
INTERACTION BETWEEN CONTINENTAL WATERS
AND THE ENVIRONMENT

Evaluating Parameters of the Distributions of Extreme Storms with Several Events per Year Taken into Account

D. E. Klimenko^{a,*}, E. S. Cherepanova^a, and A. Yu. Kuz'minykh^a

^aPerm State University, Perm, 614990 Russia

*e-mail: listopad19531@mail.ru

Received February 12, 2017; revised September 10, 2018; accepted October 18, 2018

Abstract—Applied problems of the analysis of data series of flood-forming storm precipitation (rain intensity over short time intervals), containing several events per year are discussed. The use of data containing several events per year has been shown to be justified for reliable determination of statistical characteristics of time series at a short observation period. The statistics of time series containing one or more events per year have been shown to correlate well with the frequency of the observed phenomenon in the Ural region. Recommendations have been developed for recalculating the statistics of the series containing several events per year into statistics for a single event per year; a brief comparative analysis of the methods used in Russia in this field is given.

Keywords: rain floods, showers, one or several events per year, multi-extremum series

DOI: 10.1134/S0097807819040110

INTRODUCTION

The study is focused on the applied problems of probabilistic estimates of flood-forming storm precipitation, which have a very variegated time and space nature. The central problem of the study is the estimation of the statistical parameters of time series containing several values within a calendar year and the transformation of such series into a limiting form, i.e., to a distribution in which each time interval contains a single maximum (peak-peak in N.A. Kartvelishvily's terminology [4]). Considering some factors (the sparse precipitation-gage network, the short observation series, the inadequate coverage of the territory by radar station network, the limited potential of remote-sensing methods, and the weak correlation between the series obtained by nearby gages), it is reasonable to utilize all observed maximums in excess of some threshold in the quantitative estimate of precipitation intensity over short time intervals [9, 14, 18, 19]. On the other hand, the use of methods of group analysis of time series will lead to excessive simplification of the spatial structure of shower parameter fields, which, in the case of mountain areas, makes it impossible to reveal geographic regularities.

The probabilistic nature of the hydrometeorological processes was not questioned in the XX century, starting from the practical studies of Heisen [4]. The practical needs have led to the ubiquitous employment of mathematical statistics apparatus, which have extended to engineering reference books and standards by the mid-XX century. The construction

design, in what regards the assessment of the extreme flood flow of small rivers, is based on the formula of limiting intensity [12]—the simplest implementation of the genetic theory of runoff formation. The characteristics of precipitation over short time intervals are still least known.

Methodologically, the authors of this study proceed from the hypothesis that the hydrometeorological time series are not distributed in accordance with some known statistical law, but are approximated by some statistical distribution [2, 4]. The agreement between the theoretical and the empirical distribution can be assessed, conventionally, with the use of statistical criteria. The authors have found that the series of shower intensity over short time intervals can be approximated by a two-parameter lognormal distribution (Caipatan).

The calculation of statistical characteristics based on the analysis of all maximums recorded within a year is aimed to extend the data series on storm precipitation (the series containing a single value within each year are often not enough for correct estimation) with the following evaluation of distribution statistics and their conversion into the statistics of series containing a single peak-peak.

PREVIOUS STUDIES

The use of peak-peak of the sum (h_i) or intensity (i_i) of precipitation in the calculation of flood flow over a calendar year has been fixed in the standards. Accord-

ing to [4], for time series containing several maximums per year (W -type samples), the evaluation of the exceedance probability of an event through conversion of W series to V form (V -type samples contain one event per year) has been originally associated only with the engineering tradition of statistical calculations in hydrometeorology. However, earlier, B.V. Gnedenko [3] has established a relationship between the properties of the original distribution $W(x)$ and the type of the limiting distribution $V(x)$: at large (small) x , the condition $\lim_{x \rightarrow \infty} [W(x)] = V(x)$ holds.

The issue of the correctness of the choice of a single peak-peak out of several extreme events over a calculation period T has been discussed in the literature since the 1970s [3, 4, 6]. The problem of the passage from the statistical parameters of the distribution of phenomena occurring several times a year (W -type samples), to the statistics of phenomena observed once a year (V -type samples), has been discussed in the literature since the early XX century [2, 4, 26, 27].

In this country, the time series of storms containing several extremums per year were studied by G.A. Alekseev [1], and S.N. Kritskii and M.F. Menkel' (1981) [6, 11].

The present-day world practice of the analysis of extreme values is often based on three types of Gumbel distribution (Gumbel distribution of generalized extreme values and Weibull distribution), the families of Pearson III-type and Halfen distribution, and generalized logistic and Pareto distributions. In the analysis of the series containing several events per year, the non-parametric two-component distribution and Wakeby distribution [9] are also used. The problems of assessing the series containing several events per year have been recently discussed in [19, 21–25].

In [11, 19], in the analysis of the series containing several events, a substantiation is given to the principle of calculating the maximum observed every year with a Poisson frequency in truncated series with exponential distribution.

MATERIAS AND METHODS

The source data were the observation series recorded from 1936 to 2015 at 192 weather stations in the territory of Ural DHMS, equipped with pluviographs. The analyzed characteristics were the intensity of a shower over short time intervals (<300 min) and the total precipitation (>10 mm) during all events under consideration.

The check of whether the kicks belong to a single general population (Dixon and Smirnov–Grubbs test) yielded positive results for extreme showers for all weather stations. The presence of kicks can be attributed to the shortness of the series and errors in the primary treatment of pluviograph data.

The series are independent, and individual showers have no genetic affinity. All empirical series were estimated for agreement with statistical distribution laws with the use of goodness-of-fit tests: Kolmogorov–Smirnov, Shapiro–Wilks, and Pearson χ^2 (chi-square).

The series of intensities over short intervals ($t < 300$ min) in 80% of cases can be described by a lognormal distribution (with a significance level of 5–45%) and in 20% of cases, by an exponential distribution (with a significance level of 5–11%). Lognormal distribution law was chosen to describe the time series of shower intensity in the territory of Ural DHMS.

The structure of source data on the showers allows several storms with often similar total volume and intensity to take place within a year. The use of data on all events is determined by the need to elongate the observation series, considering that regression analysis and analogy methods are inapplicable to the processing of data on storms.

As mentioned above, the approach to processing the data on several events per year and to passing to the statistics of series containing one peak-peak per year (i.e., to statistics of limit distributions) was developed by E. Gumbel and B.V. Gnedenko [2, 3, 9]. As applied to this study, the intensity is to be understood as the exceedance probability of an extreme intensity of a storm within a year out of the number of all such events in the year, and the frequency is to be understood as the number of such events in the year.

The exceedance probability function for a distribution containing one event in a year (for V sample) $P(v)$ is determined the exceedance probability function with n events in the year (for W sample) and can be expressed by the relationship included in [7] in the form:

$$P(v) = 1 - \exp(-P(w)n_w/n_v), \quad (1)$$

where n_v and n_w are the numbers of terms in the samples V and W , respectively.

This relationship, according to [11], holds for statistically homogeneous series, in which the annual number of events can be approximated by Poisson distribution. The statistical processing of series of the annual number of intense showers (Pearson's test of fit χ^2 was applied with a significance level of 0.05), the series in 80% can be approximated by the normal distribution and in 20%, by the Poisson distribution (in 80% of observations, the observed Pearson's statistics is 2–4 times the critical value). The homogeneity test for sample means and variances (Student's and Fisher's tests) yielded positive results for the series that can be approximated by the normal distribution and for 60% of normalized series that can be approximated by the Poisson distribution. These facts allow formula (1) to be used to evaluate the probability of events that occur several times per year, though with a large degree of conventionality.

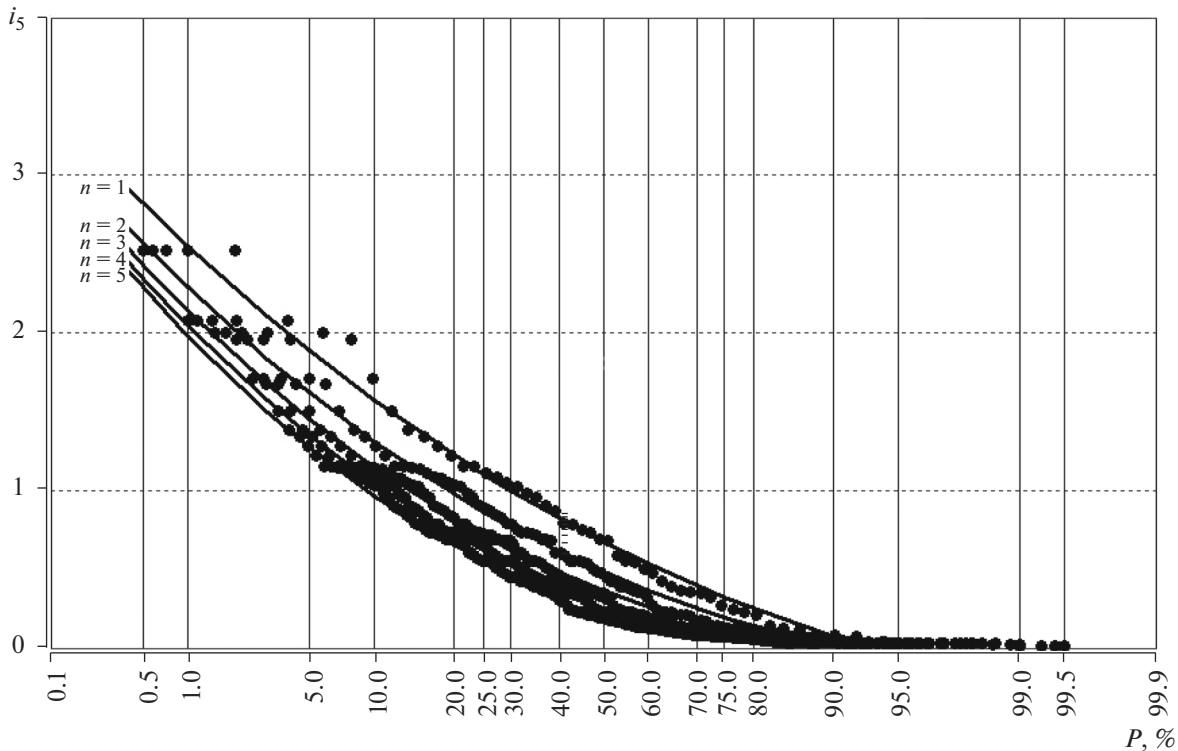


Fig. 1. Empirical curves of exceedance probability for series of maximal storm intensity over 5 min i_5 (mm/min), containing different number of maximums in year n (from 1 to 5) at weather station Verkhnee Dubrovo over period 1936–2015.

As noted in [4], the use of such simplified relationships leads to underestimation of the characteristics of rare exceedance probability and, accordingly, to overestimation of the exceedance probability when only data on peak-peaks is used and several maximums appear within the year. Contrary to that, the use of data on all events in a year without conversion of statistical parameters to the limiting distribution (one event per year) leads to even greater sinking of the exceedance probability curve (Fig. 1). In other words, the maximal and limiting values of distribution parameters (the mean and the standard deviation) can be reached only in the limit, i.e., at one event per year.

The analysis of the interrelationship between the statistical distribution parameters: the mean and the standard deviation for a normal (two-parameter) distribution of the values of i_{5i} —time series containing one event per year, with the parameters of the series containing n events per year, was carried out in the following order. First, logarithms of the independent terms of the time series were taken (this was made because the apparatus used to calculate the statistics of the lognormal distribution is cumbersome compared with the method of moments used in the case of the normal distribution). Out of the 192 analyzed observation points, data were taken for the pluviographs at which, during ≥ 20 years, in each calculation year, ≥ 5 storms took place with a limit intensity of

> 0.2 mm/min within a 5-minute time interval (observation data show that, in the territory under consideration, ≤ 8 storms with an intensity in excess of the specified value can take place within a year).

Beforehand, the extreme rains within each calculation year were ranked in the descending order of the maximal observed intensity within a 5-minute interval (in other words, the storms were combined by their ordinal number within a calendar year). Such ranking is possible as the time series of the characteristics of extremes are independent and there is no correlation within the series, as determined by the genesis of extreme precipitation (frontal or air-mass). Series containing simultaneously $n = 1, 2, 3, 4,$ and 5 events per year were compiled. The formation of time series containing n (from 1 to 5) events per year, was implemented by successively combining series of extreme intensities of a shower of the first order with series of extremums of the second, etc. orders. In this case, the chronological order of events within each calculation year was disturbed.

As the processing of samples containing several events in a year increases the probability of obtaining a heterogeneous sample, the homogeneity of the samples should be evaluated first. At the application of truncation procedure to series containing different number of events in a year, the truncation point ξ will be variable for both the series containing different

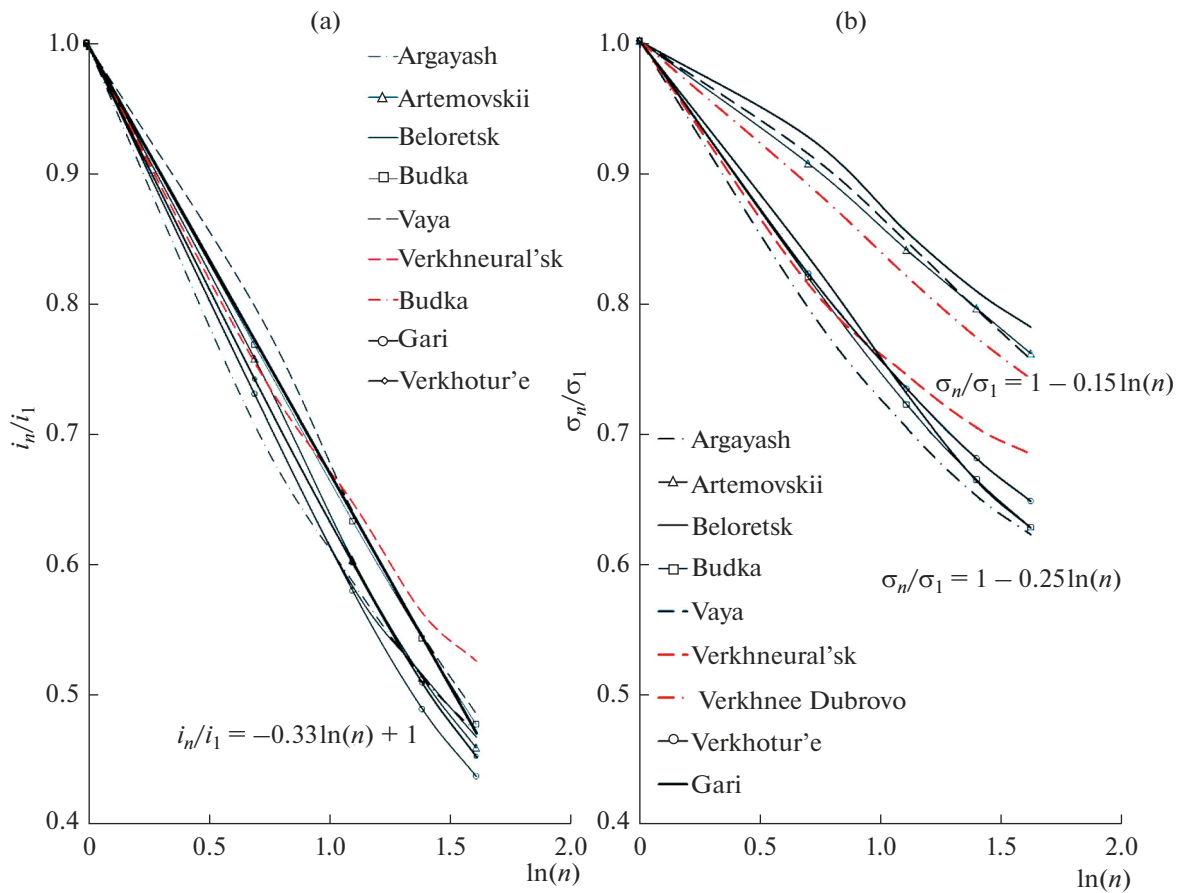


Fig. 2. Calculated nomograms: (a) the ratios of mathematical expectations of time series containing one (\bar{i}_1) and several events (\bar{i}_n) in a year (n varies from 2 to 5 events), vs. the logarithm of the number of events in the year $\bar{i}_n/\bar{i}_1 = f(\ln[n])$; (b) the ratios of RMSD of the series containing one (σ_1) and several (σ_n) events per year vs. the logarithm of the number of events in the year $\frac{\sigma_n}{\sigma_1} = f(\ln[n])$.

number of events for the same observation point and the different weather stations. Because of this, the procedure described here is developed for statistically homogeneous samples (or for samples converted into homogeneous by truncation). By their nature, the series being analyzed are independent, because individual storms have no genetic relationships with one another. The homogeneity tests for sample means and variances yielded positive results for 95% of the analyzed chronological samples, containing from 1 to 5 events in a year.

The method of moments was used to determine the statistical parameters of the series and to establish the dependences for the ratios

(1) of expectations of the time series containing one (\bar{i}_1) and several (\bar{i}_n) events in a year (n varies from 2 to 5 events) on the logarithm of the number of events in a year:

$$\frac{\bar{i}_n}{\bar{i}_1} = f(\ln[n]), \tag{2}$$

(2) root-mean-square deviations (RMSD) of the series containing one (σ_1) and several (σ_n) events in a year on the logarithm of the number of events in a year:

$$\frac{\sigma_n}{\sigma_1} = f(\ln[n]). \tag{3}$$

These relationships are represented by nomograms (Fig. 2). For the ratio (2), a single relationship was obtained in the form:

$$\frac{\bar{i}_n}{\bar{i}_1} = 1 - 0.33\ln[n]. \tag{4}$$

For the ratio (3), a nomogram was obtained described by an equation in the form:

$$\frac{\sigma_n}{\sigma_1} = 1 - a \ln[n], \quad (5)$$

where a is an empirical parameter taken equal to 0.15 for the series that show a coefficient of variation $C_{v1} > 1$ and 0.25 for other series.

Therefore, the value of σ_1 depends on both the number of events in a year and the value of C_{v1} , as shown by equation (5).

The obtained relationships and nomograms enable the passage from statistical parameters of series containing any number of events in a year to the parameters of series containing a single event in a year. The errors of such transformation, estimated by observation data, are $\leq 1.5\%$ for the mean and 5% for the RMSD of the series.

The scatter of the plots of the obtained relationships can be explained by the specific features of the annual distribution of the statistical parameters of storms and the limited observation series. The quantitative regularities in the distribution of statistical parameters of storms within a year have not been studied nor described in the literature. In other words, after the implementation of arrangement, we still cannot unambiguously establish the dependence of statistical parameters of the series containing one value for each storm n on the ordinal number of the event in the year.

To use the obtained nomograms in practice, it is reasonable to determine the mean number of storms in a year in the territory by long-term observation data. Such calculations have been carried out by the authors based on the data of pluviograph observations for ≥ 40 years over period from 1936 to 2015. The values of the mean number of storms were mapped (Fig. 3) to show a regular increase in the mean frequency of such events in the mountain region of the Urals and in some areas in its eastern piedmonts.

The appropriateness of averaging the number of storms in a year by a group of weather stations operating under similar conditions (at a short observation period) is confirmed by the existence of relationships $i_n = f(\ln[n])$ and $\sigma_n = f(\ln[n])$ for all analyzed weather stations without ranking the events within each year, and, when all events in the year taken into account—by each weather station (one point—one weather station). The relationships with data ranked over storms within a year should be considered more reliable. As mentioned above, the series of the number of single storms over a year for the weather stations under consideration can be approximated by a normal distribution. Because of this, the characteristic of the mean number of storms in a year is a good characteristic of the center of the distribution of these series. The calculation algorithm involves the formation of samples for all storms in a year with a rate > 10 mm per 1 h (whatever the number of events in each calculation year) without sampling a certain number of events in each year. The use of the mean number of storms in a

year makes it possible to determine the limiting values of the mean annual precipitation and the RMSD based on the formed series of maximal precipitation rates.

The obtained relationships and cartograms can be used to evaluate the statistical parameters of the limiting distributions based on data on storms with any number of events per year. In the presented form, the relationships and the cartogram have been used to convert the statistical parameters of the distribution of storm intensity for weather stations from the case of several events per year to the case of a single event in a year.

In the practical calculations, the following algorithm is recommended for processing short series (< 10 years):

(1) The data of pluviographic observations are used to form a sample of all single storms with a rate > 10 mm/h;

(2) The maximal rates of a storm within 5-min intervals are calculated (the passage to the rates over intervals with other length or to the total precipitation over a storm event can be made by reduction curves given in [5]);

(3) After homogeneity tests, statistical parameters of series containing several peaks per year are determined;

(4) The developed relationships (Fig. 2) are used to recalculate the statistical characteristics of the series containing several events to series containing one event per year, as is required in the engineering practice. At the initial length of a series with one event per year ranging from 10 to 15 years, an increase in the series length by a factor of 2–4 through the incorporation of all events in a year minimizes the mean square error in the estimate of the mean from 35 to 23–16%; and the error in RMSD, from 90 to 55–30%. Considering that the errors associated with the passage from statistics of multiextremum series to series with one event per year are not greater than 1.5% for the mean and 5% for the RMSD of the series, the use of all phenomena appears to be an effective method for obtaining reliable estimates of statistical parameters of the series.

DISCUSSION

Three approaches are now in use in the practice of hydrological calculations of the characteristics of storm precipitation events with rare occurrence:

(1) the analysis of outliers (in the foreign literature—method of maximization) and the moments of distributions associated with them [19];

(2) the use of the distributions of extremums (limit distributions) [2];

(3) the use of one-side-truncated distributions (in which the analysis is focused on the tails of integral

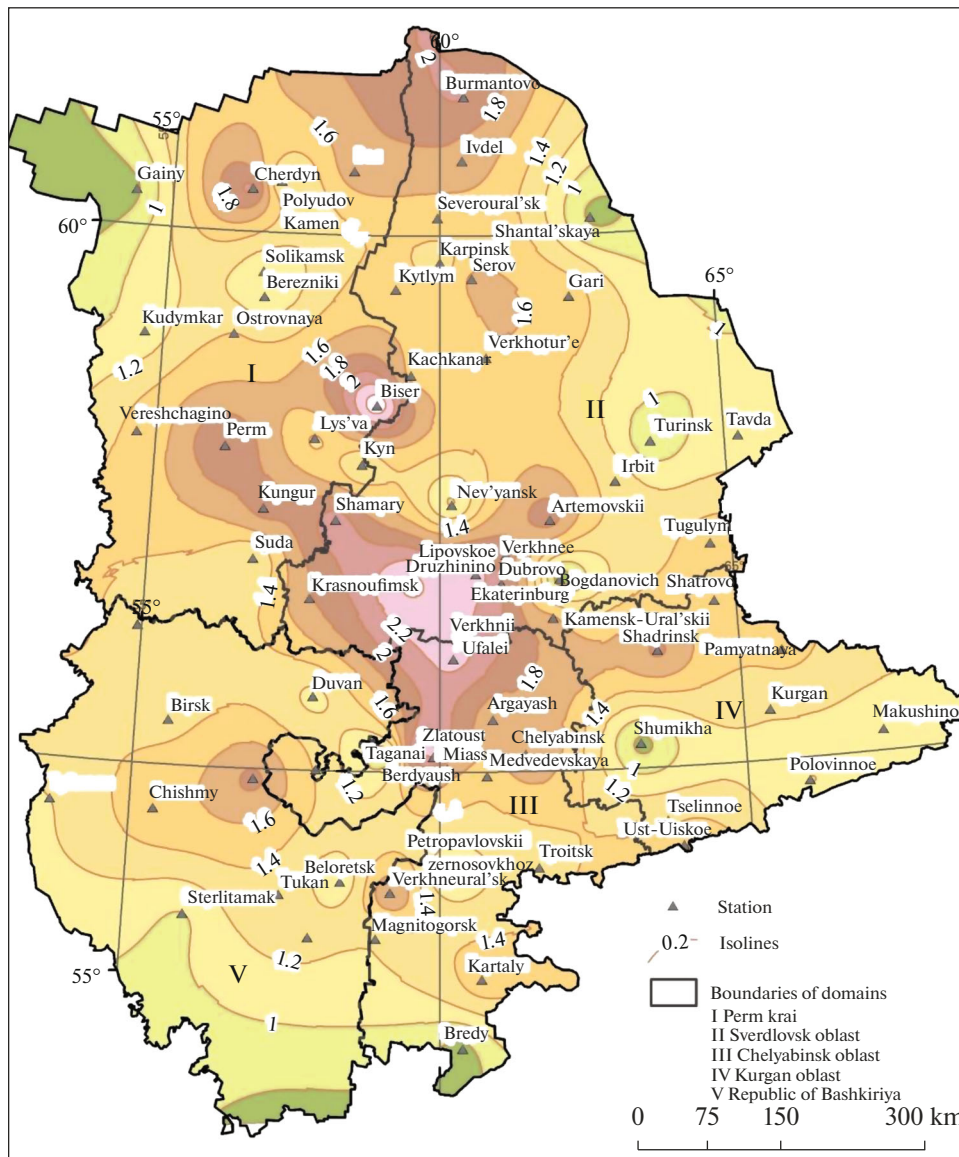


Fig. 3. Calculated map of the mean number of storms with the amount of precipitation >10 mm/h in the territory of Ural DHMS and Bashkiriya.

distributions rather than their near-mode part, as is common in the mathematical statistics) [8].

In the overwhelming majority of studies, these approaches are presented as independent methods of statistical analysis; however, in the form used in practice, they are particular cases of a realization of Gumbel limit distribution.

The second approach was developed in [1] as applied to the analysis of data on rains and extreme water discharges during rain floods, and in the recent decades, it was rarely mentioned in the studies. The present-day trend to the passage to deterministic models of river runoff demonstrates the need to use data on all extreme characteristics of storms in a year when calculating the flow of rain floods.

In [8], the exceedance probability of a truncated distribution is given in the form:

$$P(w) = P(v)P(\xi), \quad (6)$$

where $P(w)$ and $P(v)$ are distribution functions of the full and truncated samples, respectively (in the accepted denotations, W is the full sample and V is a truncated sample), $P(\xi)$ is the exceedance probability in the truncation point ξ (clearly, $P(\xi) \approx n_v/n_w$, if we rely on the volumes of samples). In other words, in the system of data processing by several event in a year, accepted by the authors, only peak-peaks would represent n_v , while n_w would be represented by all other maximums, corresponding to 2, 3, etc. events in each year. The complete identification of the technique of

truncated and limit distribution is complicated by the fact that the set of maximums relating, for example, to the first and second events in a year are, most often, overlapping over the long-term observation period. At the same time, the authors believe that combining the techniques of the use of truncated and limit distributions is a promising task in the studies of the statistics of phenomena containing several events in a year. In [19], such combination has led to the development of a two-component distribution, which can be regarded as the maximum of two extremums of different orders (the first and the second in the year) in truncated series, each having a Poisson frequency and exponentially distributed values of maximums.

CONCLUSIONS

In this study, the authors propose an approach to determining statistical parameters of distributions with any number of events per year, adapted to engineering practice. The material for the study was the data on storm intensity over short time intervals. A scheme is proposed for arrangement storms within year. The frequency of storms is mapped for the Ural territory, thus making it possible to determine and zone the transition coefficients for conversion of the statistics of distributions with any number of events in a year to statistics of limit distributions with the use of proposed nomograms.

All calculations have been carried out for lognormal distribution, which gives best approximation of the series of storm intensities.

Previously, the authors have established relationships between the statistics of phenomena for one or several events per year, grouped for all weather stations and based on the data on the mean number of storms in a year with medium intensity over 5 min and the coefficient of variation of storm intensity. For the mean values, a relationship has been obtained coinciding with (4) (the calculation error of \bar{i}_1 based on \bar{i}_n never exceeded 2.3%; the procedure proposed in this study reduced the error to 1.5%). In what regards RMSD, the relationship of the type (5) allows the error in σ_1 to be reduced to 4.8% (compared with 20%, obtained before). The relationships obtained by the arrangement of storms for each weather station considerably improve the accuracy of calculations.

The perspectives of the further development of the proposed procedure for calculating the parameters of the limit distribution of storm intensity based on the frequency of storms per year are related with the studies of the number of storms within a year and the frequency of their appearance in the months of the warm season.

REFERENCES

1. Alekseev, G.A., Evaluating the probability of hydrological and climatological events, which occur several times a year, *Tr. Gl. Geofiz. Obs. im. A.I. Voeikova*, 1954, vol. 43, no. 97, pp. 106–112.
2. Gumbel, E., *Statistics of Extremes*, Columbia Univ. Press.: New York, 1962.
3. Johnson, N.L., Kotz, S., and Balarkishnan, N., *Odnomernye nepreryvnye raspredeleniya* (Continuous Univariate Distributions), Part 2., *Teoriya veroyatnostnykh raspredelenii* (Theory of Probability Distributions), Moscow: BINOM, 2010.
4. Kartvelishvili, N.A., *Stokhasticheskaya gidrologiya* (Stochastic Hydrology), Leningrad: Gidrometeoizdat, 1975.
5. Klimenko, D.E., Eponchintseva, D.N., Korepanov, E.P., and Cherepanova, E.S., Studying the reduction curves of flood-forming storm precipitation in Transuralia, *Meteorol. Gidrol.*, 2018, no. 2, pp. 76–89.
6. Kritskii, S.N. and Menkel', M.F., *Gidrologicheskie osnovy upravleniya rechnym stokom* (Hydrological Principles of River Runoff Control), Moscow: Nauka, 1981.
7. *Mezhdunarodnoe rukovodstvo po metodam rascheta osnovnykh gidrologicheskikh kharakteristik* (International Guide on the Methods for Calculating Main Hydrological Characteristics), Leningrad: Gidrometeoizdat, 1984.
8. Ratkovich, D.Ya. and Bolgov, M.V., *Stokhasticheskie modeli kolebanii sostavlyayushchikh vodnogo balansa rechnogo basseina* (Stochastic Models of Variations of Water Balance Components in a River Basin), Moscow: Inst. Vod. Probl., Ross. Akad. Nauk, 1997.
9. *Rukovodstvo po gidrologicheskoi praktike. Sbor i obrabotka dannykh, analiz, prognozirovaniye i drugie primeneniya* (Guide on Hydrological Practice. Data Collection and Processing, Analysis, Forecasting, and Other Applications), World Meteorological Organization, WMO-168, 1994. 1997.
10. Hald, A., *Statistical Theory with Engineering Applications*, New York–London, 1952.
11. Khristoforov, A.V., Kruglova, G.V., and Samborskii, T.V., *Stokhasticheskaya model' kolebanii rechnogo stoka v pavo-dochnyi period* (A Stochastic Model of River Runoff Variations during Floods), Moscow: Mosk. Gos. Univ., 1998.
12. Chebotarev, A.I. and Serpik, B.I., Choice and substantiation of formulas for calculating maximal discharges of rain floods, in *Sb. rabot po gidrologii* (Coll. Works in Hydrology), Leningrad: Gidrometeoizdat, 1973, issue 11, pp. 3–47.
13. *Comprehensive Risk Assessment for Natural Hazards*, World Meteorological Organization. WMO/TD-No. 955. 1999.
14. *Estimation of Maximum Floods*, World Meteorological Organization, WMO-No. 233, TP 126, Techn. Note No. 98, Geneva, 1969.
15. Hershfield, D.M., Method for estimating probable maximum rainfall, *J. American Waterworks Association*, vol. 57, August, 1965, pp. 965–972.
16. Hershfield, D.M., *Rainfall frequency atlas of the United States for durations from 30 minutes to 24-hours and return periods from 2 to 100 years*, Techn. Paper 40,

- Washington, DC: US Weather Bureau, 1961, pp. 400–440.
17. *Intercomparison of models of snowmelt runoff*. Operational Hydrology Report № 23. WMO Publ. № 646. Geneva: World Meteorological Office, 1986. 440 p.
 18. *Manual for Depth–Area–Duration Analysis of Storm Precipitation*, World Meteorological Organization. WMO-No. 237, Geneva, 1969.
 19. *Manual for Estimation of Probable Maximum Precipitation*, World Meteorological Organization, Operational Hydrology Re. No.1, WMO-No. 332, Geneva, 1986.
 20. Miller, J.F., *Physiographically Adjusted Precipitation–Frequency Maps: Distribution of Precipitation in Mountainous Areas*, WMO, no. 326 (11), 1972, pp. 264–277.
 21. Pilgrim, D.M. and Cordery, I., *Flood runoff, Handbook of Hydrology*, New York, USA: McGraw-Hill, 1993.
 22. Pilgrim, D.M. and Cordery, I., Rainfall temporal patterns for design floods, *ASCE J. Hydraulic Engineering*, 101 (HY1), 1975, pp. 81–95.
 23. Pilgrim, D.H. and Doran, D.G., Practical criteria for the choice of method for estimating extreme design floods, *IAHS. Publ. no. 213*, Wallingford, UK: Inst. Hydrology, 1993.
 24. Pilgrim, D.H., *Australian Rainfall and Runoff. A Guide to Flood Estimation*, Canberra: Inst. Engineers Australia, 1998.
 25. Sevruk, B. and Geiger, H., *Selection of Distribution Types for Extremes of Precipitation*, World Meteorological Organization. Operational hydrology rep. no. 15. WMO-No. 560, Geneva, 1981.
 26. Todorovic, P. and Woolhiser, D.A., Stochastic structure of the local pattern of precipitation, *Stoch. Approach to Wat. Res.*, 1976, vol. 2, pp. 217–222.
 27. Todorovic, P. and Yevjevich, V., Stochastic processes of precipitations, *Colorado State Univ. Hydro. Paper*, 1969, vol. 35, pp. 1–61.

Translated by G. Krichevets