Can Rainfall-Runoff Models Provide Accurate Estimates of Design Flood Quantiles in Ungauged Catchments?

Andrijana Todorović[1]
Andrea Petroselli[2]
Nikola Zlatanović[3]

ABSTRACT: Accurate estimation of design floods is one of the most important tasks of applied hydrology, since design of hydraulic structures and flood protection measures heavily depends on these estimates. In hydrologically gauged catchments these estimates are usually obtained by applying frequency analyses over a series of annual maxima. In case of ungauged catchments, design floods are commonly estimated by applying an event-based rainfall-runoff models with design rainfall hyetographs (hereafter referred to as *models*). Although the design flood estimates obtained in this way are rather sensitive to every element of these models (e.g., design rainfall duration or unit hydrograph method), there is no specific guidance on the modelling decisions to obtain reliable design flood estimates. Robustness of a particular model is evaluated in gauged catchments (where observed data are available), by comparing design floods obtained with the model, to the results of the frequency analysis. The agreement between the two design flood estimates is usually quantified in terms of relative errors, which do not take into account uncertainty in the quantiles. In this paper, we propose a complementary approach to evaluation of event-based models, which implies that design flood estimates are compared to the confidence intervals of the quantiles. This approach is applied to nine models used to simulate design floods of 20-, 50- and 100-year return periods at the location of Zavlaka stream gauge on the Jadar River. The results show that taking quantile confidence intervals into consideration can provide additional insights in model performance, and, thus, should be mandatorily included in the model evaluation. Comparison across the models reveals considerable sensitivity of the design flood estimates to the models. Models that use daily design rainfall of uniform intensity systematically underestimate corresponding quantiles, while the estimates obtained with shorter rainfall durations or with rainfall of non-uniform intensities result in errors of both signs. Equifinality between the curve number and rainfall duration is also detected in the results. This study exposes challenges in evaluating event-based models, and emphasises the need for specific guidance on the application of these models for design flood estimation.

Key words: design flood; design storm; event-based models; flood frequency analysis; ungauged catchments; unit hydrograph

Да ли модели падавине-отицај могу да дају поуздане оцене меродавних великих вода на неизученим сливовима?

АПСТРАКТ: Прорачун меродавних великих вода представља један од најзначајнијих задатака инжењерске хидрологије, с обзиром да се хидротехничке грађевине и системи заштите од поплава пројектују према меродавним протоцима. На хидролошки изученим сливовима, меродавне велике воде се најчешће рачунају применом методе годишњих максимума. На хидролошки неизученим сливовима, меродавни протоци се најчешће одређују применом хидролошких модела епизода са рачунским кишама. Иако меродавни протоци добијени помоћу ових модела веома зависе од одабира сваке компоненте модела (нпр., трајање рачунске кише или модел јединичног хидрограма), не постоје конкретна упутства за формирање ових модела како би се добиле поуздане процене меродавних великих вода. Ефикасност конкретног модела може се анализирати само на изученим сливовима, тако што се меродавне велике воде добијене помоћу модела пореде са резултатима анализе

[1]Dr Andrijana Todorović, Faculty of Civil Engineering of the University of Belgrade; Uppsala University, Department of Earth Sciences; Assistant professor; atodorovic@grf.bg.ac.rs
[2] Dr Andrea Petroselli, Department of Economics, Engineering, Society and Business Organization - DEIM, Tuscia University; Associate professor; petro@unitus.it
[3] Nikola Zlatanović, Faculty of Civil Engineering of the University of Belgrade; PhD candidate; zzznikola@gmail.com

годишњих максимума. Слагање између меродавних великих вода добијених на ова два начина најчешће се квантификује кроз релативну грешку, која не узима у обзир неизвесност самих квантила. У овом раду, приказана је допунска анализа ефикасности ових модела, која подразумева поређење резултата модела са интервалима поверења квантила добијених методом годишњих максимума. Ова анализа је урађена са девет модела, који су коришћени за прорачун 20-, 50- и 100-годишњих великих вода на водомерној станици Завлака на реци Јадар. Резултати показују да разматрање ширине интервала поверења квантила омогућава боље сагледавање ефикасности модела. Оцене меродавних протока веома варирају у зависности од одабраног модела. Модели који користе дневну рачунску кишу равномерног интензитета (тзв. блок киша) систематски потцењују квантиле великих вода, док модели који користе блок кишу краћег трајања или кишу неравномерног интензитета могу и да прецењују и да потцењују квантиле. Резултати указују и на међусобну повезаност параметара модела (енгл. *equifinality*), као што су број криве (*CN*) и трајање рачунске кише. У овом раду јасно је указано на изазове у вредновању модела епизода, као и на потребу за израдом конкретних препорука за примену ових модела у циљу добијања меродавних великих вода.

Кључне речи: рачунске (меродавне) велике воде; рачунске кише; метода годишњих максимума; модели епизода; хидролошки неизучени сливови; јединични хидрограм

1 Introduction

Accurate estimates of design floods are essential for adequate design of hydraulic structures, such as dams or embankments ([1],[2]), and uncertainties in these estimates can mislead the design [3]. Furthermore, accurate estimates of design flows are needed for dam safety studies [3], as well as for flood hazard and flood risk assessments [4]. Traditionally, design floods are estimated from statistical analyses of observed flood flows. To this end, either flood frequency analyses or Peak-over-Threshold (PoT) methods are commonly applied ([1],[5],[6]). The former implies fitting theoretical distributions to probability plots of observed annual maxima, and computation of quantiles of interest (i.e., design flood flows) according to the best-fit distribution ([5],[7]). The PoT method relies on analyses of series of flood flows that exceed a certain threshold value, not taking a calendar year of their occurrence into consideration ([5], [8]).

Rainfall-runoff models, including event-based and continuous models, are also employed for design flood estimation [1]. Event-based models simulate individual floods caused by a single rainfall event, as opposed to continuous models, which simulate catchment response over long time that includes high flow periods, as well as dry periods in-between ([9],[10]). Continuous models are generally more complex because they simulate numerous processes, such as evapotranspiration, snowpack or baseflow, all of which are commonly omitted by event-based models. Furthermore, continuous models comprise many parameters that have to be estimated during the calibration process [11]. To estimate design floods, statistical analyses are performed with the outputs of continuous models ([1],[12]). On the other hand, event-based models are run with design rainfall, and the return period of resulting flood flows is assumed equal to the return period of design rainfall ([1],[5]). This approach is referred to as design storm method [13].

In order to apply an event-based model for design flood estimation, numerous modelling decisions have to be made, such as selection of the design rainfall duration and hyetograph shape, selection of methods for effective rainfall computation and runoff transformation, or setting the initial conditions, which are represented by a free model parameter in these models ([1],[3],[12]). There is a lack of specific guidance on these modelling decisions, except for design rainfall duration, which is usually selected as the one that results in the largest peak flow. Design flood estimation in ungauged catchments is limited to the application of rainfall-runoff models that do not require calibration, i.e., their parameters can be estimated from physiographic properties of the catchment, such as area or slope ([5],[23],[15],[16],[17]). Commonly, these are parsimonious models based on synthetic unit hydrographs [5], although continuous models can also be applied [18]. Alternatively, various empirical approaches and regional analyses can be used; however, their application is generally limited to location that these methods are obtained for [19].

Each of these approaches to design flood estimation is accompanied by uncertainties. For example, probability distributions and quantiles can be considerably affected by a presence of outlies in the series of annual maxima [20]. The quantiles can also be affected by the criteria for selection of the best fitted distribution [21]. Continuous rainfall-runoff models generally have low performance in extreme flows, since calibration leads to a “squeezing” of the flow distribution (i.e., shifting distribution tails closer to the mean value) [22], which leads to underestimated flood flows [23]. Furthermore, rainfall records are generally too short to allow accurate estimation of floods of long return periods [3]. Calibration is a challenging task both in continuous [24] and event-based models [25], and uncertainties in parameter estimates can eventually affect the design flood estimates [26]. Design floods obtained by applying event-based models are quite sensitive to each decision made throughout the modelling process ([27],[19]). For example, variations in initial conditions or design rainfall duration can result in design floods of return periods that can be of an order of a magnitude larger or smaller than the assumed return period of design rainfall [27], questioning thereby the strong assumption on equivalence of the two [1].

This study focuses on application of event-based models in ungauged catchments. To single out the best modelling decisions, i.e., those that yield the most credible design flood estimates, the estimates obtained with rainfall-runoff models are compared to those obtained from the statistical analyses, which are generally considered to be a standard approach in engineering practice. Although such comparisons are conducted in gauged catchments, the conclusions are assumed valid for ungauged catchments as well ([2],[13],[26],[27]). However, a comprehensive comparison that could enable proper model evaluation and identification of best modelling decisions, is challenging due to the uncertainties inherent to the quantiles. Model performance is usually assessed in terms of relative errors, commonly without any considerations of the quantile uncertainties [1]. In this paper we argue that such an approach is oversimplified and can lead to a loss of information needed for proper evaluation of models Furthermore, model evaluation based only on relative error magnitude is inevitably accompanied by subjectivity.

In this paper we propose a complementary analysis of model performance, and apply it for evaluation of nine models in the Jadar River catchment at the location of the Zavlaka stream gauge. Specifically, the models are evaluated by comparing resulting design floods not only to the corresponding quantiles obtained from the frequency analysis, but also to the confidence intervals of the flow quantiles. The model evaluation is aimed at detection of best modelling decisions for estimation of design floods. This approach to model evaluation is presented in detail in section 2, together with the catchment and the rainfall-runoff models employed in this study. The results are presented and discussed in sections 3 and 4, respectively, while section 5 concludes the paper.

2 Methodology

2.1 Catchment and Data

The Jadar River represents a 75 km long right tributary of the transboundary Drina River. The length of the Jadar River to Zavlaka amounts to 35.4 km with a mean slope of 0.7%. The area of the Jadar River catchment upstream of the Zavlaka stream gauge amounts to 313 km^2. The highest elevation areas and the steepest slopes are located in the southernmost parts of the catchment (Figure 1). Time of concentration of the catchment is estimated at approximately 9 hours. Pseudogley and podzolic soils are prevalent in the catchment. Broad-leaved forests and agricultural land prevail in the catchment, while other land use types, including urbanised areas, are present to considerably lesser extent.

For this study, maximum daily flows observed at the Zavlaka stream gauge from 1960 to 2018 (except for 2016) are used (Figure 2). The key statistics of the series are reported in Table 1. Preliminary statistical analyses do not reveal significant trends or autocorrelation in the annual maxima. Design rainfall are obtained from depth-duration-frequency (DDF) curves developed for the Loznica meteorological station [28], which is located in the immediate proximity of the catchment.

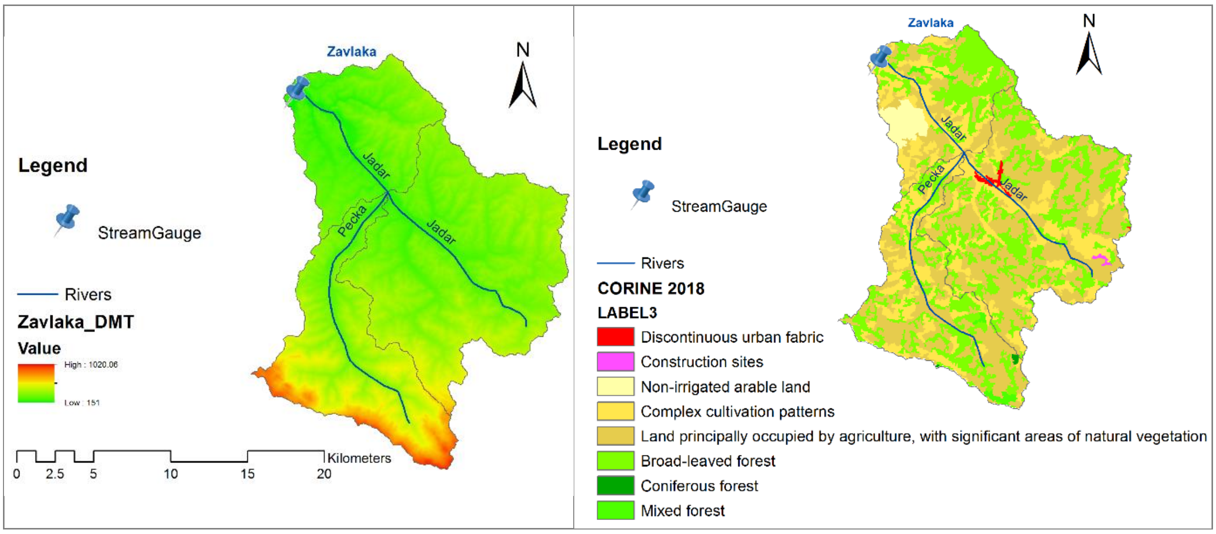

Figure 1. The Jadar River catchment upstream of the Zavlaka stream gauge: catchment digital terrain model and the stream network (left), and land use types according to CORINE 2018 [29].

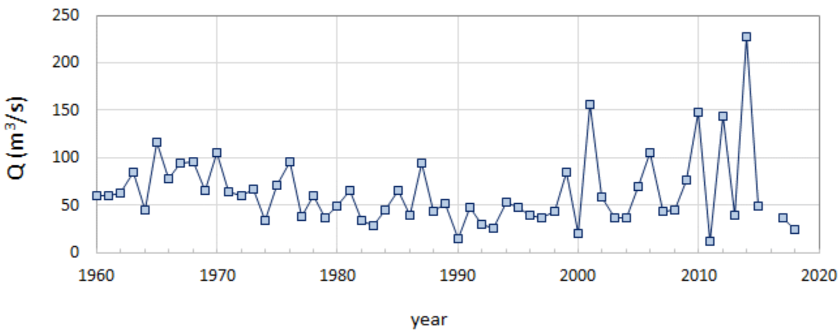

Figure 2. Observed annual maxima at the Zavlaka stream gauge.

Table 1. Statistics of annual maxima series observed at the Zavlaka stream gauge over the period 1960-2018.

Statistic	**Mean value**	**Standard deviation**	**Coefficient of variation**	**Skewness coefficient**
Series of annual maxima	63.2	38.4	0.607	1.935
Log-transformed series	1.735	0.241	0.139	-0.027

2.2 Frequency Analysis of Flood Flows at Zavlaka

To enable evaluation of performance of rainfall-runoff models, design floods are estimated by applying the frequency analysis. Several candidate distributions are considered: namely, log-normal, Gumbel, Pearson III and log-Pearson III. Distribution parameters are estimated by applying the method of moments, and fitness of the distribution is estimated by applying the Cramér–von Mises test [5]. The quantiles together with the confidence intervals are computed for return periods of 20, 50 and 100 years. Following confidence intervals are computed: 75% (α=0.25), 90% (α=0.10), 95% (α=0.05) and 99% (α=0.01). Details on the candidate distributions, confidence intervals and on the Cramér–von Mises test can be found in the literature [5].

2.3 Rainfall-Runoff Models for Design Computation of Flood

In this study several different models are used to compute design floods at the location of Zavlaka stream gauge. The models differ according to the design storms and hydrograph simulation method. The models are presented in detail in this section, while their key features are outlined in Table 2. Model acronyms in the table are created by combining acronyms of the hydrograph method, design hyetograph shape and duration. Although they could be referred to as modelling chains, the term “*model*” is used in this study for the sake of simplicity.

Design rainfall in all models is obtained straightforwardly from DDF curves at Loznica (section 2.1), with exception of model B_block_tmax. Specifically, this model uses DDF rescaled to match catchment average design rainfall, which results in slightly larger rainfall depths: for example, 100-year daily rainfall amounts to 110.6 mm, while the corresponding DDF value for Loznica amounts to 102.6 mm. Most models uses daily rainfall, however, it is discretised into finer time steps (see Table 2). In few models (acronyms in Table 2 contain “tmax”), rainfall duration is selected so that it provides the largest flood flows. The time steps at which models were run to detect the rainfall duration that yields the largest design floods are also given in the table. Three shapes of design hyetograph are used: rectangular shape, and non-uniform shapes obtained by applying the alternating block and Chicago methods [30]. Rainfall reduction with respect to catchment area is applied by all the models.

All models use the SCS-CN method to compute effective rainfall [31]. Although average antecedent conditions (AMC II, [31]) are adopted in all the models, catchment-average estimates of *CN* values slightly differ depending on the lookup tables used.

Different unit hydrograph models are used for runoff routing, as indicated in Table 2. Most models rely on the Jovanović-Brajković (JB), Brajković (B) or Ristić (R) unit hydrographs ([5],[13],[32],[33]). The unit hydrographs are defined by the time of rise T_p and time of recession T_r, while the peak ordinate u_{max} is computed from estimated runoff volume as follows:

$$u_{\max} = \frac{2A}{T_p + T_r} \tag{1}$$

where A denotes the catchment area.

The time of rise and recession time are computed from the selected design rainfall duration t_r, and also depend on the lag time t_p, i.e., time elapsed from the hyetograph centroid and peak of the hydrograph:

$$t_p = a\,t_r + t_0 \tag{2}$$

$$T_p = t_p + t_r / 2 \tag{3}$$

$$T_r = r T_p \tag{4}$$

where coefficients a and r depend on catchment properties, and recommendations on their values can be found in the literature ([30],[33]). Variable t_0 represents the lag time of the instantaneous unit hydrograph. The Jovanović-Brajković and Brajković unit hydrographs differ according to the equations used for computation of t_0 as follows:

<u>Jovanović-Brajković unit hydrograph</u>

$$t_0 = 0.4 L^{0.67} \left(\frac{L L_c}{\sqrt{I_u}} \right)^{0.086} \tag{5}$$

<u>Brajković unit hydrograph</u>

$$t_0 = 1.06 \left(\frac{L}{I_u} \right)^{0.47} \tag{6}$$

where L and L_c denote length of the river, and length from the catchment centroid to the outlet, respectively, while I_u represents mean slope of the river [33].

Ristić unit hydrograph implies immediate computation of t_p (irrespective of rainfall duration), which is used for computation of time of rise T_p, as follows:

$$t_p = 1.399 \left(\frac{L L_c}{\sqrt{I_u\, I_{mean}}} \right)^{0.315} \tag{7}$$

where I_{mean} denotes mean catchment slope [14].

Two of the models outlined in Table 2 are based on EBA4SUB, which simulated flow-weighted instantaneous unit hydrograph from the catchment digital terrain model raster [2]. In particular, the catchment instantaneous unit hydrograph (IUH) is obtained applying the Width Function (WF) framework. The so-called WFIUH is automatically calculated from digital terrain model flow paths and the estimated time of concentration, leading to the catchment travel time distribution. In detail, surface flow velocities are calculated based on the catchment slopes and land cover by employing empirical equations for hillslope cells, followed by calibration of flow velocity in channel cells to ensure that the projection of the WFIUH centre of mass on the temporal axis is equal to the basin lag time, expressed as 60% of the catchment time of concentration ([15],[16],[17]).

All models are spatially explicit, with exception of B_block_tmax (Table 2). Specifically, models created with the HEC-HMS software simulate runoff in the subcatchments (see Figure 1), as well as its routing along the river reaches by using the lag method [34].

Table 2. Key features of the rainfall-runoff models used in this study.

Model	**Rainfall Duration**	**Hyetograph Shape**	***CN* value**	**Unit Hydrograph**	**Software**
JB_block_d	1 day	rectangular	67.6	Jovanović-Brajković UH	HEC-HMS [34]
R_block_d	1 day	rectangular	67.6	Ristić UH	HEC-HMS [34]
JB_block_tmax	duration for max *Q*, analyses with 2h time steps	rectangular	67.6	Jovanović-Brajković UH	HEC-HMS [34]
R_block_tmax	duration for max *Q*, analyses with 2h time steps	rectangular	67.6	Ristić UH	HEC-HMS [34]
B_block_tmax	duration for max *Q*, analyses with 1min time steps	rectangular	79	Brajković UH	*lumped model
JB_ABM_d	1 day, 15 min discretisation	alt. block method	67.6	Jovanović-Brajković UH	HEC-HMS [34]
R_ABM_d	1 day, 15 min discretisation	alt. block method	67.6	Ristić UH	HEC-HMS [34]
EBA4SUB_block_d	1 day, 60 min discretisation	rectangular	67.6	flow-weighted instantaneous unit hydrograph	EBA4SUB [2]
EBA4SUB_Ch_d	1 day, 60 min discretisation	Chicago	67.6	flow-weighted instantaneous unit hydrograph	EBA4SUB [2]

2.4 Evaluation of Performance of Rainfall-Runoff Models in Reproducing Design Floods

The rainfall-runoff models are run with design rainfall of return periods of 20, 50 and 100 years, resulting in design floods of corresponding return periods. These design floods are compared to the quantiles computed from the frequency analysis, and the discrepancy between the two is quantified in terms of relative error. The relative errors are analysed with respect to both their magnitude and sign, to reveal if a model consistently under- or overestimates quantiles across the return periods.

The results of rainfall-runoff models are also compared to the confidence intervals of the quantiles (section 2.2). Specifically, a design flood estimate that is within the 75% confidence interval (α=0.25), it is closer to the expected value (the quantile) than an estimate obtained by another model located within the 99% confidence interval (α=0.01). Therefore, all design floods obtained by the models are assigned to five different categories in a way that a larger category value indicates greater departure from the expected value of design flood, i.e., quantile obtained from the adopted distribution. For example, estimates that are within the 75% confidence interval are categorised into group 1, estimates that are within the 90% confidence interval (but outside the 75% interval) are categorised into group 2, and so forth. The estimates that are outside the 99% confidence interval are categorised into group 5, and indicate unacceptably large errors in design flood estimates.

3 Results

3.1 Design Floods at the Zavlaka Stream Gauge

Fitted candidate distributions are presented in Figure 3, together with the probability plots of observed annual maxima at Zavlaka. The Cramér–von Mises test statistic values are given in Table 3. The critical values of the Cramér–von Mises test statistic $N\omega^2$ amounts to 0.462 for level of significance of 0.05, which is commonly adopted in engineering practice ([5],[7]). Comparison of the test statistics to the critical value shows that all candidate distributions are well fitted, and can be used for quantile estimation. In this study, the log-normal distribution is selected as the best fitted one, and the quantiles obtained from this distribution are used for evaluation of rainfall-runoff models. Selection of this distribution can also be justified by the skewness coefficient of the log-transformed series, which is approximately equal to zero (Table 1).

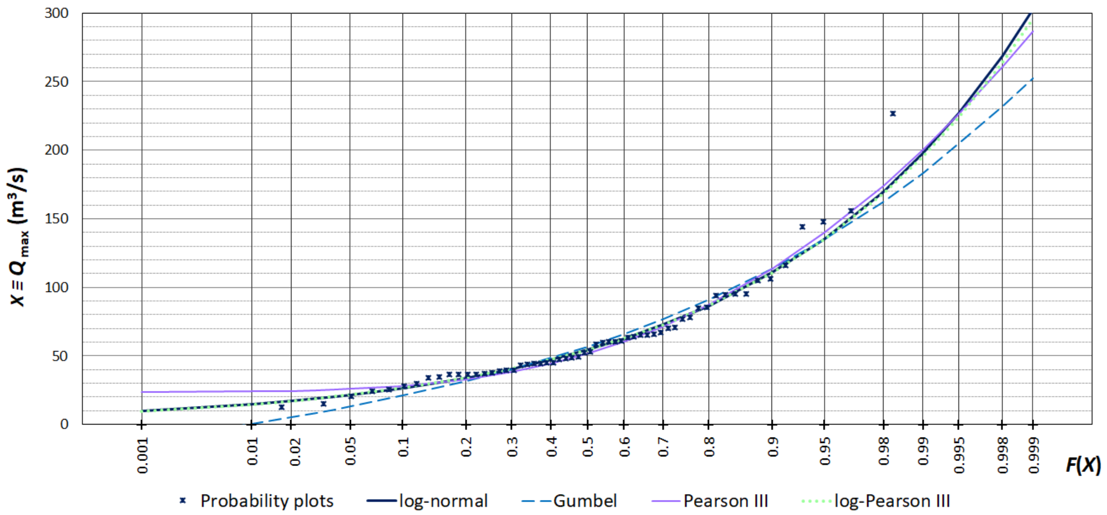

Figure 3. Fitted candidate distributions to the annual maxima observed at the Zavlaka stream gauge.

Table 3. Results of the Cramér–von Mises test.

Distribution	log-normal	Gumbel	Pearson III	log-Pearson III
Test statistic $N\omega^2$	0.046	0.138	0.076	0.047

Table 4. Design flood estimates (in m^3/s) obtained by using the rainfall-runoff models. The results for the three selected return periods are presented together with the quantiles obtained from the frequency analysis, and relative error (in %). Positive values of the relative error mean that a model overestimates the quantiles. Systematic under- and overestimation of the quantiles across the three return periods are highlighted in the table.

Model ID	Design Flood (m^3/s)			Relative Error (%)		
	20 years	50 years	100 years	20 years	50 years	100 years
quantiles – frequency analysis	135.3	169.8	197.6			
JB_block_d	99.1	133.5	162.6	-26.8	-21.4	-17.7
R_block_d	103.8	139.0	169.0	-23.3	-18.2	-14.5
JB_block_tmax	103.8	151.5	188.8	-23.3	-10.8	-4.5
R_block_tmax	114.2	166.6	206.3	-15.6	-1.9	4.4
B_block_tmax	150.5	217.4	276.6	11.2	28.0	40.0
JB_ABM_d	125.4	186.8	237.9	-7.3	10.0	20.4
R_ABM_d	139.0	207.6	264.8	2.7	22.2	34.0
EBA4SUB_block_d	109.9	141.5	163.4	-18.8	-16.7	-17.3
EBA4SUB_Ch_d	247.7	326.3	413.8	83.0	92.1	109.4

The design floods of 20-, 50- and 100-year return periods obtained by the selected rainfall-runoff models are presented in Table 4, together with the corresponding results of the quantiles computed from the frequency analysis. The design flood estimates considerably vary across the models. Models based on daily design rainfall of constant intensity (i.e., rectangular hyetographs) result in lower estimates in comparison to the remaining model. The greatest design flood estimates are obtained by the EBA4SUB_Ch_d, while the other models based on non-uniform design rainfall hyetographs do yield much smaller design flood estimates.

3.2 Performance of Rainfall-Runoff Models in Reproducing Design Floods

To facilitate comparison of design flood estimates obtained by using the rainfall-runoff models to the corresponding quantiles, relative errors are presented in Table 4. The relative errors vary considerably across the models and the return periods. No distinct patterns in change of relative errors with return period can be detected. For example, absolute values of relative errors increase with the return period in e.g., B_block_tmax or EBA4SUB_Ch_d, as opposed to JB_block_tmax or EBA4SUB_block_d, which yields fairly similar errors across the return periods.

Most models systematically either under- or overestimate the quantiles. The models run with daily rainfall of uniform intensity (i.e., with rectangular hyetographs) systematically underestimate the quantiles, as opposed to the model with Chicago hyetograph. Systematic overestimation is detected in models with different rainfall durations and hyetograph shapes. Only two models do not exhibit systematic errors across the return periods, and they also, on average, yield the lowest values of the relative errors.

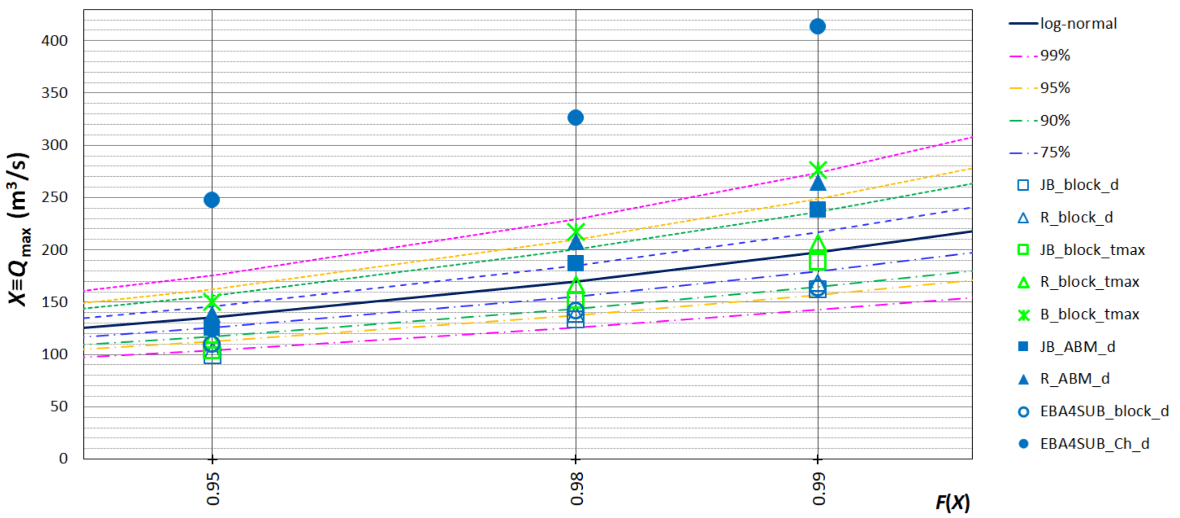

Figure 4. Fitted log-normal distribution and its 75%, 90%, 95% and 99% confidence intervals, and the design estimates obtained by the rainfall runoff models.

The design floods simulated by the models are shown together with the fitted log-normal distribution and its confidence intervals in Figure 4. The design flood estimates are categorised according to the confidence interval they are contained by (see section 2.4), and these categories are illustrated in Figure 5. These alternative approaches to the representation of design flood estimates enable a more comprehensive insight in discrepancies between the two types of design flood estimates (i.e., results of frequency analysis and rainfall-runoff modelling) than mere relative error values. For example, the relative error values of EBA4SUB_block_d are largely similar across the return periods, however, 20-year design flood is outside the 95% confidence interval, as opposed to the remaining two periods. In other words, 20-year design flood estimate obtained by this model can be considered more uncertain than the estimates of 50- and 100-year design floods, and this cannot be inferred solely from the relative error values. Similarly, 100-year design flood obtained by R_AMB_d has rather high value (34%), but it does not exceed 99% confidence interval, which is not the case for many other estimates that yield relative errors of absolute values of ~20%.

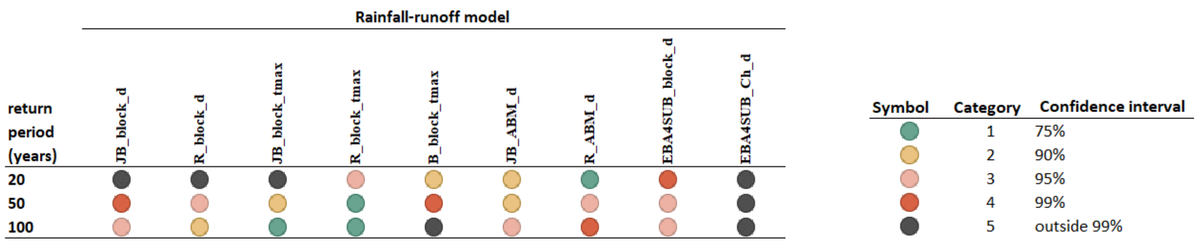

Figure 5. Model performance with respect to the confidence intervals of the quantiles of 20-, 50- and 100-year return periods. The design flood estimates obtained by the rainfall-runoff models are categorised into five groups depending on the confidence interval they are enclosed within. Lower categories implies smaller departure of the estimate from the quantile. The estimates of category 5 are outside the 99% confidence interval of quantiles.

4 Discussion

This study presents a comprehensive analysis of performance of nine different rainfall-runoff models with respect to simulation of design floods. Model performance in this regard is commonly appraised from the relative error values, not taking into account uncertainty of a quantile of interest. Our results clearly show that model performance should not be appraised solely on the values of relative errors. Specifically, the same relative error value can imply that design flood estimate is within e.g., 75% confidence interval in one, or that it exceeds 99% confidence interval in the other case. We, therefore, suggest taking the uncertainty of the quantiles obtained from the frequency analyses into consideration throughout evaluation of model performance, i.e., credibility of the resulting design floods. The relative errors can reveal systematic under- or overestimation of the quantiles; hence, considerations of quantile uncertainties should not replace them, but rather complement them.

The results obtained in this study clearly suggest that models run with daily rainfall of uniform intensity are shown to systematically underestimate design floods. This corroborates the results presented by Plavšić et al. [1]. Underestimated design floods inevitably lead to undersized hydraulic structures, therefore, that this type of error is unacceptable from the standpoint of civil engineering. These results also suggest that, if a rectangular hyetograph is used, rainfall duration is essential for accurate design flood estimation. The models that use rainfall duration that yields the largest design floods result in both under- and overestimation of the quantiles. Such behaviour can be explained by different *CN* values and time steps at which this "optimal" rainfall duration is obtained. Additionally, "optimal" runoff duration in case of B_block_tmax is approximately 2 h, and 12 h in case of JB_block_tmax and R_block_tmax. These results clearly show an interplay between optimal design rainfall duration and *CN* values, which can be considered a kind of "equifinality" [35] in these types of models. Therefore, further research is needed to provide guidance on inferring the "optimal" rainfall duration in case of rectangular unit hydrographs. It is well known, in fact, that many applications follow the hypothesis that the maximum peak discharge is caused by a rainfall with a duration equal to the catchment concentration time, but this hypothesis is debated in literature. Indeed, in many practical applications, rainfall durations 2-3 times larger than the time of concentration are often used in order to maximise the peak discharge [36].

The models that simulate time-varying rainfall intensities exhibit a wide range of behaviours in terms of relative errors, although there is a general tendency to overestimation of the quantiles, and even the upper limits of 95% or 99% confidence intervals. The overestimation is particularly pronounced in case of the model that uses Chicago design hyetographs. These results suggest that design rainfall represents a key source of uncertainty in design flood estimation with event-based models.

This study is based on nine models and only one catchment. Further research is needed to test the validity of the conclusions presented here. Such research should include a larger number of models and catchments, and should be accompanied by regional analyses. Application of information criteria for best model selection [21] or application of multi-model combination methods to estimate design flood [37] also present promising avenues of research in this field.

5 Conclusions

This study presents a comprehensive evaluation of nine event-based models formed by making different modelling decisions at every step of their development. The models are evaluated with respect how well they can reproduce design flood estimates obtained by applying a frequency analysis, which is generally considered a standard approach in engineering practice. The objectives of this study are twofold: (1) to examine if commonly used relative error is versatile enough to reveal model robustness in simulating design floods, and (2) to identify the best modelling decisions in order to obtain credible design flood estimates.

This study clearly shows that relative error values should be complemented by additional comparisons of the design flood estimates to confidence intervals of the corresponding quantiles. In this way, quantile uncertainty can be taken into account, and the models could be evaluated more thoroughly.

Design rainfall duration is essential in models that assume uniform rainfall intensity, however, it can be compensated by *CN* values, suggesting the “equifinality” between the two. Generally, daily rainfall should not be used with design rainfall of uniform intensity in catchments with times of concentration shorter than one day. Models that use non-uniform design rainfall intensity generally tend to overestimate the quantiles obtained from the frequency analysis, even with rainfall duration longer that the catchment time of concentration. This is particularly pronounced with the Chicago design hyetograph. Identification of best modelling decisions requires further research that has to include a greater number of models and catchments.

Acknowledgement

This paper represents the result of bilateral collaboration between the Tuscia University at Viterbo and Faculty of Civil Engineering of the University of Belgrade. It is also done as a part of PhD Thesis of Nikola Zlatanovic. The data on annual maxima were provided by the Republic Hydrometeorological Service of Serbia (RHMSS).

References

[1] J. Plavšić, N. Zlatanović, and A. Todorović, “Design storm duration for estimation of floods in ungauged basins,” in *CONFERENCE PROCEEDINGS 7TH INTERNATIONAL CONFERENCE CONTEMPORARY ACHIEVEMENTS IN CIVIL ENGINEERING 2019*, 2019, pp. 77–86.

[2] A. Petroselli, S. Asgharinia, T. Sabzevari, and B. Saghafian, “Comparison of design peak flow estimation methods for ungauged basins in Iran,” *Hydrol. Sci. J.*, vol. 65, no. 1, pp. 127–137, 2020, doi: 10.1080/02626667.2019.1686506.

[3] E. Paquet, F. Garavaglia, R. Garçon, and J. Gailhard, “The SCHADEX method: A semi-continuous rainfall–runoff simulation for extreme flood estimation,” *J. Hydrol.*, vol. 495, pp. 23–37, Jul. 2013, doi: 10.1016/j.jhydrol.2013.04.045.

[4] M. Jovanović, A. Todorović, and M. Rodić, “Kartiranje rizika od poplava,” *Vodoprivreda*, vol. 41, no. 1–6, pp. 31–45, 2009.

[5] J. Plavšić, “Inženjerska hidrologija,” 2019. http://hikom.grf.bg.ac.rs/wp-content/uploads/2017/10/skripta_2_ModeliranjeOticaja.pdf.

[6] H. Tabari, “Extreme value analysis dilemma for climate change impact assessment on global flood and extreme precipitation,” *J. Hydrol.*, vol. 593, no. August 2020, p. 125932, Feb. 2021, doi: 10.1016/j.jhydrol.2020.125932.
[7] N. T. Kottegoda and R. Rosso, *Applied Statistics for Civil and Environmental Engineers*, Second Edi. Oxford, United Kingdom: Blackwell Publishing, 2008.
[8] J. Plavšić, “Analiza rizika od poplava pomoću prekidnih slučajnih procesa,” Univerzitet u Beogradu, 2005.
[9] I. G. G. Pechlivanidis, B. M. Jackson, N. R. Mcintyre, and H. S. Wheater, “Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications,” *Glob. NEST J.*, vol. 13, no. 3, pp. 193–214, 2011, doi: 10.30955/gnj.000778.
[10] S. Grimaldi, F. Nardi, R. Piscopia, A. Petroselli, and C. Apollonio, “Continuous hydrologic modelling for design simulation in small and ungauged basins: A step forward and some tests for its practical use,” *J. Hydrol.*, vol. 595, p. 125664, Apr. 2021, doi: 10.1016/j.jhydrol.2020.125664.
[11] A. Todorović, M. Stanić, Ž. Vasilić, and J. Plavšić, “The 3DNet-Catch hydrologic model: Development and evaluation,” *J. Hydrol.*, vol. 568, pp. 26–45, Jan. 2019, doi: 10.1016/j.jhydrol.2018.10.040.
[12] W. Boughton and O. Droop, “Continuous simulation for design flood estimation—a review,” *Environ. Model. Softw.*, vol. 18, no. 4, pp. 309–318, Apr. 2003, doi: 10.1016/S1364-8152(03)00004-5.
[13] M. Rogger *et al.*, “Runoff models and flood frequency statistics for design flood estimation in Austria - Do they tell a consistent story?,” *J. Hydrol.*, vol. 456–457, pp. 30–43, 2012, doi: 10.1016/j.jhydrol.2012.05.068.
[14] R. Ristić, “Vreme kašnjenja oticaja na bujičnim slivovima u Srbiji,” *Glas. šumerskog Fak.*, vol. 87, pp. 51–65, 2003, [Online]. Available: https://scindeks-clanci.ceon.rs/data/pdf/0353-4537/2003/0353-45370387051R.pdf.
[15] R. Piscopia, A. Petroselli, and S. Grimaldi, “A software package for predicting design-flood hydrographs in small and ungauged basins,” *J. Agric. Eng.*, vol. 46, no. 2, p. 74, Jun. 2015, doi: 10.4081/jae.2015.432.
[16] A. Petroselli and S. Grimaldi, “Design hydrograph estimation in small and fully ungauged basins: a preliminary assessment of the EBA4SUB framework,” *J. Flood Risk Manag.*, vol. 11, pp. S197–S210, Jan. 2018, doi: 10.1111/jfr3.12193.
[17] A. Petroselli, R. Piscopia, and S. Grimaldi, “Design discharge estimation in small and ungauged basins: EBA4SUB framework sensitivity analysis,” *J. Agric. Eng.*, vol. 51, no. 2, pp. 107–118, Jun. 2020, doi: 10.4081/jae.2020.1040.
[18] S. Grimaldi, A. Petroselli, and F. Serinaldi, “A continuous simulation model for design-hydrograph estimation in small and ungauged watersheds,” *Hydrol. Sci. J.*, vol. 57, no. 6, pp. 1035–1051, Aug. 2012, doi: 10.1080/02626667.2012.702214.
[19] J. Smithers, “Methods for design flood estimation in South Africa,” *Water SA*, vol. 38, no. 4, pp. 633–646, Sep. 2012, doi: 10.4314/wsa.v38i4.19.
[20] J. Plavšić, V. Mihailović, and B. Blagojević, “Assessment of Methods for Outlier Detection and,” in *Proceedings of the Mediterranean Meeting on "Monitoring, modelling and early warning of extreme events triggered by heavy rainfalls"*, 2014, pp. 181–192.
[21] G. Di Baldassarre, F. Laio, and A. Montanari, “Design flood estimation using model selection criteria,” *Phys. Chem. Earth, Parts A/B/C*, vol. 34, no. 10–12, pp. 606–611, 2009, doi: 10.1016/j.pce.2008.10.066.
[22] W. H. Farmer, T. M. Over, and J. E. Kiang, “Bias correction of simulated historical daily streamflow at ungauged locations by using independently estimated flow duration curves,” *Hydrol. Earth Syst. Sci.*, vol. 22, no. 11, pp. 5741–5758, Nov. 2018, doi: 10.5194/hess-22-5741-2018.
[23] N. Mizukami *et al.*, “On the choice of calibration metrics for ‘high-flow’ estimation using hydrologic models,” *Hydrol. Earth Syst. Sci.*, vol. 23, no. 6, pp. 2601–2614, Jun. 2019, doi: 10.5194/hess-23-2601-2019.
[24] K. K. Yilmaz, J. A. Vrugt, H. V. Gupta, and S. Sorooshian, “Model Calibration in Watershed Hydrology,” in *Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting*, B. Sivakumar and R. Berndtsson, Eds. Singapore: World Scientific Publishing, 2010, pp. 53–105.
[25] R. Eric, A. Todorovic, J. Plavsic, and V. Djukic, “Rainfall-runoff simulations in the Lukovska River Basin with the HEC-HMS model,” *Glas. Sumar. Fak.*, no. 119, pp. 33–60, 2019, doi: 10.2298/GSF1919033E.
[26] P. Brigode *et al.*, “Sensitivity analysis of SCHADEX extreme flood estimations to observed hydrometeorological variability,” *Water Resour. Res.*, vol. 50, no. 1, pp. 353–370, 2014, doi: 10.1002/2013WR013687.
[27] A. Viglione, R. Merz, and G. Blöschl, “On the role of the runoff coefficient in the mapping of rainfall to flood return periods,” *Hydrol. Earth Syst. Sci.*, vol. 13, no. 5, pp. 577–593, May 2009, doi: 10.5194/hess-13-577-2009.

[28] S. Prohaska *et al.*, *Intenziteti jakih kiša u Srbiji*. Beograd: Institut za vodoprivredu "Jaroslav Černi," 2014.
[29] Copernicus, "CORINE 2018." https://land.copernicus.eu/pan-european/corine-land-cover/clc2018 (accessed Aug. 31, 2021).
[30] M. Wanielista, R. Kersten, and R. Eaglin, *Hydrology: water quantity and quality control*, 2nd Editio. John Wiley & Sons Ltd, 1997.
[31] Natural Resources Conservation Service, "Estimation of Direct Runoff from Storm Rainfall," in *Hydrology National Engineering Handbook Part 630*, United States Department of Agriculture, 2004, p. 79.
[32] S. Jovanović and Z. Radić, *Parametarska hidrologija*. Građevinski fakultet u Beogradu, 1990.
[33] S. Jovanović, "Hidrologija," in *Tehničar 6*, Beograd: Građevinska knjiga, 1989.
[34] A. Feldman, "Hydrologic Modeling System HEC-HMS - Technical Reference Manual," 2000.
[35] K. Beven, A. Binly, A. Binley, and A. Binly, "The Future of Distributed Models: Model Calibration and Uncertainty Prediction," *Hydrol. Process.*, vol. 6, no. 3, pp. 279–298, 1992, doi: 10.1002/hyp.3360060305.
[36] A. E. Sikorska, D. Viviroli, and J. Seibert, "Effective precipitation duration for runoff peaks based on catchment modelling," *J. Hydrol.*, vol. 556, pp. 510–522, 2017, doi: 10.1016/j.jhydrol.2017.11.028.
[37] K. Okoli, M. Mazzoleni, K. Breinl, and G. Di Baldassarre, "A systematic comparison of statistical and hydrological methods for design flood estimation," *Hydrol. Res.*, vol. 50, no. 6, pp. 1665–1678, Dec. 2019, doi: 10.2166/nh.2019.188.