Journal of Hydrology 590 (2020) 125254

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Flood frequency estimation and uncertainty in arid/semi-arid regions

Asher Metzger[a,*], Francesco Marra[a,b], James A. Smith[c], Efrat Morin[a,*]

[a] Institute of Earth Sciences, Hebrew University of Jerusalem, E.J. Safra Campus – Givat Ram, Jerusalem, Israel
[b] Institute of Atmospheric Sciences and Climate, National Research Council of Italy, CNR-ISAC, Bologna, Italy
[c] Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA

A R T I C L E I N F O

This manuscript was handled by Marco borga, Editor-in-Chief

Keywords:
Flood frequency analysis
Arid/semi-arid regions
Annual maxima series
Partial duration series
Extreme value theory
Parameter estimation

A B S T R A C T

At site flood frequency analysis (FFA) in arid/semi-arid watersheds poses unique challenges to researchers and practitioners due to the generally limited data records. This study presents a comprehensive evaluation of FFA in arid/semi-arid watersheds in relation to the unique characteristics of these regions, such as the limited number of floods occurring each year and the large variability of the flood peak discharges. Study cases in Israel and the US are examined and compared with non-arid watersheds, characterized by Mediterranean climate, and with synthetic flood records.

Results show that the tail of extreme value distributions describing arid/semi-arid watersheds is found to be heavier than the one describing Mediterranean watersheds. The number of yearly floods and the variability of flood peak discharge are shown to have a crucial impact on the accuracy of the quantile estimates with smaller number of events per year and larger coefficient of variation of flood peak discharge being related to larger errors in the estimated quantiles. Partial duration series approach provides a slightly reduced bias in the estimates, but should not be blindly preferred over annual maxima series as it presents comparable estimation uncertainty. In general, the generalized extreme value and the generalized Pareto distribution are found to be non-optimal choices for the examined arid/semi-arid watersheds.

1. Introduction

Hydrological design and flood risk management require estimates of quantiles of the peak discharges characterized by low yearly exceedance probability at a given location. Flood frequency analysis (FFA) aims at identifying analytical distributions reproducing the cumulative distribution of observed extreme flood peaks (e.g., the series of annual maximum peak) and whose tail is expected to represent the probability of exceeding extreme, and potentially still unobserved, discharges. The typical approach relies on observations and assumes stationarity of both catchment hydrological response and climatic forcing. However, despite the large amount of research devoted to the topic, there is still no consensus on a standard methodology able to provide information for an arbitrary gauged catchment and, even less, for ungauged catchments or nonstationary conditions (Coles, 2001; François et al., 2019; Katz, 2002).

A large number of studies focus on two distribution classes to describe extreme flood peak discharges, namely the generalized extreme value (GEV) distribution, for annual maxima, and the generalized Pareto (GP) distribution for the peaks exceeding a large threshold. These choices originate from the extreme value theorem (Fisher and Tippett, 1928; Gnedenko, 1943), which demonstrates that the extreme values of independent and identically-distributed random variables can only converge to these distributions. As stated by Haan (2002): "There is no direct theoretical connection between any analytical form of the frequency distribution and the underlying mechanisms governing flood flows except through the limit theorems", and these general results provide a-priori knowledge of the distribution describing the extremes thus limiting the problem to the estimation of the distribution parameters. However, reality is far from the asymptotic behavior and, even in presence of perfectly measured independent and identically distributed peaks, an infinite number of floods per year (or a large-enough threshold) are required for the theorem to hold, and practitioners sometimes rely on different distributions which seem to better represent the available observations.

However, the GEV and GP distributions are widely used for FFA worldwide as they represent a commonly accepted background (Ashkar and Ba, 2017; Ben-Zvi, 2016; Castillo and Hadi, 1997; Hosking et al., 1985; Hosking and Wallis, 1987; Katz, 2002; Martins and Stedinger, 2000; Meirovich et al., 1998; Morrison and Smith, 2002; Papalexiou and Koutsoyiannis, 2013; Rahman et al., 2013; Solari et al., 2017; Villarini and Smith, 2010). This becomes crucial for future studies

* Corresponding authors.
E-mail address: efrat.morin@mail.huji.ac.il (E. Morin).

https://doi.org/10.1016/j.jhydrol.2020.125254
Received 13 January 2020; Received in revised form 30 May 2020; Accepted 26 June 2020
Available online 05 July 2020

aiming at describing extreme flood peaks at large (global) scales, such as the ones already performed for precipitation (e.g., Papalexiou and Koutsoyiannis, 2013).

Depending on the available data series, applications may follow two approaches. The first is based on the maximum flood peak discharges observed during equally-long temporal blocks, which are expected to follow a GEV distribution. It is usually known by annual maxima series (AMS) as, due to meteorological periodicity reasons, the temporal blocks are chosen to be of a year length. The second approach, based on independent flood peaks exceeding a high pre-defined threshold, is known as partial duration series (PDS) or peaks-over-threshold. In this case, the exceedances are expected to follow a GP distribution. The choice among these stands with the practitioner and depends on the available data, the sought level of complexity in the data treatment, and a trade-off between estimation bias and uncertainty (Davison et al., 1990). Specifically, AMS makes use of limited data and allows to easily ensure independence of the flood peaks while PDS allows to include more than one flood per year, attempting to decrease the parameter estimation variance, at the price of methodological complexities, i.e., independence of flood peaks needs to be ensured and a proper threshold needs to be chosen. Among others, two methods are more frequently used for the distribution parameter estimation from the data series: the maximum likelihood (ML), and the probability weighted moments (PWM) (Greenwood et al., 1979; Sillitto, 1951). The PWM method is formally equivalent to the method of the L-moments since PWM and L-moments are linearly related (Ferreira and De Haan, 2015; Hosking et al., 1985; Hosking and Wallis, 1987; Martins and Stedinger, 2000). The ML method is generally deemed more accurate for the general case, while the PWM is renowned for a reduced estimation variance in presence of short records and reduced sensitivity to outliers (i.e., values particularly larger than the typical values in the record).

Arid/semi-arid regions cover about a third of the land area, globally, with indications that both the extent of the area and the residing population living there are increasing (Huang et al., 2016; Nicholson, 2011). Despite this, studies focusing on FFA in these regions are quite rare (Zaman et al., 2012).

In non-arid regions many studies comparing the AMS and PDS approaches are available, sometimes finding the PDS (e.g., Bezak et al. (2014) on the Sava River in Slovenia; Rahman et al. (2013) on some regions of Australia) or the AMS approach to be preferable (Rahman et al., (2013) on other regions of Australia). Interestingly, recent studies based on data from the US could not find particular improvements of the PDS over the AMS in presence of short data records (Schumer et al., 2014; Hu et al., 2019). Similarly, contrasting results are available on the parameter estimation method (Rahman et al., 2013), even if general agreement on the better performance of PWM for limited data records was reached (Martins and Stedinger, 2000; Hu et al., 2019).

Regional growth curves (i.e., dimensionless frequency curves representing the regional analogous to at-site frequency curves) of arid regions were found to be heavier-tailed than humid regions (Farquharson et al., 1992; Zaman et al., 2012), also as a consequence of the limited number of floods observed in each year. In a global-scale regional FFA study Smith et al. (2014) show that arid regional regression models perform poorer than humid regional models, presumably because arid regions are spatially more heterogeneous.

In particular, arid regions represent a particularly interesting case as they pose a number of climate-specific challenges. Arid catchments are characterized by a small number of flows per year, with frequent no-flow years, by a large variability of flood peak discharges and by limited lengths of observational records (Cohn et al., 2013; Knighton and Nanson, 2001; Schumer et al., 2014). As such, arid regions are further away from the asymptotic behavior described above; this, together with the generally limited length of observational records, might create practical problems for the parameter estimation and, in the case of PDS, for the choice of the threshold. It is thus crucial to understand the effects of these characteristics on FFA in arid/semi-arid watersheds.

To the best of our knowledge there is no comprehensive study fully exploring the impact of arid flood record characteristics on FFA. This study focuses on two study regions (Israel and US) with two climatic zones each (arid/semi-arid and Mediterranean) and uses observed flood records and synthetic records from a specifically developed stochastic flood generator to improve our understanding of these issues. Three research questions are addressed: (1) What are the tail characteristics of flood distributions in arid/semi-arid catchments in comparison to more humid regions? (2) How do sparse flood occurrence, short record length and other arid/semi-arid flood characteristics affect FFA and its uncertainty? (3) Which approach and parameter estimation methods are more adequate for FFA in arid/semi-arid catchments?

The paper is structured as follows. In Section 2 the study regions and data collection and processing are presented. In Section 3, FFA and flood generator methods are provided. In Section 4 the results of the study are presented. Discussion and conclusions, Sections 5 and 6, respectively, end the paper.

2. Study regions and data

The study focuses on watersheds characterized by arid/semi-arid climate regimes in Israel and US (Fig. 1). Mediterranean watersheds in the same two areas are used as a comparison, because they are geographically close, and thus characterized by similar precipitation patterns and seasonality, but larger yearly precipitation amounts. The climatic regions are defined following the Koppen-Geiger classification (Peel et al., 2007): arid/semi-arid watersheds are defined as Arid Hot Desert (BWh) or Arid Hot Steppe (BSh) climates, while Mediterranean watersheds by Temperate Hot and Dry Summer (Csa) climate.

Arid/semi-arid watersheds in Israel are located in the southern part of the country (Fig. 1). Mean annual precipitation ranges between 20 and 500 mm yr^{-1}. Flood occurrence and magnitude (Fig. 2A) follow the precipitation climatology, with most of the floods in the winter months (December to February), when Mediterranean lows deliver most of the annual precipitation to the area, and fewer floods, but with larger unit peak discharges, in the transition months, when strong convection is frequently brought by active Red Sea troughs and, more rarely, tropical plumes (Armon et al., 2018; Kahana et al., 2002). The Israeli watersheds are characterized by large areas of bare rock and shallow soils, such as lithosols, with mostly absence of vegetation and with the presence of debris cover and desert pavement. Some of the watersheds (mainly in the north-eastern part of the study region) have wetter upstream areas where one can find more urbanization and cultivated areas over more developed soils such as luvisols. Some sections of the channel beds are covered in alluviums (Greenbaum et al., 2006; Shentsis et al., 1999; Zoccatelli et al., 2019). Selected Mediterranean watersheds in Israel are situated in the northern part of the country, with mean annual precipitation ranging between 400 and 700 mm yr^{-1}.

Flood data obtained from the Israeli Hydrological Service have gone through the regular data quality assurance of such institutions. Stage-discharge rating curves are examined and updated when necessary stations are visited regularly. However, one should be aware that streamflow data in arid regions may suffer from inherent challenges such as changes in channel cross-sections during flood events due to incision or sediment deposition, few direct measurements at high flood stages due to limited accessibility in such situations, sub-optimal flow conditions due to the typical steep gradients and possible non-uniform flow over the cross sections. Events with corrupted or unreliable data were removed from the data set. The gaps in this study were not great and did not affect the analysis as the data for frequency analysis does not depend on sequential uninterrupted data. The obtained data were already separated into independent flood events. Data records include flood peak discharges from all flow events recorded in the hydrometric stations. After quality control, 18 stations were chosen for the arid/semi-arid region (Fig. 1 and Table A.1) and 16 for the Mediterranean region (Fig. 1 and Table A.3) based on the criterion of having at least

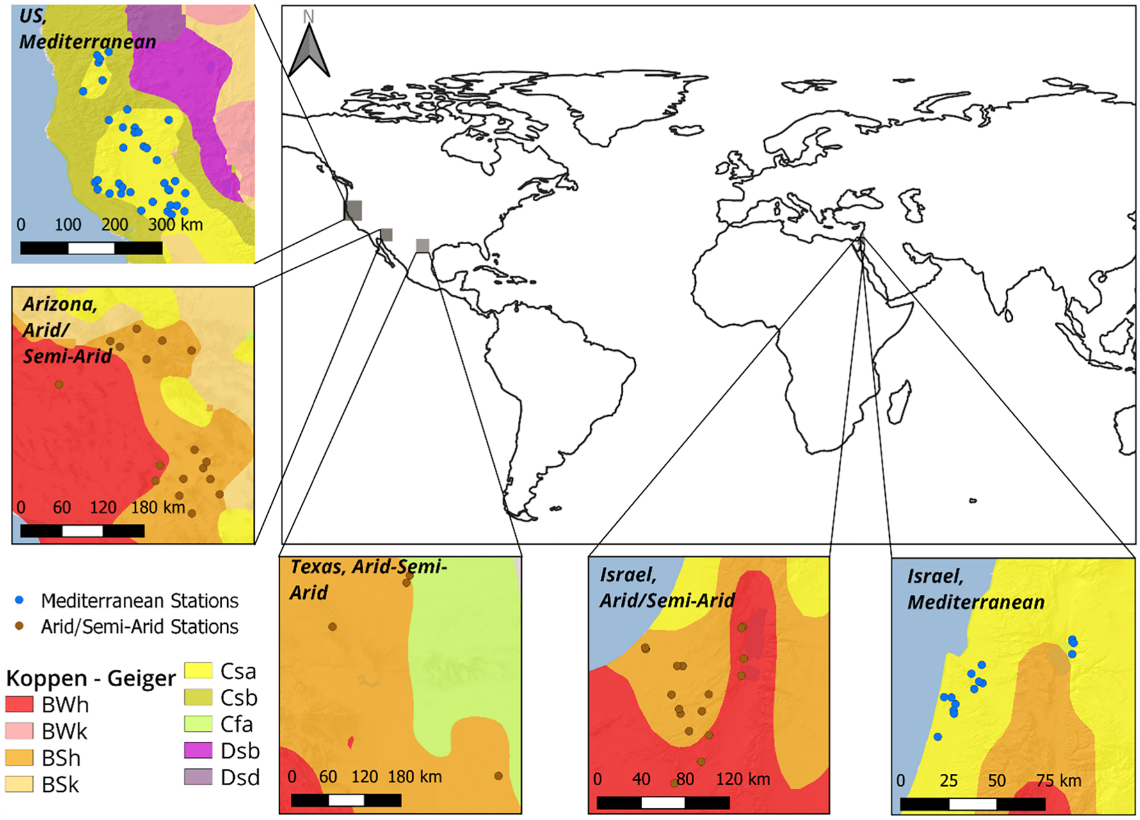

Fig. 1. The arid/semi-arid and Mediterranean study regions in Israel and the US. The Koppen-Geiger climate classification (Peel et al., 2007) and locations of hydrometric stations analyzed in this study are presented. See Tables A.1 and A.2 for stations details.

15 years of quality-controlled data.

The arid/semi-arid US study area includes watersheds clustered in two areas in the southwestern part of the country. US watersheds exhibit a broad range of hydrogeological properties. Like the Israeli watersheds, extensive regions of bare rock and shallow soils characterize a number of the watersheds. For a summary of watershed properties and their relations to flood processes in arid/semi-arid watersheds in the US, see Smith et al. (2018) and references therein. In Arizona the watersheds belong to the lower Colorado region, with mean annual precipitation between 200 and 700 mm yr^{-1}. In Texas three watersheds belong to the Texas Gulf region and one watershed to the Rio Grande River Basin. The mean annual precipitation ranges between 400 and

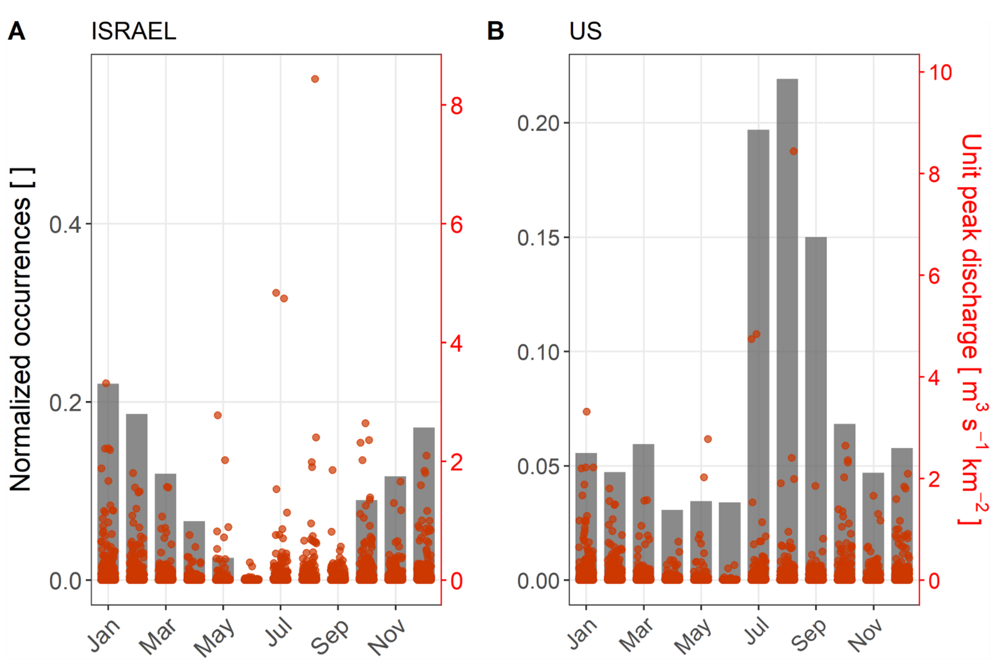

Fig. 2. Seasonality of flood occurrences (grey bars) and flood peak unit discharge (red points) in the arid/semi-arid study regions for Israel and the US. Note the different y-axes scales.

600 mm yr^{-1}. Floods mostly (> 50%) occur during summer (Fig. 2B), in phase with the largest unit peak discharges. However, large floods are also reported during winter. In the transition seasons there are fewer floods with very low unit peak discharge. The Mediterranean watersheds of the US regions are situated in the northern part of California and southern Oregon, with the mean annual precipitation ranging from 500 to over 900 mm yr^{-1}. We use US Geological Survey (USGS) stream gaging records to examine flood peak distributions (see Ryberg et al., 2017); a recent summary of USGS discharge measurements is provided in Turnipseed and Sauer (2010). Measurements of many extreme floods are made by indirect discharge methods, involving field measurements of peak water surface profiles and channel cross sections, combined with hydraulic computations (Costa and Jarrett, 2008). Indirect measurements are made for floods at stream gaging sites when the gage is destroyed or fails to operate properly. Peak discharge from indirect measurements have significant errors, especially for the most extreme flood peaks in arid/semi-arid regions (see Costa and Jarrett, 2008; Smith et al., 2019, 2018). Flood records were retrieved from the USGS, including (i) continuous data records containing instantaneous discharge values, and (ii) annual peak discharge records, that typically have longer records. In addition to the 15-year quality-controlled minimum record, watersheds draining more than 5000 km^2 or with substantial urban areas were excluded from the analysis to better match the characteristics of the Israeli watersheds. After this selection, 21 arid/semi-arid watersheds (Fig. 1 and Table A.2) and 36 Mediterranean watersheds were used for the analysis (Fig. 1 and Table A.4). Both types of records were retained for arid/semi-arid watersheds, while only the annual peak discharge records were used for Mediterranean watersheds analysis. Continuous data required separation of independent floods: events are considered independent when separated by at least 24 h with discharge lower than twice the median flow of the station. After sensitivity analysis, this threshold was chosen as the largest threshold providing consistent estimates of the distribution parameters across the watersheds.

3. Methods

3.1. Flood frequency analysis

When dealing with FFA, one should consider that identical distribution is assumed (i.e., stationarity); as the available records are typically short and over-dispersed, potential non-stationary models cannot be extracted from the available data. We will thus stand with Serinaldi and Kilsby (2015) and assume stationarity, in light of the relatively short and high variance available records.

3.2. AMS approach

AMS are prepared by taking the maximum peak discharge observed in each hydrological year (Oct 1st – Sep 30th). The GEV has the following cumulative distribution function (Coles, 2001):

$$G(x) = \begin{cases} exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \xi \neq 0 \\ exp\left\{-exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, \xi = 0 \end{cases} \tag{1}$$

where x is the data, μ is the location parameter, σ is the scale parameter and ξ is the shape parameter. A shape parameter larger than 0 indicates a power-type tail (the decay rate is slower than exponential) and a shape parameter smaller than 0 indicates upper bounded distributions.

3.3. PDS approach

PDS data consist of all the independent flood peaks exceeding a given threshold. The threshold selection is far from being straightforward (see Dupuis, 1999, and Scarrott and Macdonald, 2012 for an in-depth review) and, despite few automatic methods have been recently proposed for precipitation data (e.g., Fukutome et al., 2015; Solari et al., 2017), to the best of our knowledge there is no method that has shown superiority over others across regions and climates. Additionally, depending on the tail type, the threshold selection was shown to be a potential source of systematic bias in the results (Papalexiou et al., 2018).

Novel approaches (Ben-Zvi, 2016; Solari et al., 2017) have been tested and found to lack robustness across regions and watersheds. In particular, large instability was reported for arid/semi-arid watersheds. In order to adopt a robust methodology able to work in diverse climatic conditions, we decided to adopt a simple and commonly used approach that sets the threshold able to identify a pre-defined number of floods per year. Ben-Zvi (2016) reports that literature recommends to use between 1 and 5 floods per year. Dealing with arid watersheds, the only way to fulfill the 'high enough' requirement of EVT is to use a small factor, and therefore, following Cunnane (1979) and many others, we adopt the recommended number of 1.65 flood per year. Sensitivity analyses on our watersheds showed no significant deviation of the results for factors between 1.5 and 2.

The GP distribution is expected as the limiting distribution of the of a series of independent and identically distributed random variables exceeding a high-enough threshold (Coles, 2001; Davison et al., 1990). The GP has the following cumulative distribution function (Coles, 2001):

$$G(x) = \begin{cases} \left\{1-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \xi \neq 0 \\ \left\{1-exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, \xi = 0 \end{cases} \tag{2}$$

where x is the data, μ is the location parameter, σ is the scale parameter and ξ is the shape parameter. As for the case of the GEV, a shape parameter larger (smaller) than 0 indicates a heavy (light) tail of the distribution. The GP location parameter is also known as the threshold and $x - \mu$ is also referred to as exceedances above the threshold.

3.4. Parameter estimation methods

In this study, the ML and PWM methods are applied to derive the three parameters (μ, σ and ξ) of the GEV distribution fitting the AMS, and two (σ and ξ) of the three parameters of the GP distribution fitting the PDS, as the GP location parameter μ, i.e. the threshold, is set prior to parameter estimation. For both methods, we used the *gevFit* and *gpdFit* functions from the ***fExtremes*** R package (Wuertz, 2006) to estimate the parameters and calculate the distribution quantiles.

3.5. Uncertainty range estimation

Uncertainty is quantified using a non-parametric bootstrapping method (Hall et al., 2004; Overeem et al., 2008). As recommended by Hall et al. (2004), we used 1000 samples with replacement among the hydrological years in the record. In the case of PDS, the threshold discharge and the peaks exceeding the threshold are newly determined at each resample. The distribution parameters are then estimated and the quantiles calculated. Uncertainty is quantified as the 95% range (i.e., the range between the 2.5% and the 97.5% bounds).

For assessment of the extent of the uncertainty, a normalized uncertainty range is defined as the ratio of the uncertainty range (i.e. the difference between the 97.5% and the 2.5% bounds) by the estimated return level.

3.6. Estimation of the tail heaviness

We quantify the heaviness of the distribution tails using the tail ratio defined here as the ratio between the 100-year and the 10-year flood peaks. Heavier tails will be characterized by larger ratios, and vice versa. The method is similar to the method presented in Smith et al. (2018), but the ratio between maximal observed and 10-year flood peak was used.

3.7. *Uncertainty in shape parameter estimated from short records*

Record length is a major contributor to FFA uncertainty (Martins and Stedinger, 2000; Serinaldi and Kilsby, 2014) and, due to the lower density of monitoring stations, is a typical issue affecting arid/semi-arid regions (Nicholson, 2011). Here, we examine the effect of record length on the uncertainty in the shape parameter estimation. To do so, stations with records exceeding 50 years were selected (3 stations from Israel and 11 from US). Ensembles were created by bootstrapping (1000 repetitions) records increasing from 15 years to the full record length of the station and the uncertainty in the GEV/PWM shape parameter was computed for as the 95% interval.

3.8. *Synthetic arid/semi-arid flood generator*

A synthetic flood generator is developed to produce realistic flood records characterized by: (i) the desired distribution of the number of floods observed annually, and (ii) the desired parent distribution describing the flood peak discharges. In arid/semi-arid regions, the distribution of the annual number of floods is often over-dispersed (80% of the examined watersheds, Fig. 3). Thus, a negative binomial distribution, more adequate in over-dispersed conditions than a Poisson distribution (Lang et al., 1999), was chosen. Concerning the parent distribution describing the flood peak discharges, four distributions were examined: Weibull, Pareto, log-normal and gamma. In this case, both goodness of fit tests and graphical diagnosis were used. It was found that the log-normal distribution best fits the data of most of the arid/semi-arid stations. Where this might be the case in this study, other distributions may be better for modelling peak discharge distribution for other regions or climates. Log-normal distribution of flood peak was also found to be a good model for ordinary events by (Zhang et al., 2019). Parameters of the negative binomial and log-normal distributions were estimated at each station using the ML method in the *fitdistrplus* and *Mass* R packages (Delignette-Muller and Dutang, 2015; Venables and Ripley, 2002). The obtained parameters are presented in Fig. 4, and their median values were used in the flood generator to create arid/semi-arid flood records with the desired characteristics.

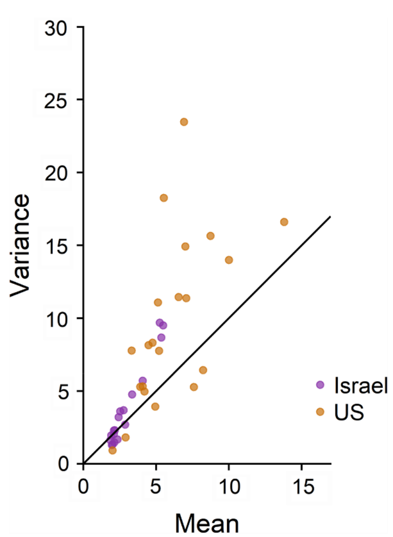

Fig. 3. Variance vs mean of annual number of flood occurrences in the US and Israel stations. The line depicts the 1:1 relation. Points above the line indicate over-dispersion.

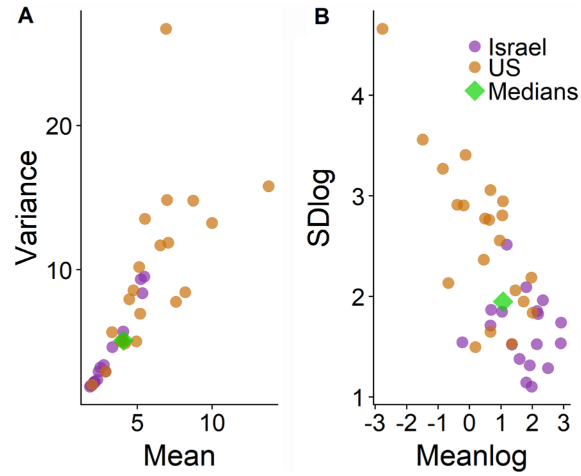

Fig. 4. Parameters of the negative binomial distributions describing the number of yearly floods (A) and of the lognormal distribution describing the floods intensity (B) of all the arid/semi-arid stations. Green diamonds show the medians of the scatter.

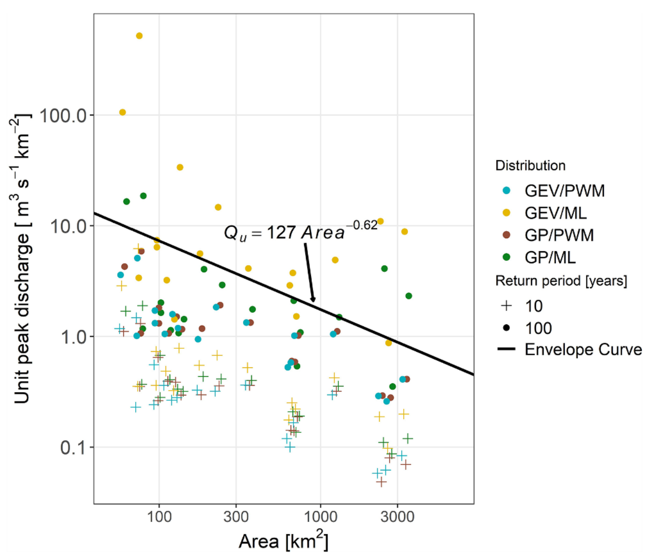

Fig. 5. Unit peak discharges for 10- and 100-year return period estimated using different distributions and parameter estimation method vs drainage area for all arid/semi-arid analyzed stations in Israel. The envelope curve for the eastern Mediterranean (Tarolli et al., 2012) is depicted by the black line.

4. Results

4.1. *FFA in arid/semi-arid regions*

Fig. 5 shows the 10- and 100-year flood unit peak discharge estimated using ML and PWM over arid/semi-arid watersheds in Israel as a function of the catchment area. The regional envelope curve for eastern Mediterranean region presented by Tarolli et al. (2012) is superimposed: by construction, envelope curves provide an estimate of the regional upper bound to the unit peak discharges to be expected as a function of the watershed area (Enzel et al., 1993). If the data record used to build the curves is complete, only particularly rare floods are expected to exceed the envelope curves. The figure shows that, in many cases, 100-year floods estimated using ML substantially exceed the regional envelope curve. This indicates that ML provides unreliable estimates for our dataset. Thus, we will focus hereafter on the PWM estimation method.

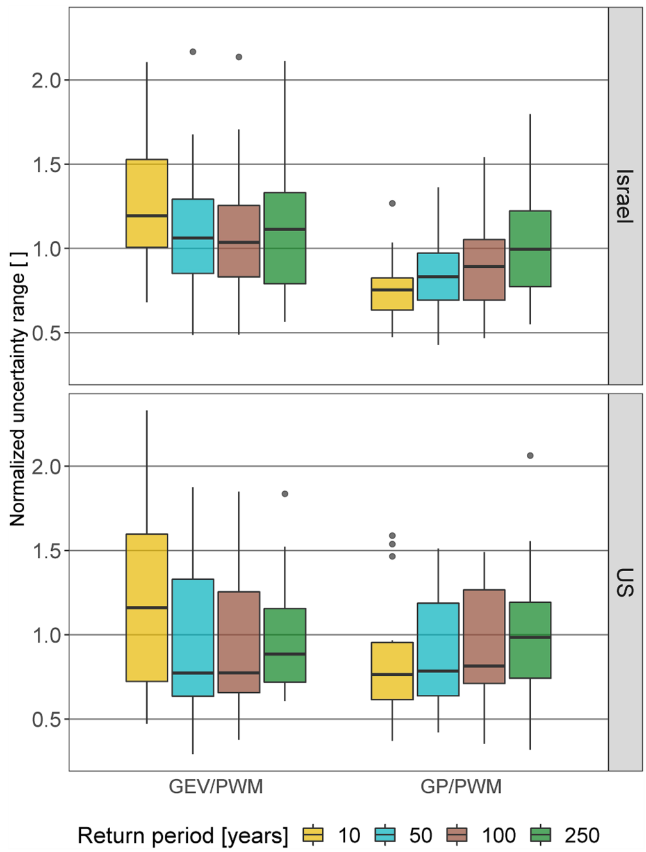

Fig. 6. Boxplots of the 95% uncertainty range (i.e., the difference between the 97.5 and the 2.5% quantiles) of the peak discharge for different return periods divided by the estimated peak discharge for the different distributions and estimation methods for all the arid/semi-arid stations. The black line in each boxplot marks the median, the boxes lower and upper borders mark the 25 and 75% quartiles, respectively, the whiskers mark the minimum and maximum values unless these values exceed 1.5 · IQR (inter quartile range – the distance between lower and upper quartiles) and the dark points mark the outliers.

Across the regions, the normalized uncertainty of lower quantiles (e.g., 10 years) is smaller for GP than for GEV (Fig. 6). The normalized uncertainty for Israel is larger than that of the US for the GEV distribution and for larger quantiles. This may be attributed to the shorter records available for stations in Israel (medians are 34.5 and 50 years for Israel and US, respectively). The normalized uncertainty tends to increase with return period for GP, while no consistent pattern is observed for GEV (Fig. 6) with the 10-year normalized uncertainty being larger than the 50-year. It is worth noting that the general behavior of the uncertainties for the two approaches is consistent among the two study regions, instilling confidence in the robustness of the results. Fig. 7 also shows how the PDS approach generally provides smaller uncertainty, in particular for shorter return periods (10- and 50-year), while less advantages are observed for longer return periods (100- and 250-year), particularly in the US. Given the recent debate on the topic (e.g., Hu et al., 2019; Marra et al., 2018; Schlögl and Laaha, 2016), we test the common assumption that the PDS approach is preferable over the AMS due to the larger amount of data points used in the former for the parameters estimation (Ben-Zvi, 2016; Castillo and Hadi, 1997; Morrison and Smith, 2002), focusing on arid/semi-arid watersheds. To do so, we examined whether PDS estimated values are within the uncertainty range of AMS estimates and vice versa (Fig. 7). All values were normalized by dividing by the estimated PDS (GP) peak discharge (Fig. 7A) and AMS (GEV) estimated peak discharge (Fig. 7B). Almost in all cases the estimated values of the GP or GEV are within the uncertainty range of the competing distribution with the only exception of the US, in which the longer return periods (e.g. 50- and 100-year) of the GP estimation is sometimes comparable to the upper GEV uncertainty bound.

A similar analysis is performed based on the synthetic flood record. A long synthetic flood record is created (10^5 years) to represent the population, 1000 series of 60 years (similar to real site record lengths) are sampled and their flood peak discharge, and uncertainty is estimated using GEV/PWM and GP/PWM (Fig. 8). In this case, the reference discharge for a given quantile can be empirically derived from the population. Results here show that GP slightly outperforms GEV. It is however worth noting how both approaches underestimate the reference peak discharge: this aspect will be addressed in detail in the following.

4.2. Comparison of arid/semi-arid vs. Mediterranean watersheds

In this section the arid/semi-arid climate watersheds are compared with the Mediterranean climate watersheds. Owing to the type of available data, only the AMS approach is used for such comparison; the results above, however, showed comparable behavior among the two approaches. PWM is chosen for the parameter estimation due to the superior performance highlighted above.

Fig. 9 presents the distributions of the tail ratios in the different regions and climates. The distributions of the tail ratio for the GEV/PWM were compared using the Welch test (Welch, 1951) followed by the ad hoc Games-Howell test (Games and Howell, 1976) for pairwise comparison. Both tests are specifically suitable for heteroscedastic samples with unequal sizes. Heteroscedasticity was determined using the Levene and Bartlett tests (Levene, 1960; Snedecor and Cochran, 1989). All differences were found to be significant, with one interesting exception: no significant difference is observed between Israeli Mediterranean watersheds and arid US watersheds. This should not come as a surprise given that the climatic classification considers many factors (including temperature and seasonality) while other factors, such as the mean annual precipitation amounts, are comparable in the two areas.

The distribution of unit peak discharge for different return periods and climatic regions is presented in Fig. 10A. The unit peak discharge for Mediterranean regions is larger for shorter return periods (i.e. 10–50 years) but peak discharge of arid/semi-arid floods of 100 and 250-year return periods are found to exceed the corresponding Mediterranean floods in the US. In addition, the ratio of the 250-year to the 10-year flood was examined. This ratio can be used to compare the upper tail across watersheds in a similar manner to the upper tail ratio that is used in Smith et al. (2018). The mean ratio for the arid/semi-arid watersheds (6.84 for Israel, 7.06 for the US) is larger than the one in Mediterranean watersheds (5.04 for Israel, 2.68 for the US).

A comparison of the normalized uncertainty ranges across the analyzed climates and regions indicates that the median normalized uncertainty range is larger for the arid/semi-arid watersheds (Fig. 10B). In particular, the uncertainty range of the US-Mediterranean watersheds is significantly smaller than the other climate-regions. This is probably due to the longer record length (median of 74 years as compared to 34.5, 50, and 42 of arid/semi-arid Israel, arid/semi-arid US and Mediterranean Israel regions, respectively), even if the possibility that GEV and GP are better suited to these watersheds than arid ones, cannot be excluded.

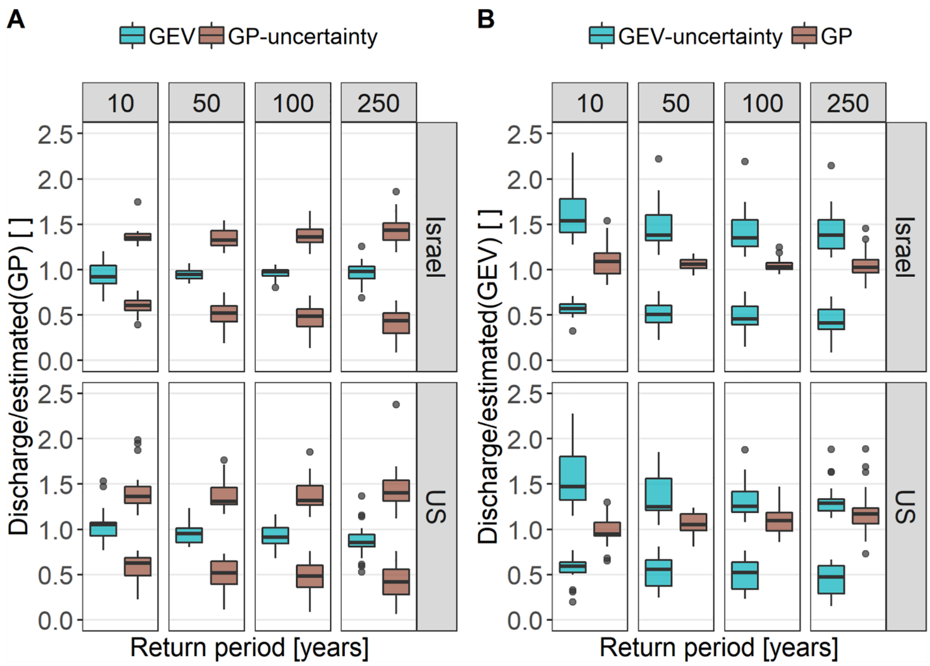

Fig. 7. A) Boxplots (all stations) of GEV estimate (cyan) and GP estimated uncertainty bounds (2.5 and 97.5% quantiles, brown), normalized over the GP estimate, for different return periods for the two regions. B) Same as A but showing GP estimate compared to GEV uncertainty bounds, normalized over the GEV estimate. The PWM parameter estimation method was used in all cases.

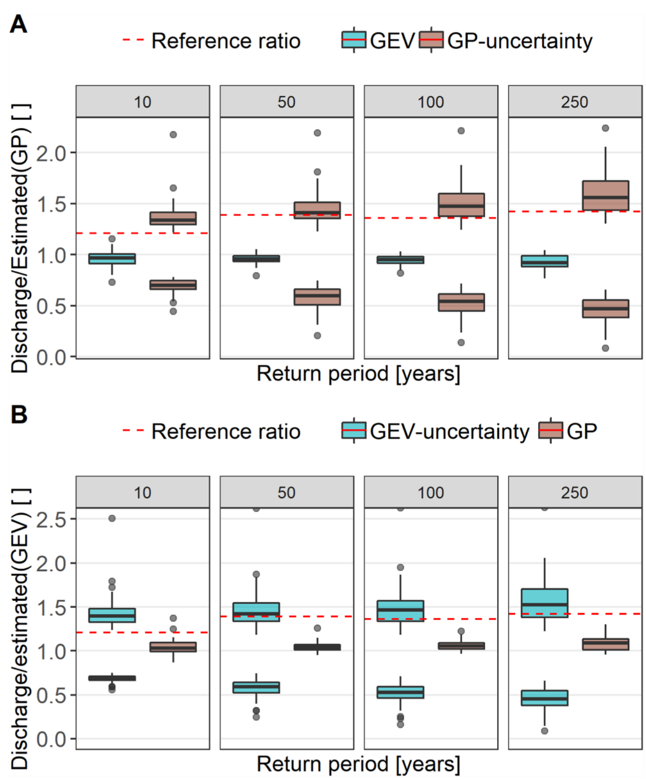

Fig. 8. Similar to Fig. 9 but based on the synthetic flood generator data. Dashed red lines represent the reference peak discharge divided by the median estimated discharge of the GP in A and by the median estimated discharge of the GEV in B.

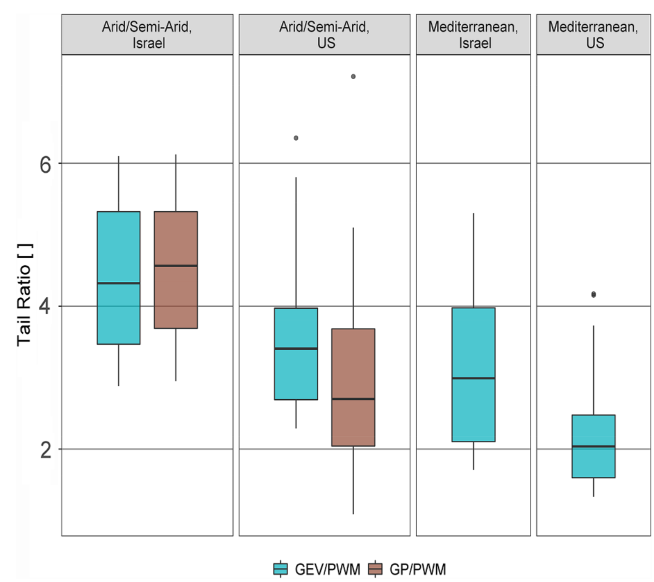

Fig. 9. Comparison of the tail ratio unit peak discharge between climates (Arid/Semi-Arid and Humid) and regions in Israel (top) and US (right) for GEV-PWM and GP-PWM.

4.3. Sensitivity to the characteristics of arid/semi-arid flood records

Fig. 11 shows the uncertainty in the GEV/PWM tail ratio estimated from record between 15 years and the full length for stations that have at least 50-year record. As expected, uncertainty range is negatively correlated with record length and its rate of change decreases with record length.

Second, we explore the effects of record length on the estimation of quantiles and parameters using the synthetic flood generator. Records of 15- to 500-year long were created and the 100-year flood magnitude was estimated for both GEV and GP. For this analysis, the reference 100-year flood is known from the full 10^5-year synthetic record and the bias, defined as the ratio between the estimated and reference peak discharge, is computed (Fig. 12A). Both distributions tend to underestimate the 100-year flood (i.e. a value of 1 is equal to no bias) for all record lengths. Longer records provide less biased estimates, but a clear

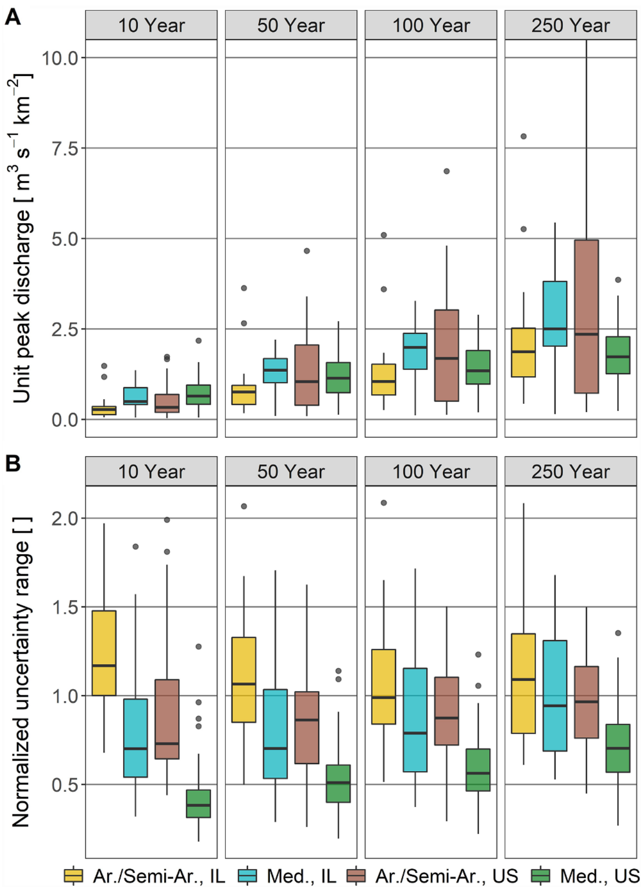

Fig. 10. A) Unit peak discharge as a function of return period for different climates (Arid and Semi-arid, Ar./Semi-Ar., and Mediterranean, Med.) and regions (Israel, IL, and United States, US). Estimation is done by the GEV/PWM. B) Uncertainty ranges of the unit peak discharges normalized by the estimated discharge for the GEV/PWM for different return periods for the different climates and regions.

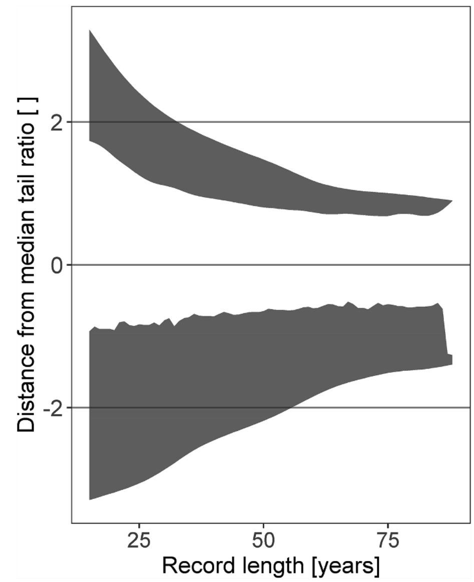

Fig. 11. Uncertainty of the shape parameter vs record length. The distance of the 2.5% and 97.5% uncertainty range curves from the median shape parameter is presented for all the stations in both the regions that have a record longer than 50 years.

convergence to the reference value is not reported even for 500-year records. This non-convergence of the estimates to the record is counter-intuitive and could be related to the fact that, in the arid watersheds of Israel, it is quite common to have one, two or even no floods in a given year (Table A.1) while for the arid/semi-arid watersheds in the US the number of floods in a year may exceed ten, but still remaining limited (Table A.2 and Fig. 12D, bottom).

In order to quantify the extent of the estimation error, we created different synthetic series with different mean number of flood events per year by altering the mean parameter of the negative binomial distribution describing the number of yearly floods. The log-normal distribution describing the individual floods intensities was kept unchanged. Fig. 12D presents the bias as a function of the mean number of events per year. The distributions of the mean number of annual floods for all stations in the study are given at the bottom of Fig. 12D (grouped by climate). As expected, the bias for all the distributions and estimation methods decreases when the mean number of events per year increases, and converges to 1 for 20–30 floods per year. Both distributions tend to underestimate the reference value by up to ~20% for less than ~5 events per year.

The occurrence of years with no flow is a common feature in the arid/semi-arid watersheds in Israel (in the US stations all years had at least one flood event). The fraction of years with no flows ranges up to 37%, with a mean of 15.3% in the Israeli arid watersheds. Using the synthetic flood generator, we quantified the effect of no flow years on FFA of the GEV distribution (Fig. 12B). In general, underestimation of the 50- and 100-year return periods is observed as a consequence of years with no-flows. The underestimation worsens as the percent of years with no flow increases but it is most notable in the extreme case of 50% years with no flow.

Another typical feature of flood discharge in arid/semi-arid regions is its large variability (Chouaib et al., 2018; Nicholson, 2011). This translates to a larger coefficient of variation (CV) of flood peak discharges than in humid regions (Salinas et al., 2014) (Fig. 12C, bottom). In order to examine what is affected by this, the accuracy of 100-year flood estimation, we produced synthetic records representing different CVs by alterations of the synthetic generator lognormal distribution and leaving the binomial distribution describing the number of yearly floods unchanged. The distributions of the CV for all stations in the study are given at the bottom of Fig. 12C (grouped by climate). As expected, the relative error increases as CV increases. Both distributions underestimate the 100-year flood for large CVs (e.g. for CV = 5 the GEV/PWM is 0.91 of the reference flood and the GP/PWM is 0.93 of the reference flood) but for a CV smaller than 3 the estimation is quite accurate. GP seems the more accurate for most CV values.

5. Discussion

Rainfall frequency studies have shown evidences of a heavier tail in arid/semi-arid regions (Ben-Zvi, 2009; Marra et al., 2017; Marra and Morin, 2015; Morin et al., 2020) and few FFA studies have shown some degree of inheritance in the flood tail characteristics (Farquharson et al., 1992; Smith et al., 2014; Zaman et al., 2012). We found the tail of flood distributions in arid/semi-arid watersheds to be significantly heavier than in Mediterranean watersheds of the same region (Fig. 9). This confirms what is reported for rainfall, and may be attributed to the climatic and physical characteristics of arid/semi-arid watersheds such

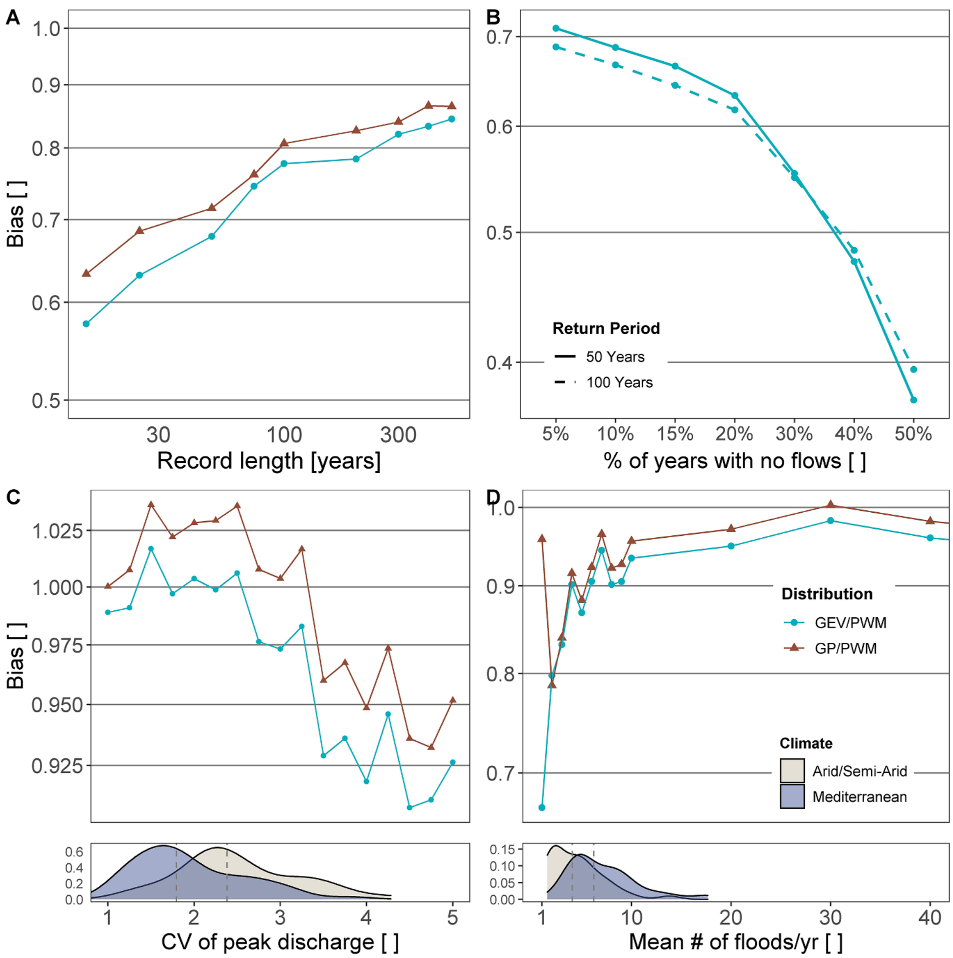

Fig. 12. A) Bias of the 100-year flood with respect to the reference 100-year flood for different approaches vs record length. The 100-year reference flood is obtained empirically from the synthetic record. B) Bias of the 50-year and 100-year floods for the GEV/PWM distribution for different percentages of no flow years of total years on record. C) Bias of 100-year flood vs the coefficient of variation values of the log-normal distribution describing the flood peaks. The density plot at the bottom shows the distribution of the coefficient of variation in all the stations in the study (grouped by climate). The dashed lines are the medians of the distributions. D) Bias of the 100-year flood vs the mean number of flood events per year. The density plot at the bottom shows the distribution of the mean number of events in all the stations in the study (grouped by climate).

as high energy, high intensity convective rainfall events (Camarasa-Belmonte and Soriano, 2014; Zaman et al., 2012), low infiltration capacities (Morin and Benyamini, 1977), and to the small scale of the examined watersheds, more likely to respond to the more skewed distributions of short duration rainfall (Marra and Morin, 2015) and of the synoptic events more characterizing the arid portions of the region (Marra et al., 2019). At the same time, statistical reasons related to the smaller number of floods per year are likely to contribute to this, and cannot be excluded a priori. Our results confirm the vital importance of quantification and communication of the uncertainty in the estimated quantiles: in arid/semi-arid watersheds, the uncertainty range may be as large as the estimated quantiles, surprisingly also for relatively short return periods (Fig. 6).

It is a common belief that, in presence of short records, the PDS approach should outperform the AMS thanks to the larger amount of data used for the parameter estimation (Bezak et al., 2014; Nagy et al., 2017). This however comes at the price of methodological complexities (threshold estimation; separation of independent floods), which might lead to biases in FFA. In arid areas, owing to the high variability of flood peak discharge, a simple threshold estimation method must be adopted. We found similar performance between GEV and GP in terms of uncertainty estimation (Fig. 7; Fig. 8), confirming previous results by Hu et al. (2019), Marra et al. (2018) and Schlögl and Laaha (2016).

Results from this study show how parameter estimation by means of ML in arid/semi-arid watersheds often leads to unrealistic values (Fig. 5). This confirms previous findings in which PWM was shown to outperform ML in the estimation of parameters from short data records (e.g., Martins and Stedinger, 2000; Hosking and Wallis, 1987; Dupuis, 1999) and stresses the particular importance of parameter estimation method for arid/semi-arid regions.

GEV and GP seem to provide poorer estimates in arid areas, due to the small number of floods per year. In addition, since the lognormal distribution is in the domain of attraction of the Gumbel distribution (i.e., the GEV with shape parameter equal to zero), but this convergence is very slow, the fact that flood peaks in these areas were found to follow such distribution, an extremely large number of floods per year are required to reach the asymptotic GEV.

More specifically, (i) the limited mean number of floods per year (Fig. 12B and D), and (ii) the large variability of flood peak discharge (Fig. 12C), are found to significantly affect peak discharge estimates. The PDS approach is found to perform slightly better since the use of a threshold is likely to remove low flows and no-flows from data series used for the parameter estimation. Regionalization approaches could be attempted to decrease the observed uncertainties, even if the lack of data makes it more difficult to define homogeneous regions in arid areas, with respect to other regions. Alternative estimation methods, such as the mixed method by Morrison and Smith (2002) as well as recently developed methodologies for extreme value analysis might be able to overcome the limitations of traditional methods in arid regions. The Metastatistical Extreme Value framework (Marani and Ignaccolo, 2015), which allows to relax the asymptotic assumption on the number of floods per year and vastly increase the data sample thus reducing the issues related to short record lengths (Marra et al., 2018; Zorzetto et al., 2016), was recently tested to model extreme flood peak discharge in a large sample of catchments in the contiguous Unites States and found to outperform traditional methods in more than 75% of the cases (Miniussi et al., 2020). More specifically, its simplified version (Marra et al., 2019) promises improved estimates of the flood peak discharge distribution in presence of limited number of floods in each year.

6. Conclusions

This study presents a comprehensive evaluation of the use of the generalized extreme value (GEV) and generalized Pareto (GP) distributions flood frequency analysis in arid/semi-arid watersheds. Special attention is given to the impact of the peculiar characteristics of these watersheds, namely (i) the limited number of floods per year, and (ii) the large variability of flood peak discharges. Study cases in Israel and the US are examined together with control watersheds characterized by Mediterranean climate and synthetic flood records reproducing the characteristics of arid/semi-arid records. The main findings of the study are as follows:

- For the examined data records, the tail ratio of extreme floods in arid/semi-arid watersheds is larger than the one describing Mediterranean watersheds, indicating a heavier tail and confirming previous observations on short-duration rainfall
- The mean number of floods per year and the coefficient of variation of flood peak discharge have a crucial impact on the accuracy of the estimates: (a) smaller number of events per year and (b) larger coefficient of variation of the flood peak discharge distributions are related to large errors in the estimated quantiles, irrespective of the adopted approach (annual maxima vs. partial duration series).
- The use of partial duration series approach provides slightly improved estimates because the use of a threshold reduces the amount of low flows and no-flows in the data series. However, no significant advantage is observed in term of estimation uncertainty: estimates obtained using one of the two approaches are consistently within the other's uncertainty range. From a practical point of view, due to the increase complexity, partial duration series should be preferred only after a careful examination of the characteristics of individual case of interest.
- Arid watersheds are further from the asymptotic behavior than catchments in more humid hydro-climatic conditions, and the GEV and GP distributions provided poorer estimates of extreme quantiles in arid than in Mediterranean. This problem was enhanced when maximum likelihood methods were used for the parameter estimation

In arid/semi-arid regions, the challenges faced by practitioners when assessing design floods through flood frequency analysis are exacerbated. In general, findings from this study, indicate that in arid/semi-arid watersheds maximal likelihood estimation method should not be used even in presence of long data records, the partial duration series approach should not be blindly preferred over the annual maxima series approach, and the uncertainty in the estimated quantiles should not be overlooked.

CRediT authorship contribution statement

Asher Metzger: Conceptualization, Data curation, Methodology, Visualization, Writing - original draft, Writing - review & editing, Formal analysis. **Francesco Marra:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **James A. Smith:** Writing - original draft, Writing - review & editing, Funding acquisition. **Efrat Morin:** Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was funded by the NSF-BSF (grant BSF 2016953), by the Israel Science Foundation (grant 1069/18), by the BARD project (grant IS-5124-18) and by a Google gift grant. This study is a contribution to the PALEX project Paleohydrology and Extreme Floods from the Dead Sea ICDP Core and is a contribution to the HyMeX program. Data for this paper is from the US geological survey (freely available at https://waterdata.usgs.gov/nwis/rt) and the Israel hydrological service.

Appendix

Table A.1
Israeli arid/semi-arid hydrometric stations.

Station no.	Name	Record length (yr)	Watershed area (km^2)	Mean annual rain (mm yr^{-1})	Mean # of floods yr^{-1}	CV of flood magnitude
23,106	Besor – Nizana Road	51	179	50–100	2.1	1.5
23,134	Beqa – Be'er-Sheva	29	97	100–200	5.2	3.3
23,135	Beqa – Be'er Sheva	37	97	100–200	5.4	2.1
23,137	Be'er-Sheva – Hazerim	36	1234	200–300	5.5	2.6
23,150	Besor – Re'im	38	2643	100–200	2.5	1.9
23,160	Gerar – Re'im	40	645	200–300	4.1	2.3
48,125	Darga	25	75	200–300	2.8	1.6
48,130	Teqoa	26	135	300–350	2.3	2.1
48,185	Rahaf	23	75	50–100	2.4	1.9
48,192	Hemar – Downstream the Cliff	22	357	50–100	2.9	1.9
55,106	Zin Upper – Avedat	25	124	50–100	2.2	2.8
55,110	Zin – Waterfall	57	232	50–100	1.8	2.5
55,140	Zin – Masos	35	674	50–100	1.9	2.4
55,165	Mamshit	37	59	50–100	3.3	2.0
56,140	Ramon	32	111	50–100	2.1	1.2
56,150	Neqarot – Upper	28	708	50–100	2.0	2.6
57,130	Paran – Western Border	21	2350	0–50	2.1	2.2
57,165	Paran – The Bottleneck	52	3324	0–50	1.9	2.3

Table A.2
US arid/semi-arid hydrometric stations.

Station No.	Name	Continuous record length (yr)	Annual peak record length (yr)	Watershed area (km^2)	Mean annual rain (mm yr^{-1})	Mean # of floods yr^{-1}	CV of flood magnitude
8,128,000	S Concho Rv At Christoval TX.	19	88	1070	400–500	5.5	4.3
8,131,400	Pecan Ck Nr San Angelo TX.	14	43	210	200–300	2.0	2.5
8,194,200	San Casimiro Ck Nr Freer TX.	27	57	1215	400–500	7.0	3.4
8,447,020	Independence Ck Nr Sheffield TX.	17	28	1976	400–500	4.9	3.7
9,471,380	Upper Babocomari River Near Huachuca City AZ.	19	16	408	500–600	4.5	2.4
9,480,500	Santa Cruz River Near Nogales AZ.	33	86	1380	500–600	4.8	2.2
9,481,740	Santa Cruz River At Tubac AZ.	24	19	3134	500–600	13.8	2.3
9,482,000	Santa Cruz River At Continental AZ.	26	72	4356	400–500	6.5	3.4
9,484,000	Sabino Creek Near Tucson AZ.	31	84	92	400–500	5.2	3.3
9,484,550	Cienega Creek Near Sonoita AZ.	18	14	513	400–500	10.0	3.1
9,484,600	Pantano Wash Near Vail AZ.	30	58	1184	700–750	8.7	2.2
9,485,000	Rincon Creek Near Tucson AZ.	31	63	116	400–500	3.9	2.2
9,486,800	Altar Wash Near Three Points AZ.	27	38	1199	500–600	7.1	1.4
9,487,000	Brawley Wash Near Three Points AZ.	24	43	2010	400–500	6.9	2.2
9,497,980	Cherry Creek Near Globe AZ.	32	50	518	400–500	7.6	3.0
9,499,000	Tonto Creek Above Gun Creek Near Roosevelt AZ.	32	75	1748	600–650	8.2	3.7
9,508,300	Wet Bottom Creek Near Childs AZ.	31	48	94	600–650	4.1	2.5
9,510,200	Sycamore Creek Near Fort Mcdowell AZ.	32	56	425	500–600	5.1	2.9
9,512,280	Cave Creek Blw Cottonwood Cr Near Cave Creek AZ.	31	35	214	500–600	4.2	2.7
9,513,780	New River Near Rock Springs AZ.	29	54	177	400–500	2.9	3.5
9,517,490	Centennial Wash At Southern Pacific Railroad Brdg	28	30	4705	400–500	3.3	2.1

Table A.3
Israeli Mediterranean hydrometric stations.

Station no.	Name	Record length (yr)	Watershed area (km^2)	Mean annual rain (mm yr^{-1})	Mean # of floods yr^{-1}	CV of flood magnitude
8126	Nahalal	26	41	500–600	6.5	1.9
8130	Ha'shofet – Hazore'a	32	12	600–700	13.0	3.0
8140	Bet Lehem	48	22	600–700	10.1	2.1
8146	Qishon – The Quarry	52	695	600–700	9.5	2.4
8155	Zippori – Tel Alil	18	246	500–600	12.0	1.6
12,130	Daliyya – Bat Shelomo	51	42	600–700	5.7	2.2
12,140	Daliyya – Tel Aviv-Haifa Road	42	69	500–600	10.0	2.7
13,105	Tanninim – 'Ammiqam	43	51	600–700	6.2	2.0
13,125	Ada – Giv'at Ada	47	18	600–700	7.8	3.0
13,135	Barqan – Kefar Glickson	46	29	500–600	8.2	2.0
15,120	Alexander – Elyashiv	47	488	500–600	9.3	1.9
31,155	Meshushim – Dardara	44	160	400–500	14.6	2.7
31,160	Yehudiya – Bet Zayda Road	22	81	400–500	8.9	1.9
31,163	Daliyot – Bet Zayda Road	22	109	400–500	8.3	2.7
31,165	Samak – 200 M Elevation	25	101	400–500	9.6	2.4
38,175	Harod – Bet She'an,Near 90 Road	21	181	400–500	9.5	1.6

Table A.4
US Mediterranean hydrometric stations.

Station no.	Name	Record length (yr)	Watershed area (km^2)	Mean annual rain (mm yr^{-1})	Mean # of floods yr^{-1}	CV of flood magnitude
11,335,000	Cosumnes R A Michigan Bar Ca	109	1388	500–600	3.6	2.6
11,336,580	Morrison C Nr Sacramento Ca	46	138	500–600	17.7	1.5
11,336,585	Laguna C Nr Elk Grove Ca	20	83	500–600	13.5	3.2
11,342,000	Sacramento R A Delta Ca	71	1101	1900–2000	5.3	1.6
11,355,500	Hat C Nr Hat Creek Ca	69	420	800–900	2.0	0.2
11,372,000	Clear C Nr Igo Ca	75	591	1300–1400	7.9	1.2
11,374,000	Cow C Nr Millville Ca	67	1101	800–900	5.8	1.3
11,376,000	Cottonwood C Nr Cottonwood Ca	75	2401	700–800	3.9	1.7
11,376,550	Battle C Bl Coleman Fish Hatchery Nr Cottonwood Ca	53	925	700–800	7.8	1.1
11,379,500	Elder C Nr Paskenta Ca	67	239	600–700	4.5	2.0
11,381,500	Mill C Nr Los Molinos Ca	87	339	900–1100	8.2	1.3
11,383,500	Deer C Nr Vina Ca	99	539	800–900	5.7	1.6

(*continued on next page*)

Table A.4 (*continued*)

Station no.	Name	Record length (yr)	Watershed area (km^2)	Mean annual rain (mm yr^{-1})	Mean # of floods yr^{-1}	CV of flood magnitude
11,390,000	Butte C Nr Chico Ca	85	381	1300–1400	5.7	1.5
11,418,500	Deer C Nr Smartsville Ca	81	219	1000–1100	7.7	2.4
11,421,000	Yuba R Nr Marysville Ca	72	3468	600–700	3.6	1.5
11,424,000	Bear R Nr Wheatland Ca	87	756	500–600	5.1	2.8
11,427,000	Nf American R A North Fork Dam Ca	74	886	1000–1100	4.7	1.9
11,446,500	American R A Fair Oaks Ca	111	4890	500–600	3.5	1.3
11,447,360	Arcade C Nr Del Paso Heights Ca	35	81	500–600	17.4	1.7
11,449,500	Kelsey C Nr Kelseyville Ca	69	95	1000–1100	4.5	1.7
11,451,000	Cache C Nr Lower Lake Ca	71	1368	900–1000	4.1	1.7
11,451,100	Nf Cache C A Hough Spring Nr Clearlake Oaks Ca	44	156	1000–1100	3.9	1.8
11,451,300	Nf Cache C Nr Clearlake Oaks Ca	26	313	700–800	4.5	2.3
11,451,715	Bear C Ab Holsten Chimney Cyn Nr Rumsey Ca	18	246	600–700	3.3	1.8
11,452,500	Cache C A Yolo Ca	113	2950	500–600	7.0	3.1
11,454,000	Putah C Nr Winters Ca	85	1487	700–800	5.2	1.6
11,461,000	Russian R Nr Ukiah Ca	65	259	1000–1100	3.8	1.6
11,461,500	Ef Russian R Nr Calpella Ca	74	239	1000–1100	6.8	1.2
11,462,500	Russian R Nr Hopland Ca	77	938	1000–1100	4.8	1.2
11,519,500	Scott R Nr Fort Jones Ca	74	1691	700–800	4.1	1.7
11,522,500	Salmon R A Somes Bar Ca	92	1945	1100–1200	5.8	1.1
14,357,500	Bear Creek At Medford, Or	98	749	500–600	11.2	1.9
14,362,000	Applegate River Near Copper, Or	77	583	700–800	4.8	1.1
14,362,250	Star Gulch Near Ruch, Or	32	41	700–800	4.7	2.7
14,366,000	Applegate River Near Applegate, Or	78	1251	700–800	4.5	1.4

References

Armon, M., Dente, E., Smith, J.A., Enzel, Y., Morin, E., 2018. Synoptic-scale control over modern rainfall and flood patterns in the Levant drylands with implications for past climates. J. Hydrometeorol.

Ashkar, F., Ba, I., 2017. Selection between the generalized Pareto and kappa distributions in peaks-over-threshold hydrological frequency modelling. Hydrol. Sci. J. 62, 1167–1180. https://doi.org/10.1080/02626667.2017.1302089.

Ben-Zvi, A., 2016. Selecting series size where the generalized Pareto distribution best fits. J. Hydrol. 541, 778–786. https://doi.org/10.1016/j.jhydrol.2016.07.038.

Ben-Zvi, A., 2009. Rainfall intensity-duration-frequency relationships derived from large partial duration series. J. Hydrol. 367, 104–114. https://doi.org/10.1016/j.jhydrol.2009.01.007.

Bezak, N., Brilly, M., Šraj, M., 2014. Comparaison entre les méthodes de dépassement de seuil et du maximum annuel pour les analyses de fréquence des crues. Hydrol. Sci. J. 59, 959–977. https://doi.org/10.1080/02626667.2013.831174.

Camarasa-Belmonte, A.M., Soriano, J., 2014. Empirical study of extreme rainfall intensity in a semi-arid environment at different time scales. J. Arid Environ. 100 (101), 63–71. https://doi.org/10.1016/j.jaridenv.2013.10.008.

Castillo, E., Hadi, A.S., 1997. Fitting the Generalized Pareto Distribution to Data. J. Am. Stat. Assoc. 92, 1609–1620. https://doi.org/10.1080/01621459.1997.10473683.

Chouaib, W., Caldwell, P.V., Alila, Y., 2018. Regional variation of flow duration curves in the eastern United States: process-based analyses of the interaction between climate and landscape properties. J. Hydrol. 559, 327–346. https://doi.org/10.1016/j.jhydrol.2018.01.037.

Cohn, T.A., England, J.F., Berenbrock, C.E., Mason, R.R., Stedinger, J.R., Lamontagne, J.R., 2013. A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series. Water Resour. Res. 49, 5047–5058. https://doi.org/10.1002/wrcr.20392.

Coles, S., 2001. An introduction to statistical modeling of extreme values. Springer Series Statist. https://doi.org/10.1007/978-1-4471-3675-0.

Costa, J.E., Jarrett, R.D., 2008. An evaluation of selected extraordinary floods in the United States reported by the US Geological Survey and implications for future advancement of flood science.

Cunnane, C., 1979. A note on the Poisson assumption in partial duration series models. Water Resour. Res. 15, 489–494. https://doi.org/10.1029/WR015i002p00489.

Davison, A.C., Smith, R.L., Journal, S., Statistical, R., Series, S., 1990. Models for exceedances over high thresholds published by: wiley for the royal statistical society. Wiley R. Stat. Soc. 52, 393–442.

Delignette-Muller, M.L., Dutang, C., 2015. fitdistrplus: an R Package for Fitting Distributions. J. Stat. Softw. 64.

Dupuis, D.J., 1999. Exceedances over high thresholds: a guide to threshold selection. Extremes 1, 251–261. https://doi.org/10.1023/A:1009914915709.

Enzel, Y., Ely, L.L., House, P.K., Baker, V.R., Webb, R.H., 1993. Paleoflood evidence for a natural upper bound to flood magnitudes in the Colorado River Basin. Water Resour. Res. 29, 2287–2297. https://doi.org/10.1029/93WR00411.

Farquharson, F.A.K., Meigh, J.R., Sutcliffe, J.V., 1992. Regional flood frequency analysis in arid and semi-arid areas. J. Hydrol. 138, 487–501. https://doi.org/10.1016/0022-1694(92)90132-F.

Ferreira, B.Y.A., De Haan, L., 2015. On the block maxima method in extreme value theory: PWM estimators. Ann. Stat. 43, 276–298. https://doi.org/10.1214/14-AOS1280.

Fisher, R.A., Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the large sample. Math. Proc. Cambridge Philos. Soc. 24, 180–190.

François, B., Schlef, K.E., Wi, S., Brown, C.M., 2019. Design considerations for riverine floods in a changing climate – a review. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2019.04.068.

Fukutome, S., Liniger, M.A., Süveges, M., 2015. Automatic threshold and run parameter selection: a climatology for extreme hourly precipitation in Switzerland. Theor. Appl. Climatol. 120, 403–416. https://doi.org/10.1007/s00704-014-1180-5.

Games, P. a, Howell, J.F., 1976. Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study. J. Educ. Stat. 1, 113–125.

Gnedenko, B., 1943. Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. Ann. Math. 44, 423–453. https://doi.org/10.2307/1968974.

Greenbaum, N., Ben-Zvi, A., Haviv, I., Enzel, Y., 2006. The hydrology and paleohydrology of the Dead Sea tributaries. Geol. Soc. Am. Spec. Pap. 401, 63–93. https://doi.org/10.1130/2006.2401(05).

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R., 1979. Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Water Resour. Res. 15, 6. https://doi.org/10.1029/WR015i005p01049.

Haan, C.T., 2002. Statistical methods in hydrology. Iowa State Press.

Hall, M.J., van den Boogaard, H.F.P., Fernando, R.C., Mynett, a. E., 2004. The construction of confidence intervals for frequency analysis using resampling techniques. Hydrol. Earth Syst. Sci. 8, 235–246. https://doi.org/10.5194/hess-8-235-2004.

Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized pareto distribution. Technometrics 29, 339. https://doi.org/10.2307/1269343.

Hosking, J.R.M., Wallis, J.R., Woo, E.F., 1985. Estimation of the generalized extreme value distribution by the method of probability weighted moments. Technometrics 27, 251–261. https://doi.org/10.1080/00401706.1985.10488049.

Hu, L., Nikolopoulos, E.I., Marra, F., Anagnostou, E.N., 2019. Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods: an evaluation over the contiguous US. J. Flood Risk Manage. 1 (13). https://doi.org/10.1111/jfr3.12580.

Huang, J., Yu, H., Guan, X., Wang, G., Guo, R., 2016. Accelerated dryland expansion under climate change. Nat. Clim. Change. 6, 166–171. https://doi.org/10.1038/nclimate2837.

Kahana, R., Ziv, B., Enzel, Y., Dayan, U., 2002. Synoptic climatology of major floods in the Negev Desert. Israel. Int. J. Climatol. 22, 867–882. https://doi.org/10.1002/joc.766.

Katz, R., 2002. Statistics of extremes in climatology and hydrology. Adv. Water Resour. 25, 1287–1304.

Knighton, A.D., Nanson, G.C., 2001. An event-based approach to the hydrology of arid zone rivers in the channel country of Australia. J. Hydrol. 254, 102–123. https://doi.org/10.1016/S0022-1694(01)00498-X.

Lang, M., Ouarda, T.B.M.J., B, 1999. Bobe e, Towards operational guidelines for over-threshold modeling, J. Hydrol 225, 103–117.

Levene, H., 1960. Contributions to probability and statistics. Essays Honor Harold Hotell. 278–292.
Marani, M., Ignaccolo, M., 2015. A metastatistical approach to rainfall extremes. Adv. Water Resour. 79, 121–126. https://doi.org/10.1016/j.advwatres.2015.03.001.
Marra, F., Morin, E., 2015. Use of radar QPE for the derivation of Intensity-Duration-Frequency curves in a range of climatic regimes. J. Hydrol. 531, 427–440. https://doi.org/10.1016/j.jhydrol.2015.08.064.
Marra, F., Morin, E., Peleg, N., Mei, Y., Anagnostou, E.N., 2017. Intensity-duration-frequency curves from remote sensing rainfall estimates: comparing satellite and weather radar over the eastern Mediterranean. Hydrol. Earth Syst. Sci. 21, 2389–2404. https://doi.org/10.5194/hess-21-2389-2017.
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. Adv. Water Resour. 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. Adv. Water Resour. 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 Resour. Res. 36, 737–744. https://doi.org/10.1029/1999WR900330.
Meirovich, L., Ben-Zvi, A., Shentsis, I., Yanovich, E., 1998. Frequency and magnitude of runoff events in the arid Negev of Israel. J. Hydrol. 207, 204–219. https://doi.org/10.1016/S0022-1694(98)00135-8.
Miniussi, A., Marani, M., Villarini, G., 2020. Metastatistical Extreme Value Distribution applied to floods across the continental United States. Adv. Water Resour. 136, 103498. https://doi.org/10.1016/j.advwatres.2019.103498.
Morin, E., Marra, F., Armon, M., 2020. Dryland precipitation climatology from satellite observations, in: Levizzani, V., Kidd, C., Kirschbaum, D., Kummerow, C., Kummerov, K., Turk (Eds), F. (Eds.), Satellite Precipitation Measurement. Springer. https://doi.org/10.1007/978-3-030-24568-9.
Morin, J., Benyamini, Y., 1977. Rainfall Infiltration into Bare Soils. Water Resour. Res. 13, 813–817.
Morrison, J.E., Smith, J.A., 2002. Stochastic modeling of flood peaks using the generalized extreme value distribution. Water Resour. Res. 38, 41-1-41–12. https://doi.org/10.1029/2001WR000502.
Nagy, B.K., Mohssen, M., Hughey, K.F.D., 2017. Flood frequency analysis for a braided river catchment in New Zealand: comparing annual maximum and partial duration series with varying record lengths. J. Hydrol. 547, 365–374. https://doi.org/10.1016/j.jhydrol.2017.02.001.
Nicholson, S.E., 2011. Dryland climatology. Dryland Climatology. https://doi.org/10.1017/CBO9780511973840.
Overeem, A., Buishand, A., Holleman, I., 2008. Rainfall depth-duration-frequency curves and their uncertainties. J. Hydrol. 348, 124–134. https://doi.org/10.1016/j.jhydrol.2007.09.044.
Papalexiou, S.M., AghaKouchak, A., Foufoula-Georgiou, E., 2018. A diagnostic framework for understanding climatology of tails of hourly precipitation extremes in the United States. Water Resour. Res. 54, 6725–6738. https://doi.org/10.1029/2018WR022732.
Papalexiou, S.M., Koutsoyiannis, D., 2013. Battle of extreme value distributions : a global survey on extreme daily rainfall. Water Resour. Res. 49, 187–201. https://doi.org/10.1029/2012WR012557.
Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644. https://doi.org/10.5194/hess-11-1633-2007.
Rahman, A.S., Rahman, A., Zaman, M.A., Haddad, K., Ahsan, A., Imteaz, M., 2013. A study on selection of probability distributions for at-site flood frequency analysis in Australia. Nat. Hazards 69, 1803–1813. https://doi.org/10.1007/s11069-013-0775-y.
Ryberg, K.R., Goree, B.B., Williams-Sether, T., Mason, R.R., J., 2017. The U . S . Geological Survey Peak-Flow File Data Verification Scientific Investigations Report 2017 – 5119. https://doi.org/10.31333/sir20175119.
Salinas, J.L., Castellarin, A., Kohnová, S., Kjeldsen, T.R., 2014. Regional parent flood frequency distributions in Europe – Part 2: climate and scale controls. Hydrol. Earth Syst. Sci. 18, 4391–4401. https://doi.org/10.5194/hess-18-4391-2014.
Scarrott, C., Macdonald, A., 2012. a Review of extreme value threshold estimation and uncertainty quantification. REVSTAT – Stat. J. 10, 33–60.
Schlögl, M., Laaha, G., 2016. Extreme weather exposure identification for road networks – a comparative assessment of statistical methods. Nat. Hazards Earth Syst. Sci. Discuss. 1 (26). https://doi.org/10.5194/nhess-2016-373.
Schumer, R., Knust, A., Boyle, D.P., 2014. Characteristics of ephemeral hydrographs in the Southwestern United States. J. Hydrol. Eng. 19, 10–17. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000643.
Serinaldi, F., Kilsby, C.G., 2015. Stationarity is undead: uncertainty dominates the distribution of extremes. Adv. Water Resour. 77, 17–36. https://doi.org/10.1016/j.advwatres.2014.12.013.
Serinaldi, F., Kilsby, C.G., 2014. Rainfall extremes: toward reconciliation after the battle of distributions. Water Resour. Res. 50, 336–352. https://doi.org/10.1002/2013WR014211.
Shentsis, I., Meirovich, L., Ben-Zvi, A., Rosenthal, E., 1999. Assessment of transmission losses and groundwater recharge from runoff events in a wadi under shortage of data on lateral inflow, Negev, Israel. Hydrol. Process. 13, 1649–1663.
Sillitto, G.P., 1951. Interrelations between certain linear systematic statistics of samples from any continuous population. Biometrika 38, 377–382. https://doi.org/10.2307/2332583.
Smith, A., Sampson, C., Bates, P., 2014. Regional flood frequency analysis at the global scale. Water Resour. Res. 51, 539–553. https://doi.org/10.1002/2014WR015814.
Smith, J.A., Baeck, M.L., Yang, L., Signell, J., Morin, E., Goodrich, D.C., 2019. The paroxysmal precipitation of the desert: flash floods in the Southwestern United States. Water Resour. Res.
Smith, J.A., Cox, A.A., Baeck, M.L., Yang, L., Bates, P., 2018. Strange floods: the upper tail of flood peaks in the United States. Water Resour. Res. 54, 6510–6542. https://doi.org/10.1029/2018WR022539.
Snedecor, G.W., Cochran, W.G., 1989. Statistical methods, 8thEdn. Ames Iowa State Univ. Press Iowa 503.
Solari, S., Egüen, M., Polo, M.J., Losada, M.A., 2017. Peaks Over Threshold (POT): a methodology for automatic threshold estimation using goodness of fit p-value. Water Resour. Res. 53, 2833–2849. https://doi.org/10.1002/2016WR019426.
Tarolli, P., Borga, M., Morin, E., Delrieu, G., 2012. Analysis of flash flood regimes in the North-Western and South-Eastern Mediterranean regions. Nat. Hazards Earth Syst. Sci. 12, 1255–1265. https://doi.org/10.5194/nhess-12-1255-2012.
Turnipseed, D.P., Sauer, V.B., 2010. Discharge measurements at gaging stations: US Geological Survey Techniques and Methods, book 3, chap.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Statistics and Computing. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-21706-2.
Villarini, G., Smith, J.A., 2010. Flood peak distributions for the eastern United States. Water Resour. Res. 46, 1–17. https://doi.org/10.1029/2009WR008395.
Welch, B.L., 1951. On the comparison of several mean values: an alternative approach. Biometrika 38, 330. https://doi.org/10.2307/2332579.
Wuertz, D., 2006. Rmetrics - Extreme Financial Market Data Author 44.
Zaman, M.A., Rahman, A., Haddad, K., 2012. Regional flood frequency analysis in arid regions: a case study for Australia. J. Hydrol. 475, 74–83. https://doi.org/10.1016/j.jhydrol.2012.08.054.
Zhang, X., Duan, K., Dong, Q., 2019. Comparison of nonstationary models in analyzing bivariate flood frequency at the Three Gorges Dam. J. Hydrol. 579, 124208. https://doi.org/10.1016/j.jhydrol.2019.124208.
Zoccatelli, D., Marra, F., Armon, M., Rinat, Y., Smith, J.A., Morin, E., 2019. Contrasting rainfall-runoff characteristics of floods in desert and Mediterranean basins. Hydrol. Earth Syst. Sci. 23, 2665–2678. https://doi.org/10.5194/hess-23-2665-2019.
Zorzetto, E., Botter, G., Marani, M., 2016. Supplementary information to : On the emergence of rainfall extremes from ordinary events. Geophys. Res. Lett. 43, 8076–8082. https://doi.org/10.1002/2016GL069445.