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 Engineering
Ovidius University of Constantza
124, Mamaia Boulevard, Constantza
ROMANIA
cgherghina@gmail.com

CARMEN MAFTEI
Faculty of Civil Engineering
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:

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

where:

- \( n \) is the sample volume,
- \( x_1, x_2, \ldots, x_n \) are the original data,
- \( x'_1, x'_2, \ldots, x'_n \) are the ordered data,
- \( \bar{x} \) is the sample mean of the data,
- \( w = (w_1, w_2, \ldots, w_n) \) or

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

- \( M \) 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 \( h \) \((h \in \mathbb{N}^+)\) is defined by:

\[
\gamma(h) = \text{Cov}(X_t, X_{t+h})
\]

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

\[
\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad h \in \mathbb{N}^+.
\]

If \( x_1, \ldots, x_n \) 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:
\[ H_0: \sigma_i^2 = \sigma_j^2 = \ldots = \sigma_k^2, \]
and its alternative:
\[ H_1: \sigma_i^2 \neq \sigma_j^2, \]
for at least one pair \((i,j)\).

The test statistic is:
\[
W = \frac{n-k}{k-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Z_{ij} - \overline{Z}_i)^2
\]
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,
- \(Z_{ij} = |X_{ij} - \overline{X}_i|\),
- \(\overline{Z}_i\) 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 \(\alpha\) if
\[
W > F_{\alpha,k-1,n-k}, \quad \text{where } F_{\alpha,k-1,n-k} \text{ is the upper critical value of the F distribution with } k-1 \text{ and } n-k \text{ degrees of freedom at the significance level of } \alpha.
\]

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.

A time series \((X_t)_{t \in N}\) is divided into \(d\) sub-series of length \(m\). For each sub-series \(n = 1, \ldots, d\):
- Find the mean, \(E_n\) and the standard deviation, \(S_n\);
- Normalize the data \((X_{in})\) by subtracting the sub-series mean:
\[
Z_{in} = X_{in} - E_n, \quad i = 1, \ldots, m;
\]
- Create a cumulative time series:
\[ Y_m = \sum_{j=1}^{i} Z_{j}, \quad i = 1, \ldots, m; \]

- Find the range
\[ R_n = \max_{j=1, \ldots, m} Y_{jn} - \min_{j=1, \ldots, m} Y_{jn}; \]

- 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
\[ Ht \cdot c \cdot t^H. \]

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:
\[ Y_k = \sum_{j=1}^{k} (X_k - \overline{X}), \quad k = 1, \ldots, N; \]
and find the range
\[ R(N) = \max_{k=1, \ldots, N} Y_k - \min_{j=1, \ldots, N} Y_k; \]
- instead of using the standard deviation to normalize \( R(N), \) he uses the following sum:
\[ S_q(N) = \sqrt{\sum_{j=1}^{N} (X_j - \overline{X})^2 + 2 \sum_{j=1}^{q} \omega_j(q) \left[ \sum_{j=1}^{N} (X_j - \overline{X})(X_{i-j} - \overline{X}) \right]} \]
where
\[ \omega_j(q) = 1 - \frac{j}{q+1}, q < N. \]

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:
\[ q = \left[ \frac{3N}{2} \right] \left( \frac{2\hat{\rho}}{1 - \hat{\rho}^2} \right)^{\frac{1}{q}}, [31] \quad (*) \]
\[ q = \left[ \frac{N}{10} \right] \left( \frac{2\hat{\rho}}{1 - \hat{\rho}^2} \right)^{\frac{1}{q}}, [32] \quad (**) \]
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
\[ X_m(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, k = 1, 2, \ldots, d \]
is calculated, as well as, its sample variance:
\[ \text{Var} X_m(k) = \frac{1}{d} \sum_{k=1}^{d} (X_m(k) - \overline{X}). \]

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 \( 2H - 2. \)

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 \( \text{Var} X_m. \)

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 \( n(H-1) \).

### 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:

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

- 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 \( (X_k)_{k=1,N} \) 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 \( \lambda \) is a frequency.

It was shown that a series with LRD should have a periodogram proportional to \( \lambda^{2-2H} \) 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>
<thead>
<tr>
<th>Station</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Elevation (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tulea</td>
<td>+45.11</td>
<td>+28.49</td>
<td>4.36</td>
</tr>
<tr>
<td>Jurilova</td>
<td>+44.46</td>
<td>+28.55</td>
<td>37.65</td>
</tr>
<tr>
<td>Corogia</td>
<td>+44.44</td>
<td>+28.30</td>
<td>219.20</td>
</tr>
<tr>
<td>Harsenov</td>
<td>+44.41</td>
<td>+27.57</td>
<td>37.51</td>
</tr>
<tr>
<td>Cemaraia</td>
<td>+44.21</td>
<td>+28.03</td>
<td>87.17</td>
</tr>
<tr>
<td>Medgidia</td>
<td>+44.15</td>
<td>+28.14</td>
<td>68.54</td>
</tr>
<tr>
<td>Constanta</td>
<td>+44.13</td>
<td>+28.38</td>
<td>12.80</td>
</tr>
<tr>
<td>Adacミシ</td>
<td>+44.08</td>
<td>+28.00</td>
<td>158.00</td>
</tr>
<tr>
<td>Mangalia</td>
<td>+43.49</td>
<td>+28.35</td>
<td>6.00</td>
</tr>
<tr>
<td>Sulina</td>
<td>+45.09</td>
<td>+29.39</td>
<td>2.08</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Station</th>
<th>Kolmogorov – Smirnov ( * )</th>
<th>Shapiro – Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adamclisi</td>
<td>0.106 30 0.200</td>
<td>0.565 39 0.220</td>
</tr>
<tr>
<td>Cernavoda</td>
<td>0.087 30 0.200</td>
<td>0.566 39 0.279</td>
</tr>
<tr>
<td>Medgidia</td>
<td>0.058 39 0.200</td>
<td>0.588 39 0.964</td>
</tr>
<tr>
<td>Harsova</td>
<td>0.128 30 0.037</td>
<td>0.504 39 0.003</td>
</tr>
<tr>
<td>Corugea</td>
<td>0.050 30 0.010</td>
<td>0.575 39 0.108</td>
</tr>
<tr>
<td>Tulcea</td>
<td>0.116 39 0.164</td>
<td>0.455 39 0.118</td>
</tr>
<tr>
<td>Salina</td>
<td>0.073 30 0.200</td>
<td>0.581 39 0.707</td>
</tr>
<tr>
<td>Jurlovaca</td>
<td>0.085 30 0.200</td>
<td>0.566 39 0.245</td>
</tr>
<tr>
<td>Constanta</td>
<td>0.124 30 0.033</td>
<td>0.574 39 0.502</td>
</tr>
<tr>
<td>Mangalia</td>
<td>0.078 30 0.200</td>
<td>0.577 39 0.606</td>
</tr>
</tbody>
</table>

\*This is a lower bound of the true significance

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).

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

<table>
<thead>
<tr>
<th>Lag</th>
<th>ACF</th>
<th>Std. err.</th>
<th>Box-Ljung Statistic</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.006</td>
<td>0.154</td>
<td>0.001</td>
<td>1 0.971</td>
</tr>
<tr>
<td>2</td>
<td>0.036</td>
<td>0.122</td>
<td>0.157</td>
<td>2 0.934</td>
</tr>
<tr>
<td>3</td>
<td>0.112</td>
<td>0.130</td>
<td>0.725</td>
<td>3 0.888</td>
</tr>
<tr>
<td>4</td>
<td>0.164</td>
<td>0.148</td>
<td>1.988</td>
<td>4 0.735</td>
</tr>
<tr>
<td>5</td>
<td>0.082</td>
<td>0.166</td>
<td>2.353</td>
<td>5 0.798</td>
</tr>
<tr>
<td>6</td>
<td>0.036</td>
<td>0.144</td>
<td>2.736</td>
<td>6 0.845</td>
</tr>
<tr>
<td>7</td>
<td>0.033</td>
<td>0.144</td>
<td>3.130</td>
<td>7 0.812</td>
</tr>
<tr>
<td>8</td>
<td>0.033</td>
<td>0.139</td>
<td>3.142</td>
<td>8 0.802</td>
</tr>
<tr>
<td>9</td>
<td>0.036</td>
<td>0.137</td>
<td>3.259</td>
<td>9 0.848</td>
</tr>
<tr>
<td>10</td>
<td>0.006</td>
<td>0.135</td>
<td>3.361</td>
<td>10 0.992</td>
</tr>
<tr>
<td>11</td>
<td>0.033</td>
<td>0.132</td>
<td>3.777</td>
<td>11 0.995</td>
</tr>
<tr>
<td>12</td>
<td>0.033</td>
<td>0.130</td>
<td>4.068</td>
<td>12 0.998</td>
</tr>
<tr>
<td>13</td>
<td>0.033</td>
<td>0.128</td>
<td>4.718</td>
<td>13 0.991</td>
</tr>
<tr>
<td>14</td>
<td>0.034</td>
<td>0.125</td>
<td>4.712</td>
<td>14 0.990</td>
</tr>
<tr>
<td>15</td>
<td>0.030</td>
<td>0.123</td>
<td>5.145</td>
<td>15 0.991</td>
</tr>
<tr>
<td>16</td>
<td>0.034</td>
<td>0.120</td>
<td>5.551</td>
<td>16 0.994</td>
</tr>
</tbody>
</table>

\* The underlying process assumed is independence (white noise).
\* 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.

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.

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&Heighnian 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

<table>
<thead>
<tr>
<th>Station</th>
<th>Buishard</th>
<th>Pettitt</th>
<th>Lee &amp; Heighnian</th>
<th>Hubert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adamclisi</td>
<td>no</td>
<td>no</td>
<td>no – Apr 2004</td>
<td>no – Nov 1966, Apr 1971</td>
</tr>
<tr>
<td>Medgidia</td>
<td>yes</td>
<td>yes</td>
<td>no – May 1997</td>
<td>no – Apr 1971, Jul 1991</td>
</tr>
<tr>
<td>Corugea</td>
<td>yes</td>
<td>yes</td>
<td>no – Aug 1972, Apr 2004</td>
<td>no – Aug 1972, Apr 2005</td>
</tr>
<tr>
<td>Tulcea</td>
<td>yes</td>
<td>yes</td>
<td>no – Apr 2004</td>
<td>no – May 1997</td>
</tr>
<tr>
<td>Sulina</td>
<td>no</td>
<td>no</td>
<td>no – Aug 1982</td>
<td>no – Jul 1972</td>
</tr>
<tr>
<td>Jurilovca</td>
<td>yes</td>
<td>yes</td>
<td>no – Jun 1985</td>
<td>no – Nov 1969</td>
</tr>
<tr>
<td>Constanta</td>
<td>-</td>
<td>yes</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mangalia</td>
<td>yes</td>
<td>yes</td>
<td>no – Aug 2005</td>
<td>no – Aug 2005</td>
</tr>
</tbody>
</table>

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

### Table VI. Results of Levene test

<table>
<thead>
<tr>
<th>Station</th>
<th>Annual</th>
<th>Monthly</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig</td>
</tr>
<tr>
<td>Adamclisi</td>
<td>2.494</td>
<td>0.123</td>
</tr>
<tr>
<td>Cernavoda</td>
<td>0.581</td>
<td>0.310</td>
</tr>
<tr>
<td>Medgidia</td>
<td>0.014</td>
<td>0.906</td>
</tr>
<tr>
<td>Harsova</td>
<td>5.777</td>
<td>0.021</td>
</tr>
<tr>
<td>Corugea</td>
<td>3.265</td>
<td>0.079</td>
</tr>
<tr>
<td>Tulcea</td>
<td>4.892</td>
<td>0.033</td>
</tr>
<tr>
<td>Sulina</td>
<td>0.575</td>
<td>0.453</td>
</tr>
<tr>
<td>Jurilovca</td>
<td>0.911</td>
<td>0.346</td>
</tr>
<tr>
<td>Constanta</td>
<td>0.466</td>
<td>0.499</td>
</tr>
<tr>
<td>Mangalia</td>
<td>0.392</td>
<td>0.532</td>
</tr>
</tbody>
</table>
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 KaotlXL [38], in Table VII.

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

<table>
<thead>
<tr>
<th>Station</th>
<th>Adamclisi</th>
<th>Cernavoda</th>
<th>Constanta</th>
<th>Corugea</th>
<th>Hanova</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>0.7207</td>
<td>0.9775</td>
<td>0.9738</td>
<td>0.9224</td>
<td>0.8028</td>
</tr>
</tbody>
</table>

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.
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

<table>
<thead>
<tr>
<th>Station</th>
<th>Adamclisi</th>
<th>Cernavodă</th>
<th>Constanta</th>
<th>Constea</th>
<th>Harsova</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>0.5847</td>
<td>0.6697</td>
<td>0.7316</td>
<td>0.6932</td>
<td>0.6675</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Station</th>
<th>Jiuviolca</th>
<th>Medgidia</th>
<th>Mangalia</th>
<th>Sulina</th>
<th>Tulcea</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>0.7303</td>
<td>0.6551</td>
<td>0.5395</td>
<td>0.7441</td>
<td>0.5723</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Station</th>
<th>Adamclisi</th>
<th>Cernavodă</th>
<th>Constanta</th>
<th>Constea</th>
<th>Harsova</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>0.0112</td>
<td>0.0125</td>
<td>0.0177</td>
<td>0.0147</td>
<td>0.016</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Station</th>
<th>Jiuviolca</th>
<th>Medgidia</th>
<th>Mangalia</th>
<th>Sulina</th>
<th>Tulcea</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>0.0139</td>
<td>0.0109</td>
<td>0.0239</td>
<td>0.0134</td>
<td>0.0125</td>
</tr>
</tbody>
</table>

## 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


