UDC 556.5

STATISTICAL ASSESSMENT OF THE RUNOFF TIME SERIES HOMOGENEITY OF THE RIVERS IN BELARUS

A. A. Volchak[1], S. V. Sidak[2], S. I. Parfomuk[3]

[1] Doctor of Geographical Sciences, Professor, Dean of the Faculty of Engineering Systems and Ecology, Brest State Technical University, Brest, Belarus; email: volchak@tut.by
[2] Assistant of the Department of Informatics and Applied Mathematics Brest State Technical University, Brest, Belarus; e-mail: harchik-sveta@mail.ru
[3] Ph.D in Engineering, Associate Professor, Head of the Department of Informatics and Applied Mathematics Brest State Technical University, Brest, Belarus; e-mail: parfom@mail.ru

Abstract
The article considers the need to evaluate the homogeneity of hydrological series using several statistical criteria, since the effectiveness of methods for testing homogeneity depends on the distribution of data. In addition to the officially recommended statistical parametric Student and Fisher criteria for this purpose, the Buishand tests, the standard normal homogeneity test and the nonparametric Pettitt test were tested to assess the homogeneity of the time series of the runoff of rivers in Belarus. Based on the application of tests for series of different types of runoff, the classification of the studied series is carried out depending on the number of tests confirming the homogeneity hypothesis.

Keywords: river runoff, homogeneity, points of decomposition, statistical criteria, classes of homogeneity.

СТАТИСТИЧЕСКАЯ ОЦЕНКА ОДНОРОДНОСТИ ВРЕМЕННЫХ РЯДОВ СТОКА РЕК БЕЛАРУСИ

А. А. Волчек, С. В. Сидак, С. И. Парфомук

Реферат
В статье рассмотрена необходимость оценивания однородности гидрологических рядов с использованием нескольких статистических критериев, поскольку эффективность методов тестирования однородности зависит от распределения данных. Кроме официально рекомендуемых для этих целей статистических параметрических критериев Стьюдента и Фишера, для оценки однородности временных рядов стока рек Беларуси апробированы тесты Буишанда, стандартный нормальный тест на однородность и непараметрический тест Петтитта. На основании применения тестов для рядов различных видов стока проведена классификация исследуемых рядов в зависимости от количества тестов, подтверждающих гипотезу однородности.

Ключевые слова: речной сток, однородность, точки разладки, статистические критерии, классы однородности.

Introduction

Statistical analysis of hydrological time series is a fundamental step that must be performed before any action on modeling hydrological processes. An important issue in the study of the statistical structure of hydrological series is the choice of a method for processing the available hydrological information. To effectively apply the methods of mathematical statistics and probability theory to the series of hydrological characteristics, it is necessary to assess the adequacy of the time series to the required prerequisites of the mathematical apparatus used. The main requirements are the homogeneity and stationary of the sample, i.e. it is necessary that all random variables of the sample are from the same distribution of the general population, and the key sample parameters (mean, variance) are invariant in time [1].

The presence of points of decomposition (years of deviation from the homogeneity of observations) in the hydrological time series, reflecting both changes in the behavior of runoff-forming factors and the itself, is the main indicator of changes in the degree of impact of anthropogenic load and climate on the process of runoff formation. In case of violation of the homogeneity and stationary of the hydrological time series, the previously performed calculations of the main hydrological characteristics become unreliable, which results in a change in the degree of vulnerability of hydraulic structures, the inability to assess the risk of dangerous hydrological phenomena, etc. In this regard, the detection of such points of decomposition should be considered as the very first and key step in the analysis of the variability of hydrological processes.

According to the technical code of established practice, for the quantitative assessment of the homogeneity of hydrological series, it is recommended to use the criteria of sharply deviating extreme values in the empirical distribution (the Smirnov-Grubbs and Dixon criteria), the homogeneity of sample variances (the Fisher criterion) and sample averages (the Student's criterion) [2]. The peculiarity of the application of the Smirnov-Grubbs and Dixon criteria is that they are applicable for the conditions of a normal symmetric distribution law of the general population and the absence of autocorrelation. However, the empirical distributions of hydrological characteristics, in particular extreme runoff, may have an asymmetry, and often in time series there may be a statistically significant autocorrelation between adjacent members of the series (r(1)) [3, 4].

Specialists of the State Hydrological Institute (Russia) have developed generalized Fischer and Student criteria that are calculated for series that have an intra-row or inter-row connection [5]. These criteria are included in regulatory documents, which provided them with a mandatory status and wide use in hydrological research. However, it is proved in [6] that the generalized Student and Fisher criteria have much more disadvantages than advantages and are not acceptable for assessing the homogeneity and stationary of hydrological series. It is also worth noting that the WMO Guidelines on hydrological Practice do not recommend the use of these criteria in hydrological studies [7]. The main disadvantage of parametric criteria is due to the fact that the classical results ensure the correctness of the application of these criteria only when the assumption that the analyzed samples belong to the normal distribution law is fulfilled, since only for this situation are the distributions of the statistics of the considered parametric criteria or the percentage point table known and runoff does not always meet these requirements [8].

In this regard, a combined approach is used in this work: in addition to the standard tests (Student, Fisher), three additional tests for homogeneity are applied. All three tests can be used to determine the points of decomposition, if there is a violation of homogeneity. Two parametric tests – Buishand Range Test (BHR) and Standard Normal Homogeneity Test (SNH) and one nonparametric test – Pettitt's Test (PT) were used.

The purpose of the study is to test and compare methods for assessing the homogeneity of the series of different types of runoff on the example of rivers in Belarus.

To achieve this goal, the following tasks have been solved:
1) Analysis of the homogeneity of the runoff series of the rivers of Belarus using the Fischer and Student criteria;
2) Testing of BHR, SNH, PT tests to assess the homogeneity of the river runoff in Belarus, determining the points of decomposition in the runoff time series;
3) Classification of the studied series by the number of tests confirming the homogeneity hypothesis.

doi.org/10.36773/1818-1112-2021-126-3-92-95

Initial Data

The study used the observation data of the State Institution "Republican Center for Hydrometeorology, Control of Radioactive Contamination and Environmental Monitoring" of the Ministry of Natural Resources and Environmental Protection of the Republic of Belarus for the current hydrological stations for the period of instrumental observations published in the materials of state cadasters. The assessment of the stationary and homogeneity of the river runoff series was performed for 6 hydrological stations of the largest rivers of Belarus: the Pripyat River at the Mozyr station, the Neman River at the Grodno station, the Western Dvina River at the Vitebsk station, the Berezina River at the Bobruisk station, the Dnieper River at the Orsha station, and the Dnieper River at the Rechitsa station. The 70-year period (1948-2017) was chosen as a representative period. The gaps in the data series were restored using the computer software complex "Hydrolog" [9].

Research Methods

The calculated value of the Student's criterion statistics (Ct) is determined by the formula:

$$Ct = \frac{\overline{Q_1} - \overline{Q_2}}{\sqrt{n_1 \cdot \sigma_1^2 + n_2 \cdot \sigma_2^2}} \cdot \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 2)}{n_1 + n_2}} \qquad (1)$$

where $\overline{Q_1}$, $\overline{Q_2}$, σ_1^2, σ_2^2 – average values and variances of two consecutive samples, respectively, n_1, n_2 – sample volumes.

To assess the homogeneity according to the Student's criterion, it is necessary to compare the calculated values of Ct obtained by the formula (1) and the critical values of the statistics t_{cr} at a given significance level [10]. As a rule, the significance level is set to 5%, which is equivalent to accepting the null hypothesis about the equality of the average two samples of the time series with a probability of 95%. If $Ct > t_{cr}$, then the hypothesis of homogeneity for the two parts of the series is rejected. Accordingly, a number of the considered hydrological characteristics are recognized as heterogeneous.

To estimate the homogeneity of the variances of two consecutive parts of the series, the value of the Fisher statistics F is used:

$$F = \sigma_1^2/\sigma_2^2 \text{ when } \sigma_1^2 > \sigma_2^2. \qquad (2)$$

If the calculated value of the criterion statistics is $F < F_{cr}$ at given degrees of freedom, then the hypothesis of homogeneity of variances is accepted at a given significance level α [10].

All additional tests for the study of homogeneity (BHR, SNH, and PT) are determined by the null hypothesis that the data are independent and equally distributed; while the alternative hypothesis indicates that the data is considered heterogeneous with the presence of a point of decomposition. The choice of these tests for an additional study of the homogeneity of hydrological series is based on the following: the PT test is used for any type of distribution of sample data; the SNH test is more sensitive to detecting points of decomposition at the beginning and end of the time series. Despite the fact that the properties of the BHR test are mainly studied for a normal distribution, this test, like the PT test, can be applied to any type of data distribution.

The Buishand Range Test is based on adjusted partial sums or cumulative deviations from the average [12]:

$$S_0^* = 0,\ S_k^* = \sum_{t=1}^{k}\left(Q_t - \overline{Q}\right), \qquad k = \overline{1, n}\ , \qquad (3)$$

where S_k^* – cumulative deviation, Q_t – observed data on river runoff, $\overline{Q} = \frac{\sum_{k=1}^{n} Q_k}{n}$ – sample average, n – sample volume.

Adjusted partial amounts (S_k^{**}) are calculated by the formula:

$$S_k^{**} = \frac{S_k^*}{\sigma_Q}, \quad k = \overline{0, n}\ , \qquad (4)$$

where $\sigma_Q = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}\left(Q_t - \overline{Q}\right)^2}$ – standard deviation.

Outlier-sensitive Statistics $\widetilde{Q}$ is calculated by the formula:

$$\widetilde{Q} = \max_{0 \le k \le n}\left|S_k^{**}\right|. \qquad (5)$$

If $\frac{\widetilde{Q}}{\sqrt{n}} < \widetilde{Q}_{кр}$ then the null hypothesis is accepted. Otherwise, the series is homogeneous.

The Standard Normal Homogeneity Test [13] compares the average values of the data for the first k years with the data for the last (n-k) years using the T_k statistics:

$$T_k = k\overline{z_1} + (n-k)\overline{z_2}, \ \ k = \overline{1, n-1}\ , \qquad (6)$$

where

$$\overline{z_1} = \frac{1}{k}\sum_{i=1}^{k}\frac{Q_i - \overline{Q}}{s}, \ \ \overline{z_2} = \frac{1}{n-k}\sum_{i=k+1}^{n}\frac{Q_i - \overline{Q}}{s}, \ \ s = \frac{\sum_{i=1}^{n}\left(Q_i - \overline{Q}\right)^2}{n-1}. \qquad (7)$$

If the statistics $T_0 = \max_{1 \le k \le n-1} T_k$ is greater than the critical value [14], then the null hypothesis about the homogeneity of the series is rejected. For the maximum value of T_k the number k corresponds to the year of the homogeneity violation.

The Pettitt's Test. Statistics U_k for the calculation of which the ranks r_i of the hydrological quantities Q_i, $i = \overline{1, n}$ are used, is calculated by the formula:

$$U_k = 2\sum_{i=1}^{k} r_i - k(n+1), \ \ k = \overline{1, n}\ , \qquad (8)$$

Next, we calculate the statistics $U_0 = \max_{1 \le k \le n}|U_k|$. The statistical significance of the fracture point is checked by comparing the value of U_0 with the critical value [15].

Results and Discussion

To assess statistical homogeneity in terms of variances and averages, respectively, according to the Fisher and Student criteria, the time series is divided into two or more consecutive samples of the same or different length, and it is desirable to associate the boundaries of the partition with the dates of the alleged violation of stationary. In this regard, the verification of homogeneity according to these criteria was carried out in two stages: the first stage – the entire period of the study was divided into two shorter ones: 1948-1965 and 1966-2017 (division is due to mass reclamation works in 1965); the second stage is the division into the periods 1948-1987 and 1988-2017 (division is due to the fact that 1988 corresponds to the beginning of an intensive increase in average annual air temperatures on the territory of Belarus [16]). The results of the study of the series of average annual, maximum and minimum river runoff rates for homogeneity according to the Student and Fisher criteria are presented in Table 1.

Table 1 – The results of checking for the time series homogeneity of different types of river runoff using the Student and Fisher criteria

River – Station	Annual		Maximum spring		Minimum summer-autumn		Minimum winter	
	Ct	F	Ct	F	Ct	F	Ct	F
	(1948-1965) – (1966-2017)							
Pripyat – Mozyr	-3.05 (-)	1.11	0.48	1.23	-3.38 (-)	1.77	-2.61 (-)	3.00 (-)
Neman – Grodno	0.08	1.97	2.96 (-)	5.62 (-)	-1.46	1.02	-0.47	1.33
Western Dvina – Vitebsk	0.10	1.36	3.33 (-)	2.01	1.29	2.83 (-)	-1.41	2.01
Berezina – Bobruisk	0.47	2.29 (-)	3.58 (-)	7.71 (-)	-2.47 (-)	1.53	-2.32 (-)	1.29
Dnieper – Orsha	-0.18	1.31	4.97 (-)	3.76 (-)	-1.24	1.02	-2.52 (-)	5.15 (-)
Dnieper – Rechitsa	0.26	0.45	4.03 (-)	0.17	-1.07	1.01	-2.74 (-)	1.53
	(1948-1987) – (1988-2017)							
Pripyat – Mozyr	-1.16	1.21	2.51 (-)	2.66 (-)	-1.07	1.00	-1.44	4.34 (-)
Neman – Grodno	0.88	1.31	3.00 (-)	7.70 (-)	1.11	1.17	-2.16 (-)	1.34
Western Dvina – Vitebsk	-1.82	1.20	2.65 (-)	1.36	-0.44	1.34	-3.85 (-)	2.63 (-)
Berezina – Bobruisk	-0.17	1.07	3.50 (-)	9.44 (-)	-1.43	1.08	-3.19 (-)	1.25
Dnieper – Orsha	-2.14 (-)	1.45	3.56 (-)	2.43 (-)	-3.73 (-)	2.00	-4.34 (-)	3.41 (-)
Dnieper – Rechitsa	-1.13	1.04	3.45 (-)	6.63 (-)	-2.16 (-)	2.02	-4.85 (-)	1.36

Note: (-) – the series is inhomogeneous.

The results obtained in Table 1 show that the use of only the Student and Fisher tests in most cases does not allow us to make an unambiguous decision about the homogeneity of the hydrological series. This result is quite predictable, since the tests are based on a comparison of different sample parameters. This justifies the need to use more tests to check the homogeneity of the series.

The results of checking the runoff series for homogeneity by SNHT, BHR, PT tests are presented in Table 2.

Table 2 – The results of the analysis of time series of river runoff for homogeneity using the BHR, SNHT, PT tests

Test / River – Station	BHR		SNHT		PT	
	$\widetilde{Q}/\sqrt{n}$	point of decomposition	T_0	point of decomposition	U_0	point of decomposition
	Annual runoff					
Pripyat – Mozyr	1.29	1968	8.09		457	1968
Neman – Grodno	0.59		3.61		258	
Western Dvina – Vitebsk	1.19		5.63		485	1984
Berezina – Bobruisk	0.55		5.64		226	
Dnieper – Orsha	1.28	1984	6.51		525	1984
Dnieper – Rechitsa	0.61		6.61		350	
	Maximum spring runoff					
Pripyat – Mozyr	1.49	1981	8.90	1981	534	1983
Neman – Grodno	1.58	1971	11.32	1970	641	1988
Western Dvina – Vitebsk	1.8	1971	14.63	1970	584	1971
Berezina – Bobruisk	1.88	1971	15.71	1971	678	1987
Dnieper – Orsha	2.19	1970	21.75	1970	747	1972
Dnieper – Rechitsa	1.95	1970	17.24	1970	700	1987
	Minimum summer-autumn runoff					
Pripyat – Mozyr	1.56	1967	11.94	1967	555	1968
Neman – Grodno	0.85		3.43		298	
Western Dvina – Vitebsk	0.83		4.16		260	
Berezina – Bobruisk	1.29	1972	9.49	1961	473	1972
Dnieper – Orsha	1.78	1983	12.78	1983	672	1981
Dnieper – Rechitsa	1.28	1979	6.04		453	1976
	Minimum winter runoff					
Pripyat – Mozyr	1.41	1969	9.25	1969	638	1969
Neman – Grodno	1.28	1992	7.11		499	1992
Western Dvina – Vitebsk	1.86	1988	14.31	1989	593	1988
Berezina – Bobruisk	1.46	1987	8.86	1990	524	1985
Dnieper – Orsha	2.14	1980	18.42	1980	757	1981
Dnieper – Rechitsa	2.09	1978	18.51	1990	764	1977

The final results obtained during the analysis for homogeneity of various types of runoff time series using five tests are presented in Figure 1. The obtained diagrams show the number of tests that reject the null hypothesis of data homogeneity for the studied hydrological series at the significance level of 5 %.

According to the results of checking the homogeneity of hydrological series for five tests, the studied series are classified depending on the number of tests that accept the hypothesis of data homogeneity at the significance level of 5% according to the following rules:

- Data series belongs to class A, if the null hypothesis is accepted by all tests or rejected by one of five tests at a 5% significance level;
- Data series belongs to Class B if two or three tests reject the null hypothesis of data homogeneity;
- Data series belongs to Class C if four or five tests reject the null hypothesis of data homogeneity at the 5% significance level.

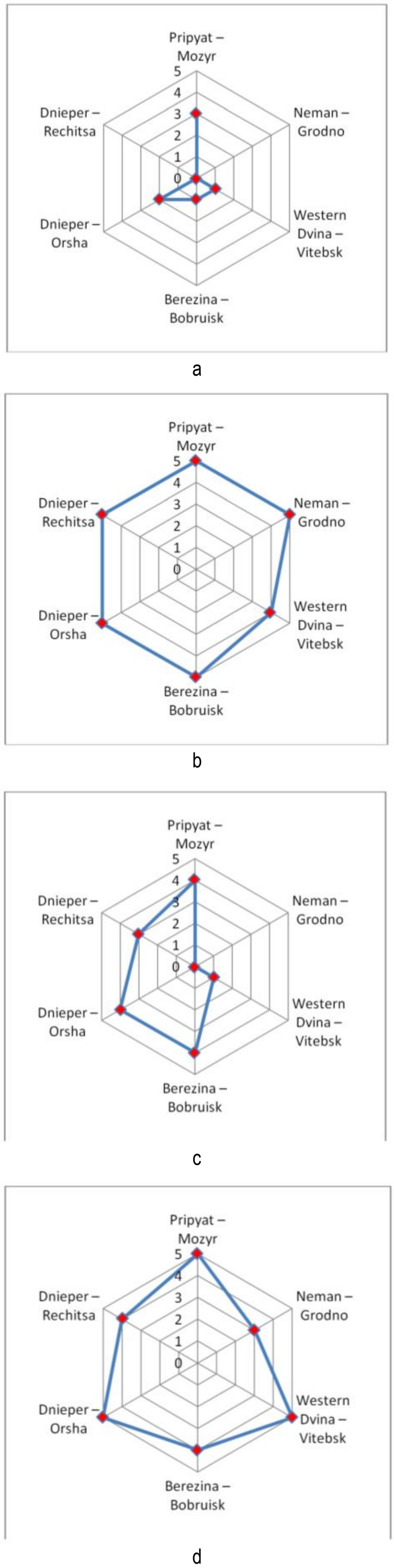

Figure 1 – The number of tests rejecting the null hypothesis of data homogeneity for the series of average annual runoff (a), maximum spring runoff (b), minimum summer-autumn runoff (c), minimum winter runoff (d)

doi.org/10.36773/1818-1112-2021-126-3-92-95

As a result of checking the series of river runoff for homogeneity, we came to the result: all the studied series of maximum runoff belong to class C (the points of decomposition in these series fall on the period 1970-1988); the series of annual runoff for the Pripyat River at the Mozyr station (with the point of separation in 1968) and the Dnieper River at the Orsha station (with the point of decomposition in 1984) belong to class B, the rest of the studied annual runoff series belong to class A; the range of the minimum winter runoff the Neman River at the Grodno station belongs to class B (with the point of decomposition in 1992), all the other studied rows of the minimum winter runoff belong to class C (the points of decomposition in these rows fall on the period 1969-1992); the rows of the minimum summer-autumn runoff for the Berezina River at the Bobruisk station, the Dnieper River at the Orsha station, and the Pripyat River at the Mozyr station belong to Class C, the Dnieper River at the Rechitsa station belongs to Class B, others belong to Class A.

Conclusion

The BHR, SNHT, and PT tests used in this study allow us to draw more reasonable conclusions about the homogeneity and points of decomposition of the hydrological series. The general conclusion from the statistical analysis of the homogeneity of the series of different types of river runoff in Belarus for the period 1948-2017 is that all the series of the maximum runoff and most of the series of the minimum winter runoff for the studied channels do not meet the conditions of homogeneity. Most of the annual runoff series are characterized by homogeneity, the minimum summer-autumn runoff series belong to different classes of homogeneity. The results of the analysis of the homogeneity of the samples of the series as a whole indicate significant changes in the formation of the water content of the rivers of Belarus, mainly in the period from the 1970s to the 1990s.

In this study, only statistical methods for estimating homogeneity are considered. However, all series belonging to Class B should be subjected to critical analysis before further use of the series, for example, for modeling. Such an analysis can serve as hydro-genetic methods for assessing homogeneity, which take into account the physical conditions of the formation of river runoff.

The work was carried out with the support of the BRFFI (grant no. X20M064)

References

1. Sikan, A. V. Methods of statistical processing of hydrometeorological information / A. V. Sikan. – St. Petersburg : RSHMU, 2007. – 278 p.
2. Calculated hydrological characteristics. The order of definition. Technical Code of Established Practice. TCP 45-3. 04-168-2009(02250) // Minsk : RUE "Stroytechnorm", 2010. – 55 p.
3. Volchek, A. A. Assessment of changes in the runoff of the Vilia River on the territory of Belarus / A. A. Volchek // Hydrometeorology and ecology. – 2016. – No. 4. – P. 32–62.
4. Volchek, A. A. Accounting for the distribution of maximum values of river runoff during the restoration of historical and cultural values / A. A. Volchek, S. V. Sidak // Restoration of historical and cultural objects as the preservation of the cultural heritage of the Republic of Belarus: a collection of articles of a scientific and technical seminar, Brest State Technical University, Brest, September 30, 2020 / editor: E. A. Tur [et al.]. – Brest: BrSTU Publishing House, 2020. – P. 35–39.
5. Rozhdestvensky, A. V. Synthesis of homogeneous criteria student's t test and Fisher to the case of correlated in time and space hydrological characteristics / A. V. Rozhdestvensky, A. V. Saharuk // Trudy GGI. – 1981. – Vol. 282. – P. 51–71.
6. Gorbachev, L. A. Evaluation of the homogeneity and stationarity of the series of annual runoff of rivers in Ukraine / L. A. Gorbachev, B. F. Khristyuk // Hydrology, hydrochemistry and Hydroecology. – 2014. – Vol. 1 (32). – P. 22–31.
7. Guide to hydrological practice. Volume II. Water resources management and the practice of applying hydrological methods / WMO No. 168. – Geneva, 2021. – 808 p.
8. Application of Homogeneity Tests: Problems and Solution / B. Y. Lemeshko [et al.] // Analytical and Computational Methods in Probability Theory. First International Conference, ACMPT 2017, Moscow, Russia, October 23-27, 2017 / Proceedings Editors: Vladimir Rykov, Nozer D. Singpurwalla, Andrey M. Zubkov (Eds.). – Cham : Springer, 2017. – P. 461–475.
9. Volchek, A. A. Automation of hydrological calculations / A. A. Volchek // Water management construction and environmental protection: Proceedings of the international scientific and practical conference on problems of water management, industrial and civil construction and economic and social transformations in the conditions of market relations / Brest polytechnic Institute. – Biberach-Brest-Nottingham, 1998. – P. 55–59.
10. Recommendations on statistical methods for analyzing the homogeneity of spatio-temporal fluctuations of river runoff. – Leningrad : Hydrometeoizdat, 1984. – 78 p.
11. Taxak, Arun Kumar. Long term spatial and temporal rainfall trends and homogeneity analysis in Wainganga basin, Central India / Arun Kumar Taxak, A. R. Murumkar, D. S. Arya // J. Weather and Climate Extremes. – 2014. – № 4. – P. 50–61.
12. Buishand, T. A. Some methods for testing the homogeneity of rainfall records / T. A. Buishand // Journal of Hydrology. – 1982. – № 58. – P. 11–27.
13. Alexandersson, H. A. A homogeneity test applied to precipitation data / H. A. Alexandersson // International Journal of Climatology. – 1986. – № 6. – P. 661–675.
14. Khaliq, M. N. On the critical values of the standard normal homogeneity test (SNHT) / M. N. Khaliq, T. B. M. J. Ouarda // International Journal of Climatology. – 2007. – № 27. – P. 681–687.
15. Pettitt, A. N. A Non-parametric Approach to the Change-point Problem / A. N. Pettitt // Journal of the Royal Statistical Society. Series C (Applied Statistics). – 1979. – № 28 (2). – P. 126–135.
16. Climate change in Belarus and their consequences / V. F. Loginov [et al.] / National Academy of Sciences of Belarus, Institute for Problems of Natural Resources Use and Ecology. – Minsk : Tonpik, 2003. – 330 p.

Accepted 26.10.2021