Flood frequency analysis using mean daily flows vs. instantaneous peak flows

Anne Bartens[1], Bora Shehu[2], Uwe Haberlandt[1]

[1]Institute of Hydrology and Water Resources Management, Leibniz University of Hannover, Germany
[2]Institute for Environmental Sciences and Geography, University of Potsdam, Germany

Correspondence to: ~~Anne Bartens (fangmann~~Uwe Haberlandt (haberlandt@iww.uni-hannover.de)

Abstract. In many cases flood frequency analysis (FFA) needs to be carried out on mean daily flow (MDF) ~~series without any available information on the~~ instead of instantaneous peak flow (IPF~~). We analyze~~), which can lead to underestimation of design flows. Typically, correction methods are applied on the MDF data to account for such underestimation. In this study, we first analyse the error ~~of~~ distribution of MDF derived flood quantiles over 648 catchments in Germany. The results show that using ~~MDFs instead of IPFs for flood quantile estimation on a German dataset and assess spatial patterns~~MDF instead of IPF data can lead to underestimation of the mean annual maximum flow (MHQ) up to 80% and ~~factors that influence the deviation of MDF floods from their IPF counterparts. The main dependence could be found for~~ is mainly dependable on the catchment area but ~~also~~ appears to be influenced as well from gauge elevation ~~appeared to have some influence. Based on the findings we propose simple~~. This relationship is shown to be different for summer vs winter floods. To correct such underestimation, different linear models ~~to correct both MDF flood peaks of individual flood events~~ based on predictors derived from MDF hydrograph and ~~overall MDF flood statistics.~~ catchment characteristics are investigated. Apart from the catchment area, a key predictor in ~~the~~these models is the event-based ratio of flood peak and flood volume ~~obtained directly from the daily flow records. This correction approach requires~~(p/V ratio) obtained by MDF data. The p/V models applied either on MDF derived events or statistics, seem to outperform other reference correction methods. Moreover, they require a minimum of data input, ~~is~~are easily applied~~,~~ and valid for the entire study area ~~and successfully estimates IPF peaks and flood statistics. The models perform~~. Best results are achieved when the L-moments of the MDF annual maximum series are corrected with the proposed model, which reduces the flood quantile errors up to 60%. This approach behaves particularly well in smaller catchments~~,~~ (<500km^2), where ~~other IPF estimation~~reference methods fall short. ~~Still~~However, the limit of the proposed approach is reached for catchment sizes below 100 km², where the hydrograph information from the daily series is no longer capable of approximating instantaneous flood dynamics, and gauge elevation below 100m, where the difference between MDF and IPF floods is very small.

1 Introduction

Common flood frequency analysis (FFA) is based on samples of maximum flows~~.~~, e.g. annual maximum flow series (AMS). The magnitude and variability of these maxima pose the baseline for the choice of probability distribution, the estimation of its parameters and eventually the deduction of flood quantiles as design criteria~~.~~ for various water works (Maidment, 1993). For FFA to be as accurate as possible, ~~it is important~~ two criteria need to ~~have~~be met; first a large number of observed peak flows is necessary to ensure an adequate selection and fitting of the probability distribution, and secondly it is important that the peak flows are measured with high precision~~, so that~~ in order to account for the best description of maximum flood magnitude and dynamics ~~are well assessable~~.

However, embracing the true dimension of a peak requires continuous measurement of the flow on a high temporal resolution~~.~~ (e.g. at 15min time steps). Such data is rarely available, or at the best case only available for short periods, which is insufficient for flood frequency analysis. Typically, long observation of floods are available as mean daily flows records and oftentimes FFA needs to be carried out on ~~average daily flow~~these records instead. The daily averaging naturally flattens the flood peak, and the true maximum becomes unknowable. Particularly for small basins, there is a considerable underestimation of flood peak by the mean daily flows (Fill and Steiner, 2003). Hence it becomes essential to develop new methods based on easily accessible data to correct the mean daily flows for a better representation of the flood peaks.

The degree of ~~this~~the above-mentioned smoothing, i.e. the difference between the true instantaneous peak flow (IPF) and the maximum mean daily flow (MDF) (herein referred to as the peak ratio), depends on the response time of a system, which is controlled by a multitude of factors. The average relationship between MDF and IPF peaks at a site depends greatly on its basin area (Fuller, 1914) and characteristics related to topography~~.~~; like altitude, relief and channel slope (Canuti and Moisello, 1982). For instance, there is a visible trend of the peak ratio to decrease with the basin area, which is expected as larger basins have higher baseflows (Ellis and Grey, 1966). Furthermore, the internal variability of the ~~MDF IPF~~peak ratio ~~within~~at a site's flow record is largely determined by the type of meteorological input causing the individual flood events (Viglione and Blöschl, 2009; Gaál et al., 2015).~~2013).~~ This means that the peak ratio of rainfall and snowmelt events are different from one another.

A variety of studies make use of the dependencies named above in order to estimate IPFs from MDFs, ~~including~~ and can be generally classified as methods based on the catchment characteristics as in Fuller (1914), Ellis & Gray (1966), Canuti & Moisello (1982), Ding et al. (2015) or including also climate characteristics as in Taguas et al. (2008), Muñoz et al. (2012) and ~~Ding et al. (~~Gaal et al. (2015). Mostly, these methods are in the form of linear models based on maximum MDFs and the selected catchment or climate predictors.~~2015).~~

Other IPF estimation methods aim at using the bare minimum of available data, i.e~~.~~., solely the available mean daily flow record~~.~~ (MDF). In these cases, ~~usually~~ the shapes of hydrographs are used to estimate the instantaneous peaks of events. The shape of a hydrograph can hold important information regarding an event's or even the entire site's flashiness and thus its ~~MDF/IPF~~peak ratio. Short flood events with steep rising and falling limbs are typical of a quickly reacting system, due to limited storage capacity and/or high intensity rainfall~~.~~ or due to moderate intensity rainfall on snow. In such ~~a case~~events, the

discrepancy between IPF and MDF will be significantly larger than for hydrographs with long durations and gentle slopes. ~~E.g.,~~For example, Ellis & Gray (1966) found that the peak ratio ~~IPF/MDF~~ distinctively decreases with increasing hydrograph width.

Several approaches use the maximum mean daily flow and the discharge of the previous and/or successive day (e.g. Langbein, 1944) to predict IPFs. Chen et al. (2017) ~~compare two~~compared three of these methods, namely those of Sangal (1983~~) and~~), Fill and Steiner (2003~~). They also propose~~) and their own new method (referred to as the slope-method). These methods are based on the rising and falling slopes of the event hydrograph, estimated from the three consecutive days around the peak~~.~~, and differ on how the information is integrated in the formula. They found that their slope-~~based~~ method and Fill and Steiner's method ~~perform well~~outperform the other two approaches and Fullers' method (Fuller, 1914) (estimation method based on basin area), and are probably applicable under a wide range of climates. However, both methods' performances ~~decrease~~deteriorate with decreasing catchment size and work best for areas larger than 500 km².

~~There naturally~~Of course, there exist more complex means to correct the divergence between MDFs and IPFs. This includes disaggregation of the daily flow series to a finer scale, as done by e.g. Stedinger and Vogel (1984), Tarboton et al. (1998), Kumar et al. (2000), Tan et al. (2007) and Acharya and Ryu (2014). Also, hydrological modelling may be applied for IPF estimation, e.g. in combination with high-resolution disaggregated rainfall (Ding et al., 2016) ~~and~~or by using regionalized model parameters (Ding and Haberlandt, 2017). Several studies have applied machine learning techniques to estimate instantaneous peaks from mean daily ~~data~~flows, including Shabani & Shabani (2012), Dastorani et al. (2013) and Jimeno-Sáez et al. (2017). While disaggregation, hydrological modeling and machine learning prove very effective in their studies, they often ~~requires~~require a number of computational steps and/or a variety of data sources. Indeed, the estimation methods based on the catchment or hydrograph characteristics remain still more desirable due to their simplicity, as they are based on easily accessible data and popular methods (i.e. linear models).

~~This study aims at analyzing the differences between IPF and MDF with focus on flood frequency. The errors in mean maximum flows, distribution parameters and flood quantiles are assessed and analyzed for spatial patterns. Based on the findings, a method is proposed that facilitates~~So far, the two main IPF estimation methods are developed distinctively from one another with no combination of both catchment and hydrograph information. In this study we propose linear models that facilitate IPF estimation using a combination of daily event hydrographs and functional dependencies with geomorphic catchment descriptors, while keeping the data input to a minimum. Key predictor in this approach is the ratio of direct event peak runoff and direct event volume. This ratio is expected to effectually describe the shape of a flood event, which in turn gives an idea about the expected instantaneous peak: the larger the daily peak and the smaller the event volume, the larger the expected difference between IPF and MDF and vice versa. We assume that the peak-volume ratio (p/V) holds important information on the general behavior of flood events (Tan et al., 2006; Gaál et al. 2015; Fischer, 2018~~),~~) and thus on the expected magnitude of the IPF~~.~~ peaks as well. Moreover, the p/V of individual events can describe the internal variability at a site by reflecting different types of floods caused by different rainfall and/or snowmelt inputs. At the same time the p/V accounts for

the variability between sites caused by local flood generating processes governed by general physiographic and climatic conditions. ~~Accordingly, the proposed method is tested for IPF estimation for individual events, which are then used for FFA, and for direct correction of site-specific distribution parameters and flood quantiles.~~

Another important point to be considered is that most of the studies mentioned before investigate the performance on IPF maximum series and pay little attention to how these methods estimate the design flows with specific return periods. The general assumption is that, if the IPF maximum series are estimated well enough on average, so are the IPF quantiles. However, a well estimated average IPF maximum may still lead to underestimation of design flows with a high return period (say 100years). It makes sense to investigate as well if linear models based on MDF- moments, parameters or quantiles are more favorable for the estimation of the IPF quantiles. Accordingly, p/V models are employed here to correct MDF information at different levels; correction of individual flood events from MDF, correction of MDF annual or seasonal maximum series, and the direct correction of MDF-derived statistics (like mean maximum flow, L-moments, distribution parameters or even flood quantiles).

In this study, the linear models based on the *p/V* as key predictor (referred here as p/V models) are developed and assessed based on flow data from 648 catchments in Germany (as described in Section 2). The description of methods and models used here for the estimation of the IPF from MDF information is given in Section 3.2. We then analyze the performance of the models in two main parts: their ability to estimate the mean maximum flow (MHQ) (see Section 4.1) and their ability to estimate probability distribution and the respective design floods (see Section 4.2). For the best model achieved, an uncertainty estimation is tackled by means of spatio-temporal resampling (see Section 4.3). Finally, the range and limitations of the proposed methodology and conclusions are given in Sections 4.4 and 5.

2 Study area and data

This study uses flow data from 648 catchments distributed over Germany as shown in Figure 1. For the analyses, continuous average daily flow (MDF) and instantaneous peak flows provided for each month as monthly peaks (IPF) are available. The selected sites represent the datasets of the federal agencies, who provide online access to both datasets (Lower Saxony, Saxony-Anhalt, Saxony, Bavaria and Baden-Württemberg; see section Data Availability).

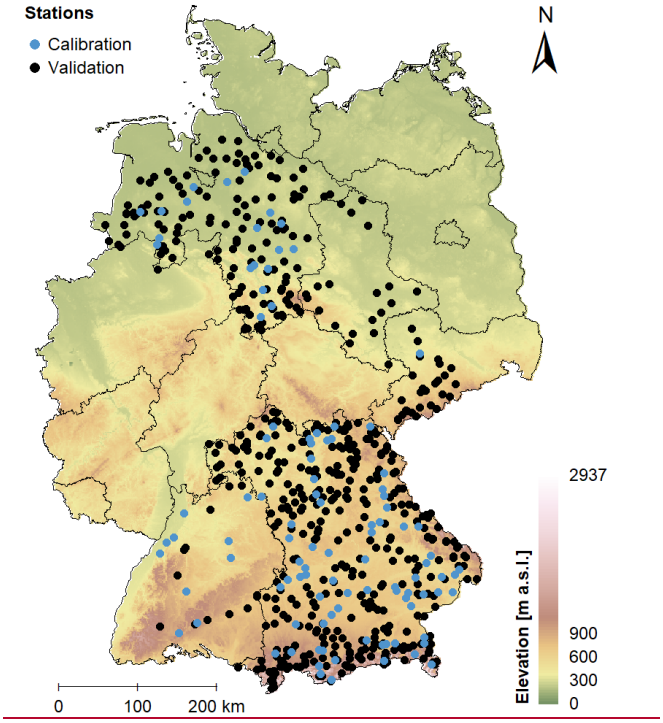


Figure 1: The spatial distribution of the 648 catchments and their respective discharge gauges employed in this study. The 103 sites used for model calibration are marked in blue. The elevation is shown in the background colors and is provided by Jarvis et al. (2008), while the borders of the German federal states are shown with black lines.

Germany poses a transition zone from an oceanic climate in the northwest to a humid continental climate in the southeast. The northwestern parts are influenced by wet air and have mild winters, while the more southeastern parts are drier and exhibit larger temperature ranges. The average temperature for the entire country is 8.9 °C, the monthly averages ranging between 0.4°C in January and 18°C in July (reference period 1981-2010; DWD). The average annual precipitation is 819 mm, where amounts generally decrease in west-east direction and in strong dependence on topography. Annual rainfall sums are generally highest over the Alps at the very Southern border and the various secondary mountain ranges. The flat continental east is the driest. Temporally, the summer months are wettest with rainfall often occurring in convective events. Snowfall occurs between October and April, where amount and depth of snow cover increase with decreasing oceanic influence and increasing altitude. Even though not the entire area of Germany is covered by the available data, the selected sites provide a cross section through the climatically and topographically distinct regions, from the flat oceanic northwest to the mountainous continental southeast. The lengths of the discharge records vary substantially from 11 to 183 years with a mean of 48.4 years (temporal span from 1831 to 2021). For the general assessment of differences in IPF and MDF floods and final model validation, all 648 sites with their variable record lengths are considered. For assessment of flood frequency criteria only those sites with at least 30 years of observations were used (486). Model fitting (herein referred as calibration) was carried out on a subset of 103 sites, whose discharge data were thoroughly checked. Also, their records were cropped to a common period from 1979 to 2012, to eliminate

potential non-stationary effects. For the 103 gauges used for calibration a catalogue of catchment descriptors is available. For the remaining gauges only rudimentary information was obtained, i.e. catchment size, geographical position and altitude of the gauges. Fig. 2 shows how the 648 discharge gauges are distributed in terms of catchment size and gauge elevation. It is evident that the majority of the sites have catchment areas below 500km^2 and gauges situated at elevations higher than 100m a.s.l..

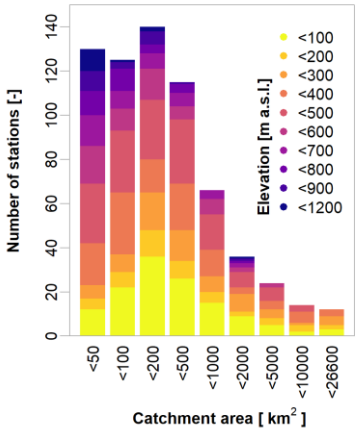


Figure 2: Distribution of catchment size and elevation for the 648 sites employed in this study.

3 Methods

3.1 Flood frequency analysis

Flood frequency analysis (FFA) is applied on the two datasets for the available catchments in Germany: mean daily flows (MDF) and instantaneous monthly peak flows (IPF). First the maximum series are extracted from each dataset either on an annual basis (annual maximum series – AMS) or for each season summer vs winter (seasonal maximum series). For extrapolation of the maximum series and estimation of floods with specific return periods, distributions were fitted to the annual and seasonal samples of both IPF and MDF datasets. This enables the direct comparison of both flood quantiles and distribution parameters. Here, the General Extreme Value distribution (GEV) of the following form was used for all samples (Maidment, 1993):

$$F(x) = \exp\left\{-\left[1 - k \cdot \frac{(x-\xi)}{\alpha}\right]^{\frac{1}{k}}\right\}, \quad (1)$$

with location parameter ξ, scale parameter α and shape parameter k. The parameters are estimated using sample L- moments (Hosking and Wallis 1997). The GEV has been proven before to be a suitable distribution for different catchments in Germany

as indicated by Haktanir and Horlacher (1993), Villarini et al. (2011), Ding et al. (2015,2016), Ding and Haberlandt (2017) and therefore has been chosen in our study as well. The goodness of fit of the distributions was determined with the Cramer-von-Mises test.

When extracting annual maximum series (AMS) different flood peaks from different genesis (i.e. from convective/stratiform rainfall, from snowmelt and so on) are mixed together and described by a single GEV distribution. However, if a certain flood type is dominating the annual maxima sample but is not typical for extremely large floods, then the fitted GEV distribution becomes misleading. To consider the different genesis in the flood peaks, maximum series are derived here for two seasons; summer (May-October) and winter (November-April). Then a mixed model is applied, which combines two GEV distributions fitted to each of these subsamples of the data, like summer and winter floods. A simple maximum mixing approach is used to combine the individual distributions in order to assess the annual non-exceedance probability of specific flood values:

$$F_{\text{mix}}(x) = \prod_{i=1}^{n} f_i(x). \quad (2)$$

with $f_i(x)$ as the annual non-exceedance probability calculated for each sub-sample (summer and winter) and $F_{mix}(x)$ as the mixed-model annual non-exceedance probability for a flood value x. This approach allows the combined estimation of flood quantiles from multiple underlying distributions and thus the assessment of errors in seasonal FFA. The approach is described in detail in Fischer et al. (2016). In their study they used thresholds to determine whether a seasonal maximum is actually a flood event, which may not be the case during dry summers. This threshold was defined as the minimum annual maximum flow. We do not censor our data with thresholds, i.e. for matters of simplicity we assume that every seasonal maximum is indeed a flood event.

~~**2 Methods**~~

~~2.1~~3.2 Analysis and estimation of ~~IPF peaks~~instantaneous peak flows (IPF)

~~In a first step, the general differences between IPF and MDF flood peaks are analyzed. Since for IPF monthly maximum flows are the only available information (see chapter 3), a direct comparison for each flood event is not possible. Instead, we focus on the analysis of flood statistics. The percentage deviation~~3.2.1 Calculation of the ~~MDF statistic MDF_{stat}~~p/V predictor from ~~the IPF statistic IPF_{stat}~~

$$\frac{\text{MDF}_{\text{stat}} - \text{IPF}_{\text{stat}}}{\text{IPF}_{\text{stat}}} * 100\ \% \quad (1)$$

~~is computed at each station for any desired quantity *stat*, like the~~ mean ~~annual maximum flow (MHQ), L-moments (Hosking, 1990), distribution parameters and flood quantiles.~~ daily flows

~~In order to improve the IPF estimation by MDF, several correction methods are applied, which make use of the~~ Motivated by the recent findings of Fischer et al. (2016) and Fischer (2018) regarding different flood types, here the flood peak-volume ratio~~.~~ *p/V* extracted from mean daily flows (MDF) is considered an important predictor that can help to estimate more accurately the IPF series from the MDF ones. This ratio is computed for ~~events in the average daily time series using the direct peak flow Q_{dir} and the direct flood volume Vol_{dir}, calculated after baseflow subtraction~~each flood event extracted from the MDF dataset as shown by Eqn. (3):

$$\mathrm{p/V}\left[\frac{1}{d}\right] = \frac{\mathrm{Q_{dir}}\,[\mathrm{m^3 d^{-1}}]}{\mathrm{Vol_{dir}}\,[\mathrm{m^3}]}\,. \qquad (~~2~~3)$$

where *p/V* is the peak-volume ratio, Q_{dir} is the direct peak flow, Vol_{dir} the direct flood volume. Both Q_{dir} and Vol_{dir} are calculated on flood events extracted from the MDF series after subtracting the baseflow.

For separation of the flood events from MDF, the initial steps of the procedure used by Tarasova et al. (2018) are carried out, which are proven effective and convenient for the German catchments. For the initial step of baseflow separation they selected the simple non-parametric algorithm by the Institute of Hydrology (1980), which was able to identify the starting points of events in daily flow series in a wide range of catchments. The same method is applied to the series of mean daily flows in our study, which involves the following steps. First, 5-day non-overlapping blocks are used to find minima, which are identified as turning points if they are more than 1.1 times smaller than their neighboring minima. The baseflow is then derived by simple linear interpolation between the turning points. Discharge peaks are subsequently determined from the flow series and for every peak the start and end of the belonging flow event is defined by the nearest surrounding turning points. To prevent false identification of events due to natural variability, events are discarded if their peak discharge is not at least 10% larger than the baseflow. Tarasova et al. (2018) suggested a second step of re-defining events with multiple peaks in an iterative procedure. This step is not carried out here, as it requires rainfall and snowmelt information, which are not available in our case. ~~The first IPF estimation method aims at correcting individual events. For calibration, all events are identified that contain a monthly maximum instantaneous peak. For these events the daily and instantaneous peaks, as well as the daily p/Vs are computed. Then a linear regression model of the following form is fitted~~

It is assumed that most events, especially the larger ones relevant for FFA, are separated correctly.

3.2.2 Estimation of instantaneous peak flows

In this study we propose linear models to estimate IPF from the MDF data, where the peak-volume ratio (as described in Eqn. 3) is one of the main predictors. Additionally, other predictors that describe the catchment physiology or climate (referred for simplicity as

$$\mathrm{IPF_{event}} = \frac{\mathrm{MDF_{event}}}{(a + b_1 * \mathrm{p/V_{event}} + b_2 * \mathrm{CD_1} + \cdots + b_{n+1} * \mathrm{CD}_n)}, \qquad (~~3~~)$$

where *CD* denotes additional catchment descriptors that may be included in the models.) are integrated and investigated. The combination of hydrograph shape and catchment characteristicsdescriptors as predictors is expected to better reproduce both the at-site and between-site variability in the IPF-MDF relationship and yield a more universal model. Several catchment descriptors describing land use, soil type, average climate variables, geographic information and catchment morphology were investigated prior to the study. Two main descriptors, namely basin area and gauge elevation, were found to be more important for the linear model and hence are included in the study shown here.

The describedSince the p/V ratio is calculated for each event, the first p/V model investigated aims at correcting individual events from MDF series. All events that contain a maximum instantaneous monthly peak are identified. For these events the daily MDF_{event} and instantaneous peaks IPF_{event}, as well as the p/V_{event} are computed. Then a linear model of the following form is fitted:

$$\mathrm{IPF_{event}} = \frac{\mathrm{MDF_{event}}}{(a+b_1*\mathrm{p/V_{event}}+ b_2*\mathrm{CD_1}+\cdots+ b_{n+1}*\mathrm{CD}_n)} \quad \text{(4)}$$

correction method will

where *CD* denotes additional catchment and climate descriptors that may be compared withincluded in the models, a and b are the parameters of the linear model fitted by the calibration procedure. The fitting of the linear model parameters is performed on the calibration set (as indicated in section 2) only for the period 1972-2012.

To assess the performance of the new methodology, we employ here also the slope correction -method developed by Chen et al. (2017). This) as a reference. The slope-method estimates an instantaneous event peak flow IPF_{event} based on the slopes of the daily peak Q_{peak} to its preceding Q_{pre} and following daily flows Q_{pre} and Q_{suc}. The IPF is thus estimated as as shown in Eqn. 5:

$$\mathrm{IPF_{event}} = \mathrm{Q_{peak}} + \frac{(\mathrm{Q_{peak}}-\mathrm{Q_{pre}})*(\mathrm{Q_{peak}}-\mathrm{Q_{suc}})}{2*\mathrm{Q_{peak}}-\mathrm{Q_{pre}}-\mathrm{Q_{suc}}} \quad \text{(5)}$$

(4,

For validation, Both event correctioncorrecting methods areneed information from the MDF hydrograph selected for each extracted flood event. Hence these methods can be applied in two ways: 1) IPFs are estimated for all separated events in the daily flow series, even if these events have small daily peaks. 2) IPFs are estimated for the annual maximum daily peak only.Then the flood frequency analysis is performed on the estimated event-based IPFs (after selecting maximum events for each year or season). 2) IPFs are estimated for the annual maximum daily peak only. This means that the event hydrograph corresponding to the annual or seasonal daily maximum is considered for the calculation of p/V in Eqn. 4 or the peak discharges in Eqn. 5. The obtained annual/seasonal maximum series is then used as a basis for the flood frequency analysis. In both cases statistics are derived from the estimated IPF series and compared to the observed IPF statisticesones. Procedure 1) is theoretically more accurate, since maxima in IPF and MDF do not necessarily occur at the same time (no temporal overlap.).

More precisely, events with maximum instantaneous peaks can have rather small mean daily peaks in some instances. Correcting only the maximum MDFs would lead to underestimation of the IPFs in these cases. ~~Procedure 2)~~ On the other hand, procedure 2) may prove more robust in cases where smaller events are not properly separated, i.e. their volumes are over- or underestimated. These events would lead to unrealistic IPF estimates, when using the *p/V* ~~correction method.~~as a primary predictor. The larger events containing the annual maximum MDF are expected to be more properly separated by the algorithm described ~~below~~above.

~~In order~~Alternatively, to the event-based estimation, the proposed p/V model can be ~~able~~ applied also directly to ~~estimate~~the MDF derived statistics with the aim to reproduce the IPF statistics ~~directly from daily records, a second type of IPF estimation methods is analyzed.~~. These involve the estimation of flood statistics, i.e. mean annual and seasonal maximum flows~~,~~ (MHQ), sample L-moments, estimated distribution parameters and derived flood quantiles based on averaged peak-volume ratios *p/*~~Vs~~V_{mean}. These average *p/V*s are obtained from all annual/seasonal maximum MDF events at each ~~station~~site. As described before, these maximum events are expected to be properly separated and although the maximum MDF events may not necessarily be identical to the maximum IPF events, their shape may hold important information on local processes. The model set up is analogous to the event correction approach:

~~$\mathrm{IPF}_{\mathrm{stat}} = \mathrm{MDF}_{\mathrm{stat}} * (a + b_1 * \mathrm{p/V}_{\mathrm{mean}} + b_2 * \mathrm{CD}_1 + \cdots + b_{n+1} * \mathrm{CD}_n)$.~~ ~~(5)~~

$$\mathrm{IPF}_{\mathrm{stat}} = \frac{\mathrm{MDF}_{\mathrm{stat}}}{(a+b_1*\mathrm{p/V}_{\mathrm{mean}}+ b_2*\mathrm{CD}_1+\cdots+ b_{n+1}*\mathrm{CD}_n)}, \quad (6)$$

where *stat* is the desired statistic being estimated, *CD* the selected catchment or climate descriptors, *a* and *b* the parameters of the model as fitted on the calibration set, p/V_{mean} is the average p/V_{event} for annual or seasonal series. The model is expected to represent the average conditions that determine the average deviation of MDF from IPF estimates. The p/V_{mean} in itself is expected to be a good predictor that reflects local conditions like spatial scale, climate, geology and other external factors that control flow variability obtainable from daily flow records. The additional inclusion of catchment descriptors is tested case by case and may contribute to the reproduction of spatial variability~~.~~ of the target variable.

An overview of all the methods employed here together with their description is given in Table 1. All methods consisting of the linear models based on the *p/V* ratio as a main predictor (p/V-method) have been optimized based on the calibration set only for the period 1972-2012. For the selection of the best model, the coefficient of determination (R^2) and the significance of model parameters (based on the p-value) are considered. For validation all sites with their respective observed period are used. Through the validation we compare and assess the ability of the proposed models to capture the mean maximum flow (MHQ) and the probability distribution and the respective design floods.

Table 1: Description of all the methods and their applications employed here for the estimation of the IPF and their statistics.

Application	*Name*	*Description*

Reference	MDF	IPFs are taken directly without correction from average daily flows MDF
Event-based analysis	Slope-events	Estimate IPF for all flood events from MDF according to Eqn. 5
	LM-events	Estimate IPF for all flood events from MDF according to Eqn. 4
AMS-based analysis	Slope-AMS	Estimate IPF as per Eqn. 5 only for events that corresponds to MDF annual/seasonal maxima
	p/V-AMS	Estimate IPF as per Eqn. 4 only for events that corresponds to MDF annual/seasonal maxima
Statistics-based analysis	p/V-Lmoms	Estimate IPF L-moments as per Eqn. 6 based on the MDF L-moments derived from the annual/seasonal maximum series
	p/V-params	Estimate the IPF GEV parameters as per Eqn. 6 based on the MEF GEV parameters derived from annual/seasonal maximum series
	p/V-quants	Estimate IPF quantiles as per Eqn. 6 based on MDF quantiles derived from annual/seasonal maximum series
	p/V-MHQ	IPF mean maxima (MHQ) estimated as per Eqn. 6 based on MHQ extracted from MDF annual/seasonal maximum series

3.2.3 Analysis of the instantaneous peak flows

Since the IPF data are not continuous rather a maximum for each month (see Section 2), a direct comparison for each flood event is not possible. Instead, we focus on the analysis of flood statistics. ~~**2.2 Event separation**~~

~~For separation of the flood events, the initial steps of the procedure used by Tarasova et al. (2018) are carried out, which has proven effective and convenient for their German dataset. For the initial step of baseflow separation they selected the simple nonparametric algorithm by the Institute of Hydrology (1980), which was able to identify the starting points of events in daily flow series in a wide range of catchments. The same method is applied to the series of mean daily flows in our study, which involves the following steps. At First, 5-day non-overlapping blocks are used to find minima, which are identified as turning points if they are more than 1.1 times smaller than their neighboring minima. The baseflow is then derived by simple linear interpolation between the turning points. Discharge peaks are subsequently determined from the flow series and for every peak the start and end of the belonging flow event is defined by the nearest surrounding turning points. In order To prevent false identification of events due to natural variability, events are discarded if their peak discharge is not at least 10% larger than the baseflow. Tarasova et al. (2018) suggest a second step of re-defining events with multiple peaks in an iterative procedure. This step is not carried out here, as it requires rainfall and snowmelt information, which are not available in our case. It is assumed that the majority of events, especially the larger ones relevant for FFA, are separated correctly.~~

~~**2.3 Distribution fitting and flood quantile estimation**~~

~~For extrapolation of the time series and estimation of floods with specific return periods, distributions were fitted to the annual and seasonal samples of both IPF and MDF. This enables the direct comparison of both the higher flood quantiles and of the~~

~~estimated distribution parameters. Here, the General Extreme Value distribution (GEV) of the following form was used for all samples~~

$$~~F(x) = e^{-\exp\left(\frac{1}{k} * \log\left(1-k*\frac{x-\xi}{\alpha}\right)\right)}~~ \quad (6)$$

~~with location parameter ξ, scale parameter α and shape parameter k. The parameters were estimated using sample L- moments. The goodness of fit of the distributions was determined with the Cramer-von-Mises test.~~

~~Additionally, for seasonal considerations, mixed models were applied, which combine two or more GEV distributions fitted to different subsamples of the data, like summer and winter floods. A simple maximum mixing approach is used to combine the individual distributions:~~

$$~~F_{mix}(x) = \prod_{i=1}^{n} F_i(x).~~ \quad (7)$$

~~This approach allows the combined estimation of flood quantiles from multiple underlying distributions and thus the assessment of errors in seasonal FFA. The approach is described in detail in Fischer et al. 2016. In their study they used thresholds in order to determine whether a seasonal maximum is actually a flood event, which may not be the case during dry summers. This threshold was defined as the minimum annual maximum flow. We do not censor our data with thresholds, i.e. for matters of simplicity we assume that every seasonal maximum is indeed a flood event.~~

For this purpose, the percentage deviation of the MDF statistic MDF_{stat} from the IPF statistic IPF_{stat} is calculated as following:

$$Error\ (\%) = \frac{\mathrm{MDF_{stat}} - \mathrm{IPF_{stat}}}{\mathrm{IPF_{stat}}} * 100\ \% \quad ~~2.4~~ (7)$$

where *Error* is computed at each site for any desired statistical quantity *stat*, like the mean annual maximum flow (MHQ), L-moments (Hosking, 1990), distribution parameters and flood quantiles.

Apart from the *Error (%)* at each site, two additional performance criteria are calculated over all sites: the normalized root mean square error *nRMSE (%)* as per Eqn. 8 and the percent *pBIAS (%)* as per Eqn. 9.

$$nRMSE\ (\%) = \frac{\sqrt{\frac{\sum_{i=1}^{N}\left(\mathrm{MDF_{stat,i}} - \mathrm{IPF_{stat,i}}\right)^2}{N}}}{\mathrm{sd(IPF_{stat})}} * 100\ \%, \quad (8)$$

$$pBIAS\ (\%) = \frac{\sum_{i=1}^{N} \mathrm{MDF_{stat,i}} - \mathrm{IPF_{stat,i}}}{\sum_{i=1}^{N} \mathrm{IPF_{stat,i}}} * 100\ \%, \quad (9)$$

where N is the number of the validation sites, $MDF_{stat,i}$ and $IPF_{stat,i}$ are the respective statistics at site i from MDF and IPF series, and $sd(IPF_{stat})$ is the standard deviation of IPF statistics over all considered sites. These criteria are computed for each of the methods described in Table 1.

3.3 Uncertainty Analysis

Since both distribution fitting and IPF estimation via ~~linear~~p/V models are approximations and not fully accurate, we eventually assess the overall level of uncertainty in the final IPF flood quantile estimates. As it will be later shown in Section 4.2 the best linear model is chosen to be the p/V-Lmoms - the model correcting directly the L-moments of the MDF series. This is done by using simple ~~bootstrapping~~resampling with replacement procedures~~.~~; resampling in time when selecting the maximum series for FFA, resampling in space when selecting the sites for the p/V model (either calibration or validation of the models) and resampling both in space and time. In a first step, the series of annual ~~maxima~~/seasonal maximum from both ~~daily~~MDF and ~~monthly maximum data~~IPF dataset are analogously resampled 1000 times with replacement~~.~~ (temporal sample and parameter uncertainty). For each resampling the desired flood quantiles are estimated using L-moments. The range of these estimates provides the baseline level of uncertainty due to ~~distribution fitting.~~ sample and parameter uncertainty. The temporal uncertainty is calculated at each site for the original MDF and IPF series (respectively MDF-bs and IPF-bs) and are considered a benchmark for comparison.

In a second step, ~~linear regression~~ p/V models are fitted to each pairing of ~~estimated IPF and~~ temporally resampled MDF ~~flood quantiles over~~and IPF series while considering all ~~stations~~sites in the study area~~. In order~~ that have more than 30 years of observations. This means that the temporal sample uncertainty is propagated through the p/V model (p/V-full). To assess the uncertainty of the ~~fitted models~~selected p/V model, another resampling is carried out, this time shuffling the set of considered ~~stations,~~sites, where original MDF site specific L-moments are resampled again 1000 times with replacement~~.~~ before fitting the p/V model (p/V-bs). Lastly the total uncertainty both in space and time is assessed by combining the temporal sample and parameter uncertainty with the uncertainty of the fitted models. This means that the maximum series are resampled 1000 times, and for each ~~station this procedure yields 1000 estimates of paired flood quantiles from both the IPF and MDF series (IPF-bs and MDF-bs),~~of these sets, the sites are resampled 1000 ~~full model quantile estimates resulting from the original p/V model fitted to each permutation (p/V-full), and~~ times as well before fitting the p/V model. So, the total uncertainty will be derived by 1000~~*~~ x 1000 quantile estimates ~~resulting from permutation of the p/V model for all IPF and MDF transpositions (p/V~~(p/V-bs-bs~~-bs).~~).

~~In order~~ To assess the overall level of uncertainty, several indices ~~will be assessed at the individual stations.~~are computed at each site. The first one is the relative width of the 95% confidence intervals (~~CI~~)*CI95%*) calculated for all aforementioned ~~bootstrap sample~~ resampling estimates of the desired flood quantile:

~~$CI_{bs} =$~~ $CI95\%_{bs}(-) = \frac{x_{bs;0.975} - x_{bs;0.025}}{x_{bs;0.5}}$,

~~(8~~(10)

where $x_{bs;0.025}$ and $x_{bs;0.975}$ are the 2.5% and 97.5% quantile and $x_{bs;0.5}$ the median of the respective sample. ~~The second one is the deviation of the individual MDF and p/V model bootstrap samples from the IPF sample, which allows the assessment of error distributions~~

The second one is the deviation of the IPF estimated samples from the IPF original sample, which allows the assessment of error distributions:

$$error_{bs} = \frac{x_{bs} - IPF_{bs}}{IPF_{bs}} * 100\%. \quad (~~9~~11)$$

where IPF_{bs} is the temporal resample of the IPF original data and x_{bs} is the resampled estimated either from the original MDF series or from the modeled IPF series. From the resulting error vector, a variety of statistics can be computed for comparison. Finally, the agreement of the 95% confidence intervals of the MDF and ~~LM~~p/V-model samples with the IPF confidence bands are determined as percentage overlap at each ~~gauge~~site:

$$overlap = \frac{\min(x_{bs;0.975}, IPF_{bs;0.975}) - \max(x_{bs;0.025}, IPF_{bs;0.025})}{\max(x_{bs;0.975}, IPF_{bs;0.975}) - \min(x_{bs;0.025}, IPF_{bs;0.025})} * 100\%. \quad (~~10~~12)$$

~~3 Study area and data~~

~~This study uses data from 653 discharge gauges distributed over Germany. For the analyses, average daily flow and maximum monthly flow are required. The selected stations represent the datasets of the federal agencies, who provide online access to both parameters (Lower Saxony, Saxony-Anhalt, Saxony, Bavaria and Baden-Württemberg; see section Data Availability).~~

where IPF_{bs} is the temporal resample of the IPF original data and x_{bs} is the resampled estimated either from the original MDF series or from the modelled IPF series.

~~Germany poses a transition zone from an oceanic climate in the northwest to a humid continental climate in the southeast. The northwestern parts are influenced by wet air and have mild winters, while the more southeastern parts are drier and exhibit larger temperature ranges. The average temperature for the entire country is 8.9 °C, the monthly averages ranging between 0.4°C in January and 18°C in July (reference period 1981-2010; DWD). The average annual precipitation is 819 mm, where amounts generally decrease in west-east direction and in strong dependence on topography. Annual rainfall sums are generally highest over the Alps at the very Southern border and the various secondary mountain ranges. The flat continental east is driest. Temporally, the summer months are wettest with rainfall often occurring in convective events. Snowfall occurs between October and April, where amount and depth of snow cover increase with decreasing oceanic influence and increasing altitude.~~

Even though not the entire area of Germany is covered by the available data, the selected gauges provide a cross section through the climatically and topographically distinct regions, from the flat oceanic northwest to the mountainous continental southeast.

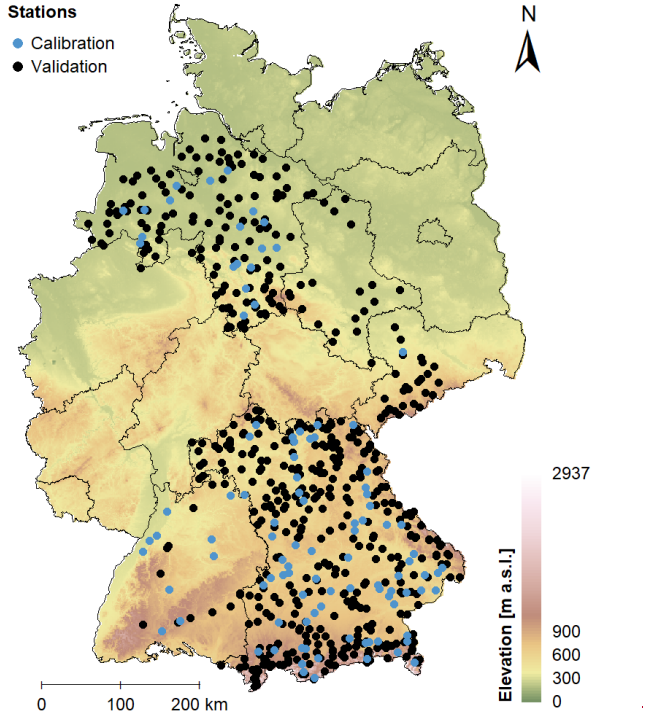
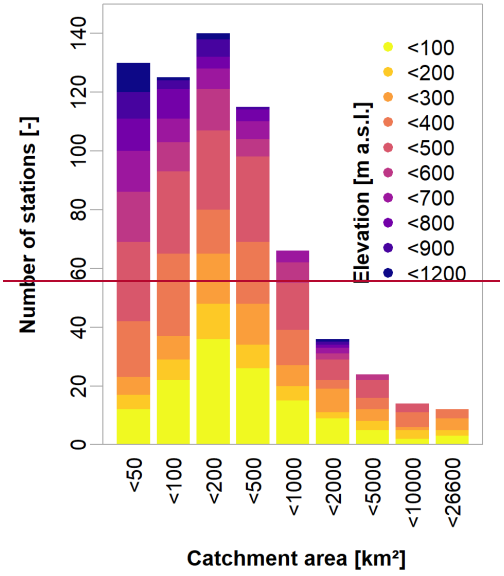


Figure 1: Location of the 653 gauges used for analysis. The 103 stations used for model calibration are marked in blue. Digital elevation data by Jarvis et al. (2008).

The lengths of the discharge records vary substantially from 11 to 183 years with a mean of 48.4 years. For the general assessment of differences in IPF and MDF floods and final model validation, all 648 stations with their variable record lengths are considered. For assessment of flood frequency criteria only those stations with at least 30 years of observations were used (490). Model fitting was carried out on a subset of 103 gauges, whose discharge series were thoroughly checked. Also, their records were cropped to a common period from 1979 to 2012, in order to eliminate potential non-stationary effects.

For the 103 stations used for calibration a catalogue of catchment descriptors is available. For the remaining stations only rudimentary information was obtained, i.e. catchment size, geographical position and altitude of the gauges. Fig. 2 shows the how the 648 discharge stations are distributed in terms of catchment size and gauge elevation.



~~Figure 2: Distribution of catchment size and elevation of the 648 gauges used for analysis.~~

4 Results and discussion

4.1 Mean Maximum Flow (MHQ)

4.1.1 Comparison of ~~MDF~~mean daily and ~~IPF peaks~~instantaneous peak flows

In theory, the relative deviation between MDF and IPF peaks depends greatly on catchment size~~.~~, as shown for instance in Fuller (1914) and Ellis and Gray (1966). Small catchments without appreciable buffering capacity react fast to even small rainfall, leading to short and steep flood waves that are hardly reproduced on coarsely averaged time scales. Factors like steep slopes, impermeable underground and short but intense rainfall contribute to the flashiness of storm events and make these even less representable through daily flow records. The effect of the catchment size is clearly visible in ~~the~~our data set. Fig. 3 demonstrates ~~by means~~the errors of the mean annual maximum flow (MHQ) estimated by MDF instead of IPF series (as per Eqn. 7). It is clear that the larger the area, the smaller the deviation between MDF and IPF. ~~Also, errors~~ This agrees well with the findings of Ellis and Gray (1966) and Chen et al. (2017) which state that for larger basins the peak-ration between MDF-IPF series converges to 1. Hence in these cases the MDFs are a good representation of the IPFs peaks. Moreover, MHQ errors shown in Fig. 3 appear to be especially large in higher altitudes. ~~Generally, the~~ This is as well expected as mountainous catchments have a fast response time and are generally more influenced by the meteorological forcing (by snowmelt processes or convective events) as also shown in Gaal et al. (2015). Overall, the MHQ error in our catchments seems to increase in north-south direction, which could be a secondary effect of both increasing altitude and decreasing catchment size.

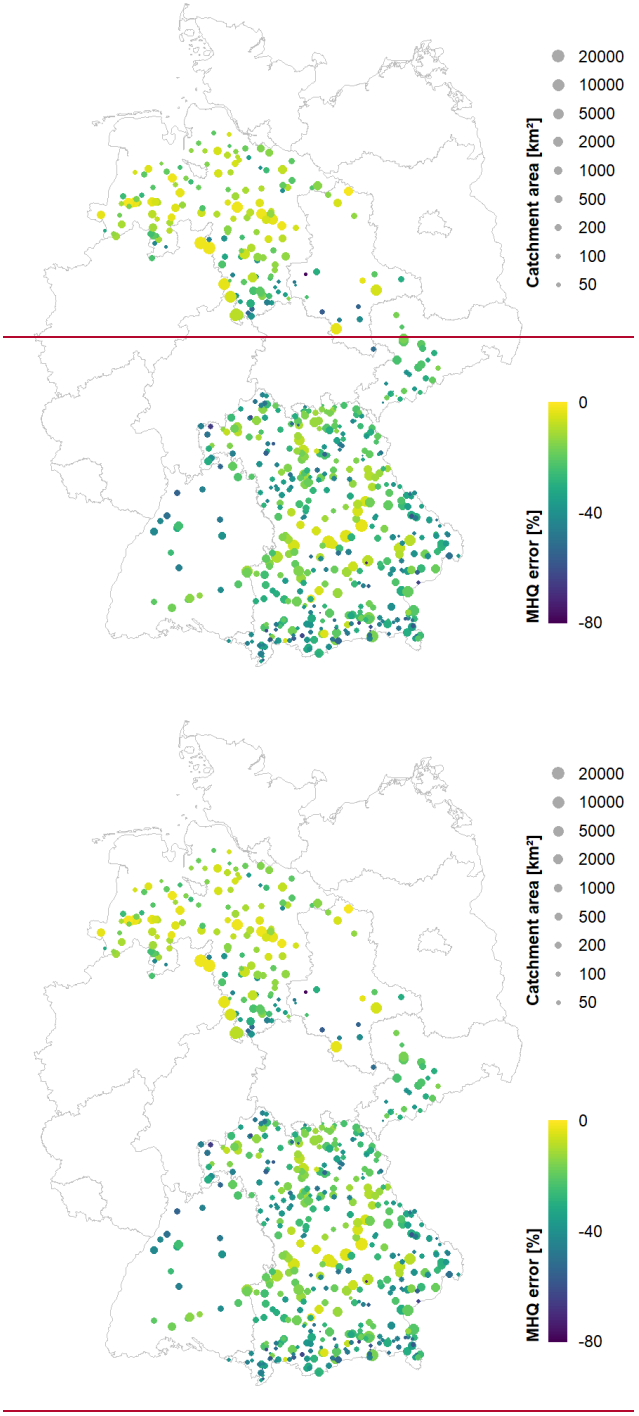


Figure 3: Spatial distribution of the ~~discrepancy~~mean annual maximum flow (MHQ) *error (%)* between mean daily (MDF) and instantaneous peak (IPF ~~in the MHQ.~~) flows obtained from all sites (calculated as per Eqn. 7)

When assessing the differences between ~~average~~mean daily and instantaneous peaks, it is also meaningful to take a closer look at different types of floods. For our German ~~dataset~~sites the two most opposite types are a) flood events induced by short intense rainfall, especially convective events~~,~~ dominant mainly in summer (May-October), and b) extended flood events with
405 significant volume, as caused by snowmelt and/or stratiform rain~~.~~ occurring mainly in winter (November-April). Presumably, the latter flood type is much better represented by ~~average~~mean daily flow than the former. In order to roughly distinguish between the two types, the flow records are divided into summer (May - October) and winter (November - April) half years. Due to the limited data availability, a clear distinction between convective, stratiform and snow-melt events cannot be achieved here. Some snowmelt events in the high alpine catchments may still occur in May/June but are classified as summer events.
410 However, the coarse division of the data into half years rather than seasons is due to the subsequent analysis of seasonal flood statistics and application of the mixed seasonal model.

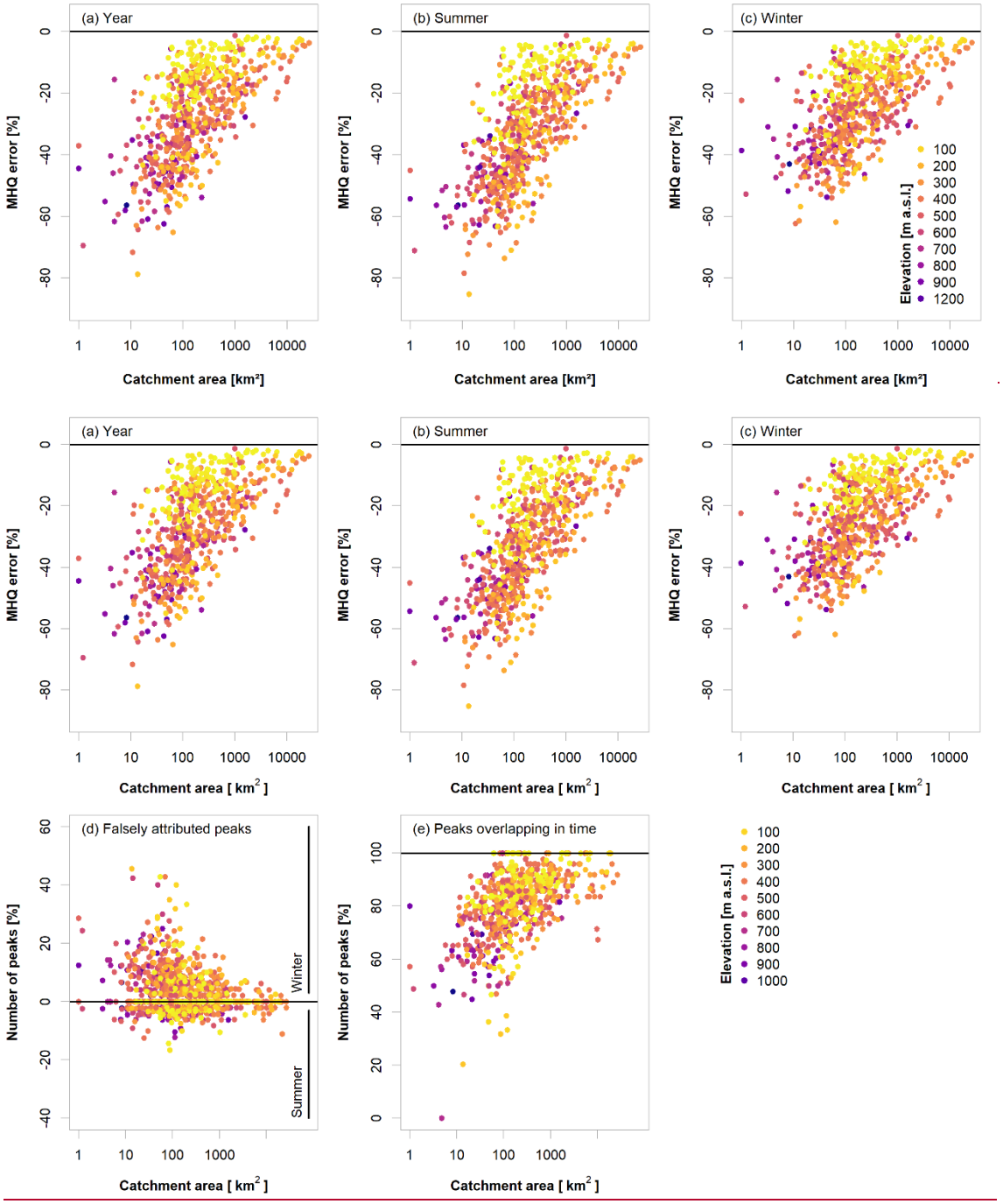
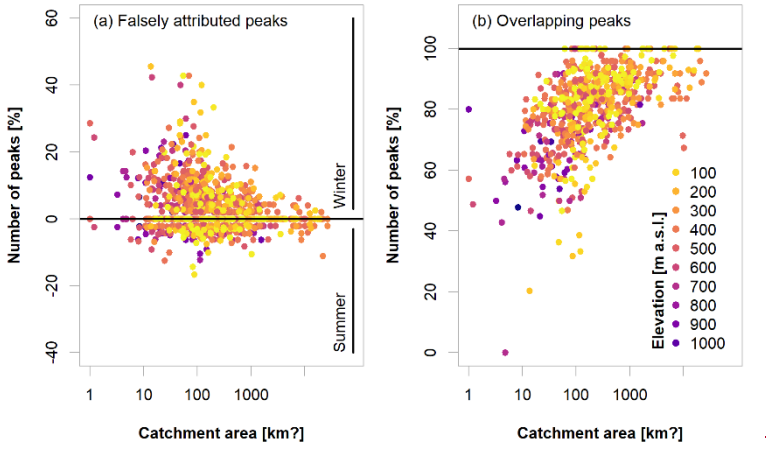


Figure 4: ~~Error~~upper row – *error (%)* in the mean maximum flow (MHQ) (as per Eqn. 7) obtained in relation to catchment size and gauge elevation for the entire year (a), summer (b) and winter (c~~).~~); lower row - percentage of peaks falsely attributed by mean daily flows (MDF) to the winter or summer half-year (d) and percentage of peaks in mean daily (MDF) and instantaneous peak (IPF) flows that overlap in time (within a 5-day buffer) (e). Results are illustrated for all sites.

In the upper row of Fig. 4 ~~the error in~~(a-c) the MHQ error is shown for the entire year and ~~the~~respectively for summer and winter ~~half year~~seasons. The relationship with catchment area is clearly visible in all three cases. Also, the effect of the elevation becomes obvious, namely sites in the lowest elevations (yellow points, below 100 m) ~~showing~~show very small errors, even for small catchment sizes down to approximately 100 km². This is the clearest stratification in the error due to elevation; the errors at higher altitudes appear less distinguishable.

There is, however, a clear distinction between summer and winter seasons. As expected, the MHQ error is overall smaller in the winter months, where snowmelt and stratiform events prevail, while the convective events in summer are poorly reproduced by MDF. The error in the annual peaks is a mixture of the two seasons~~.~~; which season contributes ~~mainly~~more to the annual peaks depends on the individual flood regimes. When looking at the IPF data, at 68.8% of the considered ~~gauges~~sites the winter floods exceed their summer counterparts on average, while 29.2% are dominated by summer floods. ~~These seasonality statistics are established on basis of the IPF.~~ When considering MDF instead, only 22.1% of the gauges are identified as having maximum peaks in summer. This indicates that the ~~average~~mean daily flow smooths significant peaks to a point where they are no longer relevant for the overall flood behavior. Fig. ~~5 a~~4 (d) shows the percentage of annual maxima at each ~~gauge~~site that are attributed to the wrong season when using MDF. Each ~~gauge~~site is represented by two dots: Negative values show the percentage of all annual maxima that are falsely attributed to summer, while positive values show the falsely attributed winter peaks. It is obvious that with decreasing catchment size an increasing number of annual maxima are falsely identified in the winter half year, while the actual instantaneous maxima occur in summer. Apart from not being able to properly identify flood magnitudes when using mean daily ~~flow series~~flows, this is a serious issue for classification of flood regimes, identification of dominating flood types and application of heterogeneous flood frequency analysis when daily data is the only available option.

Another general issue highlighted by this analysis, independent of seasonality, is that the ~~asynchronous occurrence~~peaks of both IPFs and MDFs~~.~~ do not necessarily occur at the same day (there is no temporal overlap). In their study, Chen et al. (2017) illustrated that only on 82% of the events investigated, the peaks of both IPF and MDF series occurred on the same day. This suggest that instantaneous maxima are not always identifiable in the mean daily ~~flow series~~flows, i.e. the maxima obtained from the daily series are inevitably found in other places. The temporal overlap of MDF and IPF derived peaks for our catchments are shown in Figure 4 (e). In general, the smaller the catchment, the smaller the temporal overlap between instantaneous and daily peaks~~, as seen in Fig. 5 b.~~. A weak relationship is also visible between high elevation and smaller temporal overlap between the two peaks. This problem needs to be kept in mind when attempting to estimate instantaneous peaks from daily peaks, since the two may belong to significantly different flood events (different genesis) and thus to different populations.



~~Figure 5: Percentage of peaks falsely attributed by MDF to~~4.1.2 Estimation of mean maximum flow (MHQ)

~~So far,~~ the ~~winter or summer half-year (a) and percentage of peaks~~error in ~~the~~ mean maximum flow (MHQ) between MDF and IPF ~~overlapping in time (~~is shown to be influenced by both catchment area and gauge elevation. Both of these predictors may be helpful to correct MDF for a better agreement with ~~a 5-day buffer; b).~~

~~**4.2 Estimation of mean annual IPF**~~

~~For both correction of the individual events and of~~ the ~~MHQ, linear models appeared appropriate.~~IPF data. Moreover, there seems to be a significant linear dependence between the peak ratios MDF/IPF ~~and,~~ *p/V* and the logarithm of the catchment size. We first test the suitability of various predictors to predict MHQ of IPF by fitting the p/V models to the individual events of MDFs (p/V-event), to the MDF maximum series (p/V-AMS) or lastly directly to the MDF mean maximum flow (p/V-MHQ). Various model combinations with the available ~~variables~~ predictors (catchment area, elevation and *p/V* ratio) are tested using the calibration data set~~. Table 1 lists~~ and the ~~coefficients~~respective coefficient of determination ~~of these model combinations. An asterisk indicates that not all regressors in the model~~ are ~~significant (p = 5%).~~ shown in Table 2. The selected models are marked in grey in Table 2 and ~~the~~their respective full model formulas are given in Table ~~2~~3. For most models, the majority of variance in the IPF-MDF relationship is explained by the *p/V* ratio and the catchment area. For winter, including gauge elevation appeared to improve the model slightly.

The models show a similar performance for the annual and summer peak ratios both for ~~the~~correction of individual events (p/V-event) and ~~the~~ mean maximum flow (p/V-MHQ~~.~~). For winter, the model performance seems to differ, especially ~~for~~when correcting the individual events (p/V-event ~~correction.~~). It appears that the ~~linear~~ models using the *p/V* ratio have more difficulty to estimate the winter peak ratio. This could be due to improper event separation, which will be discussed in more detail below and which leads to unrealistic *p/*~~Vs~~*V* ratios. The fact that elevation is a significant ~~regressor~~predictor in the MHQ model (p/V-MHQ) may also suggest that the peak ratios in winter are more heterogeneous.

Table ~~1:~~2: Coefficients of determination of ~~different target variables and~~ various linear model combinations~~.~~ (see Table 1 for description of models). Values are obtained by fitting the models only to the calibration set. Grey cells indicate the best ~~linear~~p/V model for each ~~target variable.~~ application and asterisks indicate at least one non-significant ~~regressors~~predictors in the ~~linear~~p/V models.

~~Target variable~~ ~~Regressors~~ Application / Predictors	Event~~-peak MDF/IPF-~~based (p/V-event)			~~Annual/seasonal maximum MDF/IPF~~AMS-based (p/V-AMS)			MHQ~~-MDF/IPF-~~based (p/V-MHQ)		
	Year	Summer	Winter	Year	Summer	Winter	Year	Summer	Winter
Area	0.14	0.19	0.12	0.30	0.26	0.25	0.42	0.42	0.38
Elevation	0.01	0.01	0.01	0.04	0.01	0.03	0.06	0.02	0.08
p/~~V~~$_{mean}$V	0.13	0.13	0.09	0.21	0.21	0.20	0.55	0.49	0.49
p/~~V~~$_{mean}$V + Area	0.23	0.26	0.17	0.39	0.36	0.35	0.66	0.65	0.63
p/~~V~~$_{mean}$V + Elevation	0.14	0.13	0.10	0.23	0.22	0.23	0.56*	0.51	0.56
p/~~V~~$_{mean}$V + Area + Elevation	0.23	0.26	0.17	0.40	0.36*	0.36	0.67*	0.65*	0.68

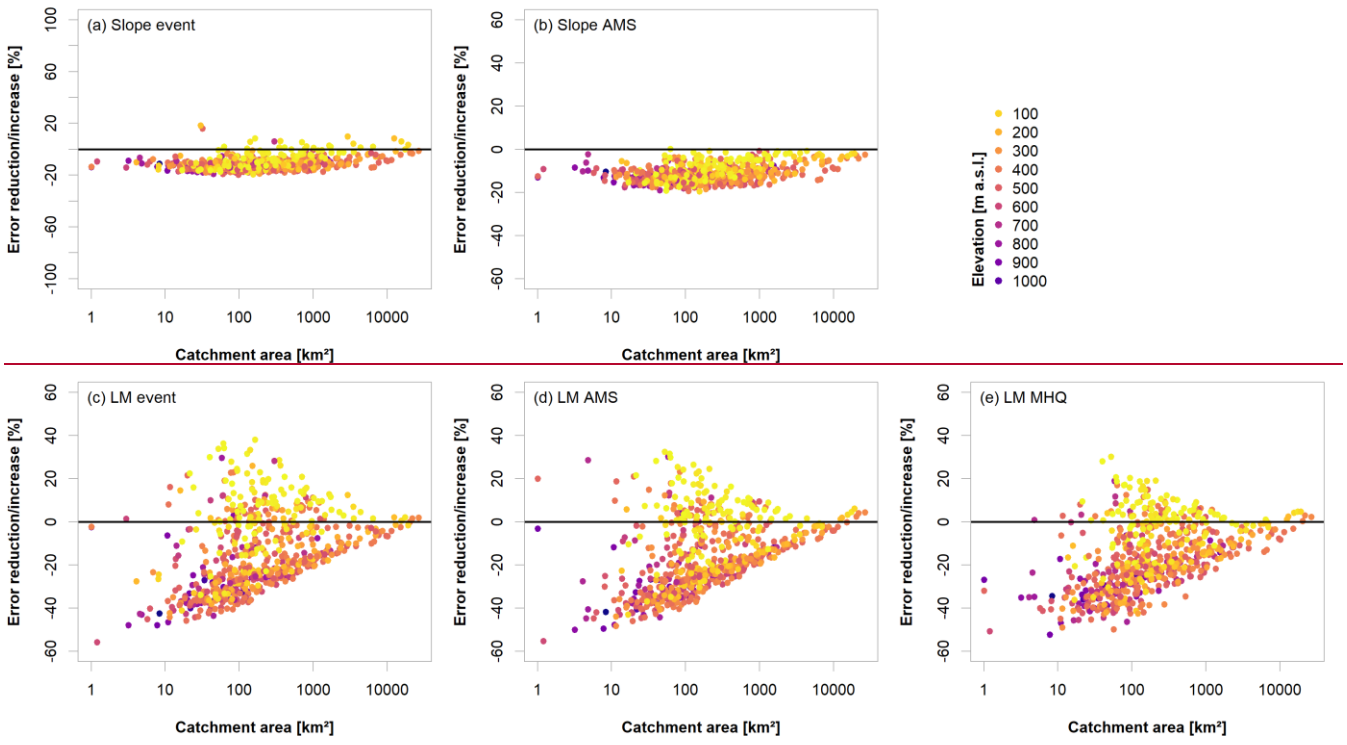
Table ~~2: Linear~~3: Best p/V models (as shown with grey in Table 2) fitted on the calibration set for correction of individual events~~,~~ (p/V-events), for the annual/seasonal ~~maxima~~maximum (LM-AMS) and the MHQ~~.~~ (LM-MHQ).

Type		Model
Events (p/V-events)	Year	MDF / (0.59 - 0.43 * p/~~V~~$_{mean}$ V$_{event}$ + 0.047 * log(area))
	Summer	MDF / (0.44 - 0.36 * p/~~V~~$_{mean}$V$_{event}$ + 0.063 * log(area))
	Winter	MDF / (0.63 - 0.35 * p/~~V~~$_{mean}$V$_{event}$ + 0.044 * log(area))
Maxima (p/V-AMS)	Year	MAX$_{MDF}$ / (0.53 - 0.42 * p/~~V~~$_{mean}$V$_{max}$ + 0.057 * log(area))
	Summer	MAX$_{MDF}$ / (0.61 - 0.73 * p/~~V~~$_{mean}$V$_{max}$ + 0.061 * log(area))
	Winter	MAX$_{MDF}$ / (0.70 - 0.68 * p/~~V~~$_{mean}$V$_{max}$ + 0.054 * log(area))
MHQ (p/V-MHQ)	Year	MHQ$_{MDF}$ / (0.74 - 0.94 * p/V$_{mean}$ + 0.043 * log(area))
	Summer	MHQ$_{MDF}$ / (0.83 - 1.19 * p/V$_{mean}$ - 0.054 * log(area))
	Winter	MHQ$_{MDF}$ / (0.99 - 1.31 * p/V$_{mean}$ + 0.035 * log(area) - 0.00012 * elevation)

Fig. ~~6~~5 shows the change in mean absolute error in the annual MHQ after correction with the different methods in relation to catchment size and elevation~~.~~: positive values indicate that the error has increased after correction, while negative values indicate that the error has decreased after correction. The slope method (Fig. 5 - a) applied to the individual events (slope-event) yields a rather constant reduction of the error independent of catchment size. However, there are several outliers produced by ~~the~~this method, which can be attributed to improper separation of smaller events. Applying the slope method only to the annual maximum MDF events~~,~~ (slope-AMS), as done in Fig. 5 (b), shows a much smoother and more constant error reduction. The ~~methods~~corrections using the ~~linear model~~p/V models proposed here (Fig. 5 c-e) yield a much larger improvement for the smaller catchments (where the original MDF error ~~is~~was generally larger than in the bigger catchments~~.~~ ~~However~~). Nevertheless, these ~~methods~~corrections simultaneously lead to an increase of the error in several cases. This deterioration appears to affect those ~~stations~~sites that have been highlighted before in section 4.1.1, namely the ones with the lowest elevations in the data set where the original MDF error was quite low.

The differences between correcting the individual events (p/V-events in Fig. 5 c) and only correcting the annual maxima (events (p/V-AMS in Fig. 5 d) using by means of the linear models model appear rather small in this case.. This suggests that even though the annual maximum from MDF does not occur at the same time as the annual maximum IPF, the method still yields an appropriate estimate of the true IPF. The method of On the other hand, directly correcting the MHQ (p/V-MHQ in Fig. 5 e) results in slightly lower error reduction for the smaller catchments but also appears to produce fewer outliers and is thus considered more robust.

It should be noted that working with large data and automatic event separation without manual post-correction leads to problems that could potentially be avoided when considering individual time series more carefully. Several events are identified as too long or too short (or not at all), so their volumes are over- or understated, respectively. This results in false *p/Vs V* ratios and in some cases to severe over- or underestimation of the peak peaks. The weight of such events is assumed to be significantly lower when correcting flood statistics based on average *p/Vs V* ratios. In addition, the overall performance can only be assessed for events that contain the monthly instantaneous maximum flow, i.e. primarily larger events. How the event correction performs for minor events cannot be analyzed here.



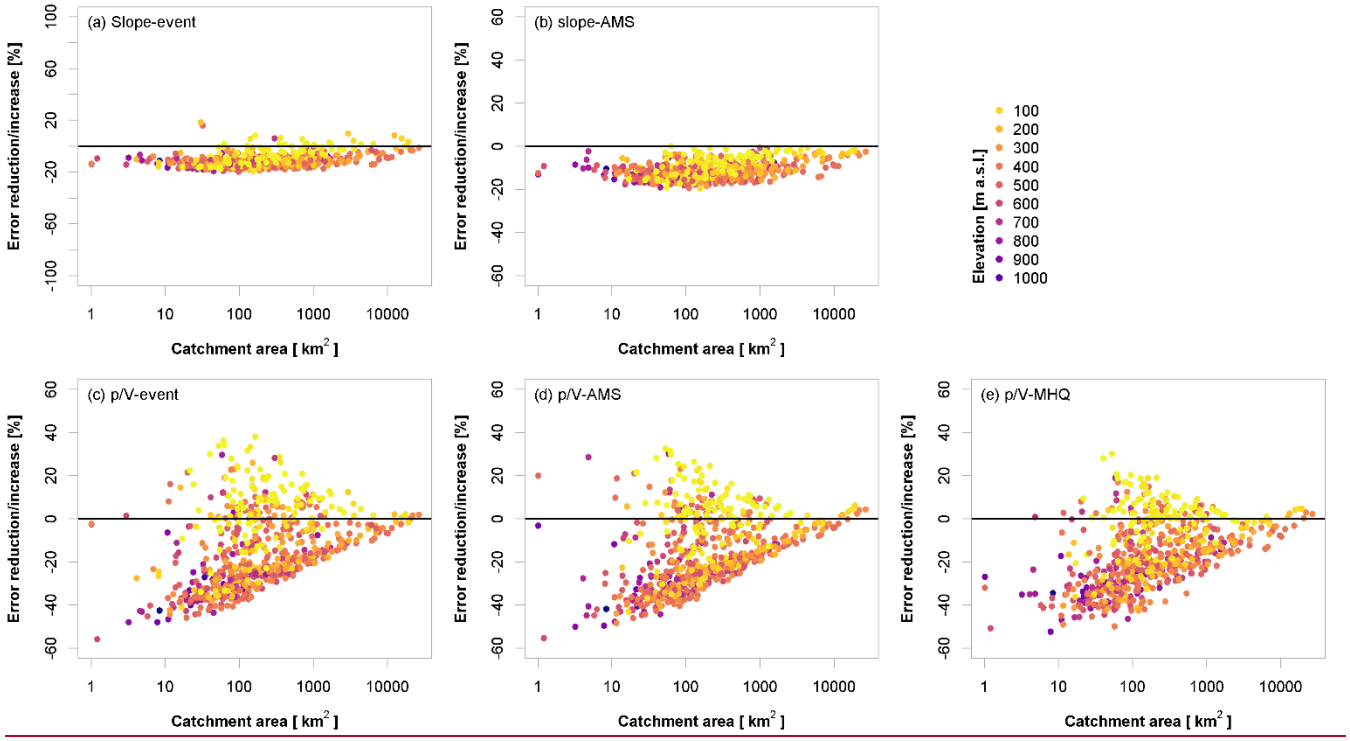


Figure ~~6.~~5: Error reduction (negative values) ~~/~~vs error increase (positive values) in the mean annual maximum flow (MHQ) for different IPF estimation methods when compared to MDF. For an overview of the methods, the reader is directed to Table 1. Values are obtained by applying the selected methods to the validation set.

Fig. ~~7~~6 summarizes the overall model performances for the mean annual/seasonal maximum flow (MHQ) at all ~~648 stations~~validation sites and compares the individual methods to the error in using MDF directly. It is obvious that all methods give significantly better IPF estimates than the mere MDFs. The slope correction ~~has~~methods (both slope-event and slope-AMS) have quite a large bias~~,~~ (median error around -10%), which is, as seen above, not only disadvantageous. Still, the overall error is smaller for the ~~LM methods,~~p/V models (p/V-events, p/V-AMS and p/V-MHQ), where the median error is at 0-2%, with fewer positive outliers produced by the ~~LM~~p/V-MHQ approach.

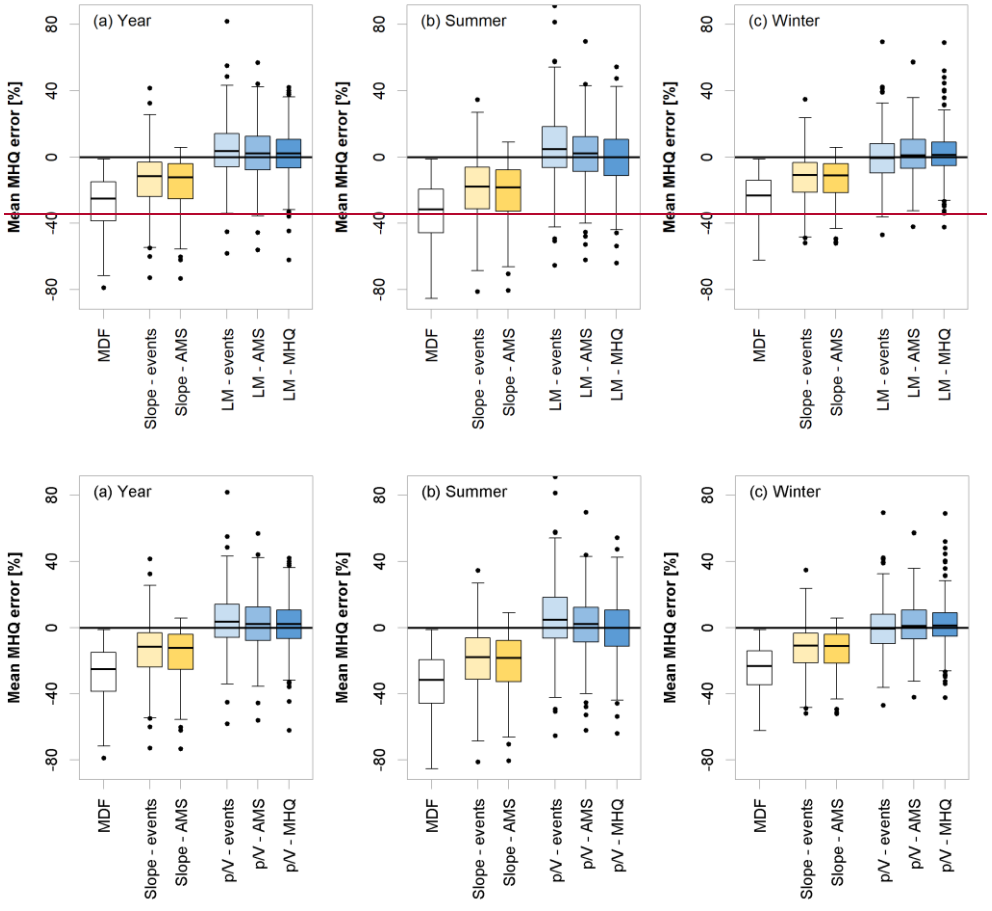


Figure ~~7:~~6: *Error (%)* comparison of ~~performances of~~ different ~~IPF estimation~~ methods to estimate the mean maximum annual flow (MHQ) for the entire year (a), summer (b) and winter(c). Values are obtained by applying the selected methods to the validation set. For an overview of the methods, the reader is directed to Table1.

Table ~~3~~4 summarizes the normalized root mean square error (*nRMSE~~)~~ (%)*) and the percentage bias (*pBIAS~~)~~ (%)*) of the ~~instantaneous MHQ~~mean annual/seasonal maximum flow (MHQ) estimated via the different model variants. In terms of *nRMSE*, the performances of the slope and ~~LM~~ p/V-methods are comparable, ~~while~~with the slope methods ~~are~~being more biased. There are a number of outliers produced by the ~~LM~~ p/V-methods, especially positive ones, that affect the overall *nRMSE*. As seen in Fig. ~~6~~5, this is primarily concerning the low elevation catchments below 100 m. The values in parentheses in Table ~~3~~4 indicate the performance criteria for gauges with catchment areas below 500 km². Here, the advantage of the ~~LM~~p/V-approaches over the slope ~~method~~methods become apparent, even though a large number of low elevation catchments fall in this category, which negatively affect the overall error.

Table ~~3: NRMSE~~4: Normalized root mean square error (nRMSE in %) and percentage bias (pBIAS in %) of estimated vs. observed instantaneous annual/seasonal mean maximum flow (MHQ) over all sites for different ~~model varaints~~methods. The values in parentheses show the performances for catchment sizes below 500 km².

	Year		Summer		Winter	
	nRMSE [%]	pBIAS [%]	nRMSE [%]	pBIAS [%]	nRMSE [%]	pBIAS [%]
MDF	17.0 *(47.9)*	-18.0 *(-32.4)*	18.1 *(49.0)*	-20.6 *(-38.1)*	14.9 *(44.1)*	-16.4 *(-28.7)*
Slope-events	8.4 *(25.0)*	-6.8 *(-15.8)*	7.9 *(29.0)*	-9.2 *(-21.8)*	9.2 *(21.3)*	-6.5 *(-13.0)*
Slope-AMS	7.4 *(31.2)*	-8.1 *(-19.3)*	8.4 *(33.6)*	-10.5 *(-25.0)*	7.2 *(28.0)*	-7.4 *(-16.1)*
~~LM~~p/V-events	9.3 *(16.7)*	-2.8 *(-1.0)*	8.4 *(16.6)*	-2.6 *(-0.5)*	10.6 *(17.6)*	-5.1 *(-4.1)*
~~LM~~p/V-AMS	10.7 *(20.4)*	-5.4 *(-2.9)*	11.0 *(19.7)*	-5.4 *(-3.7)*	9.4 *(21.4)*	-4.7 *(-2.9)*
~~LM~~p/V-MHQ	7.7 *(19.0)*	-3.9 *(-2.3)*	12.5 *(20.6)*	-6.8 *(-5.0)*	8.5 *(20.8)*	-3.8 *(-1.7)*

Table ~~4~~5 shows the average ~~deviation of~~error between the annual MHQ predicted with the ~~LM~~p/V-MHQ model ~~from~~with the observed instantaneous annual MHQ, distributed for different catchment sizes and elevations. It becomes obvious that for the smallest elevations, the instantaneous annual MHQ is overestimated, especially for smaller catchment sizes. Catchments in the range between 100 and 200 m of altitude also show quite large errors but these are mostly negative. It is also apparent that the catchments with outlets at higher elevations exhibit large negative errors in most cases.

Table ~~4:~~5: Average prediction *error* (%) of the ~~LM~~p/V-MHQ model for the annual MHQ ~~in percent~~calculated over the validation sites and shown here for ~~the~~ different ranges of area and elevation. Red shades indicate overestimation, while blue shades underestimation.

Catchment area [km²]	Elevation [m a.s.l.]									
	<100	<200	<300	<400	<500	<600	<700	<800	<900	<1200
<50	9.35	-10.77	-8.86	-6.77	1.17	-6.35	0.64	-3.14	-0.31	-15.33
<100	18.59	-16.98	-9.75	1.97	4.21	0.37	1.83	-1.05	-2.74	-19.45
<200	13.79	-10.41	1.04	1.45	8.70	-3.61	3.27	5.48	-9.16	4.96
<500	11.43	-6.57	-0.01	4.19	5.58	3.62	-4.05	-3.93	-21.83	-
<1000	8.12	-5.31	3.80	0.86	3.22	-4.66	3.53	-	-	-
<2000	5.05	-1.94	-4.45	5.29	-1.66	-3.97	-3.68	-5.91	6.42	-13.63
<5000	-0.61	-6.75	-1.96	-1.15	-2.73	-3.53	-	-	-	-
<10000	-3.47	-6.38	-0.70	-4.96	-6.50	-	-	-	-	-
<30000	-7.48	-8.37	-5.25	-5.84	-	-	-	-	-	-

4.~~3 Comparison of IPF and MDF~~2 Probability distributions and derived design flows

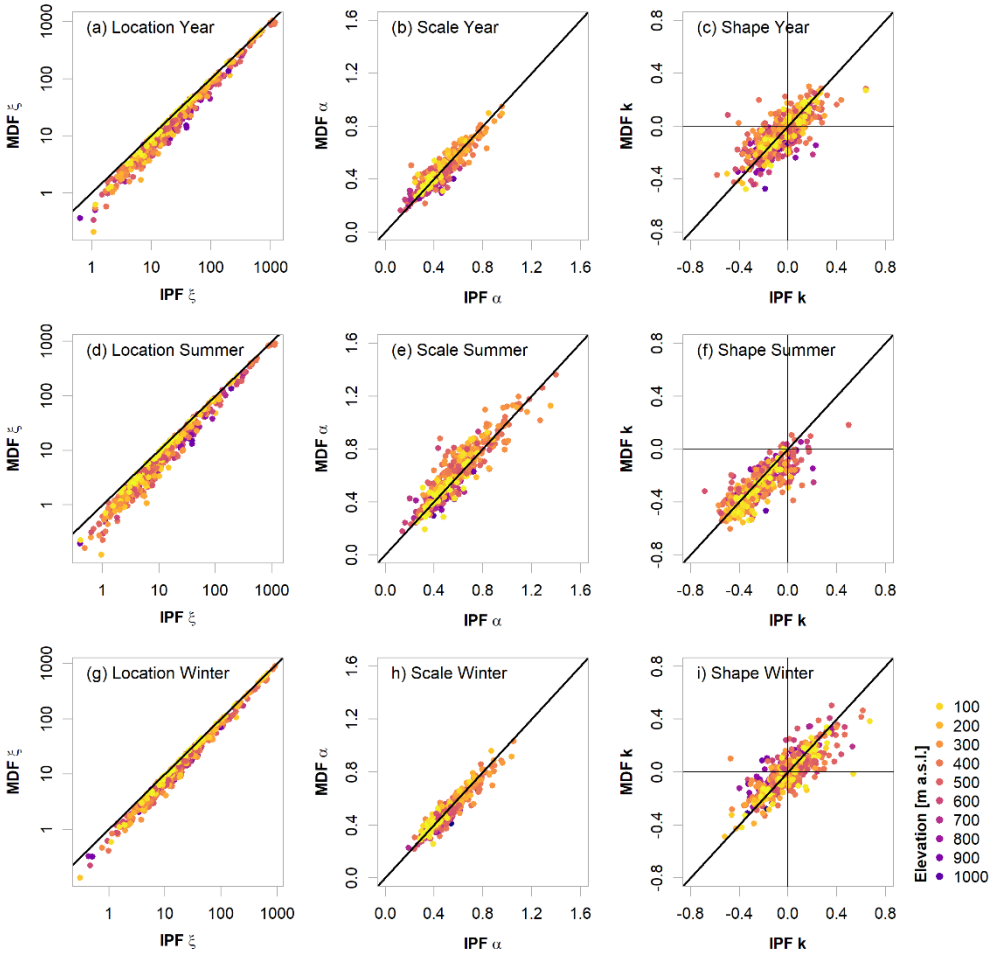
So far the proposed p/V models were analyzed in their ability to estimate better the mean maximum flow (MHQ) from MDF data. In this subsection the focus is shifted to the ability of the methods to estimate parameter distribution of the IPF and the derived flood quantiles. The GEV distribution appeared to be a generally suitable distribution for the ~~stations~~gauges in the

dataset. A Cramer-von-Mises test was carried out for the original IPF and MDF samples, as well as for the slope and ~~LM~~p/V corrected samples at each ~~station~~site and certified a good fit in all cases (p-value = 5%).

4.2.1 Comparison of mean daily (MDF) with instantaneous peak (IPF) flow distributions

A comparison between the estimated parameters for the IPF and MDF samples for the year and the seasons are shown in Fig. ~~8~~7. As expected, the location parameters are consistently underestimated by the MDF series, with the largest errors in summer. This naturally leads to an overall downward shift of the "true" distribution when estimated from MDF values. The scales, here normalized by the location, appear to be primarily overestimated in summer, leading to distributions that are steeper for MDF than for IPF samples. For the year and winter, the errors in the scale parameters appear to be balanced in their directions.

The shape parameters differ quite substantially between the seasons. In summer the vast majority of estimated parameter values is negative, both in IPF and MDF. This indicates a heavy tail behavior for the summer floods. The fact that these negative values are in many cases smaller in the MDF than in the IPF sample, suggests that the tails are overstated in the former case. ~~This in combination with the underestimation of the location parameter leads to an overall underestimation of the lower and an overestimation of the higher flood quantiles by the MDF sample.~~ For the year and winter season, again, no clear trend is visible. The distribution parameters of the low-elevation gauges appear to be very well estimated by the MDF. For the higher elevations, especially the estimation of the shape parameters seems difficult. For the whole year, the IPF shape is underestimated at a lot of gauges, while it is primarily overestimated in winter. Overall, due to the underestimation of the location parameter leads, underestimation of both the lower and higher flood quantiles by the MDF sample is expected.


(a) Location Year
(b) Scale Year
(c) Shape Year
(d) Location Summer
(e) Scale Summer
(f) Shape Summer
(g) Location Winter
h) Scale Winter
i) Shape Winter
MDF ξ
IPF ξ
MDF α
IPF α
MDF k
IPF k
1
10
100
1000
0.0
0.4
0.8
1.2
1.6
-0.8
-0.4
0.0
0.4
0.8
Elevation [m a.s.l.]
100
200
300
400
500
600
700
800
900
1000

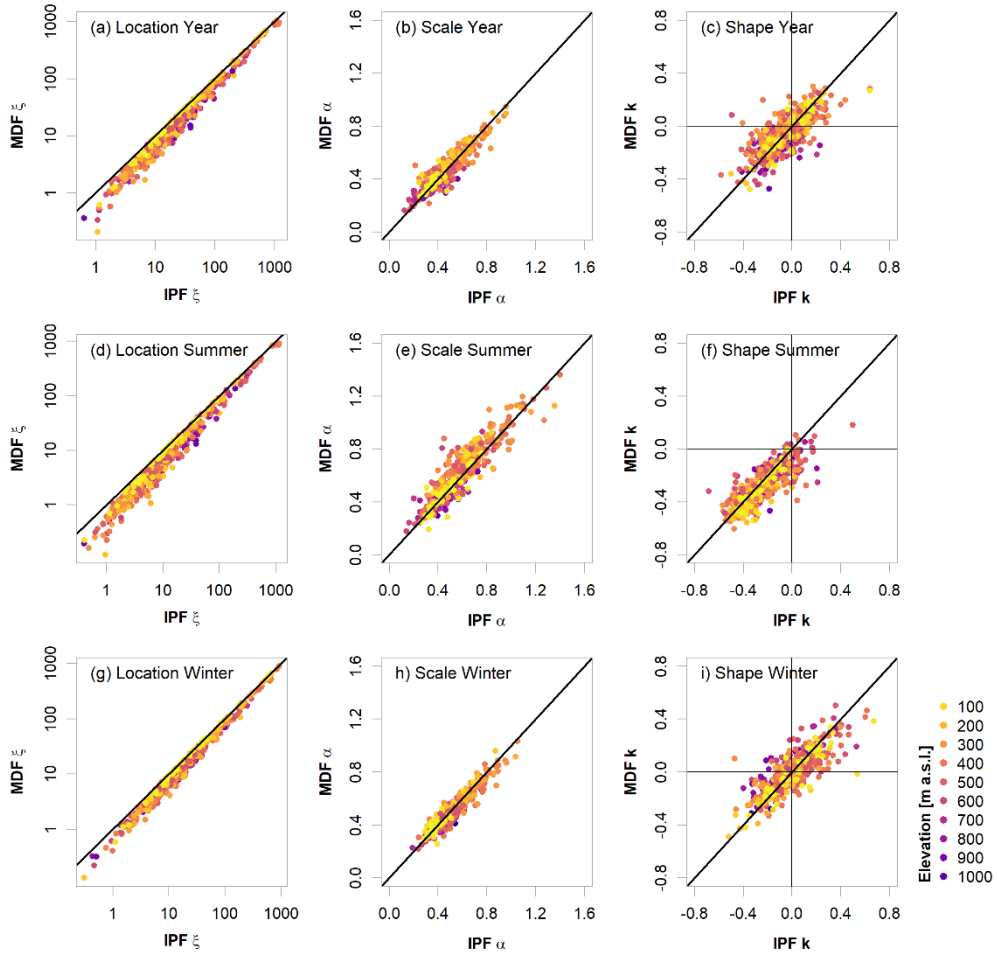


Figure ~~8: Estimated~~ 7: Observed Generalized Extreme Value (GEV) parameters from the instantaneous peaks (IPF) vs. mean daily (MDF ~~samples for the year and the two seasons.~~) annual/seasonal maximum series. Here only validation sites with observations longer than 30 years are shown.

Generally, the heavy tails of the summer distributions in contrast to the flatter tails in winter let the summer floods become dominant at higher quantiles. For a return period of 100 years, the summer floods exceed the winter peaks at 61.9% of the ~~stations~~sites. For 50 and 10 years this exceedance occurs at 51.2% and 35.7% of ~~stations~~sites, respectively. This behavior is also noticeable in the MDF but for fewer gauges, namely 53.4%, 43.2% and 21.0% for 100, 50 and 10-year return periods.

4.4~~2~~2.2 Estimation of instantaneous peak flow (IPF) distributions and quantiles

Three approaches were tested for estimating IPF flood quantiles based on MDF statistics: a) correcting the sample L-moments required for parameter estimation (~~LM~~p/V-Lmoms), b) correcting the parameters of the fitted distribution (~~LM~~p/V-params), and c) directly correcting the desired flood quantiles (~~LM~~p/V-quants). Method a) is convenient since a single model for each

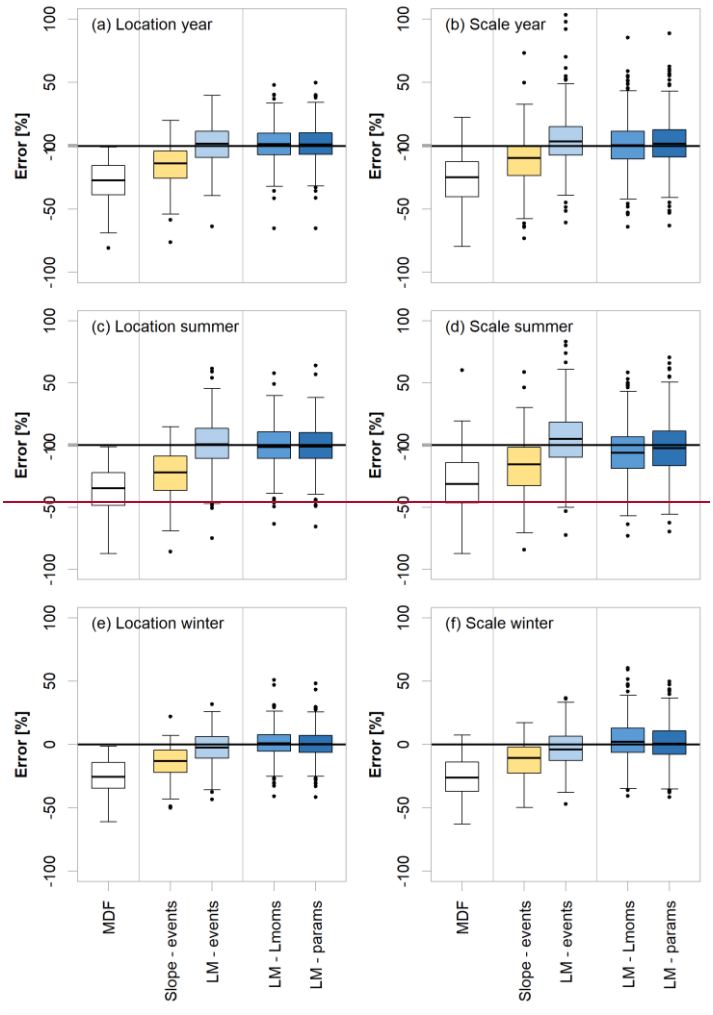
L-moment facilitates a correction of the complete distribution and hence each desired flood quantile. Estimating the L-moments has the additional advantage of not being restricted to a certain type of probability distribution. A proper distribution can be selected and fitted locally using the corrected L-moments. Still, the other methods may prove more robust and are hence tested as well. The final models for each target variable are selected according to the procedure for the MHQ (see Table ~~1~~2), using the calibration data set. For reasons of conciseness, only the final models are presented here with the coefficients of determination from the calibration (Table ~~5~~6). For further comparisons, distributions were also fitted to the annual and seasonal maxima that have been previously corrected using the slope (slope-events) and ~~LM~~ p/V-methods for events (~~LM~~p/V-events). Since the shape parameter is generally difficult to estimate, especially for such a short time period, and the models' estimates are generally close to the observed MDF shape parameter, it will not be estimated using the model variants. Instead, the MDF shape parameter estimate will be used in all instances.

Table ~~5: Linear~~6: p/V models fitted on the calibration data set for correction of ~~individual events,~~L-Moments (p/V-Lmoms), GEV-parameters (p/V-params) and flood quantiles (p/V-quants) derived from the mean daily flow (MDF) annual~~/~~ or seasonal ~~maxima and the MHQ~~maximum series. For an overview of the methods, the reader is directed to Table 1.

Type			Model	R^2
L-moments (p/V-Lmoms)	L1	Year	$L1_{MDF}$ / (0.74 - 0.94 * p/V_{mean} + 0.043 * log(area))	0.66
		Summer	$L1_{MDF}$ / (0.83 - 1.19 * p/V_{mean} + 0.054 * log(area))	0.65
		Winter	$L1_{MDF}$ / (0.99 - 1.31 * p/V_{mean} + 0.036 * log(area) - 0.00012 * elevation)	0.67
	L2	Year	$L2_{MDF}$ / (0.64 - 0.65 * p/V_{mean} + 0.048 * log(area))	0.45
		Summer	$L2_{MDF}$ / (0.71 - 0.86 * p/V_{mean} + 0.062 * log(area))	0.50
		Winter	$L2_{MDF}$ / (0.89 - 1.09 * p/V_{mean} + 0.043 * log(area) - 0.00016 * elevation)	0.53
GEV parameters (p/V-params)	ξ	Year	ξ_{MDF} / (0.77 – 1.02 * p/V_{mean} + 0.042 * log(area))	0.67
		Summer	ξ_{MDF} / (0.89 – 1.39 * p/V_{mean} + 0.052 * log(area))	0.64
		Winter	ξ_{MDF} / (0.96 – 1.36 * p/V_{mean} + 0.037 * log(area))	0.63
	α	Year	α_{MDF} / (0.67 – 0.77 * p/V_{mean} + 0.048 * log(area))	0.45
		Summer	α_{MDF} / (0.78 – 1.14 * p/V_{mean} + 0.064 * log(area))	0.42
		Winter	α_{MDF} / (0.97 – 1.24 * p/V_{mean} + 0.037 * log(area) - 0.00015 * elevation)	0.56
Flood quantiles (p/V-quants)	HQ10	Year	$HQ10_{MDF}$ / (0.72 - 0.87 * p/V_{mean} + 0.043 * log(area))	0.61
		Summer	$HQ10_{MDF}$ / (0.79 - 1.09 * p/V_{mean} + 0.058 * log(area))	0.60
		Winter	$HQ10_{MDF}$ / (0.96 - 1.23 * p/V_{mean} + 0.038 * log(area) - 0.00014 * elevation)	0.63
	HQ50	Year	$HQ50_{MDF}$ / (0.70 - 0.75 * p/V_{mean} + 0.044 * log(area))	0.52
		Summer	$HQ50_{MDF}$ / (0.73 - 0.83 * p/V_{mean} + 0.057 * log(area))	0.53
		Winter	$HQ50_{MDF}$ / (0.89 - 1.09 * p/V_{mean} + 0.043 * log(area) - 0.00016 * elevation)	0.54
	HQ50	Year	$HQ100_{MDF}$ / (0.69 - 0.70 * p/V_{mean} + 0.044 * log(area))	0.46
		Summer	$HQ100_{MDF}$ / (0.70 - 0.71 * p/V_{mean} + 0.057 * log(area))	0.46
		Winter	$HQ100_{MDF}$ / (0.87 - 1.03 * p/V_{mean} + 0.044 * log(area) - 0.00017 * elevation)	0.49

Fig. ~~9~~8 shows the errors (%) in GEV-parameter estimates for the different approaches in comparison to the original uncorrected MDF error ~~at~~(%) computed over the ~~490~~486 validation ~~stations~~sites with minimum of 30-~~year flow records.~~years of observations. Since the method p/V-quants directly corrects the MDF-quantiles, it cannot be used to estimate the GEV parameters and hence is not illustrated in the Fig. 8. All methods shown, clearly improve the estimation for the location and

scale parameters~~, where the L-Moment~~ when compared to the original MDF estimates. All corrections based on the p/V models proposed here (p/V-events, p/V-Lmoms and p/V-params) are less biased than the slope method (slope-event) proposed by Chen et al. (2017). Particularly the correction of the MDF sample L-Moments (p/V-Lmoms) shows ~~overall~~the smallest error and bias.



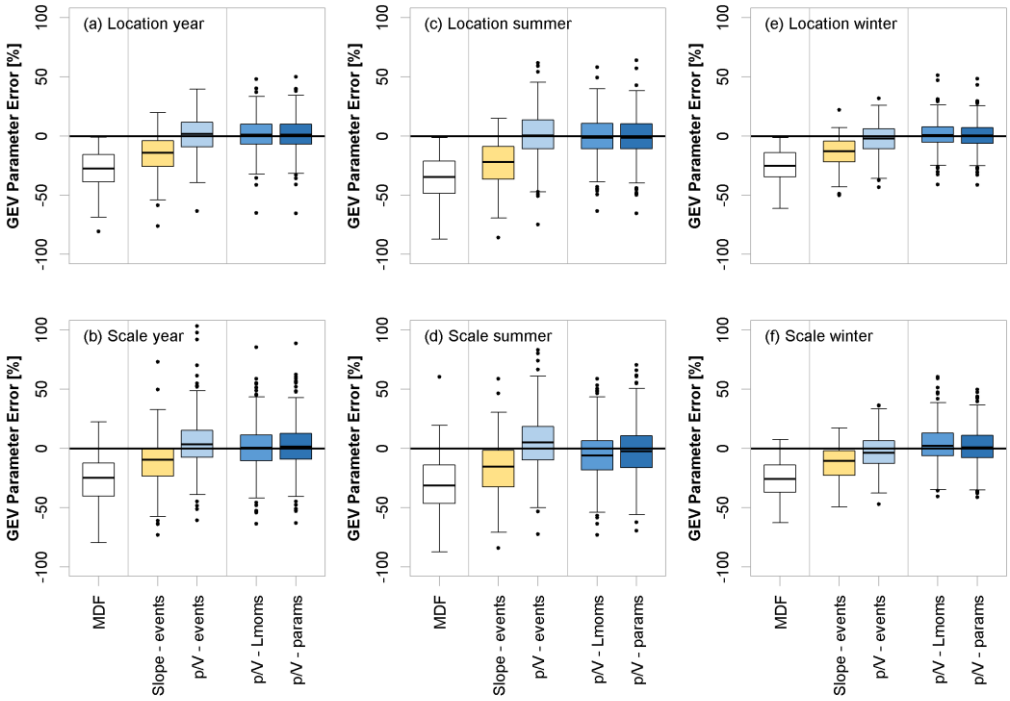


Figure ~~9: Comparison of performances of~~ 8: *Error (%)* comparison from various IPF-estimation methods (see Table 1 for explanation of methods) in their ability to estimate the Generalized Extreme Value (GEV) distribution parameters based on annual and seasonal maximum series. Only validation sites with more than 30 years of observation are used for the ~~year and the two seasons~~boxplots.

Fig. ~~10~~9 demonstrates the quality of the different correction approaches by means of the 10-, 50- and 100-year flood at the 486 validation ~~stations~~sites. With increasing return period, the performance of all correction methods appears to decline. Differences in the tails of the fitted distributions are more difficult to capture by the analyzed approaches. This turns out to be especially valid for the low-altitude catchments. The overcorrection that was observed for the mean is even more pronounced here, which leads to an average decline in model performance. Also, the general uncertainty in parameter estimation and extrapolation far beyond the time series length need to be kept in mind. Overall, even the estimation of the "true" IPF quantiles is potentially defective in itself, as will be discussed in the next section.

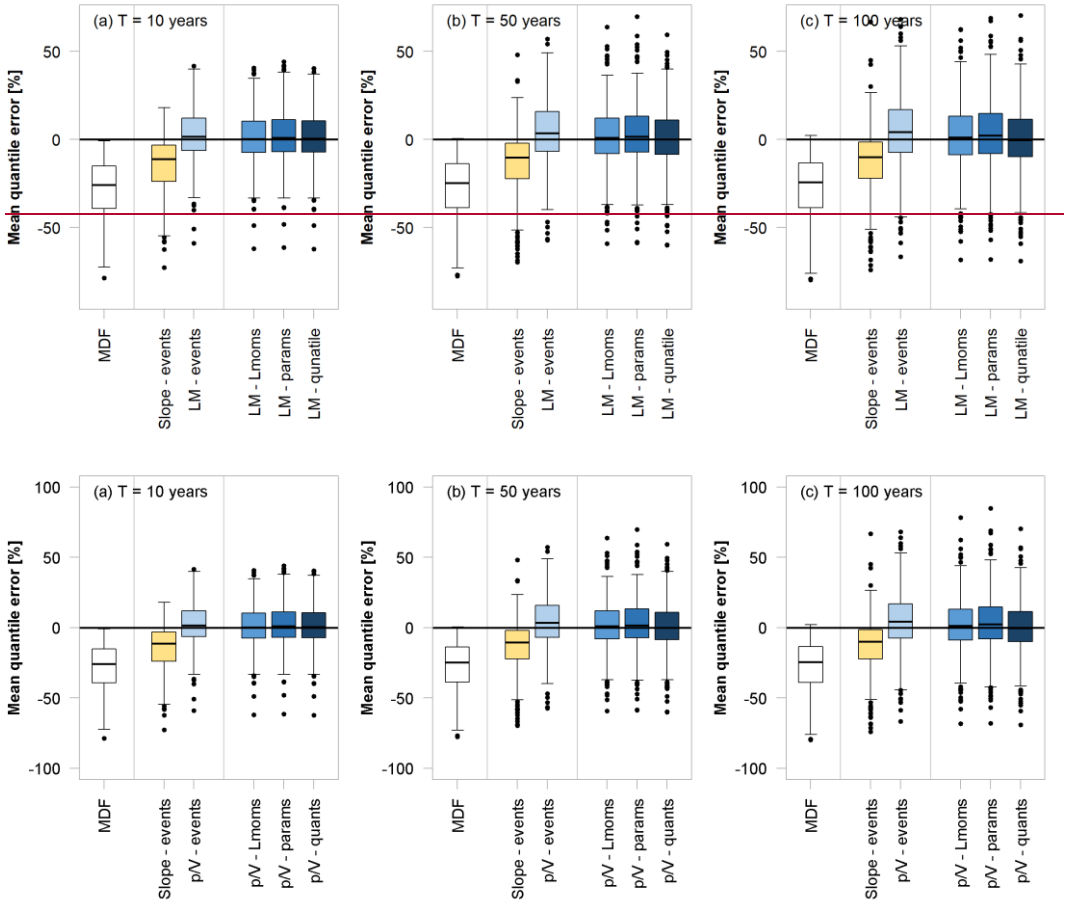


Figure ~~10:~~9: *Error (%)* comparison of ~~performances of~~ various IPF-estimation methods (see Table 1 for explanation of methods) in their ability to estimate different flood quantiles based on annual maximum series. Only validation sites with more than 30 years of observation are used for the ~~entire year~~boxplots.

Since the average *p/V* ratio is used for the direct correction of L-moments, parameters and flood quantiles, it is expected, that the performance of these methods decreases with increasing return period, ~~since~~as the ~~average *p/V*~~*V*$_{mean}$ ratio may not relate much to the higher quantiles. Still, even for the 100-year flood, these approaches appear to work just as well as the ~~LM~~p/V-event approach, as also indicated by the ~~normalized root mean square error (NRMSE)~~performance criteria (*nRMSE (%)* and ~~the percentage bias (PBIAS)~~*pBIAS (%)*) given in Table ~~6~~7. The performance of all three methods is comparable, but due to its previously named advantages, the L-moment method (p/V-Lmoms) is considered the superior approach in this setting.

Between the event correction techniques, the slope ~~method performs~~methods perform similar to the ~~LM method~~p/V methods in terms of overall error but ~~is~~are again more biased. When focusing on the catchments with areas below 500 km², the superiority of the ~~LM~~p/V-methods becomes apparent.

Table ~~6.~~7: Performances of different IPF estimation methods in terms of ~~NRMSE~~normalized root square mean error (*nRMSE* (%)) and percentage bias (*pBIAS* (%)) for different flood quantiles~~.~~ estimated from annual maximum series. The performance is computed over validation sites with more than 30 years of observations, while the values in parentheses show the performances for catchment sizes below 500 km².

	T = 10 years		T = 50 years		T = 100 years	
	nRMSE [%]	pBIAS [%]	nRMSE [%]	pBIAS [%]	nRMSE [%]	pBIAS [%]
MDF	17.8 *(50.0)*	-18.0 *(-32.9)*	17.8 *(48.1)*	-18.2 *(-32.3)*	17.9 *(47.5)*	-18.3 *(-39.1)*
Slope-events	7.0 *(30.3)*	-5.8 *(-17.5)*	8.7 *(27.7)*	-4.7 *(-15.8)*	10.3 *(27.8)*	-4.2 *(-15.2)*
~~LM~~p/V-events	7.7 *(20.1)*	-2.1 *(-1.3)*	8.5 *(19.1)*	-0.8 *(0.8)*	9.8 *(21.1)*	-0.3 *(1.8)*
~~LM~~p/V-Lmoms	8.2 *(21.1)*	-4.0 *(-3.3)*	8.7 *(21.1)*	-3.8 *(-2.8)*	9.3 *(22.8)*	-3.7 *(-2.6)*
~~LM~~p/V-params	8.1 *(20.6)*	-3.6 *(-2.3)*	8.5 *(20.7)*	-3.3 *(-1.6)*	9.1 *(22.5)*	-3.1 *(-1.2)*
~~LM~~p/V-quants	7.8 *(20.9)*	-3.6 *(-3.1)*	8.6 *(21.4)*	-4.0 *(-3.9)*	9.3 *(23.2)*	-4.3 *(-4.4)*

The distribution of the prediction error for the correction of ~~the LM~~L-moments (p/V-Lmoms~~ model~~) over the different catchment areas and elevations can be found in Table ~~7~~8. The errors are exemplary shown for the 100-year flood~~.~~ (HQ100). The error distribution is comparable to of the best estimation of MHQ shown in Table 5. Again, especially the overestimation for the lowest elevations is striking, as well as the significant underestimation at higher altitudes.

Table ~~7.~~8: Average prediction *error* (%) of the ~~LM~~p/V-Lmoms model for the 100-year flood (HQ100 ~~in percent~~) calculated over the validation sites and shown here for ~~the~~ different ranges of area and elevation. Red shades indicate overestimation, while blue shades underestimation.

		Elevation [m a.s.l.]									
		<100	<200	<300	<400	<500	<600	<700	<800	<900	<1200
Catchment area [km²]	<50	18.37	-34.61	-14.45	-12.68	2.66	-2.17	17.33	-6.80	-13.73	-15.58
	<100	24.81	-7.61	-8.01	6.46	4.91	0.72	3.98	5.26	7.49	-33.61
	<200	30.72	-9.97	-1.26	0.57	4.38	-4.76	3.33	5.37	-5.10	-
	<500	16.30	2.68	-1.65	2.56	4.34	8.03	5.78	2.56	-20.01	-
	<1000	7.95	1.56	-2.25	-1.93	6.88	3.36	3.43	-	-	-
	<2000	5.47	-8.43	-12.11	-2.63	-3.31	1.26	-6.80	1.19	16.97	-1.23
	<5000	-0.50	-7.33	-7.80	-3.66	-0.53	5.03	-	-	-	-
	<10000	-6.63	-9.86	-6.86	-4.07	-7.26	-	-	-	-	-
	<30000	-6.40	-17.31	-5.02	-0.92	-	-	-	-	-	-

Finally, the model performances of the mixed models, combining summer and winter floods, are analyzed for different flood quantiles. Their behavior is generally comparable to the annual maximum series approach, as shown in Fig. ~~11~~10. Even though the quantiles obtained with the mixed models may be more extreme and more parameters need to be estimated and corrected, there is no indication that the IPF correction will not function in this case.

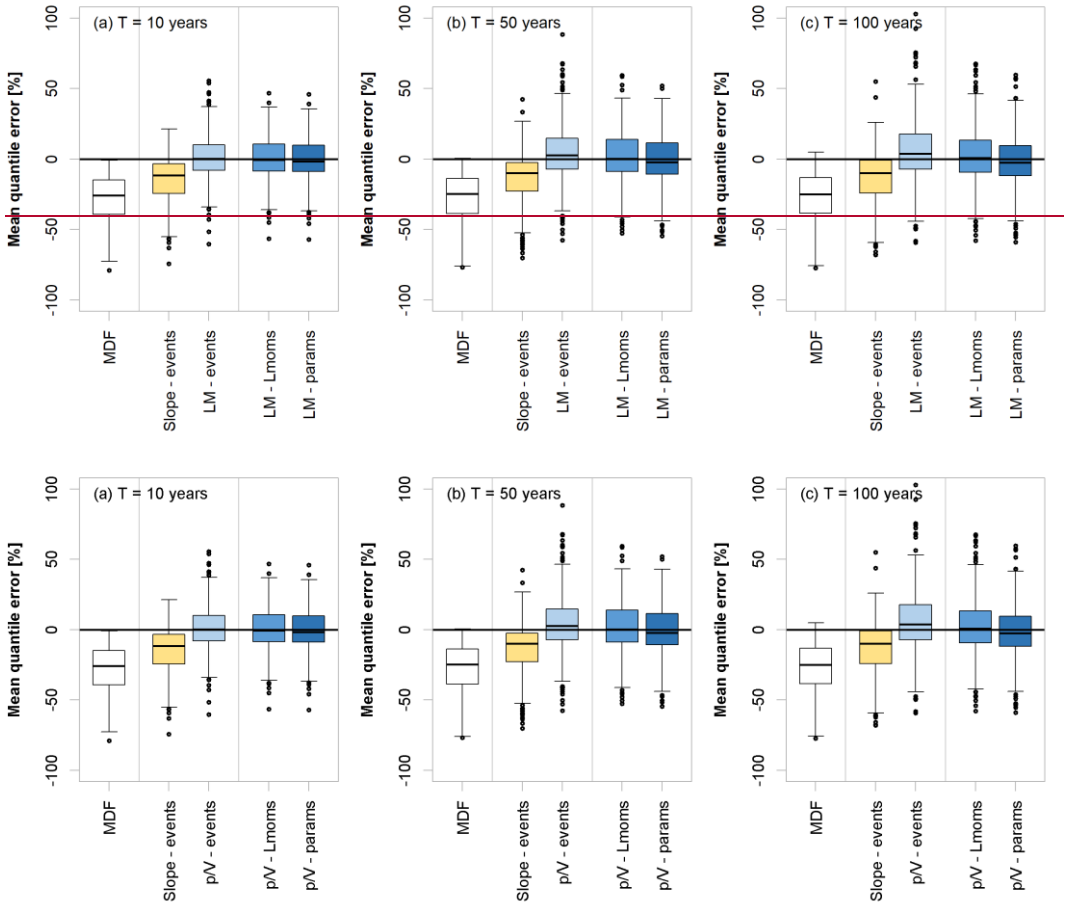


Figure ~~11:~~10: *Error (%)* comparison of ~~performances of~~ various IPF-estimation methods (see Table 1 for explanation of methods) in their ability to estimate different ~~seasonally mixed~~ flood quantiles based on mixed-models of annual maximum series. Only validation sites with more than 30 years of observation are used for the boxplots.

The *nRMSE* (%) and *pBIAS (%)* values for the mixed approach are shown in Table ~~8~~9. According to these values, the event-based correction methods appear to perform best overall. For the smaller catchments (< 500 km²) the ~~LM~~p/V methods outperform the slope ~~method~~methods.

Table 2: ~~Mixed model~~ Performances of different IPF estimation methods in terms of ~~NRMSE~~normalized root mean square error (*nRMSE* (%)) and percentage bias (*pBIAS* (%)) for different flood quantiles~~.~~ estimated from mixed models of seasonal maximum series. The performance is computed over validation sites with more than 30 years of observations, while the values in parentheses show the ~~performance~~performances for catchment sizes below 500 km².

	T = 10 years		T = 50 years		T = 100 years	
	nRMSE [%]	pBIAS [%]	nRMSE [%]	pBIAS [%]	nRMSE [%]	pBIAS [%]
MDF	17.7 *(50.2)*	-17.9 *(-32.9)*	17.5 *(48.3)*	-18.0 *(-32.3)*	17.6 *(48.0)*	-18.1 *(-32.1)*
Slope-events	8.1 *(31.7)*	-6.3 *(-18.6)*	9.6 *(28.6)*	-4.7 *(-16.6)*	10.6 *(28.5)*	-4.1 *(-15.9)*
~~LM~~p/V-events	8.1 *(21.2)*	-2.5 *(-2.7)*	8.5 *(20.8)*	-0.3 *(1.2)*	9.1 *(23.3)*	0.7 *(3.1)*

~~LM~~p/V-Lmoms	12.3 *(23.0)*	-5.7 *(-3.9)*	12.7 *(22.1)*	-5.8 *(-3.0)*	13.0 *(22.9)*	-5.8 *(-2.6)*
~~LM~~p/V-params	12.5 *(23.5)*	-6.2 *(-4.7)*	13.4 *(23.8)*	-7.3 *(-5.7)*	14.0 *(25.3)*	-7.9 *(-6.6)*

655

4.~~5~~3 Uncertainty Analysis

The results of the ~~bootstrapping~~resampling procedure used to assess uncertainty of IPF estimates are exemplary shown in Fig. ~~12~~11 for the 100-year flood (HQ100) at a single ~~station~~site with a reduced number of 100 ~~permutations~~realisations. In panel (a~~,~~), the IPF and MDF estimates for each ~~permutation~~temporal resampling of the annual maximum series are plotted against
660 each other~~.~~ (respectively IPF-bs and MDF-bs). This shows the bandwidths of both the IPF and MDF estimates as a result of sample and parameter uncertainty ~~in the distribution fitting.~~. Fig. ~~12~~ 11 (b) shows the ~~estimated~~resampled IPF flood quantiles (IPF-bs) vs. the quantiles estimated using the ~~LM models for each permutation. The dark blue points represent the full linear models using all available stations in the study area, while the light blue points represent 100 resampled~~ p/V-Lmoms model ~~estimates.~~by considering different sources of uncertainty: p/V-bs illustrates the uncertainty only due to the fitting of the p/V-
665 Lmoms model, p/V-full indicates the sample and parameter uncertainty (MDF-bs) propagated through the p/V-Lmoms model, and p/V-bs-bs combines the sample and the parameter uncertainty (MDF-bs) with the p/V-Lmoms model uncertainty (p/V-bs) to tackle the total uncertainty. In this example, it becomes obvious that the ~~range in flood quantile estimates due to permutation in the linear models~~uncertainty from the p/V model (p/V-bs) is significantly smaller than the ~~range in estimates due to distribution fitting.~~sample and parameter uncertainty (MDF-bs or even IPF-bs). This is valid for the majority of ~~stations~~sites
670 and is hardly affected from the number of realisations.

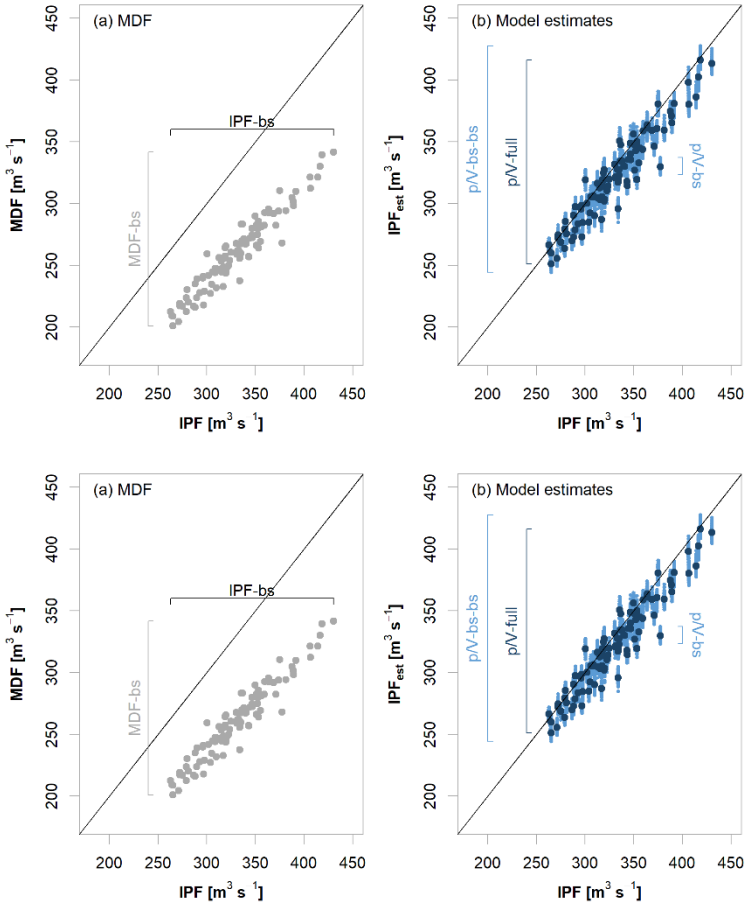
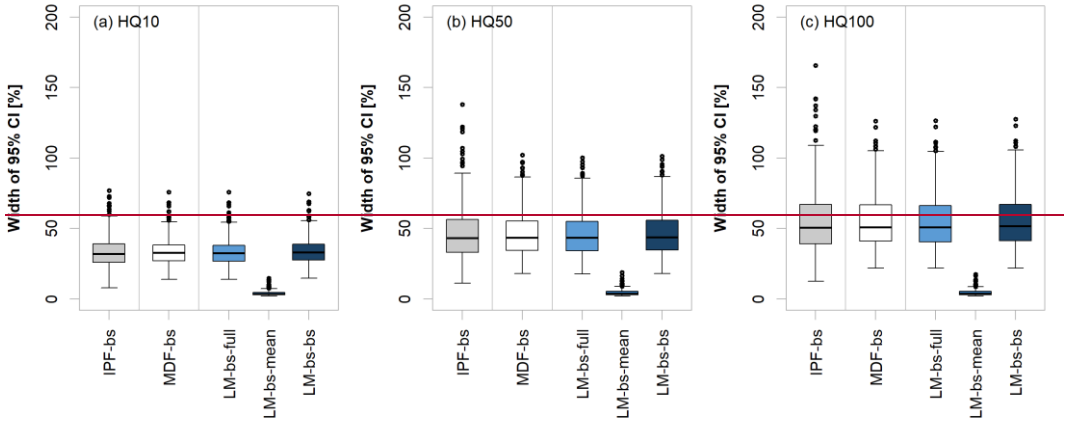


Figure ~~12:~~11: Example of ~~bootstrapping results at a station~~uncertainty ranges with 100 ~~permutations.~~realizations at a single site: (a) HQ100 from IFP-bs vs. MDF ~~for each permutation of~~ bs illustrating the ~~time series~~sample and parameter uncertainty, (b) HQ100 from IPF-bs vs. ~~100~~estimated IPF where p/V-bs illustrates the p/V-Lmoms model ~~estimates per permutation; the~~ uncertainty (shown as dark blue ~~dots represent the~~ points), p/V-full ~~model~~illustrates the propagation of sample and parameter uncertainty through the p/V-Lmoms model, and the p/V-bs-bs illustrate the total uncertainty that combines both sample and parameter uncertainty with the p/V-Lmoms model uncertainty.

Fig. ~~13~~12 shows the relative widths of the 95% confidence intervals for all ~~bootstrapping samples.~~types of uncertainty estimated. The average widths of the IPF-bs, MDF-~~bs,~~bs and p/V-full seem to be similar with each other, where the IPF sample and parameter uncertainty shows a larger variability ~~in the IPF sample.~~. The width of the average range of the ~~individual~~p/V-Lmoms model ~~permutations~~uncertainty (p/V-bs ~~mean~~) is very small at all ~~stations~~sites and therefore contributes little to the overall level of uncertainty (p/V-bs-bs). Thus the overall uncertainty of the p/V-Lmoms estimation method is mainly influenced by the sample and parameter uncertainty of the original MDF series.



~~Figure 13: Relative widths of various bootstrap samples for different flood quantiles.~~

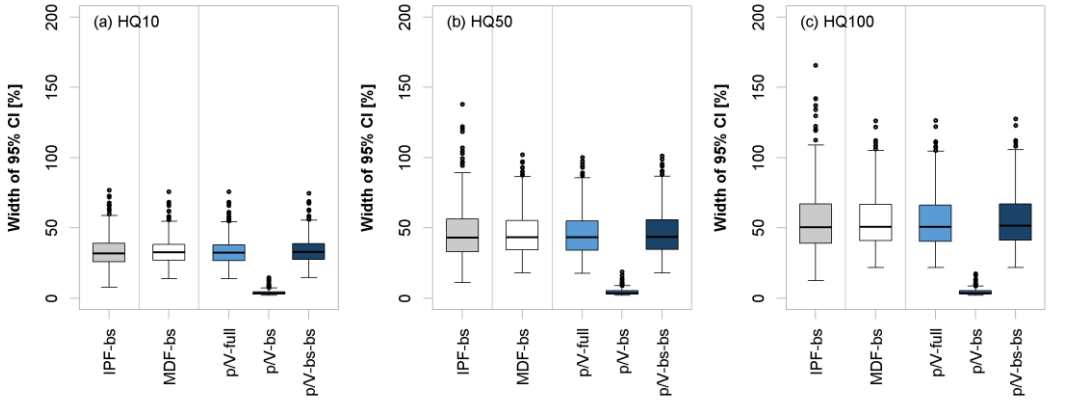


Figure 12: Relative widths of the 95% confidence interval (as per Eqn. 10) of various uncertainty types for different flood quantiles, where: IPF-bs and MDF-bs show the sample and parameter uncertainty of the original series, p/V-bs shows the sample and parameter uncertainty propagated through the p/V-Lmoms model, p/V-bs shows only the uncertainty of the p/V-Lmoms model and p/V-bs-bs shows the total uncertainty that combines both sample and parameter uncertainty with the p/V-Lmoms model uncertainty. The boxplots here are obtained for validation sites with more than 30 years of observations.

In order to assess the full bandwidth of the *errors* in the linear model estimates, they are compared to the range of errors in the MDF estimates. Here the *errors* for the uncertainty both in MDF (MDF-bs) and p/V-Lmoms (p/V-bs-bs) estimates are computed according to Eqn.11. Fig. ~~14~~13 shows the ~~mean~~median deviations of the MDF-bs and p/V-bs-bs quantiles from the ~~perturbed~~respective IPF-bs quantiles, as well as the lower and upper limits of the 95% confidence intervals of the errors for the 10-, 50- and 100-year flood quantiles. The median errors from LM-bs-bs are very similar over the three quantiles, but the higher quantiles HQ100 exhibits higher outliers. This is in agreement with the performance of the p/V-Lmoms model illustrated in Figure 9. This means that the median errors obtained over the 1,000,000 realizations are very similar with the actual model errors at each site. Moreover, it is obvious that the overall uncertainty gets larger with increasing return period, as can be seen

by the increasing distance between lower and upper confidence limits. The ~~LM model~~p/V-Lmoms estimates appear to be slightly positively skewed, which is especially noticeable in the 95% confidence interval for the HQ100. At many ~~stations~~sites there is a significant overestimation of the true IPF quantile ~~with some of~~when combining the ~~linear~~sample and parameter uncertainty with p/V-Lmoms model ~~transpositions~~uncertainty. The MDF estimates on the other hand exhibit the expected persistent underestimation.

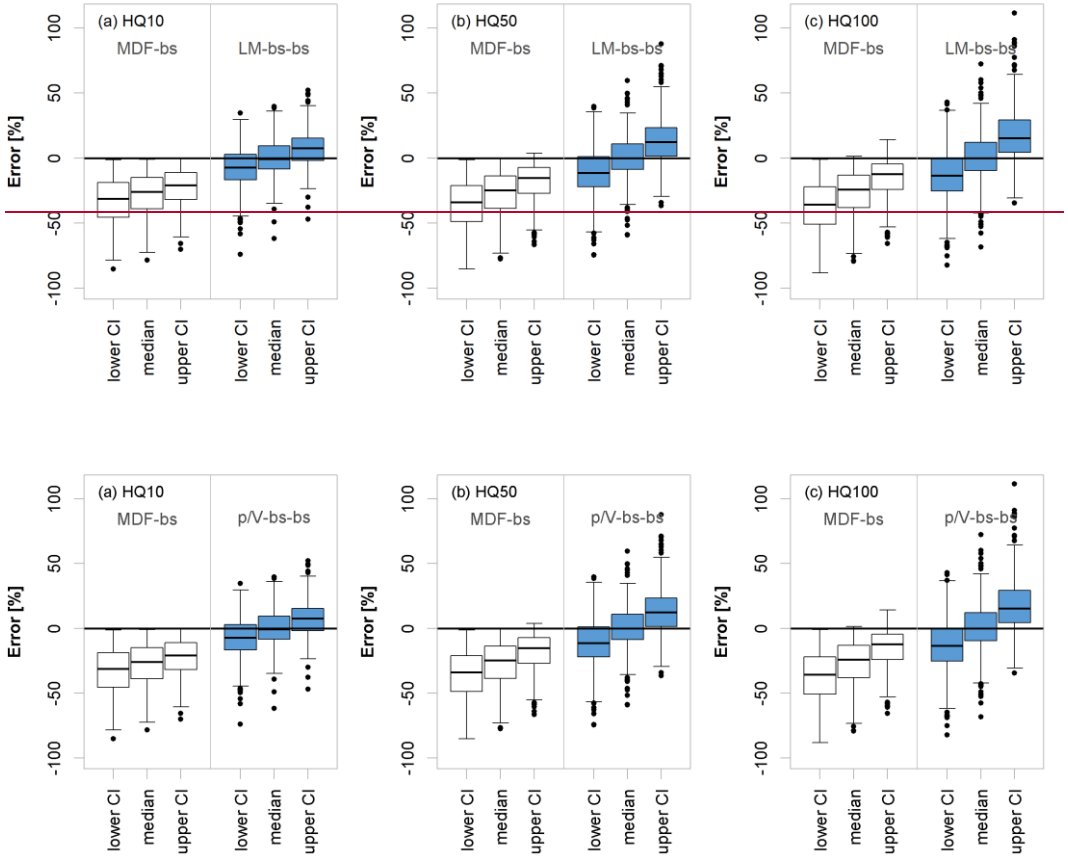


Figure ~~14:~~13: Error distribution ~~of the MDF (left) and LM (right) bootstrap samples~~obtained as per Eqn. 11 for three flood quantiles~~.~~ left) from considering the sample and parameter uncertainty of the mean daily flow (MDF) series (MDF-bs) and right) total uncertainty of the p/V-Lmoms model (p/V-bs-bs). Shown are the median errors, as well as the lower and upper limits of the 95% confidence intervals obtained at the validation sites with more than 30 years of observation.

Fig. ~~15~~14 summarizes the general overlap of the confidence intervals of MDF and estimated IPF with the confidence intervals of the observed IPF for the three flood quantiles~~.~~ (as per Eqn. 12). It becomes obvious that the agreement between IPF and the ~~LM~~ p/V-Lmoms model estimates is significantly larger than with the MDF values. This observation suggests that with high probability the ~~LM~~ p/V-Lmoms model estimates are in the range of the "true" IPF quantiles. The fact that overlaps in both the

MDF and the models increase with increasing return period suggests again the overall level of uncertainty in the higher IPF quantiles due to ~~distribution fitting~~sample and parameter uncertainty.

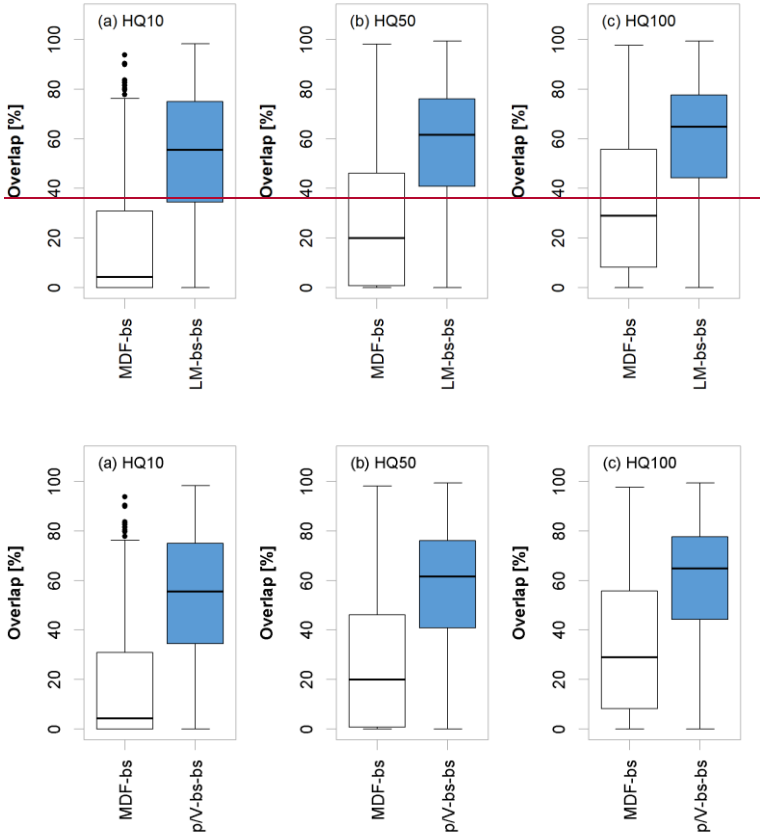


Figure ~~15:~~14: Percentage overlap ~~between~~for the three flood quantiles as per Eqn. 12 computed from the 95% confidence intervals of ~~all IPF bootstrap estimates~~mean daily flow (MDF) sample and ~~MDF~~parameter uncertainty (MDF-bs) and p/V~~ model bootstrap estimates for three flood quantiles.~~ Lmoms total uncertainty (p/V-bs-bs). The boxplots are obtained by considering validation sites with more than 30 years of observations

4.6 Range of applications and limitations

The method of correcting the error of MDF floods using the *p/V* ratio performs well and is easily applicable in our study area. However, its great simplification and mere approximation of physical flood generating processes results in some problems and limitations that will be listed and discussed here.

The first aspect that may influence the performance of the proposed IPF correction method is the event separation technique. The chosen technique determines how flood events and thus the required hydrograph characteristics are defined. The choice of baseflow separating algorithm can greatly affect the identification of start and end points of flood events. Strict independence criteria and thresholds for event recognition may lead to rejection of crucial flood events when considering daily time series.

Lax criteria, on the other hand, may create unnaturally long multi-peak events and false inclusion of small events, both leading to unrealistic hydrograph characteristics and IPF estimates. Thus, the additional step of refining multiple peak events, as suggested by Tarasova et al. (2018) should be carried out, when rainfall and snowmelt information is available. In their study, the refinement led to a reduction of multi-peak events from more than 50% to 44.7% of all identified events. In this study, the ratio of multi-peak to single-peak events is 57.9% for the year, 58.2% for summer and 58.4% for winter.

Using the p/V_{event} in order to correct individual events and then using the corrected series for FFA poses in theory a more sensible approach than using the p/V_{mean} from the annual MDF maxima. As mentioned before, maximum MDF events do not necessarily coincide with maximum IPF events, which is why correcting all events first and then selecting the annual maxima should yield a more appropriate IPF sample. But again, correcting individual events depends greatly on a very careful event separation, which could not be achieved in this case and which led to some unrealistic IPF estimates. Nonetheless, if a proper event separation is possible, the event correction method may have the larger potential. In such a case, a single model would be sufficient to account for all aspects of IPF estimation, including high flood quantiles.

A problem for IPF correction, which has been exhaustively discussed above, are gauges that exhibit little difference between MDF and IPF floods, even though their p/V ratio would suggest a much larger error. For our dataset this applies to the lowest-altitude gauges in the dataset. The MDFs at these ~~stations~~sites are overcorrected and thus exhibit severe overestimation of the true IPFs. We therefore discourage the application of the suggested correction methods at catchment outlets situated below 100 m a.s.l..

This observation may also suggest that other factors need to be considered for proper error estimation or that the parameters of the correction models need to be adjusted for different subsets of data. This is also relevant for the question of universality of the proposed method. Our data set is limited and representative of a temperate humid climate and moderate altitude. Thus, a qualitative sensitivity analysis is carried out on the full 648-~~stations~~gauge dataset in order to identify patterns that may be extrapolatable to other regions. The subsets are selected by combinations of geographical location, catchment size and gauge elevation. Target variable is the mean annual maximum IPF. Differences in the individual models due to different degrees of freedom are natural, which is why only those subsets that lead to significant deviations from the original model are mentioned here.

Two sets of ~~stations~~sites deviate noticably from the original model. The first one includes the low-altitude gauges discussed before. Here the overall error is so small that no correction yields better results than correction by the ~~linear model~~p/V approach. The second group includes the catchments with areas below 50 km². The errors for these ~~stations~~sites appear very scattered and randomly distributed. Comparing the *p/V* ratio from the daily series with the p/V ratio obtained from instantaneous events, it becomes obvious that the difference increases with decreasing catchment size and becomes excessively large and random for catchment sizes below 100 km². The correction using mean daily *p/V* ratio only functions where unknown instantaneous flood dynamics are roughly approximated by observed daily flow variability. The smaller the temporal scale of an instantaneous flood event, the poorer it is reproduced in the daily records. If instantaneous events manifest themselves

765 primarily on a ~~subdaily~~sub-daily basis, the possibility to describe their dynamics via daily flows becomes ineligible. This observation is also in accordance with the observed temporal shifts between MDF and IPF events, which is increasingly pronounced in smaller catchments. In summary, the proposed correction method founders at smaller scales below 100 km². Even though the IPF estimation leads to a general improvement at this scale, the daily flood time scale poses a poor predictor in these catchments.

770 Longitude and latitude do not appear to have any effect on the model fitting. Dividing the study area into quadrants does not result in any differences between the subsets, even when ~~equalizing the other factors~~considering similar catchment size and elevation. Also, neither record length nor period of record appear to have an influence.

The distinction between summer and winter for representation of the two most opposite flood types is particularly valid for this study area and should be adjusted where flood types are otherwise distributed. In general, even the rough distinction 775 between different flood types for IPF estimation proved meaningful in our case, as it revealed different dynamics and MDF-IPF relationships. This observation could be further exploited by more carefully defining and distinguishing flood types, as e.g. proposed by Fischer (2018) or Tarasova et al. (2020).

Finally, one should note that the type of distribution for flood quantile estimation can only be selected based on daily data and may differ from the optimal IPF distribution. For our data, the GEV proved flexible enough to be a good match in both MDF 780 and IPF but this could differ in other cases.

5 Conclusions and outlook

As in other studies before, it could be shown that the IPF-MDF relationship depends primarily on catchment size. It could also be observed that other factors, in this case gauge elevation, play a role in determining the difference between MDF and IPF floods. The relationship also appeared to differ between the two types of floods considered here, namely winter and summer 785 floods. Since summer floods are often caused by short but intense rain events and thus exhibit steep rising and falling limbs, their ~~subdaily~~sub-daily peaks are much larger than and difficult to estimate from the smoothed average daily peaks. Long, voluminous winter floods on the other hand show a much smaller IPF-MDF ratio and are easier to model.

This study has also shown that hydrograph characteristics, like the peak-volume ratio of flood events can be used to estimate instantaneous peak flows when only average daily series are available. The *p/V* ratio may be used to predict both IPFs of 790 individual events and instantaneous flood statistics, including mean annual and seasonal maximum flows and flood quantiles. Due to improper flood event separation, the event-based correction method produced some outliers in our case but may work significantly better when flood events can be defined more carefully. In general, the ~~LM~~p/V method requires a minimum of data and can be applied using mere information from the daily series itself. The performance could be marginally improved by including gauge elevation as additional predictor in some of the models.

795 The general recommendation for estimating IPF flood quantiles is to use the average *p/V* approach for correction of L-moments. This method is convenient since L-moments can be globally corrected while distributions may be locally fitted afterwards. It

turned out that the first two L-moments are easily estimated using p/V_{mean}, while higher order L-moments or L-moment ratios are more difficult to model with this approach.

There are two limitations, where the proposed method should be handled with care: a) at ~~stations~~sites with gauge elevations below 100 m, since it overestimates the true difference between IPF and MDF and b) at catchments smaller than 100 km², where it underestimates the error so that the full correction potential cannot be achieved. Still, in comparison to the slope method proposed by Chen et al. 2017, the ~~LM~~p/V approach works significantly better for smaller catchment areas, especially below 500 km². For larger catchments, the slope method appears very robust for all catchment sizes and elevations. The ~~LM~~p/V methods perform better in many larger catchments but outliers ~~my~~might be produced where the above-named restrictions are met.

For future analyses it will be meaningful to test the universality of the proposed approach in other study regions. Also, the effect of the flood event separation on the IPF estimation performance should be analyzed in more detail, especially in order to improve the event correction technique. Finally, it will be interesting to see if explicit consideration of more carefully defined flood types can improve the IFP estimation in mixed models.

Data availability

The discharge data used in this study is publicly available on the websites of the respective federal agencies.

Lower Saxony: Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz (NLWKN) http://www.wasserdaten.niedersachsen.de/cadenza/

Saxony-Anhalt: Landesbetrieb für Hochwasserschutz und Wasserwirtschaft Sachsen-Anhalt (LHW) https://gld-sa.dhi-wasy.de/GLD-Portal/

Saxony: Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie (LFULG) https://www.umwelt.sachsen.de/umwelt/infosysteme/ida/

Bavaria: Bayerisches Landesamt für Umwelt (LfU) https://www.gkd.bayern.de/de/

Baden-Württemberg: Landesanstalt für Umwelt Baden Württemberg (LUBW) https://udo.lubw.baden-wuerttemberg.de/public/

Author contribution

UH formulated the research goal. The study was designed by both authors and carried out by AB. AB and BS prepared the manuscript with contributions from UH.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This work is part of the research group FOR 2416 “Space-Time Dynamics of Extreme Floods (SPATE)” funded by the German Research Foundation (“Deutsche Forschungsgemeinschaft”, DFG).

References

Acharya, A. and Ryu, J. H.: Simple Method for Streamflow Disaggregation, J Hydrol Eng, 19, 509-519, doi: 10.1061/(ASCE)HE.1943-5584.0000818, 2014.

Canuti, P. and Moisello, U.: Relationship between the Yearly Maxima of Peak and Daily Discharge for Some Basins in Tuscany, Hydrolog Sci J, 27, 111-128, doi:10.1080/02626668209491094, 1982.

Chen, B., Krajewski, W. F., Liu, F., Fang, W. H., and Xu, Z. X.: Estimating instantaneous peak flow from mean daily flow, Hydrol Res, 48, 1474-1488, doi: 10.2166/nh.2017.200, 2017.

Dastorani, M. T., Koochi, J. S., Darani, H. S., Talebi, A., Rahimian, M. H:. River instantaneous peak flow estimation using daily flow data and machine-learning-based models, J. Hydroinform. 15, 1089–1098, doi:10.2166/hydro.2013.245, 2013.

Ding, J. and Haberlandt, U.: Estimation of instantaneous peak flow from maximum mean daily flow by regionalization of catchment model parameters, Hydrol Process, 31, 612-626, doi:10.1002/hyp.11053, 2017.

Ding, J., Haberlandt, U., and Dietrich, J.: Estimation of the instantaneous peak flow from maximum daily flow: a comparison of three methods, Hydrol Res, 46, 671-688, doi:10.2166/nh.2014.085, 2015.

Ding, J., Wallner, M., Muller, H., and Haberlandt, U.: Estimation of instantaneous peak flows from maximum mean daily flows using the HBV hydrological model, Hydrol Process, 30, 1431-1448, doi:10.1002/hyp.10725, 2016.

Deutscher Wetterdienst (DWD): Vieljährige Mittelwerte https://www.dwd.de/DE/leistungen/klimadatendeutschland/vielj_mittelwerte.html, last access: 17 September, 2021.

Ellis, W. and Gray, M.: Interrelationships between the peak instantaneous and average daily discharges of small prairie streams, Can Agr Eng–39, 1966.

Fill, H. D. and Steiner, A. A.: Estimating instantaneous peak flow from mean daily flow data, J Hydrol Eng, 8, 365-369, doi: 10.1061/(ASCE)1084-0699(2003)8:6(365), 2003.

Fischer, S.: A seasonal mixed-POT model to estimate high flood quantiles from different event types and seasons, J Appl Stat, 45, 2831-2847, doi:10.1080/02664763.2018.1441385, 2018.

Fischer, S., Schumann, A., and Schulte, M.: Characterisation of seasonal flood types according to timescales in mixed probability distributions, J Hydrol, 539, 38-56, doi:10.1016/j.jhydrol.2016.05.005, 2016.

Fuller, W. E.: Flood flows, T Am Soc Civ Eng, 77, 564– 617, doi:10.1061/taceat.0002552, 1914.

Gaál, L., Szolgay, J., Kohnová, S., Hlavčová, K., Parajka, J., Viglione, A., Merz, R., and Blöschl, G.: Dependence between flood peaks and volumes: a case study on climate and hydrological controls, Hydrolog Sc J, 60, 968-984, doi: 10.1080/02626667.2014.951361, 2015.

Haktanir, T. and Horlacher, H. B.: Evaluation of various distributions for flood frequency analysis, Hydrological Sciences Journal, 38, 15–32, https://doi.org/10.1080/02626669309492637, 1993.

Hosking, J. R.M.: L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics, Journal of the Royal Statistical Society: Series B (Methodological), 52, 105–124, https://doi.org/https://doi.org/10.1111/j.2517-6161.1990.tb01775.x, 1990.

Hosking, J. R. M. andWallis, J. R.: Regional Frequency Analysis, Cambridge University Press, https://doi.org/10.1017/CBO9780511529443, 1997.

Institute of Hydrology: Low flow studies, Report 1, Wallingford, UK, 1980.

Jarvis, A., Reuter, H. I., Nelson, A. and Guevara, E.: Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), 2008.

Jimeno-Saez, P., Senent-Aparicio, J., Perez-Sanchez, J., Pulido-Velazquez, D., Cecilia, J. M.: Estimation of instantaneous peak flow using machine-learning models and empirical formula in Peninsular Spain, Water, 9, 347, doi: 10.3390/W9050347, 2017.

Kumar, D. N., Lall, U., and Petersen, M. R.: Multisite disaggregation of monthly to daily streamflow, Water Resour Res, 36, 1823-1833, doi:10.1029/2000WR900049, 2000.

Langbein, W. B.: Peak discharge from daily records, Water Resour Bull, pp. 145, 1944.

Maidment, D. R.: Handbook of Hydrology, McGraw-Hill, New York, 1993

Muñoz, E., Arumí, J. L. and Vargas, J.: A design peak flow estimation method for medium-large and data-scarce watersheds with frontal rainfall, J Am Water Resour Assoc, 48 (3), 439–448, doi: 10.1111/j.1752-1688.2011.00622.x, 2012.

Sangal, B. P.: Practical Method of Estimating Peak Flow, J Hydraul Eng-Asce, 109, 549-563, doi: 10.1061/(ASCE)0733-9429(1983)109:4(549), 1983.

Shabani, M., Shabani, N.: Application of artificial neural networks in instantaneous peak flow estimation for Kharestan Watershed, Iran, J. Resour. Ecol. 3, 379–383, 10.5814/j.issn.1674-764x.2012.04.012, 2012.

Stedinger, J. R. and Vogel, R. M.: Disaggregation Procedures for Generating Serially Correlated Flow Vectors, Water Resour Res, 20, 47-56, doi: 10.1029/WR020i001p00047, 1984.

Taguas, E. V., Ayuso, J. L., Pena, A., Yuan, Y., Sanchez, M. C., Giraldez, J. V., and Perez, R.: Testing the relationship between instantaneous peak flow and mean daily flow in a Mediterranean Area Southeast Spain, Catena, 75, 129-137, doi: 10.1016/j.catena.2008.04.015, 2008.

Tarasova, L., Basso, S., Wendi, D., Viglione, A., Kumar, R., and Merz, R.: A Process-Based Framework to Characterize and Classify Runoff Events: The Event Typology of Germany, Water Resour Res, 56, doi:10.1029/2019WR026951, 2020.

Tarasova, L., Basso, S., Zink, M., and Merz, R.: Exploring Controls on Rainfall-Runoff Events: 1. Time Series- Based Event Separation and Temporal Dynamics of Event Runoff Response in Germany, Water Resour Res, 54, 7711-7732, doi: 10.1029/2018WR022587, 2018.

Tarboton, D. G., Sharma, A., and Lall, U.: Disaggregation procedures for stochastic hydrology based on nonparametric density estimation, Water Resour Res, 34, 107-119, doi: 10.1029/97WR02429, 1998.

Tan, K.-S., Chiew, F. H. S. and Grayson, R. B.: A steepness index unit volume flood hydrograph approach for sub-daily flow disaggregation, Hydrological Processes, 21, 2807–2816, doi: 10.1002/hyp.6501, 2007.

Viglione, A. and Blöschl, G.: On the role of storm duration in the mapping of rainfall to flood return periods, Hydrol Earth Syst Sc, 13 (2), 205–216, doi:10.5194/hess-13-205–200, 2009.

Villarini, G., Smith, J. A., Serinaldi, F., and Ntelekos, A. A.: Analyses of seasonal and annual maximum daily discharge records for central Europe, Journal of Hydrology, 399, 299–312, https://doi.org/10.1016/j.jhydrol.2011.01.007, 2011.