

## 1. DATE ȘI ANALIZA PRELIMINARĂ

S-a dispus de date privind precipitațiile lunare pe o perioadă de 41 ani (1965 - 2005), de la 10 stații principale din Dobrogea.

Analiza datelor a constatat din teste de omogenitate, normalitate, corelație și homoscedasticitate, ruptură a seriilor lunare.

După analiza seriilor de date concluziile au fost următoarele:

a. Ipoteza nulă că seriile sunt aleatoare a fost acceptată la nivelul de încredere de 99%, exceptând stația Jurilovca.

b. Ipoteza că seriile sunt normal distribuite a fost respinsă, dar ipoteza normalității poate fi acceptată pentru toate seriile, exceptând Constanța, după aplicarea unor transformări Box-Cox, cu  $\lambda \in [0.34, 0.42]$ .

c. Rezultatele testelor de ruptură sunt diferite, funcție de testele aplicate.

d. Ipoteza omogenității seriilor este respinsă după aplicarea testului Wilcoxon.

e. În majoritatea seriilor există valori aberante, care au fost eliminate înaintea determinării modelelor matematice.

## 2. REZULTATE

În cele ce urmează prezentăm comparativ rezultatele obținute pentru stațiile Medgidia și Sulina, acestea fiind reprezentative pentru evoluția precipitațiilor lunare în perioada analizată.

### 2.1. Modele multicomponente

Seria **Medgidia** este reprezentată în Fig.1. Deoarece seria are două puncte de ruptură, analiza a fost făcută pe sub – serii. Valorile înregistrate în iunie 1971 și august 1991 (165.4 mm, respectiv 150 mm) sunt tratate ca valori aberante și vor fi eliminate înaintea elaborării modelului.

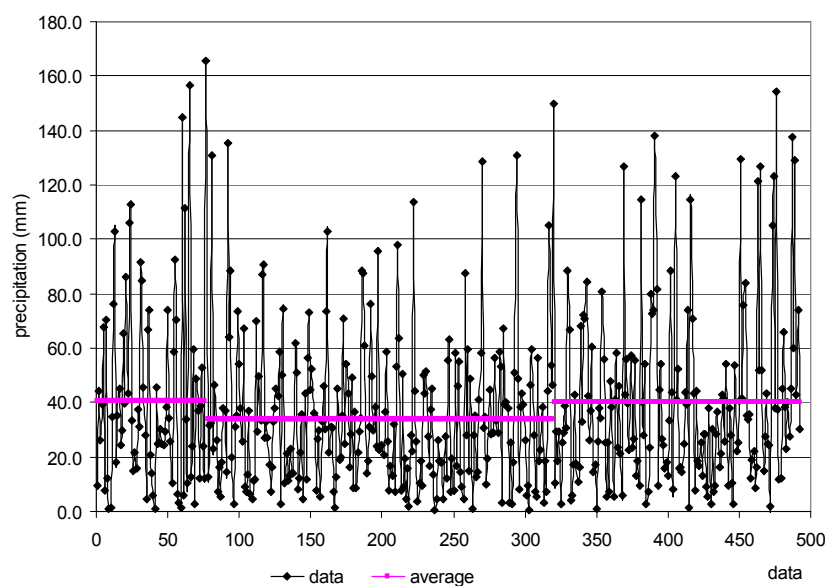


Fig. 1. Seria de precipitații Medgidia

#### I. Ianuarie 1965 - mai 1971

După aplicarea unei transformări Box-Cox cu  $\lambda=0.42$ , seriei inițiale, pentru seria rezultată,  $(X_t)$ , s-a determinat un model:

$$X_t = Z_t - 0.2242Z_{t-4}, t \in \overline{5, 76}, \quad (1)$$

unde  $(Z_t)_{t \in \overline{1, 76}}$  este un zgomot alb.

#### II. Iulie 1971 - iulie 1991

Graficul acestei sub - serii evidențiază existența unei componente sezoniere . Ca urmare, seria a fost descompusă după modelul multiplicativ:

$$X_t = Y_t \cdot S_t \cdot \varepsilon_t, \quad (2)$$

unde  $Y_t$  este trendul,  $S_t$  - componenta sezonieră și  $\varepsilon_t$  - reziduul, care este un zgomot alb.

Trendul s-a determinat prin metoda mediei mobile. Din analiza factorilor de sezonaliitate (Tabel 1) rezultă ca cele mai afectate de fluctuațiile sezoniere au fost lunile ianuarie și februarie, iar cel mai puțin, octombrie.

### III. Septembrie 1991 - decembrie 2005

Analiza ACF și PACF ale acestei sub-serii relevă prezența unor valori ridicate la decalajele 11, 22, 33, deci nu este vorba de o sezonaliitate tipică (de 12 luni). Ca urmare, s-a utilizat o diferențiere de ordin 11, modelul final propus, după înlăturarea coeficienților nesemnificativi și aplicarea criteriului verosimilității maxime, fiind de tip medie mobilă, de ordin 11:

$$X_t = Z_t - 0.6399Z_{t-11}, t \geq 12, \quad (3)$$

unde  $(Z_t)_{t \geq 1}$  este un zgomot alb.

Seriile ajustate, împreună cu datele inițiale și prognoza pe următoarele 12 luni, realizată prin aplicarea metodei Winters sunt reprezentate în Fig.2. Se observă că modelul dă rezultate bune, iar prognoza pe 2006 corespunde valorilor înregistrate în acea perioadă.

Tabel 1. Factori sezonaliitate

luna	factor sezonali (%)
1	182.2
2	144.9
3	114.1
4	75.6
5	83.3
6	102.2
7	74.2
8	64.0
9	68.0
10	62.4
11	104.9
12	124.2

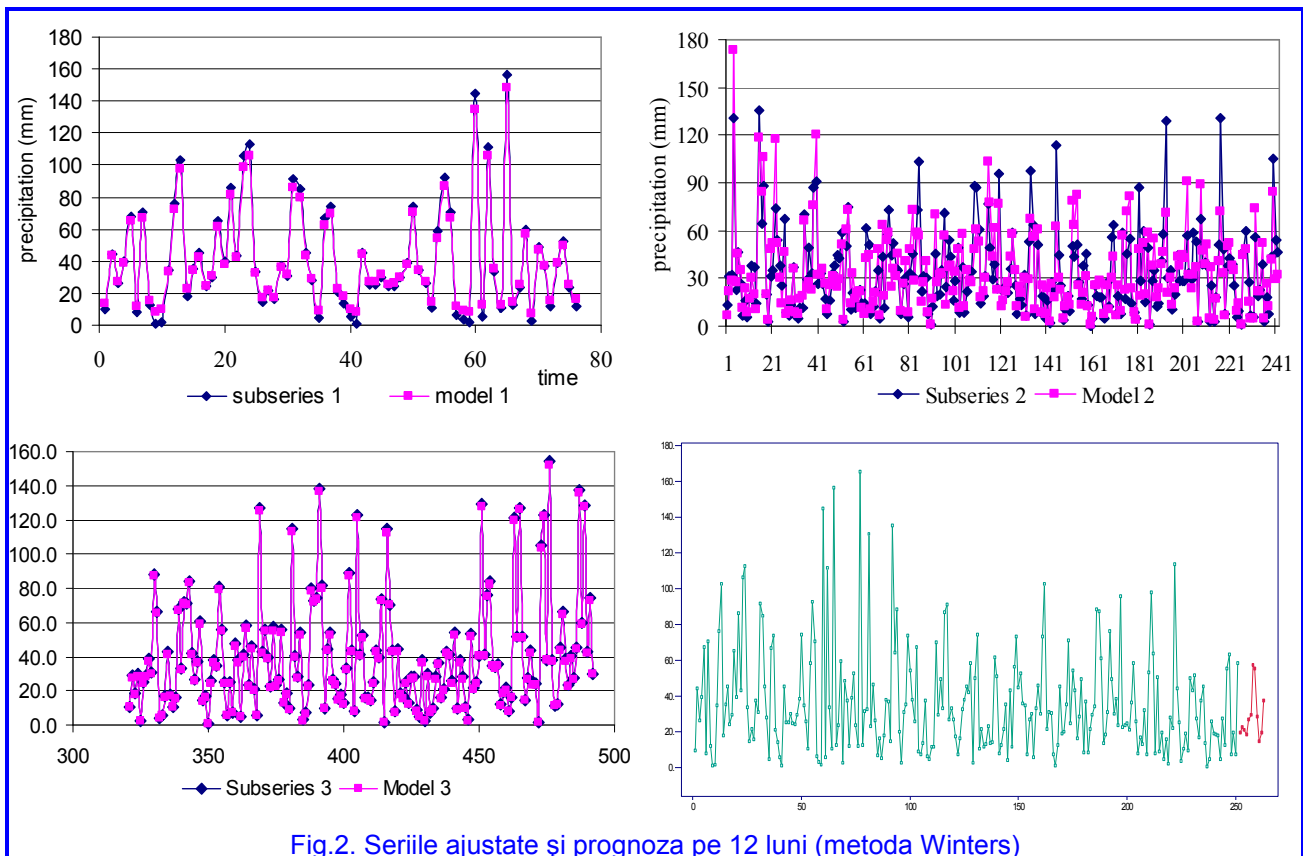


Fig.2. Seriile ajustate și prognoza pe 12 luni (metoda Winters)

Pentru seria **Sulina**, după aplicarea testului Kolmogorov - Smirnov, ipoteza că seria este normal distribuită a fost respinsă. După aplicarea aceluiași test seriei obținute după o transformare Box-Cox, cu parametrul  $\lambda = 0.34$ ,  $(Z_t)$ , se acceptă ipoteza că seria noua serie este normal distribuită. După analiza corelogramei se respinge ipoteza că seria este necorelată, la nivelul de încredere 95%. Ipoteza homoscedasticității a fost respinsă după aplicarea testului Bartlett.

I. Ianuarie 1965 – august 1982: sub – seria  $(Z_t)$  formează un zgomot alb.

II. Septembrie 1982 – decembrie 2005

Descompunerea sub – seriei  $(Z_t)$  s-a realizat după un model multiplicativ, procedându-se astfel:

- S-au determinat indicii de sezonaliitate ( $S_t$ ) prin metoda mediei aritmetice (Tabel 2);
- S-a eliminat componenta sezonieră, prin raportarea valorilor seriei la indicii de sezonaliitate corespunzători;
- S-a determinat trendul:  $\bar{Z}_t = 4.2768 + 0.6475 * \cos(0.05348t - 2.9331)$  (Fig.3);
- S-au determinat reziduurile. Varianța reziduală a fost 0.326515, iar testele făcute conduc la acceptarea ipotezei că erorile formează un zgomot gaussian.

Tabel 6. Indici de sezonaliitate

ian.	febr.	martie	aprilie	mai	iunie	iulie	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

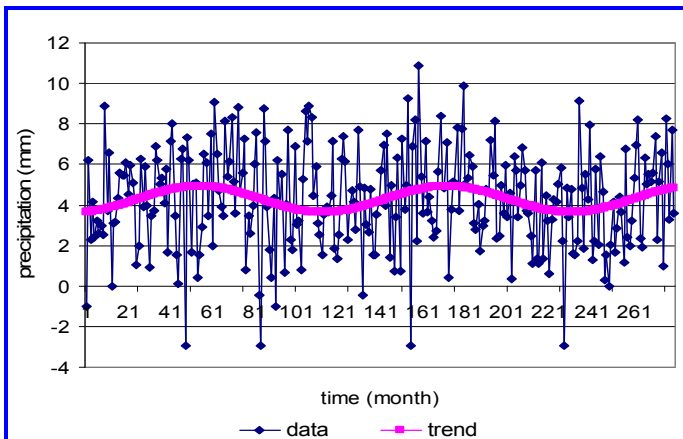


Fig. 3. Datele și trendul pentru sub-seria 2

Predicția evoluției precipitațiilor a fost făcută prin antrenarea unei rețele neuronale.

Etapele pentru realizarea predicției folosind rețelele neuronale au fost:

- *Design*: definirea numărului de noduri din fiecare layer (input, hidden, output) și a funcțiilor de transfer (*logsig*, *tansig*, *purelin*);
- *Training*, în scopul optimizării funcției eroare;
- *Testare*, pentru a analiza dacă rețeaua a găsit un echilibru între acuratețe și generalizare.

Datele din fiecare sub-serie au fost împărțite în 3 părți: o mulțime de training, una de validare și una de testare. S-a folosit o rețea „two-layer feed forward” cu un layer de output și layer ascunde având între 5 - 10 neuroni *tansig* sau *purelin* și output - ul un neuron *tansig*. Ca rutină de training s-a folosit *trainlm* din Matlab.

Rezultatele predicției apar promițătoare (Fig.5).

## 2.2. Modele frecvențiale

Ajustarea modelelor frecvențiale pentru seriile de date a fost făcută utilizând legea lui Gumbel:  $F(x) = \exp[-\exp(-(x-a)/b)]$ .

Cu variabila redusă  $u = (x-a)/b$ , distribuția se scrie ca:  $F(u) = \exp[-\exp(-u)]$  și  $u(F) = -\ln(-\ln(-F(x)))$ .

Avantajul folosirii variabilei reduse este acela că expresia quantilei este liniară:  $x_q = a + bu_q$ .

Etapele ajustării sunt următoarele:

1. Pregătirea datelor pentru trasarea poligonului frecvențelor cumulate:
  - trierea valorilor în ordine crescătoare și atașarea unui rang,  $r$ ,
  - calculul frecvențelor empirice, cu formula Hazen:

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

unde  $r$  este rangul, iar  $n$  – numărul de valori.

2. Calculul variabilei reduse Gumbel,  $u$ ;

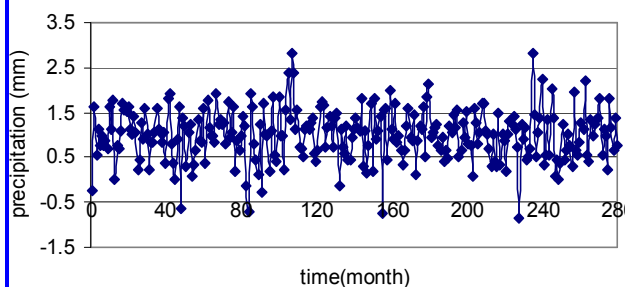


Fig. 4. Residuul în modelul multiplicativ pentru sub-seria 2

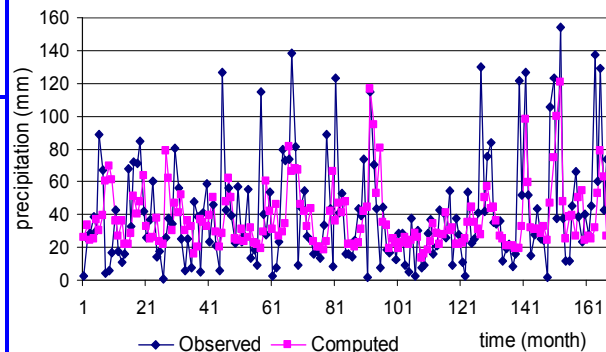


Fig. 5. Predicție pe 170 luni folosind rețele neuronale pentru seria Sulina

3. Reprezentarea grafică a perechilor  $(u_i, x_i)$  și determinarea parametrilor  $a$  și  $b$  ai dreptei de regresie, prin metoda momentelor sau a celor mai mici pătrate;

4. Pentru timpi de întoarcere diferiți,  $T$ , calculul frecvenței  $F = 1 - 1/T$  ;
5. Calculul variabilei reduse  $u(F) = -\ln(-\ln(-F(x)))$ ;
6. Calculul valorii estimate a quantilei  $x_q$  .

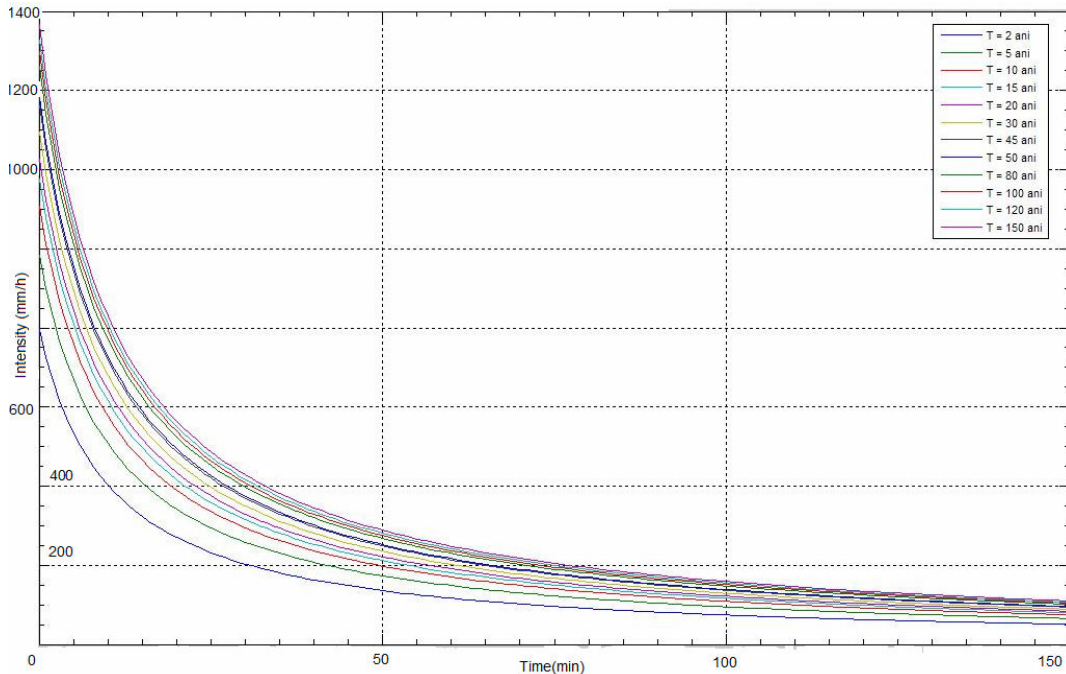


Fig.6. Curbe IDF pentru stația Medgidia

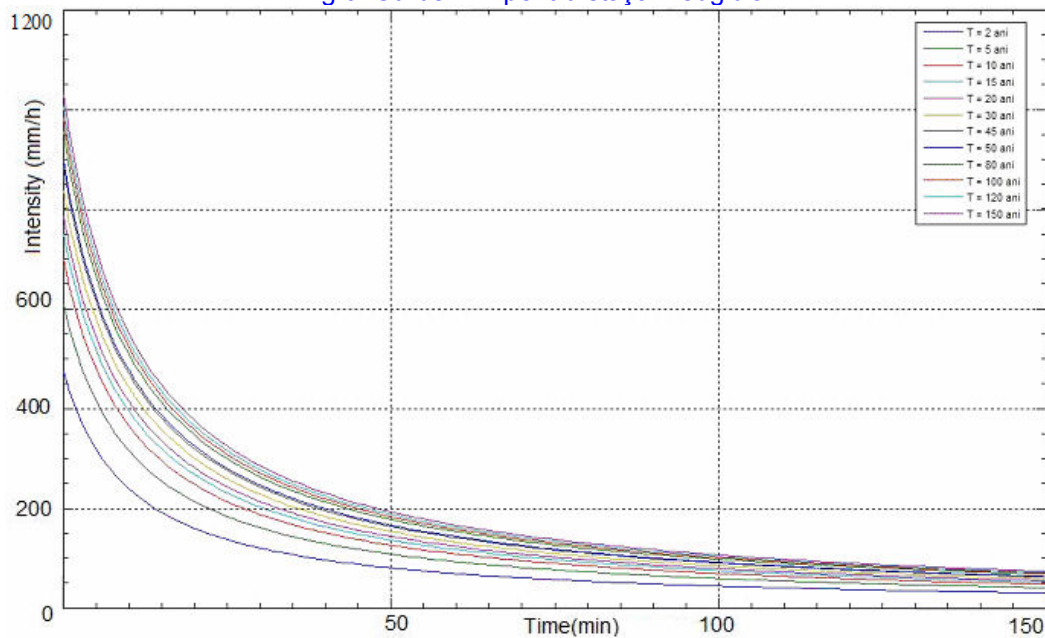


Fig.7. Curbe IDF pentru stația Sulina

### 3. Concluzii

Analizele făcute privind evoluțiile precipitațiilor lor în perioada 1965 – 2005 pe seriile de date medii lunare arată că acestea nu au înregistrat variații foarte mari de la o stație la alta. Astfel, punctele de ruptură sunt în majoritatea cazurilor prezente în anii 1970 - 1971 și 2003 – 2004. Nici una dintre serii nu este normal distribuită, însă ipoteza normalității poate fi acceptată după transformări Box-Cox, cu excepția seriei Constanța.

Având în vedere existența unor date lunare era de așteptat ca seriile să prezinte sezonabilitate. Însă, în majoritatea cazurilor periodicitatea nu este 12 pentru seria integrală, ci numai pe unele sub-serii, altele prezentând periodicitatea 11. Ca urmare, descompunerea în trend, sezonabilitate și componentă aleatoare nu poate fi aplicată cu rezultate satisfăcătoare pe seriile integrale.

Metoda de predicție de tip Holt nu au dat rezultate seriile prezentând trend. Metoda Winters au dat rezultate bune pentru predicția pe termen scurt. Utilizarea rețelelor neuronale apare promițătoare în ceea ce privește prognoza pe termen îndelungat.

Analiza curbelor IDF pentru timpi de întoarcere diferiți au dus la concluzia că intensitatea maximă a precipitațiilor este cuprinsă între 400 și 1600 mm/h, fiind mai ridicată în interiorul regiunii decât pe litoral.

În toate situațiile au fost analizate și reziduurile și s-a demonstrat că acestea sunt zgomote gaussiene.

#### 4. Bibliografie

1. H. Akaike, *Information theory and an extension of the maximum likelihood principle*, 2<sup>nd</sup> 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. Pelican, *Modeling the evolution of precipitation in a region of Romania*, 7th World Congress in Probability and Statistics – Singapore
4. A. Bărbulescu, C. Maftei, *Modeling the seasonal evolution of precipitations in Dobrudja*, EGU, Viena, 14-18.04.2008
5. A. Bărbulescu, E. Pelican, *On the Sulina Precipitation Data Analysis using the ARMA and a neural network technique*, MACMESE 08,
6. P.J. Brockwell, R.A. Davis, *Time series analysis, forecasting and control*, Holden - Day, San Francisco, 1976
7. 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.
8. 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.
9. C. Maftei, A. Bărbulescu, *Statistical analysis of climate evolution in Dobrudja region* Lecture Notes in Engineering and Computer Sciences, vol.II, p. 1082 – 1087
10. 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