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

Anne Bartens[1], Uwe Haberlandt[1]

[1]Institute of Hydrology and Water Resources Management, Leibniz University of Hannover, Germany

5 *Correspondence to*: Anne Bartens (fangmann@iww.uni-hannover.de)

Abstract. In many cases flood frequency analysis needs to be carried out on mean daily flow (MDF) series without any available information on the instantaneous peak flow (IPF). We analyze the error of using MDFs instead of IPFs for flood quantile estimation on a German dataset and assess spatial patterns and factors that influence the deviation of MDF floods from their IPF counterparts. The main dependence could be found for catchment area but also gauge elevation appeared to

10 have some influence. Based on the findings, we propose simple linear models to correct both MDF flood peaks of individual flood events and overall MDF flood statistics. Key predictor in the models is the event-based ratio of flood peak and flood volume obtained directly from the daily flow records. This correction approach requires a minimum of data input, is easily applied, valid for the entire study area and successfully estimates IPF peaks and flood statistics. The models perform particularly well in smaller catchments, where other IPF estimation methods fall short. Still, the limit of the approach is reached

15 for catchment sizes below 100 km², where the hydrograph information from the daily series is no longer capable of approximating instantaneous flood dynamics.

1 Introduction

Common flood frequency analysis (FFA) is based on samples of maximum flows. The magnitude and variability of these maxima pose the baseline for the choice of probability distribution, the estimation of its parameters and eventually the

20 deduction of flood quantiles as design criteria. For FFA to be as accurate as possible, it is important to have a large number of peak flows measured with high precision, so that 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. Such data is rarely available and oftentimes FFA needs to be carried out on average daily flow records instead. The daily averaging naturally flattens the flood peak and the true maximum becomes unknowable. The degree of this smoothing, i.e. the

25 difference between the true instantaneous peak flow (IPF) and the maximum mean daily flow (MDF) 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). The internal variability of the MDF-IPF ratio within 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., 2013). A variety

30 of studies make use of the dependencies named above in order to estimate IPFs from MDFs, including Fuller (1914), Ellis & Gray (1966), Canuti & Moisello (1982), Taguas et al. (2008), Muñoz et al. (2012) and Ding et al. (2015).

Other IPF estimation methods aim at using the bare minimum of available data, i.e. solely the available daily flow record. 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 ratio.

35 Short events with steep rising and falling limbs are typical of a quickly reacting system, due to limited storage capacity and/or high intensity rainfall. In such a case, the discrepancy between IPF and MDF will be significantly larger than for hydrographs with long durations and gentle slopes. E.g., Ellis & Gray (1966) found that the peak ratio IPF/MDF distinctively decreases with increasing hydrograph width.

Several approaches use the maximum daily flow and the discharge of the previous and/or successive day (e.g. Langbein, 1944)

40 to predict IPFs. Chen et al. (2017) compare two of these methods, namely those of Sangal (1983) and Fill and Steiner (2003). They also propose their own method based on the rising and falling slopes of the event hydrograph, estimated from the three consecutive days around the peak. They found that their slope-based method and Fill and Steiner's method perform well and are probably applicable under a wide range of climates. However, both methods' performances decrease with decreasing catchment size and work best for areas larger than 500 km².

45 There naturally 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 regionalized model parameters (Ding and Haberlandt, 2017). Several studies have applied machine learning techniques to estimate instantaneous peaks from daily data,

50 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 a number of computational steps and/or a variety of data sources.

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

55 findings, a method is proposed that facilitates 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-

60 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 the expected magnitude of the IPF. 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
65 for FFA, and for direct correction of site-specific distribution parameters and flood quantiles.

2 Methods

2.1 Analysis and estimation of IPF peaks

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
70 on the analysis of flood statistics. The percentage deviation of the MDF statistic *MDF*$_{stat}$ from the IPF statistic *IPF*$_{stat}$

$$\frac{\mathrm{MDF_{stat}} - \mathrm{IPF_{stat}}}{\mathrm{IPF_{stat}}} * 100 \ \% \tag{1}$$

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.

In order to improve the IPF estimation by MDF, several correction methods are applied, which make use of the peak-volume
75 ratio. 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

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

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
80 computed. Then a linear regression model of the following form is fitted

$$\mathrm{IPF_{event}} = \frac{\mathrm{MDF_{event}}}{(a + b_1 * \mathrm{p/V_{event}} + b_2 * \mathrm{CD}_1 + \cdots + b_{n+1} * \mathrm{CD}_n)}, \tag{3}$$

where *CD* denotes additional catchment descriptors that may be included in the models. The combination of hydrograph shape
and catchment characteristics 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.

85 The described event correction method will be compared with the slope correction method developed by Chen et al. (2017).
This method estimates an instantaneous event peak flow based on the slopes of the daily peak *Q*$_{peak}$ to its preceding and
following daily flows *Q*$_{pre}$ and *Q*$_{suc}$. The IPF is thus estimated as

$$\mathrm{IPF_{event}} = \mathrm{Q_{peak}} + \frac{(\mathrm{Q_{peak}} - \mathrm{Q_{pre}}) * (\mathrm{Q_{peak}} - \mathrm{Q_{suc}})}{2 * \mathrm{Q_{peak}} - \mathrm{Q_{pre}} - \mathrm{Q_{suc}}} \ . \tag{4}$$

For validation, both event correction methods are applied in two ways: 1) IPFs are estimated for all separated events in the
90   daily flow series, even if these events have small daily peaks. 2) IPFs are estimated for the annual maximum daily peak only.
In both cases statistics are derived from the estimated IPF series and compared to the observed IPF statistics. Procedure 1) is
theoretically more accurate, since maxima in IPF and MDF do not necessarily 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 may prove more robust in cases where
95   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. The larger events containing the annual maximum MDF are expected
to be more properly separated by the algorithm described below.

In order to be able to estimate 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, sample L-moments, estimated
100 distribution parameters and derived flood quantiles based on averaged p/Vs. These average p/Vs are obtained from all annual
maximum MDF events at each station. 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

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

105 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.

2.2 Event separation

110 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
115 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.

120 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
125 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

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

with location parameter ξ, scale parameter α and shape parameter k. The parameters were estimated using sample L-moments.
130 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:

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

135 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.

140 2.4 Uncertainty

Since both distribution fitting and IPF estimation via linear models are approximations and not fully accurate, we eventually assess the overall level of uncertainty in the final IPF flood quantile estimates. This is done using simple bootstrapping procedures. In a first step, the series of annual maxima from both daily and monthly maximum data are analogously resampled 1000 times with replacement. For each resampling the desired flood quantiles are estimated using L-moments. The range of
145 these estimates provides the baseline level of uncertainty due to distribution fitting.

In a second step, linear regression models are fitted to each pairing of estimated IPF and MDF flood quantiles over all stations in the study area. In order to assess the uncertainty of the fitted models, another resampling is carried out, this time shuffling

the set of considered stations, again 1000 times with replacement. For each station this procedure yields 1000 estimates of paired flood quantiles from both the IPF and MDF series (IPF-bs and MDF-bs), 1000 full-model quantile estimates resulting from the original p/V model fitted to each permutation (p/V-full), and 1000*1000 quantile estimates resulting from permutation of the p/V model for all IPF and MDF transpositions (p/V-bs-bs).

In order to assess the overall level of uncertainty, several indices will be assessed at the individual stations. The first one is the relative width of the 95% confidence intervals (CI) calculated for all aforementioned bootstrap sample estimates of the desired flood quantile

$$CI_{bs} = \frac{x_{bs;0.975} - x_{bs;0.025}}{x_{bs;0.5}},$$
(8)

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

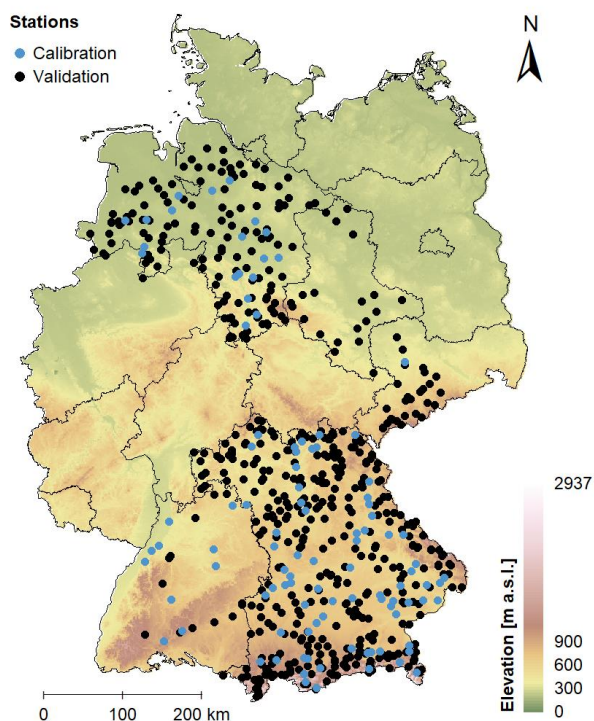
$$error_{bs} = \frac{x_{bs} - IPF_{bs}}{IPF_{bs}} * 100\%.$$
(9)

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-model samples with the IPF confidence bands are determined as percentage overlap at each gauge:

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

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). 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
175 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.

180 **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
185 (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.

190

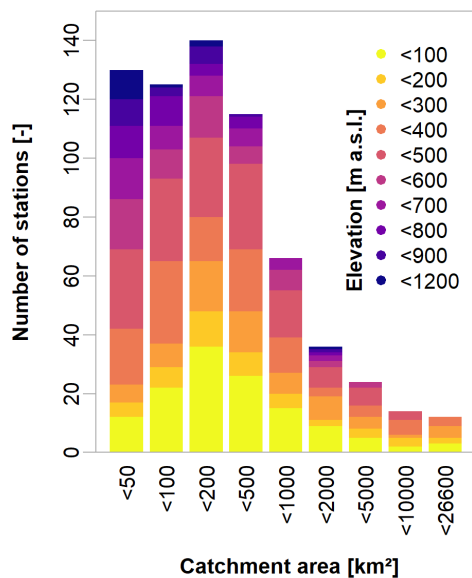



Figure 2: Distribution of catchment size and elevation of the 648 gauges used for ~~analysis~~.

4 Results and discussion

4.1 Comparison of MDF and IPF peaks

195 In theory, the relative deviation between MDF and IPF peaks depends greatly on catchment size. 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 data set. Fig. 3 demonstrates by means of the mean annual maximum
200 flow that the larger the area, the smaller the deviation between MDF and IPF. Also, errors appear to be especially large in higher altitudes. Generally, the error seems to increase in north-south direction, which could be a secondary effect of both increasing altitude and decreasing catchment size.

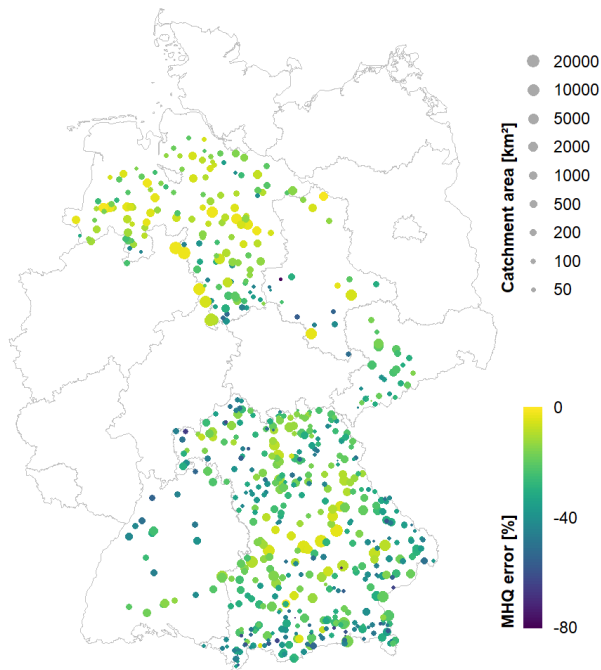



Figure 3: Spatial distribution of the discrepancy between MDF and IPF in the MHQ.

205 When assessing the differences between average daily and instantaneous peaks, it is also meaningful to take a closer look at different types of floods. For our German dataset the two most opposite types are a) flood events induced by short intense rainfall, especially convective events, and b) extended flood events with significant volume, as caused by snowmelt and/or stratiform rain. Presumably, the latter flood type is much better represented by average 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

210 - April) half years. Due to the limited data availability, a clear distinction between convective, stratiform and snow-melt events cannot be achieved here. 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.

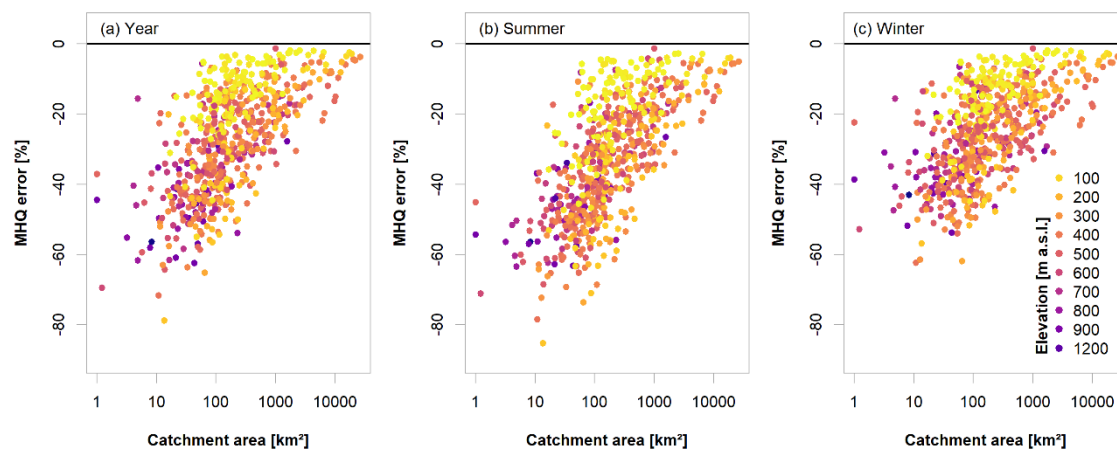
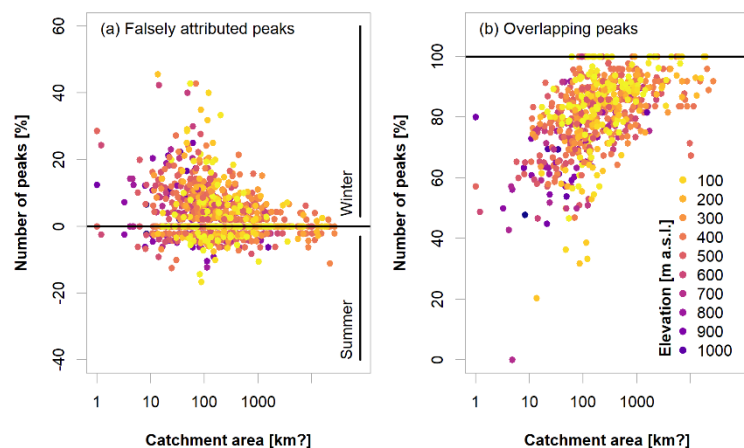



Figure 4: Error in the MHQ in relation to catchment size and gauge elevation for the entire year (a), summer (b) and winter (c).

215   In Fig. 4 the error in the MHQ is shown for the entire year and the summer and winter half year. The relationship with catchment area is clearly visible in all three cases. Also, the effect of the elevation becomes obvious, namely in the lowest elevations (yellow points, below 100 m) showing 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. As expected, the error is overall smaller in the winter
220   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 to the annual peaks depends on the individual flood regimes. At 68.8% of the considered gauges 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
225   the average daily flow smooths significant peaks to a point where they are no longer relevant for the overall flood behavior. Fig. 5 a shows the percentage of annual maxima at each gauge that are attributed to the wrong season when using MDF. Each gauge 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
230   in summer. Apart from not being able to properly identify flood magnitudes when using daily flow series, 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 the asynchronous occurrence of IPFs and MDFs. Instantaneous maxima are not always identifiable in the daily flow series, i.e. the maxima obtained from the daily series
235   are inevitably found in other places. In general, the smaller the catchment, the smaller the temporal overlap between instantaneous and daily peaks, as seen in Fig. 5 b. 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 events and thus to different populations.

240 **Figure 5: Percentage of peaks falsely attributed by MDF to the winter or summer half-year (a) and percentage of peaks in MDF and IPF overlapping in time (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. There seems to be a significant linear dependence between the peak ratios MDF/IPF and p/V and the logarithm of the catchment size. Various
245 model combinations with the available variables catchment area, elevation and p/V are tested using the calibration data set. Table 1 lists the coefficients of determination of these model combinations. An asterisk indicates that not all regressors in the model are significant ($p = 5\%$). The selected models are marked in grey and the full model formulas are given in Table 2. For most models, the majority of variance in the IPF-MDF relationship is explained by the p/V and the catchment area. For winter, including gauge elevation appeared to improve the model slightly.

250 The models show a similar performance for the annual and summer peak ratios both for the individual events and the MHQ. For winter, the model performance seems to differ, especially for the event correction. It appears that the linear models using the p/V 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. The fact that elevation is a significant regressor in the MHQ model may also suggest that the peak ratios in winter are more heterogeneous.

255

Table 1: Coefficients of determination of different target variables and various linear model combinations. Grey cells indicate the best linear model for each target variable. Asterisks indicate non-significant regressors in the linear models.

Target variable / Regressors	Event peak MDF/IPF			Annual/seasonal maximum MDF/IPF			MHQ MDF/IPF		
	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}	0.13	0.13	0.09	0.21	0.21	0.20	0.55	0.49	0.49
p/V_{mean} + Area	0.23	0.26	0.17	0.39	0.36	0.35	0.66	0.65	0.63
p/V_{mean} + Elevation	0.14	0.13	0.10	0.23	0.22	0.23	0.56*	0.51	0.56
p/V_{mean} + Area + Elevation	0.23	0.26	0.17	0.40	0.36*	0.36	0.67*	0.65*	0.68

Table 2: Linear models fitted for correction of individual events, annual/seasonal maxima and the MHQ.

Type		Model
Events	Year	MDF / (0.59 - 0.43 * p/V_{mean} + 0.047 * log(area))
	Summer	MDF / (0.44 - 0.36 * p/V_{mean} + 0.063 * log(area))
	Winter	MDF / (0.63 - 0.35 * p/V_{mean} + 0.044 * log(area))
Maxima	Year	MAX_{MDF} / (0.53 - 0.42 * p/V_{mean} + 0.057 * log(area))
	Summer	MAX_{MDF} / (0.61 - 0.73 * p/V_{mean} + 0.061 * log(area))
	Winter	MAX_{MDF} / (0.70 - 0.68 * p/V_{mean} + 0.054 * log(area))
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)

260

Fig. 6 shows the change in mean absolute error in the annual MHQ after correction with the different methods in relation to catchment size and elevation. The slope method (a) applied to the individual events yields a rather constant reduction of the error independent of catchment size. However, there are several outliers produced by the method, which can be attributed to improper separation of smaller events. Applying the slope method only to the annual maximum MDF events, as done in Fig.

265 5 b shows a much smoother and more constant error reduction. The methods using the linear model (Fig. 5 c-e) yield a much larger improvement for the smaller catchments where the error is generally larger than in the bigger catchments. However, these methods simultaneously lead to an increase of the error in several cases. This deterioration appears to affect those stations that have been highlighted before in section 4.1, namely the ones with the lowest elevations in the data set.

The differences between correcting the individual events (Fig. 5 c) and only correcting the annual maxima (Fig. 5 d) using the

270 linear models appear rather small in this case. This suggests that even though the annual maximum MDF does in many cases not occur at the same time as the maximum IPF, the method still yields an appropriate estimate of the true IPF. The method of directly correcting the MHQ (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

275 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 and in some cases to severe over- or underestimation of the peak. The weight of such events is assumed to be significantly lower when correcting flood statistics based on average p/Vs. In addition, the overall performance can only be assessed for events that contain the monthly maximum flow, i.e. primarily larger events. How the event correction performs for minor
280 events cannot be analyzed here.

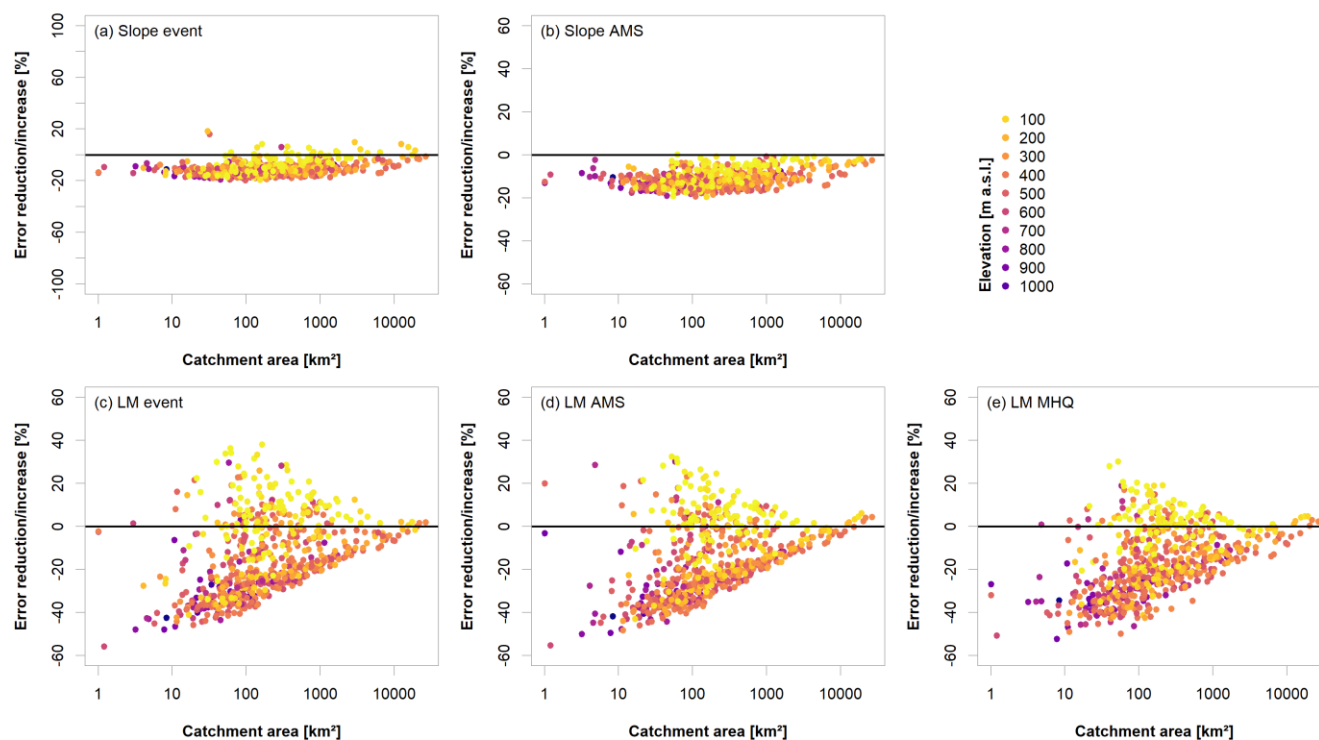



Figure 6: Error reduction (negative values) / increase (positive values) in the mean maximum flow for different IPF estimation methods when compared to MDF.

Fig. 7 summarizes the overall model performances for the mean annual/seasonal maximum flow at all 648 stations and
285 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 quite a large bias, which is, as seen above, not only disadvantageous. Still, the overall error is smaller for the LM methods, with fewer positive outliers produced by the LM-MHQ approach.

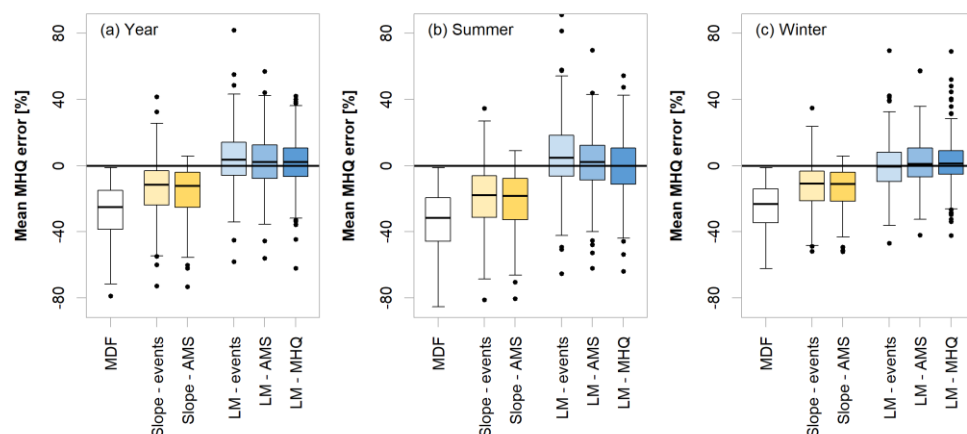



Figure 7: Comparison of performances of different IPF estimation methods for the entire year (a), summer (b) and winter(c).

290   Table 3 summarizes the normalized root mean square error (NRMSE) and the percentage bias (PBIAS) of the instantaneous MHQ estimated via the different model variants. In terms of NRMSE, the performances of the slope and LM methods are comparable, while the slope methods are more biased. There are a number of outliers produced by the LM methods, especially positive ones, that affect the overall NRMSE. As seen in Fig. 6, this is primarily concerning the low elevation catchments below 100 m. The values in parentheses in table 3 indicate the performance criteria for gauges with catchment areas below

295   500 km². Here, the advantage of the LM-approaches over the slope method become apparent, even though a large number of low elevation catchments fall in this category, which negatively affect the overall error.

Table 3: NRMSE and percentage bias of estimated vs. observed instantaneous MHQ over all sites for different model varaints. 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-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-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-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)*

300

Table 4 shows the average deviation of the MHQ predicted with the LM-MHQ model from the observed instantaneous MHQ for different catchment sizes and elevations. It becomes obvious that for the smallest elevations, the instantaneous 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

305   large negative errors in most cases.

Table 4: Average prediction error of the LM-MHQ model for the MHQ in percent for the different ranges of area and elevation.

		Elevation [m a.s.l.]									
		<100	<200	<300	<400	<500	<600	<700	<800	<900	<1200
Catchment area [km²]	<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	-	-	-	-	-	-

310

4.3 Comparison of IPF and MDF distributions

The GEV distribution appeared to be a generally suitable distribution for the stations 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 corrected samples at each station and certified a good fit in all cases ($p = 5\%$).

315 A comparison between the estimated parameters for the IPF and MDF samples for the year and the seasons are shown in Fig. 8. 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.

320 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, again, no clear trend is visible.

325 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.

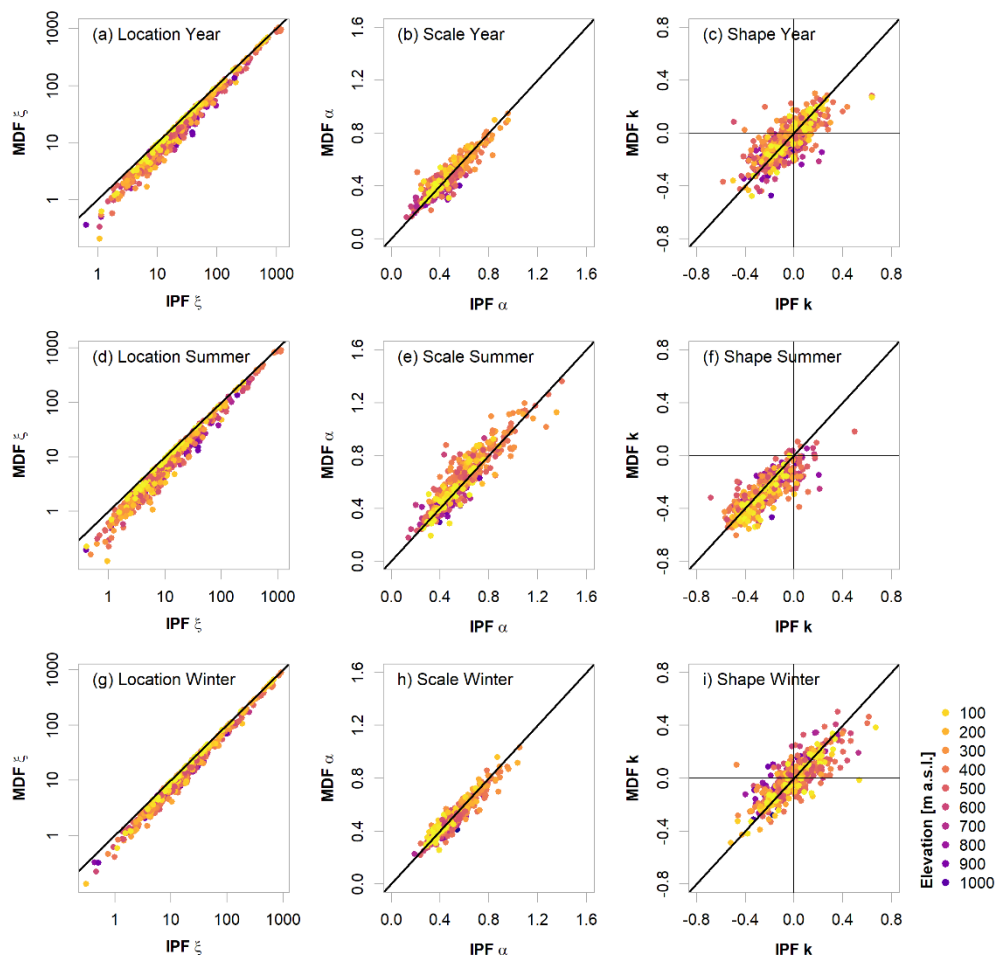



Figure 8: Estimated GEV parameters from the IPF vs. MDF samples for the year and the two seasons.

330   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. For 50 and 10 years this exceedance occurs at 51.2% and 35.7% of stations, 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 Estimation of IPF quantiles

335   Three approaches were tested for estimating IPF flood quantiles: a) correcting the sample L-moments required for parameter estimation (LM-Lmoms), b) correcting the parameters of the fitted distribution (LM-params), and c) directly correcting the desired flood quantiles (LM-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

340 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), 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). 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 methods for events (LM-events).

345 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 models fitted for correction of individual events, annual/seasonal maxima and the MHQ.

Type			Model	R^2
L-moments	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	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	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

350

Fig. 9 shows the errors in parameter estimates for the different approaches in comparison to the original uncorrected MDF error at the 490 validation stations with minimum 30-year flow records. All methods clearly improve the estimation for the location and scale parameters, where the L-Moment correction shows overall smallest error and bias.

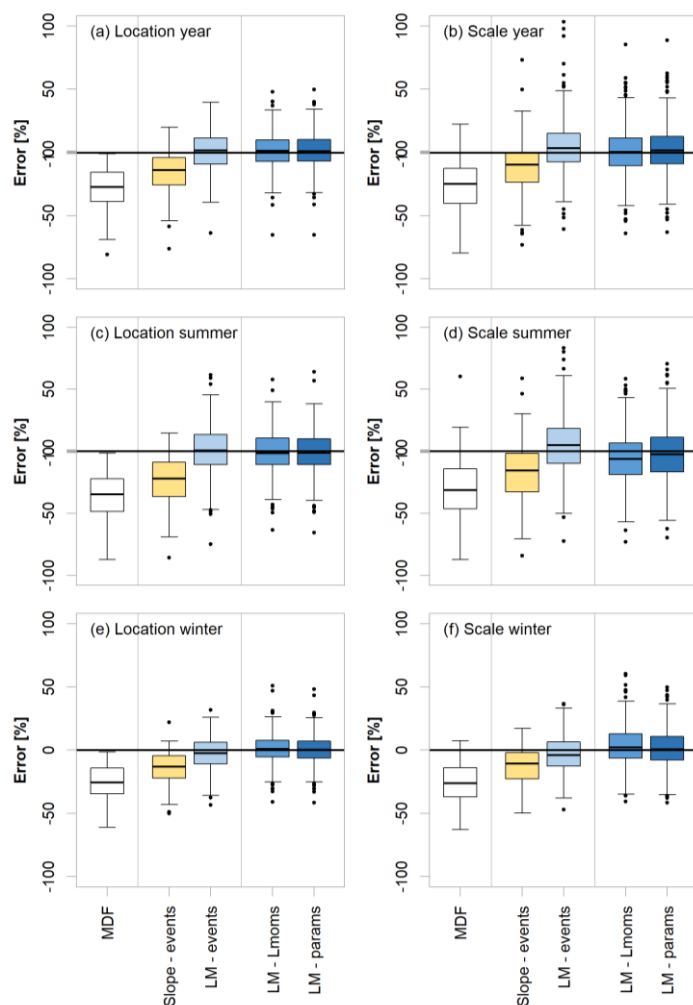



Figure 9: Comparison of performances of various IPF-estimation methods for the GEV distribution parameters for the year and the two seasons.

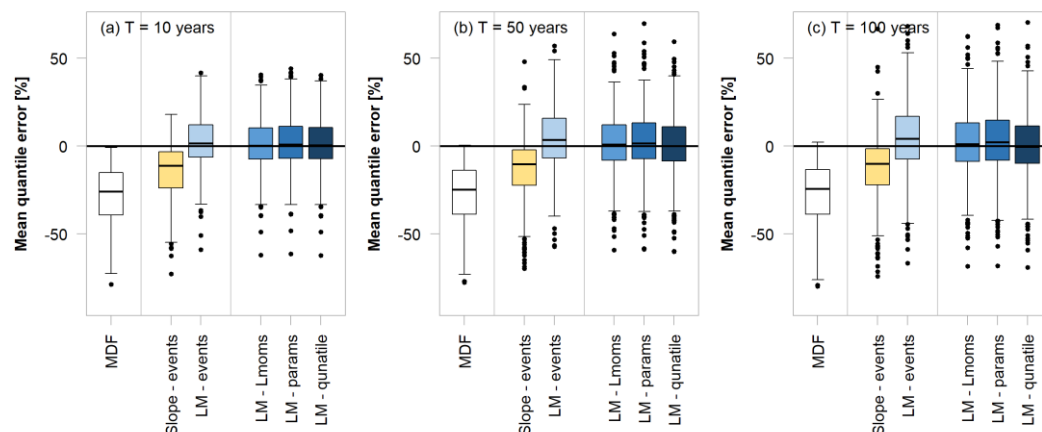
Fig. 10 demonstrates the quality of the different correction approaches by means of the 10-, 50- and 100-year flood at the 486 validation stations. With increasing return period, the performance of all correction methods appears to decline. Differences in the tails of the fitted distributions are more 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.

365

Figure 10: Comparison of performances of various IPF-estimation methods for different flood quantiles for the entire year.

Since the average p/V 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 the average p/V 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-event approach, as also

370 indicated by the normalized root mean square error (NRMSE) and the percentage bias (PBIAS) in Table 6. The performance of all three methods is comparable, but due to its previously named advantages, the L-moment method is considered the superior approach in this setting.

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

375 apparent.

Table 6: Performances of different IPF estimation methods in terms of NRMSE and percentage bias for different flood quantiles. 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-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-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-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-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)*

380 The distribution of the prediction error of the LM-Lmoms model over the different catchment areas and elevations can be found in table 7. The errors are exemplary shown for the 100-year flood. The error distribution is comparable to the MHQ. Again, especially the overestimation for the lowest elevations is striking, as well as the significant underestimation at higher altitudes.

385

Table 7: Average prediction error of the LM-Lmoms model for the HQ100 in percent for the different ranges of area and elevation.

		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. Even though
390   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.

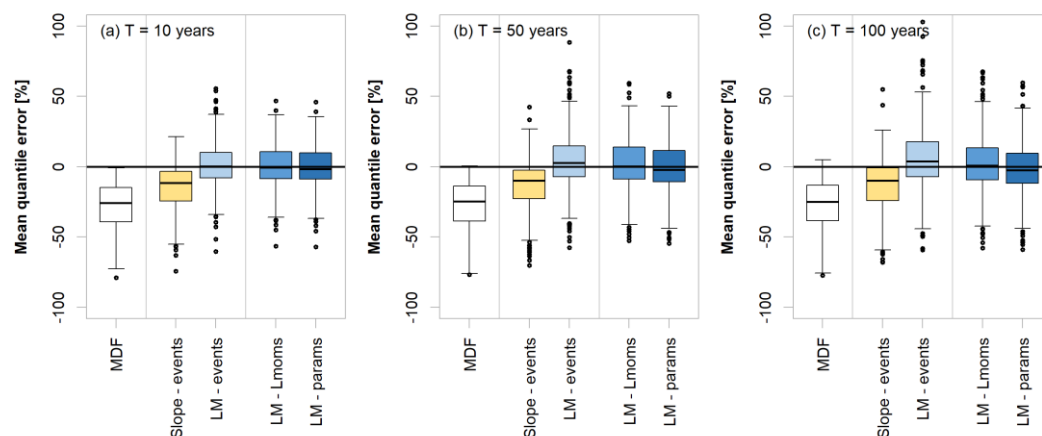



Figure 11: Comparison of performances of various IPF-estimation methods for different seasonally mixed flood quantiles.

The NRMSE and PBIAS values for the mixed approach are shown in table 8. According to these values, the event-correction
395   methods appear to perform best overall. For the smaller catchments (< 500 km²) the LM methods outperform the slope method.

Table 8: Mixed-model performances of different IPF estimation methods in terms of NRMSE and percentage bias for different flood quantiles. The values in parentheses show the performance 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-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-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-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)

4.5 Uncertainty

The results of the bootstrapping procedure used to assess uncertainty are exemplary shown in Fig. 12 for the HQ100 at a single station with a reduced number of 100 permutations. In panel a, the IPF and MDF estimates for each permutation of the annual maximum series are plotted against each other. This shows the bandwidths of both the IPF and MDF estimates as a result of uncertainty in the distribution fitting. Fig. 12 b shows the estimated IPF flood quantiles 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 model estimates. In this example, it becomes obvious that the range in flood quantile estimates due to permutation in the linear models is significantly smaller than the range in estimates due to distribution fitting. This is valid for the majority of stations.

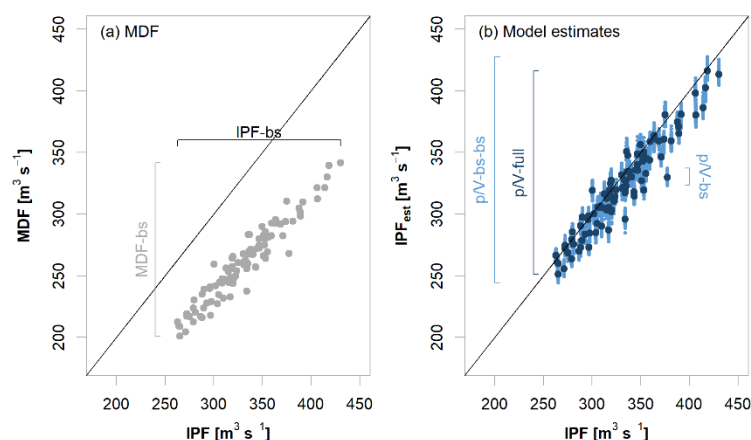



Figure 12: Example of bootstrapping results at a station with 100 permutations. (a) HQ100 from IFP vs. MDF for each permutation of the time series, (b) HQ100 from IPF vs. 100 model estimates per permutation; the dark blue dots represent the full model.

Fig. 13 shows the relative widths of the 95% confidence intervals for all bootstrapping samples. The average widths of the IPF-bs, MDF-bs$_s$ and p/V-full seem to be similar with a larger variability in the IPF sample. The width of the average range of the individual model permutations (p/V-bs-mean) is very small at all stations and therefore contributes little to the overall level of uncertainty (p/V-bs-bs).

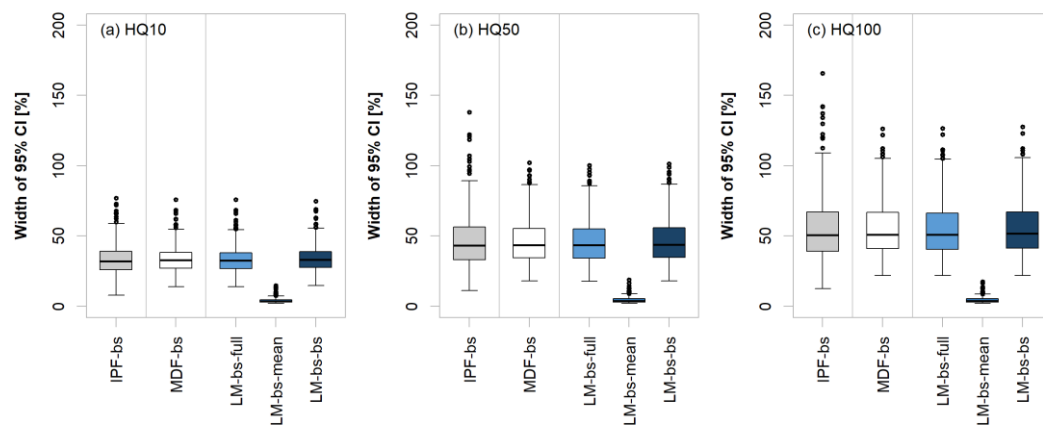



Figure 13: Relative widths of various bootstrap samples for different flood quantiles.

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. Fig. 14 shows the mean deviations from the perturbed IPF quantiles, as well as the lower and upper limits of

420   the 95% confidence intervals of the errors for the 10-, 50- and 100-year flood quantiles. 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 estimates appear to be slightly positively skewed, which is especially noticeable in the 95% confidence interval for the HQ100. At many stations there is a significant overestimation of the true IPF quantile with some of the linear model transpositions. The MDF estimates on the other hand exhibit the expected persistent underestimation.

425

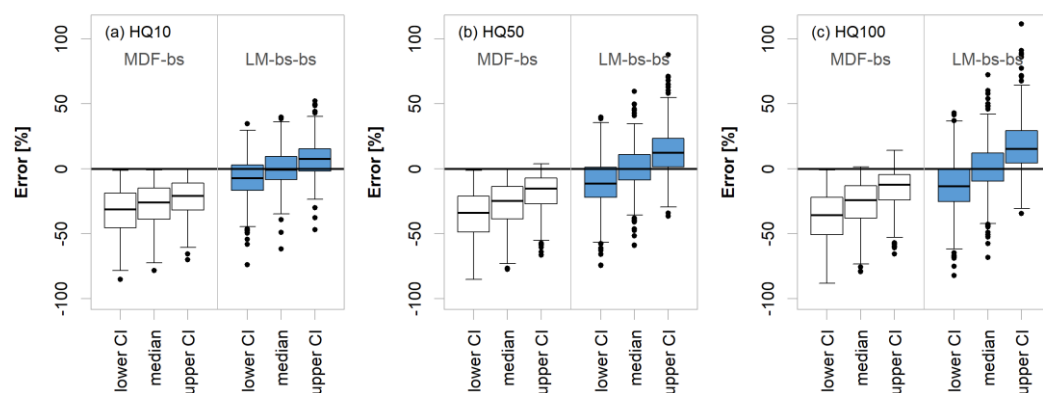



Figure 14: Error distribution of the MDF (left) and LM (right) bootstrap samples for three flood quantiles. Shown are the median errors, as well as the lower and upper limits of the 95% confidence intervals.

Fig. 15 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. It becomes obvious that the agreement between IPF and the LM- model estimates

430   is significantly larger than with the MDF values. This observation suggests that with high probability the LM-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.

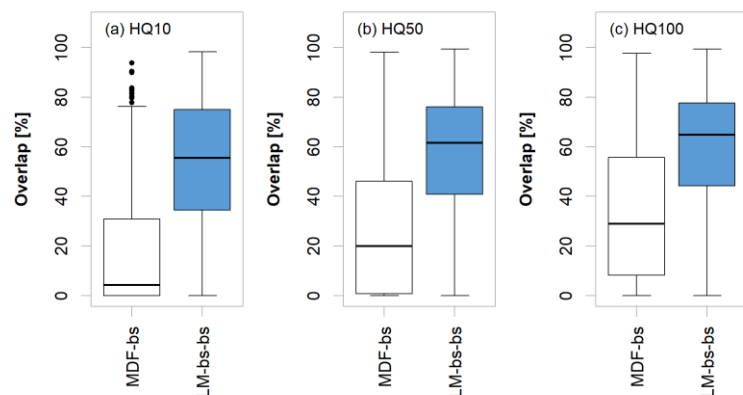



Figure 15: Percentage overlap between the 95% confidence intervals of all IPF bootstrap estimates and MDF and p/V model
435 **bootstrap estimates for three flood quantiles.**

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.

440 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
445 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
450 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
455 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 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
460   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 dataset in order to identify patterns that may be
465   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 deviate noticably from the original model. The first one includes the low-altitude gauges discussed before.
470   Here the overall error is so small that no correction yields better results than correction by the linear model. The second group includes the catchments with areas below 50 km². The errors for these stations appear very scattered and randomly distributed. Comparing the p/V from the daily series with the p/V 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 only functions where unknown instantaneous flood dynamics are roughly
475   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 primarily on a subdaily 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
480   at this scale, the daily flood time scale poses a poor predictor in these catchments.

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 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
485   this study area and should be adjusted where flood types are otherwise distributed. In general, even the rough distinction 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
490   may differ from the optimal IPF distribution. For our data, the GEV proved flexible enough to be a good match in both MDF
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
495   floods. The relationship also appeared to differ between the two types of floods considered here, namely winter and summer
floods. Since summer floods are often caused by short but intense rain events and thus exhibit steep rising and falling limbs,
their subdaily 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
500   instantaneous peak flows when only average daily series are available. The p/V ratio may be used to predict both IPFs of
individual events and instantaneous flood statistics, including mean annual and seasonal maximum flows and flood quantiles.
Due to improper flood event separation, the event-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 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
505   including gauge elevation as additional predictor in some of the models.

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.

510   There are two limitations, where the proposed method should be handled with care: a) at stations with 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, the
LM 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 methods perform better in many larger
515   catchments but outliers my 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.

520

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/

525 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/

530 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 prepared the manuscript
535 with contributions from UH.

Competing interests

The authors declare that they have no conflict of interest.

540 **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:
545   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.

550 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.

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.

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.

585   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
590   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
595   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
600   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.