

APPLICATION OF NON-PARAMETRIC CRITERIA FOR HOMOGENEITY AND NORMAL DISTRIBUTION OF ANNUAL STREAMFLOW TIME-SERIES DATA FOR THE ARDA RIVER BASIN

Abstract: The check of homogeneity and normal distribution characteristics of annual runoff is a mandatory requirement for further reliable hydrological analysis in each catchment. The aim of this work is to explore the listed statistical parameters for the annual streamflow time-series data of the Arda River basin (South Bulgaria), because the hydrometric output data, provided by the National Institute of Meteorology and Hydrology (NIMH), are modified. Basic statistical data, which include streamflow information from nine gauging stations, measured during the 1951–2015 period, are used here. Study period consists of two sub-periods (1951–1983, with original data, provided by Hydrological Reference Books, and 1984–2015, which includes the modified output information by the NIMH). Graphic methods (double-mass curves) and non-parametric criteria are applied. Basic statistical calculations are realized in the XLSTAT 2.0 (2018) application. A constant significance level ($p\text{-value} \geq 0.05$) is fixed. Results obtained show that all tested time-series are homogeneous. Exceptions of this statement are the data at the Byala River – Bostina, the Elhovska River – Rudozem and the Varbitsa River – Varli dol, according to the non-parametric criteria of Buishand and Neumann. Similar result achieves the normal distribution test. Excluding the hydrometric data from the gauging stations at the Byala River – Bostina and the Malka Arda River – Banite, all others time-series data are normal distributed. Log-normal histogram tests, approximating the empirical distribution in the listed stations, are accepted. Generally, the results established in this work show that the available hydrometric information about the annual flow in the catchment area of the Arda River, could be used for representative studies of hydrological processes.

Author information:

Kalin Seymenov

Student,
Sofia University “St. Kliment Ohridski”
✉ kalin.seimenov@abv.bg
🌐 Bulgaria

Keywords:

annual streamflow, Arda River basin,
homogeneity, normal distribution, statistical
analysis

The research was funded by the Department of Climatology, Hydrology and Geomorphology at the Faculty of Geology and Geography, Sofia University “St. Kliment Ohridski”. The article was written after winning a student scientific competition for hydro-climatic studies (A certificate of Kalin Seymenov – issued on 4th June 2018).

УВОД

Изследването на редовете от годишен речен отток за хомогенност и нормално разпределение е предварителен етап за изчисляване на средномногогодишна стойност, емпирична обезпеченост, трендове, циклични колебания и други характеристики на хидроложките процеси в даден речен басейн. Причини за нарушаване на еднородността на времевия ред от стойности на хидроложките характеристики, са промяната в условията на събиране и обработване на хидрометричната информация, изграждането на антропогенни съоръжения и др. [1]. Към тях от 80-те години на XX в. се прибавя предоставянето от Националния институт по метеорология и хидрология (НИМХ) на модифицирани данни с определени отклонения на стойностите от действителните. Последното обстоятелство аргументира целта на настоящето изследване – анализ на хомогенността и нормалното

разпределение на годишния речен отток в басейна на р. Арда, през периоди с оригинални и модифицирани изходни данни – за гранична на двата периода се приема 1983 г. – последната година с оригинална и публична информация за речния отток в България.

ИЗХОДНА ИНФОРМАЦИЯ И МЕТОДИ НА ИЗСЛЕДВАНЕ

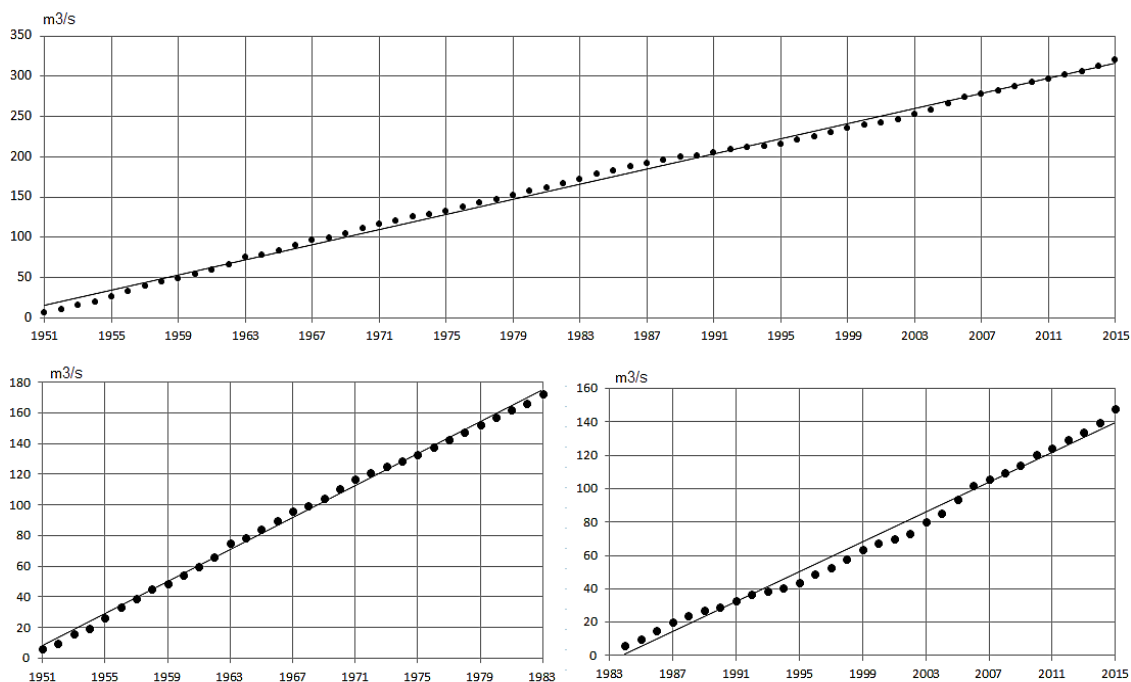
Опорната хидрометрична мрежа се състои от девет хидрометрични станции (ХМС): два водомерни поста на главната река (при Рудозем и Вехтино) и седем на притоците (три на левите – р. Бяла, Черна и Малка Арда, и четири на десните притоци – по една на р. Елховска и Крумовица и две на р. Върбица). Наблюденията започват предимно през 50-те години на ХХ в. (за начална в изследването е приета 1951 г., от която има измервания при повечето ХМС) и продължават до 2015 г., с изключение на р. Бяла – с. Бостина и Елховска – Рудозем, където ХМС са закрити през 2006 г. (табл. 1). Данните за годишния речен отток включват оригинални сведения, записани от хидрологичните справочници и годишници до 1983 г. [2, 3], а след посочената година са предоставени от НИМХ в модифициран вид.

Т а б л и ца 1 Хидрометрични станции и периоди на наблюдение в речния басейн на р. Арда

№ на ХМС	Река – ХМС	Н _{ср.} (m)	F (km ²)	Л _{изв.} (km)	Период с хидрометрични наблюдения	Използван период в изследването
61050	Бяла – с. Бостина	1384	52.9	12.1	1953–2006	1953–2006
61330	Елховска – Рудозем	1147	84.1	19.2	1950–2006	1951–2006
61350	Черна – с. Търън	1280	242.1	39.3	1951–2015	1951–2015
61400	Малка Арда – с. Баните	1172	114.0	30.5	1951–2015	1951–2015
61450	Върбица – с. Върли дол	647	471.2	54.8	1951–2015	1951–2015
61500	Върбица – сп. Джебел	584	1149.0	88.1	1936–2015	1951–2015
61550	Крумовица – Крумовград	494	497.6	37.3	1936–2015	1951–2015
61650	Арда – Рудозем	1162	260.6	38.7	1950–2015	1951–2015
61700	Арда – с. Вехтино	825	857.4	79.4	1950–2015	1951–2015

Еднородността и нормалното разпределение на годишния речен отток се изследват за периодите 1951–1983 и 1984–2015, с продължителност съответно 33 и 32 години, както и за целия период 1951–2015. Времевите редове се тестват за еднородност чрез двойносумарните криви, основаващи се на кумулативно натрупване на членовете на редицата – ако процесът е хомогенен, графиката е права линия [4], и на непараметричните тестове на Pettit, Buishand, Neumann и Standard Normal Homogeneity Test (SNHT) [5 (6), 7, 8, 9]. Нормалното разпределение се оценява по критериите Anderson Darling, Shapiro-Wilk, Lilliefors, Jarque-Bera [10, 11 (12), 13, 14]. Изчислителните процедури са осъществени в приложението XLSTAT. Нулевата хипотеза (H_0) не се отхвърля за $p\text{-value} \geq \alpha$, с приет праг на значимост $\alpha=0.05$. Избраните непараметрични тестове са прилагани за откриване на хидроморфологични промени в хидроложки редове [15, 16], и за статистически анализ на годишния отток в главните отточни области на България [17].

РЕЗУЛТАТИ



Фигура 1 Двойносумарна крива на годишния отток на р. Арда – Рудозем

Изследването за еднородност на времевите редове чрез двойносумарната крива, показва следните резултати: редиците за 1951–1983 и 1984–2015, както за 1951–2015, са хомогенни при всички ХМС (фиг. 1, прил. 1). Резултатът е потвърден от *p-value* на непараметричните тестове, с някои изключения. Такива се установяват за р. Бяла – с. Бостина, р. Елховска – Рудозем и р. Върбица – с. Върли дол, при *p-value* извън областта на допустимите значения, по критериите Buishand и Neumann1, и с препоръка за приемане на алтернативна хипотеза (H_a) (табл. 2).

Таблица 2 Статистики за хомогенност на редовете от годишен отток

Река – ХМС	Период на наблюдение											
	1951 – 2015				1951 – 1983				1984 – 2015			
	Pettit	Buishand	Neumann	SNHT	Pettit	Buishand	Neumann	SNHT	Pettit	Buishand	Neumann	SNHT
Бяла – с. Бостина	0.08	0.07	0.01	0.05	0.09	0.00	0.02	0.05	0.08	0.01	0.01	0.05
Елховска – Рудозем	0.10	0.06	0.02	0.09	0.11	0.07	0.03	0.10	0.08	0.06	0.02	0.09
Черна – с. Търън	0.11	0.07	0.07	0.08	0.11	0.08	0.08	0.09	0.10	0.06	0.07	0.08
Малка Арда – с. Баните	0.11	0.08	0.09	0.08	0.13	0.09	0.10	0.08	0.09	0.07	0.08	0.07
Върбица – с. Върли дол	0.09	0.04	0.03	0.10	0.09	0.07	0.03	0.12	0.08	0.01	0.02	0.07
Върбица – сп. Джебел	0.13	0.15	0.10	0.11	0.18	0.14	0.12	0.12	0.08	0.15	0.09	0.10
Крумовица – Крумовград	0.21	0.11	0.08	0.10	0.22	0.09	0.08	0.14	0.19	0.13	0.08	0.07
Арда – Рудозем	0.08	0.06	0.05	0.07	0.09	0.06	0.05	0.08	0.07	0.06	0.05	0.06

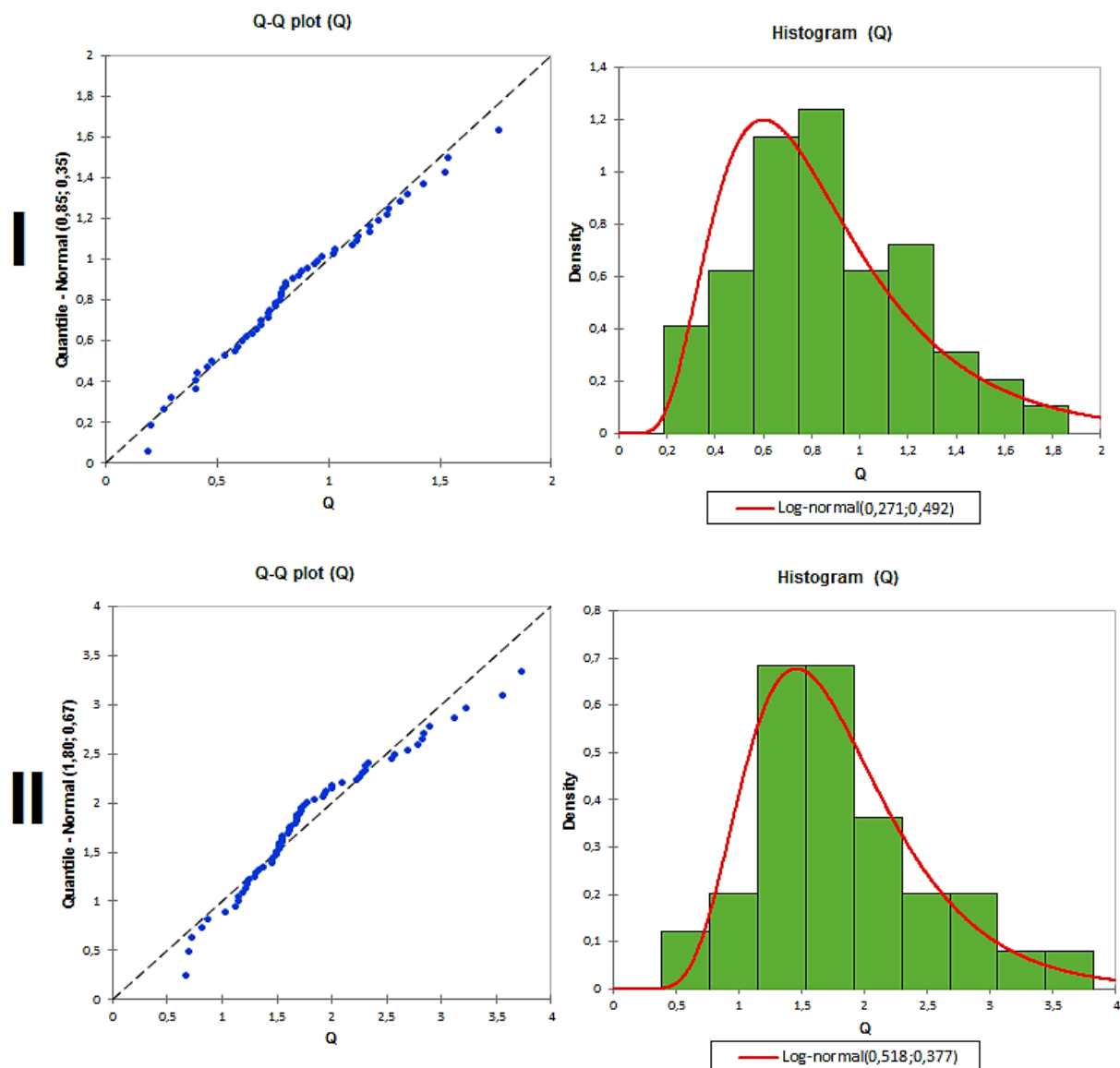
1 Посочените непараметрични критерии са максимално ефективни при брой членове в статистическата редица $n > 2000$ [18, 19].

Резултатите от непараметричните тестове за нормално разпределение на годишния отток в поречието Арда, показват че повечето редици са с нормално разпределение, и през двата изследвани периода: *p-value* попада в областта на допустими значения (табл. 3). Изключение са

Т а б л и ц а 3 Статистики за нормално разпределение на редовете от годишен отток

Река – ХМС	Период на наблюдение											
	1951 – 2015				1951 – 1983				1984 – 2015			
	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
Бяла – с. Бостина	0.10	0.04	0.11	0.02	0.12	0.06	0.08	0.03	0.09	0.02	0.12	0.01
Елховска – Рудозем	0.15	0.11	0.18	0.07	0.21	0.09	0.17	0.07	0.10	0.08	0.15	0.06
Черна – с. Търън	0.28	0.10	0.19	0.06	0.30	0.10	0.19	0.06	0.25	0.11	0.18	0.07
Малка Арда – с. Баните	0.09	0.03	0.13	0.04	0.09	0.04	0.13	0.04	0.11	0.02	0.11	0.03
Върбица – с. Върли дол	0.24	0.19	0.45	0.13	0.25	0.22	0.58	0.13	0.23	0.16	0.40	0.14
Върбица – Джебел	0.19	0.16	0.25	0.19	0.23	0.19	0.31	0.22	0.17	0.14	0.20	0.17
Крумовица – Крумвград	0.16	0.07	0.11	0.08	0.18	0.09	0.12	0.08	0.15	0.06	0.10	0.08
Арда – Рудозем	0.13	0.19	0.24	0.09	0.17	0.21	0.28	0.10	0.08	0.17	0.19	0.08
Арда – с. Вехтино	0.18	0.13	0.15	0.07	0.24	0.14	0.15	0.06	0.12	0.12	0.14	0.07

статистическите редове от данни на р. Бяла – с. Бостина и р. Малка Арда – с. Баните, по критериите на Anderson-Darling и Jarque-Bera (табл. 3). Времевиите серии не се подобряват след прилагането на Vox-Cox трансформации, което отхвърля първоначалното предположение за нормалност. Лог-нормални разпределения, апроксимиращи емпиричните, са приети за редиците на годишния отток в двете хидрометрични станции, с риск от отхвърляне на нулевата хипотеза (H_0), докато е вярна – от 71.5% до 84.1% (фиг. 2). Инструментът Chi-squared постъпва



Ф и г у р а 2 Q-Q plot; Хистограми на лог-нормално разпределение (1951–2015):
I) р. Бяла – с. Бостина; II) р. Малка Арда – с. Баните

в потвърждение на получените резултати: $p\text{-value} > \alpha$: 0.078 и 0.091. Параметърът на лог-нормалното разпределение (μ) приема максимална стойност 0.718 (при стандартна статистическа грешка ≤ 0.047) и вероятностна функция –200.334.

ЗАКЛЮЧЕНИЕ

Резултатите от проведеното изследване, дават основание за следните изводи:

- Двойносомарните криви и непараметричните критерии на Pettit's и SNHT оценяват всички изследвани хидрологични серии от данни като хомогенни. Подобни са резултатите, получени от тестовете на Buishand и Neumann, с някои изключения, за които е препоръчана алтернативна хипотеза (H_a).
- Тестовете на Shapiro-Wilk и Lilliefors показват, че данните от изследваните редици са с нормално разпределение при всички хидрометрични станции. Сходен резултат постигат критериите на Anderson-Darling и Jarque-Bera, с някои изключения, за които са възприети лог-нормални разпределения, апроксимиращи емпиричните.
- Подходът на изследване чрез разчленяване на изчислителния период на две части, не открива съществени различия. За част от станциите, в периода с модифицирани данни се наблюдава приближаване на $p\text{-value}$ до нивото на значимост, но със запазване на стойностите в областта на допустимите значения.

Времевите серии от данни осигуряват достоверност при пресмятането на основни статистически характеристики (средноаритметична стойност, стандартно отклонение, коефициент на вариация) и са репрезентативни за последвали проучвания (емпирична обезпеченост, трендове, циклични колебания и др.) на годишния отток в поречието Арда.

References:

1. Hydrological handbook, 1980, p. II – chapter. VII: Annual flow and its fluctuations – B: Sofia, 73–87 (Bg).
2. Hydrological Reference Book of the River in Bulgaria, 1981, p. II, GUHM, BAS (Bg).
3. Hydrological Annual Books 1976–1983, GUHM, BAS (Bg).
4. Searcy J. K., Hardison, C. H. 1960. Double Mass Curves. Manual of Hydrology: Part 1. General Surface, Water Techniques,” US Geological Survey, Water-Supply Paper 1541-B, 1960.
5. Pettit, A. 1979. Non-Parametric Approach to Change-Point Problem. – *Applied Statistics*, Vol. 28, 126–135
6. Pohlert, T. 2017. Non-Parametric Trend Tests and Change-Point Detection. – *Creative Commons License*, 1–18.
7. Buishand., T. 1982. Some Methods for Testing of the Homogeneity. – *Journal of Hydrology*, Vol. 58, 11–27.
8. Neumann, J. 1984. The Rank Version of von Neumann’s Ratio Test for Randomness. – *Journal of American Statistical Association*, Vol. 37, 40–46.
9. Alexanderson, H. 1986. Homogeneity Test Applied to Precipitation Test. – *Journal Climatology*, 6, 661–675.
10. Anderson, T. D. Darling. 1952. A test of Goodness of Fit. – *Journal of American Statistical Association*, Vol. 49, 765–769.
11. Shapiro, S., M. Wilk. 1965. Analysis of Variance Test for Normality. – *Biometrika*, Vol. 52, 591–611.
12. Royston, J. 1982. An Extension of Shapiro and Wilk’s Test for Normality to Large Samples. – *Journal of the Royal Statistical Society (Applied Statistics)*, Vol. 31, 115–124.
13. Lilliefors, H. W. 1967. On the Kolmogorov-Smirnov test for normality ... – *Journal of the American Statistical Association*, Vol. 62(318), 399–402.
14. Bera, A.K., Jarque, C.M. 1982. Model Specification Tests A Simultaneous Approach. – *Journal of Econometrics*, Vol. 20, 59–82.
15. Mauro, N. 2017. Fundamentals of Statistical Hydrology – Springer, 1st Edition.
16. Ruxton, G., D. Wilkinson, M. Neuhaeuser. 2015. Advice on testing the null hypothesis that a sample is drawn from a normal distribution. – *Animal Behavior*, Vol. 107, 249–252.
17. Hristova, N., I. Penkov, K. Seymenov. 2018. Annual Streamflow of Major Catchment Areas in Bulgaria in the Beginning of XXI century. – *Book of International Scientific Conference “Air and Water Components of the Environment”*, 15th–17th March 2018, Cluj-Napoca, Romania, 153–162.
18. Razali, N., Y. Wah. 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. – *Journal of Statistical Modeling and Analytics*, Vol. 2, 21–33.
19. Machiwal, D., M. Jha. 2012. Hydrological Time Series Analysis: Theory and Practice – Springer, 1st Edition.

ПРИЛОЖЕНИЯ

Приложение 1 Двойно сумарни криви на годишния речен отток в поречието Арда:

I) р. Черна – с. Търън; II) р. Малка Арда – с. Баните; III) р. Върбица – сп. Джебел; IV) р. Крумовица – Крумовград

