STUDIA UBB GEOGRAPHIA, LVIII, 2, 2013, pp. 49 - 58
(RECOMMENDED CITATION)

STATISTICAL ANALYSIS OF HYDROLOGICAL DATASETS TO DETERMINE LONG-TERM FORECAST

ŞT. BILAŞCO[1], CS. HORVÁTH[2]

ABSTRACT. – Statistical Analysis of Hydrological Datasets to Determine Long-Term Forecast. Hydrological forecasting takes various forms, from the calculation of certain runoff probabilities to statistical analysis of datasets recorded at gauging stations. If the first method of forecast refers only to punctual events, floods, inundations, making it useful for hydropower and watershed management facilities design, the statistical method allows longer forecasts by analyzing the measured datasets. For the statistical analysis of the hydrological data, we implemented the Thomas-Fiering model which is usually used for hydrological applications. The recorded monthly average runoff data was selected from Cluj and Răcătău gauging stations on a 52 year period. The model was used to generate synthetic values at monthly scale, during the 1950 - 2002 period, for 20 years between 1992 and 2012 it forecasted values and it was validated through the 10 years between 1992 and 2002.

Keywords: hydrological forecasting, statistical method, Thomas-Fiering model

1. INTRODUCTION

The statistical modeling of hydrological datasets can be classified into two categories, monthly runoff estimation models using monthly measured discharges, and monthly scale forecast models which use annual values. The practical utility of both types of models lies in accurately estimating the water resources of an area.

Methods to estimate monthly and seasonal scale runoff were used since the beginning of hydrological datasets studies; Hazen in 1914 generated a forecast for 300 years combining values from 14 river basins in a single dataset. Models to generate data sequences were made by other researchers also, among which we mention, Barnes in 1954, Sudler (1927), Haidu (1995, 1997) and Haidu & Linc Ribana (2001).

The Thomas-Fiering model fits into the first category, that of estimating the average monthly runoff values by emphasizing the seasonality of discharge in the profile of the analyzed gauging station.

---

[1] *Romanian Academy, Cluj-Napoca Subsidiary, Geography Section, 400015, Cluj-Napoca, Romania, e-mail:sbilasco@geografie.ubbcluj.ro*

[2] *Babeş-Bolyai University, Faculty of Geography, 400006, Cluj-Napoca, Romania, e-mail: csaba.horvath@geografie.ubbcluj.ro*

The model was created by Thomas and Fiering in 1962 mainly to generate modeled values at monthly scale and secondly to generate forecast. Considering the seasonal component, which is associated to the succession of seasons, in the scientific literature it is stated (Markham, 1970, Pop & Horvath, 2009, Th. Petersen *et al*., 2012) that hydrological data has a strong seasonal character, therefore the model assumes there is a statistical link between the seasonal observations and monthly runoff data in the same year, and also a link between the same observations in successive years.

2. STUDY AREA

The study area represents the upper basin of the Someşul Mic catchment which is part of the Someş – Tisa Water Branch, collecting waters from the eastern part of the Apuseni Mountains.

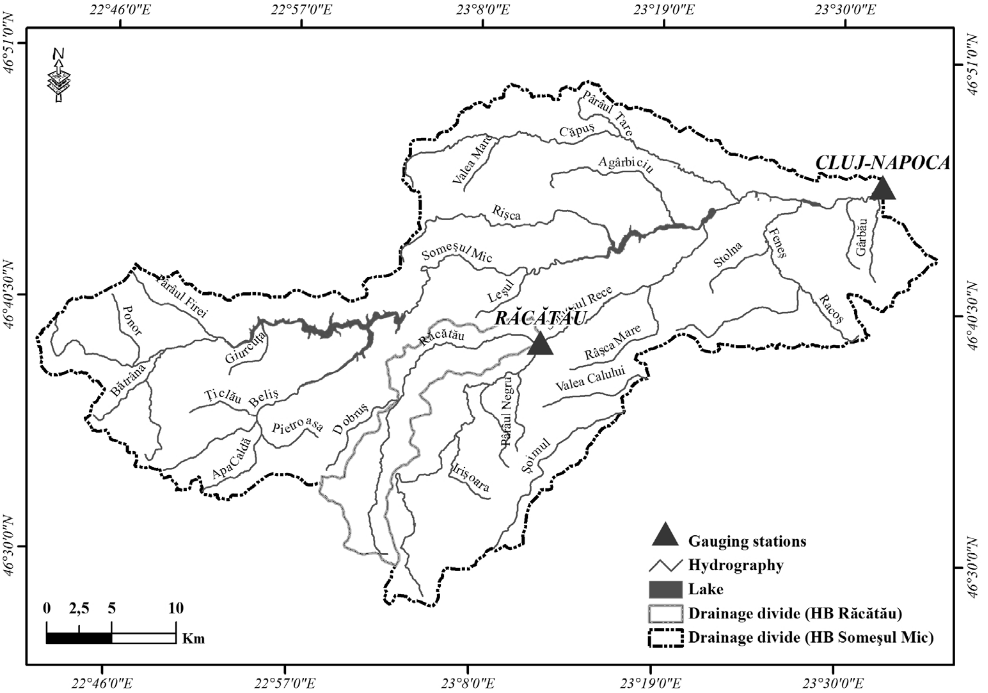

Fig. 1. Study area and gauging stations

The considered gauging stations for implementing the Thomas-Fiering model are represented by Răcătău and Cluj-Napoca stations, both with significant observation periods (fig. 1).

3. METHODOLOGY AND DISCUSSIONS

The Thomas-Fiering model uses a regression equation based on discharges measured in successive time intervals, consequently the discharge values measured in June are analyzed depending on the values measured in May and the computations are made using a linear regression function.

The model is executed using two series of hydrological datasets, Răcătău hydrometrical station with observed discharge values and a catchment area of 102 km^2 and Cluj-Napoca with reconstructed discharge and a catchment area of 1194 km^2, both represented by monthly average discharge data for a 52 years period, between 1950 and 2002.

Based on the observed data we were able to compute a series of descriptive statistical values (table 1) for Răcătău and Cluj-Napoca stations.

From the analysis of the statistical data we can observe the seasonality of the runoff with a maximum discharge value in spring presented as the maximal percentage of the annual total flow and a minimum in winter and autumn (table 1).

Statistical data for Cluj-Napoca and Răcătău stations

Table 1

Season	Winter		Spring		Summer		Autumn	
Gauging station	Răcătău	Cluj-Napoca	Răcătău	Cluj-Napoca	Răcătău	Cluj-Napoca	Răcătău	Cluj-Napoca
Average	1.087	8.362	3.441	25.324	2.441	16.858	1.263	9.030
Percent	13.21%	14.4%	41.79%	42.51%	29.65%	28.30%	15.35%	15.16%
Maximum discharge (m^3/s)	4.383	21.073	7.557	50.533	5.807	36.566	4.350	23.333
Maximal year	1996	1996	1977	2000	1979	1975	1979	1998
Minimal discharge (m^3/s)	0.307	2.143	1.243	8.268	0.816	5.937	0.396	2.311
Minimal year	1985	1954	1961	1961	1981	1952	1961	1961
Seasonal Variation Coefficient	0.599	0.475	0.358	0.343	0.446	0.429	0.583	0.533
Seasonal Asymmetry Coefficient	25.954	5.053	1.809	1.575	3.812	4.400	15.848	10.808

Using the two hydrological datasets, to complete the model we used the following equations:

$$X_{j+1,i} = \overline{X}_{j+1} + B_j\left(X_{j,i} - \overline{X}_j\right) + t_i S_{j+1}\left(1 - R_j^{\,2}\right)^{1/2} \qquad (1)$$

where:

$X_{j+1,i}$, $X_{j,i}$ – generated monthly runoff for month j+1 during year i and month j during year i, respectively

$\overline{X}_{j+1}, \overline{X}_j$ - mean monthly historic runoff record for the (j+1)th and jth months

t_i – normal random variant with mean of zero and variance of unity

B_j - regression coefficient between runoff in (j)th and (j+1)th months

$$B_j = R_j\left(S_{j+1} / S_j\right) \qquad (2)$$

S_{j+1} - standard deviation of the historic stream flow record for month j+1

R_j - serial correlation coefficient between flows in (j)th and (j+1)th months

To complete the model, the formula from equation (1) was implemented into MS Excel resulting in a macro that incorporates all the mathematical equations and formulas and also the graphical analysis and comparison of the datasets which are necessary.

The arithmetic mean for a given month (j), expresses an average result with seasonal character (1), because for the monthly values there are 12 averages for the 12 corresponding months.

The first step in completing the model is the fitting of runoff data based on the measured discharge values and the comparison of the two datasets to monitor the homogeneity of the datasets.

Visual analysis of the two figures reveals a pronounced seasonal character of runoff for the two analyzed gauging stations. Seasonality is characterized by high runoff in spring; with clearly higher recorded and simulated values at both hydrometric stations, Cluj-Napoca (fig 2 a, b) and Răcătău (fig. 3 a, b) and also, by the absolute low in winter. It should be noted that at Cluj-Napoca hydrometric station a maximum during the summer of 1974 (June-July) was measured and also caught by simulation (fig. 2 a, b). According to the 2010-2013 Someș - Tisa Water Branch Catchment Plan, the runoff that was recorded during that period was due to rainfall that exceeded significantly the average multiannual precipitation.

The comparative analyses of observed-simulated datasets highlight a very good correlation for both stations and also an exact capture of measured peaks. The extreme events are captured very well as time interval in the simulated sequence but significantly minimized in value. Most striking examples are the peaks measured at Cluj-Napoca station in May 1970, with 68.39 m^3/s correlated with a simulated 38. 1 m^3/s discharge, also in July 1974 with 57.8 m^3/s measured value and only 37.9 m^3/s modeled (fig. 2 a, b) and at Răcătău station in May 1977 with 12.70 m^3/s measured and 5.71 m^3/s (fig. 3. A b) simulated. All examples are extreme values, representing peak values in both data series, in the measured and in the simulated runoff generated by the model.

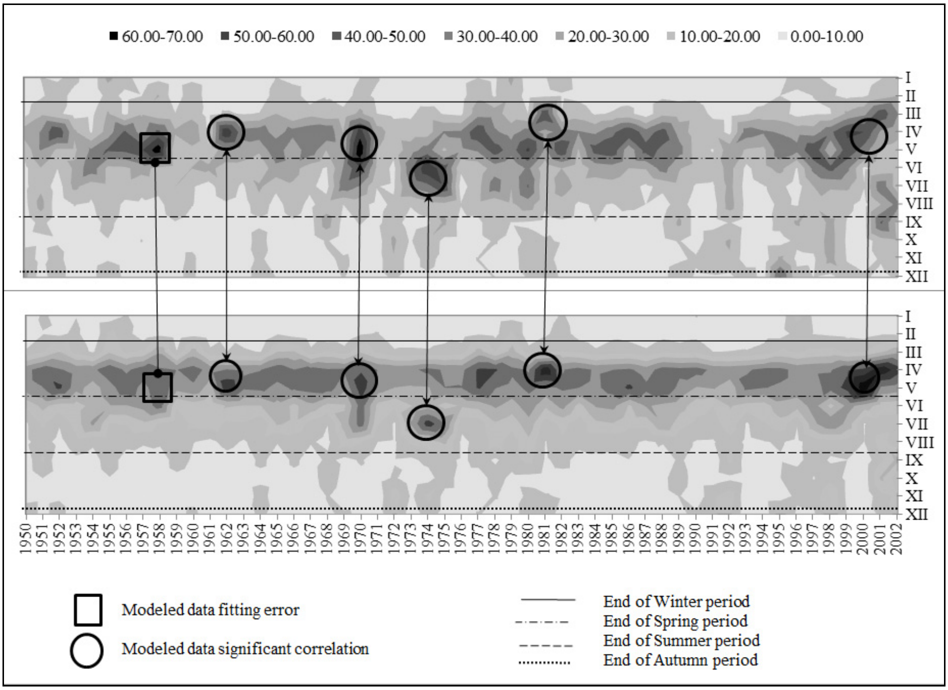

Fig. 2. Correlation between the observed and simulated runoff - Cluj-Napoca

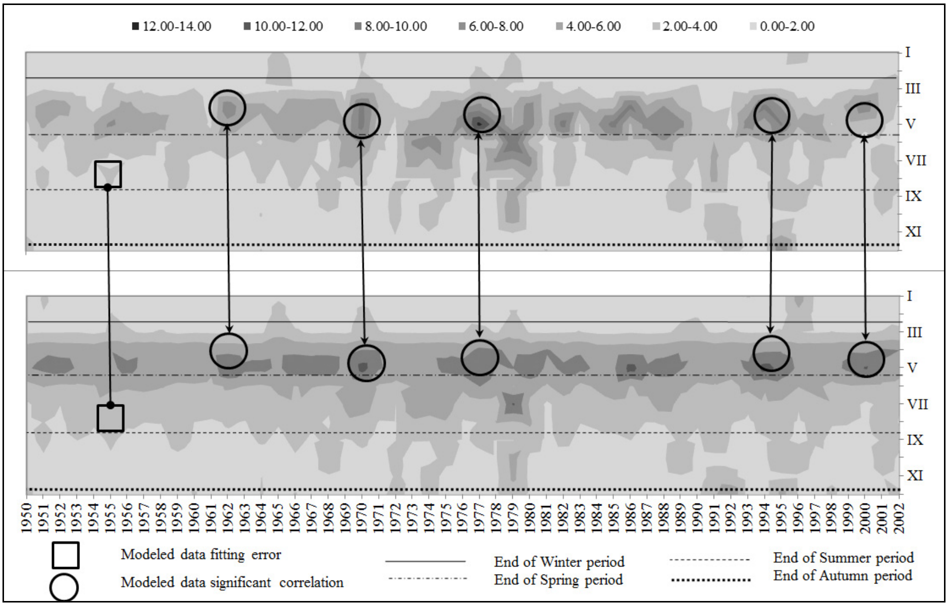

Fig. 3. Correlation between the observed and simulated runoff - Răcătău

Comparing both visually and statistically the two modeled and simulated datasets, we observe the underestimation of measured discharge values by the model.

Preliminary validation was done through percentage difference between the measured and modeled values. Very high differences are recorded, 55.7% for Cluj-Napoca station data series and 44.96% at Răcătău station, but also there are values equal to 0 (perfect fitting) and negative values which suggest simulated discharges higher than the measured ones. Overall average values enroll in the margin of error accepted for completion of hydrological studies, 8.6% at Cluj-Napoca and 5.2% at Răcătău.

Observing the two graphs, we can state that there is a good correlation between the two observed and simulated series. For both considered series, we analyzed the correlation and the residuals to validate the model.

The correlation analysis was performed using the graphical method which involves the graphical representation of the j and j +1 pairs establishing also the shape and intensity of the correlation.

The value of R_j, (3) has a double meaning. The formula expresses the correlation between the monthly datasets, respectively between the values of May and June, values between June and July, etc. In addition, the 12 R_j values express also the phenomenon of autocorrelation between x_j string (current month) and the x_{j-1} string (previous month).

$$R_j = \frac{\sum_{i=1}^{N-1}\left(X_{j,i} - \overline{X}_j\right)\left(X_{j+1,i} - \overline{X}_{j+1}\right)}{\sqrt{\sum_{i=1}^{N-1}\left(X_{j,i} - \overline{X}_j\right)^2\left(X_{j+1,i} - \overline{X}_{j+1}\right)^2}} \tag{3}$$

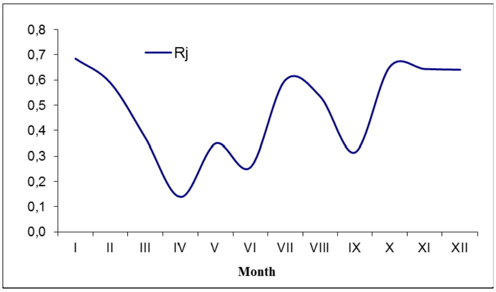

Fig. 4. Correlation at Răcătău

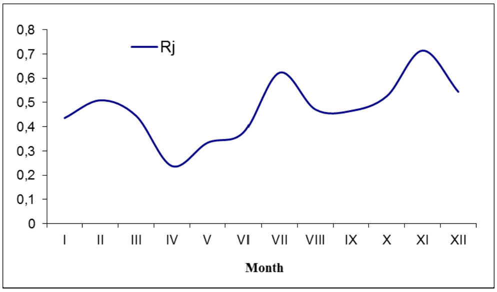

Fig. 5. Correlation at Cluj-Napoca

From the analysis of the correlation graphs one can observe a very good correlation (a strong relationship between j and j +1) at both considered gauging stations. About the same types of correlations are established for both gauging stations, better highlighted for Răcătău hydrometric station dataset (Fig.4) and less in the case of Cluj-Napoca (fig. 5). Positive correlations are determined between April-May, June-July and September-October at both hydrometric stations, and negatively correlated cases are found in February-April and August-September.

The residuals of the model after Clark (1973) are represented in the following graphics:

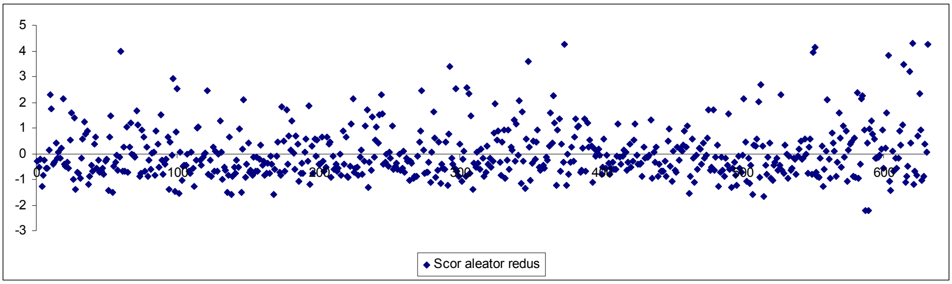

Fig. 6. Residuals of the model at Cluj-Napoca

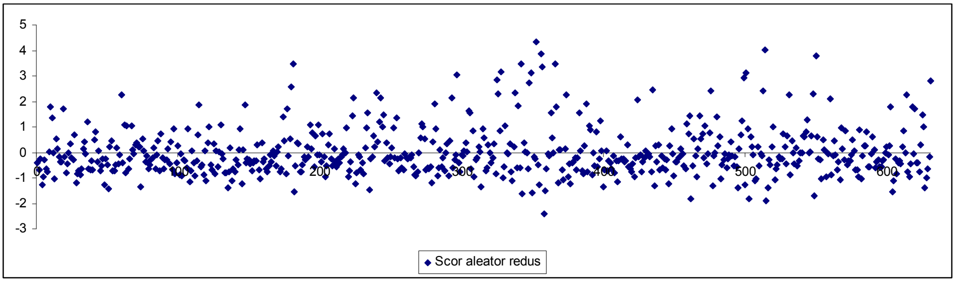

Fig. 7. Residuals of the model at Răcătău

Analyzing the 2nd and 3rd graphs one can follow the seasonality of the model parameters and also a very good correlation between the observed and simulated runoff. At the same time the 4th and 5th figure reveal a good correlation between the model parameters. Also, the positive and negative values of the model residuals presented in figures 6 and 7 with an approximately normal distribution equaling zero, supports the validation and use of the Thomas-Fiering model in researching and forecasting runoff for the two studied gauging stations and can be extrapolated to all the small catchments tributary to the Someșul Mic River upper basin.

4. RESULTS AND CONCLUSIONS

Based on the Thomas-Fiering model we forecasted the runoff at the two gauging stations at Cluj-Napoca (fig. 8) and Răcătău (fig. 9) for a period of 10 years. The comparison period of the observed discharge with the forecasted runoff is between the years 1992-2012, also the 1992 and 2002 period is for validation interval and 2002-2012 is for forecasting.

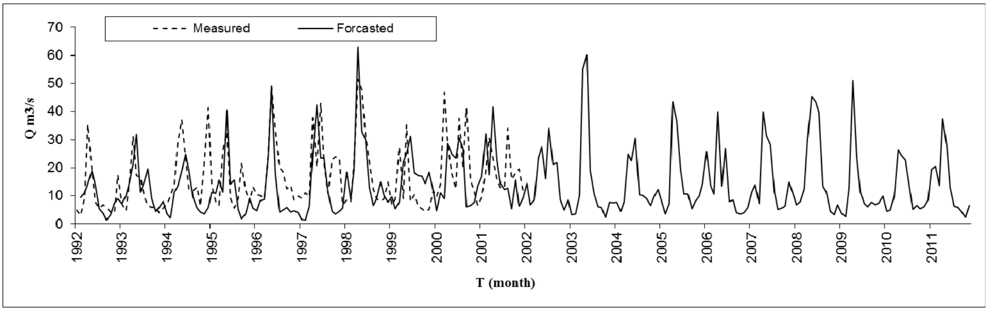

Fig. 8. Forecasted discharge – Cluj-Napoca

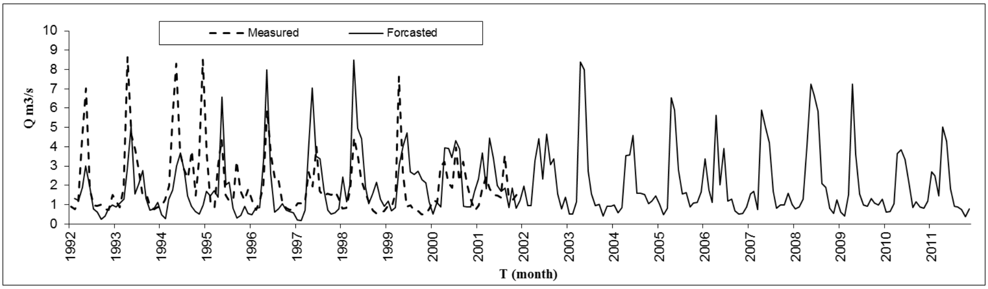

Fig. 9. Forecasted discharge – *Răcătău*

From the analysis of the forecasted runoff at Cluj-Napoca gauging station, through the Thomas-Fiering model, it can be observed that in April 2004 the forecasted (fourth month of the year 2004) maximum discharge of 30.38 m^3/s was preceded by a period of smaller runoff and also followed by a period when the discharge was declining.

Monthly average runoff at Cluj-Napoca gauging station

Table 2.

Year 2004	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
Q m^3/s measured	6,28	9,81	21,1	55,9	22,4	10,1	9,99	16,3	13,0	10,8	13,2	19,3
Q m^3/s forecasted	7.97	24.92	22.47	30.38	10.5	10.26	9.04	6.56	9.74	12.17	7.61	3.5

The forecasted data were compared with the recorded data (table 2) at the same gauging station, so we can point out that in April 2004 we measured a discharge of 55.9 m^3/s, which is comparable to the predicted values. Also, this is the highest recorded discharge during the 2004 year. Also one can observe the seasonality of the runoff from the forecasted discharge datasets presented as example the 2005 data (table 3), with high runoff values in spring and low discharge values at the end of summer and beginning of autumn.

Forecasted discharge seasonality (Cluj-Napoca 2005)

Table 3

Season	**Winter**		**Spring**			**Summer**			**Autumn**			
Mounth	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Q m^3/s	7.61	3.5	7.24	43.5	36.6	19.2	10.7	10.7	5.51	7.99	9.98	14.5

The same situation can be observed in the forecast of Răcătău gauging station. In terms of comparison between the measured and predicted runoff, also in capturing the seasonality of the forecasted flow for the same period as Cluj-Napoca gauging station, the forecasted discharge values are obviously smaller. From the above stated and the datasets analysis, the forecasted discharge for the two gauging stations presents a good correlation, so the runoff oscillations are traced from the upstream station to the downstream station.

Analyzing the results of the model and taking into account the conclusions we can say that the model can be used to generate runoff values and stochastic long-term prognosis for small catchments within the upper basin of the Someșul Mic River catchment.

REFERENCES

1. Barnes, F.B. (1954), *Storage required for a city water supply*. Inst. Engs _Aust 26: 198.
2. Markham, C.G. (1970), *Seasonality of Precipitation in the United States*, Annals of the Association of American Geographers, Vol. 60, Iss. 3.
3. Clarke, R.T. (1973), *Mathematical Models in Hydrology*, FAO Irrigation and Drainage Paper, No 19.
4. Haidu I. (1995), *Testing Thomas-Fiering Models and Derived Versions on Geophysical Time - Series.* In Annales Geophysicae, Suppl. II to vol 13., Hamburg 3-7 Apr.1995, C 588.
5. Haidu I. (1997) *Analiza seriilor de timp. Aplicații în hidrologie*. Editura *H*G*A*, Bucureşti.
6. Haidu, I., Haidu, C., Raicu, T. (1995), *Water Transfer Analysis – Preliminary Results,* Studia Universitatis “Babes- Bolyai” Cluj-Napoca, Seria Geographia, Nr. 1-2, 1995.
7. Haidu, I., Linc, Ribana (2001) *Modelarea stocastica a sezonalitatii scurgeri raurilor. Aplicatie la Culoarul Timis-Cerna*. Analele Univ. de Vest din Timisoara, Seria Geografie, Vol. XI.
8. Hazen, A. (1914), *Storage to be provided in impounding reservoirs for municipal water supply*. Trans. ASCE 77: 1539.
9. Pop, Oana Antonia, Horvath, Cs. (2009), *Assessing the average multi-annual runoff in the Tur River Basin*, Geographia Napocensis, Year III, nr. 1. Edit. Casa Cartii de Stiinta.
10. Sudler, C.E. (1927), *Storage required for the regulation of streamflow*. Trans. ASCE91: 622.

11. Petersen, Th., Devineni, N., Sankarasubramanian, A. (2012), *Seasonality of monthly runoff over the continental United States: Causality and relations to mean annual and mean monthly distributions of moisture and energy*, Journal of Hydrology, Volumes 468–469, 25 October 2012, pages 139-150, ISSN 0022-1694
12. Thomas, H.A., Fiering, M. B. (1962) *Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation*. In: Design of Water Resources Systems, (Ed. by A. Maas *et al.)* Chapter 12. Harvard University Press, Cambridge, Mass.
13. *** Planului Bazinal pentru Administratia Bazinala de Apa Somes-Tisa 2010-2013 *http://www.rowater.ro/dasomes/Planbaz/Plan%20Bazinal%202010-2013.htm*