МАТЕМАТИЧНІ ТА ІНФОРМАЦІЙНІ МОДЕЛІ В ЕКОНОМІЦІ

UDC 004.942 ; 627.512 ; 556.013

D.V. STEFANYSHYN

ON THE USE OF THE TYPE I GUMBEL DISTRIBUTION TO ASSESS RISKS GIVEN FLOODS

Abstract. *In this article, there is considered the possibility of using the type I Gumbel distribution as one among alternative forecasting models in assessing flood risks at rivers, which is carried out based on forecasting of maximum hydrological characteristics according to hydrological observations. Examples are given and analyzed are the results of forecasting of the maximum water discharges of low exceedance probabilities, which were obtained with the use of the type I Gumbel distribution and other probability distributions for the Dnieper (the Vyshgorod water level gauge) and the Stryi River (the Verkhnye Syn'ovydne water level gauge). In addition to the analytical distributions for the Dnieper River, a generalized probability distribution function obtained by modeling within the fuzzy set theory was also used in the comparative analysis. The possibility of extending the scope of the practical application of the type I Gumbel distribution in forecasting hydrological maxima by logarithmic transformation is shown.*

Keywords: *alternative, asymmetry, hydrological maxima, flood, forecasting, probability distribution, risk.*

Introduction

Floods are among the most dangerous natural disasters causing human fatalities and different losses [1]. According to estimates [2], floods were responsible for about 6,8 million deaths in the 20th century. Annual economic worldwide losses from floods reached hundreds of millions of dollars [3]. And risk of deaths and catastrophic losses due to floods has been increasing by years in particular through the accelerated urbanization of coastal areas, river basins and lakeshores despite the fear of people before the floods [4].

The reasons, phenomena, factors and events that precipitate floods are diverse, multifaceted, and interrelated [2, 5]. These are weather factors such as storms, heavy or sustained rainfalls, rapid snow melts, heavy rains combining with rapid snow melts etc. Typical human factors include breakthroughs of dams, spreading water impervious surfaces in human settlements, defects or failures of drainage systems etc. Deforestation [6], physical land features including phase state of soils of the underlying surface, the present or absence of vegetation, other river basin drainage characteristics also influence flood outcomes. It should also mention the climate changes and the overall environment degradation [2, 5–7].

In Ukraine, flood challenges are also relevant [8]. Especially often floods occur on the Carpathian rivers in the western regions of our country [8–10]. Floods also threaten to the population living in the Dnipro basin. First of all, floods pose a significant threat to the population living in the Pripyat river basin [11]. There is also a permanent danger of artificial floods due to accidents on numerous dams [12]. Therefore, forecasting risks of floods in our country is one of important problems, and its relevance even will be increasing over time.

As it is known, the basic model used to forecast hydrological characteristics (water levels, discharges, volumes of water, etc.) according to hydrological observations is the probability distribution function [5, 13–15]. It can be used to determine flood parameters and its occurrence probability, to estimate the probable losses (risks of losses) because of floods, to solve water management problems, to assess the design values of water levels and discharges having extreme exceedance probability in designing of hydraulic structures [5, 13–17].

1. The use of probability distributions in hydrology

The use of probability distributions in hydrology is based on the assumption that hydrological observations data form representative sequences of independent and identically distributed random variables [5, 13–15]. For most rivers that have not undergone significant anthropogenic loads, and in the presence of data of continuous hydrological observations within time intervals of at least 30–40 years, this assumption can be quite admitted [5, 13].

Usually, the hydrological maxima have a positive asymmetry. Sometimes this asymmetry is quite significant. As well as, they exceed zero, or some other lower limit, but, in theory, they are not limited to the upper limit.

There are a lot of analytical probability distributions that meet these conditions and in case of presence of appropriate data might be used to forecast the maximum hydrological characteristics having low exceedance probability. These are, for example, such distributions as follows: the log-normal (two- and three-parameter) distributions, the gamma family and related distributions (exponential, two-parameter distributions, the three-parameter Kritsky-Menkel distribution and the Pearson type III distribution, etc.) and the extreme value distributions, which were developed within the extreme value theory [5, 13–15].

By this theory [5, 13–15] the generalized extreme value distribution is the only possible limit distribution of properly normalized maxima of a sequence of an independent and identically distributed random variable [18, 19].

In Ukraine, when hydrological calculations being performed according to observation data, it is accepted to use the three-parameter Krytsky-Menkel gamma distribution [13]. This model involves into its composition the parameters associated by means of transcendent equations with the coefficient of variation C_V and the coefficient of asymmetry C_S, and admits various relationship between them. For the different curves, the integration with the presentation of results in a tabular form was performed. It greatly simplifies hydrological calculations.

Among the extreme value distributions the type I Gumbel distribution is the most popular one in the modern hydrology practice [5, 14, 15, 19, 20]. Its popularity in hydrological calculations is due to the fact that the Gumbel's distribution is one of the simplest and most convenient analytical models that might be used to forecast extreme values. But not only because of this has the distribution deserved attention. It should also be noted that the hydraulic maxima prediction reliability provided by the type I Gumbel distribution is sufficiently high.

Besides, this is a two-parameter probability distribution, with constant asymmetry. And, as it is well-known, the estimation errors of the coefficient of variation and the coefficient of asymmetry (See example in Table 1), which are used, for example, in the three-parameter gamma distributions, can be quite significant, greater than the errors in determining the mean value of a sample of observations and its standard deviation (they are used in the Gumbel distribution). So, the "more precise" three-parameter probability distributions may even enlarge the uncertainty of forecasting to a certain extent.

The last remark also applies to some more "advanced" four-parameter and five-parameter probability distribution models, where, in addition to the asymmetry, the excess is used, etc. [21].

Table 1 – The accuracy of calculations of the statistical characteristics for water discharges maxima (the Dnieper River, the Vyshgorod water level gauge)

Parameter	Estimation	Standard error	Relative error, %
Mean value $\bar{x}$, m^3/s	4692	180	3,8
Standard deviation σ, m^3/s	2632	128	4,9
Coefficient of variation C_V	0,56	0,06	11,0
Coefficient of asymmetry C_S	1,26	0,17	13,2

Finally, it should be noted that there is no theoretical or another strict justification for choosing an appropriate probability distribution function of a random variable when forecasting extreme characteristics [5]. In principle, therefore, any probability distribution might be considered as a working hypothesis, if the distribution meets the appropriate statistical criteria and if other considerations regarding the adequacy of simulation are taken into account [22-24].

2. The type I Gumbel distribution

The type I Gumbel distribution [25] is a particular case of the generalized extreme value distribution (the last also is known as the Fisher-Tippett or the Fisher-Tippett-Gnedenko distribution). It is also known as the log-Weibull distribution and the double exponential distribution. As well as it is related to the Gomperz distribution; when its density is first reflected about the origin and then restricted to the positive half line, a Gompertz function is obtained [18, 19].

The type I Gumbel distribution has a probability distribution density, which might be described by the function [5, 14, 15, 19, 20, 25]:

$$f(x) = \alpha \exp\{-\alpha(x-u) - e^{-\alpha(x-u)}, \ -\infty \le x \le \infty, \tag{1}$$

where, the annual exceedance probability, %, for a random X :

$$P(X \geq x) = 100 \cdot \left[1 - \exp\{-\exp[-\alpha(x-u)]\}\right], \tag{2}$$

α and u are the distribution parameters, which are determined by

$$\bar{x} = u + \frac{0{,}5772}{\alpha}, \; \sigma^2 = \frac{\pi^2}{6\alpha^2}, \tag{3}$$

$\bar{x}$ is the mean value and σ is the standard deviation of the given series of observations data.

3. Examples of hydrological maxima forecasting

3.1. *Forecasting of water discharges maxima, the Dnieper River, the Vyshgorod water level gauge*

Considered a number of observations from 1787 to 1999 (Fig. 1) with the following statistical parameters: the mean value $\bar{x}$ = 4692 m^3/s; the standard deviation σ = 2632 m^3/s; the coefficient of variation C_V = 0,56; the coefficient of asymmetry C_S = 1,26.

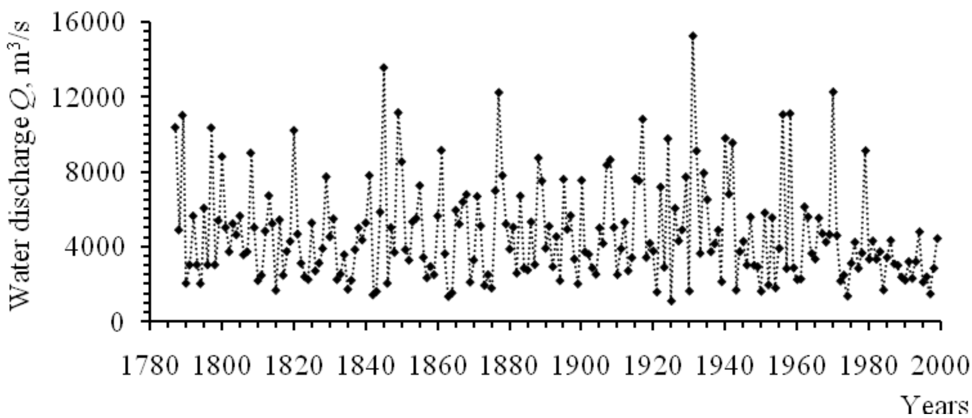

Fig. 1 – A time series of water discharges maxima, the Dnieper River, the Vyshgorod water level gauge

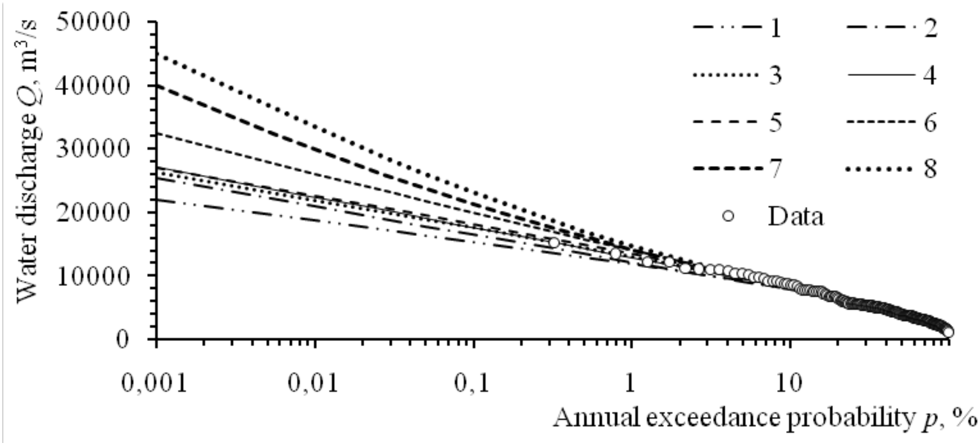

Fig. 2 – The probability distributions of water discharges maxima, the Dnieper River, the Vyshgorod water level gauge

Eight alternative probability distributions were used to forecast (See above Fig. 2), including four variants of the three-parameter gamma distribution according to the Krytsky-Menkel model: the model 1 corresponds C_V = 0,5 and $C_S = 2C_V$; the model 2 answers C_V = 0,5 and $C_S = 2{,}5C_V$; 5 satisfies C_V = 0,6 and $C_S = 2C_V$; 6 stands for C_V = 0,6 and $C_S = 2{,}5C_V$. The others distributions are as follows: the model 3 answers the Pearson III type distribution (arithmetical); 4 stands for the type I Gumbel distribution; 7 satisfies the two-parameter lognormal distribution; 8 corresponds the log-Pearson III type distribution.

After checking the statistical hypotheses by the Pearson criterion χ^2 for the significance level of 0,1%, all these probability distributions were found to be considered as the hypotheses that agreed with empirical data (Table 2). The validities $v(\chi^2)$ of the models were determined by using the Pearson criterion too.

It might be noted that the type I Gumbel distribution curve, despite the small value of its statistical validity, gave the acceptable approximation to points corresponding to observed data having low exceedance probability (less than 3%) (See also Fig. 3). The curve passes near four of them, in particular through three such points, which correspond to the observed water discharges with the least empirical exceedance probabilities.

It should also be noted that the type I Gumbel distribution (the model 4), its validity by Pearson criteria is 0,0425, in the interval of small exceedance probabilities, gave the result that is quite similar to results of "more" valid distributions (See Fig. 3); among them such distributions are as the Pearson III type distribution (arithmetical) (the model 3) with $v(\chi_i^2)$ = 0,1491 and the Krytsky-Menkel model 5 for C_V = 0,6 and $C_S = 2C_V$ with $v(\chi_i^2)$ = 0,1256.

Table 2 – Results of checking the statistical hypotheses by the Pearson criterion (the significance level of 0,1%)

Hypothesis	Probability distribution	Pearson criterion χ^2	Validity by Pearson criterion $v(\chi^2)$
1	The Krytsky-Menkel model (C_V = 0,5; $C_S = 2C_V$)	23,141	0,0418
2	The Krytsky-Menkel model (C_V = 0,5; $C_S = 2{,}5C_V$)	20,499	0,0865
3	The Pearson III type distribution (arithmetical)	18,425	0,1491
4	The type I Gumbel distribution	23,074	0,0425
5	The Krytsky-Menkel model (C_V = 0,6; $C_S = 2C_V$)	19,088	0,1256
6	The Krytsky-Menkel model (C_V = 0,6, $C_S = 2{,}5C_V$)	15,353	0,2874
7	The two-parameter lognormal distribution	15,066	0,3752
8	The Pearson III type distribution (logarithmic)	12,949	0,4530

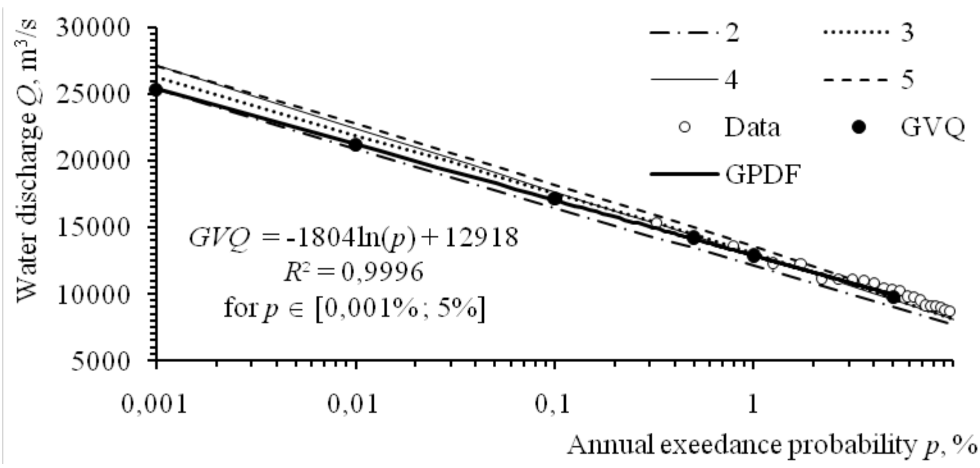

Fig. 3 – Comparison of the type I Gumbel distribution with the nearest neighbour distributions and with a generalized probability distribution function (GPDF) in the interval of small exceedance probabilities

A comparison of the type I Gumbel distribution with a generalized probability distribution function (GPDF) in the interval of small exceedance probabilities was also performed. Six generalized values (GVQ) of water discharges for annual exceedance probabilities of 5%, 1%, 0,5%, 0,1%, 0,01% and 0,001% were obtained (Fig. 3). These GVQ include "contributions" of all the chosen probability distributions (See Fig. 2). They were estimated by using methods of the fuzzy set theory according to the proposition given in [24].

The Gumbel's probability distribution showed a good coincidence with the GPDF for more likely floods having annual exceedance probabilities of 0,5% and more. In addition, the type I Gumbel distribution gives a certain margin to mitigate the risk connected with less likely floods.

3.2. *Forecasting of water discharges maxima, the Stryi River, the Verkhnye Syn'ovydne water level gauge*

The Stryi River is one among the largest left bank tributaries of the Dniester River. The river makes a significant contribution to floods taking place on the Dniester region. The presented Stryi River example is also interesting because the forecast of water discharges maxima is based on a relatively short series of observations.

There was considered a number of observations from 1951 to 1998 (See Fig. 4) with the following statistical parameters: the mean value $\overline{x}$ = 755,4 m^3/s; the standard deviation σ = 466,5 m^3/s; the coefficient of variation C_V = 0,6; the coefficient of asymmetry C_S = 1,8 ($C_S = 3C_V$).

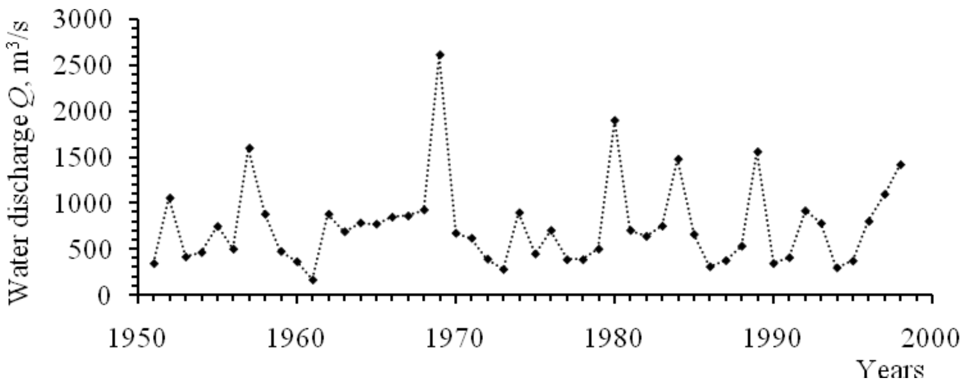

Fig. 4 – A time series of water discharges maxima, the Stryi River, the Verkhnye Syn'ovydne water level gauge

In the selected series of observations, it might be noted the high "emission" corresponding to the water discharge of 2610 m^3/s, whose empiric exceedance probability is about 2%, because of what the choice of an adequate probability distribution has essentially complicated.

Initially, the type I Gumbel distribution was considered (Fig. 5, the model 1). As well as different variants of the three-parameter gamma distribution of the Krytsky-Menkel model were used (See, for example, the model 2 with $C_S = 3C_V$ and the model 3 with $C_S = 5C_V$). None of the considered variants including the Krytsky-Menkel models gave an adequate approximation to the observed data. The Pearson III type distribution (arithmetical) with $C_S = 3C_V$ (the model 4) also turned out to be not a sufficiently adequate model to describe the observed data.

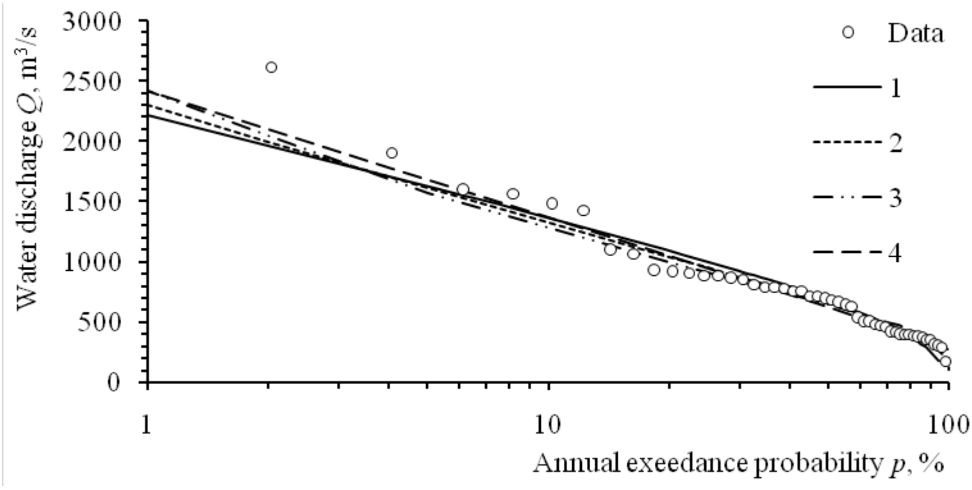

Fig. 5 – The probability distributions of water discharges maxima, the Stryi River, the Verkhnye Syn'ovydne water level gauge

It is known that the logarithmic transformation allows expanding the scope of practical application of probability distributions. This is a fairly simple and sufficiently effective technique to overcome the excessive asymmetry of observed data [26], which is carried out by replacing a random variable x by a random variable $y = \lg x$ or by a variable $y = \ln x$.

Usually, the logarithmic transformation is present in the two- and three-parameter log-normal distributions and in the log-Pearson III type distribution (See Fig. 2, the model 8). This technique may also be used regarding the type I Gumbel distribution. The results are shown in Fig. 6.

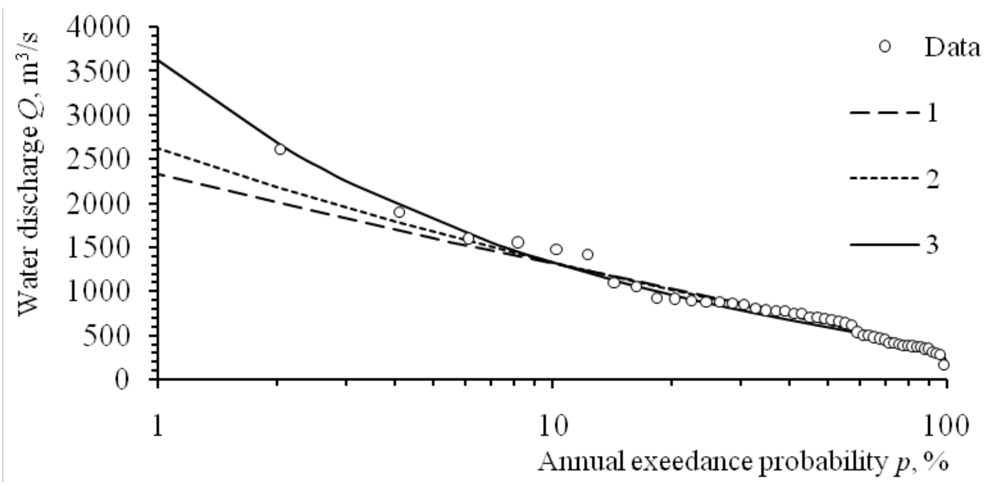

Fig. 6 – The logarithmic distributions of probability of water discharges maxima, the Stryi River, the Verkhnye Syn'ovydne water level gauge

In Fig. 6, it is shown the following probability distributions: the model 1, the two-parameter log-normal distribution; the model 2, the log-Pearson III type distribution; the model 3, the type I log-Gumbel distribution (the type I Gumbel distribution built for logarithms). It may be seen, that the logarithmic transformation for the Gumbel's distribution allowed taking into account the hidden asymmetry of data, the asymmetry, which was magnifying the uncertainty of forecasting. As a result, the less risky forecast was obtained.

Conclusions

It was considered the possibility of using the type I Gumbel distribution as one among alternative forecasting models in assessing flood risks at rivers, which is carried out based on forecasting of maximum hydrological characteristics according to hydrological observations.

It was shown, that the type 1 Gumbel distribution, both in arithmetic and logarithmic form, can be successfully applied in predicting the hydrological maxima having low exceedance probabilities. This distribution can be adequate forecasting model, which allows taking into account the individual particularities the time series of observed hydrological maxima.

REFERENCES

1. Abbott P. L. Natural Disasters. Wm. C. Brown Publishing Co. 1996. 438 p.
2. Doocy Sh., Daniels A., Murray S., Kirsch Th. D. The Human Impact of Floods: a Historical Review of Events 1980-2009 and Systematic Literature Review. URL: http://currents.plos.org/disasters/article/the-human-impact-of-floods-a-historical-review-of-events-1980-2009-and-systematic-literature-review/
3. Dobrovičová S., Dobrovič R., Dobrovič J. The Economic impact of floods and their importance in different Regions of the World with Emphasis on Europe. Business Economics and Management 2015 Conference, BEM2015. Procedia Economics and Finance 34. 2015. P.P. 649-655.
4. Rufat S., Tate E., Burton Ch. G., Maroof A. Say. Social vulnerability to floods: Review of case studies and implications for measurement. International Journal of Disaster Risk Reduction, 14, 2015. P.P. 470-486
5. Introduction to hydrology. First ed. by W. Viessman, Jr., T. E. Harbaugh, and J. W. Knapp. Intext educational publishers. New York. London. 1975. 704 p.
6. Raven P. H., Berg L. R. Environment. 5th ed, John Wiley & Sons. 2006. 406 p.
7. Briggs D., Smithson P., Addison K., Atkinson K. Fundamentals of the Physical Environment. Second Edition. London & New York, 1997. 557 p.
8. Сусідко М. М., Лук'янець О. І. Районування території України за ступенем гідрологічної небезпеки. Наук. праці УкрНДГМІ. Вип. 253. 2004. С. 196-204.
9. Кирилюк М. І. Режим формування історичних паводків в Українських Карпатах. Екологічні аспекти руслоформування малих річок (аналіз проблеми). Гідрологія, гідрохімія і гідроекологія. Наук. зб. / Відп. ред. В. К. Хільчевський. Київ : Ніка-Центр, 2001. Т. 2. С. 146-156.
10. Бойко В. М., Петренко Л. В. Стихійні гідрологічні явища на річках України в останні 10-15 років та проблеми їхнього оперативного попередження. Наук. праці УкрНДГМІ. 2006. Вип. 255. С. 272-278.
11. Сусідко М. М. Надзвичайно високі повені в басейні Прип'яті. Наукові праці УНДГМІ. 2006. Вип. 255. С. 279-282.
12. Стефанишин Д. В. Прогнозування аварій на греблях в задачах оцінки й забезпечення їх надійності та безпеки. Гідроенергетика України. № 3-4. 2011. С. 52-60.

13. Крицкий С. Н., Менкель М. Ф. Гидрологические основы управления речным стоком. Москва : Наука, 1981. 255 с.
14. Extreme Hydrological Events: New Concepts for Security (Nato Science Series: IV: Earth and Environmental Sciences).Paperback: Editors: O. F. Vasiliev, P. H. A. J. M. van Gelder, E. J. Plate, M. V. Bolgov. Springer; 1 edition. 2007. 480 p. URL: https://link.springer.com/book/10.1007%2F978-1-4020-5741-0.
15. Koutsoyiannis D. Probability and statistics for geophysical processes. National Tech. University of Athens. 2008. URL: https://www.itia.ntua.gr/en/docinfo/1322/.
16. Стефанишина-Гаврилюк Ю. Д., Стефанишин Д. В., Трофимчук О. М. Індуктивне моделювання ризиків збитків від руйнівних повеней в басейні р. Тиса за емпіричними даними з використанням моделей регресійного типу. Математичне моделювання в економіці. 2014. №1. С. 72-79.
17. Стефанишин Д. В. Обґрунтування економічно оптимального значення розрахункової витрати води малої ймовірності перевищення на водопропускних спорудах з врахуванням ризику. Математичне моделювання в економіці. 2015. № 2. С. 104-111.
18. Nasri-Roudsari D., Extreme value theory of generalized order statistics. Journal of Statistical Planning and Inference. Vol 55, 1996. P.P. 281-297.
19. Rust H. W., Kallache M., Schellnhuber H.-Jo., Kropp J. P. Confidence intervals for flood return level estimates assuming long-range dependence. Assuming Long-Range Dependence. In: Kropp J., Schellnhuber H. J. (eds). In Extremis. Springer, Berlin, Heidelberg. 2011. P.P. 213-241.
20. Okonofua S., Ogbeifun P. Flood Frequency Analysis of Osse River Using Gumbel’s Distribution. Civil and Environmental Research. Vol. 3, No. 10. 2013. P.P. 55-59.
21. Картвелишвили Н. А. Теория вероятностных процессов в гидрологии и регулирование речного стока. Ленинград : Гидрометеоиздат, 1985. 191 с.
22. Стефанишин Д. В., Грицюк П. М., Оленіч В. В. Вибір функцій розподілу кліматичних та гідрологічних параметрів з урахуванням ризику реалізації їх екстремальних значень. Штучний інтелект. 2’ 2004. ІПШІ «Наука і освіта», 2004. С. 393-397.
23. Stefanyshyn D. V. Application of risk analysis to support safety of dams and flooded territories against floods. Proc. of Int. Scientific School “Modelling and Analysis of Safety and Risk in Complex Systems”. June 24-28, 2008. Saint-Petersburg, Russia. P.P. 371-376.
24. Стефанишина-Гаврилюк Ю. Д., Стефанишин Д. В. Використання нечіткої міри для подолання невизначеності довгострокових прогнозів на основі екстраполяцій. Системні дослідження та інформаційні технології. 2013. № 4. С. 99-110.
25. Gumbel E. J. Les valeurs extrêmes des distributions statistiques. Annales de l'institut Henri Poincaré. Vol. 5. No. 2. 1935. P.P. 115-158.
26. Kuhn M., Johnson K. Applied Predictive Modeling. Springer Science+Business Media. New York, 2013. 600 p.

Been received for revision 15.02.2018.