbook of Cluster Analysis*, Handbooks of Modern Statistical Methods, Chapman and Hall/CRC, London, 2015, pp. 241–263.
- [5] P. Montero, J.A. Vilar, **TSclust**: an R package for time series clustering, *J. Stat. Softw.* 62 (1) (2014) 1–43.
- [6] P. Montero, J.A. Vilar, **TSclust**: time series clustering utilities, R package version 1.2.1, 2014, <http://CRAN.R-project.org/package=TSclust>.
- [7] E. Keogh, S. Kasetty, On the need for time series data mining benchmarks: a survey and empirical demonstration, *Data Min. Knowl. Discov.* 7 (4) (2003) 349–371.
- [8] S. Pértega, J.A. Vilar, Comparing several parametric and nonparametric approaches to time series clustering: a simulation study, *J. Classif.* 27 (3) (2010) 333–362.
- [9] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, E.J. Keogh, Experimental comparison of representation methods and distance measures for time series data, *Data Min. Knowl. Discov.* 26 (2) (2013) 275–309.
- [10] C. Döring, M.-J. Lesot, R. Kruse, Data analysis with fuzzy clustering methods, *Comput. Stat. Data Anal.* 51 (1) (2006) 192–214.
- [11] P. D'Urso, Fuzzy clustering, in: C. Hennig, M. Meila, F. Murtagh, R. Rocci (Eds.), *Handbook of Cluster Analysis*, Handbooks of Modern Statistical Methods, Chapman and Hall/CRC, London, 2015, pp. 545–573.
- [12] P. D'Urso, E.A. Maharaj, Autocorrelation-based fuzzy clustering of time series, *Fuzzy Sets Syst.* 160 (24) (2009) 3565–3589.
- [13] 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.
- [14] 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.
- [15] G.P. Aielli, M. Caporin, Fast clustering of GARCH processes via Gaussian mixture models, *Math. Comput. Simul.* 94 (2013) 205–222.
- [16] E.A. Maharaj, P. D'Urso, Fuzzy clustering of time series in the frequency domain, *Inf. Sci.* 181 (7) (2011) 1187–1211.
- [17] 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.
- [18] C.-E. Tseng, C.-Y. Peng, M.-W. Chang, J.-Y. Yen, C.-K. Lee, T.-S. Huang, Novel approach to fuzzy-wavelet ECG signal analysis for a mobile device, *J. Med. Syst.* 34 (1) (2010) 71–81.
- [19] E.A. Maharaj, P. D'Urso, D.U. Galagedera, Wavelet-based fuzzy clustering of time series, *J. Classif.* 27 (2) (2010) 231–275.
- [20] P. D'Urso, E.A. Maharaj, Wavelets-based clustering of multivariate time series, *Fuzzy Sets Syst.* 193 (2012) 33–61.
- [21] P. D'Urso, E.A. Maharaj, A.M. Alonso, Fuzzy clustering of time series using extremes, *Fuzzy Sets Syst.* (2017), <http://dx.doi.org/10.1016/j.fss.2016.10.006>, in press.
- [22] P. D'Urso, C. Cappelli, D.D. Lallo, R. Massari, Clustering of financial time series, *Physica A* 392 (9) (2013) 2114–2129.
- [23] P. D'Urso, L. De Giovanni, R. Massari, GARCH-based robust fuzzy clustering of time series, *Fuzzy Sets Syst.* 305 (2016) 1–28.
- [24] O. Linton, Y.-J. Whang, The quantilogram: with an application to evaluating directional predictability, *J. Econom.* 141 (1) (2007) 250–282.
- [25] J. Lee, S. Rao, The quantile spectral density and comparison based tests for nonlinear time series, Unpublished manuscript, Department of Statistics, Texas A&M University, College Station, USA, 2012, arXiv:1112.2759v2.
- [26] B. Lafuente-Rego, J.A. Vilar, Clustering of time series using quantile autocovariances, *Adv. Data Anal. Classif.* (2015) 1–25.

- [27] L. Kaufman, P.J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, 9th edition, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, Inc., New York, 1990.
- [28] L. Bauwens, J.V.K. Rombouts, Bayesian clustering of many GARCH models, *Econom. Rev.* 26 (2–4) (2007) 365–386.
- [29] E. Otranto, Clustering heteroskedastic time series by model-based procedures, *Comput. Stat. Data Anal.* 52 (10) (2008) 4685–4698.
- [30] G.P. Aielli, M. Caporin, Variance clustering improved dynamic conditional correlation MGARCH estimators, *Comput. Stat. Data Anal.* 76 (2014) 556–576.
- [31] E. Lazar, C. Alexander, Normal mixture GARCH(1, 1): applications to exchange rate modelling, *J. Appl. Econom.* 21 (3) (2006) 307–336.
- [32] R.T. Kipkoeh, Modeling volatility under normal and student-t distributional assumptions (a case study of the Kenyan exchange rates), *Am. J. Appl. Math. Stat.* 2 (4) (2014) 179–184.
- [33] J.L. Andrews, P.D. McNicholas, Variable selection for clustering and classification, *J. Classif.* 31 (2) (2014) 136–153.
- [34] T. Mikosch, C. Stărică, Limit theory for the sample autocorrelations and extremes of a GARCH(1, 1) process, *Ann. Stat.* 28 (5) (2000) 1427–1451.
- [35] R.A. Davis, T. Mikosch, The extremogram: a correlogram for extreme events, *Bernoulli* 15 (4) (2009) 977–1009.
- [36] A. Hagemann, Robust spectral analysis, Arxiv e-prints, arXiv:1111.1965v1, 2013.
- [37] T.-H. Li, Quantile periodograms, *J. Am. Stat. Assoc.* 107 (498) (2014) 765–776.
- [38] H. Dette, M. Hallin, T. Kley, S. Volgushev, Of copulas, quantiles, ranks and spectra: an l_1 -approach to spectral analysis, Arxiv e-prints, arXiv:1111.7205v2, 2014.
- [39] H. Han, O. Linton, T. Oka, Y.-J. Whang, The cross-quantilegram: measuring quantile dependence and testing directional predictability between time series, *J. Econom.* 193 (1) (2016) 251–270.
- [40] H.J. Skaug, D. Tjøstheim, A nonparametric test of serial independence based on the empirical distribution function, *Biometrika* 80 (3) (1993) 591–602.
- [41] Y. Hong, Generalized spectral tests for serial dependence, *J. R. Stat. Soc., Ser. B, Stat. Methodol.* 62 (3) (2000) 557–574.
- [42] S.-C. Kao, A.R. Ganguly, K. Steinhaeuser, Motivating complex dependence structures in data mining: a case study with anomaly detection in climate, in: Y. Saygin, J.X. Yu, H. Kargupta, W. Wang, S. Ranka, P.S. Yu, X. Wu (Eds.), 2009 IEEE International Conference on Data Mining Workshops, IEEE Computer Society, Los Alamitos, CA, USA, 2009, pp. 223–230.
- [43] R. Koenker, *Quantile Regression*, Econometric Society Monographs, Cambridge University Press, 2005.
- [44] Z. Bohte, D. Cepar, K. Košmelj, Clustering of time series, in: M. Barritt, D. Wishart (Eds.), *Compstat 80*, Proceedings in Computational Statistics, Physica-Verlag, Heidelberg, Vienna, 1980, pp. 587–593.
- [45] P. Galeano, D. Peña, Multivariate analysis in vector time series, *Resenhas* 4 (4) (2000) 383–403.
- [46] D. Piccolo, A distance measure for classifying ARIMA models, *J. Time Ser. Anal.* 11 (2) (1990) 153–164.
- [47] E.A. Maharaj, A significance test for classifying ARMA models, *J. Stat. Comput. Simul.* 54 (4) (1996) 305–331.
- [48] E.A. Maharaj, Clusters of time series, *J. Classif.* 17 (2) (2000) 297–314.
- [49] J. Caiado, N. Crato, D. Peña, A periodogram-based metric for time series classification, *Comput. Stat. Data Anal.* 50 (10) (2006) 2668–2684.
- [50] J.A. Vilar, S. Pértega, Discriminant and cluster analysis for Gaussian stationary processes: local linear fitting approach, *J. Nonparametr. Stat.* 16 (3–4) (2004) 443–462.
- [51] H. Tong, I. Yeung, On tests for self-exciting threshold autoregressive-type non-linearity in partially observed time series, *J. R. Stat. Soc., Ser. C, Appl. Stat.* 40 (1) (1991) 43–62.
- [52] M. Gavrilov, D. Angelov, P. Indyk, R. Motwani, Mining the stock market (extended abstract): which measure is best?, in: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'00, ACM, New York, USA, 2000, pp. 487–496.
- [53] L. Hubert, P. Arabie, Comparing partitions, *J. Classif.* 2 (1) (1985) 193–218.
- [54] W.M. Rand, Objective criteria for the evaluation of clustering methods, *J. Am. Stat. Assoc.* 66 (336) (1971) 846–850.
- [55] F. Höppner, F. Klawonn, R. Kruse, T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, Wiley, Chichester, UK, 1999.
- [56] P. D'Urso, D. Di Lallo, E.A. Maharaj, Autoregressive model-based fuzzy clustering and its application for detecting information redundancy in air pollution monitoring networks, *Soft Comput.* 17 (1) (2013) 83–131.
- [57] T.W. Liao, F. Tang, J. Qu, P. Blau, Grinding wheel condition monitoring with boosted minimum distance classifiers, *Mech. Syst. Signal Process.* 22 (1) (2008) 217–232.
- [58] J.M. Vilar, J.A. Vilar, S. Pértega, Classifying time series data: a nonparametric approach, *J. Classif.* 26 (1) (2009) 3–28.
- [59] J. Caiado, N. Crato, A GARCH-based method for clustering of financial time series: international stock markets evidence, in: C.H. Skiadas (Ed.), *Recent Advances in Stochastic Modeling and Data Analysis*, World Scientific Publishing, Singapore, 2007, pp. 542–551.
- [60] E. Otranto, Identifying financial time series with similar dynamic conditional correlation, *Comput. Stat. Data Anal.* 54 (1) (2010) 1–15.
- [61] T. Kamdar, A. Joshi, On Creating Adaptive Web Servers Using Weblog Mining, Tr-cs-00-05, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, USA, 2000.
- [62] M.-S. Yang, K.-L. Wu, J.-N. Hsieh, J. Yu, Alpha-cut implemented fuzzy clustering algorithms and switching regressions, *IEEE Trans. Syst. Man Cybern., Part B* 38 (3) (2008) 588–603.
- [63] N.R. Pal, J.C. Bezdek, On cluster validity for the fuzzy c-means model, *IEEE Trans. Fuzzy Syst.* 3 (3) (1995) 370–379.
- [64] 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.
- [65] R.L. Cannon, J.V. Davè, J.C. Bezdek, Efficient implementation of the fuzzy c-means clustering algorithms, *IEEE Trans. Pattern Anal. Mach. Intell.* 8 (2) (1986) 248–255.
- [66] F. de A.T. de Carvalho, C.P. Tenrio, N.L.C. Junior, Partitional fuzzy clustering methods based on adaptive quadratic distances, *Fuzzy Sets Syst.* 157 (21) (2006) 2833–2857.

- [67] P. D'Urso, P. Giordani, A weighted fuzzy c-means clustering model for fuzzy data, *Comput. Stat. Data Anal.* 50 (6) (2006) 1496–1523.
- [68] D. Dembélé, P. Kastner, Fuzzy c-means method for clustering microarray data, *Bioinformatics* 19 (8) (2003) 973–980.
- [69] G. Estévez-Pérez, J.A. Vilar, Functional ANOVA starting from discrete data: an application to air quality data, *Environ. Ecol. Stat.* 20 (3) (2013) 495–517.
- [70] R. Campello, E. Hruschka, A fuzzy extension of the silhouette width criterion for cluster analysis, *Fuzzy Sets Syst.* 157 (21) (2006) 2858–2875.
- [71] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (8) (1991) 841–847.
- [72] C. Hennig, T.F. Liao, How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification, *Appl. Stat., J. R. Stat. Soc.* 62 (3) (2013) 309–369.
- [73] P. D'Urso, L.D. Giovanni, Robust clustering of imprecise data, *Chemom. Intell. Lab. Syst.* 136 (2014) 58–80.