Non parametric and robust statistical test based on wavelets for time series classification

Alejandra Sánchez González[*], Henry Laniado [†], Mateo Ríos Querubín [‡]

November 14, 2018

Abstract

In this article a statistical procedure for identifying if a time series set follows the same model is developed. With the aim of supporting characterization and pattern recognition for temporal series, and inspired by the methodology of Maharaj E. A.[1], we take advantage of the wavelet coefficients properties to characterize a signal and our procedure is made by means of a randomization test on those coefficient. Our main contribution in this work is to introduce modified versions of test statistic in test for pattern recognition of time series which in general, have a great performance in terms of size and power, both being desirable features in a statistic test. It is worth pointing out that we introduce robust statistical tests whose performance are better in presence of atypical values than some techniques already studied in the literature. The methodology developed here allow us to design a new method to classify time series and atypical values identification. We implement our new methods in real and simulated cases.
Keywords: Robust estimators, Wavelet coefficients, Randomization test.

1 Introduction

Pattern recognition in time series, applicable in many areas, is an important subject due to the *power* that it gives to the data analytic. The time series appear in many science fields and hence it is important to identify those temporal processes that follow a same model, since this allow for classifying the temporal events into groups with similar characteristics. For example, in finances, one could design a portfolio with risky assets having a common model. In economics, will be possible to identify what countries have had a similar economic growing. That is why it is important to identify similarities in time series, as this contributes to decision making and forecasting.

This recognition includes the identification of atypical data or *"outliers"*, which tend to distort the information given understood as loss of money, time and credibility. The time series identification procedure introduced in this work is proposed for both stationary and non stationary processes. This second group can require some transformations in order to give satisfactory results as it is mentioned in Basawa et al [2], Diggle & Fisher [3], Guo [4], Timmer et al. [5] and Maharaj [6]. In order to work with the spectral information of the series, it is necessary to translate them from a world with a probably equally distanced sample time to the world of frequencies, where more flexible and precise analyses can be developed, what leads to the use of Wavelets functions [7] [8].

Wavelets theory includes powerful techniques that allow to study times series with different frequencies. Different applications and utilities of this theory can be found in Percival [9] where, the use of Wavelets to decompose trajectories in different ways without losing much information is explored[10]. In particular, we work with wavelets discrete transform (WDT) to compress the series information and the statistical test is executed on the coefficients that come from WDT. Note that the analysis using the WDT improves the computational performance of the identification algorithms and the recognition of changes among time and frequency over series.

The procedure studied in this paper is inspired the idea of Maharaj E. A.[1], that compares pairs of series through the use of a non parametrized test that is mainly based on wavelet coefficients of the each time series and whose

[*]Departamento de Ciencias Básicas, E-mailaddresses: alejandra.sanchez@usantoto.edu.co, Universidad Santo Tomás, Tunja - Colombia
[†]Departamento de Ciencias Matemáticas, E-mailaddresses: hlaniado@eafit.edu.co, Universidad EAFIT, Medellín - Colombia
[‡]Departamento de Ciencias Matemáticas, E-mailaddresses:mriosqu@eafit.edu.co, Universidad EAFIT, Medellín - Colombia

statistic depends on the Euclidean norm of those coefficients. This test introduced in Maharaj E. A.[1] presents good results in terms of its power. However, the results presented by Maharaj show a low performance in terms of the values obtained for the power and the size of the test, where it is presumed that the performance of the test has to decrease in the presence of outliers. Our proposal in this work is oriented to solve this problem, hence, we design a similar procedure to identify if a pair of time series follow a same model. Then, based on a non-parametric type test, since we lack any assumption on the data distribution [11], our objective and the novelty of this article, is a robust version that alows us to identify similarities in a dataset, which is highly efficient in presence of outliers data. We have even made a modification by changing the norms.

In order to improve the identification test introduced in Maharaj E. A.[1], new norms and robust statistics are included producing, in general, better results to identify times series. This new modifications assure a non parametric version, since we do not impose theoretical assumptions over data as usual in most statistical tests. In addition, we emphasize that our test is less sensible to outliers.

This new statistical test is designed under the null hypothesis that the time series follow the same model. Hence, if two time series come from the same model, then the wavelets decomposition for both series has to be the same and therefore, any transformation, for example a norm, of those coefficients must be also the same. This can be applied using different scales considered in the wavelet transformation, among other different measure as norms 1,2 and infinite, and some well known robust statistics as median and trimmed mean. The main goal here is to test if different series were produced by the same generator stochastic process.

These robust estimators are proposed as an appropriate tool for statistical inference models, basically this test must be applied for situations where deviations from the assumptions are detected or the data shows evidence of outliers, see for Ortíz. M[12]. It is here, then, that the our test shows its capability of being a process for abnormal series. This type of series might present atypical values, whose detection is quite an important matter. Grané and Veiga [13] highlight the importance of finding those values, because, even if they are of size, form or amplitude, they can have a negative influence on the estimates. Grané and Veiga [13] also propose an analysis based on the use of wavelet functions for abnormal data, whose advances might be of use for this development.

The present document is divided in five sections. The first one, after introduction, gives a brief description to Wavelet Functions, its importance and applicability. It is followed by section 3, where the proposed randomized test, grounded on the theory proposed by Maharaj E. A. [1], modified by the proposed new estimators is exposed. Afterwards, section 4 shows the methodology and simulations developed and section 5 expose the results of the project and applications of themselves respectively. Finally, conclusions and references are presented.

2 Wavelets

Wavelets theory was developed essentially through the last 120 years by authors like Alfred Haar (Haar Wavelet)[14], George Zweig (continuous Wavelet transformation), Jean Morlet (Morlets formulation of the Continuous Wavelet Transformation) and Ingrid Daubichies (*Ten Lectures on Wavelets* and visualization of themselves) among others[15][16]. Its same named transformation is used to decompose series in different forms in its different possible frequencies but also preserving its representation in its original frequency[17].

The ability of decomposing series in this manner, gives Wavelets Transformations (WT) scope in many fields of knowledge including grouping [18], prediction and animalities recognition and correction. In the field of grouping, clustering analysis using wavelets is commonly used, it can be seen in Alonso and Gouveia [19],where pollutants concentrations(C_3 y NO_2), are analyzed in Italy. Also, Alonso et al. [20] apply the same methodology with data of sea level at middle day (MSL) and works in clear and diffuse grouping methods based on a combination of wavelet characteristics exposed by D 'Urso and Maharaj [21], through the Multi-resolution analysis algorithm described in Mallat [22], later used to examine series at very different scales and resolutions in order to evaluate its different behaviors.

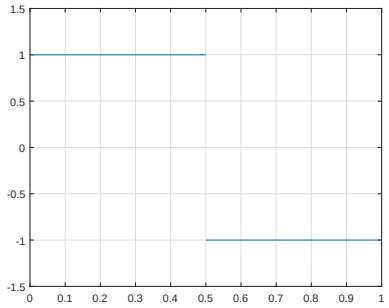
Besides its uses in clustering theory, wavelets analysis can be applied to reduce series dimensionality without losing the fundamental characteristic of themselves, by using the wavelet transform [18]. It was introduced by Grossman and Morlet in 1984, as a tool for time-scale analysis of non stationary signals, as well as a tool to study the discontinuities and non linearity of irregular data. The process proposed in this document makes use of discrete wavelet transformations in the developed algorithms, so it is strictly necessary the comprehension of its theory.The calculation for

Wavelet transform is performed through the product between the signal and a mother wavelet function [18], defined in the following way:

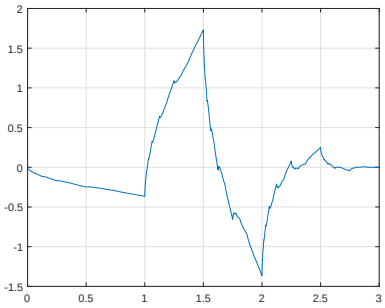
$$W(s,u) = \int_{-\infty}^{\infty} \psi_{s,u}(t) f(t)\ dt, \tag{1}$$

where $\psi_{s,u}(t)$ is a modification of a mother wavelet through translation and scaling following the form $\psi_{s,u}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)$, where s and u are a scale parameter and a translation factor respectively. The mother wavelet ψ is a localized function, which meets the following conditions: $\int_{-\infty}^{\infty} \psi(t)\, dt = 0$ y $\int_{-\infty}^{\infty} \psi(t)^2\, dt = 1$. In the current theory there are several examples of such functions, the election of which, must consider the field of application. One of the most common and applicable type of Wavelets are the Daubichies family, whose first three members are used aiming good results. Below, the behavior of this members is shown.

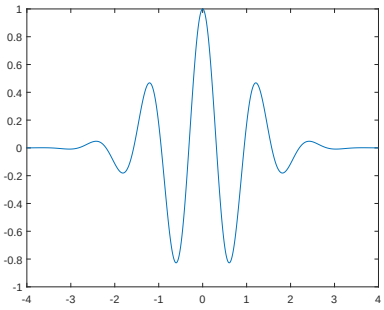
Figure 1: Funciones Wavelets



(a) Haar (Dubichies 1)



(b) Dubichies 2



(c) Morlet (Dubichies 3)

Now, continuing with the transition of equation 1, the continuous wavelet transform (CWT) is of the form:

$$CWT(s,u) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt \tag{2}$$

As it was mentioned for equation 1, the parameters for this equation are the same. The calculation of the CWT suggests the use of powerful computational methods. The poor information it delivers, since it is highly redundant for the reconstruction of the series, thus, increas the calculation time in the information processing leading to consider the use of its discrete form.

The Discrete Wavelet Transform (DWT) is a good alternative to work in discrete framework, at least for our objectives, since it allowes improves the algorithms performance, and it is capable of deliver enough information both for analysis as well as for the series reconstruction, all this with a remarkable decrease of processing time[23][24][25]. The discrete wavelet transform is of the form:

$$DWT(j,k) = \int_{-\infty}^{\infty} f(t) 2^{\frac{j}{2}} \psi(2^j t - kn) dt \tag{3}$$

The advantages of the use of DWT are important for the efficient calculation, through by means of banks of high pass filters as well as low pass filters which are divided in detail coefficients and aproximity coefficients, where the whole of the series information is preserved, allowing the reconstruction of it. The DWT contributes in the detection of sudden changes in the signal, by the identification of peaks in the observed phenomena through the detail coefficients. Works presented by Mallat [15], where the use of the DWT is linked with the identification of singularities in the signal, or the correction of data with atypical values, which reduces the asymmetry and the excess of kurtosis in the series of distributions.

The decomposition offered by the DWT for the original series in several distinct components and different scales saves the complete information of the original series. Currently there are exists several examples of its applicability such as that one exposed by Subasi [22] in its article, where an EEG signal is decomposed in several sub-bands through the DWT, which entails better results in a classification engine for diagnosis of healthy patients versus epileptic ones.

In the following section is presented in detail the methodology to identify when a pair of time series come from a the same model. We follow the same idea but changing some transformations and incorporating the robust versions. This is a good application of wavelets transformation to the statistical framework.

3 Pattern recognition analysis

The randomized test allows us to identify pairs of series belonging to the same model. Hence, we define the hypothesis test proceeding in the following manner. Let X_t and Y_t be two stationary processes either linear or non-linear, which follow models P_x and P_y, respectively. Let $X = (x_1, x_2, ..., x_n)$ and $Y = (y_1, y_2, ..., y_n)$ be observation vectors for the two stationary processes described earlier. We are interested in the following hypothesis proof:

$$H_0 : P_x = P_y \tag{4}$$
$$H_a : P_x \neq P_y \tag{5}$$

The algorithm can be summarized as follows, where $X = (x_1, x_2, ..., x_n)$ the observed data (this means, the signal or the time series, where $X_i = f(t_i)$, $t_i = i/n$, $i = 1, ..., n$ and $n = 2^j$. The discrete transform wavelet (DWT) uses orthogonal transformations to decompose X in wavelet coefficient vectors: The detailed coefficients D_1, D_2, ..., D_j and Coefficients of proximity A_j, with each set of wavelet coefficients containing $n/2^j$ data points for $j = 1, ..., j$. The main property of these detail coefficients is its extreme sensitivity to the non-fragile characteristics of the data, such as noise, jumps and spikes, The wavelet coefficients (detailed) are proportional to the difference of the average of the observations of time series in each scale, while the scale coefficients(aproximity) are proportional to the average of the original series e n the largest scale. The scale coefficient show and present a trend similar to the original series.

Then, by using Wavelet or detail coefficients, we want to test if the generation process is the same for both series. Thus, taking pairs of series $X(t)$ and $Y(t)$, which are generate through by the above process and using the Wavelet Daubechies family in levels 1 to 3, the proposed hypothesis is validated. In this work, the hypothesis (ref proof-1a and ref proof-1b) is checked through the statistic defined as the relationship between the Euclidean norm for the wavelet coefficient of the two series, as proposed Elizabeth and Maharaj cite key-1. Our contribution to this research is the proposal of the norm 1, the infinite norm and three estimators of proven robustness. The problem here is that the asymptotic distribution of the statistics indicated in the expression (ref eq: 6) is unknown, even for $p = 2$. Therefore, this test must be done through a nonparametric randomization process. The test statistic is described below:

$$R(\lambda) = \frac{\|W_x(\lambda, t)\|_p}{\|W_y(\lambda, t)\|_p}. \tag{6}$$

Where p corresponds to the selected standard. The use of robust estimators for standards contributes to the statistical calculation test. the test will be implemented with the median and the cut average at 5% and 10%. Robust

estimators, a term introduced by Box (1953), allow the development of a process insensitive to deviations from the assumptions on which the randomization test is based, which guarantees a greater insensitivity to the presence of irregularities. The statistics of tests through the proposed estimators are mentioned below:

1. Estimator by means of Norm 1.

$$R(\lambda) = \frac{\|W_x(\lambda,t)\|_1}{\|W_y(\lambda,t)\|_1} = \frac{\sum_{i=1}^{n} |W_{x_i}|}{\sum_{i=1}^{n} |W_{y_i}|} \tag{7}$$

2. Estimator by means of Norm 2.

$$R(\lambda) = \frac{\|W_x(\lambda,t)\|_2}{\|W_y(\lambda,t)\|_2} = \frac{\sqrt{\sum_{i=1}^{n} W_{x_i}^2}}{\sqrt{\sum_{i=1}^{n} W_{y_i}^2}} \tag{8}$$

3. Estimator by means of Infinite Norm.

$$R(\lambda) = \frac{\|W_x(\lambda,t)\|_\infty}{\|W_y(\lambda,t)\|_\infty} = \frac{\max_{1\leq i\leq n} |W_{x_i}|}{\max_{1\leq i\leq n} |W_{y_i}|} \tag{9}$$

4. Robust estimator based on median for norm 1 and norm 2, respectively.

$$R(\lambda) = \sqrt{\frac{median(|\mathbf{W_x}|)}{median(|\mathbf{W_y}|)}} \tag{10}$$

$$R(\lambda) = \sqrt{\frac{median(\mathbf{W_x^2})}{median(\mathbf{W_y^2})}} \tag{11}$$

5. Robust estimators based on α-trimming for norms 1 and 2, respectively.

$$R(\lambda) = \frac{\overline{W}_{\alpha_x}}{\overline{W}_{\alpha_y}} = \sqrt{\frac{\frac{1}{n-2[n\alpha]} \sum_{i=[n\alpha]}^{n-[n\alpha]} |W_{x_i}|}{\frac{1}{n-2[n\alpha]} \sum_{i=n\alpha}^{n-n\alpha} |W_{y_i}|}} = \sqrt{\frac{\sum_{i=[n\alpha]}^{n-[n\alpha]} |W_{x_i}|}{\sum_{i=n\alpha}^{n-n\alpha} |W_{y_i}|}} \tag{12}$$

$$R(\lambda) = \frac{\overline{W}_{\alpha_x}}{\overline{W}_{\alpha_y}} = \sqrt{\frac{\frac{1}{n-2[n\alpha]} \sum_{i=[n\alpha]}^{n-[n\alpha]} W_{x_i}^2}{\frac{1}{n-2[n\alpha]} \sum_{i=n\alpha}^{n-n\alpha} W_{y_i}^2}} = \sqrt{\frac{\sum_{i=[n\alpha]}^{n-[n\alpha]} W_{x_i}^2}{\sum_{i=[n\alpha]}^{n-[n\alpha]} W_{y_i}^2}} \tag{13}$$

with $\alpha \in (0, 0.5]$.

With $W_x(\lambda,t)$ and $W_y(\lambda,t)$, the detail coefficients for the series $x(t)$ and $y(t)$. The test in the absence of the assumptions for the statistical distribution, does not facilitate determining its distribution for $R(\lambda)$, hence, similar to the technique implemented in Maharaj [1], we use a non parametric to obtain an approximation of distribution $R(\lambda)$, through the calculation of a big number of $R_1(\lambda)$, $R_2(\lambda)$,..., $R_s(\lambda)$, which arise when a large number of changes among the wavelet coefficients for both series are performed. So, the p-valor of this test is obtained by determining the proportion of the values of $R_1(\lambda)$, $R_2(\lambda)$,..., $R_s(\lambda)$, which arise when $m = 4000$ exchanges between the wavelet coefficients for both series are performed (10), (11),(12) and (13).

This type of estimators are then, a contribution of effective application for the identification of series belonging to the same model. Notice that the estimator given in expression (8) is the introduced by Maharaj [1] which is our reference point to compare. While estimators (7), (9), (10), (11), (12) and (13) are those that we propose in this paper. One can observe in following section that our methodology works one better, in terms of power an size as well as when outlier data are present, than the methodology used for comparison.

The main contribution made by us to the methodology introduced in Maharaj [1], is the adaptation of the so called robust estimators, which present less sensitivity to the appearance of outlier values (Grané & Veiga [8]). The median and trimmed mean display themselves as estimators that allow to tell when a series is of any strange type, and they posses the quality of a good performance in presence of bias and variance independent of the nature of the series, thus helping the correct interpretation of the information.

4 Simulation and results

In this section we study the behaviour of the randomization test in terms of size and power for finite samples, evaluating the sensitivity of the test in several samples. A significant level of $\alpha = 0.05$ was assumed. The simulation was performed using series of different lengths. Thus, we generated integrated autoregressive processes of ARIMA moving average (p, q) and, in addition, conditional autoregressive heteroscedastic models ($GARCH$ models), which allow establishing parameters for the dependence between high order series and the evolution of volatility. key-27. Since its introduction to the literature by Engle [28] and Bollerslev [29], respectively, they have expanded in several directions.

The $ARIMA(1, d, 1)$ process simulation, with the following specification:

$$(1 - \phi B)(1 - B)^d X(t) = (1 - \theta B)\varepsilon(t) \tag{14}$$

with $\varepsilon_t \sim N(0, 1)$ as a white noise process, and where ϕ and θ are the parameters for auto-regressive and moving average respectively.

The simulation with with conditional auto-regressive heteroscedasticity models like the $GARCH(1, 1)$ model, proposed by Bollerslev (1986), with the following specification:

$$Y_t = \mu + \varepsilon_t \iota \tag{15}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{16}$$

with μ the conditional median for performance of the active Y_t, and the equation for the conditional variance σ_t^2 equal to the sum of the mean α_0, plus the information about volatility in the previous period, measured as a delay of the residuals squared in the mean equation ε_{t-1}^2 the term $ARCH$, and the last estimated period for the variance σ_{t-1}^2 the term $GARCH$.

Therefore, through a set of a thousand series generated from the previos models with different parameters, the sensitivity of the test was proved. Series lenghts of $T = 512$ were chosen, and were considered by their wavelet function characterization, thus proving both the suggested norms as well as the proposed robust estimators. For $T = 512$ the series is broken down into nine scales with a scale coefficient of 2^{j-1}, with the guarantee of having enough detail coefficients to perform the test. In this example, it was decided to decompose the series into three scales, 2^{j-1}, with $j = 1, 2$ and 3. Four thousand exchanges were used in the random assignment, the results are shown below.

The approach for the randomization test requires a sensitivity test, in which, through the calculation of power and size we determined the reliability of said test. For power calculation, pairs pertaining to different models were compared, the randomization test with one of the proposed statistics, procuring the p_value, which in this case requires values close to one. For size calculation series from the same model were evaluated, showing values close to zero. Next, the obtained results in terms of power ans size are shown.

4.1 Power and Size

The power of a test is measured as the probability of rejecting the hypothesis when it must be rejected while, the size of a test is the probability of rejecting what must not be rejected; that means that a good test is one that keep a lower size and a high power, reducing the errors of type one and two to the minimum[28].

In order to evaluate how is the performance of the test regarding power, 1000 trajectories for each process are generated so then, they are tested one by one with the ones of the other processes and the mean of the results (rejections over total amount of trajectories) is calculated. The size of the test required to randomly reorganize the set of trajectories allowing to reproduce the process of the power, with the organized and disorganized set aiming to get very low values.

All the values were obtained with a significance level of 5% and a configuration of parameters mother wavelet as daubechies 1, decomposition level 1, $T = 512$ and the chosen functions. In the case of Power reasonably significant values are evident for both sizes of the samples, in all scales and the contrast with the Euclidean norm was overcome in almost all the proposed estimators. Next, we show the power and size results using the Euclidean norm as it was done in Maharaj [1] followed by the results of the different proposals for $T = 512$.

Following Tables can be read as the estimated size of test in diagonal, while in other cases is stated the estimated power of test. We present here several frameworks where can be seen the good performance of our strategies.

Table 1: Power and size estimates for statistical based on Euclidean Standard, T=512. Maharaj [1]

	AR(1) 0.2	**AR(1)** 0.4	**AR(1)** 0.6	**AR(1)** 0.8	**MA(1)** 0.1	**ARMA(1,1)** 0.7,0.3	**ARIMA(1,1,1)** -0.5,0.4	**GARCH(1,1)** 0.5,0.3
AR(1) 0.2	0.086	0.628	0.998	1.000	0.318	1.000	1.000	1.000
AR(1) 0.4		0.096	0.908	1.000	0.940	0.998	1.000	1.000
AR(1) 0.6			0.228	1.000	1.000	0.784	1.000	0.998
AR(1) 0.8				0362	1.000	0.882	1.000	0.348
MA(1) 0.1					0.068	1.000	1.000	1.000
ARMA(1,1) 0.7,0.3						0.322	1.000	0.686
ARIMA(1,1,1) -0.5,0.4							0.904	0.996
GARCH 0.5,0.3								0.360

Table 2: Power and size estimates for statistical based on Standard L_1, T=512

	AR(1) 0.2	**AR(1)** 0.4	**AR(1)** 0.6	**AR(1)** 0.8	**MA(1)** 0.1	**ARMA(1,1)** 0.7,0.3	**ARIMA(1,1,1)** -0.5,0.4	**GARCH(1,1)** 0.5,0.3
AR(1) 0.2	0.052	0.986	1.000	1.000	0.908	1.000	1.000	0.094
AR(1) 0.4		0.060	0.983	1.000	1.000	1.000	1.000	0.747
AR(1) 0.6			0.060	1.000	1.000	0.977	1.000	0.983
AR(1) 0.8				0.084	1.000	1.000	1.000	1.000
MA(1) 0.1					0.060	1.000	1.000	0.559
ARMA(1,1) 0.7,0.3						0.086	1.000	0.993
ARIMA(1,1,1) -0.5,0.4							0.874	1.000
GARCH 0.5,0.3								0.094

Table 3: Power and size estimates for statistical based on Standard L_∞, T=512

	AR(1) 0.2	**AR(1)** 0.4	**AR(1)** 0.6	**AR(1)** 0.8	**MA(1)** 0.1	**ARMA(1,1)** 0.7,0.3	**ARIMA(1,1,1)** -0.5,0.4	**GARCH(1,1)** 0.5,0.3
AR(1) 0.2	0.026	0.796	1.000	1.000	0.483	0.826	1.000	0.095
AR(1) 0.4		0.024	0.980	1.000	1.000	1.000	1.000	0.111
AR(1) 0.6			0.016	1.000	1.000	1.000	1.000	0.511
AR(1) 0.8				0.016	1.000	1.000	1.000	0.907
MA(1) 0.1					0.034	0.048	1.000	0.502
ARMA(1,1) 0.7,0.3						0.069	1.000	0.166
ARIMA(1,1,1) -0.5,0.4							0.820	1.000
GARCH 0.5,0.3								0.040

The simulation exercise, beginning with the application of the test for standards 2 (Maharaj's proposal), norm 1 and the infinite norm. The simulation results show a better performance for standard 1 in terms of size and power. One of the first considerations of the work is to improve the proposed statistic, the process previously presented presents a substantial improvement when establishing $R(\lambda)$ by means of the quotient of the sum of absolute values between the wavelets coefficients for the two series. Next, the proposal of robust estimators is presented using some of the norms established above.

Table 4: Power and size estimates for statistical based on Median for norm 1, T=512

	AR(1) 0.2	**AR(1)** 0.4	**AR(1)** 0.6	**AR(1)** 0.8	**MA(1)** 0.1	**ARMA(1,1)** 0.7,0.3	**ARIMA(1,1,1)** -0.5,0.4	**GARCH(1,1)** 0.5,0.3
AR(1) 0.2	0.069	0.981	1.000	1.000	0.688	1.000	1.000	1.000
AR(1) 0.4		0.094	0.999	1.000	1.000	1.000	1.000	1.000
AR(1) 0.6			0.168	1.000	1.000	0.946	1.000	1.000
AR(1) 0.8				0.346	1.000	1.000	1.000	0.952
MA(1) 0.1					0.055	1.000	1.000	1.000
ARMA(1,1) 0.7,0.3						0.215	1.000	0.688
ARIMA(1,1,1) -0.5,0.4							0.912	1.000
GARCH 0.5,0.3								0.274

Table 5: Power and size estimates for statistical based on Median for norm 2, T=512

	AR(1) 0.2	**AR(1)** 0.4	**AR(1)** 0.6	**AR(1)** 0.8	**MA(1)** 0.1	**ARMA(1,1)** 0.7,0.3	**ARIMA(1,1,1)** -0.5,0.4	**GARCH(1,1)** 0.5,0.3
AR(1) 0.2	0.044	0.967	1.000	1.000	0.826	1.000	1.000	0.352
AR(1) 0.4		0.054	1.000	1.000	1.000	1.000	1.000	0.978
AR(1) 0.6			0.054	1.000	1.000	0.944	1.000	1.000
AR(1) 0.8				0.078	1.000	0.999	1.000	1.000
MA(1) 0.1					0.048	1.000	1.000	0.166
ARMA(1,1) 0.7,0.3						0.070	1.000	1.000
ARIMA(1,1,1) -0.5,0.4							0.820	1.000
GARCH 0.5,0.3								0.008

Table 6: Power and size estimates for statistical based on Trimmed Mean of norm 2, $\alpha = 5\%$ T=512

	AR(1) 0.2	**AR(1)** 0.4	**AR(1)** 0.6	**AR(1)** 0.8	**MA(1)** 0.1	**ARMA(1,1)** 0.7,0.3	**ARIMA(1,1,1)** -0.5,0.4	**GARCH(1,1)** 0.5,0.3
AR(1) 0.2	0.048	0.991	1.000	1.000	0.904	1.000	1.000	0.124
AR(1) 0.4		0.052	1.000	1.000	1.000	1.000	1.000	0.791
AR(1) 0.6			0.062	1.000	1.000	0.991	1.000	0.990
AR(1) 0.8				0.088	1.000	1.000	1.000	0.999
MA(1) 0.1					0.066	1.000	1.000	0.516
ARMA(1,1) 0.7,0.3						0.080	1.000	0.996
ARIMA(1,1,1) -0.5,0.4							0.820	0.884
GARCH 0.5,0.3								0.008

Table 7: Power and size estimates for statistical based on Trimmed Mean of norm 2, $\alpha = 10\%$ T=512

	AR(1) 0.2	**AR(1)** 0.4	**AR(1)** 0.6	**AR(1)** 0.8	**MA(1)** 0.1	**ARMA(1,1)** 0.7,0.3	**ARIMA(1,1,1)** -0.5,0.4	**GARCH(1,1)** 0.5,0.3
AR(1) 0.2	0.050	0.990	1.000	1.000	0.903	1.000	1.000	0.179
AR(1) 0.4		0.054	1.000	1.000	1.000	1.000	1.000	0.891
AR(1) 0.6			0.077	1.000	1.000	0.987	1.000	0.990
AR(1) 0.8				0.084	1.000	1.000	1.000	0.104
MA(1) 0.1					0.066	1.000	1.000	0.460
ARMA(1,1) 0.7,0.3						0.070	1.000	1.000
ARIMA(1,1,1) -0.5,0.4							0.878	0.884
GARCH 0.5,0.3								0.008

Table 8: Power and size estimates for statistical based on Trimmed Mean of norm 1, $\alpha = 5\%$ T=512

	AR(1) 0.2	**AR(1)** 0.4	**AR(1)** 0.6	**AR(1)** 0.8	**MA(1)** 0.1	**ARMA(1,1)** 0.7,0.3	**ARIMA(1,1,1)** -0.5,0.4	**GARCH(1,1)** 0.5,0.3
AR(1) 0.2	0.074	0.986	1.000	1.000	0.809	1.000	1.000	1.000
AR(1) 0.4		0.120	1.000	1.000	1.000	1.000	1.000	1.000
AR(1) 0.6			0.204	1.000	1.000	0.985	1.000	1.000
AR(1) 0.8				0.377	1.000	1.000	1.000	0.565
MA(1) 0.1					0.054	1.000	1.000	1.000
ARMA(1,1) 0.7,0.3						0.262	1.000	0.985
ARIMA(1,1,1) -0.5,0.4							0.944	1.000
GARCH 0.5,0.3								0.351

Table 9: Power and size estimates for statistical based on Trimmed Mean of norm 1, $\alpha = 10\%$ T=512

	AR(1) 0.2	**AR(1)** 0.4	**AR(1)** 0.6	**AR(1)** 0.8	**MA(1)** 0.1	**ARMA(1,1)** 0.7,0.3	**ARIMA(1,1,1)** -0.5,0.4	**GARCH(1,1)** 0.5,0.3
AR(1) 0.2	0.079	0.981	1.000	1.000	0.811	1.000	1.000	1.000
AR(1) 0.4		0.108	1.000	1.000	1.000	1.000	1.000	1.000
AR(1) 0.6			0.212	1.000	1.000	0.985	1.000	1.000
AR(1) 0.8				0.380	1.000	1.000	1.000	0.704
MA(1) 0.1					0.058	1.000	1.000	1.000
ARMA(1,1) 0.7,0.3						0.272	1.000	0.975
ARIMA(1,1,1) -0.5,0.4							0.944	1.000
GARCH 0.5,0.3								0.315

The power estimations show a desired effect for the classification of series coming from the same model, through values close to one, for the majority of estimations, for the selected scales for wavelets $db1$, $db2$ and $db3$. The power calculation improves with increasing sample size. The proposed estimators, median and trimmed mean, show a consistent behavior. It is evident that for processes with auto-regressive parameters, the proceedings show a remarkable high performance.

The size estimations tend to increase as the scale size increases, this means as the series is decomposed in a superior level. On the other hand, the values for size estimations gave non significant scores for all estimators generated through non linear processes $GARCH(1,1)$, as well as auto-regressive processes with parameter $\phi = 0.5$. They show high power for identification of pairs of series of the same model.

In general, we have a very consistent proceeding, which works well for the simulation scenarios that were considered. Plus, we believe the motivation for regarding our proceeding contributes in high degree to robust inference. In addition, any strategy introduced in this paper, as can be seen in previous Tables, has better performance in power an size than the strategy used for comparison. One also can see in the following section that performance is still better when outlier data are present.

4.2 Contaminated data

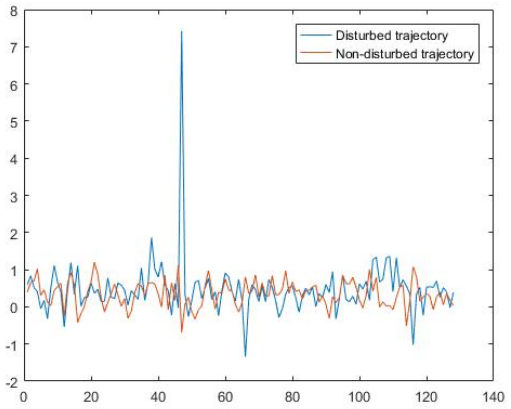
Once the behavior of the different statistics for the chosen eight processes has been analyzed, it is important to see how these measures behave in presence of contaminated data in order to qualify how robust they are. For this purpose

an experiment, below explained, is designed:

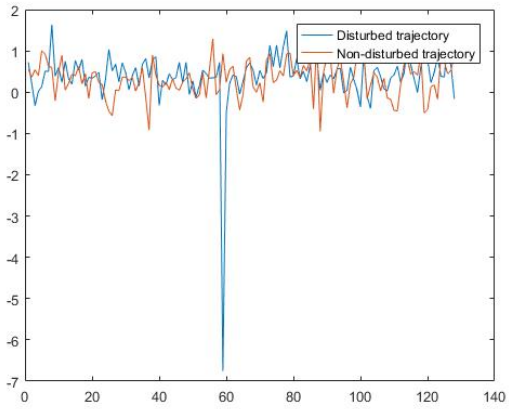
For this simulations, trajectories of $T = 128$ for each type of process, are considered (which affects performance presented in the section above) and treated with the DWT (equation 3), daubichies 1 and decomposition level 1. That defined, aiming to preserve disturbances like that, the contamination level is selected, for every case, as 5%, 10% and 20% with withe noise of high variance (compared with the ones belonging to processes) and then they are submitted to the same excersice of the previous section, where size and power are tested. It is remarked that this experiment is executed contaminating a whole set of series (X) and both sets of series (X, Y) in order to represent a reality when one series is contaminated and other where both of them are.

Below among with the results of size and power for the statistics, mentioned in equations 7 to 13, are also presented graphics of disturbed series with not disturbed ones.

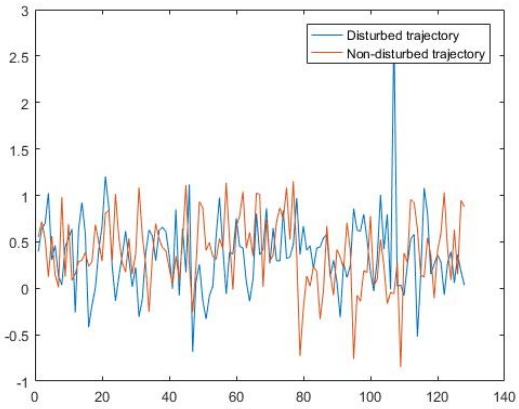
Figure 2: Contaminated series among with not contaminated ones



(a) 2% de Series Outliers



(b) 5% de Series Outliers



(c) 10% de Series Outliers

1. Norm 1

2. Median for norm 1.
   This measure seems to keep size relatively stable with no big changes as well as it preserves most of the power of the proof having as a maximum decayment a 5%

Table 10: Contamination level = 0% with median for norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0780	0.5020	0.9840	1.0000	0.2440	0.9980	1.0000	1.0000
AR(1)		0.0960	0.8100	1.0000	0.8660	0.9680	1.0000	1.0000
AR(1)			0.1680	0.9980	1.0000	0.5820	1.0000	0.9720
AR(1)				0.3100	1.0000	0.8720	1.0000	0.6600
MA(1)					0.0560	1.0000	1.0000	1.0000
ARMA(1,1)						0.2640	1.0000	0.4440
ARIMA(1,1,1)							0.8200	1.0000
GARCH(1,1)								0.3200

Source: Own elaboration by means of data simulation.

Table 11: Contamination level = 5% with median for norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0700	0.4780	0.9880	1.0000	0.1980	0.9960	1.0000	1.0000
AR(1)		0.1080	0.7880	1.0000	0.8240	0.9720	1.0000	1.0000
AR(1)			0.1680	1.0000	0.9980	0.6480	1.0000	0.9720
AR(1)				0.3360	1.0000	0.8660	1.0000	0.6440
MA(1)					0.0640	1.0000	1.0000	1.0000
ARMA(1,1)						0.2620	1.0000	0.4300
ARIMA(1,1,1)							0.8200	1.0000
GARCH(1,1)								0.2320

Source: Own elaboration by means of data simulation.

Table 12: Contamination level = 10% with median for norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0620	0.4780	0.9780	1.0000	0.2200	0.9980	1.0000	1.0000
AR(1)		0.1140	0.8020	1.0000	0.8440	0.9700	1.0000	1.0000
AR(1)			0.1560	1.0000	0.9980	0.6540	1.0000	0.9720
AR(1)				0.2940	1.0000	0.8520	1.0000	0.6460
MA(1)					0.0460	1.0000	1.0000	1.0000
ARMA(1,1)						0.2700	1.0000	0.4320
ARIMA(1,1,1)							0.7780	1.0000
GARCH(1,1)								0.2460

Source: Own elaboration by means of data simulation.

Table 13: Contamination level = 20% with median for norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0660	0.4860	0.9880	1.0000	0.2160	0.9960	1.0000	1.0000
AR(1)		0.1020	0.8080	1.0000	0.8300	0.9760	1.0000	1.0000
AR(1)			0.1400	1.0000	0.9980	0.6560	1.0000	0.9680
AR(1)				0.3400	1.0000	0.8580	1.0000	0.6520
MA(1)					0.0700	1.0000	1.0000	1.0000
ARMA(1,1)						0.2500	1.0000	0.4140
ARIMA(1,1,1)							0.8120	1.0000
GARCH(1,1)								0.2800

Source: Own elaboration by means of data simulation.

3. Trimmed mean at 5% for norm 1.

Table 14: Contamination level = 0% with trimmean 5% for norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0720	0.6180	0.9960	1.0000	0.3100	1.0000	1.0000	1.0000
AR(1)		0.1060	0.8820	1.0000	0.9220	1.0000	1.0000	1.0000
AR(1)			0.2240	1.0000	1.0000	0.6780	1.0000	0.9980
AR(1)				0.3860	1.0000	0.9060	1.0000	0.4780
MA(1)					0.0580	1.0000	1.0000	1.0000
ARMA(1,1)						0.3140	1.0000	0.7300
ARIMA(1,1,1)							0.8840	1.0000
GARCH(1,1)								0.3580

Source: Own elaboration by means of data simulation.

Table 15: Contamination level = 5% with trimmean 5% with norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0600	0.5820	0.9980	1.0000	0.2520	1.0000	1.0000	1.0000
AR(1)		0.1360	0.8740	1.0000	0.9260	0.9920	1.0000	1.0000
AR(1)			0.1860	1.0000	0.9980	0.7460	1.0000	0.9980
AR(1)				0.3380	1.0000	0.9000	1.0000	0.4440
MA(1)					0.0460	1.0000	1.0000	1.0000
ARMA(1,1)						0.2960	1.0000	0.7160
ARIMA(1,1,1)							0.8740	1.0000
GARCH(1,1)								0.3300

Source: Own elaboration by means of data simulation.

Table 16: Contamination level = 10% with trimmean 5% with norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0760	0.5840	1.0000	1.0000	0.2500	1.0000	1.0000	1.0000
AR(1)		0.1080	0.8580	1.0000	0.9160	0.9920	1.0000	1.0000
AR(1)			0.1960	1.0000	0.9980	0.7340	1.0000	0.9980
AR(1)				0.3240	1.0000	0.9060	1.0000	0.4460
MA(1)					0.0480	1.0000	1.0000	1.0000
ARMA(1,1)						0.2820	1.0000	0.7100
ARIMA(1,1,1)							0.9000	1.0000
GARCH(1,1)								0.3480

Source: Own elaboration by means of data simulation.

Table 17: Contamination level = 20% with trimmean 5% with norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0720	0.5820	1.0000	1.0000	0.2580	1.0000	1.0000	1.0000
AR(1)		0.1160	0.8760	1.0000	0.9240	0.9940	1.0000	1.0000
AR(1)			0.1720	1.0000	0.9980	0.7280	1.0000	1.0000
AR(1)				0.3580	1.0000	0.9080	1.0000	0.4560
MA(1)					0.0420	1.0000	1.0000	1.0000
ARMA(1,1)						0.2720	1.0000	0.7080
ARIMA(1,1,1)							0.8860	1.0000
GARCH(1,1)								0.3540

Source: Own elaboration by means of data simulation.

4. Trimeed mean at 10% for norm 1.

Table 18: Contamination level = 0% with trimmean 10% for norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0780	0.5300	0.9960	1.0000	0.3440	1.0000	1.0000	1.0000
AR(1)		0.1340	0.8580	1.0000	0.9220	0.9960	1.0000	1.0000
AR(1)			0.2100	1.0000	1.0000	0.7080	1.0000	0.9960
AR(1)				0.4140	1.0000	0.8980	1.0000	0.4940
MA(1)					0.0700	1.0000	1.0000	1.0000
ARMA(1,1)						0.3120	1.0000	0.7320
ARIMA(1,1,1)							0.8940	1.0000
GARCH(1,1)								0.3660

Source: Own elaboration by means of data simulation.

Table 19: Contamination level = 5% with trimmean 10% with norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0700	0.5200	0.9940	1.0000	0.3180	1.0000	1.0000	1.0000
AR(1)		0.1180	0.8500	1.0000	0.8960	0.9960	1.0000	1.0000
AR(1)			0.2120	1.0000	1.0000	0.7040	1.0000	0.9960
AR(1)				0.3680	1.0000	0.8940	1.0000	0.4700
MA(1)					0.0560	1.0000	1.0000	1.0000
ARMA(1,1)						0.2700	1.0000	0.7040
ARIMA(1,1,1)							0.9040	1.0000
GARCH(1,1)								0.3560

Source: Own elaboration by means of data simulation.

Table 20: Contamination level = 10% with trimmean 10% with norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0820	0.5220	0.9940	1.0000	0.3260	1.0000	1.0000	1.0000
AR(1)		0.1240	0.8600	1.0000	0.9120	0.9960	1.0000	1.0000
AR(1)			0.1820	1.0000	1.0000	0.7040	1.0000	0.9960
AR(1)				0.3880	1.0000	0.8900	1.0000	0.4660
MA(1)					0.0600	1.0000	1.0000	1.0000
ARMA(1,1)						0.3200	1.0000	0.7280
ARIMA(1,1,1)							0.8620	1.0000
GARCH(1,1)								0.3960

Source: Own elaboration by means of data simulation.

Table 21: Contamination level = 20% with trimmean 10% with norm 1

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0740	0.5200	0.9980	1.0000	0.3160	1.0000	1.0000	1.0000
AR(1)		0.1260	0.8580	1.0000	0.9060	0.9960	1.0000	1.0000
AR(1)			0.2020	1.0000	1.0000	0.7160	1.0000	0.9960
AR(1)				0.3840	1.0000	0.9040	1.0000	0.4680
MA(1)					0.0660	1.0000	1.0000	1.0000
ARMA(1,1)						0.3340	1.0000	0.7360
ARIMA(1,1,1)							0.9080	1.0000
GARCH(1,1)								0.3460

Source: Own elaboration by means of data simulation.

As it happens with the median for norm 1, the trimmeans for norm 1 slightly changes size but both of them preserve most of the power with minimum decreses of it (2 or 3%), which makes them considerable better options regarding robustness.

5. Norm 2.
   As it might be expected, norm 2 suffers appreciable changes when contamination appears. In this case, as the amount of disturbances grows the power of the proof reach high decayments of even 30% which makes it quite useless when there are suspicious of disturbances. This brief analysis among with the following tables highlight the bad performance of using this measure regarding robustness for this proof.

Table 22: Contamination level = 0% with norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0860	0.6280	0.9980	1.0000	0.3180	1.0000	1.0000	1.0000
AR(1)		0.0960	0.9080	1.0000	0.9400	0.9980	1.0000	1.0000
AR(1)			0.2280	1.0000	1.0000	0.7840	1.0000	0.9980
AR(1)				0.3620	1.0000	0.8820	1.0000	0.3480
MA(1)					0.0680	1.0000	1.0000	1.0000
ARMA(1,1)						0.3220	1.0000	0.6860
ARIMA(1,1,1)							0.9040	0.9960
GARCH(1,1)								0.3600

Source: Own elaboration by means of data simulation.

Table 23: Contamination level = 5% with norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0620	0.3360	0.7700	0.9960	0.1060	0.9460	1.0000	0.9940
AR(1)		0.0620	0.6160	0.9940	0.5100	0.9320	1.0000	0.9740
AR(1)			0.1300	0.9800	0.8360	0.6600	1.0000	0.9440
AR(1)				0.3180	0.9980	0.8100	1.0000	0.3460
MA(1)					0.0400	0.9540	1.0000	0.9860
ARMA(1,1)						0.2300	1.0000	0.6360
ARIMA(1,1,1)							0.8860	1.0000
GARCH(1,1)								0.3520

Source: Own elaboration by means of data simulation.

Table 24: Contamination level = 10% with norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0380	0.3340	0.8060	0.9980	0.1060	0.9560	1.0000	0.9920
AR(1)		0.0600	0.6160	0.9900	0.5280	0.8980	1.0000	0.9880
AR(1)			0.1220	0.9820	0.8220	0.6600	1.0000	0.9500
AR(1)				0.3320	1.0000	0.8020	1.0000	0.3540
MA(1)					0.0200	0.9460	1.0000	0.9920
ARMA(1,1)						0.2700	1.0000	0.6320
ARIMA(1,1,1)							0.8980	1.0000
GARCH(1,1)								0.3340

Source: Own elaboration by means of data simulation.

Table 25: Contamination level = 20% with norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0400	0.3240	0.7940	0.9980	0.1200	0.9380	1.0000	0.9840
AR(1)		0.0580	0.6420	0.9940	0.5940	0.9060	1.0000	0.9900
AR(1)			0.1160	0.9780	0.8160	0.6820	1.0000	0.9440
AR(1)				0.2980	1.0000	0.7880	1.0000	0.3660
MA(1)					0.0200	0.9600	1.0000	0.9920
ARMA(1,1)						0.2160	1.0000	0.6340
ARIMA(1,1,1)							0.9020	1.0000
GARCH(1,1)								0.3320

Source: Own elaboration by means of data simulation.

6. Median for norm 2.

Table 26: Contamination level = 0% with median for norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0820	0.5040	0.9840	1.0000	0.2320	0.9980	1.0000	1.0000
AR(1)		0.0980	0.8180	1.0000	0.8640	0.9640	1.0000	1.0000
AR(1)			0.1900	0.9980	1.0000	0.5860	1.0000	0.9720
AR(1)				0.3340	1.0000	0.8700	1.0000	0.6580
MA(1)					0.0640	1.0000	1.0000	1.0000
ARMA(1,1)						0.2360	1.0000	0.4400
ARIMA(1,1,1)							0.8140	1.0000
GARCH(1,1)								0.2920

Source: Own elaboration by means of data simulation.

Table 27: Contamination level = 5% with median for norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0620	0.4960	0.9840	1.0000	0.1880	0.9980	1.0000	1.0000
AR(1)		0.1100	0.8060	1.0000	0.8320	0.9740	1.0000	1.0000
AR(1)			0.1880	1.0000	0.9980	0.6400	1.0000	0.9680
AR(1)				0.3160	1.0000	0.8560	1.0000	0.6260
MA(1)					0.0700	1.0000	1.0000	1.0000
ARMA(1,1)						0.2420	1.0000	0.4240
ARIMA(1,1,1)							0.8000	1.0000
GARCH(1,1)								0.2660

Source: Own elaboration by means of data simulation.

Table 28: Contamination level = 10% with median for norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0680	0.4880	0.9860	1.0000	0.2080	0.9980	1.0000	1.0000
AR(1)		0.1140	0.8000	1.0000	0.8400	0.9740	1.0000	1.0000
AR(1)			0.1580	0.9980	0.9980	0.6540	1.0000	0.9680
AR(1)				0.3260	1.0000	0.8560	1.0000	0.6420
MA(1)					0.0680	1.0000	1.0000	1.0000
ARMA(1,1)						0.2660	1.0000	0.4380
ARIMA(1,1,1)							0.8200	1.0000
GARCH(1,1)								0.2400

Source: Own elaboration by means of data simulation.

Table 29: Contamination level = 20% with median for norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0580	0.4960	0.9880	1.0000	0.2020	0.9960	1.0000	1.0000
AR(1)		0.1180	0.7880	1.0000	0.8400	0.9700	1.0000	1.0000
AR(1)			0.1540	0.9980	0.9980	0.6580	1.0000	0.9700
AR(1)				0.3180	1.0000	0.8640	1.0000	0.6500
MA(1)					0.0720	1.0000	1.0000	1.0000
ARMA(1,1)						0.2480	1.0000	0.4200
ARIMA(1,1,1)							0.8180	1.0000
GARCH(1,1)								0.2740

Source: Own elaboration by means of data simulation.

7. Trimeed mean at 5% for norm 2.

Table 30: Contamination level = 0% with trimmean 5% for norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0860	0.6340	0.9980	1.0000	0.3240	1.0000	1.0000	1.0000
AR(1)		0.1140	0.9080	1.0000	0.9440	0.9980	1.0000	1.0000
AR(1)			0.2300	1.0000	1.0000	0.7800	1.0000	0.9960
AR(1)				0.3680	1.0000	0.8940	1.0000	0.4440
MA(1)					0.0580	1.0000	1.0000	1.0000
ARMA(1,1)						0.3240	1.0000	0.6560
ARIMA(1,1,1)							0.8800	1.0000
GARCH(1,1)								0.3260

Source: Own elaboration by means of data simulation.

Table 31: Contamination level = 5% with trimmean 5% with norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0660	0.5920	1.0000	1.0000	0.2380	1.0000	1.0000	1.0000
AR(1)		0.0940	0.8740	1.0000	0.9280	0.9940	1.0000	1.0000
AR(1)			0.1740	1.0000	0.9980	0.7800	1.0000	0.9960
AR(1)				0.3160	1.0000	0.8940	1.0000	0.4260
MA(1)					0.0460	1.0000	1.0000	1.0000
ARMA(1,1)						0.2920	1.0000	0.6500
ARIMA(1,1,1)							0.9060	1.0000
GARCH(1,1)								0.3240

Source: Own elaboration by means of data simulation.

Table 32: Contamination level = 10% with trimmean 5% with norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0740	0.5860	1.0000	1.0000	0.2340	1.0000	1.0000	1.0000
AR(1)		0.1220	0.8860	1.0000	0.9080	0.9980	1.0000	1.0000
AR(1)			0.1780	1.0000	0.9980	0.7940	1.0000	0.9920
AR(1)				0.3360	1.0000	0.8880	1.0000	0.4300
MA(1)					0.0420	1.0000	1.0000	1.0000
ARMA(1,1)						0.2640	1.0000	0.6340
ARIMA(1,1,1)							0.9100	1.0000
GARCH(1,1)								0.3600

Source: Own elaboration by means of data simulation.

Table 33: Contamination level = 20% with trimmean 5% with norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0640	0.5940	1.0000	1.0000	0.2420	1.0000	1.0000	1.0000
AR(1)		0.1060	0.8680	1.0000	0.9320	0.9960	1.0000	1.0000
AR(1)			0.1980	1.0000	0.9980	0.7960	1.0000	0.9960
AR(1)				0.3400	1.0000	0.8900	1.0000	0.4240
MA(1)					0.0320	1.0000	1.0000	1.0000
ARMA(1,1)						0.2620	1.0000	0.6420
ARIMA(1,1,1)							0.8960	1.0000
GARCH(1,1)								0.3400

Source: Own elaboration by means of data simulation.

8. Trimeed mean at 10% for norm 2.

Table 34: Contamination level = 0% with trimmean 10% for norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0720	0.6320	0.9980	1.0000	0.3280	1.0000	1.0000	1.0000
AR(1)		0.1440	0.9000	1.0000	0.9480	1.0000	1.0000	1.0000
AR(1)			0.2380	1.0000	1.0000	0.7640	1.0000	0.9980
AR(1)				0.3740	1.0000	0.9000	1.0000	0.4760
MA(1)					0.0720	1.0000	1.0000	1.0000
ARMA(1,1)						0.3300	1.0000	0.6380
ARIMA(1,1,1)							0.9120	1.0000
GARCH(1,1)								0.3840

Source: Own elaboration by means of data simulation.

Table 35: Contamination level = 5% with trimmean 10% with norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0620	0.6040	0.9980	1.0000	0.2640	1.0000	1.0000	1.0000
AR(1)		0.0700	0.8760	1.0000	0.9380	0.9960	1.0000	1.0000
AR(1)			0.1800	1.0000	1.0000	0.7800	1.0000	0.9940
AR(1)				0.3440	1.0000	0.8960	1.0000	0.4460
MA(1)					0.0540	1.0000	1.0000	1.0000
ARMA(1,1)						0.2600	1.0000	0.6260
ARIMA(1,1,1)							0.8900	1.0000
GARCH(1,1)								0.3420

Source: Own elaboration by means of data simulation.

Table 36: Contamination level = 10% with trimmean 10% with norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0820	0.6060	1.0000	1.0000	0.2520	1.0000	1.0000	1.0000
AR(1)		0.1200	0.8940	1.0000	0.9240	0.9960	1.0000	1.0000
AR(1)			0.1780	1.0000	0.9980	0.7780	1.0000	0.9920
AR(1)				0.3420	1.0000	0.8820	1.0000	0.4640
MA(1)					0.0480	1.0000	1.0000	1.0000
ARMA(1,1)						0.2740	1.0000	0.6280
ARIMA(1,1,1)							0.8900	1.0000
GARCH(1,1)								0.3620

Source: Own elaboration by means of data simulation.

Table 37: Contamination level = 20% with trimmean 10% with norm 2

	AR(1) 0.2	AR(1) 0.4	AR(1) 0.6	AR(1) 0.8	MA(1) 0.1	ARMA(1,1) 0.7, 0.3	ARIMA(1,1,1) -0.5, 1, 0.4	GARCH(1,1) 0.5, 0.3
AR(1)	0.0520	0.5960	1.0000	1.0000	0.2600	1.0000	1.0000	1.0000
AR(1)		0.1080	0.8700	1.0000	0.9280	0.9940	1.0000	1.0000
AR(1)			0.1700	1.0000	0.9980	0.7860	1.0000	0.9920
AR(1)				0.3280	1.0000	0.9000	1.0000	0.4520
MA(1)					0.0500	1.0000	1.0000	1.0000
ARMA(1,1)						0.2580	1.0000	0.6320
ARIMA(1,1,1)							0.8840	1.0000
GARCH(1,1)								0.3660

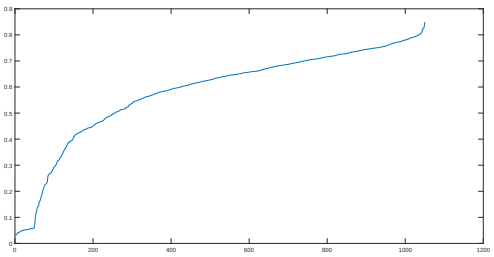
Source: Own elaboration by means of data simulation.

At this point, the behaviour of the robust versions of norm 1 and norm 2 has been seen and all of those robust measures remains almost equal even after disturbances are included. For all of this new versions of the proof size and power are well behaved are quite better than the traditional norm 2 both in size and power. That leads to suggest the use of any of this new measures for quite a better performance, especially the trimmeans, when robustness is required.

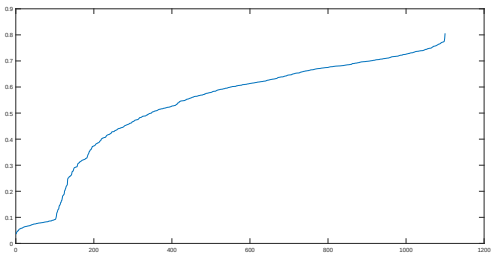
Emphasis is placed on the remarkable behaviour of the proposed measures trimmeans in presence of disturbances.The current theory suggests the need to explore techniques that allow a better analysis of strange data, or data without prior connections. The analysis of these so called strange data is based only in classical modelling techniques that do not respond to the necessities arising from the finding of atypical series, with origins in unexpected or uncontrolled interventions. This makes necessary to evaluate the effect of this type of observations, with the aim of improving the comprehension of the series being analysed, modelled, estimation, intervention analysis and prediction quality [1].

With respect to the medium estimator, which presented a high yield, two series of polluted series are presented at 5% and 10%, following the previous methodology, an analysis is made identifying the one we will call "breaking point" and that shows a different configuration, by simulating a set of 1000 series belonging to model $ARMA(p, q)$, contaminated at 5% and 10% respectfully, and based on the use of the randomization test, the vector of p_values, which is the average per row of the generated matrix for each pair of series. The graph number 3 shows the point of rupture, the exact identification of this is planted for later investigations.

Figure 3: Identification of Outliers series by means of sample contamination at different levels



(a) 5% de Series Outliers



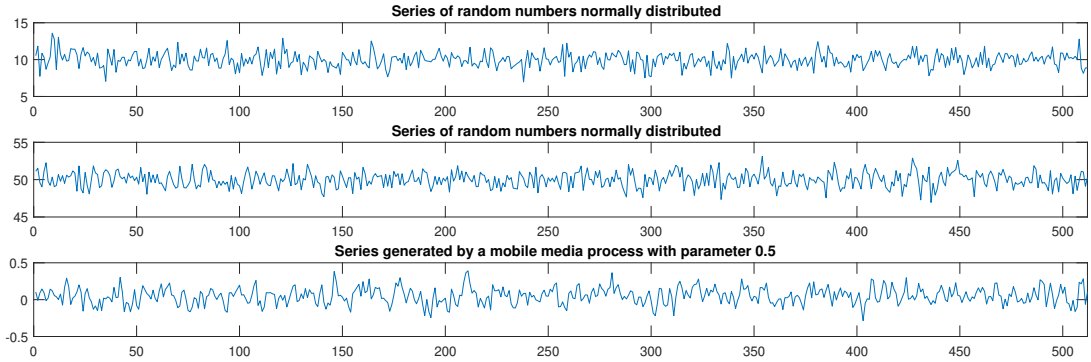
(b) 10% de Series Outliers

4.3 Clustering for time series

By means of a grouping analysis based on wavelets cite key-29, a procedure is added that allows complementing the theory previously exposed, and improves the classification of time series. Conglomerate analysis is useful when the set of series is large. The current literature offers a great variety of grouping algorithms, with the aim of connecting objects by means of a cluster under structured criteria, identifying them with measures of similarity that can be understood as relations of proximity in the data.

The prove method, is one of the most popular techniques for grouping. We can define it as a quantitative method, unsupervised, iterative and non-deterministic. The importance of this analysis lies in the necessity of handling in a different and practical way data sets of considerable length, which we can relate naturally with wavelets analysis. The K-mean algorithm works in an iterative fashion, dividing the initial data set in a K number of clusters, indicated as a parameter [30]. It is based on the minimization of the intern distance. The algorithm proceeds by choosing k centers for initial cluster(centroid), this assigns each observation of a data set to the cluster with the closest centroid by determining the distance between this data point and the group centroid. Then it determines a new mean for each group as the centroid of observations. If the data point is closest to the centroid of its oun cluster, it stays there, while, if the data point is not close to this centroid, but to a different one, it will be assigned to the closest centroid. The process is an iterative one.

Figure 4: Examples of series used in the simulation

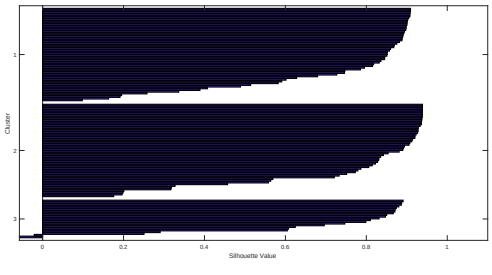


Thus, using the modification of the randomization test, and using the robust estimator of the median which showed a high efficiency, we simulated a set of 130 series $x(t)$ distributed in three groups, two of which presented a normal distribution with 50 and 30 random series respectively, and a set of 50 from model $MA(1)$ with parameter $\varphi = 0.5$, all with a length $T = 512$, the objectiv is achieve the identification of groups by using the modified test based on the discrete wavelet transform. We build a matrix which contains the p_value of size 130×130, the interactions between pairs of series taken from each one are considered their coefficients of detail. From the matrix a vector of row averages is elaborated, to which k-means is applied, taking into account a cluster selection criterion. The analysis was tested for scales with $j = 1, 2$ and 4, and taking into account functions belonging to the Daubechies family. The results are shown below:

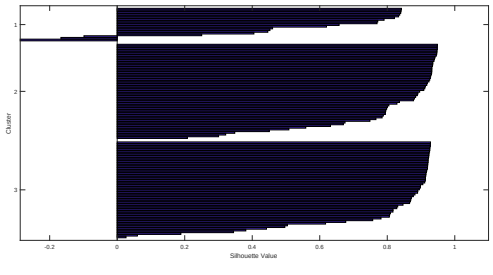
Table 38: Clustering through the use of the median

	k		**j=1**		**j=2**		**j=4**
	1	54	41.54%	29	22.31%	49	37.69%
db1	**2**	54	41.54%	54	41.54%	54	41.54%
	3	22	16.92%	47	36.15%	27	20.77%
	k		**j=1**		**j=2**		**j=4**
	1	19	14.62%	42	32.31%	52	40.00%
db2	**2**	55	42.31%	68	52.31%	58	44.62%
	3	56	43.08	20	15.38	20	15.38
	k		**j=1**		**j=2**		**j=4**
	1	65	50.00%	36	27.69%	63	48.46%
db3	**2**	41	31.54%	53	40.77%	22	16.92%
	3	24	18.46%	41	31.54%	45	34.62%

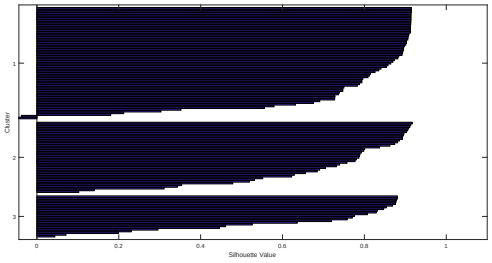
Figure 5: Wavelet shadow-graphs



(a) Daubechies 1



(b) Daubechies 2



(c) Daubechies 3

The previous table show consistent values for groups in three clusters, which are established with scale $j = 1$, it tends to lose the approximation for bigger decomposition scales.

The use of the k-means method contrasts an appropriate number of conglomerates by means of the shadow graph method proposed by Rousseeuw (1987).The diagram confirms the election for the number of clusters, its value for each point is a means of comparison for this point with points in other groups. A high shadow value indicates P_i is well paired to its own group and the opposite regarding other neighboring groups. If more points have a high shadow value, the clustering solution is appropriate. If, on the other hand, the values are low or negative, the clustering solution may have either a lot or not enough branches. Shadow-graph clustering can be used with any distance metrics.

5 Aplication

We use chronological maximum temperature series expressed in Celsius degrees, observed in 50 provinces and 2 autonomic cities in Spain for the years between 1990 and 2004. Data taken from severe heat for temperature in Mahajarah and Alonso, and D'urso [23]. Figure number 5 exhibits the usual temperature behavior with regular season variations, without remarkable trends along the years. On the other hand, figure 7 shows the Spanish provinces. At present time several authors focus their studies in the Iberian peninsula, among others we can name Alonso, De Zea Bermúdez and Scotto[33], Fernández-Montes and Rodrigo[34], Furió and Meneu [35], Brunet et al.[36], and García-Herrera et al.[37].

Temperatures in Spain depend on the complexity of geography, which implies radical differences between the regions. The year is marked by the first six months with the lowest average temperatures. The highest averages are not exclusive to July, there are regions such as the coastal, Balearic, Canary, Ceuta and Melilla observatories, in which August remains warm with a minimum temperature difference with respect to the previous month, due to the thermal inertia in deep water.

First, we present the the use of the ramdomization algorithm based on the median robust estimator, through which a series set corresponding to 5% of the more different among each other was identified (outlier series pinpointing). This is achieved by means of the p_value calculation. Therefore, the algorithm shown in previous sections generates a 52x52 matrix containing the p_value presenting the lowest significance levels, that in turn identify the temperature series with the most difference among each other. We used wavelet filters db1 (Haar), db2 and db3, at three decomposition levels. The results are shown next:

Table 39: *P_value* smallest of the set of Temperatures of Spain

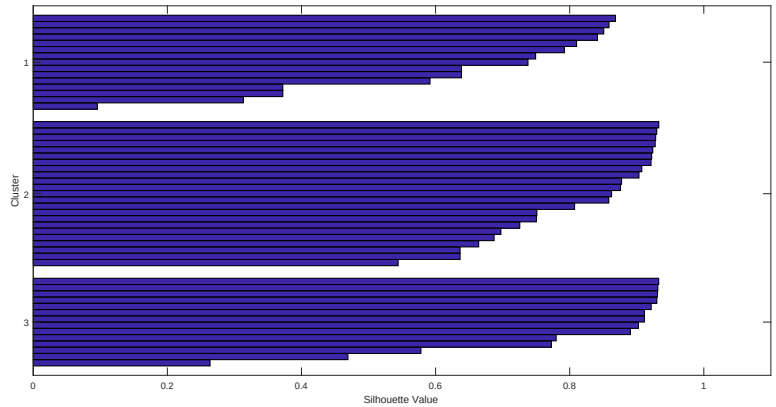
	db1		db2		db3	
	p_value	**Ciudad**	**p_value**	**Ciudad**	**p_value**	**Ciudad**
	0.0242	Cordoba	0.0241	Cordova	0.0267	Huelva
j=1	0.0252	Sevilla	0.0254	Soria	0.0269	Avila
	0.0259	Alemeria	0.0257	Huelva	0.0372	Sevilla
	0.0242	Soria	0.0246	Castellon	0.0279	Avila
j=2	0.0246	Sevilla	0.0256	Sevilla	0.0323	Huelva
	0.0247	Melilla	0.0264	Melilla	0.0338	Sevilla
	0.0242	Melilla	0.0242	Cordova	0.0266	Huelva
j=4	0.0245	Castellon	0.0258	Castellon	0.0326	Sevilla
	0.0246	Sevilla	0.0272	Melilla	0.0392	Melilla

The above table shows a consistent identification of 5%, provinces whose temperature observations are the most different among the group, taking into account three types of wavelet filters. Córdoba, Huelva and Sevilla correspond to the provinces with particularly characterized temperatures. The temperature threshold in Córdoba and Sevilla in summer time can surpass 40 Celsius degrees. Sometimes Sevilla has even registered up to 47 Celsius degress (July 7, 1959). During the heat wave in July, 1995, Sevilla and Córdoba got up to 46,6 Celsius degrees [38].

Semi-desertic Spain has its main representation in peninsular SouthEast, this is, a big part of Almería province, whose series identified the test; regarding geography of the location, which links it with an extreme pluviometric level, where the sequences built for consecutive dry days achieve long duration in south of Spain, not only in the middle warm part of the year. Historically, periods of up to 5 consecutive months have been recorded, followed by drought in provinces like Málaga, Almería and Huelva.

Identification of outliers series suggest as a complement of the analysis, cluster grouping. By using k-means method, the 52 series were assembled, through the silhouette diagram in figure 6, thus validating three groups. This graph delimits the number of established groups, since the shadow values are above 0.8, which corresponds to a nice definition of grouping. Cophenetic coefficient gives a value of 0.86, consistent in terms of grouping correlation.

Figure 6: Shadow-graph for temperature grouping in Spain



An assembly in three groups suggests a suitable climatic location, considering climatic regions in Spain, the north being characterized by a humid oceanic weather, the center with continental mediterranean weather and south with a mediterranean weather.

Figure 7: Grouping for the 52 temperature series in Spain provinces

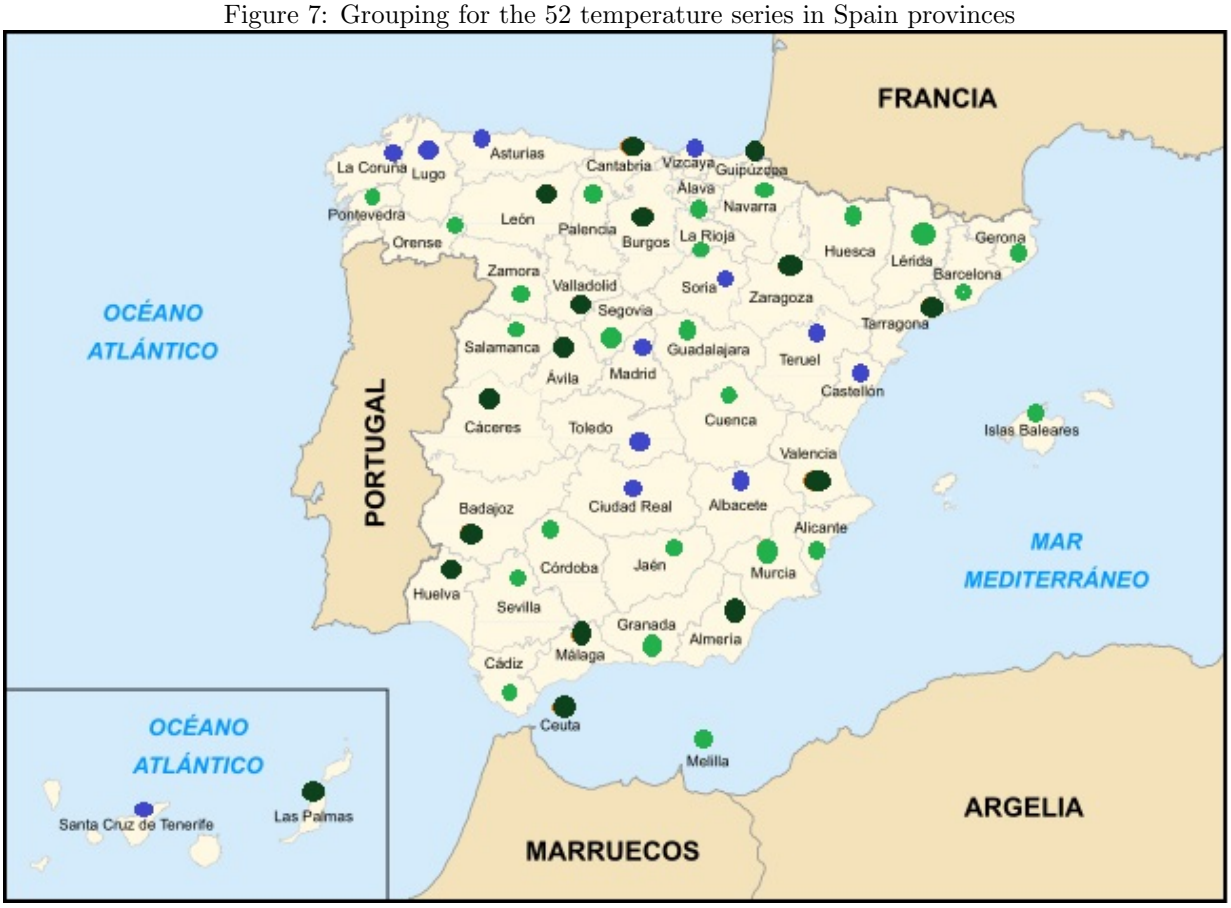
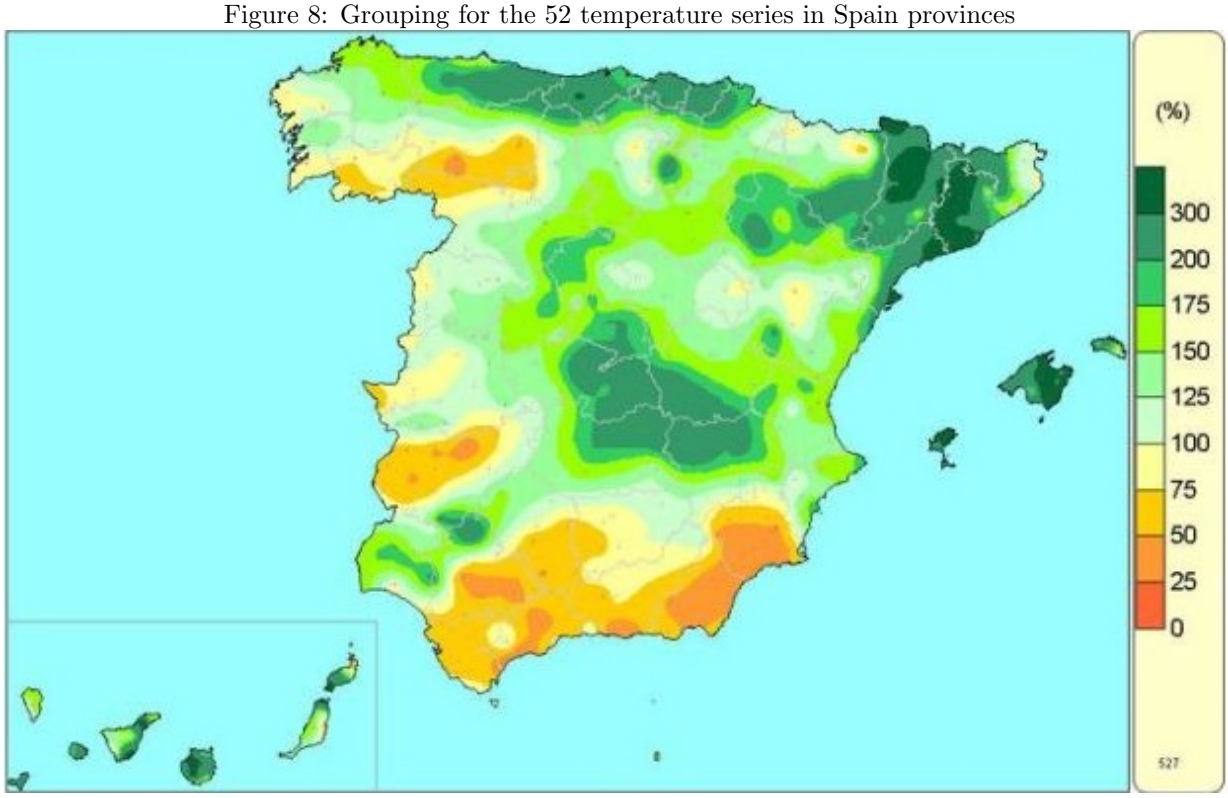


Figure 8: Grouping for the 52 temperature series in Spain provinces



The previous map shows clearly a division, three clusters locate the main bio-geographical regions for Spain. The

Euro-Siberian region is typical of the north and peninsular northern part. It has soft temperatures and humid summers. The mediterranean region corresponds to the so called *brownish Spain*, it occupies about 80% of the peninsula and the Balearic islands, characterized for dry and warm summers. Lastly, the Macaronesic region, to which the canary islands belong. It is exposed to confluent influences: On one side fresh humid air masses provided by the trade winds, and on the other side, occasionally the warm and dry saharian winds.

6 Conclusions

In conclusion, the simulation performed shows a fairly good performance, showing reasonable values of power and size, by means of the Daubichies family for the randomization test, by means of any of the proposed estimators. The test can be applied successfully to compare patterns of pairs of time series, which is a new contribution to series classification theories.

The advantages of the use of DWT are highly recognized in the current literature, its potential is correlated with the compression of long series, maintaining the relevant information of the series, as well as the application to noise filtering and the detection of singularities.

The detection of outliers, in this case the identification of the outliers series, through the use of wavelet functions and the application of the randomization test, allows the detection of series that present sudden changes, which contributes to the proper handling of these and prevents an estimation bias from occurring.

References

[1] Maharaj, E.A. (2002). Pattern Recognition of Time Series using Wavelets. "Econometrics and Business Statistics, Monash University, Australia". 1-5.

[2] Basawa, I.V., Billard, L. & Srinivasan, R. (1984). Large-sample tests of homogeneity for time series models. Biometrika 71, 203-206.

[3] Diggle, P.J. and Fisher, N.I. (1991). Nonparametric comparison of cumulative periodograms. Appl. Statist. 40, 423- 434.

[4] Guo, J. H. A non-parametric test for the parallelism of two first-order autoregressive processes. Aust. N. Z. J. Stat. 41, 59-65.(1999).

[5] Timmer, J., Lauk, M. Vach, W. & Lucking, C.H. A test for the difference between spectral peak frequencies. Comput. Statist. Data Anal. 30, 45-55.(1999)

[6] Maharaj, E.A. Clustering of time Series. J. Classification 17, 297-314.(2000)

[7] Cheong, C. W., LEE, W. W. & Yahaya, N. A. Wavelet-based temporal cluster analysis on stock time series. In Proceedings of the International Conference on Quantitative Sciences and Its Applications (ICOQSIA). (2005).

[8] Liabotis,I.,Theodoulidis,B.& Sareee,M.Improving similarity searchintime series using wavelets. Int. J. Data Warehou. Min. 2, 2, 55–81.(2006)

[9] Percival D. Wavelet methods for time series analysis. Cambridge University Press, Cambridge. Walden AT (2000)

[10] J. Walter. "A primer on wavelets and their scientific applications". University of Winconsin-Eau Claire, Hall/CRC. (1999).

[11] Badii, M.H., A. Guillen & L.A. Araiza. Estimaciones estadísticas: Un acercamiento analítico. Daena. 5(1): 237-255.(2010).

[12] Ortiz,M. "Inferencia Estadística Robusta", Universidad Veracruzana, México.

[13] Grané, A. & Veiga, H. (2009). Wavelet-based detection of outliers infinancial time series. Computational Statistics and Data Analysis. 2580–2593.

[14] Radomir, S.S & Bogdan J.F. (2000). The Haar wavelet transform: its status and achievements. School of Electrical and Electronic Engineering, Nanyang Technological University, Block S1, Nanyang Avenue. 1-20.

[15] Daubechies,Ten lectures on wavelets Society for Industrial and Applied Mathematics Philadelphia.(1992).

[16] Vonesch, C., Blu, T & Unser, M. Generalizer Daubechies Wavelet Families. IEEE Transactions on Signal Processing. Volume:55. (2007). 4415-4429.

[17] P. Faúndez. "Procesamiento digital de señales acústicas utilizando wavelets". Memoria de titulación de Ingeniería Acústica. Universidad Austral de Chile. Valdivia, Chile. (1999).

[18] Zhao, X., Milan, Z., Taylor, C.C & Barber, S. Classification tree methods for panel data using wavelet-transformed time series. Computational Statistics and Data Analysis.(2018).

[19] Alonso, A.M., Gouveia, S., Scotto,M.G & Monteiro, A. Wavelet-Based Clustering of air quality monitoring sites. Sea Level Records. Computational Statistics and Data Analysis, 1-14.(2015)

[20] Alonso, A.M., Gouveia, S., Scotto,M.G & Barbosa S.M. Wavelet-Based Clustering of Sea Level Records. Computational Statistics and Data Analysis, 51, 762–776.(2016).

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

[22] Mallat,S.,(1989). Multiresolution approximationsand wavelet orhtonormal basesof $L^2(R)$. Transactions of the American Mathematical Society. 315,69–87.

[23] Li, X., Dong, S.& Yuan, Z. Discrete wavelet transform for tool breakage monitoring. Int. J. Mach. Tools Manufact. 39, 1935–1944.(1999).

[24] Jensen, A. & Cour-Harbo, A. L. Ripples in Mathematics: The Discrete Wavelet Transforms. Springer.(2001).

[25] Chaovalit, P., Gangopadhyay, A., Karabatis, G., and Chen, Z. (2011). Discrete wavelet transform-based time series analysis and mining. ACM Comput. Surv. 43, 2, Article 6 (January 2011), 37 pages.

[26] Subasu, A. (2005). Epileptic seizure detection using dynamic wavelet network. Exp. Syst. Appl. 29, 343–355.

[27] Elizabeth A. Maharaj, Andrés M. Alonso & Pierpaolo D'Urso (2015). Clustering seasonal time series using extreme value analysis: An application to Spanish temperature time series, Communications in Statistics: Case Studies, Data Analysis and Applications, 1:4, 175-191

[28] Alonso, A.M. & Maharaj, E.A. Comparison of time series using subsampling. Computational Statistics and Data Analysis. 2589–259.(2005).

[29] Engle,R.,(1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation.Econometrica 50,987–1008.

[30] Bollerslev, T., A conditionally heteroskedastic time series model for speculative prices and rates of return. Review of Economic and Statistics 69, 542–547.(1987).

[31] Aslan, S. Iyigun, C. & Yozgatligil, C. Temporal clustering of time series via threshold autoregressive models: application to commodity prices. In: Annals of Operations Research. (2018).

[32] Arora1, P., Deepali2, D.& Varshney, S. Analysis of K-Means and K-Medoids Algorithm For Big Data. International Conference on Information Security & Privacy (ICISP2015), 11-12. Nagpur, INDIA. (2015).

[33] Alonso, A. M., P. De Zea Bermudez, and M. G. Scotto. "Comparing Generalized Pareto Models Ftted to Extreme Observations: An Application to the Largest Temperatures in Spain. "Stochastic Environmental Researchand Risk Assessment.(2014). 1221–1233.

[34] Fernández-Montes,S.% Rodrigo,F.S. "Trends in Seasonal Indices of Daily Temperature Extremes in the Iberian Peninsula. "International Journal of Climatology. 32:2320–2332.(2005).

[35] Furió, D., and V. Meneu. 2011. "Analysis of Extreme Temperatures for Four Sites Across Peninsular Spain." Theoretical andAppliedClimatology104:83–99.

[36] Brunet, M., P. D. Jones, J. Sigró, O. Saladié, E. Aguilar, A. Moberg, P. M. Della-Marta, D. Lister, A. Walther, and D. López. "Temporal and Spatial Temperature Variability and Change Over Spain. (2007).

[37] García Herrera, R.,J.Díaz, R.M.Trigo, & E.Hernández. "Extreme Summer Temperaturesin Iberia:Health Impacts and Associated Synoptic Conditions."Annals ofGeophysics 23:239–251.(2005).

[38] Castro, M., Vide, J.M, & Alonso, S. El clima de España :Pasado, presente y escenarios del clima para el siglo XXI. Ministerio de medio ambiente. (2005).

[39] Alonso, A.M. & Maharaj, E.A. (2005).Comparison of time series using subsampling. Computational Statistics and Data Analysis. 2589–259.