Romanian Reports in Physics **73**, 703 (2021)

STATISTICAL APPROACH OF THE BEHAVIOR OF HAMCEARCA RIVER (ROMANIA)

A. BARBULESCU and C. MAFTEI

"Transilvania" University of Brasov, 5 Turnului Str., 500152, Brasov, Romania
E-mails: *alinadumitriu@yahoo.com*; *cemaftei@gmail.com*

Received May 2, 2020

Abstract. In this article, we provide a statistical comparison of the flow series measured at two different hydrological stations on Taita River, which is one of the most important rivers from Dobrogea (Romania). This study has a practical purpose since Taita is among the few rivers from Dobrogea that don't suffer from dying, so its water can be used for irrigation. This study emphasizes the concordance between the results of the statistical tests and the field data concerning extreme events. Long-range dependence of the events is also noticed.

Key words: breakpoint, distribution, flow, LRD, trend.

1. INTRODUCTION

Detecting the trend of flow series is important for the engineering projects and studies of water resources, and in the researches about the impact of the climatic changes on the water resources. Therefore, many articles analyzed the spatio-temporal variability of the flow series, at different scales. All over the world, the scientists report some tendency in the hydrological data: Burn *et al.* [1] identified a negative trend in the streamflow of Canadian Prairies that could be due to changes in temperature, Fanta *et al.* [2] emphasized a negative trend in the annual river flow of southern Africa Region, Lins and Slack [3], and Douglas *et al.* [4] found ascendant trends in the low flow of studied rivers in the United States. Similar analyzes have been conducted in some regions of Asia and Europe. For example, Miao *et al.* [5] detected a negative trend in the streamflow of Yellow River Basin, Giakoumakis and Baloutsos [6] found a decrease in the precipitation records from Greece, Stahl *et al.* [7] studied streamflows from 441 small catchments in 15 European countries during the interval 1962–2004 and concluded that there exist negative trends in southern and eastern regions and generally positive trends elsewhere.

Recently, a new concern regarding changes in the flow series refers to the detection of a possible presence of so-called long memory of the data. The phenomenon known as the Hurst phenomenon represents the trend of wet years to cluster in wet periods and drought years in the dry periods. Its existence was studied by Hurst and verified on series as the flow of Warta [8] and the Nile [9] Rivers, the monthly and annual inflows in Maggiore Lake, Italy [10], the annual streamflow records across the continental United States [11], the annual flow of the Nemunas River [12], the average daily flow of Danube River [13], the Boeotikos Kephisos River runoff [14].

Last period the research performed in Dobrogea, the region of Romania where the studied catchment is situated, has been focused on modeling the precipitation or temperature evolution [15–20].

Taita River (Fig. 1) drains a small catchment (681 km^2) situated in the Northern part of Dobrogea region (Romania), in a moderate continental climate. The annual mean temperature in the region is 10.9^{o}C and the annual precipitation varies between 400 and 500 mm. Precipitation is the main source of supply of Taita River – 74%, compared to the underground supply – 24%. The river is a tributary of Babadag Lake.

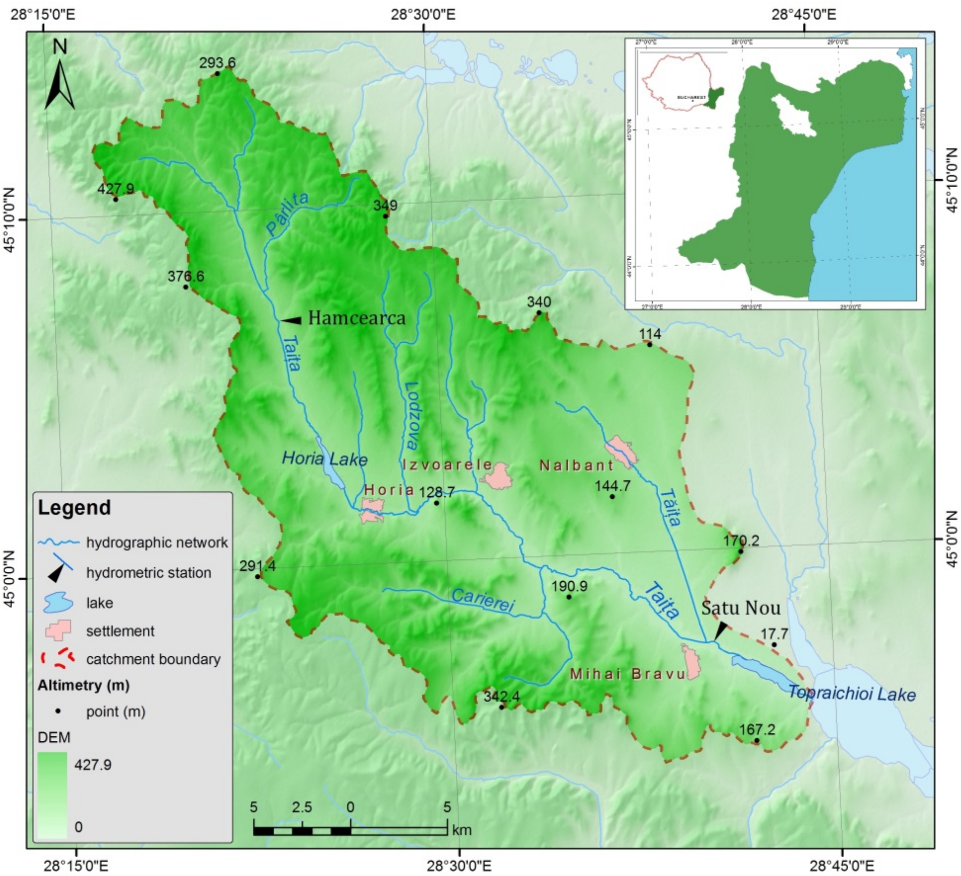

Fig. 1 – Taita catchment [20].

This basin is enhanced with different hydraulic works. An attenuation dam (in operation since 1971) is situated upstream of Horia village. It is in use to irrigate over 78 ha of land, the irrigation system being operational since 1975.

The series analyzed in this study are the monthly discharges from the hydrometric stations Hamcearca and Satu (Fig. 1) for the period January 1967–December 2010 (Fig. 2). Taita hydrographical basin has been chosen because the analyzed river (Taita) doesn't dry up in summer (even if prolonged drought periods are recorded in Dobrogea) and the studied series covers a long period (starting from 1955). Even if some articles treat the water quality on the Black Sea Littoral and the Danube Delta [21–23], our study comes to fill up a gap in the knowledge about the streamflow trend of the rivers in Dobrogea.

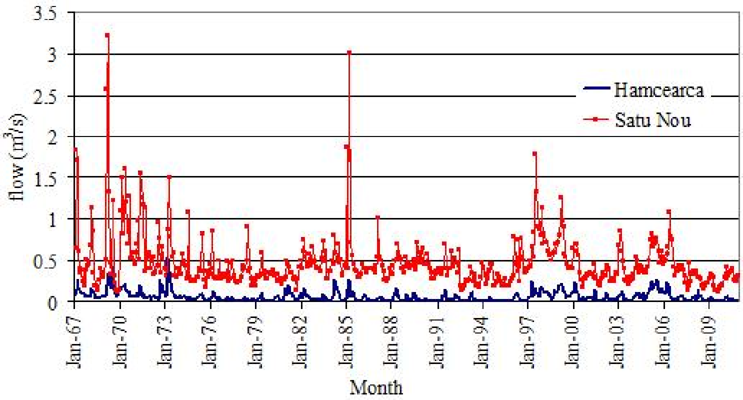

Fig. 2 – The studied series.

2. EXPERIMENTAL

The steps to be followed for analysis of hydrological data are: obtaining and preparing a suitable dataset, exploratory analysis of the time series, application of statistical tests, and results' interpretation.

Data treated here are accurate because they are provided by the National Institute of Hydrology and Water Management (INGHA). The measurement frequency was regularly done daily, smaller time steps being considered in case of floods.

The following statistical hypotheses have been tested on the monthly series, at the significance level of 0.05:

i. The hypothesis H_0: Both series (recorded at Hamcearca and Satu Nou) have the same distribution, against the alternative that they have different distributions.

ii. The hypothesis H_0: The series is Gaussian *vs.* the alternative H_1: The series is not Gaussian.

iii. The hypothesis H_0: Data are randomly distributed *vs.* the alternative H_1: Data are not randomly distributed, by the rank test [24]. The autocorrelation function was also computed for double-checking the null hypothesis, H_0.

iv. The hypothesis H_0: The series present outliers, *vs.* the inexistence of aberrant values.

The outliers' presence is related to the existence of extreme events, in this care periods with high precipitations, and consequently, floods.

v. The hypothesis H_0: There is no change in the series' mean *vs.* the alternative H_1: The series' mean is increasing (or decreasing) over time.

The modifications that appear in a time series' evolution can be investigated by different statistical tests. Since our series are not Gaussian, non-parametric methods were employed. The change point tests look for step – change of the mean/median of a series, change of the trend, of the mean/median, or change of its distribution.

Here we employed the Kendall tau and the seasonal Kendall tau for the data series [25]. The slope's estimation has been done by the nonparametric method of Sen [26]. The seasonal test computes the Mann-Kendall statistics on each group of months (Januarys, Februarys, etc.) separately, and then combines the results. The Kendall statistic is the sum of Kendall statistics of all the months.

The null hypothesis is rejected at the significance level α if $|\sigma_{Sk}| > Z$, where Z is the standard normal distribution value with an exceedance probability of $\alpha/2$,

$$\sigma_{Sk} = \sqrt{\sum_{i=1}^{m} (n_i/18)(n_i - 1)(2n_i + 5)},$$

and n_i is the number of the values in the i^{th} month.

Alternatively, the null hypothesis is rejected if the *p*-value associated with the test is less than 0.05.

We also test the hypothesis that the series has a constant unspecified mean, and an unspecified constant variance, by using a CUSUM procedure [27].

vi. The series stationarity was tested by using the ADF test. If the *p*-value associated with the test is less than 0.05, the alternative hypothesis could be accepted, so one can accept the series' stationarity [28].

vii. Detecting the long-range dependence (LRD) property.

The main way to quantify LRD is through the Hurst exponent (H): if $0 < H < 0.5$, the series is long-range anti-correlated, if $0.5 < H < 1$, it is long-range correlated, $H = 0.5$ corresponds to a Brownian noise, and $H = 1$ indicates $1/f$ noise, typical of systems in a self-organized critical state [20]. To estimate the Hurst coefficient we used different methods, as R/S, Aggregated variance, Absolute moments, the variance of residuals, the Abry-Veich, and the Whittle estimators [28–33].

3. RESULTS

i. Even if the two series contain the flow measured on the same river, their distribution is not the same, as it resulted after performing the Kolmogorov-Smirnov test and by comparison of the cumulative distributions of Hamcearca and Satu Nou series (Fig. 3).

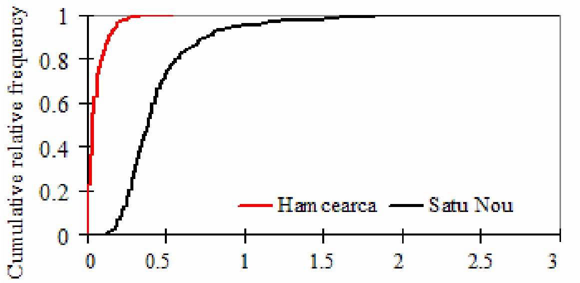

Fig. 3 – Cumulative distribution of Hamcearca and Satu Nou series.

ii. The normality tests rejected the normality hypothesis. The distributions are right-skewed, as resulted from their histograms (Fig. 4).

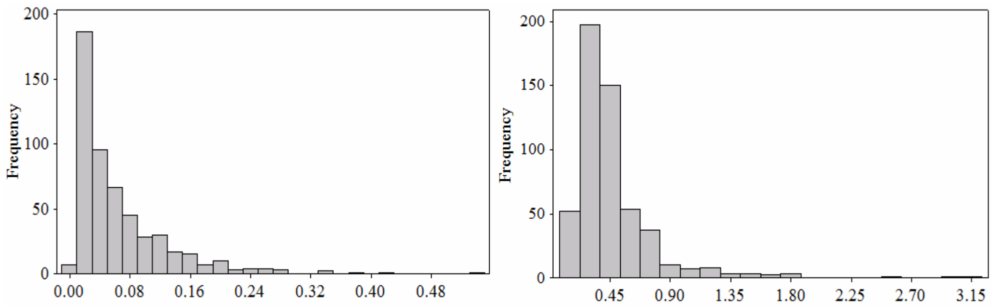

Fig. 4 – Histograms of Hamcearca and Satu Nou series.

iii. The rank test rejected the null hypothesis, a fact confirmed by the correlograms (ACFs (Fig. 5)), so data are not randomly distributed.

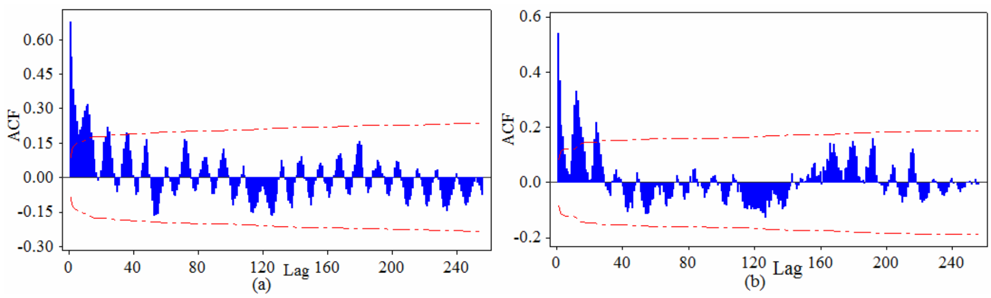

Fig. 5 – ACF of: a) Hamcearca; b) Satu Nou series.

High values of the autocorrelation function are noticed for lags between 0 and 30, which is conformity with the field reality.

iv. Testing for outliers' presence, it resulted that March 1967, February, March, April, May, and July 1969, July 1972, February, March, April, May 1973, March and April 1984, March 1985, April 1997, February, March and April 1999, February 2000, February, March, May, June and July 2005, March 2006 are outliers for Hamcearca series. February and March 1967, February, March, and April 1969, January and March 1985 are outliers for Satu Nou series (Fig. 6).

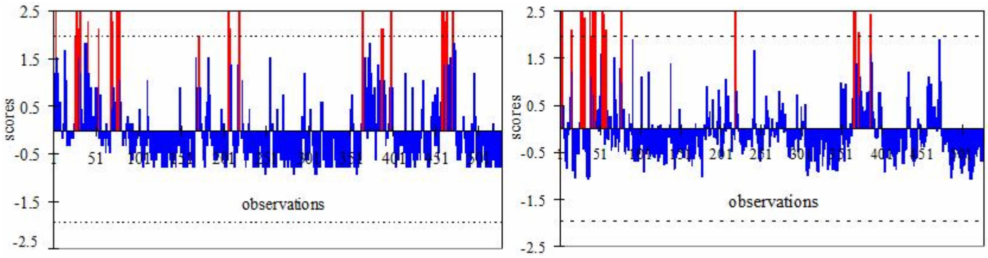

Fig. 6 – Outliers of Hamcearca and Satu Nou series.

This means that during the reported month, high quantities of precipitations were recorded, at both stations, and a high accumulation of water from different affluents was accumulated at Satu Nou. The months found as change points are exactly those when high floods were recorded all over Romania. Again, the theory and field results are concordant.

v. Using the Bayesian information criterion (BIC) for the selection of the segments, the mDP algorithm provided 20 change points for Hamcearca and 35 for Satu Nou, a big part of them being due to the outliers' presence [17]. After their removal, the change points detected by mDP are October 1980, January 1997, December 2003, and June 2006, for Hamcearca and March 1997, July 1998 and July 2006, for Satu Nou (Figs. 7, 8). So, the series have different change points. Their existence was also confirmed for both series, by CUSUM analysis (Fig. 9).

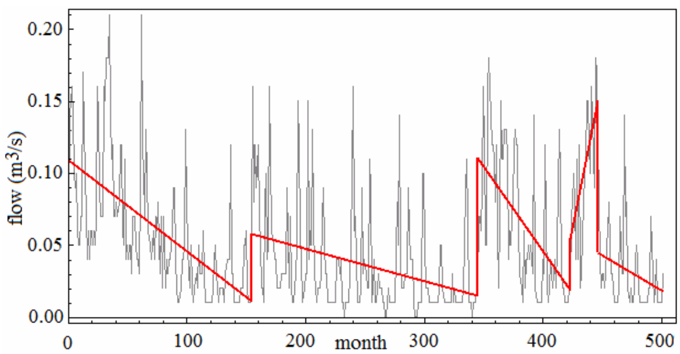

Fig. 7 – Segments of Hamcearca series after outliers' removal.
The lines represent the linear trend of the sub-series.

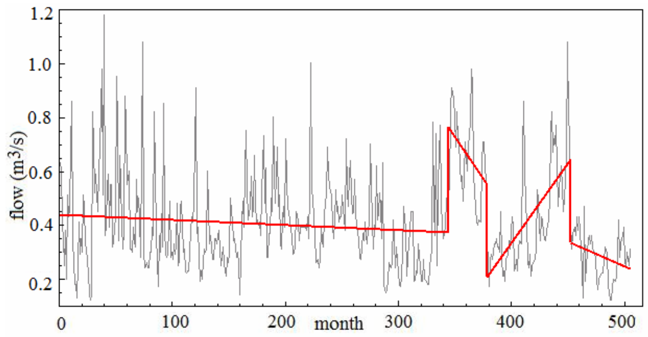

Fig. 8 – Segments of Satu Nou series after outliers' removal.
The lines represent the linear trend of the subseries.

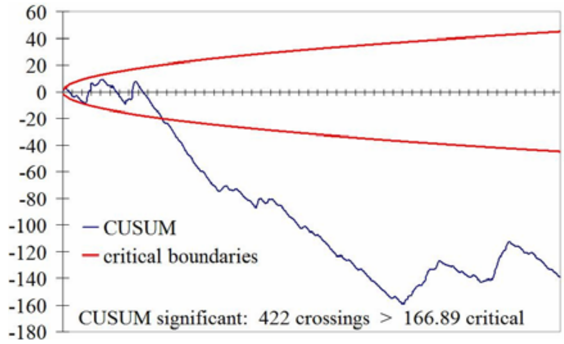

Fig. 9 – CUSUM chart of Hamcearca series.

The results of Kendall's tau and seasonal Kendall's tau tests, together with Sen's slope estimation are presented in Table 1 for both series. Since the *p*-values are less than 0.05, the null hypothesis was rejected in all cases, so both Hamcearca and Satu Nou series present trends. The slope of the linear trend is respectively –1.629E-4 for Hamcearca, and –4.425E-5 for Satu Nou.

Table 1

Results of Kendall's tau test with Sen's slope for Hamcearca and Satu Nou series

	Kendall's tau trend test		Seasonal Kendall's tau/ period =12	
	Hamcearca	Satu Nou	Hamcearca	Satu Nou
Kendall's tau	– 0.152	–0.096	–0.110	–0.166
S	–20127.000	–13289.006	–1237.000	–1757.000
Var(S)	16146745.667	16393245.006		
p-value	< 0.0001	0.001	0.000	< 0.0001
Sen's slope	–1.629E-4	–4.425E-5		
C.I.	(–0.010, 0.007)	(–0.002, 0.002)		

Since the slopes of the entire series are very small, the existence of the subseries slopes was tested. The results for Hamcearca sub-series are presented in Table 2, from which we deduce that all slopes are significant.

The slopes of the subseries of the Satu Nou series were also found to be significant, and their values were respectively –1.031E-4, –0.021, 8.221E-4, and –0.02.

Table 2

The results of Kendall's tau test with Sen's slope for the sub-series of Hamcearca series

Subseries	H1	H2	H3	H4	H5
Kendall's tau	–0.53	–0.31	–0.44	0.40	–0.47
S	–5949.00	–5056.00	–1271.00	107.00	–942.00
Var(S)	405562.33	742040.00	53152.33	1606.33	31582.00
p-value	<0.0001	<0.0001	<0.0001	0.01	<0.0001
Sen's slope	–5.263E-4	–1.351E-4	–0.001	0.004	–0.001
alpha	0.05	0.05	0.05	0.05	0.05

vi. The values of the Dickey-Fuller test statistic are –4.509 and –6.162, for Hamcearca and Satu Nou, the *p*-values being 0.01 in both cases. Since the *p*-values are less than 0.05, the null hypothesis was rejected, so the series are stationary (in trend). We can conclude that the water flow is decreasing, as the effect of climate change, so there is the danger of the apparition of the river drying phenomenon.

vii. The results of estimating the Hurst coefficient (*H*) before and after shuffling are presented in Tables 3 and 4 for the Hamcearca and Satu Nou series, respectively for the series without outliers. All the methods, but Abry-Veich and Whittle, return the values of correlation coefficient (%), whereas the last two procedures provide the confidence intervals, at 95% confidence level.

Table 3

Results of estimation of Hurst's coefficient

Method	Hamcearca			
	Initial series		Shuffled (bucket lenght = 16)	
	H	*r* (%)/95% C.I.	*H*	*r* (%)/95% C.I.
R/S	0.740	98.31	0.770	98.06
Aggregated Variance	0.709	93.83	0.722	93.28
Variance of residuals	0.989	98.52	0.992	98.81
Abry-Veich	0.918	[0.926, 1.110]	0.663	[0.571, 0.755]
Whittle	0.933	[0.874, 0.991]	0.715	[0.658, 0.772]
Method	Satu Nou			
	Initial series		Shuffled (bucket lenght = 16)	
	H	*r* (%)/95% C.I.	*H*	r (%)/95% C.I.
R/S	0.506	91.95	0.572	94.10
Aggregated Variance	0.687	99.77	0.701	97.18
Variance of residuals	0.969	97.11	0.928	97.45
Abry-Veich	0.919	[0.827, 1.011]	0.517	[0.426,0.609]
Whittle	0.850	[0.792, 0.909]	0.624	[0.568, 0.680]

Table 4

Results of estimation of Hurst's coefficient for the series without outliers

Method	Hamcearca			
	Initial series		Shuffled (bucket lenght = 16)	
	H	*r* (%)/95% C.I.	*H*	*r* (%)/95% C.I.
R/S	0.634	97.78	0.666	97.50
Aggregated Variance	0.824	97.45	0.843	95.53
Variance of residuals	0.891	97.92	0.928	98.06
Abry-Veich	0.998	[0.851, 1.067]	0.560	[0.302, 0.618]
Whittle	0.927	[0.844, 1.010]	0.685	[0.605, 0.766]
Method	Satu Nou			
	Initial series		Shuffled (bucket lenght = 16)	
	H	*r* (%)/95% C.I.	*H*	*r* (%)/95% C.I.
R/S	0.457	88.46	0.439	87.10
Aggregated Variance	0.383	79.59	0.398	78.64
Variance of residuals	0.713	96.40	0.804	95.32
Abry-Veich	0.785	[0.627, 0.943]	0.710	[0.552,0.868]
Whittle	0.764	[0.673, 0.836]	0.620	[0.541, 0.700]

For Hamcearca raw series and the series without outliers, the *H* coefficients don't differ significantly, while for Satu Nou, a small decrease of *H* values is noticed for the series without aberrant values. After shuffling the Hamcearca series, only the Abry-Veich and Whittle estimations of *H* significantly decreased, remaining above 0.5. So, the Hamcearca series present an LRD behavior. The Hurst coefficients for the Satu Nou raw series are above 0.5, with a decrease of the values computed by the Abry-Veich and Whittle methods of about 30%, after shuffling. After removing the outliers, the values computed by the R/S and Aggregated Variance are below 0.5, while the others remain above 0.62. So, the Satu Nou series present an LRD behavior.

4. CONCLUSIONS

In this article, we studied the statistical characteristics of the monthly series of the Taita River streamflow, recorded at Hamcearca and Satu Nou hydrological station. The results show a complex behavior of the series, the existence of a decreasing trend over the studied period and subperiods, and the long-range dependence. The flow variability must be taken into consideration in the context of water resources management, having deep implications in the economic activity that depends on the river resources.

REFERENCES

1. D. H. Burn, L. Fan, and G. Bell, Identification and quantification of streamflow trends on the Canadian Prairies, Hydrol. Sci. J. **53**(3), 538–549 (2010).

2. B. Fanta, B.T. Zaake, and R. K. Kachroo, A study of variability of annual river flow of the southern African region, Hydrol. Sci. J. **46**(4), 513–524 (2009).
3. H. F. Lins, and J. R. Slack, Streamflow trends in the United States, Geophys. Res. Lett. **26**(2), 227–230 (1999).
4. E. M. Douglas, R. M. Vogel, and C. N. Kroll, Trends in floods and low flows in the United States: impact of spatial correlation, J. Hydrol. **240**, 90–105 (2000).
5. C. Y. Miao, W. Shib, X. H. Chen, and L. Yang, Spatio-temporal variability of streamflow in the Yellow River: possible causes and implications, Hydrol. Sci. J. **57**(7), 1355–1367 (2012).
6. S. G. Giakoumakis, and G. Baloutsos, Investigation of trend in hydrological time series of the Evinos River basin, Hydrol. Sci. J. **42**(1), 81–88 (1997).
7. K. Stahl, H. Hisdal, J. Hannaford, L. M. Tallaksen, H. A. J. Van Lanen, E. Sauquet, S. Demuth, M. Fendekova, and J. J´Odar, Streamflow trends in Europe: evidence from a dataset of near-natural catchments, Hydrol. Earth Syst. Sci. **14**, 2367–2382 (2010).
8. M. Radziejewski, and Z. W. Kundzewicz, Fractal analysis of flow of the river Warta. J. Hydrol. **200**, pp. 280–294 (1997).
9. E. A. B. Eltahir, El Niño and the natural variability in the flow of the Nile River, Water Resour. Res. **32**(1), 131–137 (1996).
10. A. Montanari, R. Rosso, and M. S. Taqqu, Fractionally differenced ARIMA models applied to hydrologic time series, Water Resour. Res. **33**(5), 1035–1044 (1997).
11. R. M. Vogel, Y. Tsai, and J. F. Limbrunner, The regional persistence and variability of annual streamflow in the United States, Water Resour. Res. **34**(12), 3445–3459 (1998).
12. G. Sakalauskiene, The Hurst phenomenon in hydrology, Environ. Res. Eng. Manag, **3**(25), 16–20 (2003).
13. P. Pekárová, P. Miklánek, and J. Pekár, Long-term Danube monthly discharge prognosis for the Bratislava station using stochastic models, Meteorologický časopis **10**, 211–218 (2007).
14. D. Koutsoyiannis, The Hurst phenomenon and fractional Gaussian noise made easy, Hydrol. Sci. J. **47**(4), 573–595 (2003).
15. A. Bărbulescu, A new method for estimation the regional precipitation, Water Resour. Manag. **30**(1), 33–42 (2016).
16. A. Bărbulescu and J. Deguenon, Models for trend of precipitation in Dobrudja, Environ. Eng. Manag. J. **13**(4), 873–880 (2014).
17. A. Bărbulescu and J. Deguenon, Change point detection and models for precipitation evolution. Case study, Rom. J. Phys. **59**(5-6), 590–600 (2014).
18. A. Bărbulescu and J. Deguenon, About the variations of precipitation and temperature evolution in the Romanian Black Sea Littoral, Rom. Rep. Phys. **67**(2), 625–637 (2015).
19. A. Bărbulescu and C. Maftei, Modeling the climate in the area of Techirghiol Lake (Romania). Rom. J. Phys. **60**(7–8), 1163–1170 (2015).
20. C. Maftei, A. Bărbulescu, and A. A. Carsteanu, Long-Range Dependence in the Time Series of Taiţa River Discharges. Hydrolog. Sci. J. **61**(9), 1470–1474 (2016).
21. A. Bărbulescu and L. Barbeș, Mathematical models for inorganic pollutants in Constanța area, Romania, Rev. Chim.-Bucharest **64(7**), 747–753 (2013).
22. A. Bărbulescu and L. Barbeş, Assessment of surface water quality Techirghiol Lake using statistical analysis, Rev. Chim.-Bucharest **64**(8), 868–874 (2013).
23. A. Bărbulescu and L. Barbes, Assessing the Danube River water quality by statistical methods, *Environ. Earth Sci.* **79**, 2020, 122.
24. J. D. Gibbons and S. Chakraborti, Nonparametric statistical inference, Marcel Dekker, New York, USA, 1992.
25. M. G. Kendall, Rank Correlation Methods, Griffin, New York, USA, 1970.
26. P. K. Sen, Estimates of the regression coefficient based on Kendall's tau, J. Am. Stat. Assoc. **63**, 1379–1389 (1968).

27. A. Bărbulescu, Studies on time series. Applications in Environmental Sciences, Springer, 2016.
28. J. Deguenon and A. Bărbulescu, Fractal characterization of rainfall in Benin, Int. J. Ecol. Ec. Stat. **3**(30), 46–55 (2013).
29. T. Higuchi, Approach to an irregular time series on the basis of the fractal theory, Physica D**31**, 277–283 (1988).
30. H. E. Hurst, Long-term storage of reservoirs: an experimental study, T. Am. Soc. Civ. Eng. **116**, 770–799 (1951).
31. C. K. Peng, J. Mietus, J. M. Hausdorff, S. Havlin, H. E. Stanley, and Al. Goldberger, Long-range anticorrelations and non-Gaussian behavior of the heartbeat, Phys. Rev. Lett. **70**, 1343–1346 (1993).
32. M. S. Taqqu, V. Teverovsky, and W. Willinger, Estimators for long range dependence: an empirical study, Fractals **3**(4), 785–788 (1995).
33. T. Karagiannis, M. Faloutsos, and M. L. Molle, A user-friendly self-similarity analysis tool, ACM SIGCOMM Comp. Comm. Rev. **33**(3), 81–93 (2003).