EFFECT OF THE DATA LENGTH AND SEASONALITY ON THE ACCURACY OF T-YEAR DISCHARGES ESTIMATION: CASE STUDY ON THE TOPĽA RIVER

Veronika Bačová Mitková

The paper deals with the effect of two factors on the accuracy of T-year discharge estimation resp. fluctuations in the estimation of these discharges. As input data the series of daily discharges and peak discharges on the Topľa River at Hanušovce nad Topľou for the period of 1931–2015 were used. To estimate the T-year maximum discharges, the maximum annual discharges (AM) method was used with theoretical probability distributions that are among the most widely used in Slovak hydrological practice (Log-Pearson III, Gamma and Log-Normal). We analysed the effect of the time series length and the effect of seasonality (winter, summer) on the accuracy of T-year maximum discharges estimation.

KEY WORDS: the Topľa River, AM method, peak discharge, data length, seasonality, T-year discharge

Introduction

Flood frequency analysis plays a major role in the design of hydraulic structures and flood control management. One way of estimating the design discharges is the flood frequency analysis and solution of the relationship between peak discharges of the flood waves and probability of their return period (T). Directive 2007/60/EC of the European Parliament of 23 October 2007 concerning the assessment and management of flood risks requires member States to draw up flood hazard maps of floods with very long return periods T (500 to 1000 years). All methods of estimating floods with a very long return period are associated with great uncertainties. Determining the specific value of a 500- or 1000-year flood for engineering practice is extremely complex. Nowadays hydrologists are required to determine not only the specific design value of the flood, but it is also necessary to specify confidence intervals in which the flow of a given 100-, 500-, or 1000-year flood may occur with probability, for example, 90%. The correct estimations of potential culminations of floods require the inclusion of the longest data series of observations, as well as the inclusion of historic pre-instrumental data to statistically analysed data series (Gaal et al., 2010; Elleder et al., 2013; Kjeldsen et al., 2014). Brazdíl et al. (2006) studied historic hydrological materials in order to estimate floods threat in Europe. Estimation of the uncertainty at the design discharges was investigated for example by Merz and Thieken (2009) or Rogger et al. (2012). In addition to this factor, the type of theoretical probability distribution that is used to estimate maximum (extreme) values has an impact on the estimation of T-year discharges. Bačová (2019) compared the two most commonly used methods in estimating T-year maximum discharges, AM method and POT method. The author analysed effect of the threshold level value and various data set (peaks, mean daily discharges) on estimated values of Q_T. In El Adlouni et all (2008), Malamud and Turcotte showed that, the most commonly used distributions in hydrology can be divided into four groups: the normal family (normal, Lognormal), the general extreme value family (GEV, Gumbel, Fréchet, reverse Weibull), the Pearson type III family (Gamma, Pearson type III, Log-Pearson type III), and the Generalized Pareto distribution. In practice, all these models are fitted to data and compared using conventional goodness-of-fit tests. Having a data set of maximum annual values of discharges different statistical tests, like Kolmogorov-Smirnov, Anderson-Darling and Chi-Squared tests (Ang and Tang, 2007) are used to select the suitable continuous distribution. When the sample volume is not very large, the volume can be extended by numerical simulation of random variable based on the inverse method. The maximum discharges Q_{T,95} corresponding to the probability of exceedance P% are not unique values, but they depend on aleatory and epistemic uncertainty (Merz and Thieken, 2009). Aleatory uncertainty is mainly due to the time variability and the length of the maximum discharges series, while the epistemic uncertainty is the consequence of the incomplete knowledge of the hydrological system. The paper presents an estimation of the T-year maximum dis-
charges by the AM method and analyses the effect of the time series length and seasonality (winter, summer) on the accuracy of $T$-year maximum discharges estimation. In this approach Log-Pearson theoretical probability distribution type III was used. Subsequently, estimated $T$-year maximum discharges were compared with other two theoretical distribution types used in Slovakia: Log-normal and Gamma probability distribution. The set of daily discharges and peak discharges on the Topľa River at Hanušovce nad Topľou for the period of 1931–2015 was used as input data for our case study.

Method and material

Annual maximum discharges method (AM)

In estimating $T$-year maximum discharges, the annual maximum method (AM) is generally the first and most widely used method. It aims to estimate the $Q_T$ quantiles, it means such annual maximum discharges that their probability of exceedance is $1/T$, where $T$ can be e.g. 10, 20, 50, and 100, 500 or 1000 or more years. These quantiles are determined from the distribution function of the maximum annual discharges. Thus, $Q_T$ is the maximum discharges with a probability $P$ that occurs, on average, once every $T$-years. When interpreting the results, it should be borne in mind that the estimated values of $T$-year maximum discharges with very high return periods are extrapolated and that each statistical method is burdened with some uncertainty. Different types of theoretical distributions are used to evaluate the $T$-year maximum discharges. Appropriate choice of the theoretical probability distribution type should represent quite precisely the uncertainty and variability of the problem.

In the world literature, there are a number of scientific papers dealing with the selection and testing of the suitability of theoretical probability distributions in estimating the maximum values of hydrological characteristics. In our analysis we use one type of the theoretical probability distribution the Log-Pearson distribution type III (LP III). The advantage of this particular technique is that extrapolation can be made of the values for events with return periods well beyond the observed flood events. To estimate the distribution parameters, the method described in Bulletin17B was used. Bulletin 17B was issued in USA in 1981, and re-issued with minor corrections in 1982 in the Center for Research in Water Resources of the University of Texas at Austin (IACWD, 1982). Bulletin 17B provided revised procedures for weighting station skew values with results from a generalized skew study, detecting and treating outliers, making two station comparisons, and computing confidence limits about a frequency curve. Bulletin 17B is based on Bulletins 15, 17, 17A (http://iacwi.gov/hydrology/Frequency/minutes/index.html). The cumulative distribution function and probability distribution function according Hosking and Wallis (1997) are defined as:

\[
F(x) = G\left(\frac{x - \xi}{\beta}\right)/\Gamma(\alpha),
\]

\[
f(x) = \frac{\left(x - \xi\right)^{\alpha - 1}e^{-\left(x - \xi\right)/\beta}}{\beta^{\alpha}\Gamma(\alpha)}
\]

If $\gamma < 0$ then

\[
F(x) = 1 - G\left(\frac{x - \xi}{\beta}\right)/\Gamma(\alpha)
\]

\[
f(x) = \frac{\left(x - \xi\right)^{\alpha - 1}e^{-\left(x - \xi\right)/\beta}}{\beta^{\alpha}\Gamma(\alpha)}
\]

where

$\mu$ – location parameter;
$\sigma$ – scale parameter;
$\gamma$ – shape parameter;
$\Gamma$ – Gamma function.

Subsequently, the LP III probability distribution was compared with other recommended probability distributions (Gamma and Log-normal) according to OTN ŽTP 3112:1: 03. To verify the accuracy of theoretical distributions, we used a non-parametric Kolmogorov-Smirnov goodness of fit test for the significance level $\alpha=0.05$.

Topľa River basin and input data

The Topľa is upland/lowland type of river in eastern Slovakia. The catchment drainage area is 1 506 km² with length of 129.8 km (Fig. 1). The long-term mean daily discharge amounts in Hanušovce nad Topľou was 8.1 m³ s⁻¹ during period 1931–2015 (runoff height was 244.2 mm). The maximum discharge during the analysed period was 449 m³ s⁻¹ (06.04.1932) in the station Hanušovce nad Topľou. Figure 1 also shows the exceeding probabilities of the maximum annual discharges according to Log-Pearson Type III. probability distribution (LP III). The advantage of this particular technique is that extrapolation can be made of the values for events with return periods well beyond the observed flood events. This theoretical distribution belongs to the family of Pearson distributions, so called three parametric Gamma distributions, with logarithmic transformation of the data. We compared the LP III distribution with the theoretical probability distributions that were (and still are) most widely used hydrological practice in Slovakia: Gamma distribution and Log-normal distribution Table 1. From Table 1 we can see relatively small differences in the values of estimated $T$-year maximum discharge values in comparison with other two types of theoretical probability distributions used in hydrological analyses of extremes in the Slovakia. The lowest values of estimated $T$-year maximum discharges, achieved Gamma theoretical probability distribution, especially for discharges with high return periods.

Peak annual discharges (points), linear trend (red line), and 4-years moving averages for the Topľa River at Hanušovce nad Topľou during the period 1931–2015 are shown in Figure 2. In the analysed period, two dry periods of 1954–1964 and 1990–1999 were occurred.
While wet periods can be described only as years with extreme flood events (e.g., 1932, 1948, 1952, or 1980), a relatively prolonged wet period was in 2004–2010. Annual maximum discharges show a decreasing trend for the period of 1931–2015.

Results

**The effect of time series length on the T-year discharge estimation**

For analysing the effect of the length of the data series on the estimation of $T$-year discharges, the period 1931–2015 was divided into two shorter periods: 1931–1973 and 1974–2015. We had chosen this approach because for the frequency analysis is recommended the length of the observation series $5T$ (FEH, 1999). If $T=50$ years, then a 250-member observation series is required for a reliable estimate of $Q_{50}$. Such a length of data series (AM) is practically absent. Therefore, the probability of a reliable estimate of $T$-year maximum discharge for short-range river basins is relatively low. In the case of the 50-year observation series, the probability of $Q_{100}$ is 39% and in the case of the 100-year series is 63% (Viessman et al., 1977). The estimated values of the $Q_{T_{\text{max}}}$ for shorter periods of the data series are listed in Table 2. There is compared the LPIII distribution with other frequently used and recommended hydrological distributions in hydrological practice in Slovakia: Log-normal distribution and Gamma distribution. The exceedance probabilities of the annual peak discharges for two shorter periods of the Topľa River at Hanušovce nad Topľou according the LPIII distribution are presented in Figure 3a-b.

**The effect of the seasonality on the T-year discharge estimation**

For dividing the year into seasons, we proceeded from the analysis of the occurrence of floods and from the evaluation of the Topľa runoff regime during the year. In terms of the type of runoff regime, Topľa belongs to the highland-lowland area with rain-snow runoff with the culmination of river runoffs in the month of March, respectively April. The distributions of the mean monthly discharges in 10-year periods are presented in Figure 4a-b. The Figure 4a-b shows that in some 10-year periods lower values of mean monthly discharges were achieved in months of occurrence of high water levels. The long-term mean monthly discharge in the month of March reached value $11.82 \, m^3/s$ and in month of September reached value $3.25 \, m^3/s$ during the period of 1931–2015. The occurrence of annual maximum discharges is presented in Figure 5. The number

![Fig. 1. A scheme of the Topľa River basin (left) and exceedance probability of the annual peak discharges of the Topľa River: Hanušovce nad Topľou within 1931–2015 period (right).](image1)

![Fig. 2. Peak annual discharges (points), linear trend (red line), and 4-years moving averages for the Topľa River at Hanušovce nad Topľou during the period 1931–2015.](image2)
of annual peak discharges during the period of 1931–2015 occurred in the month of March was 31 peaks (Fig. 5). In terms of the Topľa runoff regime, the measured data were divided into two seasons:

- Summer season is from May to October, when peak discharges occur only from heavy rainfall (Fig. 6a).
- Winter season is from November to April, when peak discharges occur by combining heavy rainfall in the form of snow and rain as well as snow melting in the area (Fig. 6b).

The statistical data series were supplemented with maximum discharges in the given season, so that there are 85 measurements per season.

The estimated values of the $Q_{T \text{max}}$ for summer season and winter season are listed in Table 3. There is compared the LPIII distribution with other frequently used and recommended hydrological distributions in hydrological practice in Slovakia: Log-normal distribution and Gamma distribution. The exceedance probabilities of the maximum seasonal discharges of the Topľa River at Hanušovce nad Topľou according the LPIII distribution are presented in Figure 7a-b.

Comparisons of the estimated maximum discharges with a return period of 100 and 1000 years according to the selected procedures are shown in Figure 8. The highest estimated values of $Q_T$ were achieved according the LPIII distribution.

Table 1. $T$-year maximum discharges $Q_{T \text{max}}$ [m$^3$ s$^{-1}$] on the Topľa River: Hanušovce nad Topľou (1931–2015) estimated from maximum annual discharges $Q_{T \text{max}}$

| $T$ [year] | 2 | 5 | 10 | 50 | 100 | 200 | 500 | 1000 |
|-----------|---|---|----|----|-----|-----|-----|------|
| $P$ [%]   | 39| 18| 9.5| 2  | 1   | 0.5 | 0.2 | 0.1  |
| Log-Pearson III |
| $Q_{T \text{max}}$ [m$^3$ s$^{-1}$] | 139| 193| 249| 398| 473 | 556 | 679 | 783 |
| $P_{\text{value}}$ K-S | 0.999 |
| Log-normal |
| $Q_{T \text{max}}$ [m$^3$ s$^{-1}$] | 146| 199| 252| 380| 439 | 502 | 589 | 660 |
| $P_{\text{value}}$ K-S | 0.987 |
| Gamma |
| $Q_{T \text{max}}$ [m$^3$ s$^{-1}$] | 148| 204| 255| 360| 404 | 446 | 506 | 541 |
| $P_{\text{value}}$ K-S | 0.744 |

Table 2. Comparison of the estimated $Q_{T \text{max}}$ for shorter periods on the Topľa River: Hanušovce nad Topľou

| $T$ [year] | 2 | 5 | 10 | 50 | 100 | 200 | 500 | 1000 |
|-----------|---|---|----|----|-----|-----|-----|------|
| $P$ [%]   | 39| 18| 9.5| 2  | 1   | 0.5 | 0.2 | 0.1  |
| Log-Pearson III |
| $Q_{T \text{max}}$ [m$^3$ s$^{-1}$] | 143| 198| 257| 427| 519 | 627 | 797 | 949 |
| $P_{\text{value}}$ K-S | 0.813 |
| Log-normal |
| $Q_{T \text{max}}$ [m$^3$ s$^{-1}$] | 161| 202| 249| 357| 408 | 457 | 527 | 583 |
| $P_{\text{value}}$ K-S | 0.787 |
| Gamma |
| $Q_{T \text{max}}$ [m$^3$ s$^{-1}$] | 169| 219| 272| 321| 428 | 472 | 529 | 571 |
| $P_{\text{value}}$ K-S | 0.44 |

| $T$ [year] | 2 | 5 | 10 | 50 | 100 | 200 | 500 | 1000 |
|-----------|---|---|----|----|-----|-----|-----|------|
| $P$ [%]   | 39| 18| 9.5| 2  | 1   | 0.5 | 0.2 | 0.1  |
| Log-Pearson III |
| $Q_{T \text{max}}$ [m$^3$ s$^{-1}$] | 125| 180| 239| 408| 499 | 602 | 760 | 899 |
| $P_{\text{value}}$ K-S | 0.74 |
| Log-normal |
| $Q_{T \text{max}}$ [m$^3$ s$^{-1}$] | 137| 181| 233| 362| 423 | 488 | 580 | 654 |
| $P_{\text{value}}$ K-S | 0.74 |
| Gamma |
| $Q_{T \text{max}}$ [m$^3$ s$^{-1}$] | 144| 189| 236| 336| 376 | 416 | 467 | 505 |
| $P_{\text{value}}$ K-S | 0.55 |
Fig. 3. The exceedance probabilities of the annual maximum discharges according to Log-Pearson Type III probability distribution (LPIII) a) period 1931–1973 and b) period 1974–2015 on the Topľa River: Hanušovce nad Topľou.

Fig. 4. The distributions of the mean monthly discharges in 10-year periods of the Topľa River: Hanušovce nad Topľou.
Fig. 5. Month of the annual maximum discharges occurrence (period of 1931–2015) on the Topľa River at Hanušovce nad Topľou.

Fig. 6. Occurrence of extreme discharges in month of the summer season (left) and winter season (right), with supplemented maximum discharges in a given season (number of measurements is still 85 per season).

Table 3. Comparison of the estimated $Q_{T_{\text{max}}}$ in selected two season a) summer and b) winter on the Topľa River: Hanušovce nad Topľou

| Season          | $T$ [year] | $P$ [%] | $Q_{T_{\text{max}}}$ [m$^3$s$^{-1}$] | $P_{\text{value K-S}}$ |
|-----------------|------------|---------|-------------------------------------|-------------------------|
| Summer season   | 2          | 39      | 68                                  | 0.953                   |
|                 | 5          | 18      | 124                                 |                          |
|                 | 10         | 9.5     | 194                                 |                          |
|                 | 50         | 2       | 423                                 |                          |
|                 | 100        | 1       | 558                                 |                          |
|                 | 200        | 0.5     | 718                                 |                          |
|                 | 500        | 0.2     | 975                                 |                          |
|                 | 1000       | 0.1     | 1209                                |                          |
|                 |            |         | Log-Pearson III                      |                         |
|                 |            |         | $Q_{T_{\text{max}}}$ [m$^3$s$^{-1}$] | $P_{\text{value K-S}}$ |
|                 |            |         | 75                                  | 0.922                   |
|                 |            |         | 124                                 |                          |
|                 |            |         | 193                                 |                          |
|                 |            |         | 422                                 |                          |
|                 |            |         | 555                                 |                          |
|                 |            |         | 714                                 |                          |
|                 |            |         | 971                                 |                          |
|                 |            |         | 1202                                |                          |
|                 |            |         | Log-normal                           |                         |
|                 |            |         | $Q_{T_{\text{max}}}$ [m$^3$s$^{-1}$] | $P_{\text{value K-S}}$ |
| Winter season   | 2          | 39      | 86                                  | 0.79                    |
|                 | 5          | 18      | 134                                 |                          |
|                 | 10         | 9.5     | 189                                 |                          |
|                 | 50         | 2       | 318                                 |                          |
|                 | 100        | 1       | 373                                 |                          |
|                 | 200        | 0.5     | 428                                 |                          |
|                 | 500        | 0.2     | 501                                 |                          |
|                 | 1000       | 0.1     | 556                                 |                          |
|                 |            |         | Gamma                                |                         |
|                 |            |         | $Q_{T_{\text{max}}}$ [m$^3$s$^{-1}$] | $P_{\text{value K-S}}$ |
|                 |            |         | 117                                 | 0.94                    |
|                 |            |         | 172                                 |                          |
|                 |            |         | 228                                 |                          |
|                 |            |         | 372                                 |                          |
|                 |            |         | 443                                 |                          |
|                 |            |         | 518                                 |                          |
|                 |            |         | 626                                 |                          |
|                 |            |         | 715                                 |                          |
|                 |            |         | Winter season                         |                         |
|                 |            |         | $Q_{T_{\text{max}}}$ [m$^3$s$^{-1}$] | $P_{\text{value K-S}}$ |
|                 |            |         | 108                                 | 0.73                    |
|                 |            |         | 151                                 |                          |
|                 |            |         | 204                                 |                          |
|                 |            |         | 347                                 |                          |
|                 |            |         | 419                                 |                          |
|                 |            |         | 498                                 |                          |
|                 |            |         | 613                                 |                          |
|                 |            |         | 710                                 |                          |
|                 |            |         | Gamma                                |                         |
|                 |            |         | $Q_{T_{\text{max}}}$ [m$^3$s$^{-1}$] | $P_{\text{value K-S}}$ |
|                 |            |         | 115                                 | 0.78                    |
|                 |            |         | 159                                 |                          |
|                 |            |         | 208                                 |                          |
|                 |            |         | 315                                 |                          |
|                 |            |         | 360                                 |                          |
|                 |            |         | 404                                 |                          |
|                 |            |         | 462                                 |                          |
|                 |            |         | 504                                 |                          |
Fig. 7. Exceedance probabilities of the maximum discharges according to Log-Pearson Type III probability distribution (LPIII) a) summer season and b) winter season on the Topľa River: Hanušovce nad Topľou.

Fig. 8. Comparisons of the estimated maximum discharges with a return period of 100 and 1000 years according to the selected procedures.
Conclusions and discussion

The first part of the paper deals with the estimation of the $Q_T$ from annual peak discharges on the Topľa River at Hanušovce nad Topľou (1931–2015). Results of this part showed relatively small differences in the values of estimated $T$-year maximum discharge values in comparison with other two types of theoretical probability distributions used in hydrological analyses of extremes in the Slovakia. The lowest values of estimated $T$-year maximum discharges, achieved Gamma theoretical probability distribution, especially for discharges with high return periods. Phien and Jivajirajah (1984) dealt with the use of the Log-Person III distribution to estimate the maximum annual rainfall and discharges. They concluded that this distribution is more suitable for discharges with a higher return period, but for the annual floods the existence of an upper bound of the distribution, in some cases may cause some higher uncertainties. Comparing the suitability of several types of probability distributions (GEV, LPIII and Gumbel) for estimating $T$-year discharges was discussed in Millington et al. (2011). Authors do not prefer any of the selected distributions as better and recommend further research. Results of our analysis indicate that the LPIII distribution is suitable distribution for $T$-year discharge estimation with a higher return period. Estimation of flood magnitudes to be used as a basis to design the hydraulic structures and flood control management is therefore of crucial importance. Therefore the paper also presented an estimation of the $T$-year maximum discharges by the AM method and analysed the effect of the time series length and seasonality (winter, summer) on the accuracy of $T$-year maximum discharges estimation. Results showed that not only the selection of the distribution function to estimate $T$-year discharges but also the processing of the statistical series affect the results of the estimation. The shorter periods showed higher estimations of the $T$-year discharges. The highest estimated values according the LPIII distribution was achieved for summer season. The lowest estimated value according the LPIII distribution was achieved for winter season. When interpreting the results, it should be borne in mind that the $T$-year maximum discharges are related to the length of the analysed series and therefore estimated values with very high return periods are extrapolated and that each statistical method is burdened with some uncertainty that may be caused by alone method, but also the data, which may be burdened by a certain measurement error.

Acknowledgements:

This work was supported by the project VEGA 20004/19 and results from the project implementation of the “Centre of excellence for integrated flood protection of land” (ITMS 26240120004) supported by the Research & Development Operational Programme funded by the ERDF.
