Statistical analysis and evaluation of Hurst coefficient for annual and monthly precipitation time series

ALINA BĂRBULESCU
Faculty of Mathematics and Computers Science
Ovidius University of Constantza
124, Mamaia Boulevard, Constantza
ROMANIA
alinadumitriu@yahoo.com http://www.alina.ilinc.ro

CRISTINA SERBAN (GHERGHINA)
Faculty of Civil Engineereing
Ovidius University of Constantza
124, Mamaia Boulevard, Constantza
ROMANIA
cgherghina@gmail.com

CARMEN MAFTEI
Faculty of Civil Engineereing
Ovidius University of Constantza
124, Mamaia Boulevard, Constantza
ROMANIA
cmaftei@univ-ovidius.com

Abstract: - A major issue in time series analysis and particularly in the study of meteorological time series is to model and to predict their behaviour. The first step to reach this goal is the data analysis. It is what we shall do in this article, focusing on the long range dependence (LRD) property. Various estimators of LRD have been proposed. Their accuracy have been generally tested using simulated time series since sometimes only their asymptotic property are known, or worse, no asymptotic property have been proved. For ten annual and monthly data series collected in Dobrudja region, for 41 years, we did the statistical analysis and we compare the results.

Key-Words: - Normality, homoscedasticity, long – range dependence, Hurst coefficient.

1 Introduction

Weather modification is a topic of substantial worldwide interest for all countries. Modeling the precipitation and temperature evolution have been done for different region of the world [1] – [3]. Building models and testing their validity is a step in understanding and predicting the weather evolution, which is a challenge for all scientists.

Only a small number of studies is devoted to weather evolution in different regions of Romania, including the Black Sea coast [4–7]. Therefore, this article comes to complete the analyses made for a region of Romania, called Dobrudja.

Dobrudja is situated in the South – East of Romania, between the Black Sea and the lower Danube River. Dobrudja's structure (excluding the Danube Delta) is that of a plateau with hilly aspect, with an average altitude between 100 and 180 m. Its climate is temperate - continental.

It has been shown showed that the frequency of droughty years is 89 %, the longest rainless period in this area being registered in the South of Dobrudja and the Black Sea coast [8].

2 Methodology

In order to study the time series, the following steps were done:

- The normality study;
- The correlation study;
- The homoscedasticity study
- The analysis of break presence;

- The analysis of long-range dependence property.

2.1. Normality tests

The null hypothesis is:

H_0: The process is normally distributed.

The *Kolmogorov – Smirnov test* [9] is known, so we don't present it.

The Shapiro Wilk test [10]:

The statistic test is defined as:

$$W = \frac{\left(\sum_{i=1}^{n} w_i x_i'\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where:

- n is the sample volume,
- $x_1, x_2, \dots, x_n$ are the original data,
- $x_1', x_2', \dots, x_n'$ are the ordered data,
- $\bar{x}$ is the sample mean of the data,
- $w' = (w_1, w_2, \dots, w_n)$ or

$$w' = MV^{-1}\left[(M \quad ^{-1})(V^{-1}M)\right]^{-1/2},$$

- M denotes the expected values of standard normal order statistics for a sample of size n,
- V is the corresponding covariance matrix.

Small values of W indicate non – normality.

The Q-Q plot [9].

For a Q - Q plot of a sample versus a theoretical model, one can estimate *goodness of fit* and *parameters* of the model.

The data are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line. Departures from this straight line indicate departures from normality.

2.2. Study of correlation

The autocorrelation [11] of a random process describes the correlation between values of the process at different points in time, as a function of the two times or of the time difference.

Let (X_t) be a time series. The autocovariance function of (X_t) at lag h ($h \in \mathrm{N}^*$) is defined by:

$$\gamma(h) = Cov(X_t, X_{t+h})$$

and the autocorrelation function of (X_t) at lag h, as:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)},\ h \in \mathbf{N}^*.$$

If $x_1, \dots, x_n$ are observed data, from the previous formula we obtain the empirical autocorrelation function, denoted by ACF.

Together with ACF we determine the confidence limits at the confidence level of 95%. If for the selected sample, the values of ACF are inside the corresponding confidence interval, we conclude the data are not correlated.

2.3. Break points detection

A break in a time series is a change of probability low at a certain moment [12].

Some break tests permit to detect a change in a time series mean.

The methods used to detect break points were: the Pettitt test [13] and the segmentation procedure of Hubert [14, 15], since they work even if the series are not normally distributed.

In the situations of contradictory results, Buishard test and change point analysis have also been performed.

The null hypothesis is:

H_0: There is no change in the time series.

The Buishard test works in the hypothesis that the series is normal, the break absence representing the null hypothesis, H_0.

The Pettitt test is a nonparametric test, which can be used even if the time series distribution is unknown.

The Hubert segmentation procedure detects the multiple breaks in time series and the moments of their apparition. The principle is to cut the series in m segments (m>1) such that the calculated means of the neighbours sub-series significantly differ.

CUSUM procedure [16]

CUSUM charts are constructed by calculating and plotting a cumulative sum based on the data. CUSUM charts show the cumulative sum of differences between the values and the average and are used to determine changes in average.

2.4. Tests of homoscedasticity

The Bartlett test is used to test the hypothesis that k groups have equal variances. Since this test is sensitive to departures from normality, the *Levene* test [17] is an alternative to the Bartlett test, which is less sensitive to departures from normality.

In this case, the null hypothesis is:

$$H_0:\ \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2,$$

and its alternative:

$$H_1:\ \sigma_i^2 \neq \sigma_j^2 \text{, for at least one pair } (i, j).$$

The test statistic is:

$$W = \frac{n-k}{k-1} \cdot \frac{\sum_{i=1}^{k} n_i (\overline{Z_{i.}} - \overline{Z_{..}})^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (Z_{ij} - \overline{Z_{i.}})^2}$$

where:

- n is the sample volume,
- k is the number of groups in which the sample is divided,
- n_i is the sample size of each group,
- X_{ij} is the element j in the i - th group,
- $\overline{X_{i.}}$ is the mean of the i-th group,
- $Z_{ij} = \left| X_{ij} - \overline{X_{i.}} \right|$,
- $\overline{Z_{i.}}$ is the mean group of Z_{ij},
- $\overline{Z_{..}}$ is the overall mean of Z_{ij}.

The Levene test rejects the hypothesis that the variances are equal at the significance level α if $W > F_{\alpha,k-1,n-k}$, where $F_{\alpha,k-1,n-k}$ is the upper critical value of the F distribution with k - 1 and $n - k$ degrees of freedom at the significance level of α.

2.5. Measuring Hurst exponent

The study of time series with long-range dependence have been extensively developed for applications in nature sciences, as well as in DNA sequences, cardiac dynamics, internet traffic [18] and finance [19].

The Hurst [20] exponent provides a measure for long term memory and fractality [21] of a time series. For details on the topics, see the volumes of Beran [22], Embrechts and Maejima [23], and Palma [24] and the collections of Doukhan et al. [25] and Robinson [26].

The values of the Hurst exponent range between 0 and 1. Based on the Hurst exponent value H, the following classifications of time series can be realized:

- $H = 0.5$ indicates a random series;
- $0 < H < 0.5$ indicates an anti - persistent series, which means an up value is more likely followed by a down value, and vice versa;
- $0.5 < H < 1$ indicates a persistent series, which means the direction of the next value is more likely the same as current value.

Let $(X_t)_{t \in \mathrm{N}}$ be a time series, shortly denoted by (X_t). We say that (X_t) it is weakly stationary if it has a finite mean and the covariance depends only on the lag between two points in the series.

A time series (X_t) has the long range dependence property if it has correlations that persist over all time scales.

An equivalent definition is:

The time series (X_t) has the long range dependence (LRD) property if $\sum_{h=-\infty}^{\infty} \rho(h)$ diverges.

LRD can be thought of in two ways [27]:

- In the time domain it manifests as a high degree of correlation between distantly separated data points.
- In the frequency domain it manifests as a significant level of power at frequencies near zero.

The following methods are used for LRD analysis:

- R/S and Lo's modified R/S statistic;
- Aggregated Variance;
- Absolute Moments;
- Detrended fluctuation analysis (variance of residuals);
- Ratio of variance of residuals;
- Periodogram;
- Whittle's approximate MLE and Local Whittle estimators;
- Wavelets.

2.5.1. R/S analysis

The Hurst exponent can be calculated by rescaled range analysis (R/S analysis) [20].

To study the LRD in a time series, the following algorithm is used.

A time series $(X_k)_{k \in \overline{1,N}}$ is divided into d sub-series of length m. For each sub-series $n = 1, \dots, d$:

- Find the mean, E_n and the standard deviation, S_n;
- Normalize the data (X_{in}) by subtracting the sub-series mean:

$$Z_{in} = X_{in} - E_n,\ i = 1, ..., m;$$

- Create a cumulative time series:

$$Y_{in} = \sum_{j=1}^{i} Z_{jn}, \; i = 1,\dots,m;$$

- Find the range

$$R_n = \max_{j=1,m} Y_{jn} - \min_{j=1,m} Y_{jn};$$

- Rescale the range R_n / S_n;
- Calculate the mean value of the rescaled range for all sub-series of length *m*:

$$(R/S)_m = \frac{1}{d}\sum_{n=1}^{d} R_n / S_n .$$

Hurst found that (R/S) scales by power - law as time increases, which indicates

$$(R/S)_t = c \cdot t^H .$$

In practice, in classical *R/S* analysis, *H* can be estimated as the slope of log-log plot of $(R/S)_t$ versus *t.*

Although Mandelbrot [28] gave a formal justification for the use of this test, Lo [29] showed that this statistic was not robust to short memory dependence and modified this statistic.

Lo defined *modified R/S* statistic by:

- instead of considering multiple lags, only focus on lag *N*, the length of the series:
  After finding the overall mean

$$\overline{X} = \sum_{j=1}^{N} X_j ,$$

create the cumulative time series:

$$Y_k = \sum_{j=1}^{N} (X_k - \overline{X}), \; k = 1,\dots,N;$$

and find the range

$$R(N) = \max_{k=1,N} Y_k - \min_{j=1,N} Y_k ;$$

- instead of using the standard deviation to normalize $R(N)$, he uses the following sum:

$$S_q(N) = \sqrt{\sum_{j=1}^{N}(X_j - \overline{X})^2 + 2\sum_{j=1}^{q}\omega_j(q)\left[\sum_{i=j+1}^{N}(X_j - \overline{X})(X_{i-j} - \overline{X})\right]}$$

where

$$\omega_i(q) = 1 - \frac{j}{q+1}, q < N.$$

The Lo's modified *R/S* statistic is defined by:

$$V_q(N) = \frac{1}{\sqrt{N}} R(N)/S_q(N).$$

Lo uses the interval [0.809, 1.862] as the 95% asymptotic acceptance region for testing the null hypothesis:

H_0: the absence of LRD,

against the alternative:

H_1: the presence of LRD.

As discussed in [30], the right choice of *q* in Lo's method is essential. For study, the following values are used:

$$q = \left[\left(\frac{3N}{2}\right)^{\frac{1}{3}}\left(\frac{2\hat{\rho}}{1-\hat{\rho}^2}\right)^{\frac{2}{3}}\right], [31] \qquad (*)$$

$$q = \left[\left(\frac{N}{10}\right)^{\frac{1}{4}}\left(\frac{2\hat{\rho}}{1-\hat{\rho}^2}\right)^{\frac{2}{3}}\right], [32] \qquad (**)$$

where:

- *N* is the length of the series,

- $\hat{\rho}$ is the estimated first order correlation coefficient,

- [ ] is the greatest integer function.

2.5.2. Aggregated variance method [30]

A series of length *N* is divided into *d* sub-series of length *m*. For each subseries, the aggregated series, formed by the means

$$X^m(k) = \frac{1}{m}\sum_{i=(k-1)m+1}^{km} X_i, k = 1,2,\dots,d$$

is calculated, as well as, its sample variance:

$$VarX^{(m)} = \frac{1}{d}\sum_{k=1}^{d}(X^{(m)}(k) - \overline{X})^2 .$$

For successive values of *m*, the sample variance is plotted against *m* on a log-log plot. Fitting a least squares line to the points of the plot, the Hurst coefficient is calculated, knowing that the straight line slope is 2*H* - 2.

In order to distinguish between the nonstationarity and LRD, Teverovski and Taqqu [27] proposed to use this method, together with the study of successive differences of variances or fitting a function $C_1 + C_2 m^{2H-2}$ to the $VarX^{(m)}$.

2.5.3. Absolute Moments Method [30]

This method is analogous to 2.5.2., but instead of

$VarX^{(m)}$, the n^{th} absolute moments are calculated for the aggregated series,

$$AM_n^{(m)} = \frac{1}{d}\sum_{k=1}^{d}\left|X_k^{(m)} - \overline{X}\right|^n .$$

For successive values of *m*, the sample absolute moment is plotted versus *m* on a log-log plot. Fitting a least squares line to the points of the plot, the Hurst coefficient is calculated, knowing that the straight line slope is $n(H\text{-}1)$.

2.5.4. Detrended fluctuation analysis (DFA)

Detrended fluctuation analysis was originally proposed as a technique for quantifying the nature of long-range correlations by Peng *et al.* [33]. It was introduced in order to permit the detection and quantification of long-range correlations in DNA sequences.

In recent years the DFA method has been applied in analysis of different time series that appear in different fields as DNA sequences [34], heart rate study [35], human gait, meteorology, economics, and physics.

DFA method, involves the following steps [36]:

- Starting with a time series, (X_t) with the length *N*, it is integrated, obtaining:

$$Y_k = \sum_{i=1}^{k}(X_k - \overline{X}) .$$

- The integrated series in divided into *d* sub-series of equal length *m*.
- In each sub-series, fit X_t, using a polynomial function of order *l* which represents the *trend* of that sub-series. The ordinate of the fit line in each box is denoted by $y_m(k)$.
- The integrated series is detrended by subtracting the local trend $y_m(k)$ in each sub-series of length *m*.
- For a given sub-series length, *m*, the root mean-square fluctuation for the integrated and detrended series is calculated:

$$F(m) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[Y_k - y_m(k)\right]^2} .$$

- The above computation is repeated for a broad range of scales (*m*) to provide a relationship between $F(m)$ and the box size *m*.

A power-law relation between the average root-mean-square fluctuation function $F(m)$ and the box size *m* indicates the presence of scaling: $F(m) \sim m^{2H}$.

2.5.5. Ratio of Variance of Residuals

This method estimates the parameter alpha characterizing the intensity of heavy tails, instead of estimating the long-range dependence parameter *H*. The method is based on the Variance of Residuals method. It calculates the Variance of Residuals in two ways, and takes their ratio to obtain a statistic, and then fits a least-squares line to the logarithm of that statistic. This enables one to estimate alpha.[30]

2.5.6. Periodogram

Geweke and Porter-Hudak [37] proposed a semi-parametric approach to test for long-memory memory of a fractionally integrated process. The fractional difference parameter *d* can be estimated by regression equations.

Let $(X_k)_{k\in\overline{1,N}}$ be a time series, and

$$I(\lambda) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N} X_j e^{ij\lambda}\right|^2 ,$$

where λ is a frequency.

It was shown that a series with LRD should have a periodogram proportional to $|\lambda|^{1-2H}$ in the origin neighbourhood, so a log-log plot of a periodogram against the frequency should give the coefficient $1-2H = d$.

Modified *periodogram*, as well *cumulative periodogram* have also been used [30].

3. Results and discussions

The studied data represent mean annual (Fig.1) and mean monthly precipitation collected for a period of 41 years at 10 meteorological stations, starting to 1965. The coordinates of these stations and the mean annual precipitation registered are given in Table I and represented in Fig.1.

Table I. The coordinates of meteorological stations

Station	Latitude	Longitude	Elevation (m)
Tulcea	+45:11	+28:49	4.36
Jurilovca	+44:46	+28:53	37.65
Corugea	+44:44	+28:20	219.20
Harsova	+44:41	+27:57	37.51
Cernavoda	+44:21	+28:03	87.17
Medgidia	+44:15	+28:16	69.54
Constanta	+44:13	+28:38	12.80
Adamclisi	+44:08	+28:00	158.00
Mangalia	+43:49	+28:35	6.00
Sulina	+45:09	+29:39	2.08

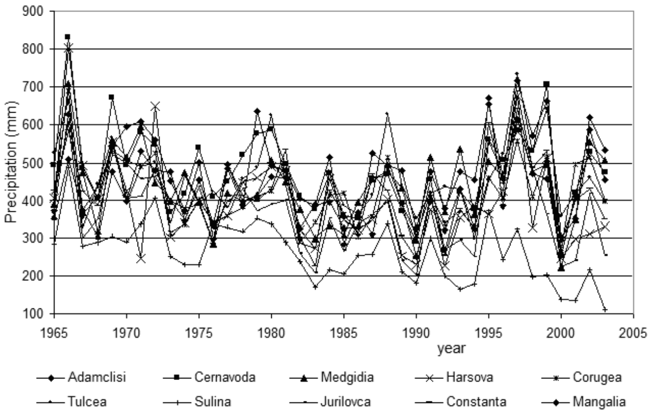

Fig.1. The mean annual precipitation of the studied stations in the period 1965 – 2003

3.1. Statistical analyses

3.1.1. The results of normality tests

The results of normality tests for annual series are presented in Table II, where:

- "Statistic" is the value of statistic associated respectively to Kolmogorov – Smirnov and Shapiro – Wilk tests,
- df is the degree of freedom,
- Sig is the significance level. A value of Sig less than 0.05 indicates a deviation from normality.

Table II. Results of normality tests for annual series

Station	Kolmogorov – Smirnov[a]			Shapiro - Wilk		
	statistic	df	Sig.	statistic	df	Sig.
Adamclisi	0.106	39	0.200*	0.963	39	0.220
Cernavoda	0.087	39	0.200*	0.966	39	0.279
Medgidia	0.068	39	0.200*	0.989	39	0.964
Harsova	0.128	39	0.037	0.904	39	0.003
Corugea	0.124	39	0.010	0.954	39	0.108
Tulcea	0.116	39	0.164	0.955	39	0.118
Sulina	0.073	39	0.200*	0.980	39	0.707
Jurilovca	0.085	39	0.200*	0.964	39	0.245
Constanta	0.124	39	0.033	0.974	39	0.502
Mangalia	0.078	39	0.200*	0.977	39	0.606

*This is a lower bound of the true significance
[a] Lilliefors significance correction

Since the Kolmogorov – Smirnov and Shapiro – Wilk tests gave contradictory results for Corugea and Constanta annual series, Jarque – Bera test has also been performed, as well as the Q - Q plot and the histogram analysis [13].

For Corugea series, the value of Jarque – Bera statistic was $JB = 3.825$ and the corresponding significance level of 0.1477 > 0.05. Thus we admit that this series is normally distributed. We also accept the null hypothesis analysing the histogram of Constanța series (Fig.2).

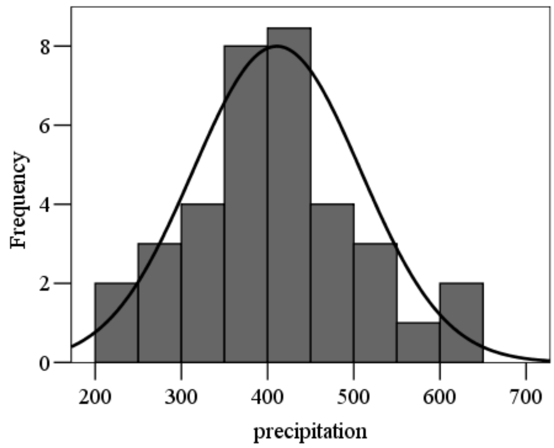

Fig.2. Histogram of Constanta series

So, between the annual series only Harsova is not normally distributed.

Performing normality tests for monthly series, we conclude that no series is normally distributed, but in nine of ten cases the normality can be reached by Box-Cox transformation, with the parameters from Table III.

Table III. Parameters of Box-Cox transformation

Series	Adamclisi	Cernavoda	Medgidia	Harsova	Corugea	Tulcea	Sulina	Jurilovca	Constanta	Mangalia
Coef.	0.45	0.48	0.39	0.36	0.42	0.47	0.34	0.36	-	0.39

3.1.2. Autocorrelation results

Analysing *the autocorrelation*, we remark that Adamclisi, Cernavoda, Medgidia, Harsova, Corugea, Tulcea annual series are not correlated.

Table IV. ACF and Box-Ljung statistic for Corugea series

Lag	ACF	Std. error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	0.006	0.154	0.001	1	0.971
2	0.056	0.152	0.137	2	0.934
3	0.115	0.150	0.723	3	0.868
4	0.166	0.148	1.988	4	0.738
5	-0.088	0.146	2.353	5	0.798
6	0.086	0.144	2.708	6	0.845
7	-0.092	0.141	3.130	7	0.873
8	-0.035	0.139	3.194	8	0.922
9	0.056	0.137	3.359	9	0.948
10	0.005	0.135	3.360	10	0.972
11	-0.085	0.132	3.777	11	0.976
12	-0.124	0.130	4.688	12	0.968
13	-0.022	0.128	4.718	13	0.981
14	0.004	0.125	4.719	14	0.989
15	-0.080	0.123	5.148	15	0.991
16	-0.054	0.120	5.351	16	0.994

[a] The underlying process assumed is independence (white noise).
[b] Based on the asymptotic chi-square approximation.

For example, for Corugea (Table IV), the value of Ljung – Box (5.351, for the lag 16), together with the corresponding Sig. value comes o confirm the hypothesis that the series is independent.

The values of autocorrelation functions (ACF) of Sulina and Jurilovca annual series are inside confidence interval, excepting the first one, and Mangalia presents a pick at the lag of 12 (that could be interpreted as a periodicity sign).

The same analysis for all monthly data, but Mangalia (Fig.3), reveals their periodicity.

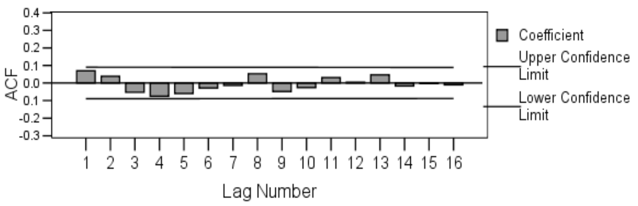

Fig.3. ACF of Mangalia monthly series

3.1.3. The results of break points tests

The results of *Pettitt test and Hubert segmentation procedure* for annual data were:

- seven series don't present break points;
- there is a break in Sulina series (1981);
- contradictory for Corugea and Harsova series.

Buishard test and CUSUM chart for Corugea series indicate a change point in 1972, in concordance to Hubert procedure. After CUSUM analysis for Harsova annual series (Fig.4), we accept the null hypothesis.

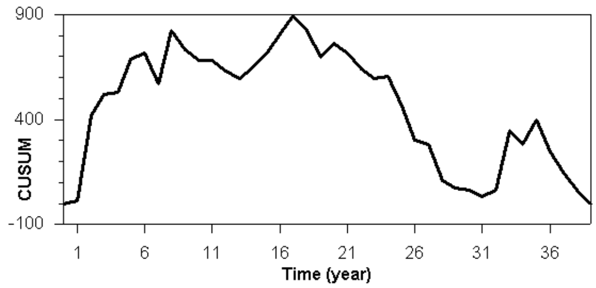

Fig.4. CUSUM of Harsova annual series

The results of break tests for monthly series are also contradictory for eight of ten series. For all the series, Buishard and Pettitt tests gave the same results, but Lee&Heghinian test and Hubert segmentation procedure lead to the conclusion that there are breaks in the time series (Table V).

These contradictions could be the result of the big number of outliers (Fig.5). As consequence, for modeling purposes the outliers could be removed, the tests reiterated and the models built on sub - periods, if necessary.

Table V. Break tests for monthly data

Station	Buishard	Pettitt	Lee & Heghinian	Hubert
Adamclisi	no	no	no – Apr. 2004	no – Nov. 1969, Apr. 1971
Cernavoda	yes	yes	no – Apr. 2004	no – Dec. 1965, Apr. 1971, July 1999, Apr. 2004
Medgidia	yes	yes	no – Apr. 2004	no – Apr. 1971, July 1991
Harsova	yes	yes	no – Apr. 2004	no – Dec. 1965, July 1972, June 1997, Apr.n2004
Corugea	yes	yes	no – Apr. 2004	no –August 1972, April 2004, August 2005
Tulcea	yes	yes	no – Apr. 2004	no – May 1997
Sulina	no	no	no – Aug. 1982	no – July 1972
Jurilovca	yes	yes	no – Jan. 1965	no – Nov.1969
Constanta	-	yes	-	no – July 2004
Mangalia	yes	yes	no – Aug. 2005	no –Aug.2005

yes means that the null hypothesis is accepted
- means that the test can not be applied

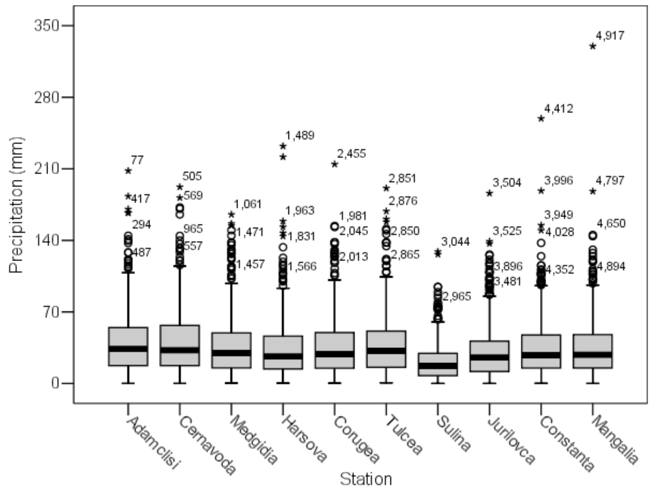

Fig.5. Box-plot of monthly series

3.1.4. The results of homoscedasticity test

The results of Levene test for annual and monthly data are summarized in Table VI.

Table VI. Results of Levene test

Station	Annual		Monthly	
	F	Sig	F	Sig
Adamclisi	2.494	0.123	109.950	0.000
Cernavoda	0.581	0.310	88.436	0.000
Medgidia	0.014	0.906	146.607	0.000
Harsova	5.777	0.021	105.202	0.000
Corugea	3.265	0.079	100.330	0.000
Tulcea	4.892	0.033	124.588	0.000
Sulina	0.575	0.453	147.933	0.000
Jurilovca	0.911	0.346	138.394	0.000
Constanta	0.466	0.499	111.390	0.000
Mangalia	0.392	0.532	88.339	0.000

Some remarks are necessary:

- For Corugea annual series, for which the results of break tests were contradictory, the homoscedasticity hypothesis was accepted based on the Levene test, which was applied twice:

a. Dividing the initial series in two sub-series, before and after the 1972;

b. Dividing the series in three sub-series, with the same number of elements (13).

The value of W statistics associated to it was respectively:

$$W = 1.15 < F_{0.05,1,37} = 4.105,$$

$$W = 2.12 < F_{0.05,2,36} = 3.259.$$

- Taking the square root of Tulcea annual data, the new series is homoscedastic.
- For all monthly series the homoscedasticity hypothesis was rejected.

3.1.5. Long –range dependence analysis

For R/S analysis on annual data, a part of results is given in [7] and those obtained using KaotiXL [38], in Table VII.

Table VII. Hurst coefficients for annual series, calculated by KaotiXL (rescaled method)

Station	Adamclisi	Cernavoda	Constanta	Corugea	Harsova
H	0.7207	0.9775	0.9738	0.9224	0.8028
Station	Jurilovca	Medgidia	Mangalia	Sulina	Tulcea
H	0.9301	0.8727	0.7049	0.9567	0.7057

The R/S charts and the evolution of V - statistics associated are exemplified in Figs. 6 – 11, where R^2 is the determination coefficient.

For Lo's modified statistic, the values used for q were calculated by (*) and (**). As result of formula (*), q was respectively: 3 – for Sulina and Jurilovca series and 2 – for the rest. Using formula (**), the coefficients were not very different.

The results are presented in Table VIII and those obtained by periodogram method, in Table IX.

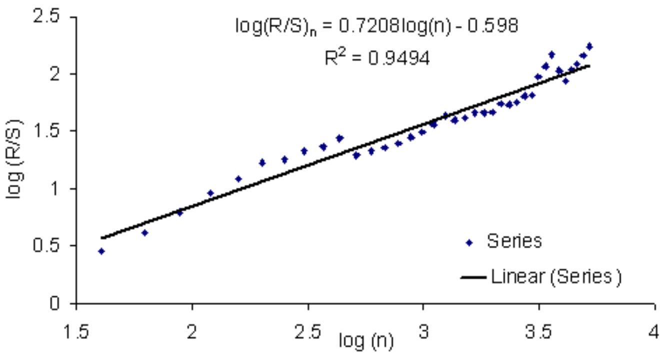

Fig.6. The R/S chart of Adamclisi annual series

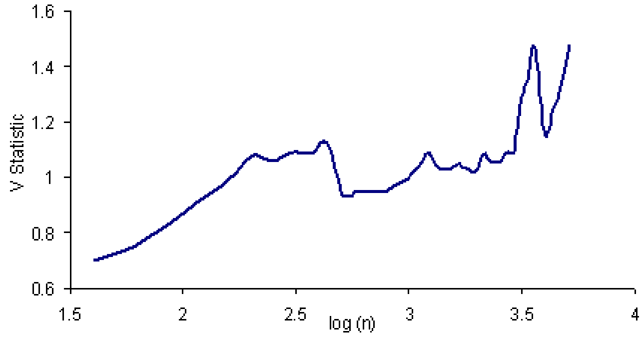

Fig. 7. The V - statistic associated to Adamclisi annual series

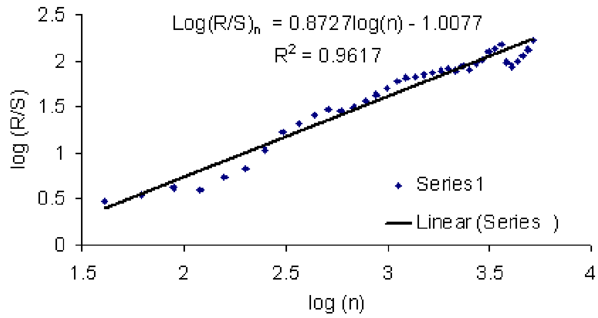

Fig.8. The R/S chart of Medgidia series

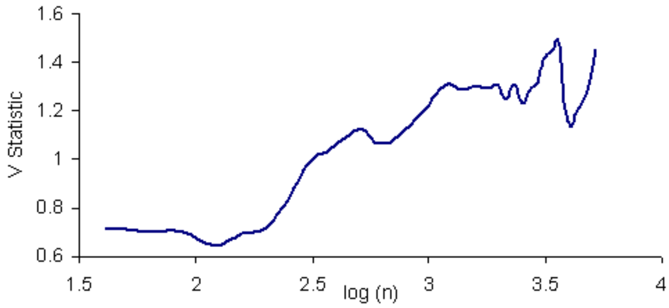

Fig.9. The V- statistic associated to Medgidia series

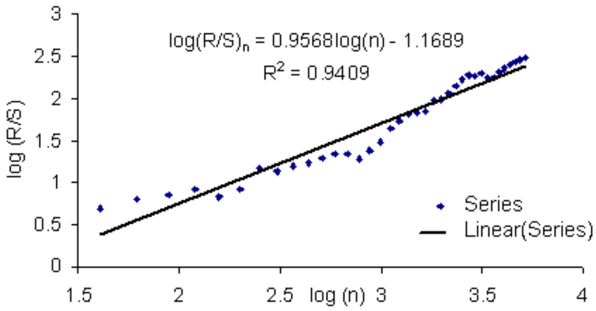

Fig.10. The R/S chart of Sulina annual series

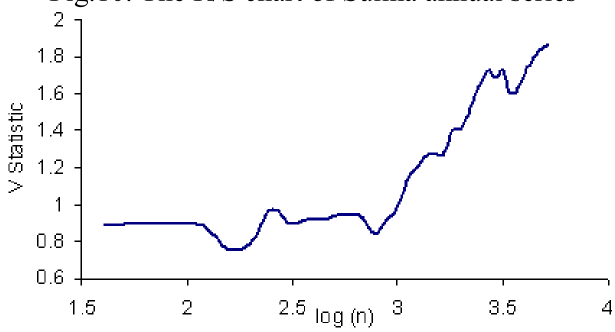

Fig. 11. The V - statistic associated to Sulina annual series

Table VIII. Hurst coefficients for annual series, calculated by Lo's method

Station	Adamclisi	Cernavoda	Constanta	Corugea	Harsova
H	0.0988	0.1177	0.11	0.1262	0.1167
Station	Jurilovca	Medgidia	Mangalia	Sulina	Tulcea
H	0.1104	0.1304	0.1165	0.119	0.1121

Table IX. Hurst coefficients for annual series, calculated by periodogram method

Station	Adamclisi	Cernavoda	Constanta	Corugea	Harsova
H	0.44505	0.4491	0.36745	0.2826	0.55745
Station	Jurilovca	Medgidia	Mangalia	Sulina	Tulcea
H	0.46905	0.3967	0.4559	0.32415	0.58925

It can be remarked that the results for each series are very different and a correct conclusion on the long range dependence property of these time series could not be extract applying only one method.

Discussing, for example, about Adamclisi, Cernavodă or Medgidia series, after the application of normality, homoscedasticity and correlation tests, it was concluded that they are Gaussian white noises [7]. But not all the values from the previous tables are in concordance with these conclusions. In [7] a FARIMA model was found for Sulina series. But the values of H, in Tables 2 and 3 are in discordance with the LRD property of Sulina series.

The reasons for which these results are so different could be: q is too small in the case of Lo's statistic (and it does not account for the autocorrelation of the process) and the small number of data.

The results of DFA are closer to those of Lo's method (for example, for Adamclisi series), other being closer to those of periodogram method.

The results are different also in the case of monthly series, as we exemplify in Tables IX and X.

Table IX. Hurst coefficients for monthly series, calculated by R/S method

Station	Adamclisi	Cernavoda	Constanta	Corugea	Harsova
H	0.5847	0.6697	0.7316	0.6932	0.6675
Station	Jurilovca	Medgidia	Mangalia	Sulina	Tulcea
H	0.7303	0.6551	0.5395	0.7441	0.5723

Table X. Hurst coefficients for monthly series, calculated by Lo's method

Station	Adamclisi	Cernavoda	Constanta	Corugea	Harsova
H	0.0132	0.0125	0.0177	0.0147	0.016
Station	Jurilovca	Medgidia	Mangalia	Sulina	Tulcea
H	0.0139	0.0109	0.0219	0.0134	0.0125

4. Conclusion

In this article we presented the statistical analysis and the results of LRD analysis for some precipitation time series, collected in Dobrudja region, for 41 years. We remark that the LRD results depends on the method used and they are not always concordant with the results of statistical tests. Therefore, it is indicated to use with circumspection different LRD analysis methods, since sometimes the statistics attached to them have properties that have been proved only for some well established time series or their properties are not known.

Acknowledgements: This article was supported by CNCSIS – UEFISCSU, project number PNII – IDEI 262/2007.

References:

[1] S. L. Badjate, S. V. Dudul, *Multi Step Ahead Prediction of North and South Hemisphere Sun Spots Chaotic Time Series using Focused Time Lagged Recurrent Neural Network Model*, WSEAS Transactions on Information Science and Applications, Issue 4, Vol. 6, 2009, pp. 684 - 693

[2] M. Shaharuddin, A. Zaharim, M. Jailani, M. Nor, O. A. Karim, K. Sopian, Application of Wavelet Transform on Airborne Suspended Particulate Matter and Meteorological Temporal Variations, WSEAS Transactions on Environment and Development, Issue 2, Vol. 4, February 2008, pp.89-98

[3] K. Sreelakshmi, P.R. Kumar, *Short term wind speed prediction using support vector machine model*, WSEAS Transactions on Computers, Issue 11, Vol. 7, 2008, pp. 1828 -1837

[4] A. Bărbulescu, E. Băutu, *Meteorological Time Series Modelling Based on Gene Expression Programming*, Recent Advances in Evolutionary Computing, WSEAS Press, 2009, pp. 17 – 23

[5] A. Bărbulescu, E. Băutu, *ARIMA Models versus Gene Expression Programming In Precipitation Modeling*, Recent Advances in Evolutionary Computing, WSEAS Press, 2009, pp.112 – 117

[6] A. Bărbulescu, E. Băutu, Mathematical models of climate evolution in Dobrudja, *Theoretical and Applied Climatology*, DOI 10.1007/s00704 – 009 – 0160 – 7

[7] A. Bărbulescu, E. Pelican, *ARIMA models for the analysis of the precipitation evolution*, Recent Advances in Computers, WSEAS Press, 2009, pp.221 – 226

[8] C. Maftei and A. Barbulescu, Statistical analysis of climate evolution in Dobrudja region, *Lecture Notes in Engineering and Computer sciences*, WCE 2008, vol.II, IAENG, pp. 1082-1087

[9] D.J. Seskin, *Handbook of parametric and nonparametric statistical procedures*, Chapmann and Hall/CRC, Boca Raton, 2007, ch. 12

[10] S. S. Shapiro, M. B. Wilk, An Analysis of Variance Test for Normality (Complete Samples), *Biometrika*, Vol. 52, No. 3/4. (Dec. 1965), pp. 591-611.

[11] P. Brokwell, R.A. Davis, *Introduction to time series and forecasting*, Springer, 2002, ch. 5 and ch.10

[12] H. Lubes, J.M. Masson, E. Servat, J.-E. Paturel, B. Kouame, J.F. Boyer, Caractérisation de fluctuations dans une série chronologique par applications de tests statistiques, *Rapport* no 3, *Programme* ICCARE, 1994, 21 pp.

[13] A. N. Pettitt, A non - parametric approach to the change-point problem, *Applied Statistics* vol. 28, no.2, 1979, pp. 126 – 135

[14] P. Hubert, J. P. Carbonnel, A. Chaouche, Segmentation des séries hydrométéorologiques. Application à des séries de précipitations et de débits de l'Afrique de l'Ouest, *Journal of Hydrology*, 110, 1989, pp. 349 – 367

[15] P. Hubert and J. P. Carbonnel, Segmentation des séries annuelles de débits de grands fleuves africains, *Bulletin de liaison du CIEH* 92, 1993, pp. 3 – 10

[16] M. Basseville, I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, Englewood Cliffs, N.J, 1993, http://www.irisa.fr/sisthem/kniga/.

[17] H. Levene, Robust tests for equality of variances, *Contributions to Probability and Statistics*: Essays in Honor of Harold Hotelling, I. Olkin et al. eds., Stanford University Press, 1960, pp. 278-292.

[18] W. Willinger et all, Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements, *Statistical Science,* 10, 1995, pp.67-85

[19] A. Lo, Fat tails, long memory, and the stock market since the 1960s, Economic Notes, Banca Monte dei Paschi di Siena, 2001.

[20] H. E. Hurst, Long-term storage of reservoirs: an experimental study, *Transactions of the American society of civil engineers,* 116, 1951, pp. 770 – 799

[21] B.B. Mandelbrot, J. R. Wallis, Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence, *Water Resources*, 5, 1969, pp. 967 – 988.

[22] J. Beran, *Statistics for Long - Memory Processes*, Chapman and Hall, New York, 1994

[23] M. Embrechts, P. Maejima, *Self - similar processes*. Princeton, University Press, Princeton and Oxford, 2002

[24] W. Palma, *Long-Memory Time Series Theory and Methods*, Wiley - Interscience, 2007

[25] P. Doukhan, G. Oppenheim, and M. Taqqu, *Theory and Applications of Long-Range Dependence*, Birkhauser, 2003

[26] P. M. Robinson, *Time Series with Long Memory*, Oxford University Press, 2003.

[27] V. Teverovsky, M. S. Taqqu, W. Willinger, On Lo's modified R/S statistic, *Preprint*, 1997

[28] B. Mandelbrot, Statistical methodology for non - periodic cycles: from the covariance to R/S analysis, *Annals of Economic and Social Measurement*, vol.1, 1972

[29] A. W. Lo, Long term memory in stock market prices, *Econometrica*, vol. 59, 1991, pp.1279 - 1313

[30] M. S. Taqqu, V. Teverovsky, W. Willinger, Estimators for long range dependence: an empirical study, *Fractals,* vol.3, no.4, 1995, pp.785 – 788

[31] D. W. K. Andrews, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59, 1991, pp. 817–858

[32] W. Wang et all, Detecting long-memory: Monte Carlo simulations and application to daily streamflow processes, www.hydrol-earth-syst-sci-discuss.net/3/1603/2006/

[33] C.- K. Peng et all, Mosaic organisation of DNA nucleotides, *Physical Review E*, vol. 49, no.2, 1994, pp. 1685 – 1689

[34] S. V. Buldyrev et all. Long-Range Correlation Properties of Coding and Noncoding DNA Sequences: GenBank Analysis, *Physical Reviews E*, vol. 51, 1995, pp. 5084 – 5091

[35] Y. Ashkenazy et all., Magnitude and Sign Correlations in Heartbeat Fluctuations, *Physical Review Letters*, vol. 86, 2001, pp. 1900-1903

[36] http://www.physionet.org/physiobank/database/synthetic/tns/paper2/node2.html

[37] J. Geweke, S. Porter-Hudak, The estimation and application of long memory time series models, Journal of Time Series Analysis, 4, 1983, pp. 221–238

[38] http://www.soft32.com/Download/free-trial/Kaoti XL/ 4-191370-1.html