Received: 10 August 2022 | Revised: 3 January 2023 | Accepted: 19 January 2023

DOI: 10.1111/jfr3.12891

ORIGINAL ARTICLE

CIWEM Chartered Institution of Water and Environmental Management | Journal of Flood Risk Management | WILEY

Toward an improved estimation of flood frequency statistics from simulated flows

Lanxin Hu[1] | **Efthymios I. Nikolopoulos**[2] | **Francesco Marra**[3,4] | **Emmanouil N. Anagnostou**[1]

[1]Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut, USA

[2]Department of Civil and Environmental Engineering, Rutgers University, New Brunswick, New Jersey, USA

[3]Department of Geosciences, University of Padova, Padova, Italy

[4]Institute of Atmospheric Sciences and Climate, National Research Council (CNR-ISAC), Bologna, Italy

Correspondence
Emmanouil N. Anagnostou, Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT, USA.
Email: emmanouil.anagnostou@uconn.edu

Funding information
Eversource Energy Center at the University of Connecticut; CARIPARO Foundation

Abstract

The estimation of extreme flood frequency for ungauged or poorly gauged catchments is a longstanding problem of great practical importance. Simulated streamflow derived from distributed hydrological models can be used to address this issue, but their representation of extreme flood peaks is often prone to large biases. This study evaluates the potential of a nonasymptotic statistical approach able to consider all the independent flood peaks instead of extremes only, the Simplified Metastatistical Extreme Value (SMEV), for the estimation of extreme flood frequency from time series of simulated streamflow. We examined 28 years of simulated daily streamflow across the contiguous United States and compared SMEV to traditional statistical models based on annual maxima. Our results suggest that when its assumptions are met SMEV can moderate the impact of hydrological model biases in the quantification of extreme flood frequency. SMEV exhibits a lower relative difference between quantiles derived from observations and simulations for all return periods and forcing dataset. Quantiles estimated from simulated streamflow time series (28-year records) using SMEV are usually in better agreement with the estimates based on 70-year-long observations. Geographical variations in the results of SMEV are noticed, with a better performance of SMEV in the east and west coasts (California, New England, and Mid-Atlantic) and in the southwestern regions (Texas-Gulf). These results indicate that the potential of SMEV for flood frequency analyses in ungauged and poorly gauged basins deserves further investigations.

KEYWORDS

flood frequency analysis, simulated streamflow, SMEV, uncertainty

1 | INTRODUCTION

Quantifying extreme flood peaks with low yearly exceedance probability is crucial for managing flood risk and designing hydraulic structures (Apel et al., 2008; Stedinger, 1983). Flood frequency analysis (FFA) usually addresses this need by means of statistical distributions: extremes (either annual maxima or the values exceeding

a high threshold) are extracted from observations, an analytical function describing their distribution is identified and used to extrapolate information about rare, and potentially still unobserved, flood peaks—such as those expected to occur only once in 100 years (1% yearly exceedance probability) (Lam et al., 2017; St. George & Mudelsee, 2019). For FFA to be successful, long and accurate observational records are required (Franks & Kuczera, 2002; Hu et al., 2020; Metzger et al., 2020; Ruiz-Bellet et al., 2017; Srinivas et al., 2008): long records are needed to have enough data points for the identification of the statistical distribution and the estimation of its parameters; accuracy of the extremes is needed to avoid biases in the derived quantities. When records with these characteristics are not available, such as in the case of ungauged or poorly gauged catchments, FFA becomes an even more challenging task to address.

Several approaches have been proposed to address the issue. For example, regionalization methods are based on the idea that some hydrological and thus statistical properties can be transferred from nearby gauged catchments (Pallard et al., 2009; Rahman et al., 2014; Smith et al., 2015). Commonly used techniques include the rational method, the index flood method, and quantile regression analysis (Griffis & Stedinger, 2007; Haddad et al., 2012; Sharifi Garmdareh et al., 2018). These methods rely on the identification of homogeneous regions, and on the assumption that catchments within such regions share some of the statistical properties of extreme floods. However, this often neglects key processes dominating the local flood generation. Moreover, it is very difficult to extend these methods to broader scales (Ayalew & Krajewski, 2017). As such, FFA in ungauged (or poorly gauged) basins remains among the grand challenges of the hydrological community (Blöschl et al., 2019).

Distributed and process-based hydrological models (Beven, 1986, 1987), as well as recent machine learning-based hydrological models (Kratzert et al., 2019; Rasheed et al., 2022; Yang et al., 2020), can be implemented to simulate discharge at ungauged locations of a river network. Continuous simulation approaches based on hydrological models were first demonstrated by Beven (1986, 1987) and applied widely in FFA on the basis of design event models (Ahn et al., 2014; Ball et al., 2016; Bradley & Potter, 1992; Passerotti et al., 2020; Saghafian et al., 2014; Yang et al., 2004). These early studies used hydrological models with stochastic rainfall generators. The recent availability of more robust, long-term and high-resolution reanalysis datasets, which provide physically consistent forcing, enhanced considerably the ability of model simulations to represent the hydrological variability of real systems (Kan et al., 2017). Several recent studies showed that these simulations can provide useful data to reconstruct flood peaks (AghaKouchak et al., 2011; Bell et al., 2007; Chilkoti et al., 2017; Kay et al., 2009; Moretti & Montanari, 2008; Sun et al., 2015; Yang et al., 2004). However, the simulated time series can be associated with considerable systematic and random errors related to the meteorological forcing, the land surface parameterization, and the model structure (Ehsan Bhuiyan et al., 2019; Renard et al., 2006). Moreover, it is difficult to calibrate the models to well reproduce extremes, because by definition only few extremes are available in the data records (Haberlandt & Radtke, 2014; Lamb, 1999). These uncertainties eventually translate to important errors in the simulated extremes, which can bias the subsequent analyses (Mirzaei et al., 2014). For example, despite the good model performance in reproducing mild flow peaks, a FFA study in Western Connecticut based on simulated flows from the Coupled Routing and Excess Storage distributed hydrological model (Shen & Anagnostou, 2017), showed systematic underestimation in the estimated quantiles as a result of the biases in the extreme flow peaks simulated by the model (Hardesty et al., 2018).

In this study, we evaluate the potential of a novel non-asymptotic statistical approach based on ordinary, as opposed to extreme, flood peaks (Miniussi, Marani, & Villarini, 2020) for FFA of hydrologic simulations in poorly-gauged or even ungauged basins. Our working hypothesis is that, since simulations of extreme flood peaks are generally characterized by larger errors than simulations of ordinary flood peaks, this method could be less sensitive to the issues that currently hinder FFA based on simulated streamflow. We use data from 671 catchments located across diverse hydroclimatic settings in the Contiguous United States (CONUS) to evaluate the potential of this approach by comparing it to traditional extreme-value methods based on annual maxima.

2 | STUDY AREA AND DATA

This study is based on the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS) dataset (Addor et al., 2017; Newman et al., 2015), which includes 671 catchments (see Figure 1) spanning a wide range of hydroclimatic conditions across CONUS. The dataset consists of hydrometeorological time series from different sources as well as several catchment attributes that influence catchment response and hydrological signature, such as topographic characteristics and climatic indices (Newman et al., 2015). The catchment scales span several orders of magnitude, ranging from 22 to 25,680 km^2 (median 470 km^2).

FIGURE 1 Stations with simulated flows from CAMELS used in this study. The figure shows a total of 671 stations with 28-year simulations, indicated by blue and orange dots; the orange dots mark the 308 stations with 70-year USGS observations or more data records. Solid lines indicate hydrological units (Seaber et al., 1987) across CONUS.

Streamflow time series include both observations from the U.S. Geological Survey (USGS) and model simulations derived from three different forcing: Daymet (Thornton et al., 2014, 2020), NLDAS (North American Land Data Assimilation System, Xia et al., 2012), and Maurer (Maurer et al., 2002). Daymet comprises gridded estimates of daily weather parameters and climatological summaries derived from daily meteorological observations. NLDAS constructs a forcing dataset from a daily gauge-based precipitation analysis, bias-corrected shortwave radiation, and surface meteorology reanalysis to produce outputs. Maurer presents a model-derived dataset of land surface states and fluxes for the CONUS.

The corresponding simulated streamflow time series are produced following the National Weather Service standard, using the SNOW-17 and Sacramento soil moisture accounting (SAC-SMA) hydrological model and shuffled complex evolution optimization calibration procedure (Duan et al., 1994). Streamflow derived from Daymet and NLDAS are available from January 1980 to December 2014, while Maurer covers from January 1980 to December 2008. Here, we focus on the 28-year hydrological period from October 1980 to September 2008.

Among the 671 USGS stations, we identified a subset of 308 stations for which at least 70 years of observed daily records are available. The full 70-year record is used here as an optimal and fair reference for quantile estimates and the sensitivity evaluation of statistical methods on simulated records.

3 | METHODS

3.1 | Statistical approach

Marani and Ignaccolo (2015) proposed a nonasymptotic approach to estimate extreme quantiles based on the analysis of the so-called ordinary events, which are all the independent realizations of a process of interest. The idea is that extremes are realizations of these underlying ordinary events, which are sampled a different and finite number of times every year. When the cumulative distribution function of the ordinary events is known, the description of extremes becomes a function of (i) this underlying distribution and (ii) the occurrence frequency of the ordinary events (see details in Zorzetto et al., 2016). Provided the underlying assumptions are met, these methods based on ordinary events showed reduced stochastic uncertainties in the estimation of extreme precipitation (Marra et al., 2018; Zorzetto et al., 2016) and flood (Miniussi, Marani, & Villarini, 2020) quantiles with respect to traditional extreme-value methods based on annual maxima or peaks over threshold, and proved less sensitive to systematic and random errors in the observed (or simulated) extremes (Marra et al., 2018). In fact, this approach includes a larger amount of data in the estimation of extreme quantiles with respect to traditional extreme-value methods, because all the independent flood peaks are used instead of annual maxima or threshold exceedances only.

A simplified version of this approach, termed Simplified Metastatistical Extreme Value (SMEV, Marra, Nikolopoulos, et al., 2019; Marra, Zoccatelli, et al., 2019), provides two important advantages: (i) improved parameter estimation accuracy; (ii) reduced sample-size issues related to the limited number of independent flood peaks in individual years (Marra et al., 2020; Miniussi, Marani, & Villarini, 2020; Vidrio-Sahagún & He, 2022). SMEV resembles the ordinary statistics of finite-size block maxima under independence (Serinaldi et al., 2020); here we will use the term SMEV for consistency with recent studies.

It is important to point out that ordinary-event based methods rely on different assumptions with respect to traditional extreme-value methods. Crucially, the class of distribution describing the ordinary events tail (F in the

following) needs to be known or reasonably assumed. Given the cumulative distribution function of the ordinary events $F(x;\theta)$ described by the array of parameters θ, and the average yearly number of ordinary events n, the SMEV cumulative distribution of extreme intensities x can be written as:

$$\zeta(x) = F(x;\theta)^n \tag{1}$$

Miniussi, Marani, and Villarini (2020) applied for the first time these methods to USGS observations across CONUS, finding that the Gamma distribution is the most suitable (among Weibull, Generalized Pareto, and Gamma) for describing ordinary flood peaks. This means that Gamma-distributed ordinary flood peaks were able to best explain the distribution of annual maximum peak discharges in most of the examined catchments. Following these results, here we adopt the Gamma distribution to describe ordinary flood events in this study.

In order to consistently apply the method, it is essential to identify ordinary events as independent. We follow the recommendation of Lamb (1999), according to which independent flood peaks should be separated by time blocks longer or equal to $T = 10$ days $+ (\log(A) - 0.41)$, where A is the basin area in square kilometers. Meanwhile, the smallest discharge value between two consecutive peaks below a threshold equal to 75% of the lower peak should be rejected. Parameters of the Gamma distribution are estimated using these independent peaks (ordinary events) using the method of the L-Moments (Hosking, 1990).

3.2 | Traditional extreme-value methods

Widely used traditional extreme-value methods used for FFA are based on both annual maxima and peaks over threshold: Generalized Extreme Value (GEV) fit of the annual maxima and Generalized Pareto (GPD) fit of the peaks exceeding a high threshold, as suggested by asymptotic statistical theory (Fischer & Tippett, 1928; Gnedenko, 1943), and Log-Pearson III (LP3) fit of the annual maxima, as recommended by the USGS. Previous work by Hu et al., 2020 for FFA over CONUS showed that, among these, the LP3 method provides more robust performance, and therefore was chosen as a benchmark for comparison with the SMEV. The parameters of the LP3 distribution are estimated using the method of the L-moments, owing to its lower sensitivity to outliers with respect to maximum likelihood methods, and to its ability to better estimate the tail heaviness in the presence of relatively short data records (Hosking, 1990).

3.3 | Quantification of uncertainty

Uncertainty is quantified using the bootstrap approach suggested by Overeem et al. (2008). A thousand M-year synthetic records, where M is the data record length, are generated via random sampling with replacement of the years in the record. Flood quantiles are then estimated for the synthetic records using the methods described above (SMEV and LP3).

3.4 | Biases in annual maxima and in FFA of simulated streamflow

To quantify how errors in the simulated extreme flood propagate to the estimated quantiles using the different statistical models, we studied the relationship between the bias in the simulated annual peak flow maximum series (AMS) and the one in the estimated quantiles. Errors are here quantified as absolute biases normalized over the catchment area:

$$Bias_{AMS} = \frac{1}{A * M} \cdot \sum_{j=1}^{M} \left(X^j_{sim} - X^j_{obs} \right) \tag{2}$$

where X^j_{sim} is the simulated annual peak flow maximum for the year j, X^j_{obs} the corresponding observed maximum, M is the number of years in the records (in our case the 28 years of simulated flows), and A is the catchment area.

3.5 | Evaluation of FFA of simulated streamflow

To evaluate the performance of the statistical methods using a robust reference, we conduct an analysis based on 308 stations with at least 70 years of observations, in order to ensure a robust reference. Based on these 308 stations, the performance of SMEV applied on 28-year simulated streamflow was evaluated across CONUS. We also compared SMEV with LP3 by calculating the ratios between the corresponding relative differences (RD), defined as:

$$RD = \frac{Q_{sim} - Q_{obs}}{(Q_{sim} + Q_{obs})/2} \times 100\% \tag{3}$$

where for each examined return period and realization, Q_{sim} is the quantile estimated from model-simulations using a given statistical method and Q_{obs} is the corresponding quantile estimated from observed streamflow using the same statistical method.

We quantify the estimation accuracy using the Skill Score (Hashino et al., 2006; Miniussi, Marani, & Villarini, 2020; Murphy & Winkler, 1992):

$$SS = \rho^2_{Q_{sim}Q_{obs}} - \left[\rho_{Q_{sim}Q_{obs}} - \left(\frac{\sigma_{Q_{sim}}}{\sigma_{Q_{obs}}}\right)\right]^2 - \left[\frac{\left(\mu_{Q_{sim}} - \mu_{Q_{obs}}\right)}{\sigma_{Q_{obs}}}\right]^2 \quad (4)$$

where ρ_{xy} is the correlation coefficient between x and y, and σ_x and μ_x are the standard deviation and mean of x, respectively. A perfect value for SS is 1.

4 | RESULTS

In the following, we present our results organized into three parts: First, we show the relation between bias in the simulated AMS and bias in the estimated flood quantiles. This will quantify the error propagation from simulated AMS to estimated flood quantiles (Section 4.1). Then, we evaluate the ability of the different methods to reproduce quantiles based on streamflow simulations by comparing them with the quantiles obtained from observed streamflow records; last, we assess possible geographic dependence in the accuracy of these methods.

4.1 | Propagation of errors from simulated annual maxima to estimated quantiles

The models reproduce flood occurrence reasonably, although a general tendency toward the underestimation of the number of flood peaks is noticed (median ∼20%, see Appendix A). This implies that the models reproduce flood occurrence reasonably. The relation between biases in simulated vs. observed AMS (as in Equation 2) for the case of NLDAS-driven simulated streamflow against the corresponding biases in the quantile estimated using different statistical methods is reported in Figure 2. Arguably, the bias of the quantiles exhibits a significant dependence on return period. Most of the bias for the traditional method (LP3) is within 50 mm day^{-1} for the 10-year return period, and increases at the higher quantiles (>100 mm/day for the 100-year return period). The bias for SMEV-derived quantiles shows a smaller dispersion, which remains similar across return periods. This indicates that SMEV-based results are less impacted by the (large) biases which affect the simulated AMS, especially when longer return periods are sought. This comes as a consequence of using ordinary events instead of extremes only, and confirms our working hypothesis. The largest peak flows, which are typically associated with larger simulation errors, are weighted less in the

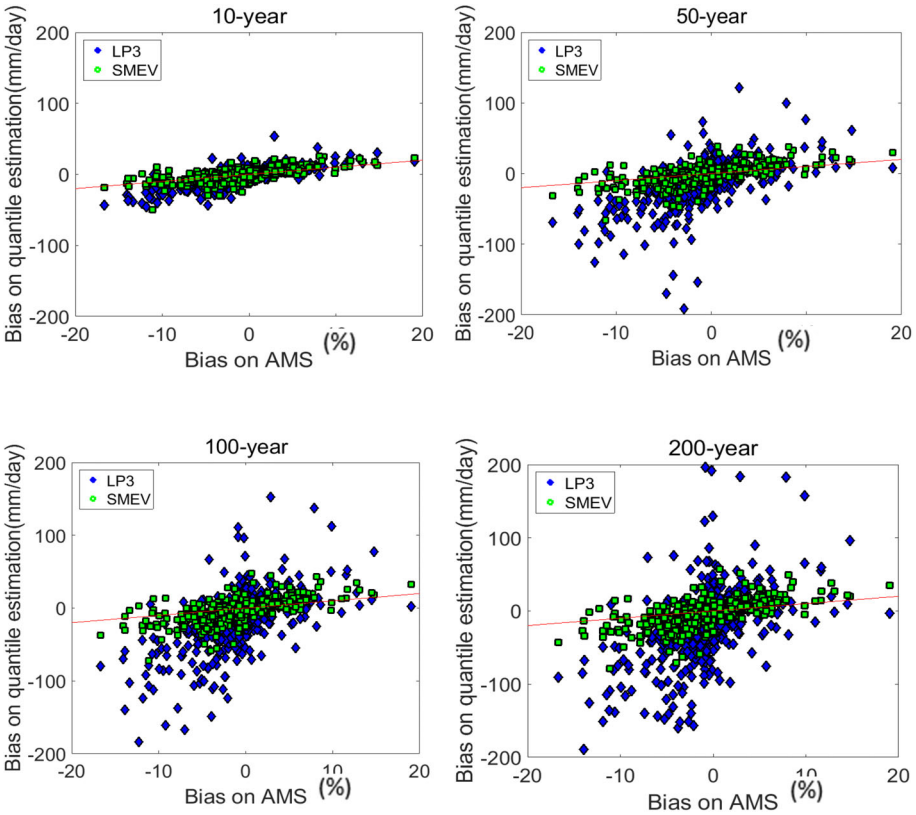

FIGURE 2 Bias on the 10-, 50-, 100-, and 200-year quantiles estimated from NLDAS-driven simulated streamflow using the four methods (LP3 and SMEV) as a function of the bias on the annual maxima series.

statistical inference, while the milder peak flows, which often contribute more to the model calibration, are weighted more.

It is worth pointing out that SMEV approaches allow also to explicitly exclude annual maxima (or any other specific event of the record) from the parameter estimation procedure in case those are not trusted (Marra et al., 2022), a possibility that can be explored in future applications.

4.2 | Evaluation of FFA from simulated streamflow

The evaluations of SMEV-based quantiles estimated using simulated streamflow time series (Daymet, Maurer, and NLDAS) are reported in Figure 3. SMEV exhibits a lower RDs between quantiles derived from observations and simulations for all return periods and forcing dataset (Figure 4). This is true both in terms of systematic (median value) and random (interquartile range and values range) components. In particular, it is interesting to notice that the dependence of RD on the return period is smaller for SMEV than for traditional methods. We use a Kolmogorov–Smirnoff test to check if the distributions of RD from SMEV and LP3 in Figure 3 are significantly different. The results show that the RD from SMEV and LP3 are significantly different at 5% significance level for all return periods except for the 2-year return period. In addition, we conducted qualitative evaluation on the interquartile ranges and whisker-to-whisker ranges of the boxplots in Figure 3. The whisker-to-whisker ranges of LP3 are 4%–96.5% higher than that of SMEV, and the interquartile ranges of LP3 are 5.5%–93.1% higher than that of SMEV, demonstrating that SMEV is less dependent on the return period.

This comes as a consequence of the reduced parameter estimation uncertainty, especially for what concerns the tail parameter of the distribution: in traditional extreme value methods, estimating the tail parameter (e.g., the shape parameter of the LP3 or of the commonly used GEV and GPD) is a well-known issue (Martins & Stedinger, 2000).

While the skill of LP3 quickly decreases with increasing return period, dropping from SS > 0.9 for 2-year quantiles to SS = ∼0.6 for 100-year quantiles, SMEV exhibits the highest Skill Score (SS = ∼0.9) consistently across return periods longer than 5 years (Figure 6). This follows what reported above and in Figure 2, and implies that, when the assumption about the ordinary flood events distribution is met (here, when the ordinary flood peaks are Gamma-distributed), using SMEV on simulated streamflow leads to estimates of high quantiles which are significantly more accurate than those that can be derived from traditional methods. This is due to the larger simulation biases characterizing extreme flows with respect to ordinary flow peaks.

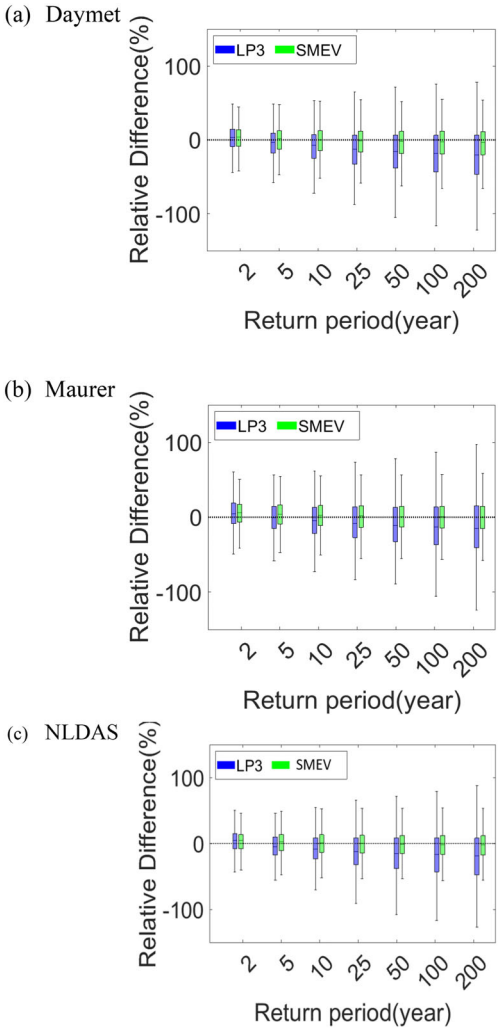

FIGURE 3 Boxplots of the RDs between quantiles estimated based on 28-year simulated streamflow against quantiles estimated based on 28-year observations; horizontal bars indicate the median, boxes indicate the interquartile range, and whiskers indicate the full range of values. Three forcing data are presented: Daymet (a), Maurer (b), and NLDAS (c).

4.3 | Geographic dependence of SMEV-based FFA accuracy

Results so far focused on the 28-year records for which simulated streamflow time series are available. Here, we provide a deeper evaluation of LP3 and SMEV-based methods using long-term observational records (≥70 years) as a reference, and we explore the possible dependence of their performance on hydroclimatic

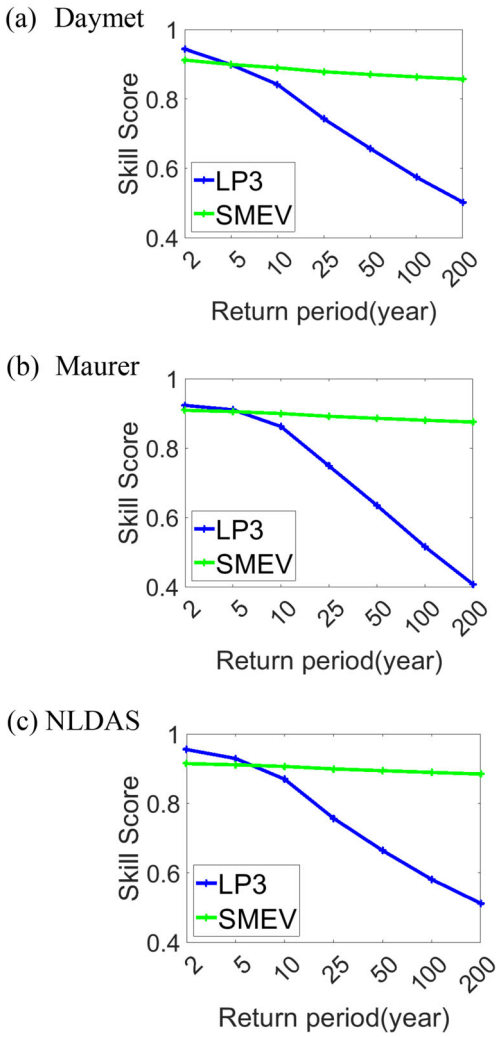

FIGURE 4 Skill score of the estimated quantiles using different extreme value methods from three forcing data: Daymet (a), Maurer (b), and NLDAS (c).

conditions and catchment area. Figure 5 and Table 1 present RDs from SMEV computed across CONUS. Using SMEV, over 70% and 65% stations exhibit less than 20% bias in the estimated 10-year and 100-year quantiles, respectively. Only less than 30 stations that exhibit more than 50% bias. Those results show that the low sensitivity of SMEV to record length is rather consistent across the majority of stations examined. Figure 6 and Table 2 further show the ratio between the RDs of LP3 and SMEV. Values larger than 1 (red colors in the figure) imply that SMEV estimates based on simulated streamflow are more similar to the estimates based on long-term observations and SMEV, while values smaller than 1 (blue colors in the figure) imply that LP3 estimates based on simulated streamflow are more similar to the estimates based on long-term observations and LP3. Comparison of traditional methods (here LP3) shows that quantiles estimated from simulated streamflow time series (28-year records) using SMEV are usually in better agreement with the estimates based on 70-year-long observations (66% of the cases for 10-year quantiles, 72% for 100-year quantiles). This is more frequent in the east and west coasts (California, New England, and Mid-Atlantic) and in the southwestern regions (Texas-Gulf) of CONUS (red dots in Figure 6), while a better performance of LP3 is reported in the Midwest and in the southeastern hydrological units.

In general, nonasymptotic approaches applied on simulated streamflow exhibit better performance in most of the climatic zones (Figure 7), with the notable exceptions of Mediterranean hot summer climates (Csa), cold semi-arid climates (Bsk), and continental climates at higher altitudes (Dsb). Interestingly, worse performances of nonasymptotic approaches in these climates were also reported for the case of precipitation (Marra et al., 2018); it was suggested that the presence of multiple mechanisms underlying the generation of extreme precipitation could be among the causes for this, but further research is needed to quantify the potential impact of other causes, such as nonstationarity of the ordinary flood events distribution (Vidrio-Sahagún & He, 2022).

5 | CONCLUSIONS

In this study, we make a further step toward reliable flood frequency analysis (FFA) in ungauged or poorly gauged basins by presenting a framework that increases the use of information from simulated streamflow series for robust estimation of extreme flood quantiles. Specifically, we evaluate the potential of applying a novel FFA approach, based on nonasymptotic ordinary statistics, on simulated streamflow data: we compare the Simplified Metastatistical Extreme Value approach (SMEV) with traditional methods based on extreme values (we used LP3 fit of annual maxima as recommended by the USGS and as deemed as the most representative based on past research). We use 308 stations with at least 70 years of observations in the CONUS and simulated streamflow from three different forcing datasets from CAMELS.

Our results confirm the working assumption of having SMEV-based estimates of extreme quantiles to be less sensitive to errors on the simulated extreme flows than those of traditional methods. This causes SMEV-based estimates of extreme quantiles to be less biased and characterized by smaller uncertainty than those from traditional methods, in particular for the estimation of higher quantiles (e.g., 100-year quantiles).

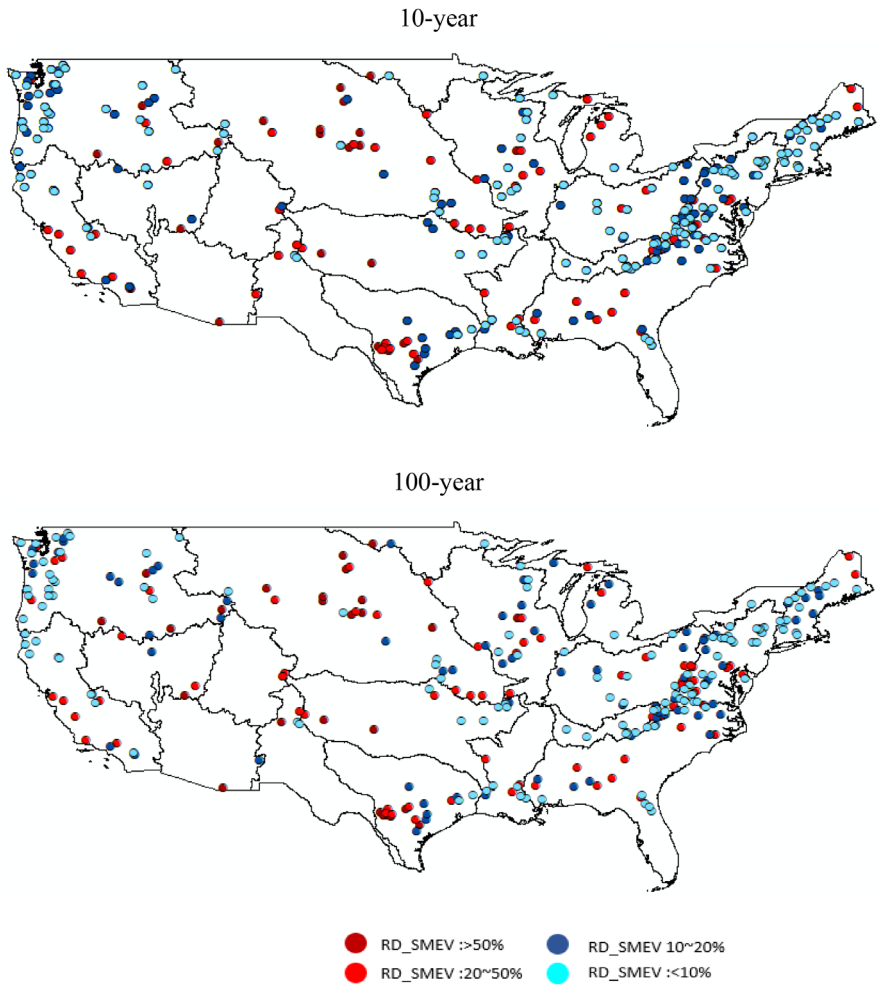

FIGURE 5 RDs from SMEV computed for the 10- and 100-year return periods at 308 stations in CONUS. Three hundred and eight stations with 70-year observations are considered as reference. The deep red and red dots indicate those stations at which SMEV presents more bias.

TABLE 1 The number of stations in RD_SMEV.

	Number of stations			
RD_SMEV (%)	**10-year return period**	**Total**	**100-year return period**	**Total**
<10	144	219	131	207
10–20	75		76	
20–50	66	89	73	101
>50	23		28	

When compared to analyses based on more than 70 years of data, FFA of simulated streamflow from SMEV presented smaller biases with respect to LP3 in 66% (72%) of the stations across CONUS in the estimation of 10-year (100-year) quantiles. Nevertheless, this is not systematic: SMEV is found to exhibit higher accuracy in the east and west coasts (California, New England, and Mid-Atlantic) and in the southwestern regions (Texas-Gulf) of CONUS, while LP3 seems preferable in the Midwest and in the southeastern catchments.

It should be recalled that both SMEV and LP3, as used here, rely on the assumption of stationarity. The possible violation of this assumption during the period examined (simulated records are 28 years long), however, is to be considered negligible with respect to the methodological uncertainty (see for instance Villarini et al., 2009). Should this violation become important, nonstationary implementations of SMEV are available, such as the one by Vidrio-Sahagún and He (2022).

Our quantitative results depend on the accuracy of our assumptions for what concerns the distribution of the ordinary events. We used a Gamma distribution following the results by Miniussi, Marani, and Villarini (2020)

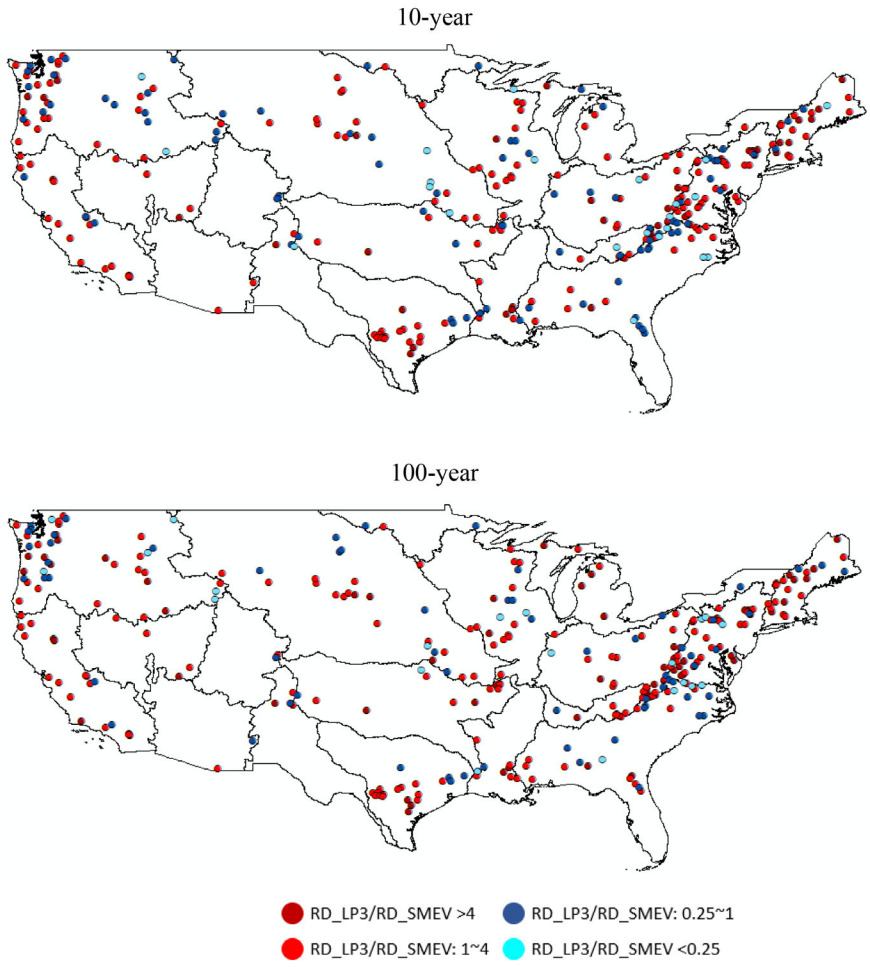

FIGURE 6 Ratio between the RDs from LP3 and those from SMEV, computed for the 10- and 100-year return periods at 308 stations in CONUS. The deep red and red dots indicate those stations at which SMEV provides more accurate estimation.

TABLE 2 The number of stations in the classifications of the ratios.

	Number of stations			
Ratio	**10-year return period**	**Total**	**100-year return period**	**Total**
<0.25	22	104	21	86
0.25–1	82		65	
1–4	164	204	158	222
>4	40		64	

and Mushtaq et al. (2022), but other study areas (or even some catchments within CONUS) could be better described by different distributions. To this end, specific tests should be used to define the optimal distribution case by case (Marra et al., 2022). To avoid this potential issue, our results are based on the comparison between simulated streamflow and observations using corresponding methods: SMEV (LP3) applied to simulated streamflow is compared to SMEV (LP3) applied to observations. Therefore, our results imply that, whatever distribution is the most appropriate for extremes, using ordinary events is a viable way to have FFA with low bias and small uncertainties using hydrological model simulations, especially if the hydrological model is not able to simulate high flow peaks accurately.

While not directly solving the problem of ungauged catchments, these results suggest a viable way for running FFA over poorly gauged catchments: provided that the distribution of the ordinary events is known, a hydrological model that can be calibrated using few years of data can be effectively used to simulate continuous streamflow time series based on atmospheric forcing datasets.

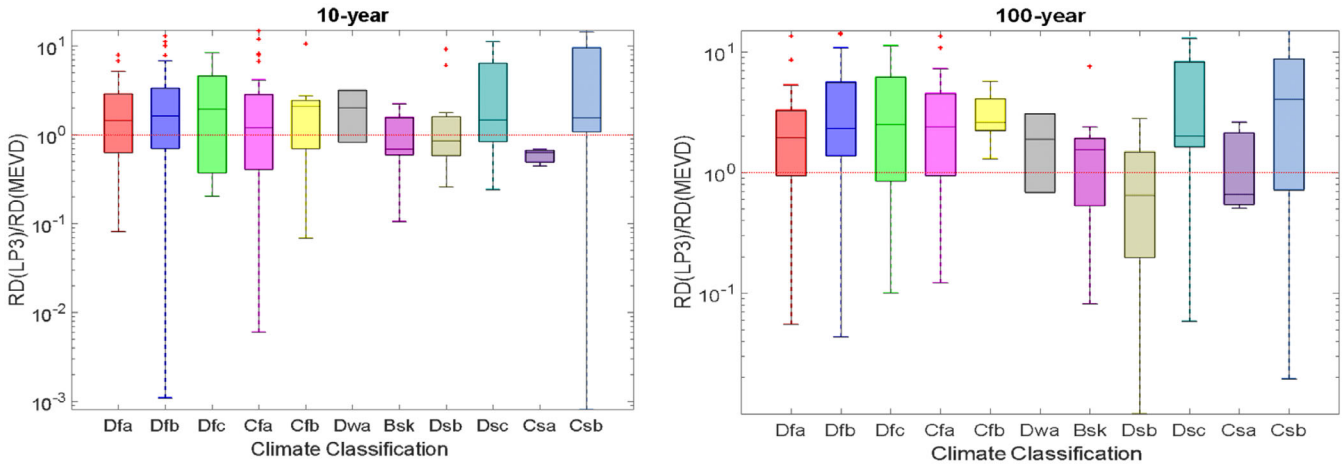

FIGURE 7 Ratio between the RDs from LP3 and those from SMEV, computed for the 10- and 100-year return periods at 308 stations across different climate conditions in CONUS.

ACKNOWLEDGMENTS

This study was supported by the Eversource Energy Center at the University of Connecticut. FM was partially supported by CARIPARO Foundation through the Excellence Grant 2021 to the "Resilience" Project.

DATA AVAILABILITY STATEMENT

All data generated or analysed during this study are included in this published article (and its supplementary information files).

ORCID

Efthymios I. Nikolopoulos https://orcid.org/0000-0002-5206-1249

Emmanouil N. Anagnostou https://orcid.org/0000-0002-1622-0302

REFERENCES

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*, 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Aghakouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research: Atmospheres*, *116*, D02115. https://doi.org/10.1029/2010JD014741

Ahn, J., Cho, W., Kim, T., Shin, H., & Heo, J.-H. (2014). Flood frequency analysis for the annual peak flows simulated by an event-based rainfall-runoff model in an urban drainage basin. *Water*, *6*, 3841–3863. https://doi.org/10.3390/w6123841

Apel, H., Merz, B., & Thieken, A. H. (2008). Quantification of uncertainties in flood risk assessments. *International Journal of River Basin Management*, *6*, 149–162. https://doi.org/10.1080/15715124.2008.9635344

Ayalew, T. B., & Krajewski, W. F. (2017). Effect of river network geometry on flood frequency: A tale of two watersheds in Iowa. *Journal of Hydrologic Engineering*, *22*, 06017004. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001544

Ball, J., Babister, M., Nathan, R., Weinmann, P., Weeks, W., Retallick, M., & Testoni, I. (2016). *Australian rainfall and runoff—A guide to flood estimation*. Commonwealth of Australia.

Bell, V., Kay, A., Jones, R., & Moore, R. (2007). Use of a grid-based hydrological model and regional climate model outputs to assess changing flood risk. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *27*, 1657–1671. https://doi.org/10.1002/joc.1539

Beven, K. (1986). Runoff production and flood frequency in catchments of order *n*: An alternative approach. Scale problems in hydrology. In V. K. Gupta, I. Rodríguez-Iturbe, & E. F. Wood (Eds.), *Scale problems in hydrology. Water science and technology library* (Vol. 6). Springer. https://doi.org/10.1007/978-94-009-4678-1_6

Beven, K. (1987). Towards the use of catchment geomorphology in flood frequency predictions. *Earth Surface Processes and Landforms*, *12*, 69–82. https://doi.org/10.1002/esp.3290120109

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., … Živković, N. (2019). Changing climate both increases and decreases European river floods. *Nature*, *573*, 108–111. https://doi.org/10.1038/s41586-019-1495-6

Bradley, A. A., & Potter, K. W. (1992). Flood frequency analysis of simulated flows. *Water Resources Research*, *28*, 2375–2385. https://doi.org/10.1029/92WR01207

Chilkoti, V., Bolisetti, T., & Balachandar, R. (2017). Climate change impact assessment on hydropower generation using multi-model climate ensemble. *Renewable Energy*, *109*, 510–517. https://doi.org/10.1016/j.renene.2017.02.041

Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology*, *158*, 265–284. https://doi.org/10.1016/0022-1694(94)90057-4

Ehsan Bhuiyan, M. A., Nikolopoulos, E. I., Anagnostou, E. N., Polcher, J., Albergel, C., Dutra, E., Fink, G., Martínez-De La

Torre, A., & Munier, S. (2019). Assessment of precipitation error propagation in multi-model global water resource reanalysis. *Hydrology and Earth System Sciences*, *23*, 1973–1994. https://doi.org/10.5194/hess-23-1973-2019

Fischer, R., & Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*, 180–190.

Franks, S. W., & Kuczera, G. (2002). Flood frequency analysis: Evidence and implications of secular climate variability, New South Wales. *Water Resources Research*, *38*, 20-1-20-7. https://doi.org/10.1029/2001WR000232

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, *44*, 423–453.

Griffis, V., & Stedinger, J. (2007). Log-Pearson type 3 distribution and its application in flood frequency analysis. II: Parameter estimation methods. *Journal of Hydrologic Engineering*, *12*, 492–500. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:5(492)

Haberlandt, U., & Radtke, I. (2014). Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows. *Hydrology and Earth System Sciences*, *18*, 353–365.

Haddad, K., Rahman, A., & Stedinger, J. R. (2012). Regional flood frequency analysis using Bayesian generalized least squares: A comparison between quantile and parameter regression techniques. *Hydrological Processes*, *26*, 1008–1021. https://doi.org/10.1002/hyp.8189

Hardesty, S., Shen, X., Nikolopoulos, E., & Anagnostou, E. (2018). A numerical framework for evaluating flood inundation hazard under different dam operation scenarios—A case study in Naugatuck River. *Water*, *10*, 1798. https://doi.org/10.3390/w10121798

Hashino, T., Bradley, A., & Schwartz, S. (2006). Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrology and Earth System Sciences Discussions*, *3*, 561–594. https://doi.org/10.5194/hess-11-939-2007

Hosking, J. R. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*. https://www.jstor.org/stable/2345653, *52*, 105–124.

Hu, L., Nikolopoulos, E. I., Marra, F., & Anagnostou, E. N. (2020). Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods: An evaluation over the contiguous United States. *Journal of Flood Risk Management*, *13*, e12580. https://doi.org/10.1111/jfr3.12580

Kan, G., Tang, G., Yang, Y., Hong, Y., Li, J., Ding, L., He, X., Liang, K., He, L., Li, Z., Hu, Y., & Cui, Y. (2017). An improved coupled routing and excess storage (CREST) distributed hydrological model and its verification in Ganjiang River Basin, China. *Water*, *9*, 904. https://doi.org/10.3390/w9110904

Kay, A., Davies, H., Bell, V., & Jones, R. (2009). Comparison of uncertainty sources for climate change impacts: Flood frequency in England. *Climatic Change*, *92*, 41–63. https://doi.org/10.1007/s10584-008-9471-4

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, *55*, 11344–11354. https://doi.org/10.1029/2019WR0260659

Lam, D., Thompson, C., Croke, J., Sharma, A., & Macklin, M. (2017). Reducing uncertainty with flood frequency analysis: The contribution of paleoflood and historical flood information. *Water Resources Research*, *53*, 2312–2327.

Lamb, R. (1999). Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation. *Water Resources Research*, *35*, 3103–3114.

Marani, M., & Ignaccolo, M. (2015). A metastatistical approach to rainfall extremes. *Advances in Water Resources*, *79*, 121–126. https://doi.org/10.1016/j.advwatres.2015.03.001

Marra, F., Borga, M., & Morin, E. (2020). A unified framework for extreme subdaily precipitation frequency analyses based on ordinary events. *Geophysical Research Letters*, *47*, e2020GL090209. https://doi.org/10.1029/2020GL090209

Marra, F., Levizzani, V., & Cattani, E. (2022). Changes in extreme daily precipitation over Africa: Insights from a non-asymptotic statistical approach. *Journal of Hydrology*, *X*, 100130.

Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., Bárdossy, A., & Morin, E. (2019). Precipitation frequency analysis from remotely sensed datasets: A focused review. *Journal of Hydrology*, *574*, 699–705. https://doi.org/10.1016/j.jhydrol.2019.04.081

Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., & Morin, E. (2018). Metastatistical extreme value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors. *Advances in Water Resources*, *117*, 27–39. https://doi.org/10.1016/j.advwatres.2018.05.001

Marra, F., Zoccatelli, D., Armon, M., & Morin, E. (2019). A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes. *Advances in Water Resources*, *127*, 280–290. https://doi.org/10.1016/j.advwatres.2019.04.002

Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, *36*, 737–744. https://doi.org/10.1029/1999WR900330

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., & Nijssen, B. (2002). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *Journal of Climate*, *15*, 3237–3251. https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2

Metzger, A., Marra, F., Smith, J. A., & Morin, E. (2020). Flood frequency estimation and uncertainty in arid/semi-arid regions. *Journal of Hydrology*, *590*, 125254. https://doi.org/10.1016/j.jhydrol.2020.125254

Miniussi, A., Marani, M., & Villarini, G. (2020). Metastatistical extreme value distribution applied to floods across the continental United States. *Advances in Water Resources*, *136*, 103498. https://doi.org/10.1016/j.advwatres.2019.103498

Mirzaei, M., Huang, Y. F., Lee, T. S., El-shafie, A., & Ghazali, A. H. (2014). Quantifying uncertainties associated with depth duration frequency curves. *Natural Hazards*, *71*, 1227–1239.

Moretti, G., & Montanari, A. (2008). Inferring the flood frequency distribution for an ungauged basin using a spatially distributed rainfall-runoff model. *Hydrology and Earth System Sciences*, *12*, 1141–1152. https://doi.org/10.5194/hess-12-1141-2008

Murphy, A. H., & Winkler, R. L. (1992). Diagnostic verification of probability forecasts. *International Journal of Forecasting*, *7*, 435–455. https://doi.org/10.1016/0169-2070(92)90028-8

Mushtaq, S., Miniussi, A., Merz, R., & Basso, S. (2022). Reliable estimation of high floods: A method to select the most suitable ordinary distribution in the metastatistical extreme value

framework. *Advances in Water Resources*, *161*, 104127. https://doi.org/10.1016/j.advwatres.2022.104127

Newman, A., Clark, M., Sampson, K., Wood, A., Hay, L., Bock, A., Viger, R., Blodgett, D., Brekke, L., & Arnold, J. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, *19*, 209–223. https://doi.org/10.5194/hess-19-209-2015

Overeem, A., Buishand, A., & Holleman, I. (2008). Rainfall depth-duration-frequency curves and their uncertainties. *Journal of Hydrology*, *348*, 124–134. https://doi.org/10.1016/j.jhydrol.2007.09.044

Pallard, B., Castellarin, A., & Montanari, A. (2009). A look at the links between drainage density and flood statistics. *Hydrology and Earth System Sciences*, *13*, 1019–1029.

Passerotti, G., Massazza, G., Pezzoli, A., Bigi, V., Zsótér, E., & Rosso, M. (2020). Hydrological model application in the Sirba River: Early warning system and GloFAS improvements. *Water*, *12*, 620. https://doi.org/10.5194/hess-13-1019-2009

Rahman, A., Haddad, K., & Eslamian, S. (2014). Regional flood frequency analysis. In S. Eslamian (Ed.), *Handbook of Engineering Hydrology: Modeling, Climate Change and Variability* (pp. 451–469). http://handle.uws.edu.au:8081/1959.7/548753

Rasheed, Z., Aravamudan, A., Gorji Sefidmazgi, A., Anagnostopoulos, G. C., & Nikolopoulos, E. I. (2022). Advancing flood warning procedures in ungauged basins with machine learning. *Journal of Hydrology*, *609*, 127736.

Renard, B., Lang, M., & Bois, P. (2006). Statistical analysis of extreme events in a non-stationary context via a Bayesian framework: Case study with peak-over-threshold data. *Stochastic Environmental Research and Risk Assessment*, *21*, 97–112. https://doi.org/10.1007/s00477-006-0047-4

Ruiz-Bellet, J. L., Castelltort, X., Balasch, J. C., & Tuset, J. (2017). Uncertainty of the peak flow reconstruction of the 1907 flood in the Ebro River in Xerta (NE Iberian Peninsula). *Journal of Hydrology*, *545*, 339–354. https://doi.org/10.1016/j.jhydrol.2016.12.041

Saghafian, B., Golian, S., & Ghasemi, A. (2014). Flood frequency analysis based on simulated peak discharges. *Natural Hazards*, *71*, 403–417.

Seaber, P. R., Kapinos, F. P., & Knapp, G. L. (1987). *Hydrologic unit maps*. U.S. G.P.O. https://doi.org/10.3133/wsp2294

Serinaldi, F., Lombardo, F., & Kilsby, C. G. (2020). All in order: Distribution of serially correlated order statistics with applications to hydrological extremes. *Advances in Water Resources*, *144*, 103686. https://doi.org/10.1016/j.advwatres.2020.103686

Sharifi Garmdareh, E., Vafakhah, M., & Eslamian, S. S. (2018). Regional flood frequency analysis using support vector regression in arid and semi-arid regions of Iran. *Hydrological Sciences Journal*, *63*, 426–440. https://doi.org/10.1080/02626667.2018.1432056

Shen, X., & Anagnostou, E. N. (2017). A framework to improve hyper-resolution hydrological simulation in snow-affected regions. *Journal of Hydrology*, *552*, 1–12. https://doi.org/10.1016/j.jhydrol.2017.05.048

Smith, A., Sampson, C., & Bates, P. (2015). Regional flood frequency analysis at the global scale. *Water Resources Research*, *51*, 539–553. https://doi.org/10.1002/2014WR015814

Srinivas, V., Tripathi, S., Rao, A. R., & Govindaraju, R. S. (2008). Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. *Journal of Hydrology*, *348*, 148–166. https://doi.org/10.1016/j.jhydrol.2007.09.046

St. George, S., & Mudelsee, M. (2019). The weight of the flood-of-record in flood frequency analysis. *Journal of Flood Risk Management*, *12*, e12512.

Stedinger, J. R. (1983). Confidence intervals for design events. *Journal of Hydraulic Engineering*, *109*, 13–27. https://doi.org/10.1061/(ASCE)0733-9429(1983)109:1(13)

Sun, W., Ishidaira, H., Bastola, S., & Yu, J. (2015). Estimating daily time series of streamflow using hydrological model calibrated based on satellite observations of river water surface width: Toward real world applications. *Environmental Research*, *139*, 36–45. https://doi.org/10.1016/j.envres.2015.01.002

Thornton, M. M., Shrestha, R., Wei, Y., Thornton, P. E., Kao, S., & Wilson, B. E. (2020). *Daymet: Daily surface weather data on a 1-km grid for North America, Version 4*. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1840

Thornton, P. E., Thornton, M. M., Mayer, B. W., Wilhelmi, N., Wei, Y., Devarakonda, R., & Cook, R. B. (2014). *Daymet: Daily surface weather data on a 1-km grid for North America, Version 2*. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1219

Vidrio-Sahagún, C. T., & He, J. (2022). Hydrological frequency analysis under nonstationarity using the metastatistical approach and its simplified version. *Advances in Water Resources*, *166*, 104244.

Villarini, G., Serinaldi, F., Smith, J. A., & Krajewski, W. F. (2009). On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resources Research*, *45*, 8. https://doi.org/10.1029/2008WR007645

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., Alonge, C., Wei, H., Meng, J., & Livneh, B. (2012). Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research: Atmospheres*, *117*, 3110. https://doi.org/10.1029/2011JD016051

Yang, D., Koike, T., & Tanizawa, H. (2004). Application of a distributed hydrological model and weather radar observations for flood management in the upper Tone River of Japan. *Hydrological Processes*, *18*, 3119–3132. https://doi.org/10.1002/hyp.5752

Yang, S., Yang, D., Chen, J., Santisirisomboon, J., Lu, W., & Zhao, B. (2020). A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data. *Journal of Hydrology*, *590*, 125206.

Zorzetto, E., Botter, G., & Marani, M. (2016). On the emergence of rainfall extremes from ordinary events. *Geophysical Research Letters*, *43*, 8076–8082. https://doi.org/10.1002/2016GL069445

How to cite this article: Hu, L., Nikolopoulos, E. I., Marra, F., & Anagnostou, E. N. (2023). Toward an improved estimation of flood frequency statistics from simulated flows. *Journal of Flood Risk Management*, e12891. https://doi.org/10.1111/jfr3.12891

APPENDIX A

Figures A1 and A2

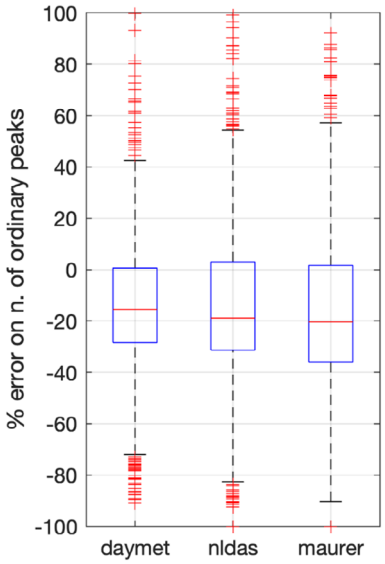

FIGURE A1 RDs between number of peaks estimated based on 28-year simulated streamflow against number of peaks estimated based on 28-year observations; horizontal bars indicate the median, boxes indicate the interquartile range, and whiskers indicate the full range of values. Three forcing data are presented: Daymet, NLDAS, and Maurer.

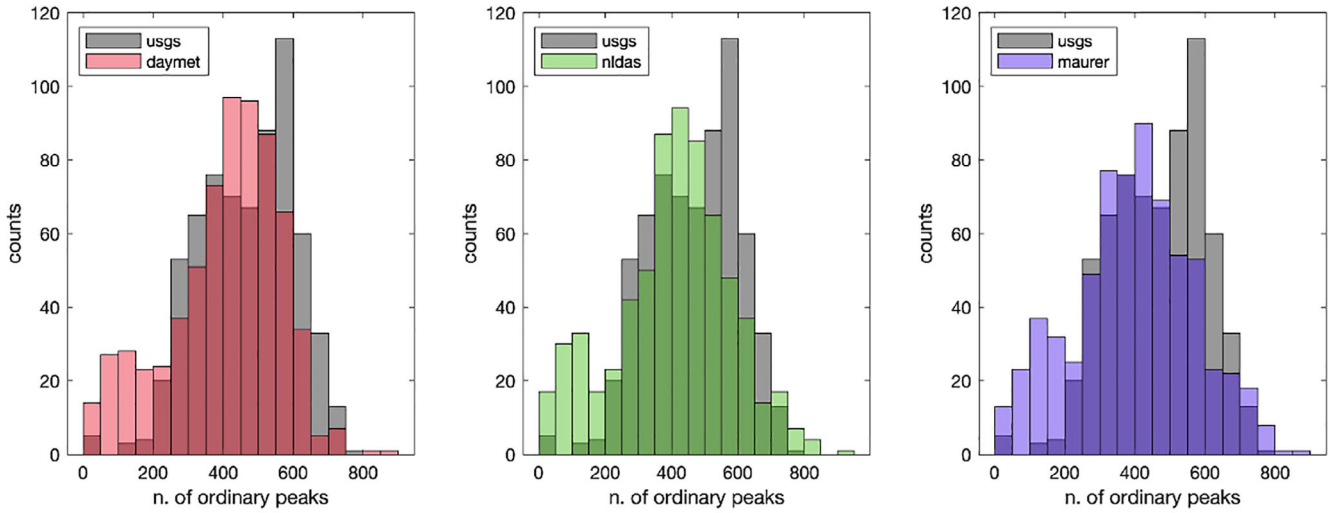

FIGURE A2 Histograms between number of peaks estimated based on 28-year simulated streamflow against number of peaks estimated based on 28-year observations; three forcing data are presented: Daymet, NLDAS, and Maurer.