

1. INTRODUCERE

Folosind datele din fazele precedente (2007, 2008) s-au parcurs următoarele etape pentru a obține modele de tip ARIMA/FARIMA pentru seriile de precipitații:

- Analiza dependenței lungi a seriilor - determinarea coeficientului Hurst;
- Analiza existenței punctelor de ruptură;
- Testarea ipotezelor că seriile sunt zgomote Gaussiene;
- Determinarea unor modele pentru seriile care nu sunt în situația de la punctul precedent. Dacă seriile au prezentat puncte de ruptură, s-au determinat modele pentru seriile integrale și subserii.

2. REZULTATE

2.1. Analiza dependenței lungi

Aceasta a fost realizată prin calculul coeficientului Hurst, H , cu ajutorul unui program realizat de noi, după algoritmul descris în [12], [13].

Dacă procesul este zgomot alb, atunci diagrama este o dreaptă cu panta 0.5. Dacă procesul are proprietatea LRD, panta este mai mare ca 0.5, iar dacă nu este persistent în timp, mai mică de 0.5.

Dintre seriile de date medii anuale numai Sulina ($H = 0.6$) prezintă proprietatea de LRD, iar dintre cele lunare, Sulina ($H = 0.611$) și Adamclisi ($H=0.577$).

2.2. Analiza existenței punctelor de ruptură

În cele ce urmează prezentăm rezultatele pentru seriile anuale (Tabel 1), cele pentru datele lunare fiind prezentate în raportul pe anul 2008.

Tabel 1. Teste de ruptură

Stație	Testul Pettitt	Procedura Hubert
Adamclisi	da	da
Cernavodă	da	da
Medgidia	da	nu, 1972
Corugea	da	da
Constanța	da	nu, 1972
Hârșova	da	da
Jurilovca	da	da
Mangalia	da	da
Tulcea	da	da
Sulina	nu, 1981	nu, 1981

da semnifică faptul că ipoteza absenței rupturilor este acceptată (la nivelul de încredere de 95%, în cazul testului Pettitt), iar *nu*, urmat de un an reprezintă anul în care este prezentă ruptura.

În 70% dintre cazuri nu există puncte de ruptură în seriile de timp. Întrucât pentru seriile Corugea și Cernavodă numai un test a respins ipoteza existenței rupturilor și există numai puține date înainte de punctele de ruptură determinate prin procedura de segmentare a lui Hubert, modelele nu au luat în considerare două sub-perioade.

2.3. Testarea ipotezelor că seriile sunt zgomote Gauss – iene

Rezultatele testelor de normalitate sunt date în Tabelul 2, unde:

- coloanele 2 și 5 conțin valorile statisticilor Kolmogorov – Smirnov și Shapiro – Wilk,
- df este numărul de grade de libertate,
- Sig este nivelul de semnificație.

Dacă $Sig < 0.05$, ipoteza normalității este respinsă.

După teste suplimentare (Jarque Bera, Q-Q plot) asupra seriilor Corugea, Constanța și Hârșova a rezultat că primele două sunt normal distribuite.

Studiul funcțiilor de autocorelație a dus la concluzia că 4 serii sunt independente.

Deci, seriile Adamclisi, Corugea, Cernavodă și Medgidia sunt zgomote Gaussiene - iene.

După extragerea radicalului din datele seriei Tulcea, se acceptă ipoteza că acesta este independentă și normal repartizată.

Tabel 2. Teste de normalitate

Statie	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistica	df	Sig.	Statistica	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. Determinarea unor modele de tip ARIMA sau FARIMA

Seria *Constanța* este corelată și homoscedastică (are aceeași varianță). Pentru seria transformată prin logaritmare, notată (Y_t), cel mai bun model (obținut folosind criteriul Akaike) a fost de tip AR(1):

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

unde (Z_t) este un zgomot alb, cu varianța 0.045.

Seria *Hârșova* este corelată și prezintă 3 valori aberante (anii 2, 8, 33). După înlocuirea acestora din urmă cu medii ale celorlalte date și logaritmare, modelul obținut a fost:

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

unde (Z_t) este un zgomot alb Gaussian cu varianță 0.202.

Pentru seria *Jurilovca*, după logaritmare, modelul obținut a fost:

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

unde (Z_t) este un zgomot alb Gaussian cu varianță 0.083

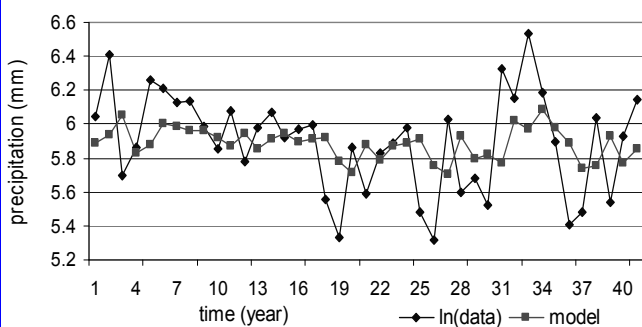


Fig.1. Modelul pentru seria Jurilovca

Pentru seria *Mangalia*, după logaritmare, modelul obținut a fost:

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

unde (Z_t) este un zgomot alb Gaussian cu varianță 0.1079.

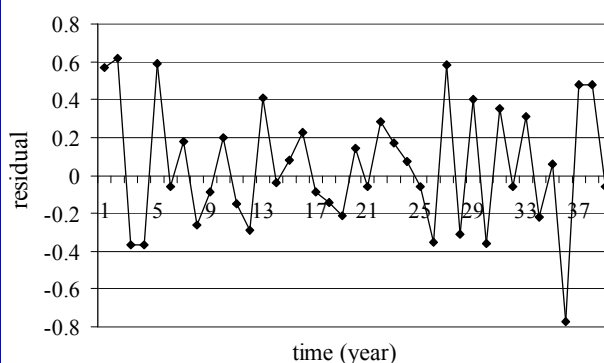


Fig.2. Residuul în modelul seriei Mangalia

Seria *Sulina* are proprietatea LRD și prezintă ruptură în 1981. În consecință, s-au determinat: un model pentru seria întreagă și două modele pentru perioadele subsecvente: 1965 - 1981 și 1982 - 2003.

Pentru seria întreagă au fost propuse două modele:

a. După logaritmare și extragerea mediei:

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

cu (Z_t) zgomot alb. (Fig.3)

Acesta este un model de tip AR(1).

b. După extragerea mediei:

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

cu (Z_t) zgomot alb cu varianța 4618.11.

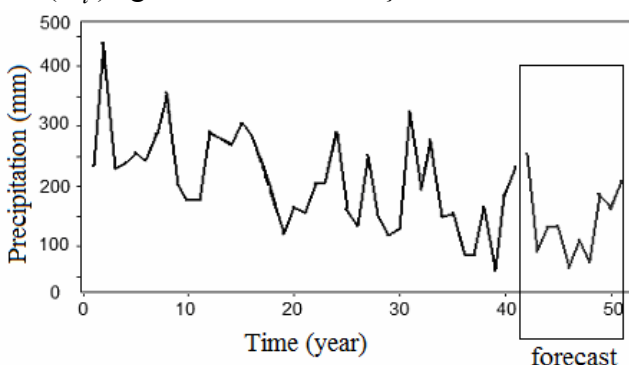


Fig.4. Predicția evoluției precipitațiilor - Sulina

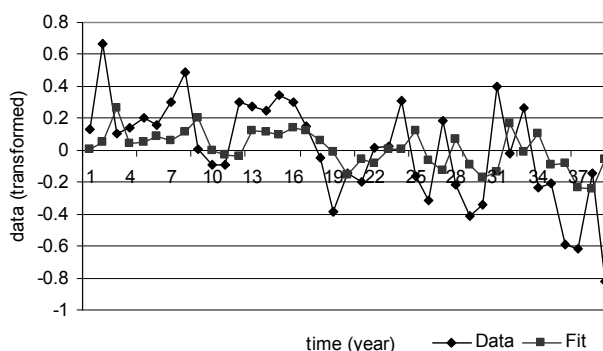


Fig.3. Model AR(1) pentru seria Sulina

Al doilea este de tip FARIMA, descrie mai bine evoluția seriei, fiind folosit și pentru predicție (Fig.4)

Sub-seria Sulina_1 (1965 - 1981) este Gaussiană, independentă și identic distribuită, cu varianța 3795.05.

Sub-seria Sulina_2 (1982 - 2003), după extragerea mediei, este zgomot gaussian, cu varianța 4296.75.

2.5. Determinarea unor modele alternative (GEP, ARIMA-AdaGEP)

Deoarece în general seriile meteorologice prezintă variabilitate mare și sunt cazuri în care evoluția neliniară nu poate fi captată prin modelele de tip ARIMA sau FARIMA, o idee a fost obținerea unor modele în care s-au combinat metodologia Box-Jenkins cu Gene Expression Programming, obținând modele hibride, care îmbunătățesc modelele obținute printr-un algoritm genetic adaptiv, elaborat de noi. (Fig.5)

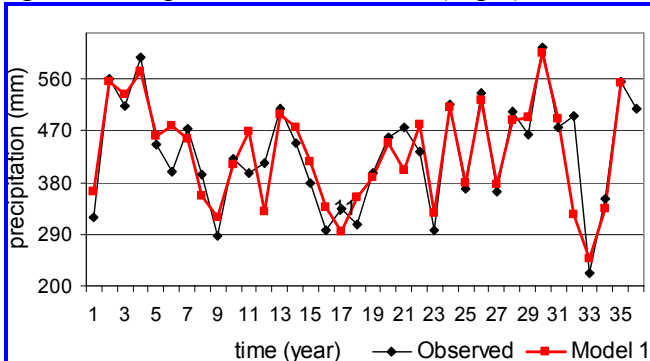


Fig.5. Model AdaGEP-AR pentru seria anuală Medgidia

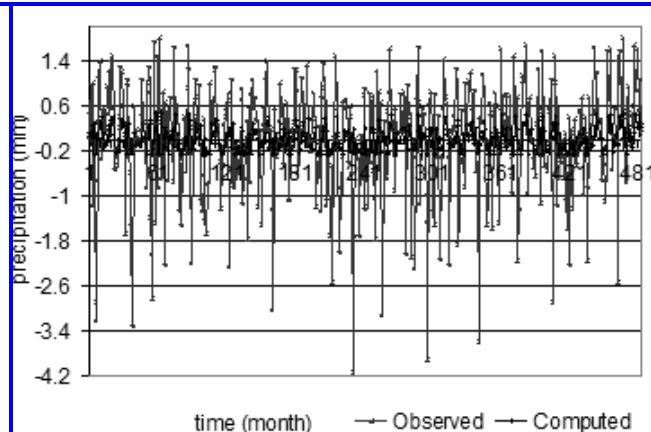


Fig.6. Model AdaGEP pentru seria lunară Sulina

Prezentăm cazul seriei lunare Sulina, care are două puncte de ruptură și pentru care un model obținut prin AdaGEP nu este satisfăcător. (Fig.6)

După extragerea mediei, cel mai bun model al seriei integrale a fost un model ARMA(2, 2):

$$Y_t = 0.9577Y_{t-1} - 0.9915Y_{t-2} + Z_t - 0.929Z_{t-1} + 0.9914Z_{t-2},$$
 cu (Z_t) zgomot alb de varianță 0.9517 (Fig.7).

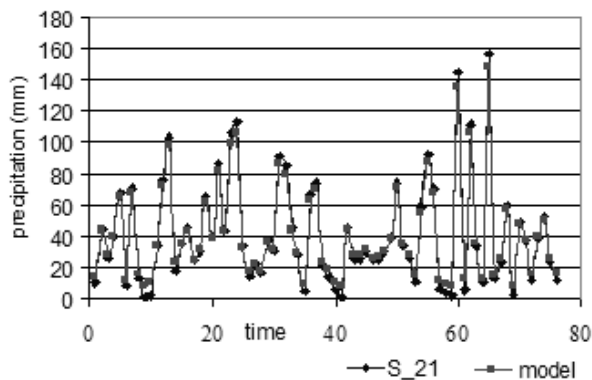


Fig.8. Model MA(4) pentru S1

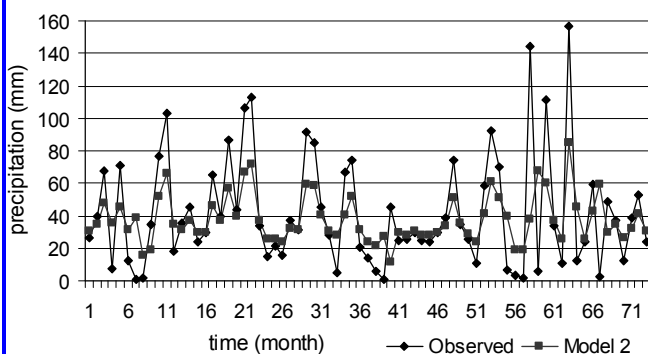


Fig.9. Model combinat pentru S1

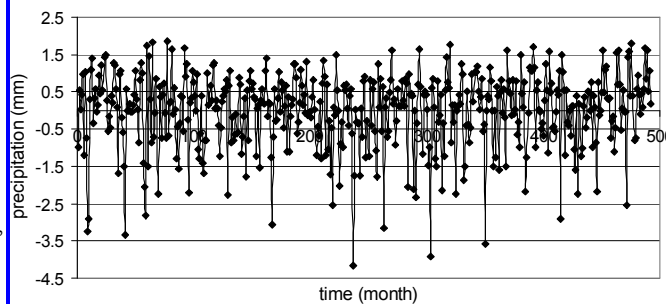


Fig.7. Model ARMA(2,2) al seriei lunare Sulina

Pentru subseriile (S1, S2, S3), delimitate de punctele de ruptură, s-au obținut respectiv modelele:

- pentru S1 (MA(4)):

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

cu $(\varepsilon_t)_{t \in \overline{1, 76}}$ zgomot alb (Fig.8)

În Fig.9 se găsește modelul combinat.

- pentru S2 nu s-a găsit un model satisfăcător de tip ARIMA, cel mai bun fiind de tip AdaGEP (Fig.10)

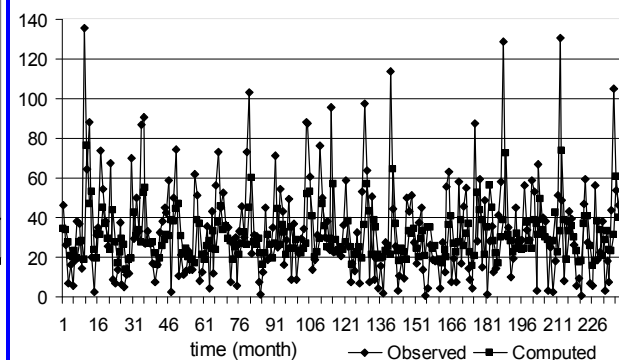


Fig.10. Model AdaGEP pentru S2

- Pentru S3, cel mai bun model a fost cel combinat, AR_GEP(Fig.11)

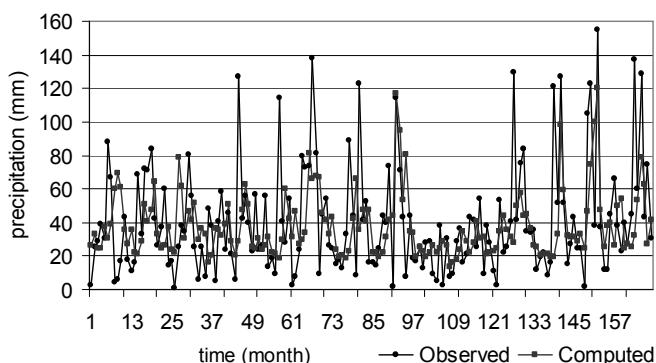


Fig.11. Model combinat al S3

4. CONCLUZII

➤ Analiza datelor a constatat din teste de omogenitate, normalitate, necorelație și homoscedasticitate, ruptură a seriilor de timp. Dintre seriile anuale analizate, patru au fost zgomote gaussiene, una a fost zgomot Gaussian, după o transformare, iar una a prezentat proprietate de LRD.

➤ Pentru calculul coeficientului Hurst s-a elaborat un program de calcul.

- Pentru seriile care nu au fost zgomote Gaussiene s-au determinat modele autoregresive, respectiv FARIMA (pentru seria Sulina).
- Selecția modelelor a fost făcută conform criteriului lui Akaike. Reziduurile au fost totdeauna zgomote Gaussiene.
- S-au propus și modele hibride (combinație a unui algoritm genetic adaptiv - AdaGEP și ARIMA) pentru îmbunătățirea modelelor obținute.
- A fost făcută predicția evoluției precipitațiilor, pe baza modelelor determinate.

5. BIBLIOGRAFIE

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, Iași, 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