PHEV! The PHysically-based Extreme Value distribution of river flows

S Basso, G Botter, R Merz, and A Miniussi

1 Department of Catchment Hydrology, Helmholtz Centre for Environmental Research - UFZ, Halle (Saale), Germany
2 Department of Civil, Environmental and Architectural Engineering, University of Padova, Padua, Italy
3 Institute of Geosciences and Geography, Martin-Luther University Halle-Wittenberg, Halle (Saale), Germany
* Author to whom any correspondence should be addressed.
E-mail: stefano.basso@ufz.de

Keywords: flood hazard, runoff dynamics, extreme events, flood frequency, large scale benchmarking

Abstract
Magnitude and frequency are prominent features of river floods informing design of engineering structures, insurance premiums and adaptation strategies. Recent advances yielding a formal characterization of these variables from a joint description of soil moisture and daily runoff dynamics in river basins are here systematized to highlight their chief outcome: the PHysically-based Extreme Value (PHEV) distribution of river flows. This is a physically-based alternative to empirical estimates and purely statistical methods hitherto used to characterize extremes of hydro-meteorological variables. Capabilities of PHEV for predicting flood magnitude and frequency are benchmarked against a standard distribution and the latest statistical approach for extreme estimation, by using both an extensive observational dataset and long synthetic series of streamflow generated for river basins from contrasting hydro-climatic regions. The analyses outline the domain of applicability of PHEV and reveal its fairly unbiased capabilities to estimate flood magnitudes with return periods much longer than the sample size used for calibration in a wide range of case studies. The results also emphasize reduced prediction uncertainty of PHEV for rare floods, notably if the flood magnitude-frequency curve displays an inflection point. These features, arising from the mechanistic understanding embedded in the novel distribution of the largest river flows, are key for a reliable assessment of the actual flooding hazard associated to poorly sampled rare events, especially when lacking long observational records.

1. Introduction
Reliable estimates of magnitude and frequency of river floods are crucial for a wide range of social and economic activities. For instance, they inform design of engineering structures, insurance premiums and adaptation strategies. But despite hundreds of years of efforts, flooding is still the most common natural disaster [1]. Hazard assessment is indeed hampered by processes that might be more variable than suggested by observed records, let alone ongoing global changes [2, 3]. This is a problem, as both purely statistical methods hitherto used for characterizing extremes of hydro-meteorological variables [4–7] and complex hydrological models encoding state-of-the-art knowledge of physical processes [8–11] heavily rely on observations. Tools pairing physical understanding of the mechanisms producing extreme events with easily tractable mathematical descriptions of them and less demanding data requirements are therefore vital to a more accurate flood risk estimation [12, 13]. This letter systematizes and tests advances in the mechanistic-stochastic description of soil moisture and runoff dynamics in river basins to emphasize what might be such a tool: the PHysically-based Extreme Value (PHEV) distribution of river flows. Perks of this physically-grounded alternative to statistical methods hitherto used for characterizing extremes of hydro-meteorological variables are its being simple as a statistical distribution, mechanistic as a fully-fledged model, and stochastic as nature is. Capabilities of PHEV for predicting flood magnitude and frequency are here benchmarked against leading flood hazard assessment methods, a pivotal milestone
to further adoption of this tool among researchers, professionals and policy makers.

2. Methods

2.1. The physically-based extreme value distribution of river flows

The physically-based extreme value distribution of river flows arises from advances in the mathematical description of catchment-scale daily soil moisture and runoff dynamics in river basins [14–16]. These well-established scientific theories represent rainfall as a marked Poisson process with frequency $\lambda_0[T^{-1}]$ and exponentially distributed depths with average $\alpha[L]$. Infiltration of rainfall into the soil determines stochastic increments of the moisture content, whereas evapotranspiration causes loss of soil moisture from the root zone. Losses linearly vary with the soil moisture between the wilting point and a critical upper threshold akin to the water holding capacity of the soil. The exceedance of this threshold triggers runoff pulses occurring with frequency $\lambda < \lambda_0[T^{-1}]$ and whose magnitudes follow an exponential distribution with average $\alpha [L]$. These pulses feed a single catchment hydrologic storage, which is finally drained by the river network as streamflow. A nonlinear storage-discharge relation echoes a hydrological response which varies with the catchment water content and mimics the joint effect of different flow components [17].

The mechanistic-stochastic description of runoff generation processes summarized above yields the following expression for the probability distribution of daily streamflows $q$ [18]:

$$p(q) = C_1 q^{-a} \exp \left( \frac{\lambda q^{1-a}}{K(1-a)} - \frac{q^{2-a}}{\alpha K(2-a)} \right)$$

(1)

which, in addition to $\lambda$ and $\alpha$, depends on the coefficient $K$ and exponent $a$ of a power law function (alike the storage-discharge relation mentioned earlier) used to describe hydrograph's recessions. $C_1$ is a normalization constant.

It also provides a physically-grounded mathematical form for the probability distribution of ordinary peak flows (sensu [19], i.e. local flow peaks occurring as a result of streamflow-producing rainfall events) [20]:

$$p_j(q) = C_2 q^{-a} \exp \left( \frac{\lambda q^{1-a}}{K(1-a)} - \frac{q^{2-a}}{\alpha K(2-a)} \right)$$

(2)

which depends on the same set of parameters. $C_2$ is also a normalization constant.

By further postulating independence of the peak flows, the physically-based probability distribution of flow maxima (i.e. maximum values in a specified time frame) finally emerges as [20]:

$$p_M(q) = \lambda \tau \exp \left( - \lambda \tau D_j(q) \right) p_j(q)$$

(3)

where $\tau[T]$ is the duration in days of the chosen time frame, $p_j(q)$ is the probability distribution of peak flows (equation (2)), and $D_j(q) = \int_q^\infty p_j(q') dq'$ is the exceedance cumulative probability of peak flows.

PHEV is derived from the solution of the master equation for the probability distribution of streamflow in a catchment [18]. As such, the approach resembles what is done when large simulation ensembles of deterministic hydrological models are jointly considered to account for varied hydro-meteorological scenarios and antecedent catchment conditions. PHEV operates in the same way, with the advantage that an analytical solution (equation (3)) is made available thanks to its low-dimensionality.

The description of precipitation and runoff generation mechanisms embedded in PHEV does not encompass all existing rainfall–runoff processes. In particular, intense snow dynamics entailing shift of water volumes across seasons [21] and precipitation mechanisms producing event duration comparable to the time frame of analysis (e.g. monsoonal rainfall, seasonally dry climates) are disregarded. Nonetheless, a large body of literature [17, 22–29] demonstrated the key relevance of the considered mechanisms in a variety of hydro-climatic and physiographic settings.

Equations (1)–(3) form a consistent and physically-grounded set of expressions to characterize the probabilistic structures of daily flows, ordinary peak flows and flow maxima. They describe their magnitudes, likelihoods of occurrence and the related flood hazard based on a set of four physically meaningful parameters ($\alpha, \lambda, a, K$) which embody climatic and landscape attributes of the considered catchment. They also enable the possibility to account for changes of climatic conditions and land cover, as shown by [23, 29, 30]. These probability distributions might therefore constitute a sound alternative to purely statistical distributions commonly used to characterize hydrological variables and their maxima.

2.2. Benchmarking PHEV against leading methods for flood hazard assessment

We benchmark the performance of PHEV against two leading statistical models for flood hazard assessment, the generalized extreme value (GEV) [31] and meta-statistical extreme value (MEV) [32] distributions. The GEV distribution is a standard tool [33, 34] traditionally applied to samples of block maxima under the assumption that the number of events within each block tends towards infinity [35]. The MEV distribution is instead a recently proposed method to estimate extremes from the features of ordinary events, which is currently gaining momentum [36–42]. It relaxes the mentioned assumptions which lie at the heart of the GEV distribution and regards as random variables both the number of independent ordinary events occurring in the considered time interval and the parameters of the distribution used to describe their
Figure 1. (A)–(D) Quantile-quantile plots of normalized flood magnitudes (i.e. seasonal maxima divided by the long term average daily flow, \(<q>\)) estimated for all catchments and seasons by means of Weibull plotting position of observations in the validation samples (horizontal axes) and either (vertical axes) (A) and (D) the Phyically-based (PHEV), (B) the Generalized (GEV) or (C) the Metastatistical (MEV) Extreme Value distributions calibrated on 10 years long samples. Median results among 100 realizations of the calibration samples are plotted. Notice that only quantiles with return periods longer than the length of the calibration samples (i.e. \(T > 10\) years) and until the second longest \(T\) which could be estimated from the validation sample are shown. Warmer colors in panels (A)–(C) indicate higher density of points and mark the most frequent performances of the methods. Dark green markers in panel (D) identify case studies for which a strong relation between recession coefficient and ordinary peak flows (see panel (F)) was detected. The current mechanistic description of the catchment hydrologic response embedded in PHEV does not consider such a relation and prevents it from providing satisfactory performances in these cases, which can be identified ex ante through hydrograph recession analyses. Light green markers in panel (D) instead indicate catchments where such a relation is weak (see panel (E)). In these cases the mechanistic conceptualization of hydrological processes underpinning PHEV is suitable and the PHEV distribution of river flows guarantees satisfactory performances. The insets of panels (B)–(D) represent the median performances for all case studies in the figures belonging to ten different hazard categories (i.e. the deciles of the observed normalized flood quantiles). (E) and (F) Exemplary case studies displaying (E) weak (USGS ID: 03 253 500, spring) and (F) strong (USGS ID: 03 504 000, summer) relations between recession coefficients and ordinary peak flows. (G) Location of gauges from the MOPEX dataset analyzed in this study (grey dots). Red markers display a subset of river basins from varied hydro-climatic regions that have been selected to perform analyses on 1000 years long synthetic time series (see figure 2).
long return periods (say, 100–1000 years) must be extrapolated on the basis of relatively short observational samples. To this aim, we utilized a stochastic rainfall generator and a parsimonious hydrological model (see details in [47]) compliant with the main physical assumptions of PHEV (being MEV and GEV purely statistical models which do not require specific assumptions on the underlying processes) to produce 1000 years long synthetic time series of daily rainfall and streamflow for the select catchments [48]. The series were used as a reference and compared to estimates derived from shorter samples extracted from the 1000 years long data series. In this way we created a controlled experiment which enables analyzing the robustness of the estimation methods to non-ergodic data samples only.

Furthermore, we employed a cross-validation procedure [19, 38] to analyze the ability of the different methods to extrapolate information outside the range covered by the observations. To do so, we first divided the available series of observed (synthetic) data into two complementary parts by randomly selecting from them (through resampling without substitution) 100 (1000) subsets of 5 years. These are the training sets used to calibrate the theoretical distributions, whereas the remaining parts constitute the validation samples. The procedure yields calibration and validation samples with no systematic variability. We used $S = 10$ years for the observations and repeated the process for $S = 10$, 20, 30 and 60 years for the long synthetic data series to explore how the different models perform with various sizes of the calibration sample.

We employed L-moments [49] to fit GEV on the calibration sample of seasonal maxima and to estimate the parameters of a Gamma distribution adopted to describe the whole set of ordinary flow peaks in the MEV approach. Further details on the calibration of GEV and MEV are available in [38]. Parameters of PHEV were instead either directly derived from daily rainfall and streamflow series ($\alpha$, $\lambda$ and $a$) or obtained through maximum likelihood calibration on the sample of seasonal maxima ($K$) [20]. In particular, we computed $\alpha$ as the average amount of precipitation in rainy days, $\lambda$ as the ratio between long-term average streamflow and $\alpha$, and $a$ as the median value of the recession exponents obtained by fitting a power law function to $dq/dt = q$ pairs observed for each hydrograph recession. The maximum likelihood approach here used to calibrate the parameter $K$ is known to provide frail estimates when small calibration samples are available, as in this application. Although aware of the handicap thus imposed to the performance of PHEV, we were compelled to adopt this method due to the current unavailability of more suitable approaches.

We finally used metrics such as the relative error (appendix C) and quantile-quantile plots [50] to evaluate and compare performances of PHEV, GEV and MEV when both observed data and long synthetic series are analyzed.

3. Results and discussion

Figures 1(A)–(C) summarizes the performances of PHEV, GEV and MEV for predicting observed flood quantiles with return periods longer than the length of the calibration samples (i.e. 10 years) in all the case studies. Warmer colors indicate higher density of the point clouds and mark the most frequent performances of the methods. PHEV (panel A) tends to slightly underestimate flood magnitudes in the record (i.e. the red-to-yellow core of the point cloud stands below the 45-degree line). Higher variability of its performances compared to GEV and MEV is manifest, especially in the case of high quantiles for which overestimation often occurs. On the contrary, both GEV (panel B) and MEV (panel C) accurately estimate lower flood magnitudes (i.e. the point clouds are centered on the 45-degree lines for low quantiles) but tend to systematically underestimate the larger floods.

In the case of MEV, this behavior is likely due to using a Gamma distribution to describe ordinary events, in agreement with [51]. In fact, such a light tailed distribution might underestimate ordinary peaks characterized by low probability, eventually resulting in underestimation of maxima with long return periods. Although the optimization of the MEV methodology is outside the scope of this work, we recognize that the choice of a site-specific distribution of ordinary peaks might reduce underestimation issues in the MEV approach [38].

As for PHEV, previous studies scantily reported about the possible existence of a relation between recession coefficient $K$ and peak flow and its capability to interfere with the characterization of daily flow distributions provided by the mechanistic-stochastic model underlying it [17, 52]. Was there such a relation, the parameters describing the hydrologic behavior of the catchment would vary along with the peak flow magnitude, thus precluding the use of a single effective $K$ to portray the catchment hydrologic response and the resulting magnitude of floods. At the same time, such a relation would also suggest caution about the dependability of purely statistical flood estimates beyond the range of recorded magnitudes, as ostensibly good performances of these methods might indicate their higher flexibility to represent conditions which occurred in the observed time frame rather than robustness in estimating uncharted flood magnitudes [53].

To investigate whether this may be the reason for the higher variability of PHEV performances displayed in panel A, we computed the recession coefficient $K$ by fitting a power law function with exponent set equal to $a$ (i.e. the median exponent across all recessions in a case study) to $dq/dt = q$ pairs observed for each hydrograph recession [54], and investigated
the possible existence of a relation between $K$ and peak flow in the MOPEX basins. Panels E and F display examples of catchments where we either did not or identified such a relation. We then classified all the case studies into two groups based on the absolute value of the slope of their $K$—peak flow relation, rendering them in panel D with either light or dark green markers if the relation was respectively weak (i.e. slope $\leq 0.4$) or strong (i.e. slope $> 0.4$). Notice that choosing alternative thresholds around this value does not substantially impact the key results of the study.

The performance of PHEV remarkably improves in case studies exhibiting weak relations between $K$ and peak flow (i.e. the majority of light green markers lie just below the 45-degree line in panel D, whereas dark green markers mainly stand above it). The Two-sample Kolmogorov-Smirnov test for the probability distribution of the relative estimation errors confirms that the two groups are significantly different ($p$-value $<0.01$). The larger variability of performances previously exhibited by PHEV (panel A) substantially diminishes for the set of cases featuring weak $K$-peak flow relations (the interdecile and interquartile ranges of the relative error decrease by 25% and 48%). Although PHEV tends to underestimate flood magnitudes in the range of observed quantiles, the bias of the estimate seems to be relatively stable for increasing quantiles compared to GEV and MEV.

The proposed analysis of hydrograph recession characteristics is thus an effective method to distinguish ex ante in what conditions PHEV shall be applied, namely in all cases (268 out of 483 in our dataset, which will be referred in the following analyses) when the recession coefficient $K$ is fairly constant across recedions and thereby agrees with the mechanistic description of runoff generation processes underlying PHEV. Possible causes of the variability of $K$ across events and basins are a much debated topic in recent years [55–62]. The geomorphological theory of recession flows [63] suggests that enhanced variability of $K$ is linked to the existence of manifold hydrologic storages which are differentially activated across events, either because of their uneven distribution along river networks or as a result of variable degrees of saturation preceding the events [64, 65]. This behavior is conceivably related to the existence of assorted triggers of runoff events and floods [66].

The contrasting behaviors of PHEV, GEV and MEV for the prediction of the highest quantiles are further analyzed in the insets of figures 1(B)–(D), which display the median performances among case studies belonging to ten different hazard categories (i.e. the deciles of the observed flood quantiles). For PHEV (inset of panel D, light green hue) markers lie just below and parallel to the 45-degree line, indicating a slight tendency to underestimation that is independent of the flood magnitude. The markers instead diverge from the 45-degree lines in the insets of panels B and C (i.e. for GEV and MEV), indicating that the estimation errors associated to these distributions are low for small floods but increase with increasing quantiles.

To better visualize these traits and their implications for estimating magnitudes of events with long return periods, we plotted the median relative errors for each hazard class as a function of the observed quantiles (figure 2(A)). We then extrapolated the performances of the different extreme distributions to higher quantiles by fitting regression lines using only half the observed errors, randomly extracted a hundred times from the whole set of them by means of resampling without substitution. This procedure yields an evaluation of the uncertainty of the extrapolations (shaded areas in panel A) and concurrently allows for inspecting their robustness against the observations themselves. The performance of PHEV (light green shaded area) is lesser affected by increasing flood quantiles, whereas it ceaselessly deteriorate with the magnitude of floods for GEV and MEV (red and blue shaded areas). This finding is especially relevant for the estimation of high quantiles associated to very long return periods. In this situation, errors of PHEV shall remain fairly limited while those of GEV and MEV are expected to further increase.

We provide a stricter assessment of the performances of the three extreme value distributions for very high quantiles in the remainder of figure 2. There we summarize the analyses performed on 1000 years long synthetic time series of rainfall and streamflow for eight select catchments from different US hydroclimatic regions which exhibit weak relations between $K$ and peak flow. Figure 2(B) displays the relative errors of PHEV, GEV and MEV for a calibration sample of 20 years and a set of large return periods commonly used in the practice. A relative error equal to zero indicates a perfect match between quantiles predicted by the model and those estimated from the validation sample. Although PHEV is affected by a slight underestimation (the median of the box plot is always below the zero line), it shows the smallest interquartile range of the relative error, which is comparable to the one of MEV and lower than GEV, for all considered return periods. The amount of outliers is rather constant for MEV while it increases with the return period for both PHEV and GEV. In particular, large numbers and extent of outliers among GEV estimates appear for return periods as low as 50 years. Conversely, the capability of PHEV to provide reliable estimates does not deteriorate as strongly for large return periods.

A similar pattern is also highlighted by the relative errors between estimated and observed maxima, plotted for increasing values of the ratio between return period $T$ and calibration sample size $S$ (panel C). Here, results for every $S = 10, 20, 30$ and 60 years are pooled together to give a compendium of the
analyses. Negative values of the median relative error for MEV reflect its tendency to underestimate high quantiles in this dataset when a Gamma distribution is used to describe the ordinary peaks. Notwithstanding comparable performances of GEV and PHEV for what concerns median errors, the latter exhibits lower error variance than the former in all analyzed ranges of $T/S$. Also in this case the performance of PHEV is rather constant with the return period, thus confirming the results of figure 2(A).

The capabilities of the three distributions and an empirical estimation method to reproduce the reference flood magnitude-frequency curve (grey dots) when a limited calibration sample ($S = 20$ years) is available are exemplified in panels D–I. We display results for these two case studies to emphasize how the shape of the flood frequency curve might affect the performance of the methods. Namely, we chose one case (panels D–F) where the flood frequency curve is rather predictable (i.e. the streamflow magnitude smoothly rises with increasing return period) and one case (panels G–I) characterized by the opposite behavior (i.e. the flood magnitude markedly increases above a certain return period).

The black lines and shaded areas span the breadth of empirical, PHEV (light green), GEV (red) and MEV (blue) estimates. The empirical method reveals its fickleness, as its estimates highly depend on the specific set of flood events sampled in the time frame $S$ even for short return periods (i.e. the uncertainty of the maximum associated to an assigned frequency is considerable for return periods as low as ten years). The use of a theoretical extreme value distribution reduces this variability (i.e. the shaded areas span smaller ranges across the y-axes at short return periods). However, GEV estimates (panels E and H) are as variable for just slightly larger return periods (i.e. 50 years). Conversely, the uncertainty of MEV is limited (i.e. blue shades in panels F and I are narrow) for the whole range of investigated return periods,

Figure 2. (A) Relative flood estimation errors for increasing quantiles resulting from the use of PHEV (light green), GEV (red) and MEV (blue) distributions. The left-hand side of the panel encompasses the range of quantiles for which errors associated to the use of different extreme value distributions could be evaluated from observations. Extrapolations of these errors for higher quantiles (i.e. less frequent floods) are displayed as shaded areas on the right-hand side of the plot. (B) Relative errors of PHEV, GEV and MEV for all select case studies and a set of large return periods commonly used in the practice when 20 years calibration samples are used. (C) Relative errors of PHEV, GEV and MEV as a function of the ratio between return period $T$ and calibration sample size $S$. Only return periods longer than the sample size used for calibration ($S = 10, 20, 30$ and $60$ years) are considered to compute the performance. Solid lines represent the median errors, shaded areas enclose the 25th–75th percentiles of the performances. (D)–(I) Normalized (i.e. seasonal maximum divided by the long term average daily flow, $<q>$) flood magnitude-frequency curves estimated by means of synthetic data series and the three extreme value distributions. Results for USGS gauge 03 075 500 in spring (panels D–F) and USGS gauge 05 471 500 in summer (panels G–I) are here displayed as exemplary case studies. Grey dots present the estimates obtained by means of Weibull plotting position of 1000 years long synthetic time series generated for the considered basins. Black solid lines and shaded areas enclose the ranges of normalized flood magnitude-frequency curves obtained when shorter data series of 20 years length are used for estimation either by means of Weibull plotting position of the shorter series (black solid lines) or the support of (D) and (G) the PHysically-based (PHEV, light green areas), (E) and (H) the Generalized (GEV, red areas) and (F) and (I) the Metastatistical (MEV, blue areas) Extreme Value distributions. The plots highlight superior reliability of PHEV for long return periods, especially when discontinuities in the flood magnitude-frequency curve exist.
but the method tends to underestimate magnitudes of rare events with return periods longer than a few hundred years, in particular when the flood frequency curve is characterized by the presence of an inflection point (panel I). PHEV (panels D and G) provides dependable estimates of flood magnitude-frequency curves which are less affected by the specific set of flood events available in the sample (i.e. the amplitude of the light green shade is as well limited). Albeit more uncertain than MEV, the pattern of PHEV estimates closely resembles that of the reference flood magnitude-frequency curve (grey dots) for return periods above a hundred years, especially when the curve trends upward. Remarkably, knowledge of the magnitude of events with return periods substantially longer than the available data series is required at most in the practice.

Notwithstanding the highlighted perks of adopting a mechanistic-stochastic approach to estimate magnitudes and frequencies of floods from daily streamflow dynamics, work remains to be done. Stable calibration methods, such as L-moments, shall be developed for PHEV as done in the past decades for standard probability distributions now routinely used in extreme value analyses. The search for approaches to estimate the parameter $K$ without resorting to calibration on observed maxima (thus achieving entirely calibration-free predictions of flood statistics) shall also continue. Given the apparent dependence of $K$ on the wetness conditions of river basins [56, 57], satellite-derived water storage and soil moisture products are promising allies in this regard, as recently shown by [67] and [68].

Using the physically-based distribution of peak flows underpinning PHEV to describe ordinary events in the MEV framework constitutes another promising research avenue whereby capitalizing on the assets of both methods, namely the mechanistic description of runoff generation processes embedded in PHEV and the possibility to explicitly account for the inter-annual variability of parameters provided by MEV. The mechanistic-stochastic conceptualization of runoff generation processes underlying PHEV might as well be further advanced to explicitly include the variability of recession coefficients in its mathematical expression, whose lack currently constitutes the main limitation of the method. Seminal works in this sense [17, 69] demonstrated visible improvements of high flow estimates when this variability is embedded in the mechanistic-stochastic description of daily flows. Addressing these aspects would advance the PHEV distribution of river flows towards ripeness for a standard application in the practice.

4. Conclusions

The PHEV distribution of river flows emerges from a mechanistic-stochastic description of basin-scale soil moisture dynamics and runoff generation processes, which provides consistent analytical expressions of the probability distributions of daily flows, peak flows and flow maxima. It constitutes a physically-based alternative to purely statistical methods hitherto used to characterize extremes of hydrological variables, whose parameters are all but one measurable from daily rainfall and streamflow records. A novel method relying on flow recession analyses and the identification of an inverse relation between recession coefficients and peak flows allows for the domain of applicability of PHEV to be defined a priori. In the absence of such relation PHEV guarantees remarkable performances which are comparable to those of two leading statistical models for flood hazard assessment. In particular, PHEV estimates entail lesser increasing bias with larger flood quantiles than it occurs for the benchmark approaches, and lower uncertainty. These features ensure more reliable appraisal of the magnitude of rare extreme floods from relatively short data series, an especially valuable achievement for the community of professionals involved in environmental hazards assessment. Our findings finally hint at the possibility to leverage the reliability of PHEV estimates for very long return periods to pinpoint inflection points of flood magnitude-frequency curves, whose occurrence entail abrupt increases of the magnitude of high flows, amplified hydrological hazard and a propensity of rivers to generate extreme floods.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: Hydro-climatic time series from the MOPEX data set are available at https://hydrology.nws.noaa.gov/pub/gcip/mopex/US_Data. Synthetic hydro-climatic time series for the select catchments used in this study to assess performances for very long return periods are available at: www.hydroshare.org/resource/abd449fbced146eb8d0ec8ad95754f1e.

Acknowledgments

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project Number 421396820 ‘Propensity of rivers to extreme floods: climate-landscape controls and early detection (PREDICTED)’ and Research Group FOR 2416 ‘Space-Time Dynamics of Extreme Floods (SPATE)’. The financial support of the Helmholtz Centre for Environmental Research - UFZ is as well acknowledged. We thank Andrea Domin for her contribution for producing the long synthetic time series used in this study. We finally thank two anonymous reviewers for their comments to the manuscript.
Appendix A. The generalized and MEV distributions

The extreme value theory states that if a random variable $x$ follows a distribution $F(x)$, the distribution of the maximum of $n$ independent and identically distributed random variables ($Y_n = \max(X_1, \ldots, X_n)$) can be written as $H_n(x) = F(x)^n$. In the traditional framework the assumption that $n \to \infty$ gives $H_n(x) \to H(x)$, where $H(x)$ is one of three limiting distributions [31, 70] unified in the generalized extreme value (GEV) distribution [35, 71]. The non-exceedance cumulative distribution function of GEV reads as:

$$H_{\text{GEV}}(x) = \exp \left( - \left(1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}} \right) \quad \text{(A.1)}$$

with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and shape parameter $\xi \in \mathbb{R}$.

The MEV framework [32] relaxes the assumption of $n \to \infty$ and considers as random variables both the parameters $\theta$ of the ordinary distributions $F(x; \theta)$ and the number of events $n$ occurring in each considered time frame (e.g. a year or a season). Following [38], we use here the Gamma distribution to describe the ordinary events, and neglect the inter-annual variability of the parameters to reduce their estimation uncertainty. We instead consider as time-varying the number of events during each time frame (note that the subscript $j$ appears in the exponent $n$ in equation (A.2)). The resulting expression of the non-exceedance cumulative distribution function of MEV [19, 32] is:

$$H_{\text{MEV}}(x) = \frac{1}{S} \sum_{j=1}^{S} \left[ \frac{\gamma(\alpha, \frac{x}{\beta} \xi)}{\beta^\alpha \Gamma(\alpha)} \right] \quad \text{(A.2)}$$

where $S$ is the length of the calibration sample, $\alpha > 0$ and $\beta > 0$ are the shape and scale parameters of a Gamma distribution, and $\gamma(\cdot)$ and $\Gamma(\cdot)$ are respectively the lower incomplete and the complete Gamma functions.

Appendix B. Subset of river basins from different US hydro-climatic regions

Information concerning the eight select catchments from different US hydro-climatic regions, which are used in this study to assess performances of the theoretical extreme distributions for very long return periods ($O(10^3)$), are available in table B1.

Appendix C. Error metric

In this study we use the relative error [72] as a metric to evaluate and compare the performances of the considered extreme value approaches. The relative error $\varepsilon$ is a non-dimensional error between observed ($q_{\text{obs}}$) and estimated ($q_{\text{est}}$) maxima computed for the same return period $T$:

$$\varepsilon(T) = \frac{q_{\text{obs}}(T) - q_{\text{est}}(T)}{q_{\text{obs}}(T)} \quad \text{(C.1)}$$

A relative error equal to zero indicates a perfect match between estimated and observed quantities. Negative values of $\varepsilon$ indicate underestimation of the model (i.e. the model predicts maxima with lower magnitudes than the observed ones), whereas positive values denote overestimation (i.e. the model predicts maxima with larger magnitudes than the observed ones). In this study we computed the relative error in a predictive fashion, i.e. by only considering in the computation quantities corresponding to return periods $T$ longer than the sample size $S$ used to calibrate the parameters of the distributions.

ORCID iDs

S Basso  https://orcid.org/0000-0003-3551-9155
G Botter  https://orcid.org/0000-0003-0576-8847
R Merz  https://orcid.org/0000-0002-3698-2339
A Miuissi  https://orcid.org/0000-0002-9765-3985

References

[1] Wallemacq P and House R 2018 Economic Losses, Poverty & Disasters 1998-2017 United Nations Office for Disaster Risk Reduction (UNDRR) and Centre for Research on the Epidemiology of Disasters (CRED)
[2] The Economist 2017 The likelihood of floods is changing with the climate—The Economist
[3] Bevere L, Gloor M and Sobel A 2020 Natural catastrophes in times of economic accumulation and climate change Swiss Re Institute
[4] Katz R W, Parlange M B and Naveau P 2002 Statistics of extremes in climatology and hydrology Adv. Water Resour. 25 328–341
[5] Morrison J E and Smith J A 2002 Stochastic modeling of flood peaks using the generalized extreme value distribution Water Resour. Res. 38 1305
[6] England F J et al 2019 Guidelines for determining flood flow frequency—Bulletin 17C US Geological Survey
[7] Metzger A, Marra F, Smith J A and Morin E 2020 Flood frequency estimation and uncertainty in arid/semi-arid regions J. Hydrol. 590 125254
[8] Maxwell R M and Miller N L 2005 Development of a coupled land surface and groundwater model J. Hydrometeor. 6 233–47
[9] Van Der Knijff J M, Younis J and De Roo A P J 2010 LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation Int. J. Geogr. Inf. Sci. 24 189–212
[10] Hirpa F A et al 2018 Calibration of the global flood awareness system (GloFAS) using daily streamflow data J. Hydrol. 566 591–606
[11] Kuffour B N O, Engdahl N B, Woodward C S, Condon I E, Kollet S and Maxwell R M 2020 Simulating coupled surface–subsurface flows with ParFlow v3.5.0: capabilities, applications and ongoing development of an open-source, massively parallel, integrated hydrologic model Geosci. Model Dev. 13 1373–97
[12] Klemes V 1989 The improbable probabilities of extreme floods and droughts Proc. of the World Meteorological Organization Conf. Held in Geneva (London: World Meteorological Organization) pp 43–51
[13] Barth N, Villarini G and White K 2019 Accounting for mixed populations in flood frequency analysis: bulletin 17C perspective J. Hydrol. Eng. 24 04019002
[14] Laio F, Porporato A, Ridolfi L and Rodriguez-Iturbe I 2001 Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: II. Probabilistic soil moisture dynamics Adv. Water Resour. 24 707–23
[15] Porporato A, Daly E and Rodriguez-Iturbe I 2004 Soil water balance and ecosystem response to climate change Am Nat. 164 625–32
[16] Botter G, Porporato A, Rodriguez-Iturbe I and Rinaldo A 2007 Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: slow, learning-prone components of the hydrologic response Water Resour. Res. 43 W02417
[17] Basso S, Schirmer M and Botter G 2015 On the emergence of heavy-tailed streamflow distributions Adv. Water Resour. 82 98–105
[18] Botter G, Porporato A, Rodriguez-Iturbe I and Rinaldo A 2009 Nonlinear storage-discharge relations and catchment streamflow regime Water Resour. Res. 45 W10427
[19] Zorzetto E, Botter G and Marani M 2016 On the emergence of rainfall extremes from ordinary events Geophys. Res. Lett. 43 8076–82
[20] Basso S, Schirmer M and Botter G 2016 A physically based analytical model of flood frequency curves Geophys. Res. Lett. 43 9070–6
[21] Schaefli B, Rinaldo A and Botter G 2013 Analytic probability distributions for snow-dominated streamflow Water Resour. Res. 49 2701–13
[22] Botter G, Peratoner F, Porporato A, Rodriguez-Iturbe I and Rinaldo A 2007 Signatures of large-scale soil moisture dynamics on streamflow statistics across US climate regimes Water Resour. Res. 43 W11413
[23] Botter G, Basso S, Porporato A, Rodriguez-Iturbe I and Rinaldo A 2010 Natural streamflow regime alterations: damming of the Piave river basin (Italy) Water Resour. Res. 46 W06522
[24] Cea A, Botter G, Bertuzzo E, Porporato A, Rodriguez-Iturbe I and Rinaldo A 2010 Comparative study of ecohydrological streamflow probability distributions Water Resour. Res. 46 W09502
[25] Mejia A, Daly E, Rossel F, Jovanovic T and Gironás J 2014 A stochastic model of streamflow for urbanized basins Water Resour. Res. 50 1984–2001
[26] Douloutaryi B, Betterle A, Basso S, Biswas B, Schirmer M and Botter G 2015 Predicting streamflow distributions and flow duration curves from landscape and climate Adv. Water Resour. 83 285–98
[27] Santos A C, Portela M M, Rinaldo A and Schaeffl B 2018 Analytical flow duration curves for summer streamflow in Switzerland Hydrol. Earth Syst. Sci. 22 2377–81
[28] Araújo R, Toyoda Y and Kazama S 2021 Runoff recession features in an analytical probabilistic streamflow model J. Hydrol. 597 125745
[29] Müller M F, Roche K R and Dralle D N 2021 Catchment processes can amplify the effect of increasing rainfall variability Environ. Res. Lett. 16 084032
[30] Botter G, Basso S, Rodriguez-Iturbe I and Rinaldo A 2013 Resilience of river flow regimes Proc. Natl Acad. Sci. USA 110 12925–30
[31] Gneedenko B 1943 Sur la distribution limite du terme d’une série aléatoire Ann. Math. 44 423–53
[32] Marani M and Ignaccolo M 2015 A metastatistical approach to rain extremes Adv. Water Resour. 79 121–6
[33] Kjeldsen T R, Jones D A and Bayliss A C 2008 Improving the FEH Statistical Procedures for Flood Frequency Estimation (Bristol: European Environment Agency (EEA))
[34] DWA 2012 Ermittlung von Hochwasserwahrscheinlichkeiten Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall (DWA)
[35] Coles S 2001 An Introduction to Statistical Modeling of Extreme Values (Springer Series in Statistics) (London: Springer)
[36] Marra F, Nikolopouloos E I, Anagnostou E N and Morin E 2018 Metastatistical extreme value analysis of hourly rainfall from short records: estimation of high quantiles and impact of measurement errors Adv. Water Resour. 117 27–39
[37] Schellander H, Lieb A and Hell T 2019 Error structure of metastatistical and generalized extreme value distributions for modeling extreme rainfall in Austria Earth Space Sci. 6 1616–32
[38] Minnissi A, Marani M and Villarini G 2020 Metastatistical extreme value distribution applied to floods across the continental United States Adv. Water Resour. 136 103498
[39] Hosseini S R, Scacioni M and Marani M 2020 Extreme Atlantic hurricane probability of occurrence through the metastatistical extreme value distribution Geophys. Res. Lett. 47 2019GL086138
[40] Zorzetto E and Marani M 2020 Extreme value metastatistical analysis of remotely sensed rainfall in ungauged areas: spatial downscaling and error modelling Adv. Water Resour. 135 103483
[41] Minnissi A and Marani M 2020 Estimation of daily rainfall extremes through the metastatistical extreme value distribution: uncertainty minimization and implications for trend detection Water Resour. Res. 56 e2019WR026535
[42] Marra F, Borga M and Morin E 2020 A unified framework for extreme subdaily precipitation frequency analyses based on ordinary events Geophys. Res. Lett. 47 e2020GL090209
[43] Schaefer J, Duan Q, Andrássy V, Francks S, Hall A and Levesley G 2006 The model parameter estimation experiment (MOPEX) J. Hydrol. 320 1–2
[44] Wang D and Hejazi M 2011 Quantifying the relative contribution of the climate and direct human impacts on mean annual streamflow in the contiguous United States Water Resour. Res. 47 W00J12
[45] Durrans S R, Eiffe M A, Thomas Jr W O and Goranflo H M 2003 Joint seasonal/annual flood frequency analysis J. Hydrol. Eng. 8 181–9
[46] Merz B et al 2021 Causes, impacts and patterns of disastrous river floods Nat. Rev. Earth Environ. 2 592–609
[47] Garbin S, Celegon Alessi E, Fanton P and Botter G 2019 Hydrological controls on river network connectivity R. Soc. Open Sci. 6 181428
[48] Miniussi A, Domin A and Basso S 2021 Synthetic time series for the manuscript "PHEV! The Phvysically-based Extreme Value distribution of river flows" by S.Basso, G.Botter, R.Merz and A.Miniussi HydroShare
[49] Hosking J R M 1990 L-Moments: analysis and estimation of distributions using linear combinations of order statistics J. R. Stat. Soc. B 52 105–24
[50] Barnett V 1973 Probability plotting methods and order statistics Appl. Stat. 24 95–108
[51] Slater L and Villarini G 2017 Evaluating the drivers of seasonal streamflow in the US Midwest Water 9 2016EA000557
[52] Ghosh D K, Wang D and Zhu T 2016 On the transition of base flow recession from early stage to late stage Adv. Water Resour. 88 8–13
[53] Kirchner J W 2006 Getting the right answers for the right reasons: linking measurements, analyses and models to advance the science of hydrology Water Resour. Res. 42 W03S04
[54] Biswal B 2021 Decorrelation is not dissociation: there is no means to entirely decouple the Brutsaert-Nieber parameters in streamflow recession analysis Adv. Water Resour. 147 103822
[55] Rupp D E, Schmidt J, Woods R A and Bidwell V J 2009 Analytical assessment and parameter estimation of a low-dimensional groundwater model J. Hydrol. 377 143–54
[56] Shaw S B, McHardy T M and Riha S J 2013 Evaluating the influence of watershed moisture storage on variations in base flow recession rates during prolonged rain-free periods in medium-sized catchments in New York and Illinois, USA Water Resour. Res. 49 6022–8
[57] Bart R and Hope A 2014 Inter-seasonal variability in baseflow recession rates: the role of aquifer antecedent storage in central California watersheds J. Hydrol. 519 205–13
[58] Biswal B and Nagesh Kumar D 2015 Estimation of drainable storage—A geomorphological approach Adv. Water Resour. 77 37–43
[59] Patnaik S, Biswal B, Kumar D N and Sivakumar B 2015 Effect of catchment characteristics on the relationship between past discharge and the power law recession coefficient J. Hydrol. 528 321–8
[60] Dralle D N, Hahn W J, Rempe D M, Karst N J, Thompson S E and Dietrich W E 2018 Quantification of the seasonal hillslope water storage that does not drive streamflow Hydro. Process. 32 1978–92
[61] Tashie A, Scaife C I and Band L E 2019 Transpiration and subsurface controls of streamflow recession characteristics Hydro. Process. 33 2561–75
[62] Tashie A, Pavelsky T and Emanuel R E 2020 Spatial and temporal patterns in baseflow recession in the continental United States Water Resour. Res. 56 e2019WR026425
[63] Biswal B and Marani M 2010 Geomorphological origin of recession curves Geophys. Res. Lett. 37 L24403
[64] Biswal B and Marani M 2014 Universal recession curves and their geomorphological interpretation Adv. Water Resour. 65 34–42
[65] Mutchner R et al 2013 Geomorphic signatures on Brutsaert base flow recession analysis Water Resour. Res. 49 5462–72
[66] Tarasova L, Basso S and Merz R 2020 Transformation of generation processes from small runoff events to large floods Geophys. Res. Lett. 47 e2020GL090347
[67] Sharma D, Patnaik S, Biswal B and Reager J T 2020 Characterization of basin-scale dynamic storage–discharge relationship using daily GRACE-based storage anomaly data Geosciences 10 404
[68] Basso S, Ghazanchaei Z and Tarasova L 2021 Characterizing hydrograph recessions from satellite-derived soil moisture Sci. Total Environ. 756 143469
[69] Botter G 2010 Stochastic recession rates and the probabilistic structure of stream flows Water Resour. Res. 46 W12527
[70] Fisher R A and Tippett L H C 1928 Limiting forms of the frequency distribution of the largest or smallest member of a sample Math. Proc. Cambridge Phil. Soc. 24 180–90
[71] Von Mises R 1936 La distribution de la plus grande de n valeurs Rev. Math. Union Interbalcanique I 141–60
[72] Abramowitz M and Stegun I A E 1972 Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables (New York: Dover)