## 1. INTRODUCTION

Using the data from the previous phases of the project (2007, 2008), the following stages were followed in order to obtain ARIMA/FARIMA models for the precipitations series:

- The long range dependence (LRD) analysis – the determination of Hurst coefficient;
- The analysis of break points;
- Testing the hypothesis that the series are Gaussian noises;
- The determination of models for the series that do not satisfy the previous condition. If the series present break points, the models were determined for the integer series and subseries.

## 2. RESULTS

### 2.1. The long range dependence analysis

The LRD was determined by the calculation of Hurst coefficient, $H$, using a program made by us, using the algorithm described in [12], [13].

If the process is a white noise, then the diagram is a straight line, with the slope 0.5. If the process has the LRD property, the slope is bigger than 0.5 and if it is not persistent, less than 0.5.

Between the annual series, only Sulina ($H$ = 0.6) has the LRD property; between the monthly series, Sulina ($H$ = 0.611) and Adamclisi ($H$=0.577) have this property.

### 2.2. Analiza existenţei punctelor de ruptură

In the followings we present the results for the annual series (Table 1)

Table 1. Results of break tests

| Station | Pettitt test | Hubert procedure |
|---|---|---|
| Adamclisi | yes | yes |
| Cernavodă | yes | yes |
| Medgidia | yes | no, 1972 |
| Corugea | yes | yes |
| Constanţa | yes | no, 1972 |
| Hârşova | yes | yes |
| Jurilovca | yes | yes |
| Mangalia | yes | yes |
| Tulcea | yes | yes |
| Sulina | no, 1981 | no, 1981 |

*Yes* signify that the hypothesis that the series doesn't have break points is accepted (at the confidence level of 95%, in the case of Pettitt test). *No*, followed by an year signify that the series presents a break point in that year.

In 70% of cases there is no break point. Since for the series Corugea and Cernavodă only a test rejected the hypothesis of the break absence and there is a small number of data before the break points determined by Hubert segmentation procedure, the models were built only for the entire series.

### 2.3. Testing the hypothesis that the series are Gaussian noises

The results of normality tests are given in Table 2, where:

- the column 2 and 5 contain the values of Kolmogorov – Smirnov and Shapiro – Wilk statistics,
- df is the number of degrees of freedom,
- Sig is the significance level.

If Sig > 0.05, the normality hypothesis is rejected.

After new tests (Jarque Bera, Q-Q plot) on Corugea Corugea, Constanţa and Hârşova series the hypothesis that the first two series are normally distributed was accepted.

If Sig < 0.05, the normality hypothesis is rejected.

After new tests (Jarque Bera, Q-Q plot) on Corugea asupra seriilor Corugea, Constanţa şi Hârşova series the hypothesis that the first two series are normally distributed was accepted.

The stuy of ACF concludes that 4 series are independent.

So, Adamclisi, Corugea, Cernavodă şi Medgidia are Gaussian noises.

After the square root subtraction from the data of Tulcea series, we accept the hypothesis that the new serie is also a Gaussian noise.

Table 2. Normality tests

| Station | Kolmogorov-Smirnov | | | 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 |
| Corugea | 0.124 | 39 | 0.010 | 0.954 | 39 | 0.108 |
| Constanta | 0.124 | 39 | 0.033 | 0.974 | 39 | 0.502 |
| Harsova | 0.128 | 39 | 0.037 | 0.904 | 39 | 0.003 |
| Jurilovca | 0.085 | 39 | 0.200* | 0.964 | 39 | 0.245 |
| Mangalia | 0.078 | 39 | 0.200* | 0.977 | 39 | 0.606 |
| Tulcea | 0.116 | 39 | 0.164 | 0.955 | 39 | 0.118 |
| Sulina | 0.073 | 39 | 0.200* | 0.980 | 39 | 0.707 |

## 2.4. The determination of ARIMA or FARIMA models

*Constanţa series* is correlated and homoscedastic. For the series transformed by taking logarithms, and denoted by ($Y_t$), the best model that has been obtained is:

$$Y_t = 0.9783Y_{t-1} + Z_t,$$

where ($Z_t$) is a white noise with the variance 0.045.

*Hârşova* is correlated and presents 3 outliers. After their removal and taking logarithms, the resulted series, ($Y_t$), is correlated and Gaussian. In these conditions, the following AR(1) model was determined:

$$Y_t = 0.9786Y_{t-1} + Z_t,\ t \geq 2,$$

where ($Z_t$) is the residual, which is a Gaussian white noise with the variance 0.202.

For *Jurilovca* series, after taking logarithms, the model obtained was:

$$Y_t = 5.885 + 0.316Y_{t-1} + Z_t,\ t \geq 2,$$

where ($Z_t$) is a Gaussian noise with the variance 0.083

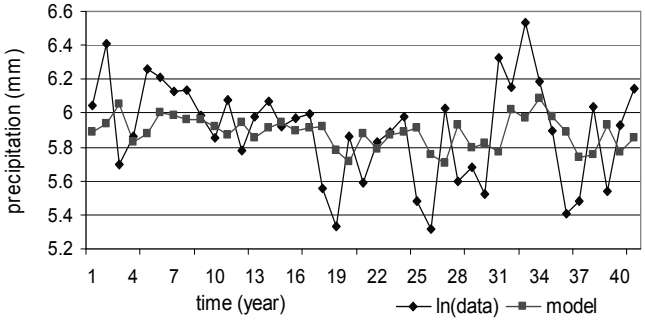


Fig.1. Model for Jurilovca series

For *Mangalia* series, after taking logarithms, the model obtained was:

$$Y_t = 0.09698Y_{t-1} + Z_t,\ t \geq 2,$$

where ($Z_t$)is a Gaussian noise with the variance 0.1079.

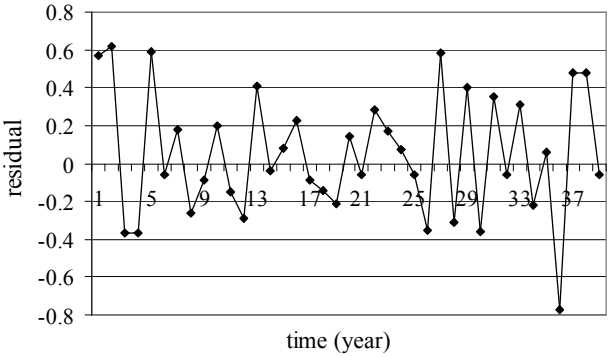


Fig.2. Residual in the model for Mangalia series

Two models were determined for the entire *Sulina* series:

a. After taking logarithm and the mean subtraction:

$$X_t = 0.4021 \cdot X_{t-1} + Z_t,$$

with $(Z_t)$ a white noise (Fig.3)

b. After the mean subtraction:

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

with $(Z_t)$ a white noise with the variance 4618.11.

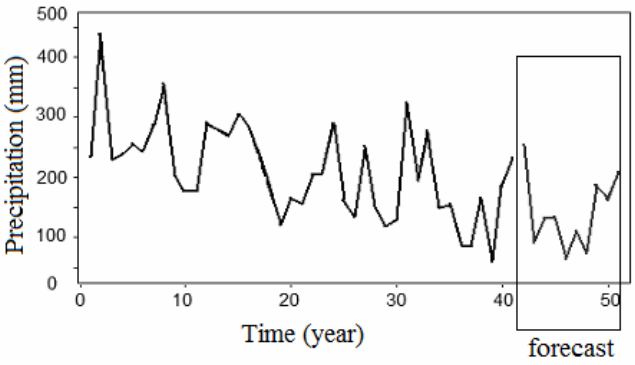


Fig.4. The prediction of precipitations evolution – Sulina series

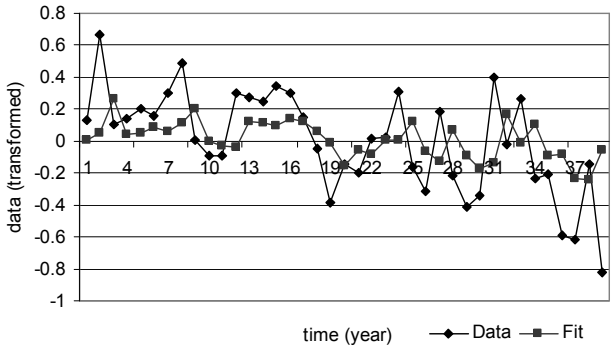


Fig.3. AR(1) model for Sulina series

The second model is of FARIMA type and better describe the series evolution, being used also for prediction (Fig.4)

The sub-series Sulina_1 (1965 - 1981) is Gaussian, independent and identically distributed, with the variance 3795.05.

The sub-series Sulina_2 (1982 – 2003), after the mean subtraction, is a Gaussian noise, with the variance 4296.75.

## 2.5. The determination of alternative models (GEP, ARIMA-AdaGEP)

Since generally the meteoreological series present high variability and their nonlinear evolution can not be captured by ARIMA or FARIMA models, we tried to determine some models combining Box-Jenkins methods and Gene Expresion Programming (hybrid models), that improve those obtained by a genetic adaptive algorithm (Fig.5)

We present the case of Sulina monthly series, that has two break points and for which the AdaGEP model is not satisfactory (Fig.6)

After the mean extraction, the best model for the integral series was an ARMA(2, 2):

$$Y_t = 0.9577Y_{t-1} - 0.9915Y_{t-2} + Z_t - 0.929Z_{t-1} + 0.9914Z_{t-2},$$

where $(Z_t)$ was a white noise with the variance 0.9517 (Fig.7).

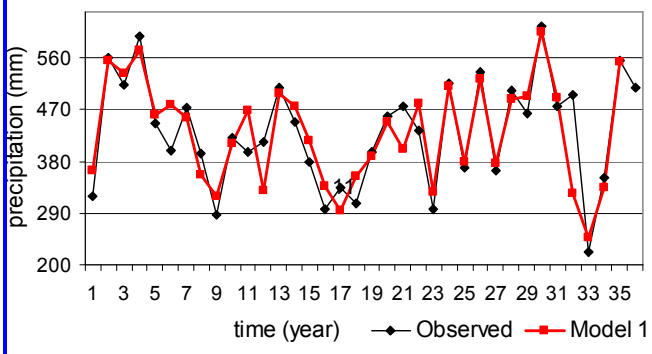


Fig.5. AdaGEP-AR model for Medgidia annual series

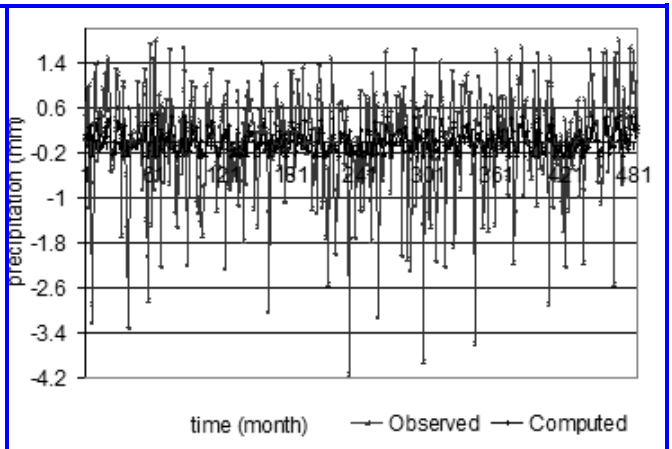


Fig.6. AdaGEP model for Sulina monthly series

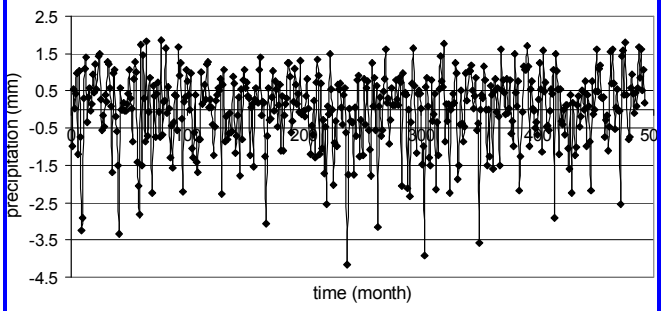


Fig.7. ARMA(2,2) model for monthly Sulina series

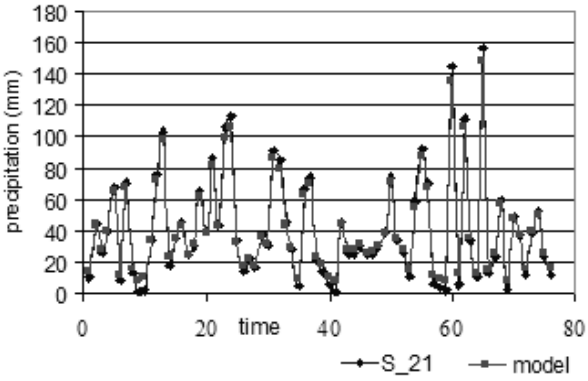


Fig.8. MA(4) model for S1

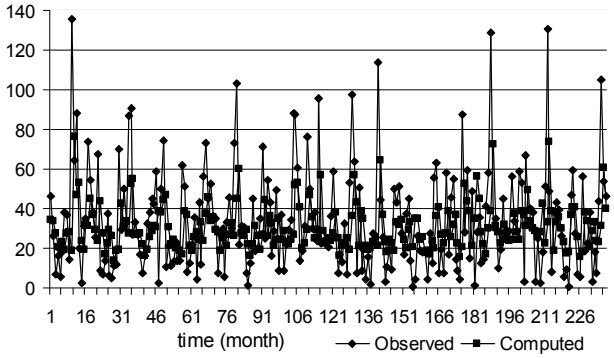


Fig.10. AdaGEP model for S2

For the sub-series S1, S2, S3, delimited by the break points, the following modles were built

- for S1 (MA(4)):

$$X_t = \varepsilon_t - 0.2242\varepsilon_{t-4}, t \in \overline{5,76}, (\varepsilon_t)_{t\in\overline{1,76}},$$

where $(\varepsilon_t)_{t\in\overline{1,76}}$ was a white noise (Fig.8)

In Fig.9 the combined model is found.

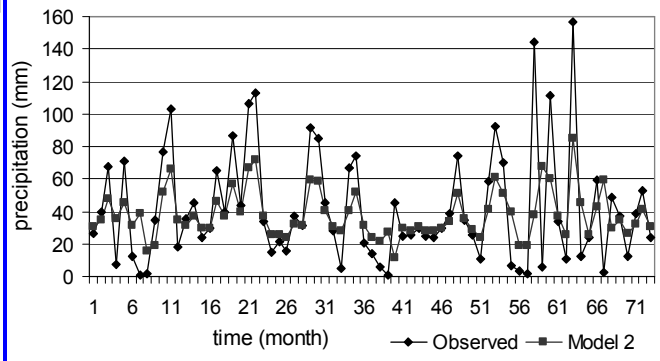


Fig.9. Combined model of S1

- for S2 a good model of ARIMA type wasn't determined, the best being an AdaGEP (Fig.10)

- For S3, the best model was the combined one, AR_GEP (Fig.11).

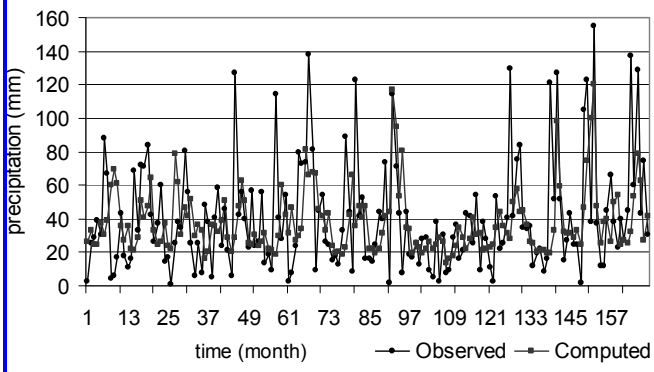


Fig.11. Combined model of S3

## 4. CONCLUSIONS

- After the annual data analysis we decided that four series are Gaussian white noises and one has the LRD property.
- Models of ARIMA/FARIMA type were determined and used for prediction.
- The models selesction was done using Akaike criterion. All the residuals were Gaussian white noises.
- Hybgrid models wee also proposed to improve the ARIMA models. They appear to be an alternative for the classical approach.

## 5. REFERENCES

1. 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

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

3**.** A. Bărbulescu, E. Băutu, *Mathematical models of climate evolution in Dobrudja*, Theoretical and Applied Meteorology, DOI 10.1007/s00704 – 009 – 0160 – 7, ISSN 0177 – 798X (Print) 1434 – 4483 (Online)

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

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

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

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

10. 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

11. 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

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

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