ГОДИШНИК НА УНИВЕРСИТЕТА ПО АРХИТЕКТУРА, СТРОИТЕЛСТВО И ГЕОДЕЗИЯ СОФИЯ

Том Volume 53 | 2020 | Брой Issue 1

ANNUAL OF THE UNIVERSITY OF ARCHITECTURE, CIVIL ENGINEERING AND GEODESY SOFIA

Получена: 20.12.2019 г.
Приета: 22.01.2020 г.

THE EFFECT OF MIXED PEAK DATA ON THE FLOOD QUANTILE ESTIMATES IN A SINGLE STATION ANALYSIS: CASE STUDY

B. Blagojević[1], V. Mihailović[2], A. Mulaomerović-Šeta[3]

Keywords: *mean daily flow, instantaneous flow, Sangal's procedure, flood frequency assessment, Bulletin 17b analysis, Bulletin 17c analysis*

ABSTRACT

In the flood frequency assessment, engineers often face the problem of mixed peak data in the data record: some peaks are instantaneous daily flows (IDF), while the others are mean daily flows (MDF). If the problem is recognized, correction factor to the MDFs in the dataset is usually applied. In our research, we use flood flow datasets from 8 hydrologic stations with catchment area up to 1000 km^2 in the Sava River basin. Four hydrologic stations are located in Bosnia and Herzegovina (BiH) and four in Serbia. We apply Sangal's procedure to establish a correction factor for a single station analysis to produce the IDFs from MDFs. We use three annual maxima datasets at each hydrologic station, comprising: 1) IDF only, 2) MDF only, and 3) mixed IDF and MDF (officially available). We subject each dataset to flood frequency assessment using USACE HEC-SSP Bulletin 17b and 17c analysis. The results show the diversity of flood quantile estimates at each station, with the most significant differences obtained according to expected probability curve in Bulletin 17b analysis. The highest uncertainty shown as mean square error of skewness coefficient is observed at stations with large data gaps, and large number of detected low outliers. The case study revealed a potential for significant underestimation or overestimation of flood quantiles, when FFA is performed on mixed dataset, especially in the domain of rare flood events.

[1] Borislava Blagojević, Assist. Prof. Dr. Eng., University of Niš, Faculty of Civil Engineering and Architecture, 14 Aleksandra Medvedeva, Niš 18 000, Serbia, e-mail: borislava.blagojevic@gaf.ni.ac.rs
[2] Vladislava Mihailović, Assist. Prof. Dr. Eng., University of Belgrade, Faculty of Forestry, 1 Kneza Višeslava, Belgrade 11030, Serbia, e-mail: vladislava.mihailovic@sfb.bg.ac.rs
[3] Ajla Mulaomerović-Šeta, Assist. M.Sc. Eng., University of Sarajevo, Faculty of Civil Engineering, 30 Patriotske Lige, Sarajevo 71 000, Bosnia and Herzegovina, e-mail: ajla.mulaomerovic@gf.unsa.ba

1. Introduction

Hydrologic input is the basis for various applications including water resources management, natural hazard issues and hydraulic engineering solutions. When it comes to flood characteristics, the most represented hydrologic input is flood quantile estimates obtained by flood frequency assessment (FFA). The base for FFA is statistical analysis at gauge sites (hydrologic stations), or gauged basins. It is usually performed on the datasets comprising peak flows – annual flow maxima. The input data is provided by responsible institution for hydrologic services and checked before issue against errors in measurement, recording, and rating curve. Still, there are problems in the datasets that may compromise FFA results. In contrast to obvious problems with datasets such as gaps in recording and insufficient data length, there are problems of inhomogeneous data.

When performing commonly used testing for detecting inhomogeneity in the annual flow maxima datasets prior to FFA, engineers in practice are often not aware of how the data is collected. Therefore, one major problem that may not be revealed by statistical tests is inhomogeneous data caused by using different measuring devices. This happens when part of the data is measured discretely, while the rest is measured continuously. The result is a mixed dataset, partly made up of mean daily flows (MDF) and partly of instantaneous daily flows (IDF) or daily extremes.

In large gauged basins, datasets comprising MDF only are not an issue in FFA because the associated IDF is almost the same. However, in small and medium basins with short time-to-peak of flood hydrographs, mixed data would lead to underestimated flood quantiles. The question of how to adjust datasets that contain only or partially MDFs is bothering hydrologists for a long time.

There are several different approaches in the literature to calculate IDFs from MDFs, recently listed and grouped by Chen et al. [1]. The first group uses only MDF sequence to estimate IDF. The simplest approach in this group is transforming MDF to IDF using the IDF/MDF ratio. A better approach is to use the weighted average of MDFs (for example [1, 3, 5]). The second group relates to methods that find regression relationships between the ratio of IDF to MDF and physiographic characteristics of the basin (usually drainage area). The third group comprises advanced methods, including hydrologic modeling and machine learning. Those methods are the most demanding in the terms of quantity and quality of input data.

In the situation when data quantity is limited and regional regression equations do not exist, a reasonable suggestion for engineering practice is to use the simple and practical Sangal’s method. Considering the available data, in this paper we apply Sangal’s procedure to establish the correction factor for a single station analysis to produce the IDFs from MDFs. This method has been applied recently in several studies and papers in its original (for example [6, 7]) or modified form [1, 3].

Our research goal is to find the implication of IDF presence/absence in annual maxima datasets on flood quantile estimates in selected small to medium catchments in the Sava River basin, when FFA is conducted according to Bulletin 17b [4] and 17c analysis [2].

In the Methodology part we briefly show Sangal’s method and applied FFA steps, as well as information about catchments and respective datasets selected for our case study. The next section is dedicated to results and discussion, organized around the main research topics: obtained station average base factor for translating MDF to IDF, change in detected outliers, relative difference of quantile estimates, and overall estimation uncertainty. In the final, concluding section, we present key findings in our case study related to station average base factor values and flood quantile estimates.

2. Methodology

2.1. Study Area and Input Data

The study area of our research is in the Sava River basin (Fig. 1 Above), represented by catchments in the Bosna River sub-basin in Bosnia and Herzegovina (BiH) and the Kolubara River sub-basin in Serbia (SRB). Out of 8 hydrologic stations with catchment area in the range of 140 km^2 to 960 km^2, four station locations are in BiH and four in SRB, presented in Fig. 1 Below.

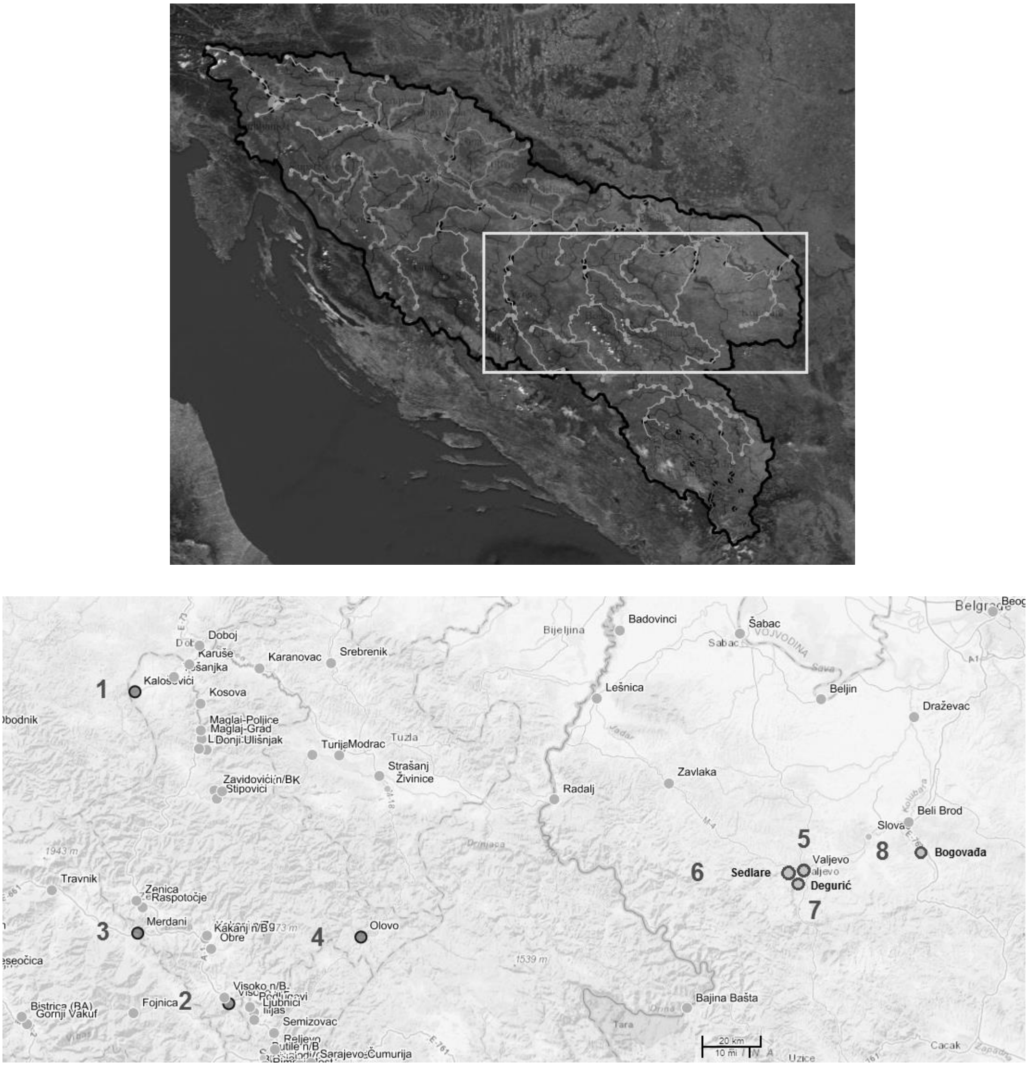

Figure 1. Above: Study area within the Sava River basin (source [11]); Below: Location of hydrologic stations for case study (source [12])

Periods with available peak flow data at each station are illustrated by the bar chart (Fig. 2). In the years with incomplete data in the case of SRB stations, peak data was accepted if less than 60 days out of the high flow period in the annual record has been missing. All flow peak datasets comprise mixed IDF and MDF data, as shown in Fig. 3.

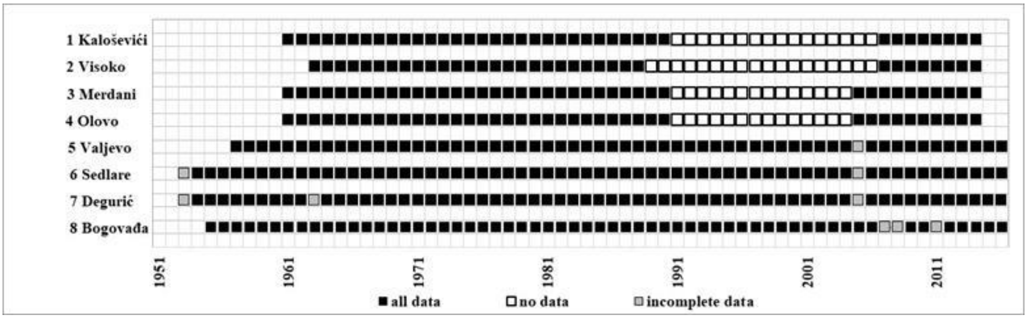

Figure 2. Available annual maxima data at all stations, covering period from the beginning of operation until 2014 for BiH stations (No. 1-4) and until 2016 for SRB (No. 5-8).

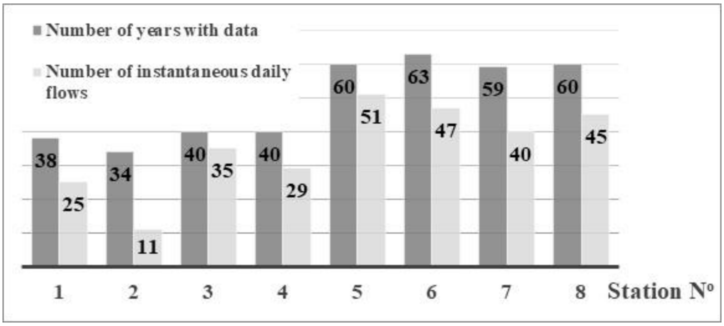

Figure 3. Graphical representation of total number of years with data and total number of IDF in the datasets at each station

2.2. Sangal's Method

Sangal [5] proposed practical method for estimating peak flow. The method uses average daily flows from three consecutive days: the day of annual maximum average daily flow occurrence, the preceding and following day, based on the assumption of a triangular hydrograph shape (Fig. 4).

First, for years with both IDF and MDF, data base factor *K* values are estimated using Eqn. 1. Second, the average value of the base factor, 'station base factor', *Ka*, is calculated for each station. Station base factor is then used to estimate missing IDF values, using Eqn. 2.

$$K = \frac{4Q2 - 2Q1 - 2Q3}{2QP - Q1 - Q3} . \quad (1)$$

$$QP' = \frac{Q1 + Q3}{2} + \frac{2Q2 - Q1 - Q3}{Ka} . \quad (2)$$

where $Q2$ is average daily discharge on a day of maximum instantaneous discharge, m^3/s;

$Q1$ and $Q3$ – average daily discharges on preceding and following day respectively m^3/s;
K – estimated base factor, dimensionless;
Ka – average base factor – station base factor, dimensionless;
QP – recorded annual maximum instantaneous daily flow (IDF), m^3/s;
QP' – estimated annual maximum IDF, m^3/s.

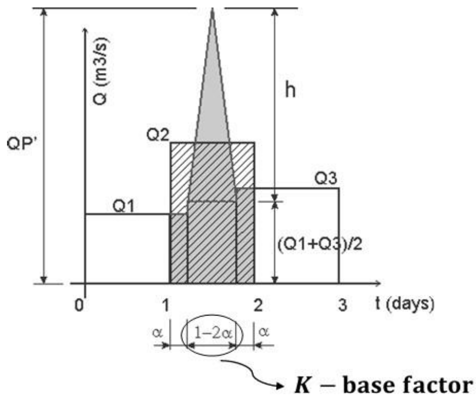

Figure 4. Schematic diagram of Sangal's method (after [5])

After the application of Sangal's procedure, in addition to mixed IDF and MDF dataset as officially available, we prepare two more peak flow datasets at each hydrologic station, comprising IDF only, and MDF only.

2.3. Flood Frequency Assesment

The set of procedures for FFA we use in this research relies on "Guidelines for Determining Flood Flow Frequency" published in 1982 [4] and the current version of these guidelines from 2019 [2]. The former procedures and underlying concepts are denoted Bulletin 17b, the latter, Bulletin 17c. Both Bulletins are developed around the basic assumption on the log-Pearson Type III distribution fitness to empirical distribution in the case of peak flows.

The tool we use to perform FFA is the U.S. Army Corps of Engineers (USACE) Hydrologic Engineering Center Statistical Software Package (HEC-SSP) version 2.1.1. [8] with the associated User's Manual [9].

While specific options for FFA according to Bulletin 17b and 17c are explained in the subsections below, to provide for comparability of the results, the common features in our analyses are the following:

- 90% Confidence interval;
- Flood quantiles of 1000, 500, 200, 100, and 50-year return period, corresponding to 0,001, 0,002, 0,005, 0,01, and 0,05 exceedance probability, respectively;
- Single site analysis using station skew.

2.3.1. Bulletin 17b Analysis

Plotting position formula selected in our FFA by Bulletin 17b is the one of Weibull's, and the low outlier test is Single Grubbs-Beck test. Together with the common features listed above, this constellation of procedures in FFA is typically used by engineers in practice.

Although the method of expected probability is no longer in use by USACE [10], there are situations when design flood flows need to be adopted from the assessed expected probability curve [9]. Therefore, we include flood quantiles assessed by this method in our research.

In the case of missing data, practicing engineers rarely perform gap-filling, especially in the case of single-site FFA when resources are restricted, including budget, time, and information related to flood data for larger area, high water marks, etc. Bulletin 17b analysis in HEC-SSP may be conducted without missing data, as frequently done in practice.

The data representation in Bulletin 17b analysis is shown in Fig. 5.

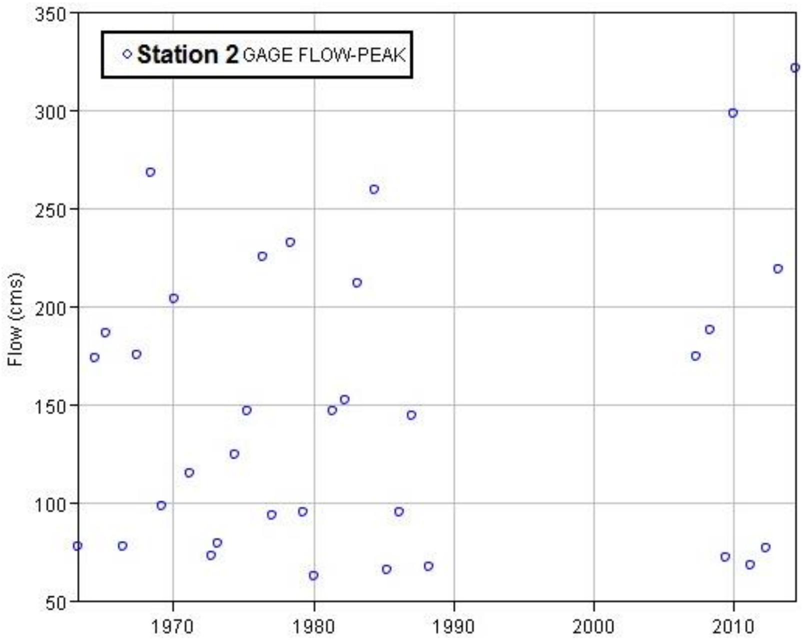

Figure 5. The data representation for Bulletin 17b analysis of IDF&MDF dataset at Station 2

2.3.2. Bulletin 17c Analysis

When performing FFA according to Bulletin 17c in HEC-SSP, plotting position is set to Hirsh/Stedinger, and low outlier test to Multiple Grubbs-Beck, because the method of moments with the Expected Moments Algorithm (EMA) is used to estimate the parameters of the distribution from station data, and adjustments are made for potentially influential low floods [2]. EMA also adjusts for missing values from an incomplete or broken record. Therefore, data representation differs from the one in Bulletin 17b, and analysis cannot be performed without information related to peak flow in all years in the time window selected for analysis.

We select Perception Threshold for data representation in years without peak flow data, and set upper threshold to infinity, and lower to the highest observed peak in the official dataset.

Data representation for one of the BiH stations with large data gap is shown in Fig. 6.

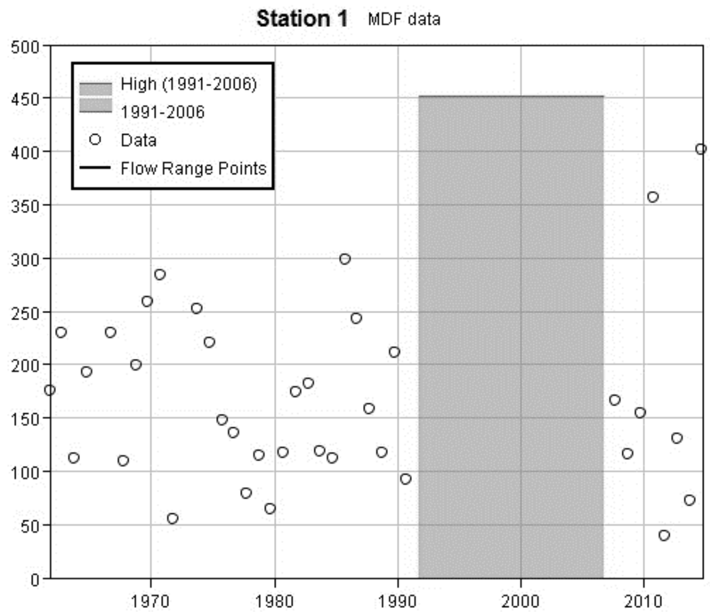

Figure 6. The data representation for Bulletin 17c analysis of MDF dataset at Station 1, with perception thresholds applied in the missing data period 1991-2006

2.4. Analysis Steps

Simplified flow chart in Fig. 7 shows in detail steps performed in our research related to Sangal's method, and FFA as final step.

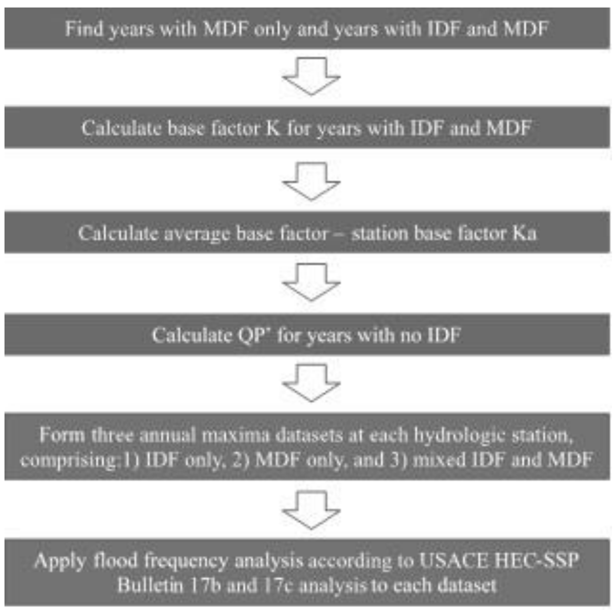

Figure 7. Calculation steps applied in our research methodology

2.5. Uncertainty

The uncertainty indication in our research is taken from the report files of Bulletin 17b and 17c analyses.

Both Bulletin 17b and 17c analysis report confidence interval as uncertainty measure of flood quantile estimates and mean square error (MSE) of at-site skew coefficient (G-at site in our case), as uncertainty of distribution moment.

Bulletin 17c report also shows EMA estimate of MSE for G-at site, and effective record length, important for the datasets with missing data.

It should be noted that the expected probability in Bulletin 17b analysis is an adjustment to account for a bias introduced in the distribution curve due to shortness of dataset.

3. Results and Discussion

3.1. Station Average Base Factor After Sangal's Procedure

First, the base factor following Sangal's procedure is calculated for each station. We noticed several years where the date of maximum MDF mismatches the date of the annual peak – IDF. In this situation, we chose IDF date and the corresponding MDF value in Sangal's procedure.

In Fig. 8-left, station base factor *Ka* versus catchment area for each station is shown, while the box plot in Fig. 8-right shows a distribution of the base factor. There is no significant difference between the mean and the median for the base factor *K*, at all stations. The set of station base factors *Ka* obtained in our research corresponds to the results of other authors [1, 3, 5 – 7].

For SRB stations (No 5-8) an expected correlation between the catchment area and *Ka* exists, according to visual inspection of the Fig. 8-left, which is not the case with BiH ones. In BiH stations (No 1-4), the catchment areas are larger and more uniform, and no relation of *Ka* to the catchment area can be observed (Fig. 8-left). For station No. 6 with the smallest catchment area, *Ka* value is rather small, due to several recorded flood waves that had a peak flow much higher than the mean daily flow, presumably, the duration of these waves was quite short. This is expected for rainfall triggered floods in smaller catchments [5]. It may be seen in the box plots on Fig. 8-right that the minimum of the annual series of estimated K for all SRB stations (No 5-8) are rather small.

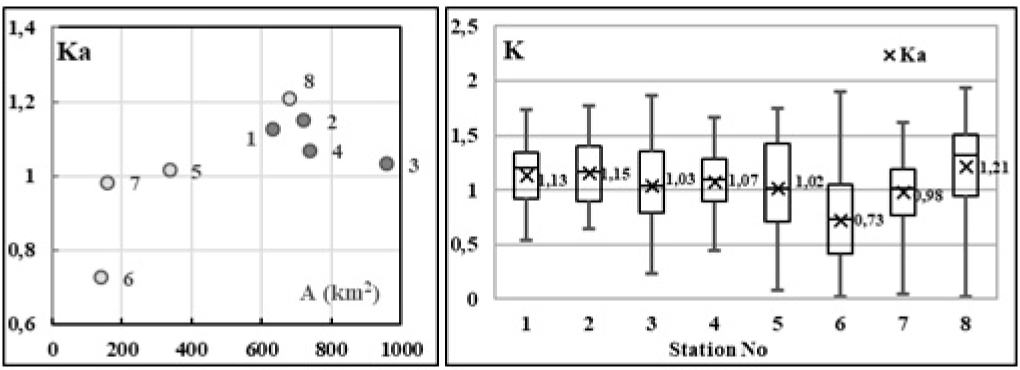

Figure 8. Left: Station base factor *Ka* versus catchment area for each station; Right: Box plot for station average base factor *Ka*

3.2. Flood Quantile Estimates and Related Uncertainties

The results of FFA are shown in Tab. 1 and Tab. 2. These are flood quantile estimates (*QT*) of selected return periods *T* in years from three datasets, obtained according to three distribution curves: Bulletin 17 b (B17b), Bulletin 17 b expected probability (B17b e.), and Bulletin 17c (B17c).

Table 1. Flood quantile estimates *QT* for 1000 and 500-year return period

Stat. No.	A, km²	Analysis	*Q1000*, m³/s			*Q500*, m³/s		
			IDF&MDF	MDF	IDF	IDF&MDF	MDF	IDF
1	633	B17b	918	672	792	845	620	755
		B17b e.	1017	743	838	924	676	796
		B17c	829	674	765	769	620	727
2	721	B17b	654	554	765	584	499	690
		B17b e.	775	646	889	672	567	783
		B17c	401	506	506	389	460	486
3	959	B17b	824	503	875	689	435	735
		B17b e.	1040	605	1096	829	505	880
		B17c	747	462	769	630	404	654
4	737	B17b	1100	1070	1620	934	852	1300
		B17b e.	1351	1435	2151	1105	1076	1631
		B17c	994	963	1400	854	780	1140
5	340	B17b	677	480	537	564	401	469
		B17b e.	785	557	599	637	453	514
		B17c	629	481	537	534	401	469
6	140	B17b	491	286	746	394	234	588
		B17b e.	584	337	898	454	265	684
		B17c	491	286	746	394	232	588
7	159	B17b	295	301	525	253	248	430
		B17b e.	336	354	622	281	283	493
		B17c	295	300	521	253	247	426
8	679	B17b	727	209	658	603	199	565
		B17b e.	848	216	744	684	205	626
		B17c	761	193	688	627	186	587

Table 2. Flood quantile estimates *QT* for 200, 100 and 50-year return period

Stat. No.	*A*, km^2	Analysis	*Q200*, m^3/s			*Q100*, m^3/s			*Q50*, m^3/s		
			IDF&MDF	MDF	IDF	IDF&MDF	MDF	IDF	IDF&MDF	MDF	IDF
1	633	B17b	748	551	700	674	498	654	598	444	602
		B17b e.	803	590	732	714	527	679	626	464	621
		B17c	687	549	671	623	495	623	578	440	572
2	721	B17b	498	431	596	436	381	528	378	334	462
		B17b e.	553	474	656	473	411	569	402	353	489
		B17c	370	402	454	352	359	426	329	317	395
3	959	B17b	539	357	578	445	304	478	363	257	392
		B17b e.	617	398	660	491	330	528	390	273	421
		B17c	499	335	524	415	288	439	342	245	364
4	737	B17b	745	629	967	621	497	766	512	390	601
		B17b e.	843	743	1138	683	562	865	549	425	656
		B17c	691	586	866	582	469	697	485	372	555
5	340	B17b	439	312	386	358	254	328	288	204	275
		B17b e.	480	341	413	384	273	347	303	215	287
		B17c	423	312	386	350	254	328	285	204	275
6	140	B17b	290	172	422	226	136	323	173	105	242
		B17b e.	321	191	472	245	147	352	184	111	258
		B17c	290	172	422	226	136	323	173	105	242
7	159	B17b	204	189	325	171	152	259	141	121	204
		B17b e.	221	208	359	182	164	280	148	127	216
		B17c	204	188	323	171	152	258	141	120	203
8	679	B17b	467	185	457	381	173	385	307	160	320
		B17b e.	512	189	493	408	177	408	323	163	334
		B17c	480	175	470	388	166	393	311	155	325

The diversification of *QT*s is reported in Tab. 3 by a relative difference of estimated flood quantiles to the reference flood quantile, calculated according to Eqn. 3:

$$RD = \frac{QT - QTref}{QTref} 100(\%), \qquad (3)$$

where RD is relative difference of estimated flood quantile, %;

QT – flood quantile of T-year return period, m^3/s;

$QTref$ – flood quantile estimate of the same T-year return period, obtained from the IDF&MDF dataset from the Bulletin 17b analysis distribution curve (B17b), m^3/s.

We use this particular selection of *QTref*, because it would be a FFA result typically obtained by practicing engineers.

The relative difference in Tab. 3 is not shown for the MDF dataset because corresponding *QT*s are lower to significantly lower (see Tab. 1 and Tab. 2) compared to the ones obtained from the other two datasets in all stations and return periods, except for the 1000-year quantiles at station No. 7.

Table 3. Relative difference of flood quantile estimates QT obtained according to three distribution curves. Relative difference ranges are also highlighted in shades of gray according to the legend shown in the last row

Stat. No.	*A*, km^2	Analysis	***RD(Q1000)*, %**		***RD(Q500)*, %**		***RD(Q200)*, %**		***RD(Q100)*, %**		***RD(Q50)*, %**	
			IDF&MDF	**IDF**	**IDF&MDF**	**IDF**	**IDF&MDF**	**IDF**	**IDF&MDF**	**IDF**	**IDF&MDF**	**IDF**
1	633	B17b	ref	-13,7	ref	-10,7	ref	6,4	ref	-3,0	ref	0,7
		B17b e.	10,8	-8,7	9,4	-5,8	7,3	-2,2	5,9	0,7	4,6	3,8
		B17c	-9,7	-16,7	-9,0	-14,0	-8,2	-10,3	-7,6	-7,6	-3,3	-4,3
2	721	B17b	ref	17,0	ref	18,2	ref	19,7	ref	21,1	ref	22,2
		B17b e.	18,5	36,0	15,1	34,1	11,0	31,8	8,6	30,5	6,3	29,3
		B17c	-38,7	-22,6	-33,4	-16,8	-25,7	-8,8	-19,3	-2,3	-13,0	4,5
3	959	B17b	ref	6,2	ref	6,7	ref	7,2	ref	7,5	ref	7,9
		B17b e.	26,2	33,0	20,3	27,8	14,4	22,3	10,6	18,8	7,5	15,8
		B17c	-9,3	-6,7	-8,5	-5,1	-7,5	-2,8	-6,6	-1,2	-5,9	0,2
4	737	B17b	ref	47,3	ref	39,2	ref	29,8	ref	23,3	ref	17,4
		B17b e.	22,9	95,6	18,3	74,6	13,2	52,8	9,9	39,3	7,1	28,1
		B17c	-9,6	27,3	-8,6	22,1	-7,2	16,2	-6,3	12,2	-5,3	8,4
5	340	B17b	ref	-20,7	ref	-16,8	ref	-12,1	ref	-8,4	ref	-4,5
		B17b e.	15,9	-11,5	13,0	-8,9	9,3	-5,8	7,1	-3,1	5,3	-0,4
		B17c	-7,1	-20,7	-5,3	-16,8	-3,6	-12,1	-2,2	-8,4	-1,0	-4,5
6	140	B17b	ref	51,9	ref	49,2	ref	45,5	ref	42,9	ref	39,9
		B17b e.	18,8	82,9	15,1	73,7	10,8	62,8	8,3	55,7	6,2	49,3
		B17c	0,0	51,9	0,0	49,2	0,0	45,5	0,0	42,9	0,0	39,9
7	159	B17b	ref	78,0	ref	70,0	ref	59,3	ref	51,5	ref	44,7
		B17b e.	13,8	110,7	11,3	94,8	8,2	76,2	6,2	63,8	4,8	53,1
		B17c	0,0	76,6	0,0	68,4	0,0	58,3	0,0	50,9	0,0	44,0
8	679	B17b	ref	-9,5	ref	-6,3	ref	-2,1	ref	1,0	ref	4,2
		B17b e.	16,7	2,4	13,4	3,7	9,6	5,6	7,1	7,0	5,3	8,7
		B17c	4,7	-5,4	4,0	-2,7	2,8	0,6	1,8	3,1	1,3	5,9
Legend:		RD range, %		>100	100÷80	80÷60	60÷40	40÷20	20÷0	0÷-20	-20÷-40	

There are two stations exhibiting significant positive *RD* in all return periods for *QT*s assessed from the IDF datasets, stations No. 6 and 7, followed by station 4, where *RD* is lower, but still rather high.

The lowest negative *RD* is found at station No. 2, followed by No. 5.

Stations No. 1 and No. 8 have 'an acceptable' level of *RD*, negative prevailing in station 1 and positive in station 8.

The results for station 3 show higher positive *RD* for *QT*s obtained according to B17b expected probability curve, compared to neglectable *RD* for *QT*s from two other probability curves.

The detected number of outliers by both Grubbs-Beck test versions is shown in Tab. 4.

Table 4. Number of low and high outliers detected in all examined datasets

Stat. No.	*A*, km^2	Test	High outliers, number			Low outliers, number		
			IDF&MDF	MDF	IDF	IDF&MDF	MDF	IDF
1	633	B17b, S G-B	0	0	0	0	0	0
		B17c, M G-B	0	0	0	0	0	0
2	721	B17b, S G-B	0	0	0	0	0	0
		B17c, M G-B	0	0	0	14	0	14
3	959	B17b, S G-B	1	1	1	0	0	0
		B17c, M G-B	0	0	0	0	0	0
4	737	B17b, S G-B	0	1	1	0	0	0
		B17c, M G-B	0	0	0	0	0	0
5	340	B17b, S G-B	0	1	0	1	0	0
		B17c, M G-B	0	0	0	1	0	0
6	140	B17b, S G-B	0	0	0	0	0	0
		B17c, M G-B	0	0	0	0	0	0
7	159	B17b, S G-B	0	0	0	0	0	0
		B17c, M G-B	0	0	0	0	0	0
8	679	B17b, S G-B	1	0	1	0	1	0
		B17c, M G-B	0	0	0	0	8	0

There are three stations (1, 6 and 7) without any detected outliers.

High outliers are detected at all datasets of station 3, MDF and IDF dataset at station 4, mixed dataset and IDF only dataset at station 8, and for MDF only at station 5. All these outliers are detected in Bulletin 17b analysis by Single Grubbs-Beck test.

There are two stations where low outliers are detected by both Single and Multiple Grubbs-Beck test, stations 5 and 8. In the latter case, the number of detected low outliers is different. The station No. 2 has 14 low outliers detected in two series.

The probability plots for stations No. 8 and No. 3 (Fig. 9 and Fig. 10) show the situation with presence of outliers in the studied datasets. Both station 3 and 8 are not among the stations with the largest *RD* of *QT*s.

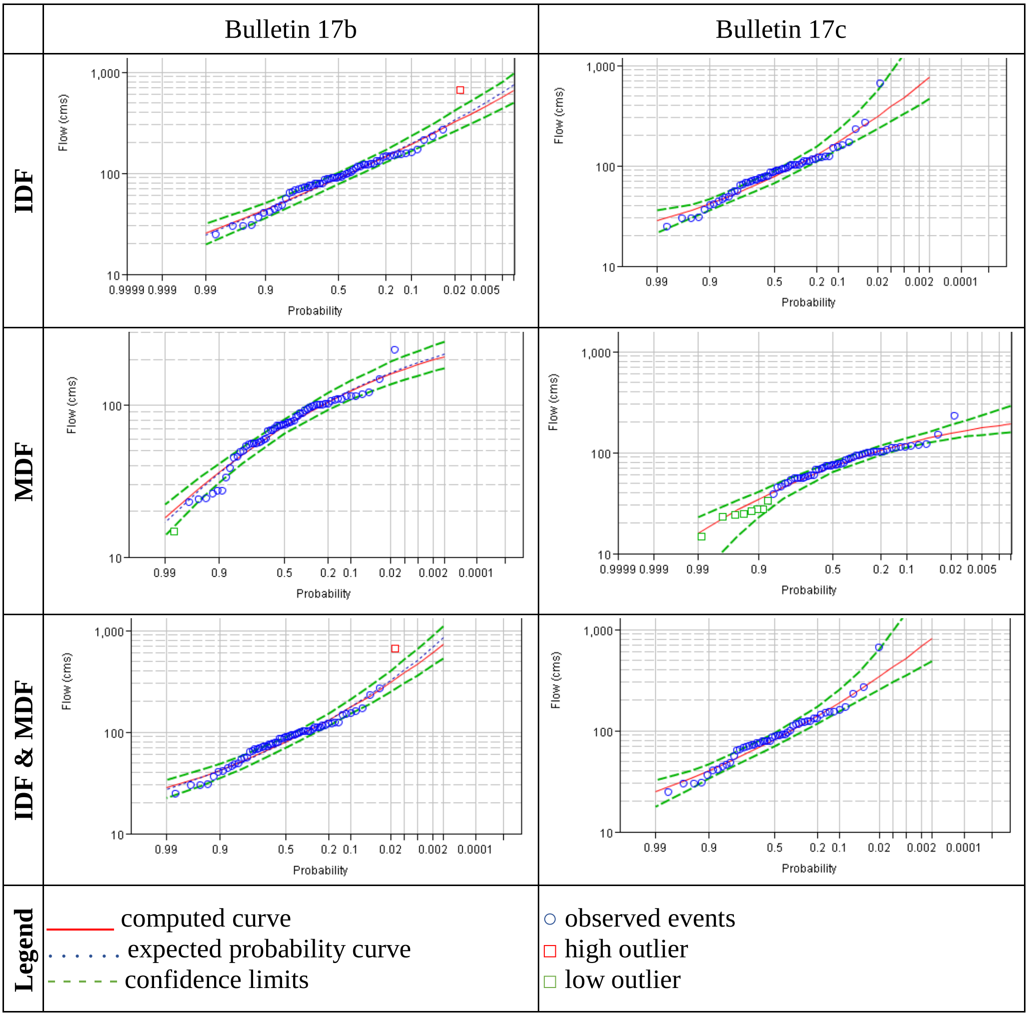

Figure 9. Probability plots at station No. 8

Probability plots for station 8 in Fig. 9 also illustrate change of skewness coefficient (*G*) sign in the datasets, as shown in Tab. 4 related to uncertainty. These two MDF datasets are negatively skewed and their mean square errors (*MSE*) slightly differ between two algorithms applied for distribution moments estimation in Bulletins 17b and 17c. This slight difference may also be attributed to 2 missing data in systematic record compared to historic period of observations.

In both Fig. 8 and Fig. 9, the 90% confidence interval is shown. Flood quantile estimation uncertainty by Bulletin 17c EMA considers more sources of uncertainty compared to Bulletin 17b [10]. Therefore, confidence intervals shown in the right columns of Fig. 8 and Fig. 9 are wider, compared to the ones in the left column.

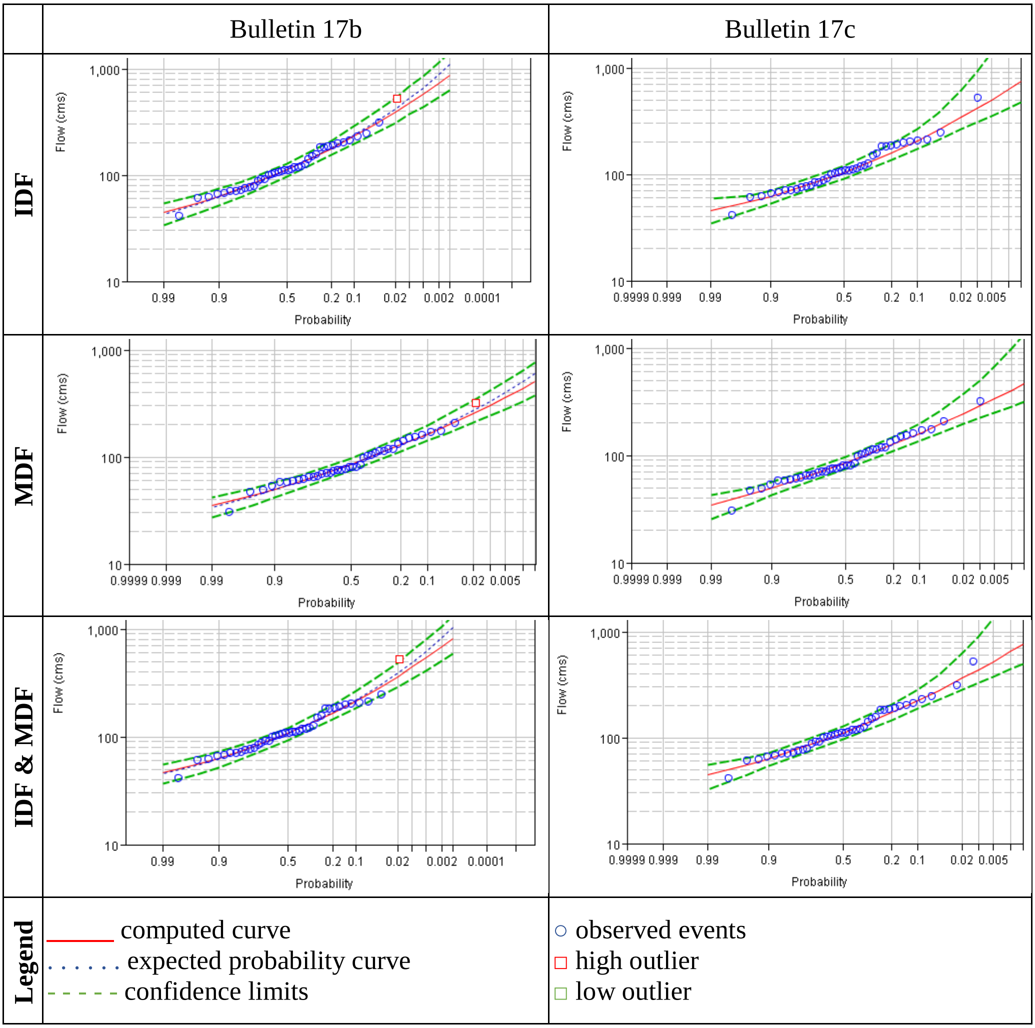

Figure 10. Probability plots at station No. 3

In Tab. 5 the source of larger difference in confidence intervals computed by Bulletin 17b and 17c can also be observed following *MSE* of *G* and especially EMA *MSE(G)*. In EMA, the value of historical information is recognized. Therefore, in the last column, the number of years in historic period and the number of years in systematic record is given for each station.

The differences in *MSE* for all datasets are highlighted in grey, showing the consequences of the large data gaps on *G* estimates in stations 1 – 4, and a few missing data at stations 7 and 8.

The importance of the information provided in Tab. 5 lies in the design flood quantile selection process. There is no straight forward, generally applicable conclusion about the best FFA result or analysis type in the presented case study. Each station requires patient examination of RD (Tab. 3), while consulting outlier information (Tab. 4), looking at probability plots, and taking into consideration the said information in Tab. 5. Therefore, further discussion is provided on the station-by-station basis.

Table 5. Skewness coefficient estimates (G) and associated uncertainty

Stat. No.	Analysis	*G*			*MSE(G)*			EMA *MSE(G)*			Rec., years
		IDF&MDF	MDF	IDF	IDF&MDF	MDF	IDF	IDF&MDF	MDF	IDF	Historic / Systematic
1	B17b	-0,367	-0,344	-0,714	0,162	0,160	0,195	/	/	/	52
	B17c	-0,411	-0,321	-0,667	0,166	0,166	0,198	0,145	0,142	0,171	38
2	B17b	0,053	-0,004	-0,07	0,152	0,148	0,153	/	/	/	52
	B17c	-1,024	-0,069	-0,793	0,269	0,153	0,220	0,202	0,128	0,161	34
3	B17b	0,727	0,439	0,568	0,189	0,161	0,173	/	/	/	54
	B17c	0,643	0,379	0,512	0,188	0,156	0,168	0,188	0,147	0,163	40
4	B17b	0,335	1,064	0,788	0,152	0,251	0,195	/	/	/	54
	B17c	0,278	0,991	0,725	0,148	0,233	0,188	0,137	0,237	0,189	40
5	B17b	0,320	0,259	-0,021	0,107	0,103	0,088	/	/	/	60
	B17c	0,162	0,259	-0,021	0,096	0,103	0,088	0,096	0,103	0,088	60
6	B17b	0,300	0,247	0,291	0,101	0,098	0,101	/	/	/	63
	B17c	0,300	0,246	0,291	0,101	0,098	0,100	0,101	0,098	0,100	63
7	B17b	0,152	0,381	0,335	0,097	0,113	0,109	/	/	/	62
	B17c	0,151	0,380	0,330	0,097	0,112	0,109	0,097	0,112	0,108	59
8	B17b	0,535	-0,601	0,223	0,123	0,128	0,100	/	/	/	60
	B17c	0,589	-0,767	0,263	0,131	0,146	0,106	0,131	0,143	0,106	58

3.3. Station-By-Station Discussion

In case all the data is available for the analysis, one would select flood quantile assessed by Bulletin 17c analysis based on the dataset comprising IDF only. The discussion in this subsection is based on data availability and comparison of *QTref* to *QT* obtained for IDF dataset by Bulletin 17c. The closest case of full data availability in our study is represented by stations 5 and 6.

In station 5, one low outlier is detected, in station 6, there are none. In these two stations, MSE of *G* is the smallest of all stations, but in station 5, the sign of *G* for IDF datasets changes to negative, compared to the positive sign in the mixed dataset. The situation with value and sign of *G* at station 6 is steady. Therefore, station 6 is the best candidate to observe RD of flood quantile estimates. It could be concluded that by adopting flood quantiles based on Bulletin 17b FFA on the mixed dataset, one would underestimate *QT* for 50% on average. Furthermore, this station is the one with smallest *Ka* applied according to Sangal's procedure, and RD may be larger, as found by Sangal: *'The formula will underpredict the peak of rainfall floods from small basins*' [5].

According to RDs calculated for station 5, design floods for 100- and 50-year return period would be negligibly underestimated, while the 1000-year flood would be 20% underestimated.

Station 8 with almost all the data in the dataset, ¾ of the recorded IDFs, and detected outliers by both analyses, arrives to almost the same flood quantiles. It is noted that station 8 and station 1 with similar catchment area have similar flood quantile differences, but of different sign. In addition, station 1 does not have any outliers detected. In both stations, signs of *G* in the mixed and IDF datasets are the same, while values are significantly changed.

The station with the largest underestimation of *QT*s by Bulletin 17b is station 7. Less than 5% of data is missing in the historic record of this station, and 2/3 IDFs are known. This is the second smallest station according to catchment area, with *Ka* closest to 1. The station is free from outliers. Therefore, this station may show the closest to real situation related to effect of mixed peak data in flood quantile estimation. Similar results in the terms of underestimation of quantiles, are obtained for station 4, the second best of BiH stations, comprising 72% IDFs in the systematic record. The data gap in this station is ¼ data, and there is one high outlier accounted in Bulletin 17b analysis. Here, signs of *G* in the mixed and IDF datasets are the same, while values are significantly changed.

The best BiH station regarding data is station 3, with ¾ of the data in historic record, and 88% of IDFs measured. The station has one high outlier, and interesting situation regarding *G*: it is of the same sign in both mixed and IDF datasets, but *G* is smaller in IDF datasets. Flood quantiles are practically the same – within 10% difference in all return periods. Ka in this station is 1,03 the smallest of all BiH values, for the largest catchment area.

A special case in BiH stations is station 2, not only because it is the poorest data-wise with 65% of the systematic record in the historic period and 1/3 of the IDFs, but because 14 low outliers are detected in Bulletin 17c analysis, and not a single one in Bulletin 17b analysis. It has the highest *Ka* of all BiH stations. While 50- and 100-year flood quantiles do not exhibit significant RD, rarer flood events tend to be overestimated by Bulletin 17b analysis, according to RD interpretation. It should be noted here that *G* estimates by Bulletin 17b and 17c are quite different both in sign and value. This is the case when with negative *G*, found by Bulletin 17 c, the upper limit of the LPT3 distribution occurs. In the cases like this, it is recommendable to calculate the value of the upper distribution limit and assure reasonable value is obtained. If the value is not reasonable from the flood quantile range standpoint, this might lead to adopting Bulletin 17b or even expected curve results for quantile estimates.

4. Conclusion

Given the data availability, the aim of the paper was to implement Sangal's practical and simple procedure for calculating the missing IDF values. The IDFs were calculated as a linear combination of MDFs of three consecutive days. The method has shown satisfactory results in the considered cases. The possibilities of applying an improved version of this procedure should be explored in the future. For example, Fill & Steiner [3] proposed variable coefficients (that can be determined over a region using historical data) in the Sangal’s equation for the IDF calculation, and found that correction factor should be applied to that equation to obtain better agreement between estimated and observed data. Chen et al. [1] proposed a slope-based method, an empirical method which uses not only MDFs but also the rising and falling limb slopes to describe the shape of a MDF hydrograph.

The available datasets at the stations selected for the research encountered majority of issues in engineering practice, including mixed IDF and MDF, data gaps, missing data and presence of outliers. In our case study, detected number of high and low outliers has not been increased due to application of Sangal's method in forming datasets comprising IDFs only. The significance of outliers is in the potential influence on a skewness coefficient, reflected on flood estimates in the case of large number of low outliers and consequent adjustment of distribution curve in Bulletin 17c procedure. One of such cases is further discussed, and further treatment is recommended.

Flood quantiles obtained by Bulletin 17b expected curve method are generally overestimated both in IDF&MDF datasets and IDF datasets. Flood quantiles assessed by Bulletin 17c EMA, tend to be smaller compared to the ones of Bulletin 17b in associated datasets with significant number of missing data, when the threshold is set to maximum observed flow.

There is a potential for significant underestimation or overestimation of flood quantiles, when FFA is performed on mixed dataset comprising both MDF and IDF, especially in the domain of rare flood events. The only way to come to an informed choice when deciding about design flood is to conduct FFA according to both Bulletin 17b and 17 c procedures and methods, and compare the results. This is a favorable approach in the situation when data quantity is limited and single station analysis is conducted in the absence of regional regression equations.

For flood studies in engineering practice, supposing data availability is not a limitation, one should correct mixed datasets, conduct FFA, and compare the quantiles with regional values and along the watercourse. The decision on design flow after FFA should then be brought in the regional context.

Acknowledgements

The authors are grateful to the Federal Hydrometeorological Service of Bosnia and Herzegovina and the Republic Hydrometeorological Service of Serbia for making data available for this research. The work for this study is partially financed from the research project TR37005 of the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

1. *Chen, B., Krajewski, W. F., Liu, F., Fang, W., Zongxue, X*. Estimating Instantaneous Peak Flow From Mean Daily Flow. Hydrology Research, 2017, 1474-1488.

2. *England, J. F. Jr., Cohn, T. A., Faber, B. A., Stedinger, J. R., Thomas, W. O. Jr., Veilleux, A. G., Kiang, J. E., Mason, R. R. Jr.* Guidelines for Determining Flood Flow Frequency, Bulletin 17C. Chapter 5 of Section B, Surface Water, Book 4, Hydrologic Analysis and Interpretation https://pubs.usgs.gov/tm/04/b05/tm4b5.pdf, visited on 05.01.2020.

3. *Fill, H. D., Steiner, A. A.* Estimating Instantenaous Peak flow from Mean Daily Flow Data. Journal of Hydrologic Engineering, 2003, 365-369. doi:10.1061/(ASCE)1084-0699(2003)8:6(365).

4. Interagency Committee on Water Data. 1982, Guidelines for Determining Flood Flow Frequency, Bulletin 17B: Interagency Committee on Water Data. Hydrology Subcommittee, Technical Report https://water.usgs.gov/osw/bulletin17b/dl_flow.pdf visited on 05.01.2020.

5. *Sangal, B. P*. Practical Method of Estimating Peak Flow. Journal of Hydraulic Engineering, 1983, 109, 549–563.

6. *Sanocki, C. A., Williams-Sether, T., Steeves, P. A., Christensen V. G.* Techniques for Estimating the Magnitude and Frequency of Peak Flows on Small Streams in the Binational U.S. and Canadian Lake of the Woods–Rainy River Basin Upstream from Kenora, Ontario, Canada, Based on Data Through Water Year 2013., Scientific Investigations Report 2019–5012, Prepared U.S. Geological Survey, 2019.

7. *Sehgal, K.* A Unified Ontario Flood Method–Regional Flood Frequency Analysis on Ontario Streams Using Multiple Regression. Department of Civil Engineering University of Toronto, Master Thesis, 2016. http://hdl.handle.net/1807/76843, visited on 05.01.2020.

8. https://www.hec.usace.army.mil/software/hec-ssp/download.aspx, visited on 05.01.2020.

9. https://www.hec.usace.army.mil/software/hec-ssp/documentation/HEC-SSP_21_ Users_Manual.pdf visited on 05.01.2020.

10. https://acwi.gov/hydrology/Frequency/minutes/major_differences_in_bulletin_17B_ and_bulletin_17C_050415.pdf visited on 05.01.2020.

11. http://www.savahis.org/ visited on 05.01.2020.

12. http://www.savagis.org/ visited on 05.01.2020.