SYNTHESIS OF THE RESEARCH RAPPORT 2007 - 2010

CONTENT

1. OBJECTIVES

- Statistical description of data series.
- Building models and prediction of precipitation evolution.
- Frequential analysis.
- Study of LRD property of data series.
- Determination of factionary dimensions of data series. Multifractal analysis of precipitation series.
- Models for extreme precipitation.

2. GENERAL DATA

Dobrudja is a region situated in the South-East of Romania between the Black Sea and the lower Danube river. Its structure (without Danube Delta) is of a plateau with hilly aspect.

Generally, Dobroudja's climate is temperate - continental and is divided in 2 units:

- (I) which contains the Danube Delta, its south, the too lagoons (Razim lake and Sinoe lake) and the eastern region (a part of 10 - 20 km width along the sea)

- (II) which contains the rest of territory.

The air average temperature is over 11^0 C/ year towards the littoral area and in the Danube floodplain, and less than 10^0C and 11^0C in the north and centre.

The studied data were collected between January 1965 and December 2005 at ten meteorological stations and consists of mean annual and mean monthly precipitation series.

3. RESULTS

3.1. Statistical analysis

The characteristic variables were the average monthly and annual data and the extreme precipitations.

After the graphical representation, the normality hypothesis has been verified by Kolmogorov – Smirnov, Shapiro – Wilk, Jarque – Bera tests and Q-Q plot. The conclusion was that the annual series are normally distributed and the monthly series don't have this property.

The homogeneity detection has been done by Wilconxon test, at the significance level of 1% and 5%. Mangalia annual series is homogenous. The homogeneity hypothesis has been accepted at 1% and rejected at 5% for the others annual series. The homogeneity hypothesis has been rejected for the monthly series.

The results of break tests were different. For the annual series Sulina, the break moment was 1981 and for the other annual series, the Buishand and Pettitt tests rejected the break hypothesis, but Lee & Heginian and Hubert tests lead us to accept 2003 as a break moment for the annual series.

All the monthly series present oultliers, that were removed before the modelling process.

The study of data correlation has been performed by the autocorrelation function, and the conclusion was that between the annual series, four are correlated and the monthly series are dependent.

The isohyets analysis reveal an increasing in the mean annual precipitation in the period 1995 - 2005, excepting in the Danube Plain and at Sulina meteorological station.

3.2. Mathemtical models

The models types determined for the precipitation series were:

I. Multiplicative models, of the type:

$$X_t = Y_t \cdot S_t \cdot \varepsilon_t,\ t \geq 0,$$

where: X_t is the series, Y_t - the trend, S_t - the seasonal component and ε_t - the residual.

II. Models of ARIMA/ FARIMA type.

III. Models obtained by neural networks methods.

IV. Models obtained by using genetic algorithms (GEP, AdaGEP) and hybrids (ARIMA – AdaGEP).

V. Frequential models (IDF curves).

VI. Nonparametric models.

In the following we shall exemplify.

1. For the monthly series *Medgidia*, that presents the break points May 1971 and July 1991, the following models have been determined (Fig.1):

For the period January 1965 - May 1971, after a Box - Cox transformation of parameter λ = 0.42, the resulted series has been modelled by:

$$X_t = Z_t - 0.2242 Z_{t-4}, t \in \overline{5,76},$$

where $(Z_t)_{t \in \overline{1,76}}$ is a whitw noise.

- For the period July 1971 - July 1991 a multiplicative model has been determined. The trend was calculated by the moving average method; the seasonality factors are given in Table 1.

Table 1. Seasonality factors (%)

Jan.	Febr.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
182.2	144.9	114.1	75.6	83.3	102.2	74.2	64.0	68.0	62.4	104.9	0.991

- For the period September 1991 - December 2005, after considering the first order difference, the proposed model is of MA(11):

$$X_t = Z_t - 0.6399 Z_{t-11},\ t \geq 12,$$

where $(Z_t)_{t \geq 1}$ is a white noise.

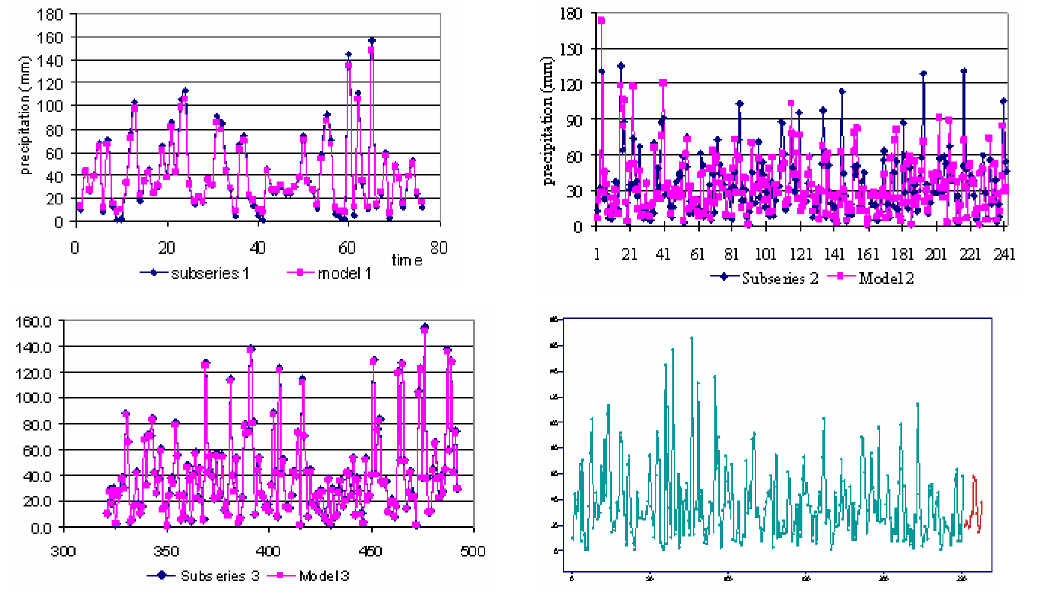

Fig.1. Medgidia monthly series: the sub-series and their models.
The prediction of 12 month (Winters method)

2. For the monthly series *Sulina*, after a Box - Cox transformation with the parameter $\lambda = 0.34$, the hypothesis that the new series (X_t) is normally distributed was accepted. The correlation hypothesis is accepted, at the confidence level of 95%. The homoscedasticity hypothesis was rejected after the application of Bartlett test, at the same confidence level.

- The sub – series (X_t) on the period January 1965 – August 1982 forms a white noise.
- A multiplicative model has been determined for the sub – series (X_t) on the period September 1982 – December 2005; the trend is given by:

$$\overline{X_t} = 4.2768 + 0.6475 * \cos(0.05348t - 2.9331) \text{ (Fig.2)},$$

The residual variance being 0.326515, and the residual forms a Gaussian noise.

Table 2. Seasonality factors (%)

Jan.	Febr.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
0.834	0.875	0.888	0.911	0.978	1.430	0.858	0.956	1.082	1.048	1.148	0.991

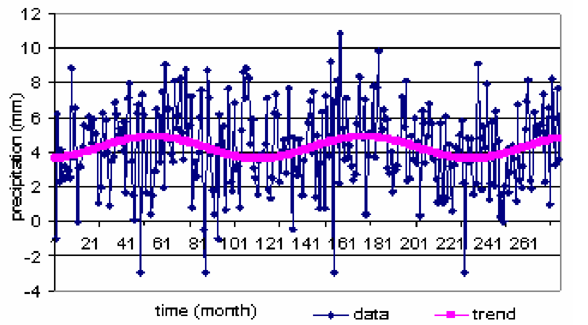

Fig. 2. Data and trend of the sub-series September 1982 – December 2005

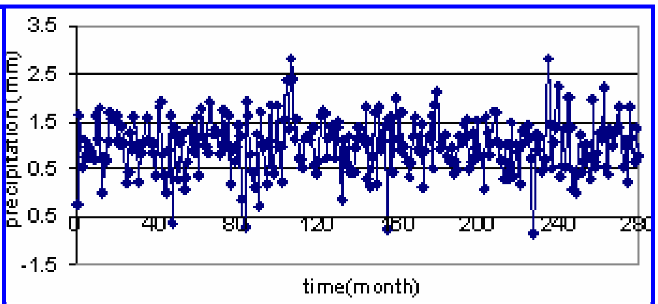

Fig. 3. Residual in the multiplicative model for the sub - series September 1982 – December 2005

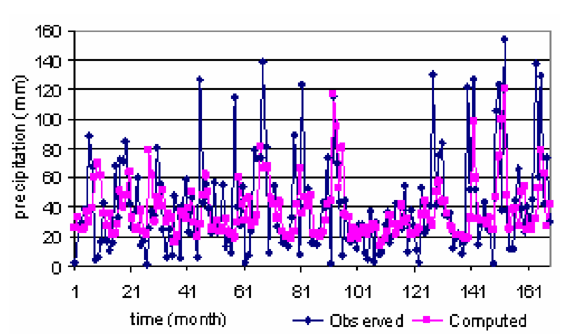

Fig. 4. Prediction on 170 month using neural networks for Sulina monthly series

Predicţia evoluţiei precipitaţiilor s-a făcut prin antrenarea unei reţele neurale (Fig.4). Datele din fiecare sub-serie au fost împărţite în 3 mulţimi: de training, de validare şi de testare. S-a folosit o reţea „two-layer feed forward" cu un layer de output şi layere ascunse având între 5 - 10 neuroni *tansig*sau *purelin* şi output - ul un neuron *tansig*.

3. For Sulina annual series, the best model determined (FARIMA), after the mean subtraction, has the equation

$$(1-B)^{0.28} X_t = Z_t ,$$

where (Z_t) is a white noise. (Fig.5)

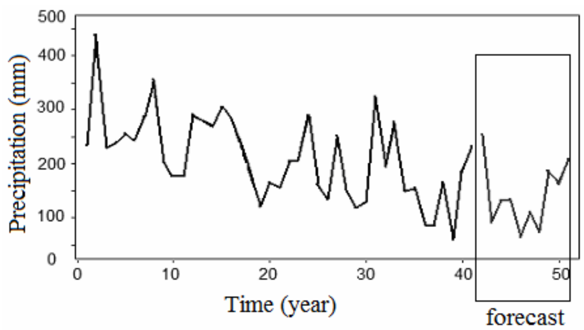

Fig. 5. Sulina annula series

4. Since the meteorological series present high variability and the non-linear evolution can not be always be captured by ARIMA/FARIMA models, GEP has been used to determine alternative models (Fig.6).

5. Box-Jenkins methodology and GEP have been combined, obtaining hybrid models, AdaGEP – ARIMA, that improve those built by an adaptive genetic proposed (AdaGEP) (Fig.7).

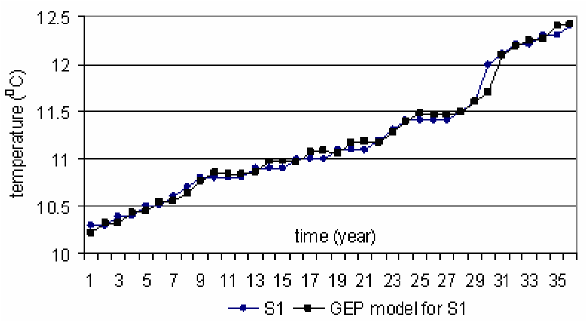

Fig.6. GEP model for Sulina annual series

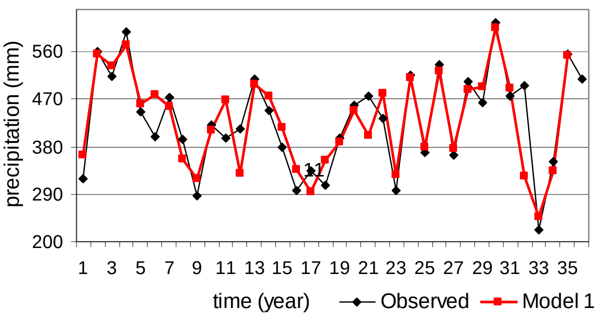

Fig.7. AdaGEP-AR model for Medgidia annual series

6. Buiding IDF curves (Figs 8,9) has been done by Gumbel law:

$$F(x) = \exp[-\exp(-(x-a)/b)].$$

With the reduced variable, $u = (x-a)/b$, the repartition function can be written:

$$F(u) = \exp[-\exp(-u)]$$

and
$$u(F) = -\ln(-\ln(-F(x))).$$

The modelling stages are:

i. Data preparation, to draw the polygon of cumulated frequencies:

- increasingly ordering the values and attaching a rank, *r*, to each one,

- the calculation of the empirical frequencies, by Hazen formula:

$$F(x_r) = (r-0.5)/n,$$

where *n* is the values number.

ii. Calculus of Gumbel's reduced variable, *u*;

iii. The representation of the pairs (u_i, x_i) and the determination of the parameters *a* and *b* of the regression line, by the momemts methods or by the least square method;

iv. Calculus of the frequency $F = 1-1/T$ for different return periods, T;

v. Calculus of $u(F)$;

vi. Calculus of the estimated value of the quantile $x_q = a + bu_q$.

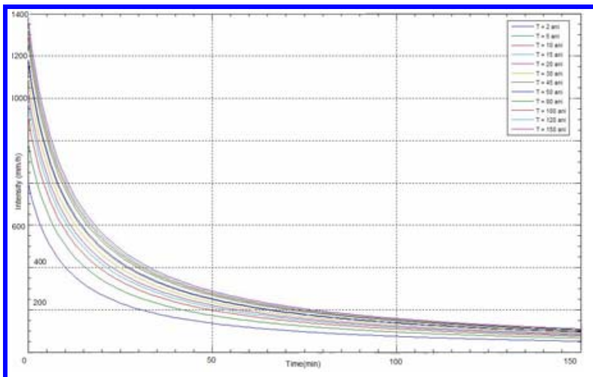

Fig.8. IDF curves for Medgidia series

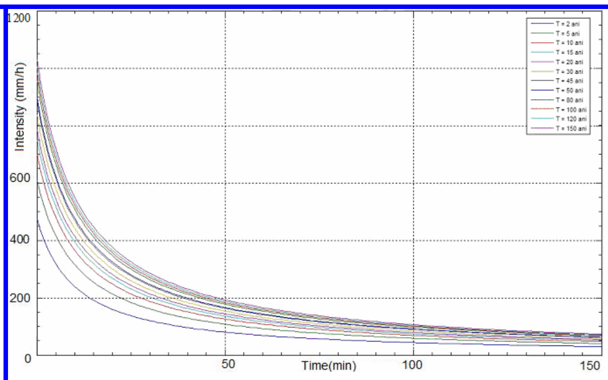

Fig.9. IDF curves for Sulina series

3.3. LRD analysis

The LRD analysis has been performed by the calculus of Hurst coefficient, *H*, by different methods, as:

- R/S and Lo's modified R/S statistic;
- Aggregated Variance method;
- Absolute Moments method;
- Detrended fluctuation analysis (DFA);
- GPH;
- Ratio of variance of residuals;
- Periodogram;
- Wavelets.

If the process is a random series, then $H = 0.5$. $0 < H < 0.5$ indicates an anti - persistent series, and $0.5 < H < 1$ indicates a persistent series.

The results obtained for the annual data were different, function of the used methods. They are given in Tables 3-5.

Table 3. Hurst coefficient of annual series, calculated by R/S 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

Table 4. Hurst coefficient of 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 5. Hurst coefficient of 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

Table 6. Hurst coefficient of annual series, calculated by GPH metod

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

A program that calculates the coefficients by the mentioned methods have been built. (Fig.10).

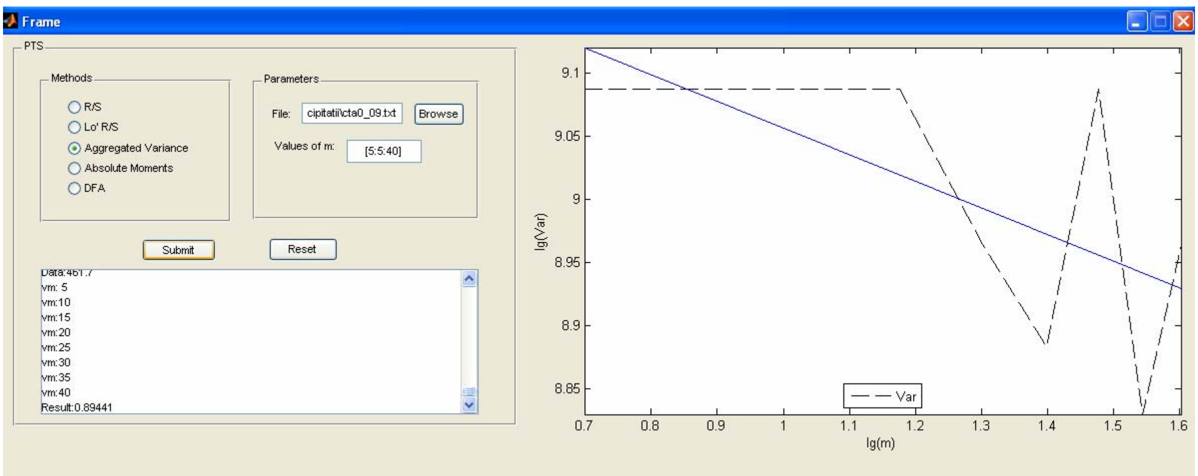

Fig.10. Print - screen of the program interface

In the next figured we present the R/S chart and the evolution of V-statistic associated to Adamclisi series. Analogous charts heve been drawn for all studied series.

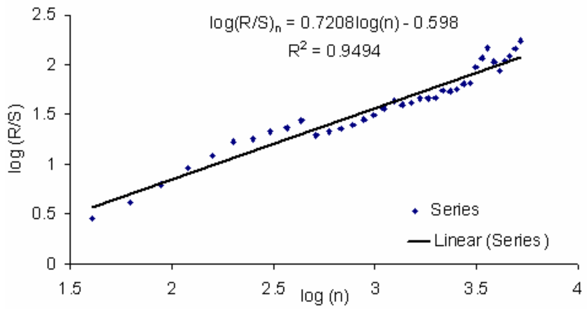

Fig. 11. R/S chart of Adamclisi monthly series

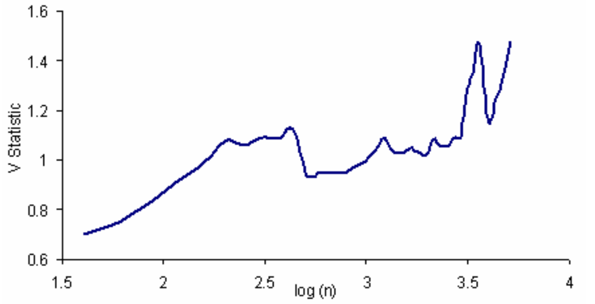

Fig. 12. The V - statistic associated to Adamclisi monthly series

3.4. Characterisation of precipitation series by factionary dimensions

The study of fractal character of precipitation series can be done using the fractionary dimensions, as: ca *box, ruler, informational, fragmentation dimensions* etc. In our case, from Table it results the homogeneity of box dimensions, approx. 1.9, for the monthly data series.

Table 6. Box dimension of monthly series

	Box dimension	Std. deviation
Adamclisi	1.89359	0.0058726
Cernavoda	1.89878	0.0078325
Constanta	1.90286	0.0055319
Corugea	1.89716	0.0055399
Harsova	1.89729	0.0054955
Jurilovca	1.90063	0.0057935
Mangalia	1.90238	0.0056315
Medgidia	1.88917	0.0062289
Sulina	1.90197	0.0057572
Tulcea	1.89563	0.0055636

3.5. Multifractal analysis of precipitation series

The study of multifractal character of monthly precipitation series imply the determination of the scaling exponent, $\tau(q)$ and of the singularity spectrum, $f(\alpha)$. Two techniques have been used (box-counting and wavelets transform maxima modulus - WTMM).

The results are presented in an article submitted [A. Bărbulescu s.a., 2010].

3.6. Models for extreme precipitation

The study of maximum daily and monthly precipitation has been realized and GPD models has been built, the results being presented in two articles, one, in print at IJMC and another submitted at Int. Journal of Climatology. See also the book [Bărbulescu s.a., NOVA Publishing, 2010]. The diagnostic plots of the proposed model for Tulcea series is given in Fig.15.

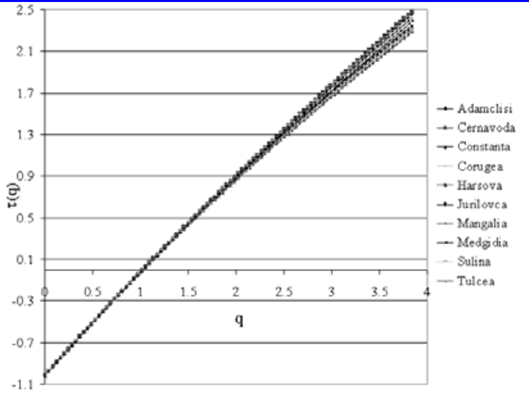

Fig.13. $\tau(q)$ determined by "box-counting method"

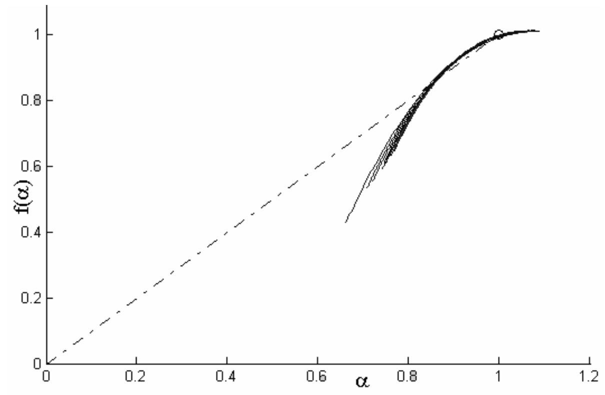

Fig.14. $f(\alpha)$ - spectrum for the normalized data

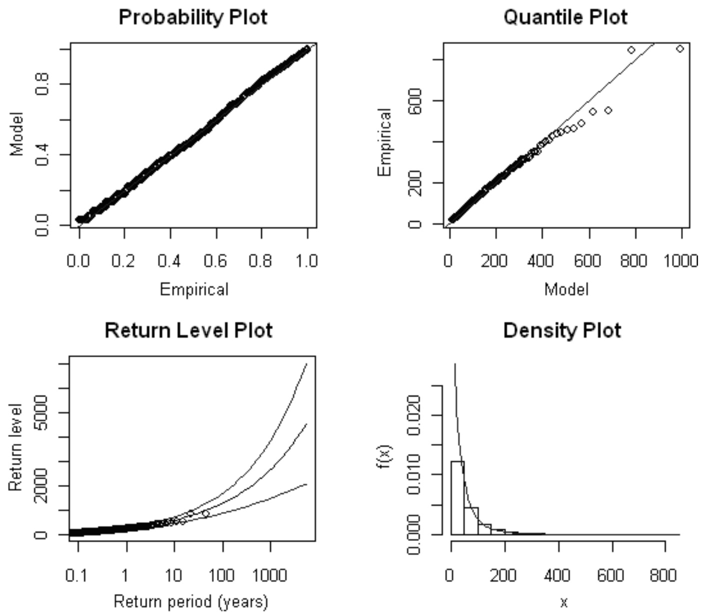

Fig.15. GPD model for the daily series Tulcea [IJMC, 2010]

Also, the PMP study has been realised (Fig.16), the results being in print in our book [Bărbulescu s.a., LAP Publishing, 2010].

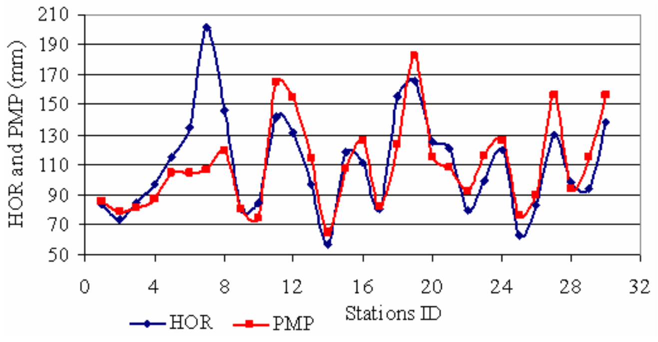

Fig.16. Results of Hershfield's method for maximum precipitation in 24 hours

4. CONCLUSION

A systematic study of the precipitation from Dobrudja region has been developed in the present research grant. The obtained models can be successfully used to predict the precipitation evolution in this region. Also, the regional model proposed by us, could be inserted in a global model, that could describe the climatic evolution in the South – East of Europe.

5. REFRENCES

H. Akaike, Information theory and an extension of the maximum likelihood principle, 2nd International Symposium on Information Theory, B. N. Petrov and F. Csaki (eds.), Akademia Kiafo, Budapest, pp. 267 - 281

A. Bărbulescu, *Time series with applications*, Junimea, Iaşi, 2002

A. Bărbulescu, E. Băutu, Mathematical models of climate evolution in Dobrudja, *Theoretical and Applied Climatology*, vol.**100**, nos.1 – 2/March, 2010, pp. 29 - 44, DOI 10.1007/s00704 – 009 – 0160 – 7

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

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

A. Bărbulescu, E. Băutu, ARIMA and GEP models for climate variation, *International Journal of Mathematics and Computation*, June 2009, Volume 3, No J09, pp. 1-7

A. Bărbulescu, E. Băutu, Time Series Modeling Using an Adaptive Gene Expression Programming, *International Journal of Mathematical Models and Methods in Applied Sciences*, Issue 2, Volume 3, 2009, pp. 85 – 93

A. Bărbulescu, J. Deguenon, D. Teodorescu, *Study of water resources in the Black Sea region. Some mathematical models*, Nova Publishing, USA, 2010 (in print)

A. Bărbulescu, C. Maftei, E. Băutu, *Modeling the hydro-meteorological time series. Applications to Dobrudja region*, Lambert Academic Publishing GmbH & Co. KG**,** Germany, 2010 (ISBN 978-3-8433-6578-9)

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

A. Bărbulescu, D - C. Toncu, Modelling Precipitation Influence on Tăbăcărie Lake Water Quality, *Mathematical Methods, Computational Techniques, Intelligent Systems*, 2010, pp.152 - 158

A. Bărbulescu, C. Şerban (Gherghina), C. Maftei, Evaluation of Hurst exponent for precipitation time series, *Latest Trends on Computers*, Vol.**II**, 2010, pp. 590 – 595

A. Bărbulescu, The analysis of correlation of some ions concentration in rainwater in an urban area, *International Journal of Mathematical Models and Methods in Applied Sciences*, Issue **2**, Vol. 4, 2010, pp.105 – 112

A. Bărbulescu, C. Șerban (Gherghina), C. Maftei, Statistical Analysis and Evaluation of Hurst Coefficient for Annual and Monthly Precipitation Time Series, *WSEAS Transactions on Mathematics*, Issue **10**, Vol. 9, October 2010, pp. 791 – 800 (ISSN 1109 - 2769)

A. Bărbulescu, C. Șerban, A. Cârsteanu, Statistical and multifractal analysis of rainfall of Romania, 2010, submitted

P. J. Brockwell, R.A. Davis, *Time series analysis, forecasting and control*, Holden - Day, San Francisco, 1976

T.A. Buishand, Some methods for testing the homogeneity of rainfall records, *Journal of Hydrology* **58**, 1982, pp. 11-27.

T. A. Buishand, Tests for detecting a shift in the mean of hydrological time series, *Journal of Hydrology* **58**, 1984, pp. 51-69.10.

A. Chhabra, R. Jensen, Direct determination of the f(alpha) singularity spectrum, *Physical Review Letters* **62** (12), 1989, pp.1327–1330.

J. Deguenon, A. Bărbulescu, Study of Extreme Daily Rainfall Using GPD Model, IJMC, 2010, în curs de apariție

J. Deguenon, A. Bărbulescu, GPD Models for extreme rainfall with cyclic trend detection over ten stations in Dobrudja, International Journal of Climatology, 2010, submis 5.08.2010

P. Hubert et al, 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

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

Khronostat, http://www.hydrosciences.org/spip.php?article239&lang=en

A. F. S. Lee, S. M., Heghinian, A Shift of the Mean Level in a Sequence of Independent Normal Random Variables - A Bayesian Approach, *Technometrics* **19**, 4, 1977, pp. 503-506

H. Lubes-Niel et al., Caractérisation de fluctuations dans une série chronologique par applications de tests statistiques - Etude bibliographique, *Rapport interne ICCARE* **3**, 1994, ORSTOM – Hydrologie

C. Maftei, A. Bărbulescu, Computing the intensity - duration - frequency curves for a Romanian catchments, *Proceedings of The third International Conference of Mathematical Sciences*, 3-6.03.2008, Al Ain, p.1178-1186.

A. N. Pettitt, A non - parametric approach to the change-point problem, *Applied Statistics* **28**, n°2, 1979, pp.126 - 135.

M. Taqqu, V. Teverovsky, W. Willinger, Estimators for long range dependence: an empirical study, Fractals, vol.**3**, no.4, pp.785-788.

W. Taylor, Change point analyser 2.0 shareware program, Taylor Entreprises, Libertyville, Illinois, 2003, http://variation.com/cpa

V. Venugopal, S.G. Roux, E. Foufoula-Georgiou, A. Arneodo, Revisiting multifractality of high resolution temporal rainfall using a wavelet-based formalism, *Water Resources Research*, **42**, 2006, W06D14, doi:10.1029/2005WR004489

R. Weron, Estimating long range dependence: finite sample properties and confidence intervals, arXiv: cond-mat/0103510v2 9 May 2001