UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova
Dipartimento di Ingegneria Civile, Edile e Ambientale (ICEA)

Scuola di Dottorato di Ricerca in
Scienze dell'Ingegneria Civile e Ambientale
XXXII ciclo

The Metastatistical Extreme Value Distribution for Rainfall and Flood Frequency Analysis with External Drivers

Supervisore: Ch.mo Prof. Marco Marani

Direttore della Scuola: Ch.mo Prof. Marco Marani

Dottorando: Arianna Miniussi

Il presente progetto di Dottorato è stato finanziato dalla Fondazione CARIPARO

Anno Accademico 2019-2020

Sommario

Una stima accurata degli estremi idrologici è fondamentale per le sue molteplici implicazioni nella progettazione ingegneristica, per la quantificazione e mappatura di eventi di piena, nell'industria assicurativa, a supporto decisionale. I metodi tradizionali, di cui la distribuzione Generalizzata dei Valori Estremi (GEV) è il riferimento, si basano su assunzioni spesso sottovalutate o che non vengono verificate, non sfruttano in maniera efficace i dati disponibili e sono inadatti per tenere conto della variabilità inter-annuale. Con l'obiettivo di perfezionare la stima di estremi con elevato tempo di ritorno, questa tesi si concentra sulla Distribuzione Metastatistica dei Valori Estremi (MEVD), un approccio introdotto per superare alcuni dei limiti della teoria tradizionale. Il lavoro qui proposto dapprima analizza la definizione di MEVD ottimale in funzione di fattori climatici locali e delle proprietà statistiche della precipitazione alla scala giornaliera. La conclusione che viene tratta attribuisce alla variabilità inter-annuale delle proprietà statistiche della precipitazione un ruolo chiave per la definizione della finestra di stima ottimale. Nella maggior parte dei casi considerati, ad eccezione di climi molto secchi, la finestra di stima dovrebbe essere mantenuta il più breve possibile per risolvere la variabilità delle distribuzioni tra gli anni. L'utilizzo di finestre brevi rende la MEVD adatta allo studio di estremi in un clima che cambia, permettendole di risolvere la variabilità inter-annuale. Finora, la MEVD è stata applicata principalmente alla precipitazione (a scala giornaliera e oraria). Qui, per la prima volta, la MEVD viene utilizzata per studiare valori di portata, sviluppando un'analisi di frequenza delle piene implementata utilizzando la MEVD su serie di picchi di portata negli Stati Uniti Continentali. Inoltre, viene valutato l'impatto di El Niño Southern Oscillation (ENSO) sui regimi di piena. Rispetto alla GEV, la MEVD fornisce stime più accurate in $\sim$76% delle stazioni analizzate, con una significativa riduzione dell'errore di stima soprattutto per tempi di ritorno molto più elevati della dimensione del campione utilizzato per la stima dei parametri. Invece, la distinzione dei picchi a seconda delle fasi di ENSO ha portato a miglioramenti trascurabili nella stima di piene estreme. Infine, sfruttando l'interessante proprietà della MEVD di includere distribuzioni miste nella sua formulazione in maniera naturale, diverse aree metropolitane americane sono state scelte come casi studio per l'applicazione di una MEVD che distingue tra pioggia ciclonica e non ciclonica. L'effetto dei cicloni tropicali sulla precipitazione è ben riconoscibile, e l'utilizzo di un approccio MEVD misto è risultato vantaggioso in alcuni casi. A causa dell'effetto prolungato nel tempo che i cicloni tropicali hanno sulla precipitazione, il vantaggio della MEVD mista nella riduzione dell'incertezza di stima è maggiore quando si considerano valori di pioggia cumulati su finestre temporali più lunghe del giorno singolo.

Abstract

An accurate estimation of hydrologic extremes is fundamental for its many implications on engineering design, flood quantification and mapping, insurance and re-insurance purposes, policy-making. Traditional methods, hinging on the Generalized Extreme Value (GEV) distribution, are founded on often-overlooked and untested assumptions, make an ineffective use of the available data, and are ill-suited for accounting for inter-annual variability. With the aim of improving the estimation accuracy of high return period extremes, this dissertation focuses on the Metastatistical Extreme Value Distribution (MEVD), an approach introduced to relax some of the limitations of the traditional Extreme Value Theory. The present work first analyzes the definition of the optimal MEVD formulation as a function of local climatic factors and of key statistical properties of rainfall at the daily scale. It concludes that the inter-annual variability of rainfall statistical properties plays an important role in the definition of the optimal time window to be used for parameter estimation. In the largest amount of cases examined, except for very dry climates, with few rainy days, the analysis window should be kept to the minimum of 1 year in order to resolve the time variability of the distributions. The use of short time windows also makes the MEVD a suitable approach to study extremes in a changing climate, as it contributes to its ability to resolve inter-annual variability. Up to now, the MEVD has been applied mainly to rainfall (at the daily and hourly scale). Here, for the first time, the MEVD is used to study streamflow data, developing a flood frequency analysis MEVD-based on series of flow peaks in the Continental United States. Moreover, the impact of El Niño Southern Oscillation (ENSO) on flood regimes is evaluated. In the comparison with the GEV, results show the outperformance of the MEVD in $\sim$76% of the analyzed stations, with a significant reduction in the estimation error especially when considering return periods much higher than the size of the sample used to estimate the distributional parameters. Yet, a negligible improvement in the estimation of extreme floods was found when stratifying peaks according to ENSO phases. In the end, leveraging the appealing property of the MEVD to naturally include mixtures of distributions in its formulation, a MEVD that distinguishes between non-Tropical Cyclones (TCs) and Tropical Cyclones-induced rainfall is applied to several American metropolitan areas. The impact of TCs on rainfall is well distinguishable, and the use of a mixed MEVD approach resulted beneficial in several cases. Its advantage in the reduction of the estimation error when compared to the single-distribution MEVD was found to be more significant when considering cumulative values of rainfall over consecutive days, due to the prolonged impact TCs have on rainfall over time.

Contents

List of Figures

List of Acronyms

Acronym	Full name
AM	Annual maxima
CONUS	Continental United States
DI	Dispersion Index
ENSO	El Niño Southern Oscillation
EV	Extreme Values
EVT	Extreme Value Theory
EW	Estimation window
FFA	Flood frequency analysis
FSE	Fractional Standard Error
GEV	Generalized Extreme Value
GHCN	Global Historical Climate Network
GPD	Generalized Pareto Distribution
MEVD	Metastatistical Extreme Value Distribution
NHC	National Hurricane Center
NOAA	National Oceanic and Atmospheric Administration
N_r	number of realizations
POT	Peak Over Threshold
PR	Puerto Rico
PWM	Probability Weighted Moments
SS	Skill Score
T	return period
TC	Tropical Cyclone
TRMM	Tropical Rainfall Measuring Mission
U.S.	United States
USGS	United States Geological Survey
USWRC	United States Water Resources Council

Chapter 1

Introduction and motivation of the work

Extreme events in water and Earth sciences are of fundamental importance in several fields, such as engineering design, insurance and re-insurance, environmental and urban planning, policy-making, etc. Their accurate quantitative estimation, i.e. accurately linking the magnitude of an extreme event and its frequency of occurrence, remains difficult, partly because extremes are, by definition, poorly sampled, partly due to the changing dynamics of the Earth system.
The traditional Extreme Value Theory (EVT) is grounded in assumptions that limit their application and, due to their ineffective use of observations, are ill-suited for taking into account the inter-annual variability that characterizes historical records.

The traditional statistical theory of extremes was developed and applied in the most diverse areas, e.g. sea water level and storm surges (Coles (2001); Haigh et al. (2010); Muis et al. (2016)), air quality (Roberts (1979); Tobías and Scotto (2005); Martins et al. (2017)), droughts (Fleig et al. (2006); Sousa et al. (2011)), earthquakes (Pisarenko and Sornette (2003); Pisarenko et al. (2014)), wind speed in hurricanes (Coles and Casson (1998); Heckert et al. (1998)), traffic (O'Connor and O'Brien (2005); Zhou et al. (2016)).
With a more specific focus on rainfall, precipitation extremes have been widely studied, e.g. for their implication on floods (see for example Smith and Baeck (2015); Zhang and Villarini (2017)) and flash floods (Rebora et al., 2013), on agriculture (Rosenzweig et al. (2002); Powell and Reinhard (2015)), on water resources (Dettinger et al., 2011), on sewage systems and stormwater infrastructure (Rosenberg et al., 2010), for being responsible of the spread of waterborne diseases (Curriero et al. (2001); Cann et al. (2013)), for triggering landslides (D'Odorico and Fagherazzi, 2003).

To provide a quantitative perspective on the relevance of the topic, Figure 1.1 (from Wallemacq and House (2018)) shows the number of disasters by major category (climate-related and geophysical, upper panel) and by type (lower panel), at the global scale in the period 1998-2017. Climate/water-

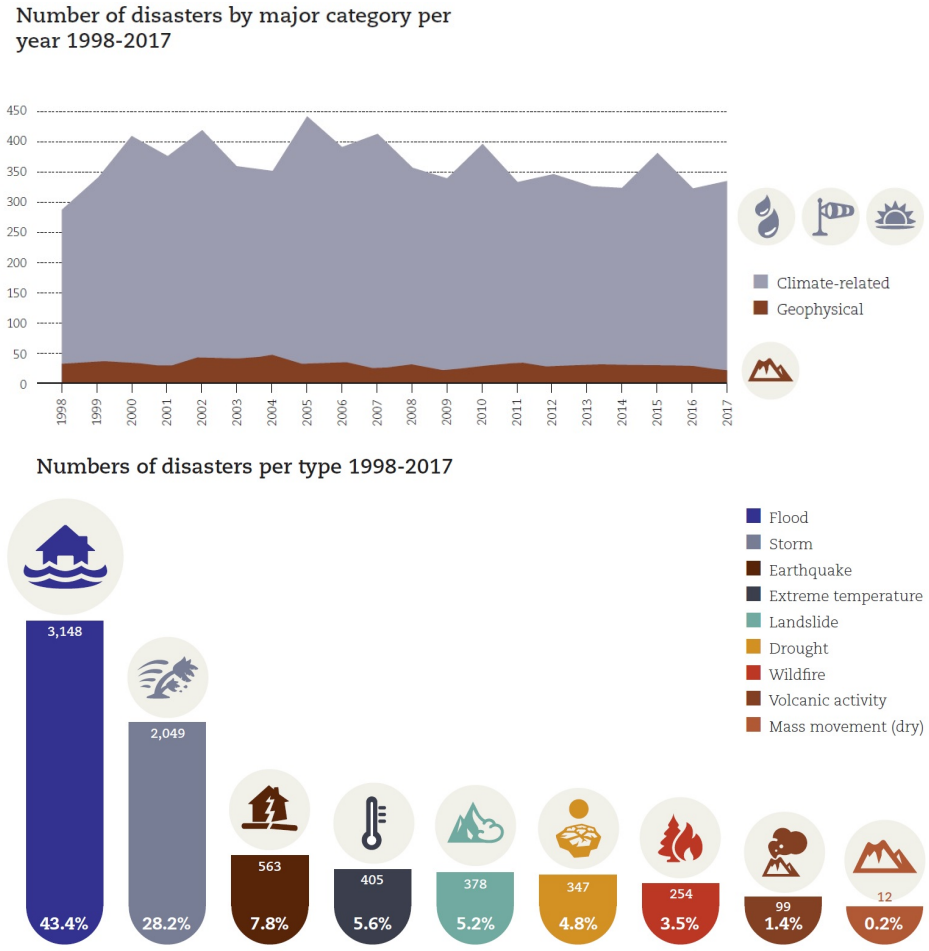


Figure 1.1: Global number of disasters per major category (upper panel) and type (lower panel) in the period 1998-2017 (from: Wallemacq and House (2018)).

related events, despite not as deadly as geophysical ones, still cause an unacceptable large number of fatalities every year (Figure 1.2). Moreover, globally they cause most of the economic losses (Figure 1.3), hence emphasizing the need of refining the existing methods or developing new ones, in order to most accurate determine their frequency and magnitude.

The return period (T), or average recurrence time of an event (see Chapter 2), is a key concept when dealing with extremes, as it is used in many regulations of engineering mitigation and adaptation measures, and summarizes in a single and readily communicated (though often misunderstood) quantity the frequency of occurrence of events exceeding an assigned magnitude. Water structures are designed for events corresponding to return periods that vary according to national regulations, but orders of magnitudes are

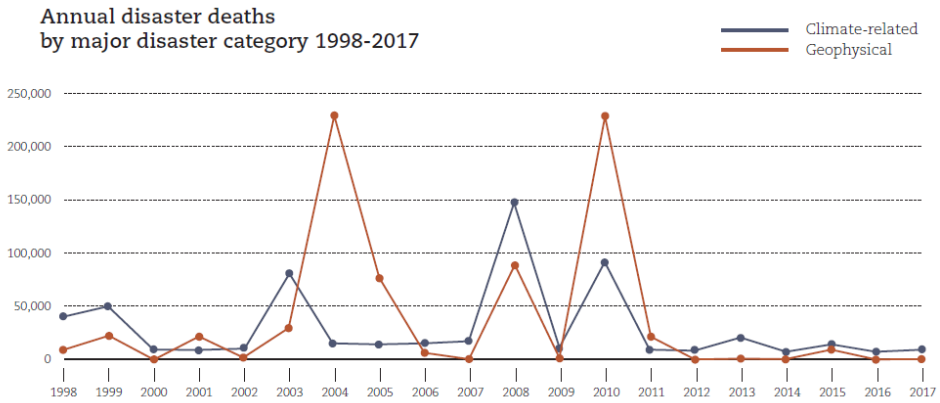


Figure 1.2: Deaths due to climate-related (blue line) and geophysical disasters (brown line) (from: Wallemacq and House (2018)).

similar for several different countries. For example, emergency spillways of a concrete dam are designed with reference to a flow peak discharge of the order of T=1000 years, for a earth-fill dam T=3000 years; typical values of T for river levels in the case of levees are near 200 years, for bridge piers around 100 years. These values of T are much higher than the length of the available observational time series, such that the estimation of the corresponding events is very uncertain.

Realistic sizes of observational records are some tens of years, exceptionally they can reach 100 years or more. This means that the value of design variable that needs to be estimated has a return period several times higher with respect to the length of the time series (see Figure 1.4). If one empirically associates to the maximum value in a time series a return period equal to the series length, this situation means that extreme value estimation is an extrapolation exercise well beyond the range of the observations (blue area in Figure 1.4). Hence, the estimation can only be performed with the help of a statistical model bridging the gap between observations and target extremes. As a result, the key question is which statistical model is most appropriate to describe the quantity of interest in the "extrapolation range". An important, though often overlooked issue, is that models that exhibit acceptable goodness of fit within the observational range, may lead to large estimation errors in the extrapolation range, for large return periods.

Recently, a new statistical approach that aims to improve over the traditional Extreme Value Theory by relaxing some of its founding hypotheses, and with a view to considering non-stationary processes, has been introduced, the Metastatistical Extreme Value Distribution (MEVD) (Marani and Ignaccolo, 2015). This approach uses most of the available observations to better constrain the shape of the distributional tail in the extrapolation

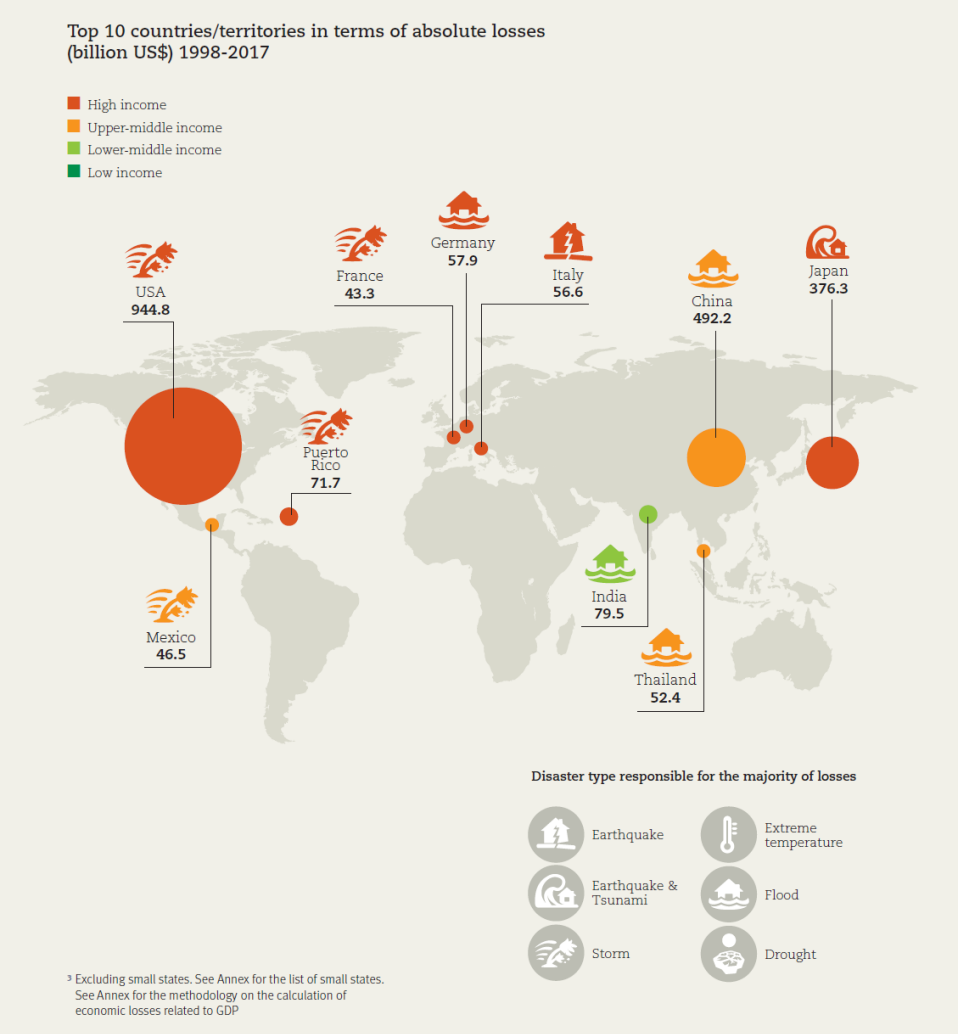


Figure 1.3: Top 10 countries for economic losses, due to the disaster type that is responsible of them (from: Wallemacq and House (2018)).

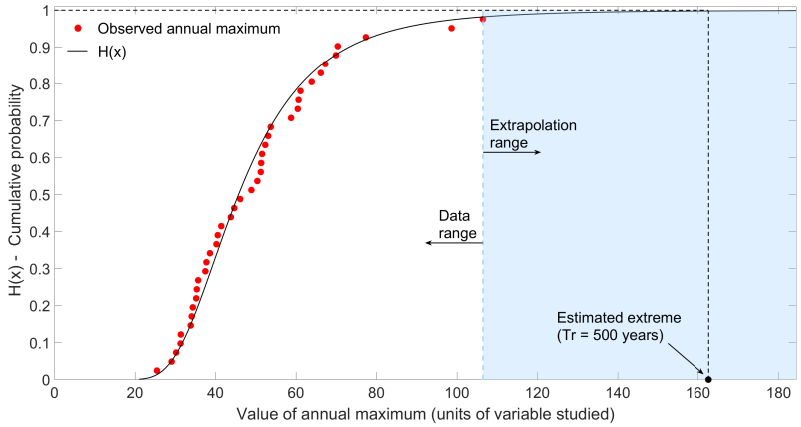


Figure 1.4: Visual explanation of the issues emerging when estimating values associated to high return periods, using historical series limited in time. Red dots represent annual maxima (extremes, following the definition of the EVT), the black line is an EV distribution fitted on data. The data range (white area) and the extrapolation range (blue area) are, respectively, the range in which data are available and the one in which a statistical model is needed to estimate values characterized by a return time higher than the one associated to the highest observed maximum. The target value that needs to be estimated is represented as a black dot on the x-axis.

range, and will be described in some detail in the next Chapter.

The MEVD approach proposed in this thesis is developed for at point analyses. Regional approaches, that are necessary in ungaged areas, are not explored here. However, the MEVD formulations studied in the present dissertation are suitable candidates for analyses in ungaged sites, since they would not spatially interpolate directly the distribution of maxima, which is affected by high uncertainty. A spatial interpolation of the ordinary values (i.e., all the observations, regardless their degree of extraordinariness), indeed, is likely to be more robust, thanks to the larger amount of information available. Even though the topic is of big interest, it does not fit this thesis purposes, leaving room for future work.
One of the limitation of the traditional EVT are the issues about data availability, and specifically short time series. Nowadays, developing technologies, e.g. satellites and radar as well as more innovative ones that leverage daily-use devices, such as cell phone links (van het Schip et al., 2017) or smartphones (Allamano et al., 2015) are increasing the information available, complementing the one provided by gages. Two points of strengths can be stressed about the MEVD approach with respect to the classical methods for extremes estimation: 1) it was shown to outperform the traditional EVT in estimating high return period quantiles also when only short series of data are available (with an error reduction up to the 50%, see Zorzetto et al. (2016)), and 2) after an appropriate downscaling, it can be used for

better estimate extremes from satellite data (for example, those from the Tropical Measurement Mission, which consists of only 18 years of data, see Zorzetto and Marani (2019)). The above-mentioned advantages make therefore the MEVD approach the ideal candidate for estimating extremes when only short time series are available or in areas where gages are scarce and the most reliable information is provided by satellites.
Undoubtedly, the advantage brought by the MEVD is lost when the observational records include only annual maxima, which were the common data historically recorded. It is yet worth noting that much effort is made in the direction of including historical information from geological formations, stratigraphy, sediments or from archives (also using supervised/unsupervised deep learning techniques, e.g., Nayman et al. (2019)), or to extract data from where annual maxima were then derived. The joint work of hydrologists, geomorphologists and historians to reconstruct long-term records (of extreme floods, for examples, as in Sheffer et al. (2003)) can provide valuable information that can complement the one already available.

The work presented here expands the formulation of the MEVD and explores its application to a variety of water extremes, including daily rainfall and flood peak discharge, focusing on a practically robust application of the MEVD and on incorporating the possible simultaneous presence of different physical mechanisms generating events with different statistical characteristics. Hence, this dissertation, after briefly recalling the traditional EVT and the MEVD framework, starts from first refining the application methodology of the MEVD approach, focusing on the optimal size of the window that should be used for parameter estimation with reference to daily rainfall extremes. Afterwards, it develops a MEVD tailored for flood frequency analysis and, with the aim of reintroducing into statistical practice the explicit consideration of the physical mechanisms that generate hydrological processes, it proposes a mixed MEVD to take into account the effects of El Niño Southern Oscillation (ENSO) on flood regimes over the Continental United States (CONUS). The final Chapter describes an approach that leverages the appealing property of the MEVD to naturally include mixtures of distributions to distinguish among different rainfall types to analyze the role of Atlantic Tropical Cyclones (TCs) in generating extreme rainfall.

Chapter 2

Extreme Value Theory

The Extreme Value Theory (EVT) defines extremes as *block maxima*, i.e. the maximum value, x, among all ("ordinary") values s_i $(i = 1, \ldots, n)$ occurred within a period/block of time of prescribed duration (usually 1 year). Under the assumption that the n ordinary events in each block are independent and identically distributed (i.i.d.) according to the same cumulative distribution $F(s)$, the cumulative distribution of block maxima is $H_n(x) = [F(x)]^n$.
In this context, the probability that the block maximum exceeds a value x in each year, $p = 1 - H_n(x)$, can be used to express the return period, as $T = 1/p = 1/(1 - H_n(x))$ (where time is expressed in multiples of the block length, which will be taken to be 1 year hereafter). The return period, T, of a specified event magnitude s is a key concept when dealing with extreme values. Here, T will be used in its definition as the average time between two successive exceedances of s (see Volpi et al. (2015) for a good discussion of possible definitions of T).
When the return period of interest is much smaller than the length (say, S) of the observational time series, the quantile corresponding to T can be estimated directly from observations, e.g. by ordering yearly maxima in ascending order and by substituting the cumulative probability with its frequency estimation, such that $T_k = 1/(1 - H_k)$ (where $H_k = k/(S + 1)$ is the cumulative frequency corresponding to x_k, the *k-th* value in the ordered list of yearly maxima). Volpi et al. (2019) recently proposed a procedure, called Complete Time-series Analysis, that allows extreme value analyses using all available data (as opposed to just annual maxima, as in the traditional approach) when the return period of interest is $T <= S$.

When $T > S$ it is necessary to adopt a parametric expression for $H_n(x)$, to be fitted to observations, in order to extrapolate quantile estimates beyond the range of observed values. The choice of a probability distribution $H_n(x)$, and the methods used to estimate its parameters, define the accuracy with which high quantiles can be estimated, formally as $x_T = H_n^{-1}(1 - 1/T)$.

The quantification of this accuracy is of central interest both theoretically

and for practical applications (Castillo, 1988). The hydrological literature traditionally addresses this problem through goodness-of-fit measures, comparing the performance of different candidate distributions in describing observed samples (e.g., Papalexiou and Koutsoyiannis (2013)). However, goodness-of-fit metrics do not reflect the accuracy with which the return period of an extreme that was not yet observed in the past can be estimated. This thesis will instead focus on cross-validation approaches, that provide out-of-sample estimates of uncertainty, which more accurately describe predictive uncertainty.

When $n \to \infty$:

$$\lim_{n\to\infty}[F(x)]^n = \begin{cases} 1, & \text{if } F(x) = 1 \\ 0, & \text{if } F(x) < 1 \end{cases}$$

which means that the limit distribution degenerates (it takes only values 0 and 1).
In order to avoid this degeneration problem, a linear transformation is applied:

$$Y = a_n + b_n x \tag{2.1}$$

where a_n and b_n are appropriate constants depending on n such that the limit distribution

$$\lim_{n\to\infty} H_n(a_n + b_n x) = \lim_{n\to\infty} F^n(a_n + b_n x) = H_n(x) \tag{2.2}$$

becomes non-degenerated.

2.1 Traditional Extreme Value Theory

According to the Three Types Theorem (Fisher and Tippett (1928); Gnedenko (1943); Gumbel (2004)), the only three types of non-degenerated distributions satisfying Eq.(2.2) are:

1. the Frechet distribution
2. the Gumbel distribution
3. the reverse-Weibull distribution

These three limit distributions are unified in the Generalized Extreme Value (GEV) distribution (Von Mises (1936); Coles (2001)) depending on the value of its shape parameter ξ:

$$H(x;\mu,\psi,\xi) = \exp\Big\{ -\big[1 + \xi\frac{(x-\mu)}{\psi}\big]_+^{-\frac{1}{\xi}} \Big\} \tag{2.3}$$

where $\{x : 1 + \xi(x - \mu)/\psi\} > 0$, $\mu \in (-\infty, \infty)$ is the location parameter, $\psi > 0$ is the scale parameter and $\xi \in (-\infty, \infty)$ is the shape parameter. For $\xi > 0$, $\xi = 0$ and $\xi < 0$ the GEV distribution leads itself back to the Frechet, the Gumbel and the reverse-Weibull distributions, respectively.

As the threshold increases, the Pickands-Balkema-de Haan theorem (Balkema and de Haan (1974); Pickands (1975)) states that the distribution of the exceedances over a set threshold u (peaks over threshold, POT) follows a Generalized Pareto Distribution (GPD):

$$F(x) = 1 - \big(1 + \xi \frac{x - u}{\sigma}\big)^{-\frac{1}{\xi}} \tag{2.4}$$

where ξ and σ are the shape and the scale parameters, respectively.
If the number of exceedances over the threshold u in any one year follows a Poisson distribution with mean λ (the so called Poisson-GPD model), this model leads again to the GEV distribution for annual maxima (Davison and Smith, 1990).
Serinaldi and Kilsby (2014) analyzed two global datasets of daily rainfall and found that as the threshold increases, the variance of the shape parameter of the GPD reduces with the increasing record length and the mean values tend to be positive, which is linked to a heavy tail behavior (if $\xi < 0$, instead, the distribution presents an upper-end point).

The interested reader is addressed to Katz et al. (2002) for a review of the statistical methods for extremes in hydrology and climatology and to Papalexiou et al. (2013) for a recent overview of the history of the EVT.

While the traditional EVT is quite appealing, as it provides a conceptual justification for the use of a well-defined distributional form (the GEV distribution), the assumptions on which it is grounded are rather specific. For example, the justification for assuming a large number of events/year (ideally tending to infinity) is dubious, such that is in practice impossible to determine whether the actual distribution of extremes is close to an asymptotic GEV form (see for example Cook and Harris (2004); Koutsoyiannis (004a)). This circumstance makes it difficult to determine whether the distribution of extremes presents a heavy tail, i.e. a tail that decays algebraically slow, with significant conceptual and practical implications (Koutsoyiannis (004b); Papalexiou and Koutsoyiannis (2013); Papalexiou et al. (2013)). The assumptions underlying the POT approach also present limitations, as the hypotheses of a Poisson arrival of exceedances and of GP-distributed magnitudes of the exceedances strictly hold for a "large enough" threshold and under the same asymptotic assumptions on n as for the derivation of GEV (Pickands, 1975).
In practice, the theoretical basis outlined above greatly complicates the selection of a suitable threshold, which must simultaneously be large enough

for the approach to hold, and low enough to retain a sufficient number of events/year, questioning the effectiveness of the POT approach. In addition to the above theoretical limitations, the traditional EVT does not make good use of the available information and effectively bases extreme value estimates on a small fraction of the available observations (annual maxima or a few values over a high threshold). This is a missed opportunity in the search of optimizing quantile estimation uncertainty, which calls for minimizing uncertainty given the available observations, or for minimizing the length of the observational time series needed to achieve an acceptable level of uncertainty. The latter objective is of great interest in the presence of non-stationarity (WMO, 2009), when a stochastic process may be assumed to be piece-wise stationary, provided that sufficiently short subsets of the observational time window are considered. In the limit, one would like to be able to track variations of T-year extremes from one year to the next. The traditional EVT precludes a time series study of this type, since GEV parameters cannot be estimated on a single yearly maximum or on the few exceedances over a high threshold occurring in a year (unless a specific parametric time variability, or co-variability with explanatory variable, of GEV parameters is assumed, thereby partly prescribing the time-variations which are being sought). The estimation of GEV parameters on sliding windows with length $S \sim$ a few years is also not viable, as GEV estimation uncertainty grows quickly as $T >> S$ (Zorzetto et al., 2016).

In order to overcome the problem of data wastage in the estimation of extremes, and with a view to tackling non-stationary processes, Marani and Ignaccolo (2015) proposed an alternative approach towards extreme values: the Metastatistical Extreme Value distribution (MEVD). This approach exploits all the available data, irrespective of their magnitude, to build the distribution of yearly maxima, $H_n(x)$, from the distribution of ordinary values, $F(x)$. Zorzetto et al. (2016) showed that the MEVD outperforms GEV when $T >> S$, reducing uncertainty by about 50% on average. This property makes the MEVD a good candidate for extreme value analysis in non-stationary rainfall, as it can achieve a relatively low estimation uncertainty ($\sim$ 20%, see Zorzetto et al. (2016)) based on short observational windows. For a visual comparison of the data used by the three methodologies (GEV, POT and MEVD) see Figure 2.1.

2.2 The Metastatistical Extreme Value Distribution

As originally introduced by Marani and Ignaccolo (2015), the MEV cumulative distribution for annual maxima is derived from the probability distributions of the ordinary values. The MEVD approach considers both the

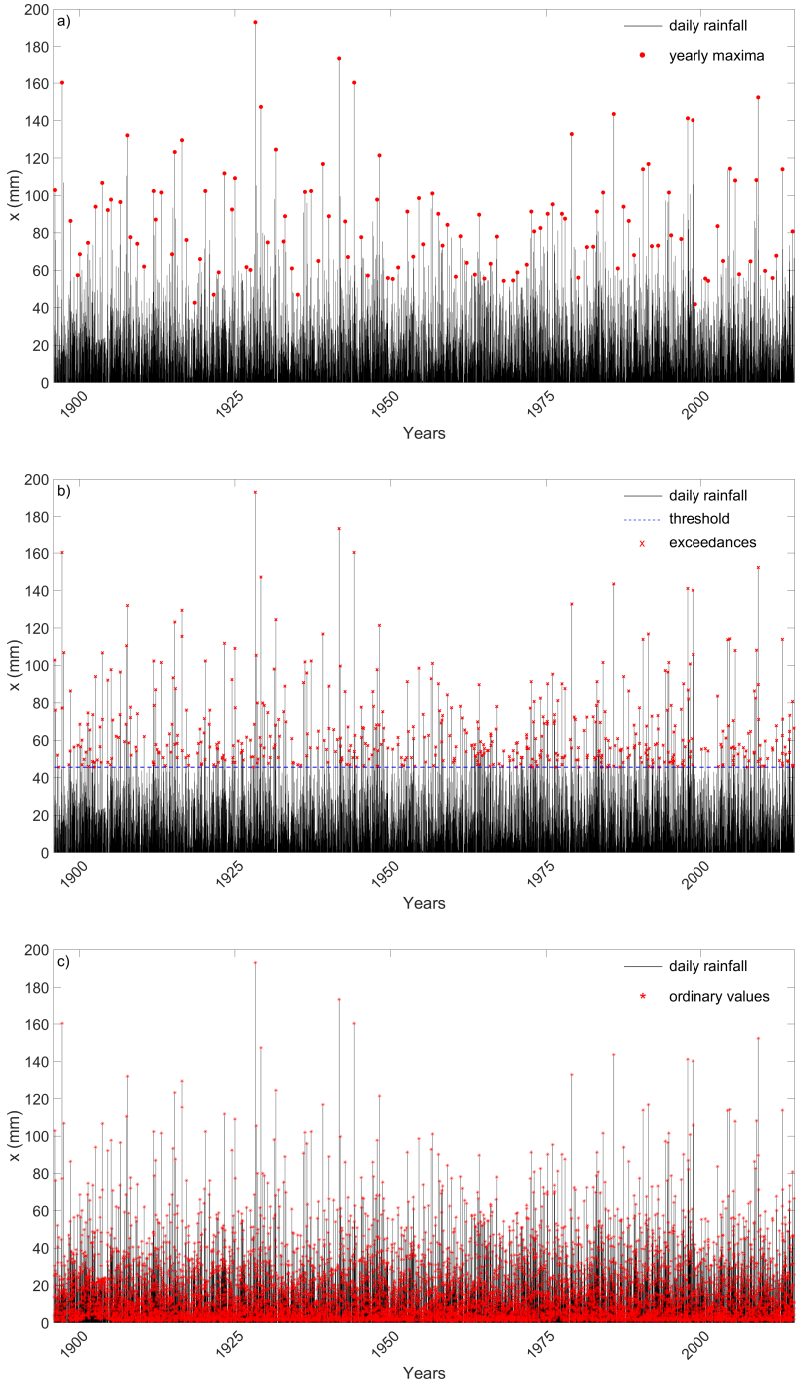


Figure 2.1: Events on which the three approaches are fitted (GEV in panel a, POT in panel b and MEVD in panel c), showed on the record of daily rainfall of the station of Albany, USA. GEV uses only annual maxima (red dots), POT identifies the exceedances (red stars) over a threshold (dashed blue line, representing the 95^{th} percentile of the rainfall values) and MEVD exploits all the ordinary values (red plus).

distributions describing the ordinary values ($F(x;\vec{\theta})$) and the number of occurrences in each year (n) as random variables. A similar type of approach, applied to non-equilibrium dynamical systems (Beck and Cohen, 2004), water dynamics in the soil-plant system (Porporato et al., 2006), fluctuation in river streamflow (Botter et al., 2013), time series modelling (Tjøstheim, 1986), modelling of ordinary fine-scale rainfall (Thayakaran and Ramesh, 2017) is also known as superstatistics or doubly stochastic processes.

The MEVD assumes that ordinary events are independent and writes the probability that a yearly maximum is smaller or equal to a specified value x as follows:

$$H(x) = \sum_{n=1}^{\infty} \int_{\Omega_{\vec{\Theta}}} [F(x;\vec{\theta})]^n g(n,\vec{\theta}) d\vec{\theta} \tag{2.5}$$

where $g(n,\vec{\theta})$ is the joint probability distribution (discrete in N and continuous in $\vec{\Theta}$) of the number of events in a year, n, and of the parameter vector $\vec{\theta}$); $\Omega_{\vec{\Theta}}$ is the population of the parameters values.
Marani and Ignaccolo (2015) propose to fit the distribution of the ordinary events to data in each year on record and to estimate the ensemble average in Eq.(2.5) as the corresponding sample average:

$$\zeta(x) = \frac{1}{S} \sum_{j=1}^{S} [F(x;\vec{\theta_j})]^{n_j} \tag{2.6}$$

For daily rainfall, Marani and Ignaccolo (2015) propose to adopt a Weibull distribution for $F(x;\vec{\theta})$, based on a global study by Wilson and Toumi (2005). Marra et al. (2018) make a similar assumption for hourly rainfall. The MEVD-Weibull can hence be written as:

$$\zeta(x) = \frac{1}{S} \sum_{j=1}^{S} \Big[1 - \exp\Big(- \big(\frac{x}{C_j}\big)\Big)^{w_j}\Big]^{n_j} \tag{2.7}$$

where:

1. S is the number of years for which observations are available
2. C_j and w_j are the scale and shape parameter of the Weibull distribution for year j
3. n_j is the number of events in year j

The reader is addressed to Figure 2.2 for a visualization of the application procedure of the MEV approach.

Several studies have shown the advantages of using the MEVD over the traditional EVT for rainfall data. Zorzetto et al. (2016) first analyzed a large

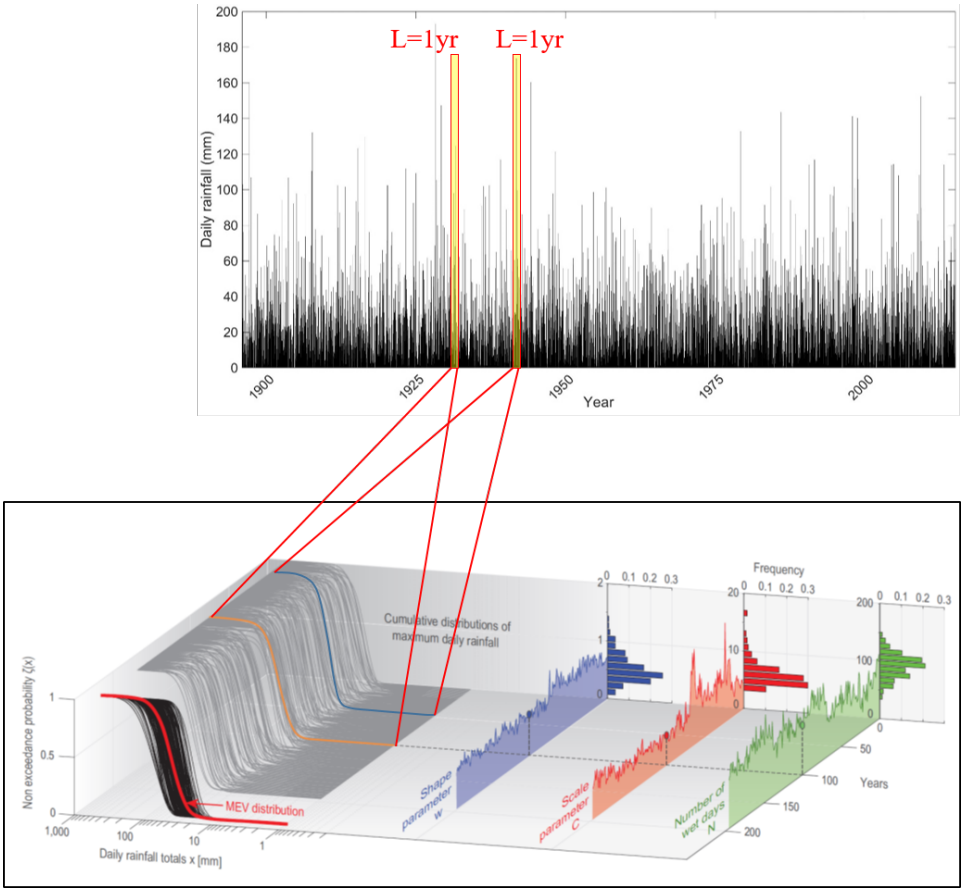


Figure 2.2: Explanation of the application methodology of Eq.(2.6) on a series of daily rainfall. Every year, the scale (C) and shape (w) parameters of the yearly Weibull distributions are estimated by fitting them on the ordinary values; the number of events n is the number of wet days within each year. The yearly cumulative distribution of maximum daily rainfall can therefore be computed as in the expression within the summation in Eq.(2.6). The MEVD is finally computed as the average over all the years of the cumulative distributions of annual maxima (represented here as a red line). From: (Zorzetto et al., 2016).

dataset of long daily rainfall records and found that the MEVD outperforms the GEV distribution when applied to daily rainfall data, especially when dealing with return periods much larger than the length of the observation sample (the average MEVD error reaches a 50%-60% reduction of the GEV estimation error). Marra et al. (2018) found comparable results in the application of the MEVD to hourly rainfall data and then proposed a Simplified MEVD (SMEV) to take into account different distributions underlying the ordinary events (Marra et al., 2019). A smooth modeling using the SMEV approach was developed on Austrian daily rainfall records by Schellander et al. (2019). Zorzetto and Marani (2019) proposed a downscaling procedure to recover the spatial correlation and the probability density function of daily rainfall at the point (gage) scale from satellite estimates and combined this methodology with the MEVD approach, finding that this new approach provides an improvement in the estimation of high return period quantiles even if only a very short record of observations is available (19 years, in the case of the Tropical Rainfall Measuring Mission). Considering streamflow records, Chapters 4 of this thesis shows the outperformance of the MEVD with respect to the traditional GEV in flood frequency analysis performed over the CONUS.

As written in Eq.(2.6), the distributions underlying the ordinary events are fitted on a yearly basis; moreover, Eq.(2.6) considers all the values in the record as generated by the same physical mechanism. In Chapter 3 the possibility of using different sizes of the estimation window is explored, while Chapters 4 and 5 formulate a mixed version of the MEVD, to take into account for different physical phenomena generating hydrological processes (ENSO and TCs, respectively).

Chapter 3

Optimal estimation of rainfall extremes through the Metastatistical Extreme Value Distribution and implications for trend detection

This Chapter studies the predictive uncertainty associated with MEVD estimates of large rainfall quantiles as a function of local rainfall statistical properties. The study defines a MEVD formulation providing the minimum predictive uncertainty in the presence of short observations as a function of statistical properties characterizing the number and variability of wet days/year and the inter-annual variability of the statistical properties of rainfall events. Using these results, the advantages of applying the optimized MEVD formulation to short sliding windows to accurately characterize the possible presence of trends in extreme events within long rainfall time series is shown through the estimation of the daily rainfall quantile associated to a return period of 100 years (h_{100}) on the long daily rainfall series of Padova station. MEVD-estimates are then compared to GEV-estimates, in order to evaluate the suitability of the two approaches to trend detections.

3.1 Introduction

An overview on the traditional EVT and the MEVD approach has been presented in Chapter 2. Marani and Ignaccolo (2015) propose to fit the distribution of the ordinary events to data in each year on record and to estimate the ensemble average in Eq.(2.5) as the corresponding sample average. Here it is noted that $F(x;\vec{\theta})$ may in general be fitted to multiple years, especially in the presence of a dry climate, for which several years may be needed to provide representative samples for the estimation of $\vec{\theta}$. Marra et al. (2019), for example, used the whole calibration sample for parameter estimation when considering stratifying observations among different rainfall types. If observations for S years are available, and $\vec{\theta}_k$ is the parameter vector estimated in the k-th estimation window with length EW (years), the

probability distribution of annual maxima Eq.(2.5) may be approximated as:

$$H(x) = \frac{1}{S}\sum_{j=1}^{S}[F(x;\vec{\theta}_{k(j)})]^{n_j} \tag{3.1}$$

where *k(j)* indicates the estimation window containing year *j*.

In the case of daily rainfall, following the reasoning by Wilson and Toumi (2005); Marani and Ignaccolo (2015), Eq.(3.1) can be written using a Weibull distribution for the $F(x;\vec{\theta}_{k(j)})$

$$H(x) = \frac{1}{S}\sum_{j=1}^{S}\Big[1-\exp\Big(-\big(\frac{x}{C_{k(j)}}\big)\Big)^{w_{k(j)}}\Big]^{n_j} \tag{3.2}$$

After Zorzetto et al. (2016), parameter estimation of the ordinary Weibull distributions is performed here using the Probability Weighted Moments (PWM) method. This method has been shown to provide reduced bias and uncertainty in the presence of small samples, and is less affected by outliers Greenwood et al. (1979). Furthermore, PWM's give more weight to the tail of the distribution than to its central part.
Coherently with the guidelines of the World Meteorological Organization (Klein Tank, A. M. G. Zwiers and Zhang, 2009), a *wet day* is defined as a period of 24-hours in which at least 1 mm of rainfall was measured. Tuning the threshold might provide a better description of the ordinary events on the one hand, but it would result in a case-dependent choice, this issue becoming even more relevant when comparing for example data from rain gages and satellites, which are characterized by different detection thresholds (Zorzetto and Marani, 2019). Hence, in the present dissertation, which analyzes a wide variety of location characterized by different climatic features, ordinary events are defined as the exceedances over a fixed threshold, set to the widely-used 1 mm value indicated by the WMO. This choice is different from Marra et al. (2019), who focused on a more confined area and left-censored observed values using a threshold equal to the 75^{th} quantile computed on the observed data.
The selection of the length EW of the estimation window represents a trade-off between making sure that a sufficient number of events is used to estimate distributional parameters and keeping EW to a minimum to resolve inter-annual fluctuations in rainfall regimes. The optimal choice of EW is thus potentially a function of the average number of events/year, its inter-annual variability, the values and inter-annual variability of Weibull parameters. The original formulation of the MEVD, as well as subsequent rainfall applications (Zorzetto et al. (2016); Marra et al. (2018); Zorzetto and Marani (2019)) assume EW=1 year. This Chapter seeks to determine how the choice of the EW affects the uncertainty in MEVD quantile estimation.

3.1.1 Some properties of the MEVD

Marra et al. (2019) propose a Simplified MEVD (SMEV in short), originally devised to account for multiple distributions and to reduce the computational effort required to invert Eq.(2.7) and estimate the quantile corresponding to a fixed value of the cumulative probability. The SMEV distribution makes two assumptions: 1) it assumes a single probability distribution can represent all the ordinary values on record; 2) it replaces the variable number of events/year with one single average value. Originally, Marra et al. (2019) propose to account for the possible presence of different rainfall-generating mechanisms (e.g. local convective storms vs. large-scale stratiform perturbations). Here it is considered the simplest SMEV distribution expression, which may be written as:

$$SMEV(x) = F(x;\vec{\theta})^{\bar{n}} \tag{3.3}$$

where $F(x;\vec{\theta})$ is the ordinary-value distribution fitted to the whole sample and $\bar{n}$ is the average yearly number of events.
The simplicity of the SMEV is appealing, but it is important to explore the implications of neglecting the inter-annual variability in the number of events and in rainfall regimes. Some of these implications are here investigated theoretically and, subsequently, through applications.

In order to address the sensitivity of the MEVD to the variability in the n_j's, let's consider a generic term, $F_j(x)^{n_j} = F_j(x;\vec{\theta})^{n_j}$, in Eq.(2.6). The sensitivity of this term to changes in its exponent can be expressed as the ratio of the fractional change in $F_j(x)^{n_j}$ to the fractional change in n_j:

$$\frac{d[F_j(x)^{n_j}]}{F_j(x)^{n_j}} = dn_j \cdot \log[F_j(x)] \tag{3.4}$$

These terms all tend to 0 when $F_j(x) \to 1$, i.e. for large (extreme) x. Therefore, it can be concluded that the sensitivity of all the terms appearing in the MEVD to the changes in the n_j's (which may be due to internal variability, to climate changes, or to observational errors) decreases as the return period of the event of interest increases. This line of reasoning also indicates that the sensitivity to changes in the n_j's is likely to be smaller in dry climates, where small values of $\bar{n}$ prevent a large variability in the exponents n_j's. For the question immediately at hand, regarding the impacts of adopting the Simplified MEVD, this finding implies that the effect on the value of $H(x)$ of substituting $\bar{n}$ in place of the n_j's in Eq.(2.6) decreases for large x, even more so since the terms $dn_j = \bar{n} - n_j$ tend to be of both signs and of similar magnitudes, thus leading to a compensation in the sum in Eq.(2.6). In general, the importance of the role of $\bar{n}$ is a function of the return period: the greater the return period of interest the smaller the role of the number of

wet days. The implications for climate modelling are worth noting: the previous result suggests that the accurate modelling of the exact distribution of the number of rainy days/year at a site is not as important as reproducing the correct distribution of the magnitudes of ordinary events, particularly if the focus is on large extremes.

After analyzing the impact of the variability in the number of yearly events, the importance of the inter-annual variability of the probability distributions of ordinary events is now addressed. Originally, the SMEV approach was proposed by Marra et al. (2019) by simply neglecting the inter-annual variability in the probability distribution of rainfall depth. Here it is shown that the SMEV expression can be fully retrieved, for large extremes, if the number of events/year and the parameter vector of the ordinary value distribution are independent.
The derivation starts from introducing the survival function $G(x;\vec{\theta})$, such that $F(x;\vec{\theta}) = 1 - G(x;\vec{\theta})$. For large values of x, one can expand the n-th power of $F(x)$ around $G_0 = lim_{x\to\infty} G(x;\vec{\theta}) = 0$, as follows:

$$\begin{aligned} F(x)^n &= [1 - G(x;\vec{\theta})]^n \cong 1 - n(1 - G_0)^{n-1} \cdot G(x;\vec{\theta}) + o(G(x;\vec{\theta})^n) \\ &\sim 1 - nG(x;\vec{\theta}) \end{aligned} \tag{3.5}$$

Under the condition of independence of n and θ, Eq.(2.5) can now be rewritten as:

$$\begin{aligned} H(x) &= \sum_{n=1}^{\infty} \int_{\Omega_{\vec{\Theta}}} [F(x;\vec{\theta})]^n g(\vec{\theta}) \cdot p(n) d\vec{\theta} = \\ &\int_{\Omega_{\vec{\Theta}}} [1 - \big(\sum_{n=1}^{\infty} p(n) \cdot n\big) \cdot G(x;\vec{\theta})] g(\vec{\theta}) d\vec{\theta} = \\ &\int_{\Omega_{\vec{\Theta}}} [1 - < n > \cdot G(x;\vec{\theta})] g(\vec{\theta}) d\vec{\theta} \end{aligned} \tag{3.6}$$

If the ensemble average $< n >$ is approximated with the sample average $\bar{n}$ and the ensemble average over $\Omega_{\vec{\Theta}}$ is approximated with a sample average over the parameter values, $\vec{\theta_j}$, estimated from suitable subsets of the data, it can be written:

$$H(x) = \frac{1}{S} \sum_{j=1}^{S} [1 - \bar{n} G(x;\vec{\theta_j})] \sim \frac{1}{S} \sum_{j=1}^{S} F(x;\vec{\theta_j})^{\bar{n}} \tag{3.7}$$

which is a first simplification of the MEVD expression, which neglects the variability in the yearly number of events, but still accounts for the inter-annual variability of the distributions. Eq.(3.7) can be further simplified, by assuming that the parameters $\vec{\theta_j}$ are time-invariant and equal to a constant $\vec{\theta}$, which yields to the SMEV expression as in Eq.(3.3). This version of

the MEVD can be retrieved under the assumption of independence between n and θ or by assuming that ordinary rainfall depths come from a time-independent distribution. Whether these hypotheses hold in practice, and what their impacts may be on the estimation of large extremes needs to be tested. It will be done in the following.

3.1.2 Implications for the detection of trends in extreme events

The detection of possible trends in extreme rainfall events is of great interest, though it remains difficult due to the large uncertainties involved in traditional approaches to estimating extremes (Easterling et al., 2000). Estimation uncertainty, in fact, arguably hides subtle changes in extremes and new approaches to better quantify changing extremes are emerging (e.g. Chavez-Demoulin and Davison (2005); Khaliq et al. (2006); Aryal et al. (2009); Westra and Sisson (2011); Beguería et al. (2011); Ouarda and El-Adlouni (2011)). The study of long observational time series can also be of help in the detection of possible trends. In a recent paper, Papalexiou and Montanari (2019) performed a global analysis of daily precipitation in the 1964-2013 period and found an increasing trend in the frequency of extremes that is unlikely under the assumption of stationarity. With a view to practical applications Wright et al. (2019) recently highlighted the need to define intensity-duration-frequency (IDF) curves, used in engineering design, that account for expected changes in extremes over the lifetime of water infrastructures.

The use of improved-accuracy extreme value analysis methods, allowing a reduced uncertainty even for short analysis windows can significantly contribute to our ability of quantifying changing extremes, in the observed past and in the projected future. This Chapter shows that the improved MEVD discussed here can provide such a contribution, by comparing results from MEVD and the traditional EVT methods applied to moving window analyses over a long observational time series.

3.2 Materials and Methods

3.2.1 Datasets

Observations

The dataset studied in Zorzetto et al. (2016) is used here; it consists of 37 daily rainfall stations widely distributed geographically and selected on the basis of their long observational periods (average series length is 135 years). These two features allow exploring different climatic conditions and quantifying the accuracy of the estimations for high return periods, using

different calibration sample sizes and experimenting with different values of the length of the estimation window.

For extreme-event trend studies the focus is on the daily rainfall observations recorded in Padova (Italy, 1725-today) (Marani and Zanetti, 2015), the longest among the stations considered here. The Padova time series consists of 277 observational years (gaps: 1765-1767, 1815-1825 and 1838-1840).

The observational dataset representing a variety of actual rainfall statistics is a fundamental reference for the quantification of uncertainty in a context tightly anchored to reality. However, in order to be able to design analysis experiments in which the range of key statistical properties expected to affect the optimal extreme value analysis approach can be pushed "to the limit", synthetic data designed after some of the "most extreme" sites in the dataset are also used. Synthetic data allow to flexibly change the values of some key statistical properties. One such property is the mean value of events/year, $\bar{n}$, which in the dataset considered here varies from 23 days/year (Mesa, Arizona, USA) to 173 days/year (Säntis, in the Swiss Alps). A second candidate is the dispersion index, DI_n, defined as the ratio between the variance of the number of events/year, s_n^2, and $\bar{n}$. DI_n is equal to 1 events/year for a Poisson process, which is the reference non-clustered process. However, rainfall is typically found to be inter-annually clustered: the average value of DI_n in the dataset studied here is 2.1 events/year, the minimum value is 1 events/year (Asheville), while the maximum value is 3.7 events/year (Mount Aigoual). Ideally, one would like to explore an even wider range of $\bar{n}$ and DI_n, hence the choice of generating synthetic data to complement analyses of observations.
In addition to this, as noted by Marra et al. (2019), the inter-annual variations of the ordinary events distribution parameters are co-caused by three effects: (i) inter-annual variations of the characteristics of the ordinary events; (ii) parameter estimation uncertainty; and (iii) varying proportions in the occurrence of different types of ordinary events, if the identical distribution assumption within each year fails. Here, it is noted that a fourth source of uncertainty is given by (iv) measurement errors, which characteristic of the observational setup. Parameter estimation uncertainty is an intrinsic feature of the data, and increases as the number of events/year decreases. The analysis of synthetic time series of rainfall, which are uncertainty-free, can provide indications that can be used to infer implications related to observational records. Lastly, in this Chapter, precipitation values are assumed to be identically distributed, while in the analyses performed in Chapter 5 the different rainfall mechanisms will be taken into account.

Synthetic daily rainfall time series

The first set of experiments aims to determine the dependence of the value of the optimal length of the estimation window, EW, in the $(\bar{n}, DI_n)$ plane. The second set of experiments is designed to evaluate the dependence of the optimal choice of EW on the inter-annual variability of the ordinary event probability distribution. In order to perform these experiments, synthetic time series are generated as follows:

1. two stations are selected, Mesa and Zürich, which are characterized by the highest and lowest variability, among all stations, in the yearly values of the shape parameter (w) of the fitted Weibull distributions ($\sigma_w = 0.16$ for Mesa and $\sigma_w = 0.06$ for Zürich). The focus is on the shape parameter because it is the main control of the sub-algebraic tail of the Weibull distribution. The probability density functions of the yearly shape parameters for these two stations are shown in Figure 3.1. By keeping the original values of yearly fitted scale and shape parameters, new Weibull-distributed synthetic series of daily rainfall depth are generated. The number of wet days in each synthetic year is generated from a negative binomial distribution (Fisher, 1941) where $\bar{n}$ and DI_n are varied to explore a wide range of their possible values. The choice of a binomial distribution is supported by a satisfactory agreement between fitted and observed values of n (see Quantile-Quantile (QQ) plots in Figure 3.2). The synthetic series are generated with an average number of rainy days between 5 and 365 (steps of 5 days/year) and with values of the dispersion index between 1.25 and 4 events/year (0.25 events/year steps);

2. with reference to the station of München, for which the correlation coefficient among the estimated scale (C) and shape (w) parameters of the yearly fitted Weibull distributions is low ($\rho \sim 0.1$), and while keeping the original series of C values, the variance of w was varied by multiplying and dividing the original value by a factor of 4. While doing so, a range of the number of rainy days per year from 5 to 200 is explored. The values of w are generated from a Gamma distribution, while the values of n are drawn from a negative binomial distribution (both distributions are fitted to observations, see QQ plots in Figure 3.3).

3.2.2 Cross-validation procedure

The uncertainty in the estimation of high return period quantiles is quantified by means of a cross-validation procedure, performing a Monte Carlo experiment with a number of realizations N_r=1000. The cross-validation procedure is structured as follows:

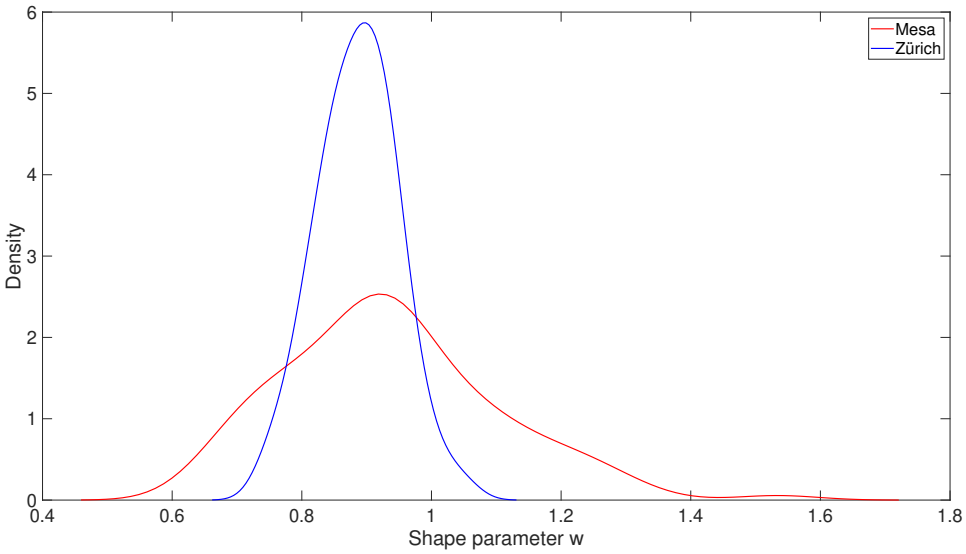


Figure 3.1: Kernel density function for the yearly shape parameters of the Weibull distributions estimated on the records of daily rainfall of Mesa (red) and Zürich (blue) stations.

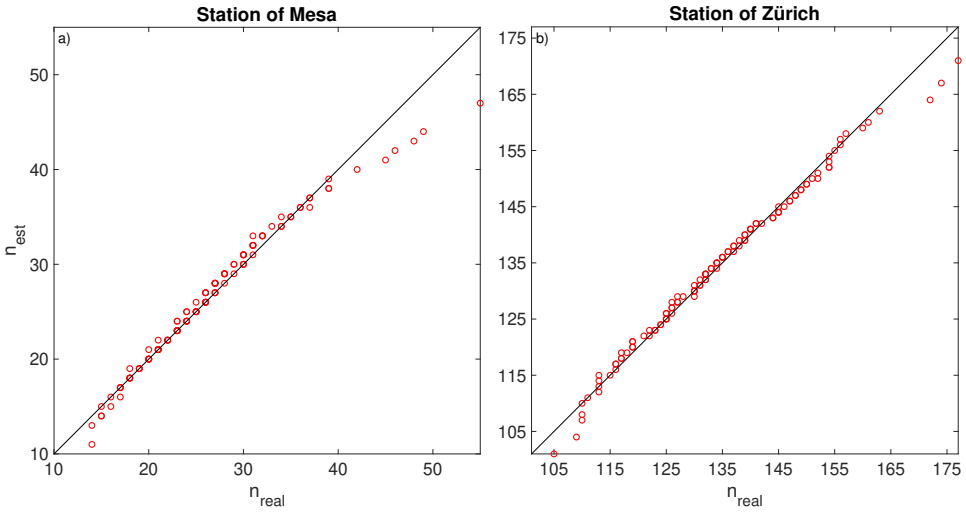


Figure 3.2: Quantile-quantile (QQ) plot of the estimated vs real number of wet days for Mesa (panel a) and Zürich (panel b) stations, in order to evaluate the goodness of fit of the negative binomial distribution on the original number of events in these two stations.

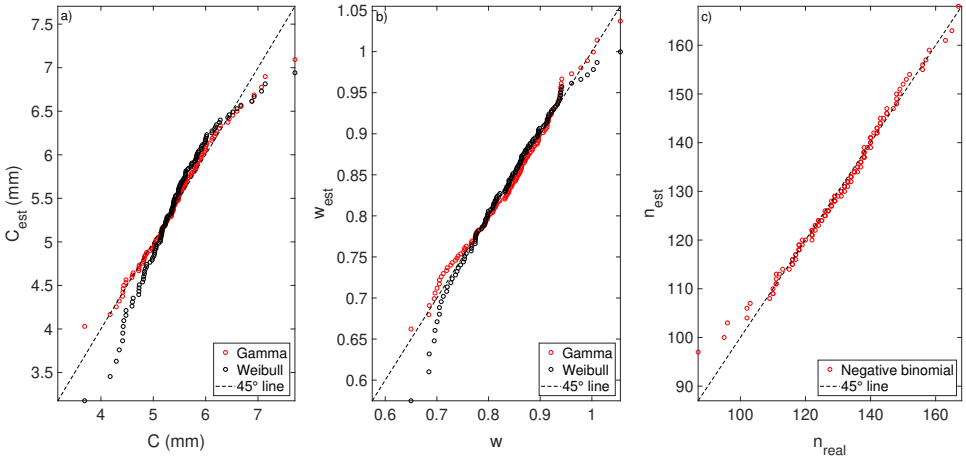


Figure 3.3: Features of München station. Gamma and Weibull distributions (represented in red and black respectively) were tested for the scale and shape parameter and the negative binomial distribution was used for the number of rainy days. Panel a (b) shows the QQ plot of the estimated C (w) values with the two tested distributions, while panel c) represents the QQ plot of the estimated number of wet days using the negative binomial distribution.

1. the observational (or synthetic) years on record is randomly reshuffled (keeping all observations in their original year to preserve both their yearly frequency distributions and the distribution of the number of events/year) to generate a realization displaying no systematic variability;

2. the reshuffled time series is divided into two independent samples, randomly selecting S years as calibration sample, which is used for estimating the values in the parameter vector $\vec{\theta}_{k(j)}$, and keeping the remaining years (the test sample, with size $L = L_{tot} - S$, L_{tot} being the length of the whole series) to test the accuracy of the estimated quantiles;

3. the empirical cumulative frequencies are estimated using the Weibull plotting position formula F_k =*k/(L+1)*, where $k = 1, \ldots, L$ is the position of value x_k in and ascending order listing of the test sample. MEVD quantiles are estimated by solving $H(x_k) = F_k$ (where $H(x)$ is given by Eq.(2.7));

4. the accuracy by comparing MEVD estimates to the statistics from the test sample is estimated according to the evaluation metrics described in 3.2.3;

5. the above process is repeated N_r=1000 times to obtain a full statistical description of error metrics. This whole process is repeated for

different calibration sample sizes (S=5, 10, 20 and 30 years) and for different estimation window lengths (EW=1, 2, 3, 4, 5 and S years).

This cross validation approach is also applied using GEV estimates (for comparisons with the current standard approach) and using the simplified versions of the MEVD.

3.2.3 Evaluation Metrics

The evaluation of the estimation accuracy is performed using two different metrics, tailored for different analysis purposes. The attention is focused on the uncertainty in the evaluation of high quantiles and on a global performance metric:

1. Non-dimensional error. For the *j-th* realization, the non-dimensional error is computed as

$$\varepsilon_j(S,T) = \frac{x_{est,j}(S,T) - x_{obs,j}(S,T)}{x_{obs,j}(S,T)} \tag{3.8}$$

between estimated and observed maxima for a specific calibration sample size (S) and return period (T). The Monte Carlo approach allows the construction of a full frequency distribution of the non-dimensional error. The Fractional Standard Error (FSE) is then computed over all the Monte Carlo realizations of the non-dimensional error:

$$FSE(S,T) = \frac{1}{N_r}\Big[\sum_{j=1}^{N_r} \varepsilon_j(S,T)^2\Big]^{\frac{1}{2}} \tag{3.9}$$

2. Skill Score (SS$\in(-\infty,1]$) (Murphy and Winkler (1992); Hashino et al. (2006)). The SS is computed for each Monte Carlo realization in order to provide a global index of accuracy of the MEVD quantile estimates. Because the interest is here on large return periods, a modified version of the SS is used, computed only on quantiles with return period greater than the calibration sample size ($T/S \geq 1$, since for $T<S$ there is no need to fit a statistical model):

$$\begin{aligned} SS(x_{est,j}, x_{obs,j}) &= \rho^2_{x_{est,j},x_{obs,j}} - \Big[\rho_{x_{est,j},x_{obs,j}} - \Big(\frac{\sigma_{x_{est,j}}}{\sigma_{x_{obs,j}}}\Big)\Big]^2 + \\ &\quad - \Big[\frac{(\mu_{x_{est,j}} - \mu_{x_{obs,j}})}{\sigma_{x_{obs,j}}}\Big]^2 \end{aligned} \tag{3.10}$$

where $\rho_{x_{est,j},x_{obs,j}}$ is the correlation between the estimated values ($x_{est,j}$) and the observations ($x_{obs,j}$); $\sigma_{x_{est,j}}$ and $\sigma_{x_{obs,j}}$ ($\mu_{x_{est,j}}$ and $\mu_{x_{obs,j}}$) represent the standard deviation (mean) of the estimates and the observations, respectively. The second term in Eq.(3.10) measures the conditional bias, and the third term the unconditional bias.

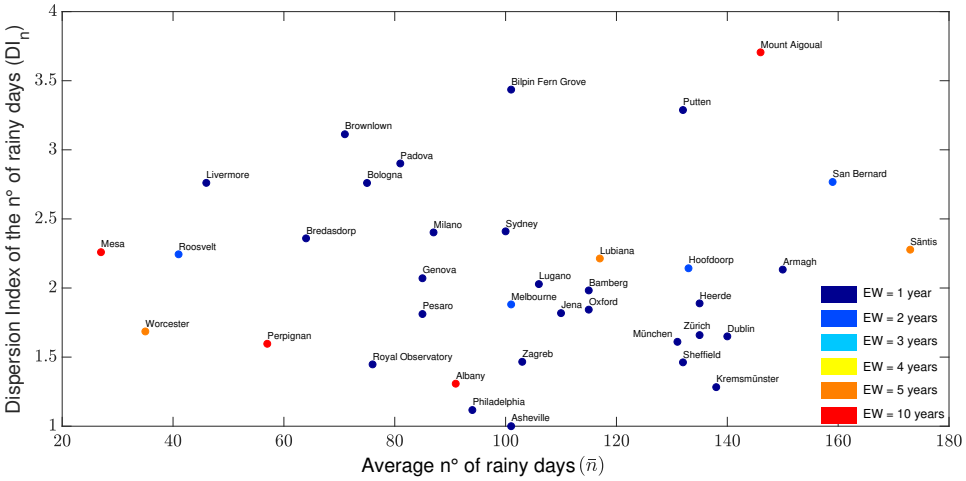


Figure 3.4: Stations arranged according to values of average yearly number of rainy days and Dispersion Index of n. Colors represent the optimal estimation window length based on the median value of SS, computed from quantiles corresponding to $T > S$ (S=10 years here). The color scale indicates the optimal estimation window length.

3.3 Results

3.3.1 Optimal Estimation Window Length

For most stations EW=1 year maximizes the value of the SS (dark blue circles in Figure 3.4).
A pattern in the plane $(\bar{n}, DI_n)$ can be hypothesized: 1) in a region delimited by small values of $\bar{n}$ and a relatively small DI_n ($DI_n < 2.5$), the optimal EW>1 year; 2) in a central area, characterized by higher values of $\bar{n}$ (60-160), irrespective of the value of DI_n, the optimal EW=1 year for the largest majority of the stations; 3) three stations (Mount Aigoual, San Bernard and Säntis) with high $\bar{n}$ and DI_n, for which the optimal EW>1 year.

At a first sight, the hypothesized regions do not seem to have clear-cut boundaries, especially with reference to region 2. In this region, indeed, four stations exhibit an optimal estimation window greater (Lubiana and Albany) or equal (Melbourne and Hoofdoorp) to two years, even if they lie in an area of the plot in which most of the other stations require a parameter estimation on a yearly basis. One possible explanation could be the fact that the actual gain obtained by using an estimation window longer than one year is not significant. Therefore, the plot in Figure 3.4 is complemented with the box plots in Figure 3.5, in order to evaluate the actual differences in the SS values obtained using the optimal estimation window and the one 1 year long.
As it can be seen in Figure 3.5, for three out of four of these stations (Hoof-

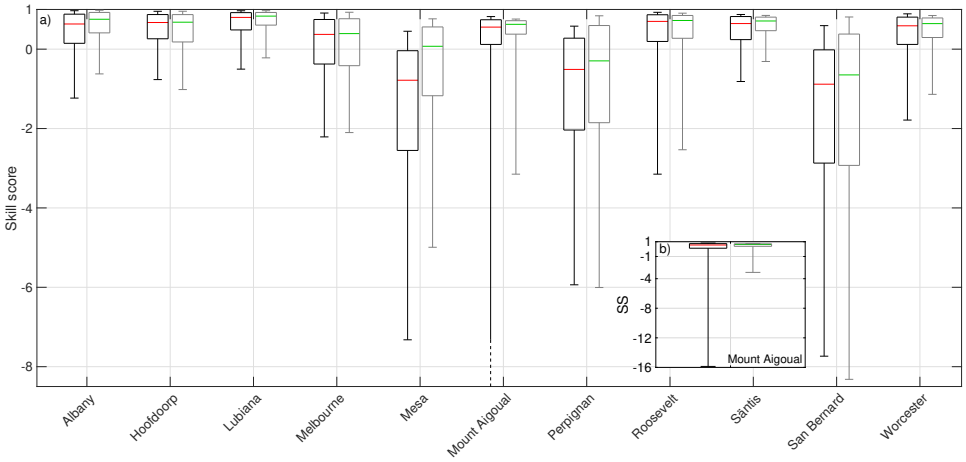


Figure 3.5: Difference between the SS obtained with the optimal estimation window (gray boxes with a green colored median) and the SS from a 1-year estimation window (black boxes with a red colored median) for stations in Figure 3.4 for which the optimal estimation window is $EW > 1$ year. The whiskers represent the 5^{th} and 95^{th} percentiles, the boxes the 25^{th} and 75^{th} percentiles and the colored lines are the median values. Subplot b) within the main plot represents the box plots for Mount Aigoual station, for which the 5^{th} percentile is much smaller than the other values. It is plotted separately to avoid the compression of all the other values along the y-axis.

doorp, Lubiana and Melbourne) no significant differences in the median value and in the percentiles of the SS can be appreciated, meaning that there is a negligible improvement when using a EW >1 year (i.e., these stations could also be represented in blue in Figure 3.4). The SS somewhat improves for Albany when EW=10 years with respect to EW=1 year. However, it can be noted that Albany is near the hypothesized boundary below which greater values of EW reduce the estimation error.

The three stations in the set, Mount Aigoual, San Bernard and Säntis, which require an estimation window larger than one year despite their large average number of rainy days per year, are all located in high-altitude mountainous areas: 1567 m asl for Mount Aigoual, 2472 m asl for San Bernard and 2502 m asl for Säntis. Looking at their comparative box plots in Figure 3.5, the station of Mount Aigoual is the one that is characterized by a more significant difference (and a much wider uncertainty) when using a 1-year window or the optimal one. This might suggest that other explanatory variable related with orographic forcing or to the mixture of solid/liquid precipitation might play a role, but they are not accounted for in this analysis.

The SS is a global metrics that provides one single number to univocally identify the optimal EW, yet it has the drawback that computing the differences between the SS of the optimal EW and the EW= 1 year does not lead to a straightforward quantification of the actual advantage conferred by the use of one EW instead of another. On the contrary, the FSE as estimated in

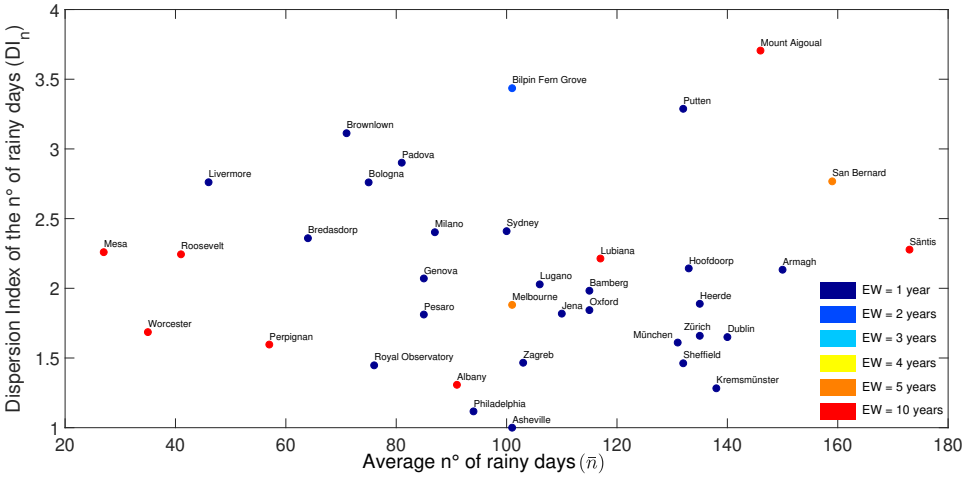


Figure 3.6: Stations arranged according to values of average yearly number of rainy days and Dispersion Index of n. Colors represent the optimal estimation window length based on the FSE value of SS computed for the quantile associated to the highest return period ($T_{max} = L_{tot} - S + 1$). The color scale indicates the optimal estimation window length.

Eq.(3.9) refers to single return periods, but allows to immediately quantify the reduction in the estimation uncertainty obtained by using the optimal EW.
Figure 3.4 is complemented identifying the optimal EW on the basis of the FSE computed for the quantile with the maximum return period ($T_{max} = L_{tot} - S + 1$) in each station. The record lengths all differ one from the other, but being the stations all characterized by very long historical series, the highest quantiles that can be estimated using S=10 years are all associated to return periods greater than 10 times the calibration sample size.
The same pattern that was highlighted in Figure 3.4 is found in Figure 3.6, in terms of comparison between EW=1 year or longer (slight differences can be instead seen when looking at the precise size of the EW). Again, the interest is in evaluating to which extent the estimation accuracy is reduced by using an EW=1 year instead of the optimal one. The FSE is a good candidate to make this comparison. As it can be seen in Figure 3.7, the four stations characterized by the lower values of $\bar{n}$, in combination with a low DI_N, are taking a more significant advantage in the use of a long EW, being the FSE steadily decreasing while widening the EW. Albany, Bilpin Fern Grove and Lubiana show very little difference in the FSE(EW=1yr) and FSE(EW=optimal), whereas for the remaining stations the highest reduction of the FSE value occurs between EW=1 years and EW=2 years, the advantage becoming negligible for larger estimation windows.

So far, the dependence of the optimal estimation window size has been explored in the $(\bar{n}, DI_n)$ "phase space". Other factors are likely important in

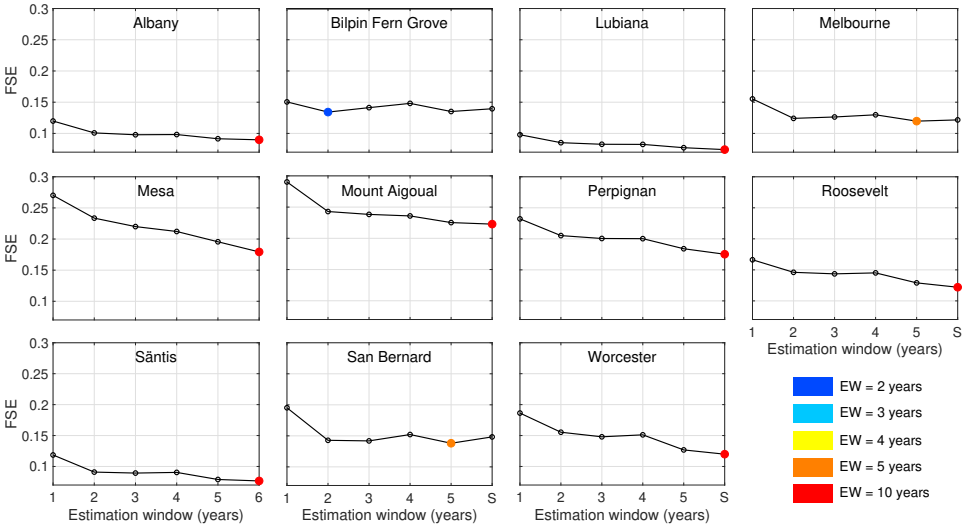


Figure 3.7: FSE values plotted against the EW length for the maximum quantile computed in each series (i.e., the one associated with a return period $T_{max} = L_{tot} - S + 1$, with S=10 years), for the stations for which the optimal EW was found longer than one year on the FSE basis. In each subplot, the colored dot represents the optimal EW size, according to the color scale in the legend.

these analyses, in particular the inter-annual variability in the distributions of the ordinary events, which will now be explored. Furthermore, the available stations do not represent all values of potential interest in the $(\bar{n}, DI_n)$ plane (Figure 3.4). In particular, the areas corresponding to very small and very large values of n remain unrepresented.

Therefore, as described in Section 3.2.1, synthetic series of daily rainfall are generated using Weibull parameters from Mesa and Zürich, while changing $\bar{n}$ and DI_n. The two cases, large (Mesa) and low (Zürich) inter-annual variability in the distributions of ordinary events, exhibit marked differences (Figure 3.8, complemented by Figure 3.9). In the presence of high inter-annual variability of rainfall magnitudes (Figure 3.8a), EW=1 year is optimal starting at low values of $\bar{n}$. When the probability distributions of rainfall magnitudes are more uniform across the years (Figure 3.8b), optimal values of EW may be large even for large values of $\bar{n}$. This finding points to the importance of the variability in the rainfall distributions that seems to exert a stronger control with respect to the average number of events/year, except for very small values of $\bar{n}$ (indicatively less than 20-25 events/year). In these dry climates a large EW size is optimal, independently of the variability of the probability distributions. In the case of low inter-annual variability, differences in the performance using the optimal EW or EW=1 year are relatively small (see Figure 3.9). In this case, using EW=1 year size is advisable, since there is little to gain from searching for the optimal

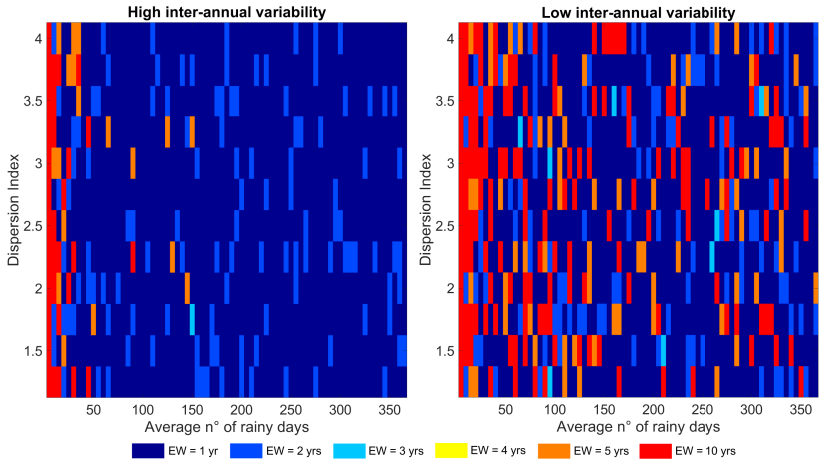


Figure 3.8: Optimal estimation window length for the synthetic series generated starting from the parameters of Mesa (Zürich), represented in panel a (b). Each synthetic station is represented as 5 days/year x 0.25 days/year bin centered in the $\bar{n}$ and DI_n values used to generate the synthetic data. The color scale indicates the optimal estimation window.

EW size. This choice also has the advantage of allowing the use of short sliding windows to more sharply detect changes in trends in non-stationary processes.

To further explore the role of the inter-annual variability in ordinary value distributions, the variability of the shape parameter is now changed in order to explore much lower and higher variability than the observed one. This is obtained by multiplying the original values of the variance of w by factors equal to 0.25 and 4, respectively. Further changes to the variability of ordinary value distributions are explored by varying the average number of events/year from 5 to 200.
Figure 3.10 (complemented by Figure 3.11 for all the values of $\bar{n}$) and Figure 3.12 show the median SS value, computed over the $N_r = 1000$ Monte Carlo realizations. For large values of the variance of w, EW=1 year yields a better performance with respect to longer estimation window lengths. No clear dependence on the average number of wet days is identifiable as the value of SS for a given EW does not show a relation on $\bar{n}$ as represented by the size of circles in Figure 3.10. When the variance of w is decreased to 1/4 of the original value, EW=10 years outperforms other estimation window sizes, but differences between EW=10 years and EW=2 years are quite small (see Figure 3.11). Again, no clear dependence on the average number of rainy days/year.

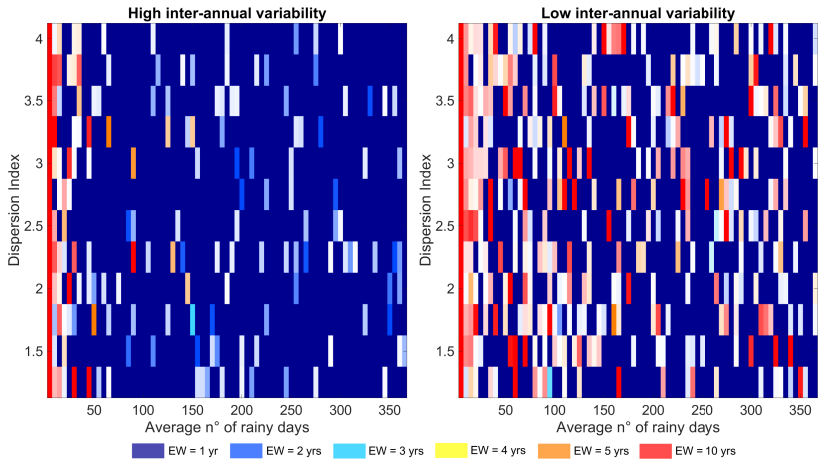


Figure 3.9: Optimal estimation window length for the synthetic series generated starting from the parameters of Mesa (Zürich), represented in panel a (b). Each synthetic station is represented as 5 days/year x 0.25 days/year bin centered in the $\bar{n}$ and DI_n values used to generate the synthetic data. The color scale indicates the optimal estimation window. Opacity values represent the differences between the SS of the optimal EW and the EW=1 year. Synthetic stations for which the optimal EW is 1 years have full opacity.

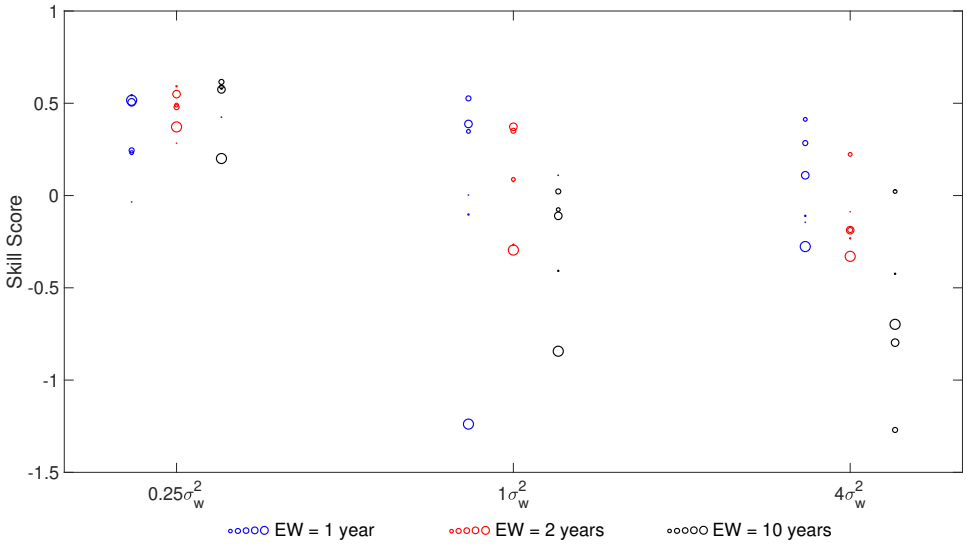


Figure 3.10: Median values of SS for the München-like station in which 1) the variance of the shape parameter of the yearly Weibull distributions fitted to the original record is modified, and 2) the average number of wet days per year is changed. The variance is multiplied by factors of 0.25, 1, and 4. The colors refer to the EW=1 year (blue), EW=2 years (red), and EW=S=10 years (black). The size of the dots is proportional to the yearly average number of rainy days. Here, to keep the plot clear, only values of $\bar{n}$ = 5, 25, 75, 100, 150, 200 are shown (all the values on $\bar{n}$ are shown in Figure 3.11).

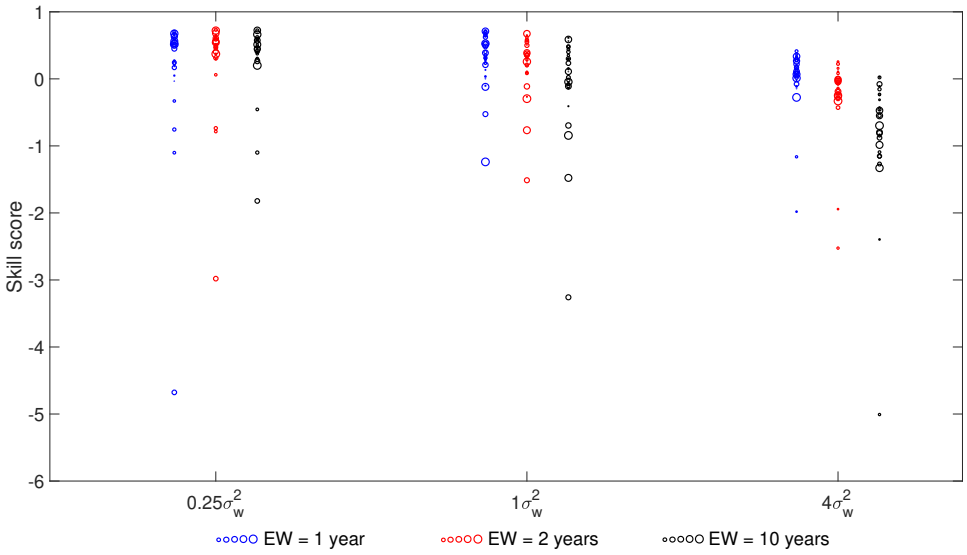


Figure 3.11: Median values of SS for the München-like station in which 1) the variance of the shape parameter of the yearly Weibull distributions fitted to the original record is modified, and 2) the average number of wet days per year is changed. The variance is multiplied by factors of 0.25, 1, and 4. The colors refer to the EW=1 year (blue), EW=2 years (red), and EW= 10 years (black). The size of the dots is proportional to the yearly average number of rainy days, which is changed from 5 to 100 with a 5 days step and from 100 to 200 with a 10 days step.

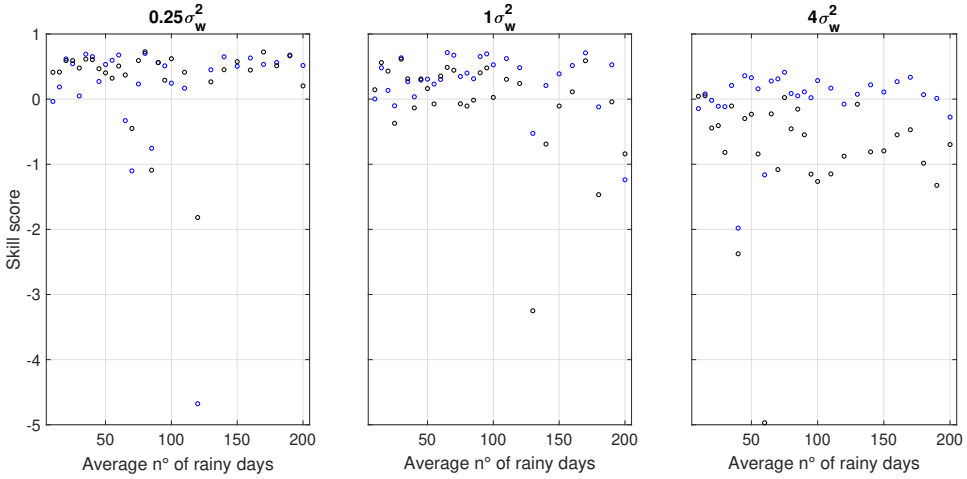


Figure 3.12: Plots showing the median skill score for the MEVD with an EW of 1 year (blue dots) and the EW of 10 years (black dots) for three values of the variance of w. On the left, the original variance is divided by 4, in the central sub-plot the variance is the original one and on the right the variance is multiplied by 4. The average number of days is from 5 to 200, with a step of 5 until 100 and with an increased step of 10 between 100 and 200. The calibration sample size is 10 years.

3.3.2 "Full" MEVD and Possible Simplifications

The performance of the SMEV formulation (Marra et al., 2019) is extensively explored as a function of the statistical characteristics of the arrival process and of the variability in the probability distributions of ordinary days. The aim of this analysis is to understand the sensitivity of the two versions of the MEVD with respect to changes in the sample sizes and when considering high ratios of the return period (T) to the calibration sample size (S). Schellander et al. (2019), for example, found that differences in the errors provided by the two approaches are negligible, of the order of 10^{-4}.

In the first row of Figure 3.13 it can be seen how the SMEV estimation accuracy generally decreases with respect to that of the MEVD (with EW=1 year) as large values of T/S are considered. Moreover, when the calibration sample size is increased (from 10 to 30 years), the results from the two approaches behave differently. In the case of the MEVD with EW=1 year the FSE decreases as S is increased. In the case of the simplified MEVD the use of a larger value of S does not necessarily lead to an improvement in the estimation error (FSE). This result may be explained by the fact that, when the calibration sample size is increased, data from a greater number of years are fitted with just one single distribution. Deviations from this single, "average", distribution therefore increase and offset the advantage of using more data in fitting the distribution. This result, obtained from a dataset including very different climates, ranging from quite dry climatic conditions to very wet ones, and characterized by varying degrees of variability in the distributions, highlights the fact that a parameter estimation on a yearly basis allows to better describe the possible inter-annual variations in statistical properties.

The sensitivity of MEVD estimates to the values of the yearly number of events is now evaluated. In particular, results from the "full" MEVD are compared with those obtained from substituting the original number of events/year with the average $\bar{n}$ (second row in Figure 3.13). This comparison is performed using different calibration sample sizes (S = 5, 10, 20 and 30 years) on the whole historical dataset. As expected, the values of the FSE of the two approaches are really similar one to the other and differences become slightly more negligible when moving towards high ratios of T over S.

The negligible effect of the change from the original number of events to its average value suggests that the differences found in the comparison between the "full" MEVD and the SMEV are completely driven by the inter-annual variability. Hence, the computation of the FSE for a MEVD-$EW = S$ years is performed keeping the original number of wet days and it is compared to

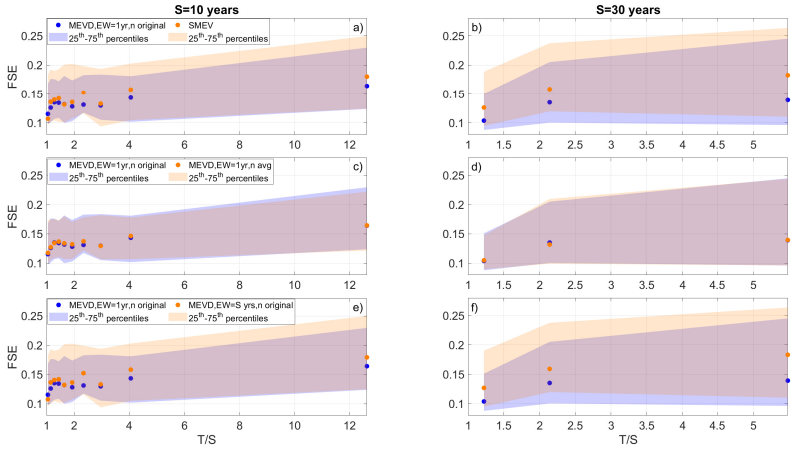


Figure 3.13: Comparison between the Fractional Standard Error (FSE) computed over the $N_r = 1000$ realizations for all the stations belonging to the historical dataset, plotted as a function or the ratio between the return period (T) and the calibration sample size (S) from the MEVD with the parameters estimation performed on a yearly basis and keeping the original number of events ("full" MEVD, in blue) and possible simplifications (orange). These are respectively: the SMEV (first row), the MEVD with a yearly parameters estimation and the approximation of the original number of wet days by their average number ("MEVD-EW=1yr-n avg", in the second row) and the MEVD with an EW as large as the calibration sample size and the original number of wet days ("MEVD-EW=Syrs-n original", third row). The two columns refer to the different sizes of the calibration sample: $S = 10$ years on the left and $S = 30$ years on the right. Dots represent the median value of the FSE in each bin and the shaded areas the 25^{th} and 75^{th} percentile. To keep the same number of values on which computing the statistics of the FSE for all the calibration sample sizes (and to have more than one bin in the case of $S = 30$ years), the binning is done in a way such that there are at least 40 values within each bin.

the "full" MEVD (third row in Figure 3.13). Results are very similar to (almost indistinguishable from) the comparison between the "full" MEVD and the SMEV. This experiment confirms the key role played by the inter-annual variability of the distributions and the negligible impact of the substitution of the original number of events with its average value.

3.3.3 Detectability of extreme value changes over time

The MEVD formulation with EW=1 year was shown to be, in wet climates, a practical solution since the adoption of a longer, though optimal, EW size does not afford significant estimation improvements. Having exhaustively characterized the estimation error dependence on the sample size, the benefit of leveraging the improved MEVD accuracy to resolve the inter-annual variability in daily rainfall extremes is now illustrated. The relevant case of the multi-centennial time series recorded in Padova (Italy) is examined; it allows the analysis of long-term changes in rainfall regimes (Marani and Zanetti, 2015). The focus is on the estimation of the daily rainfall depth, h_{100}, corresponding to a 100 year return period. Equivalently, since the analysis concerns changes in extremes that occur on time scales shorter than 100 year, h_{100} can be thought of as the value that is exceeded, in any given year, with a probability equal to $p = 0.01$. Estimates of h_{100} are performed with both MEVD and GEV (Weibull parameters for the MEVD are fitted on a yearly basis using PWM's, see Greenwood et al. (1979) and GEV parameters on annual maxima using L-Moments, see Hosking (1985); Hosking (1990)), over time windows of 10 and 30 years sliding over the time series with steps of 1 year.

It is interesting to notice potential trends in h_{100} as estimated with the two approaches. For the 10-year moving window (Figure 3.14a), h_{100}^{MEVD} smoothly fluctuates within the interval $h_{100}^{MEVD} \in [75 : 150]$, whereas the values of h_{100}^{GEV} presents several abrupt upward and downward shifts. This happens because the GEV distribution, being fitted to 10 yearly maxima, is very sensitive to the possible presence of a few very large or very low maxima in the analysis window (see the series of yearly maxima in Figure 3.15). The MEVD, on the contrary, uses all the daily rainfall values in the analysis window and is thus not very sensitive to a few large or low values, thereby leading to more stable estimates of the 100-year event. This time coherence of extremes afforded by the use of the MEVD is entirely analogous to the spatial coherence noted when estimating maps of extreme events from rainfall remote sensing estimates (Zorzetto and Marani, personal communication).
Fluctuations in h_{100}^{GEV} are more contained for the 30-year window (Figure 3.14b), due to the longer sample used for parameter estimation and because short fluctuation frequencies are dampened by the use of longer analysis windows (which is also visible in the case of MEVD estimates). The results

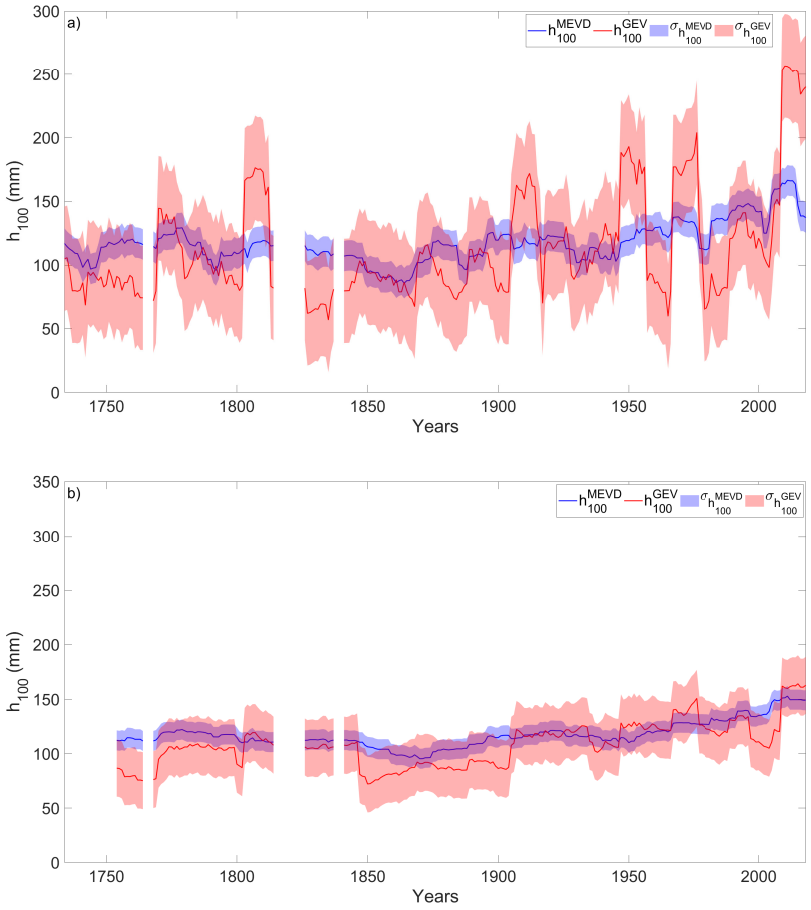


Figure 3.14: The plot in panel a (b) shows the estimation of the quantile associated to a return period of 100 years for the MEVD (blue line) and the GEV distribution (red line) on 10 (30)-year windows sliding over the whole time series of Padova station (step=1 year). The value of h_{100} obtained from each data window is associated with the last year in the window. The shaded areas represent a one standard deviation confidence interval computed by means of the Monte Carlo cross-validation process illustrated earlier. The first year for which h_{100} is computed is 1734 when using sliding windows of 10 years and 1754 in the case of the 30-year windows.

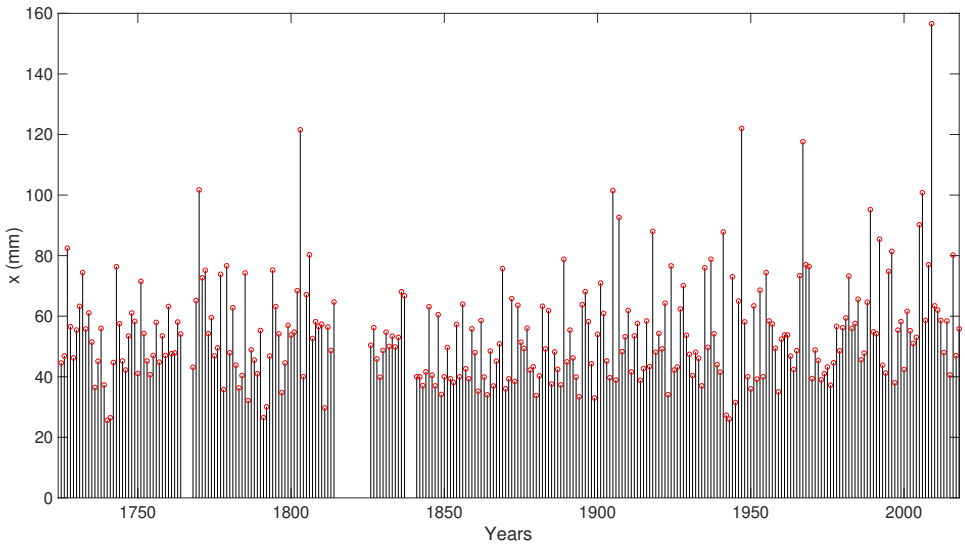


Figure 3.15: Annual maxima (red circles) for the series of daily rainfall of Padova station, from 1725 to 2018 (missing years: 1765-1767, 1815-1825, 1838-1840).

in Figure 3.14 highlight important features in MEVD and GEV estimates. Fluctuations in h_{100}^{GEV} (and its uncertainty bounds) are of the same size of the overall difference between the 100-year extreme at the end of the time series (modern times) and its estimate for the mid 18th century. h_{100}^{MEVD}, on the contrary, displays fluctuations (and confidence intervals) that are smaller than the difference in extremes between the end and the beginning of the time series. The narrower confidence interval characterizing MEVD also for short analysis window lengths is an additional advantage when change and trend analysis is the objective. In this case, the resolution with which estimates are produced are of central importance, and define the time scale at which change can be detected. Our results show, in concrete, that GEV provides very uncertain estimates when extreme value changes over time scales of 10 to 30 years are investigated.
Despite by means of a Mann-Kendall test (Mann (1945); Kendall (1975)) the h_{100} values estimated with the two approaches both show a non-stationarity, GEV estimates present abrupt steps that are unlikely to be realistic, hence making this approach not suitable for performing trend analyses.

3.4 Discussion

In this Chapter, the climatic factors determining, to first-order, the optimal MEVD approach to daily rainfall extreme value analysis have been identified. The MEVD approach has then been compared with the recently-proposed Simplified MEVD to quantify the possible impacts of such simplifications on the estimation uncertainty. Finally, the quantification of the

uncertainty of the MEVD was leveraged and its optimal formulation was adopted to show the benefits of its use in the study of long-term changes in daily rainfall extremes. In the following, the key points that can be highlighted from the analyses performed are summarized.

The average number of events and their variability throughout the years are not enough to identify an univocal correspondence with the length of the optimal estimation window (EW), unless the number of events within a series is very little (and this effect is sharpened by an increasing variance) or relatively small in combination with a small variance. In the last two cases, since the parameter estimation performed on the yearly basis would lead to high uncertainties due to the small sample available for a robust fit of a statistical distribution, the optimal length of the EW is larger than one year. As the number of events per year increases, the size of the calibration sample increases, with a consequent reduction in the uncertainties when performing a yearly parameters estimation.

The previous reasoning is not linearly linked with the average number of events/year and their variance. Through synthetic experiments, it was shown that the inter-annual variability plays a key role in the choice of the optimal EW. A high/low variability in the yearly distributions has an impact on the selection of the EW that maximizes the accuracy in the estimations. When the records show a high variability throughout the years, the transition from a EW of one year to a longer one occurs for low values of average number of rainy days, whereas there is not a clear shift when the statistical properties of the records are more uniform. These properties are therefore what really controls the choice of the EW. When their variability is high, considering aggregating years to widen the sample size used for parameters estimation is not beneficial; values belong to distributions that might be very different one from each other, hence the use of a single distribution to fit them is smoothing these differences over with the resulting failing of the advantage of having a large sample on which, theoretically, more robustly fitting a statistical model.

The "full" MEVD (i.e., the MEVD in which the parameters estimation is performed with a EW of one year and the number of events is the original one) is then compared with three simplified versions, the SMEV (Marra et al., 2019), the MEVD with a EW=1 year and the number of events substituted by its average ("MEVD-EW=1yr-avg n"), and the MEVD with the original number of events but using an EW of the same size of the calibration sample ("MEVD-EW=Syrs-n original"). The SMEV approach is undoubtedly appealing and it allows reducing the quantiles computational effort, but its use becomes relevant in dry climates or when there is the need of stratifying events in different types. Differences in the FSE comparison

between the "full" MEVD and the "MEVD-EW=1yr-avg n" were found to be negligible, confirming the importance of the inter-annual variability. The main contribution in the differences that were found in the FSE from the "full" MEVD and the SMEV is hence given by the use of a single distribution fitted on the whole calibration sample, as it was shown by comparing the FSE from the "full" MEVD and the one from the MEVD-EW=Syrs-n original.

Following the important finding of the crucial role exerted by the inter-annual variability, the property of the "full" MEVD to resolve the time-variability to study long-term changes in daily rainfall extremes was leveraged to compute the rainfall height associated to a return period of 100 years (h_{100}) using sliding windows on the long series of daily rainfall of Padova (1725-today) with both the MEVD and the GEV distribution. MEVD estimates are characterized by smooth fluctuations, while GEV estimates present some abrupt changes (especially when using a moving window of 10-years, due to the limitation of this approach to estimate values using short samples) that are unlikely to be realistic. Furthermore, in times during which lot of emphasis is laid by the scientific community on detecting trends and possible changes in the climate, the GEV is not a reliable method due to its incapability of following climatic variability (e.g., large scale oscillations) that must be carefully understood in order to assess possible climatic changes. The MEVD approach, instead, being able to resolve inter-annual fluctuations by using short time moving windows without being affected by the large errors of the traditional EVT, is a suitable model for trend detection.
With more practical implications, being the question of whether/to which degree updating design variables for engineering purposes widely discussed nowadays, it was shown here that the traditional EVT is not a suitable methodology to be applied.

Chapter 4

Metastatistical Extreme Value Distribution applied to floods across the continental United States: Use of mixed distributions and the impact of ENSO on flood regimes

This Chapter develops a MEVD approach tailored for flood frequency analysis (FFA) and, given the increasing interest in understanding and linking hydrological processes to the physical phenomena by which they are generated, a mixed-MEVD approach based on ENSO phases is proposed. In this application, the possible impact of ENSO on extreme flood regimes is evaluated.

4.1 Introduction

Globally, in the period 1998-2017, floods have been the most frequent disaster (43.4% of the natural disasters) and have caused more than 140,000 deaths (representing 11% of the fatalities due to natural disasters of all types) (Wallemacq and House, 2018). Within this global context, during the 20th century floods in the United States were the number-one natural disaster in terms of the number of lives lost and property damage (Perry, 2000), the costliest (Miller et al., 2008) and affected the largest number of people (Strömberg, 2007). They are also the second weather-related hazard in terms of fatalities in the United States, with 4,586 reported deaths between 1959 and 2005, mainly due to flash floods caused by heavy precipitation (Ashley and Ashley, 2008). Reliable flood frequency estimation methods are the basis to devise and implement strategies for the mitigation of these societal and economic impacts, with applications in a number of fields, from the design of hydraulic structures, to environmental management and planning, to flood insurance. A key concept is the design flood peak value, typically set in national regulations by specifying an average recurrence interval, or return period, T, associated with a probability of being exceeded in each year equal to $p=1/T$. The T-year flood, in turn, is estimated based on the analysis of past floods, requiring the selection of a

probability distribution to perform this inference using a sample with size $S \ll T$ (Benson, 1962).

Flood frequency analyses are commonly performed as a part of engineering and planning projects, but they too often represent a mere statistical fitting exercise, and do not attempt to incorporate a representation of the underlying physical processes. As noted by Klemeš (1974); Klemeš (1988); Klemeš (1993), the standard approach of extreme event probability estimation is moving towards a higher mathematical abstraction, renouncing any leveraging or understanding of the different flood-generating mechanisms at play. The hydrological-process information in the observations is thus often neglected, and the selection of an optimal statistical model through a goodness-of-fit metric remains the main focus of these types of analyses.

The U.S. federal guidelines themselves (Bulletin 17-B, IACWD (1982)), and its updated version 17-C (England, Jr. et al., 2018) recognize that the assumption under which stream gage records are generated by one single flood-generating mechanism may not always be realistic. They highlight the need to understand and more accurately identify these physical mechanisms and list the identification and treatment of mixed distributions to represent their diversity as a research and application priority.

Even though the idea of a more process-driven flood frequency analysis is not necessarily new (e.g. Hirschboeck (1987)), there has been a renewed interest in recent years in process-based formulations of extreme flood distributions (e.g., Alila and Mtiraoui (2002); Smith et al. (2011); Villarini and Slater (2017); Barth et al. (2019)). Flood-generating processes can quite naturally be analyzed using mixed distributions (e.g., Alila and Mtiraoui (2002)), but the determination of which flood peaks result from the different processes and of when the use of mixed distributions is beneficial remains an open problem (e.g., Villarini and Slater (2017)).

There are several different hydrological mechanisms that can drive the occurrence of flood events, including snowmelt, frontal systems, local convective processes, monsoons, and intense tropical cyclones (see Villarini (016a); Zhang et al. (2017)). Slater et al. (2015) compared hydrologic and geomorphic drivers in flood hazard, while Berghuijs et al. (2016) analyzed the dominant flood generating mechanisms across the United States. Barth et al. (2017); Barth et al. (2019) investigated the role of atmospheric rivers in the generation of flood peaks across the western United States and suggested a weighted mixed population approach to perform a process-driven flood frequency analysis to reflect the differences in flood agents.

Within the context of mixed distributions, this Chapter formulates a novel flood frequency distribution and uses it to investigate the role of El Niño

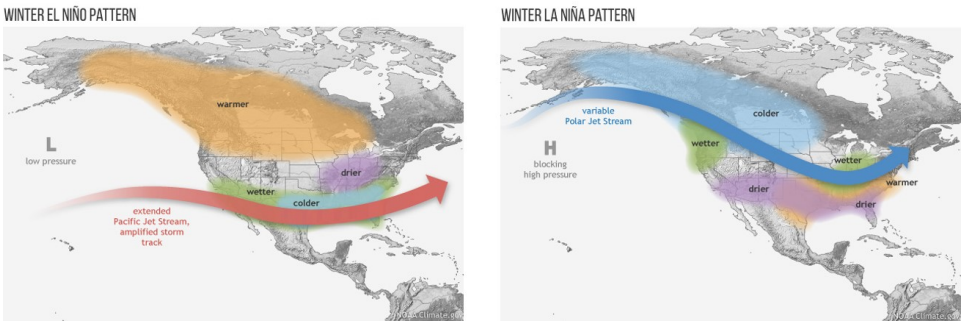


Figure 4.1: El Niño and La Niña winter patterns over North America. From: NOAA Climate.gov

Southern Oscillation (ENSO) in the generation of floods, and the detectability of its signature in observed records. ENSO is a major mode of variability of the coupled atmosphere-ocean system associated with episodes of above-normal (El Niño) and below-normal (La Niña) sea surface temperature in the tropical Pacific Ocean, with impacts on seasonal winds, rainfall, and temperature across the globe. See Figure 4.1 for a visualization of the effects of El Niño and La Niña on North America.

Here the use of a novel approach is proposed, the Metastatistical Extreme Value Distribution (MEVD), which can naturally incorporate mixed distributions to represent flood magnitudes generated by different mechanisms. The MEVD has been introduced by Marani and Ignaccolo (2015) and has been applied mostly to rainfall (Zorzetto et al. (2016); Marra et al. (2018); Zorzetto and Marani (2019)), for which it was shown to provide significantly smaller estimation uncertainty when compared to traditional approaches, especially when considering return periods that are larger than the sample size used for distribution estimation.

Currently, the MEVD has yet to be applied to flood magnitudes and flood frequency analysis. Here the first such application is provided and the following relevant questions are asked: does the MEVD outperform the traditional Generalized Extreme Value (GEV) distribution in flood frequency analysis? Does the incorporation of mixed probability distributions representing different types of flood events associated with different ENSO phases improve the estimation of event magnitudes with high return periods?

To answer these questions, the MEVD approach is applied to daily records from stream gage stations across the continental United States, examining the role played by mixtures of distributions associated with different ENSO phases. The results are compared and contrasted against those from the GEV distribution, providing qualitative and quantitative evaluations of their relative predictive performance.

4.2 Data and methodology

Daily records from 5,311 U.S. Geological Survey (USGS) stream gages across the continental United States are analyzed (Figure 4.2, panel a). The focus is on water years, defined to run between October 1 and September 30, and only sites where at least 30 complete (i.e. with more than 330 daily observations/year) years of observation exist are selected, and where no statistically significant trends are found (at the 5% level, based on the Mann-Kendall test; Mann (1945); Kendall (1975)). The historical time series selected cover the period 1916-2017, with record lengths between 31 and 101 years (Figure 4.2, panels b and c). Consult Figure 4.3 for an overview of the spatial distribution of the lengths of the historical time series.

Flood frequency analysis requires the identification of independent events; here, two approaches have been tested for the independent peaks selection:

1. the largest flood peaks within time windows of length equal to $T = 10 \cdot days + log(A)$, where A is the drainage area in square miles (Lang et al., 1999) are identified. Additionally, the smallest discharge peak within any pair of consecutive peaks is discarded if the minimum flow between them does not drop below a threshold equal to 75% of the lower of the two (Water Resources Council, USWRC (1976)). This additional condition is necessary to eliminate secondary peaks occurring during recession periods of previous floods;

2. a threshold equal to the 75^{th} percentile of the historical discharge value is set, flood events are identified between flow up-crossing and down-crossing of the threshold and the maximum value within the flood event is selected.

The entire set of peak discharge values of uncorrelated events resulting from the two selection processes is here called the set of "ordinary events" to denote that it contains all the independent events that have occurred in the record, irrespective of their magnitudes. One example of peaks selected with the two methods (for the station of St. John River at Ninemile Bridge) is shown in Figure 4.4.

The second method was yet discarded, since the spatial distribution of the average number of peaks per year does not reflect the different climatic areas in the CONUS and smoothes them over; it indeed brings to the selection of very few peaks in most of the country (see Figure 4.5 for the number of peaks selected with the first (panel a) and second (panel b) method).

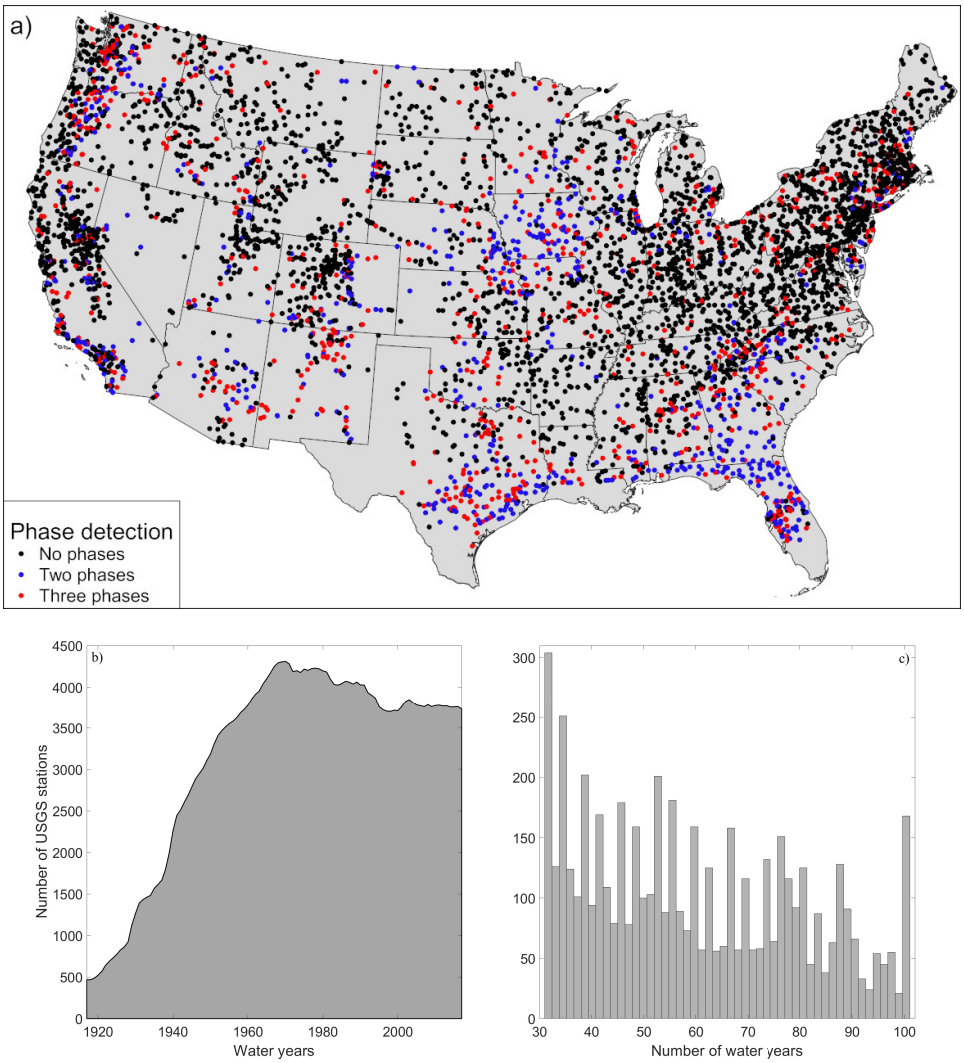


Figure 4.2: Panel a: Spatial distribution of the selected stream gages. The colors refer to the statistical detectability of ENSO phases in the distribution of discharge peak values: black dots indicate stations where the frequency distributions of ordinary peak discharge values in different ENSO phases are indistinguishable from one another; stations for which three different phases are detected are displayed in red; in blue, stations where two out of three ENSO phases can be distinguished. Panel b shows the cumulative number of stations with data in each water year, while the histogram in panel c summarizes the number of stations in terms of the number of years on record.

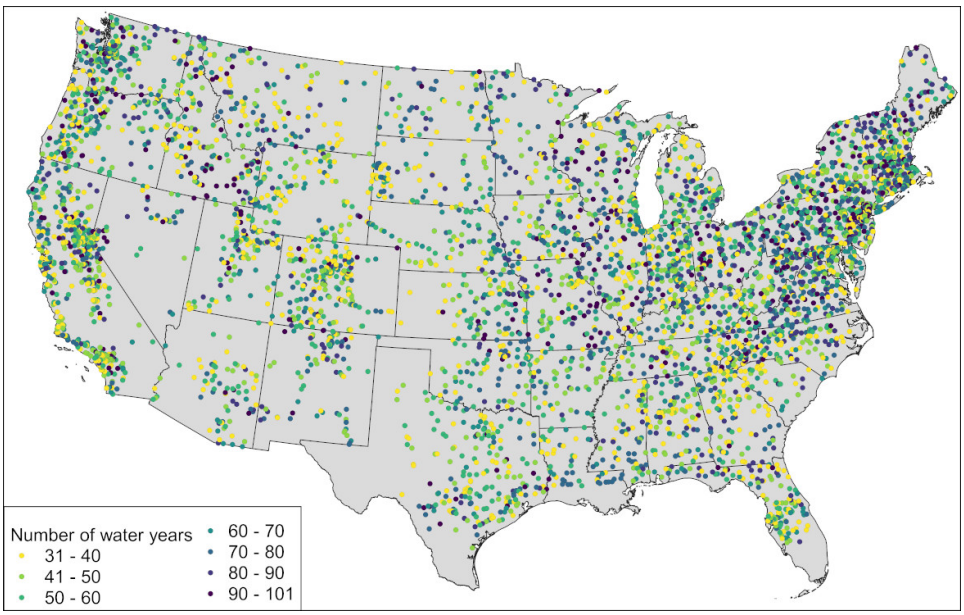


Figure 4.3: Number of water years belonging to each historical series. Colors refer to different ranges of length.

4.3 MEVD approach

As recalled in Chapter 1, the MEVD, originally introduced by Marani and Ignaccolo (2015), is explicitly formulated on the basis of the probability distribution(s) of the ordinary values, from which the distribution of extremes (annual maxima) is then derived. Hence, the expression of the MEVD describing the extreme events is identified by estimating its parameters using the entire set of observed ordinary events (see Eq.(2.5) and Eq.(2.6)). This is quite different from the assumptions at the basis of the traditional Extreme Value Theory (EVT), which focuses on fitting a distribution to the annual maxima or to relatively few values above a high threshold.
Here, the MEVD is applied to peak discharges and modified to account for ordinary values belonging to different populations, corresponding, in the present case, to different ENSO phases.
The cumulative distribution function $\zeta(x)$ of the mixed-MEVD can be written as follows:

$$\zeta(x) = \frac{1}{M} \sum_{j=1}^{M} \prod_{p=1}^{n_{ph}} \left[F_p(x; \vec{\theta_j}) \right]^{n_{j,p}} \tag{4.1}$$

where n_{ph} is the number of phases that induce statistically different distributions of the ordinary events and should therefore be considered separately; F_p is the yearly (or time-window, when the low number of events/year requires parameter estimation to be performed of multi-year windows) cumulative distribution of the ordinary values in phase p; $n_{j,p}$ is the original

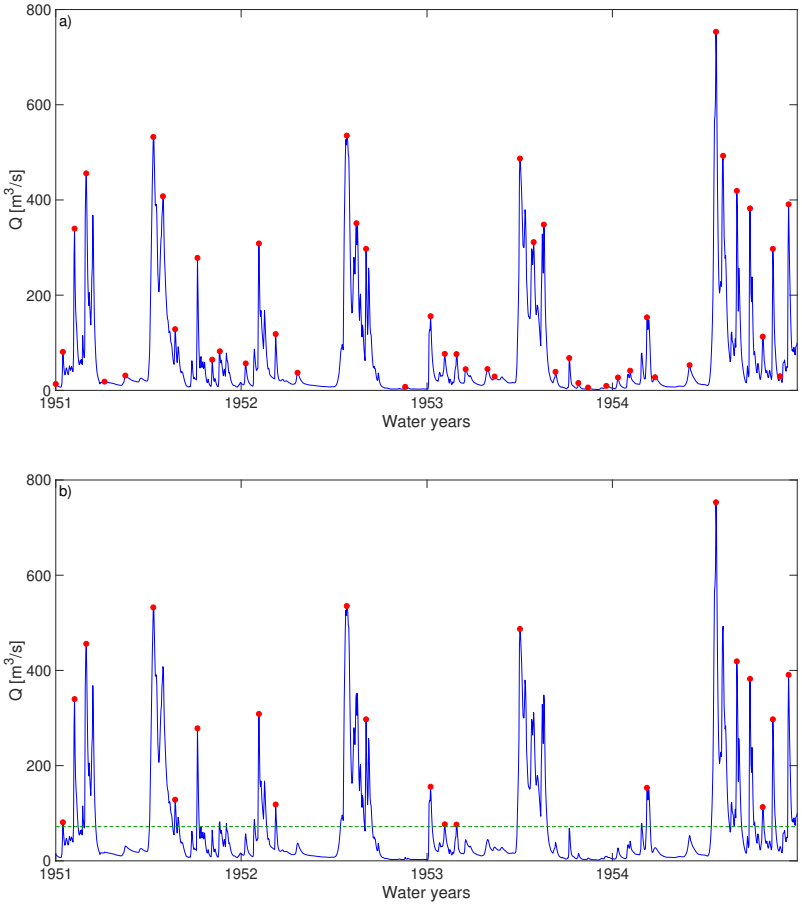


Figure 4.4: Peaks selected (red dots) within a 4 water years time window (1951-1954) using the historical record of St. John River at Ninemile Bridge station. The plot in panel a shows the peaks selected applying the first method, i.e., criterium of Lang et al. (1999) and the additional control from the USWRC (1976) are applied. The plot in panel b shows instead the peaks selected using the second method (the threshold equal to the 75^{th} percentile of the flow values is represented by the green dashed line).

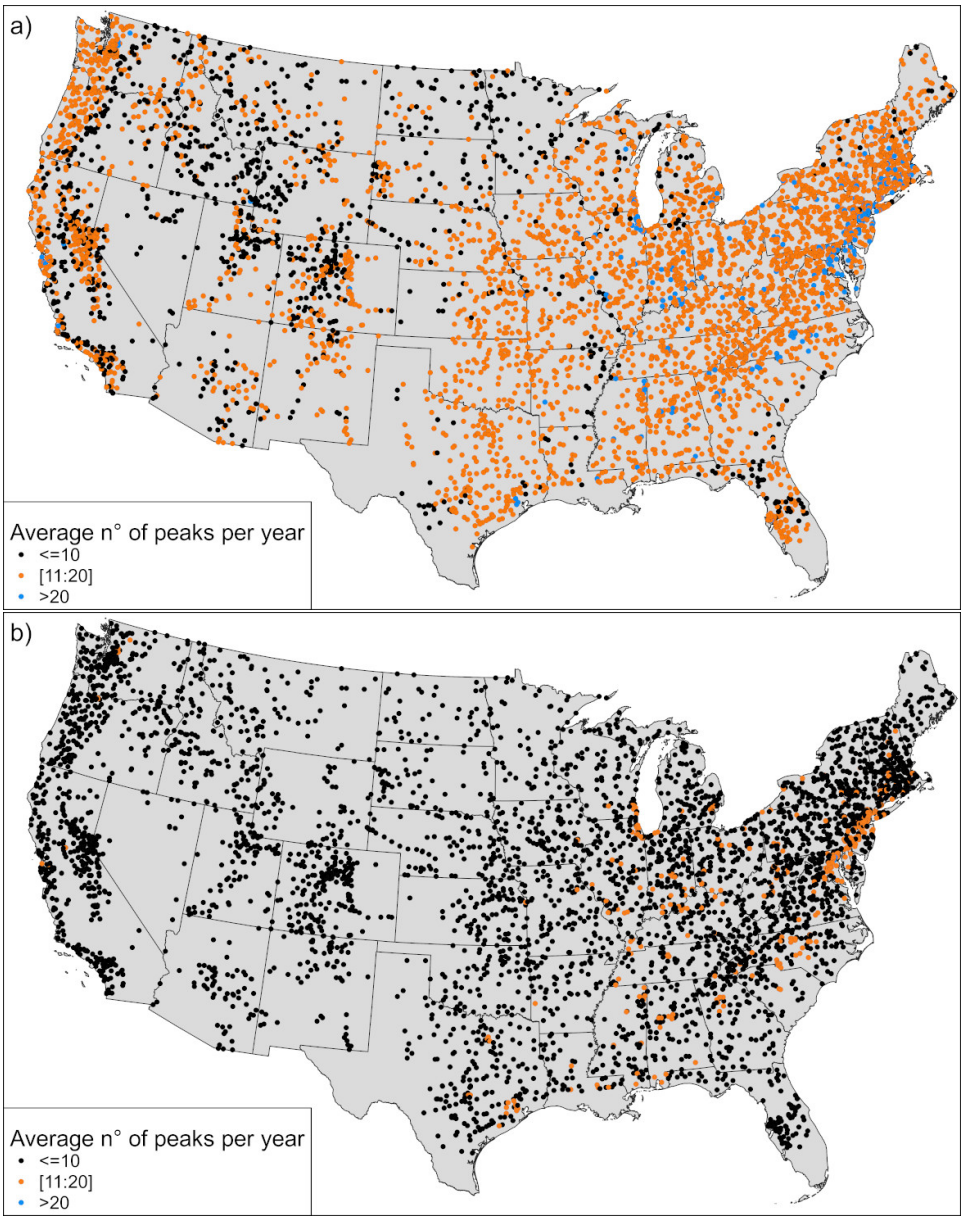


Figure 4.5: Panel a (b) shows the average number of peaks per year selected with the first (second) method. Black dots represent those stations for which the number of peaks per year is lower or equal than 10 (1243 vs 4249 stations), orange dots those with an average yearly number of peaks between 11 and 20 (3801 vs 320 stations) and light-blue dots are the stations with a number of peaks per year greater than 20 (267 vs 2 stations).

yearly number of the peaks in phase p and year j; M is the number of years for which observations are available. Eq.(4.1) reduces to the original formulation in Marani and Ignaccolo (2015) when only one phase is present.
The mixed MEVD formulation in Eq.(4.1) is similar to an approach proposed in Marra et al. (2019), which, however, does not account for the inter-annual variability of the events (i.e. there is no dependence on j in Eq.(4.1)).

The first step in the application of the MEVD approach is identifying a suitable parametric distribution to represent the ordinary events. Three candidate distributions for the $F(x;\vec{\theta_j})$ in Eq.(2.6) are evaluated: Weibull, Generalized Pareto, and Gamma distributions. The most suitable distribution is selected on the basis of the Skill Score (see Section on the evaluation metrics for its definition), comparing the estimated quantiles to the observed ones. In the present analyses, the Gamma distribution was the best performing one (see Section 4.5 for further details).

Because the average number of flood events in a year is small for many stations (e.g., in 1243 of the 5311 analyzed stations the average number of peaks/year is smaller than or equal to 10; see Figure 4.5a), parameters estimation is performed using five-year windows or the entire sample. Unlike the application of the MEVD to the analysis of daily rainfall (Zorzetto et al. (2016)), the use of a yearly estimation window is not considered here due to the potential inaccuracies in the estimation of the parameters when few values are available. This happens because the autocorrelation in discharge is much larger than in precipitation, leading to the need to use an inhibition window to identify independent flood events as previously described. This problem is further exacerbated when stratifying the data into different components of the mixture of distributions: the use of multi-year windows for parameter fitting is then necessary to make sure that, in periods in which multiple ENSO phases are present, there is a sufficient number of flood events in each phase to allow a robust parameter estimation.
Following the results presented in Chapter 3, the estimation window was kept as short as possible, provided that the number of events/block was large enough to robustly estimate parameters (hence the choice of 5 years as the minimum size for the blocks). The same sizes of the blocks used for parameter estimation were then used when dividing the peaks according to the different ENSO phases the mixed-MEVD. As discussed in the following, there is not optimal a priori choice of the size of the blocks, but the "optimal MEVD", is defined as the one that minimizes the estimation uncertainty.

4.4 Fitting Procedure and Cross-Validation

4.4.1 Fitting Procedure

GEV fitting is performed on annual maxima using L-Moments (Hosking, 1990). Zorzetto et al. (2016) found that the cross-validation performance of the Peak-Over-Threshold GEV fits is indistinguishable from the performance of Maximum Likelihood or L-Moments GEV fits on annual maxima. Hence, we only present here results from the application of the latter. The parameters of the yearly distributions in the MEVD (Eq.(4.1)) are estimated on data from five-year windows or the whole sample. MEVD-Gamma fitting is performed on the selected independent peaks via L-Moments (Hosking (1990)).

4.4.2 Evaluation Metrics

To identify the possible signature of ENSO phases in the distributions of ordinary flood peaks, each event is assigned to one of the three ENSO phases based on the Extended Multivariate ENSO Index (`https://www.esrl.noaa.gov/psd/enso/past_events.html`). The phases are defined with a monthly time span by means of an index: -1 for El Niño, 1 for La Niña and 0 for the neutral phase. Consequently, a test on whether the distributions of ordinary flood peaks for each phase are different from one another is done using the Kolmogorov-Smirnov test with the Bonferroni correction (Bonferroni (1936)) to account for multiple hypotheses testing (the three possible combinations among the phases, in this case). If the distributions of the peak magnitudes belonging to two separate ENSO phases are not statistically different at the 5% level, all discharge peak values from both phases are combined. Hence, each time series is classified in the dataset depending on whether: 1) three separate ENSO phases are distinguishable in the empirical distribution of ordinary flood peaks; 2) two separate ENSO phases are distinguishable (i.e. peaks from two of the phases were merged); 3) no ENSO phases are statistically different from each other in the set of ordinary events (see Figure 4.6 for an example of peaks distinction).

The values of the empirical cumulative frequency associated with observed peak discharge in the test sub-sample are estimated using the Weibull plotting position ($F_k = k/(L+1)$, where k denotes the *k-th* peak discharge value, Q_k, in an ascending order ranking). The estimated quantiles corresponding to each value F_k are computed using the MEVD and GEV distribution by solving $MEVD(Q_k^{MEVD}) = F_k$ and $GEV(Q_k^{GEV}) = F_k$, respectively.

To evaluate goodness-of-fit and estimation accuracy two metrics are used:

1. estimated quantiles are compared with observed ones through the computation of the Skill Score (SS $\in (-\infty; 1]$) (Murphy and Win-

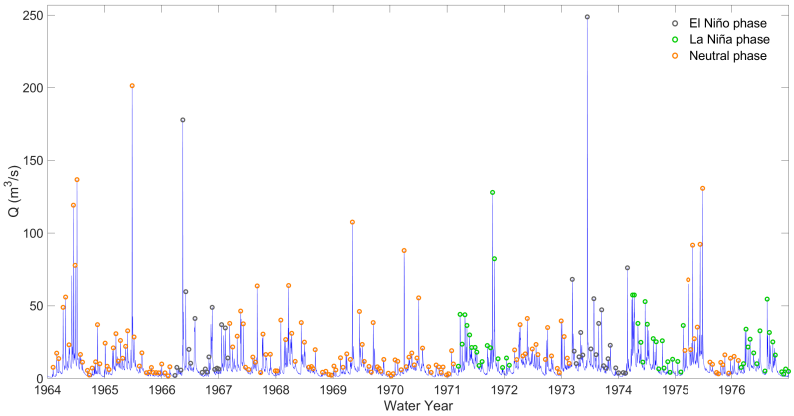


Figure 4.6: Portion of the hydrograph for Little River Above Townsend (TN) station for which the empirical distributions of the the three phases have been detected as statistically different. Gray, green and orange circles indicate the peaks respectively occurred during El Niño, La Niña and the neutral phase.

kler (1992); Hashino et al. (2006)), which provides a global metric of estimation accuracy:

$$SS(Q_{est}, Q_{obs}) = \rho^2_{Q_{est},Q_{obs}} - [\rho_{Q_{est},Q_{obs}} - (\sigma_{Q_{est}}/\sigma_{Q_{obs}})]^2 + -[(\mu_{Q_{est}} - \mu_{Q_{obs}}/\sigma_{Q_{obs}})]^2 \quad (4.2)$$

where $\rho_{Q_{est},Q_{obs}}$ is the correlation between the estimated values (Q_{est}) and the observations (Q_{obs}); $\sigma_{Q_{est}}$ and $\sigma_{Q_{obs}}$ ($\mu_{Q_{est}}$ and $\mu_{Q_{obs}}$) represent the standard deviation (mean) of the observations and estimations, respectively. The SS accounts for the potential skill (i.e., coefficient of determination) as well as conditional and unconditional biases. The SS is used both in the context of ordinary values fitting and of extreme values estimation evaluation. In the latter case, to provide a measure of the estimation of high quantiles, the terms in the skill score definition (Eq.(4.2)) are computed only on quantiles with return period $T_k = (1 - F_k) - 1 > S$, i.e. greater than the length of the dataset used for calibration. This reflects application needs, which target the estimation of extremes with return period much greater than the length of the observational time series available (estimation of quantiles with $T_k \leq S$ can be performed empirically, without the need to assume a specific probability distribution);

2. the non-dimensional error is computed as:

$$\varepsilon_j(S,T) = \frac{[Q_{est,j}(S,T) - Q_{obs,j}(S,T)]}{Q_{obs,j}(S,T)} \quad (4.3)$$

for which the whole frequency distribution is estimated based on the $N_r = 1000$ Monte Carlo realizations.
Over these realizations, the Fractional Standard Error (FSE) is finally computed:

$$FSE(S,T) = \Big[\frac{1}{N_r}\sum_{j=1}^{N_r}\varepsilon_j(S,T)^2\Big]^{\frac{1}{2}} \tag{4.4}$$

4.4.3 Cross-Validation

The uncertainty in estimating high quantiles associated with the use of the MEVD (in its single- or multi-phase versions) and of the GEV distribution is quantified by means of a cross-validation procedure involving Monte Carlo simulations (with $N_r = 1000$ realizations for each station) as follows:

1. the observational sample is divided into two sub-samples obtained by randomly selecting S years from the original time series of length L_{tot}: this sub-sample is used for parameter estimation, while data in the remaining $L = L_{tot} - S$ years are used for testing;

2. in every realization, both the SS and the FSE between the estimated and observed quantiles are computed, as described in the Section about the estimation metrics;

3. the whole procedure above is performed for different calibration sample sizes ($S = 10$, 20, and 30 years), to evaluate how estimation uncertainty varies jointly with return period and calibration sample size.

4.5 Results

4.5.1 Ordinary Values

First, the most appropriate parametric distribution for ordinary peak discharge values is selected, based on the SS computed for the ordinary peaks. Among the stations analyzed, the distribution that provides the highest SS in most of the cases (71%) is the Gamma distribution (Figure 4.7), which has thus been selected for use in comparative analyses over the whole CONUS. This is consistent with other studies in the literature (e.g., Hann (1977); Palynchuk and Guo (2008); Villarini and Strong (014c); Slater and Villarini (2017)).
For most of the stations (3718 or 70% of the total), no statistically different distributions were detected. At 883 sites (about 17%) the three phases are all different from one another. For the remaining 13% of the stations, common distributions were found to be shared by El Niño and the neutral phases or by La Niña and the neutral phase (Figure 4.2). In most of the

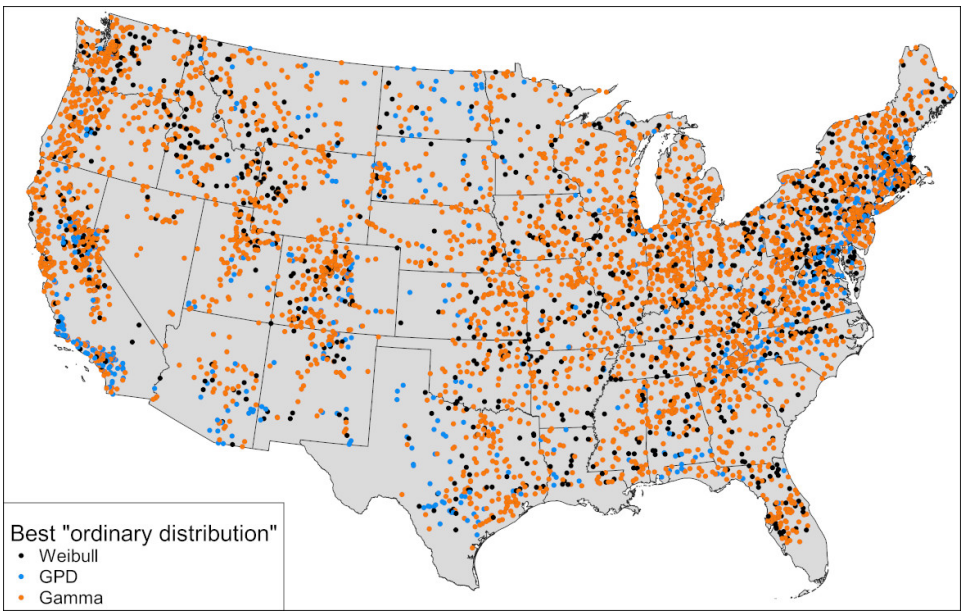


Figure 4.7: The map shows the best distribution for the ordinary events based on the skill score. A black dot means that for that station a MEVD-Weibull would be the optimal one, a light-blue dot indicates a MEVD-GPD as the best performing distribution and the orange dots are those stations for which a MEVD-Gamma is the best MEVD.

analyzed cases, different ENSO phases are detectable in ordinary peak discharge values in areas located in the eastern and southern United States, which are known to be more strongly affected by ENSO (e.g., Emerton et al. (2017); Mallakpour and Villarini (2017)). Detectable phases were found also in a group of stations in the U.S. Midwest, which are not usually areas that are affected by the effects of ENSO, but other processes might play a role.

4.5.2 Extreme Values Analysis

Now the question of evaluating the predictive performance of the MEVD-Gamma formulations comes into play; the predictive performances of these two new approaches is compared with those from the traditional GEV distribution. Then, the potential benefits of including ENSO phases in extreme flood estimation is quantified.

When comparing the "optimal MEVD" (i.e., the MEVD formulation based on the number of phases that yielded the maximum SS value computed on yearly maxima), the MEVD outperforms the traditional approach in ∼78% of the stations based on the SS metric; Figure 4.8 shows the results from a S=10 years calibration period in terms of the relative difference between the SS from the optimal MEVD and the SS from the GEV distribution, divided by the absolute value of the SS for the GEV distribution here assumed as a

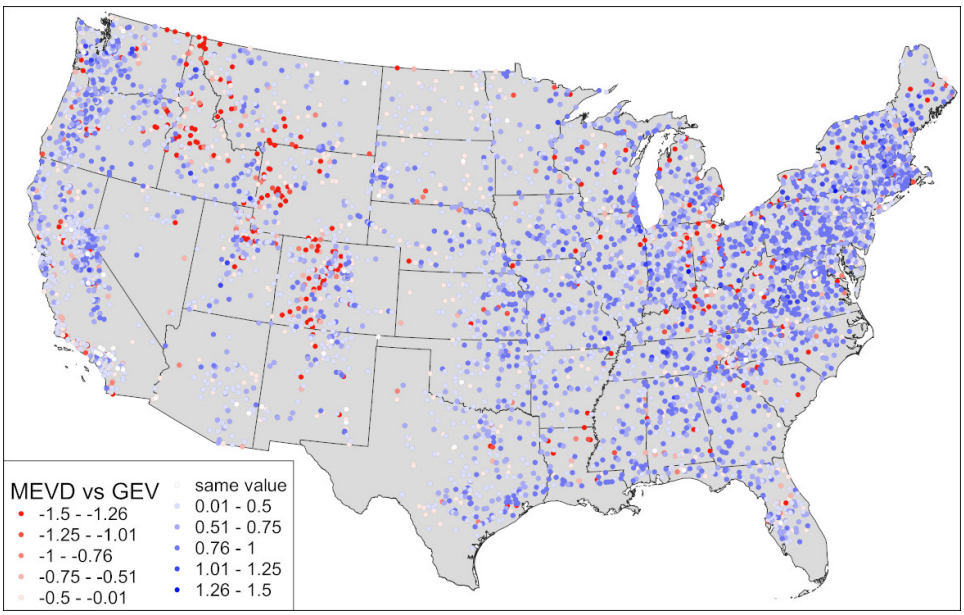


Figure 4.8: Comparison between the optimal MEVD (i.e. the one yielding the highest SS, regardless of the approach that is applied) and the GEV distribution based on the SS values, averaged over all the 1000 Monte Carlo realizations; colors indicate which distribution provides the highest values of the SS for each station (tones of blue for the MEVD and tones of red for the GEV distribution, respectively). Differences are computed between the SS of the MEVD and the one of the GEV distribution, with respect to the absolute value of the SS of the GEV: $(SS_{MEVD}–SS_{GEV})/|SS_{GEV}|$). White dots indicate the stations for which the SS values are equal, at the first decimal digit.

reference. Some areas where the performance of the GEV distribution is generally higher can be identified. They are characterized by a small number of uncorrelated flow peaks/year (e.g. the Rocky Mountains and southern California; see Figure 4.5a), usually in combination with short historical time series.

In addition to characterize one method's predictive performance globally, there is also an interest in focusing on the prediction accuracy for high return periods, being this the case for most practical applications. The return period associated with the maximum value in each test sub-series is estimated as $T_{max} = L_{tot} - S + 1$, where L_{tot} represents the length of the historical series: it is variable among the analyzed stations and ranges between 22 and 92 years. When looking at the FSE(S=10 years, T_{max}) computed for the highest return period from the MEVD and GEV approaches, the MEVD outperforms the GEV distribution in about 76% of the analyzed stations (Figure 4.9). The information provided by Figure 4.9 is complemented for all return periods in Figure 4.10, where the FSE is plotted as a function of the ratio between return period and calibration sample size S=10, 20 and 30 years in the different rows). Focusing on the first row of Figure 4.10

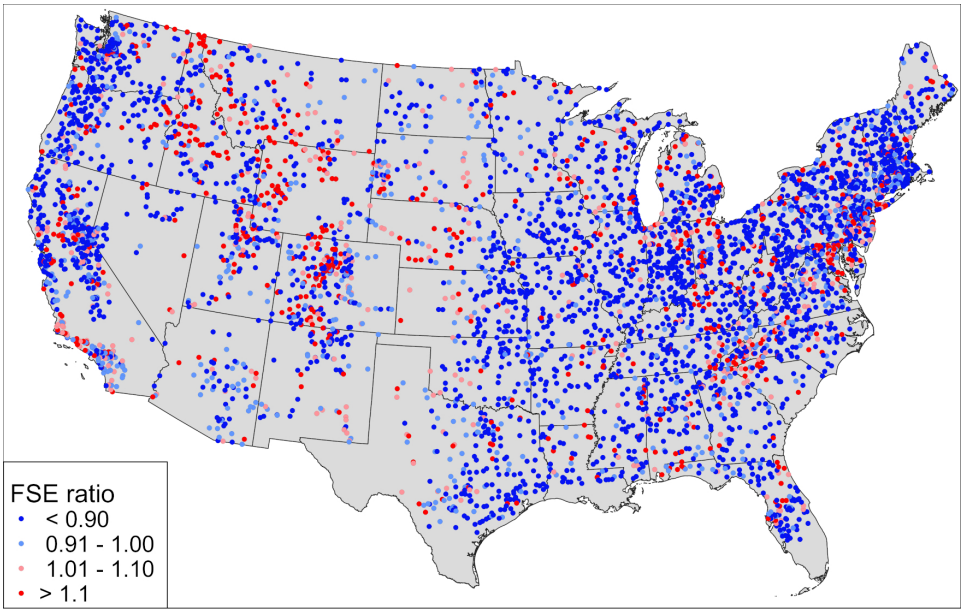


Figure 4.9: Ratio between the Fractional Standard Error from the optimal MEVD (defined on the basis of the SS between the single-phase and multi-phase approach) and the FSE from the GEV distribution, computed for the highest return period at each station. Blue dots represent sites where the MEVD outperforms the GEV distribution (the ratio between the two FSEs is lower than one), while red dots indicate those stations in which the GEV distribution is providing a more accurate estimation. Shaded colors indicate small differences between the two approaches.

(S=10 years), for small values of the ratio between the return period (T) and the calibration sample size (S) the errors in the estimations computed with the two EV approaches are comparable both in terms of average value and uncertainty. When higher values of T/S are considered, the estimates provided by the traditional GEV distribution are less accurate than those provided by the MEVD approach, which shows a 30% improvement with respect to GEV estimates. Many engineering applications are almost exclusively focused on high return periods (i.e., return periods much larger than the span of observational time series): the uncertainty of the GEV-based estimates shows a steadily increasing trend of the FSE with increasing return period, while the MEVD estimation error stabilizes around a value of about 0.32 for high values of T/S. This result suggests that, when estimating quantiles corresponding to return periods much larger than values that have been observed, the estimation errors of the traditional EVT approach will become very large, and much larger than for the MEVD-based estimates. The results of the FSE computed with calibration sample sizes of 20 and 30 years (second and third row in Figure 4.10) are consistent with those from S=10 years, yet limited to smaller ratios of T over S.

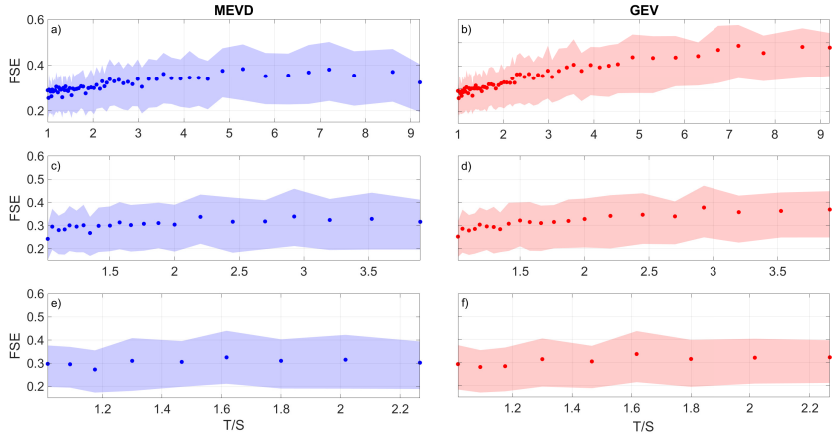


Figure 4.10: Fractional Standard Error (FSE) for the MEVD (in blue) and the GEV distribution (in red) averaged across all the stations for return periods greater than the calibration sample size, plotted as a function of the ratio between the return period (T) and the length of the calibration sample size (S=10 years in this case). The dots represent the mean values computed across all the stations and over T/S bins that include at least 300 values. The shaded areas are limited by the 25^{th} and 75^{th} percentiles. The different rows are related to the calibration sample sizes (S = 10, 20 and 30 years).

The average number of peaks/year exhibits a large variation across space (from 3 to 31) suggesting that different analysis approaches may be differently effective in dry areas, where extremely few flood events are observed, and more humid areas, where the larger number of events/year makes available larger quantities of data. Also in consideration of the spatial pattern identified in Figure 4.8, it is thus interesting to analyze the possible dependence of the estimation performance associated with different EV approaches with respect to the number of floods/year. Figure 4.11 shows the FSE plotted as a function of T/S for two groups of stations representing two end-member cases: sites with less than 10 events/year and sites with more than 17 events/year (limits are defined in such a way that both groups include about 1000 stations). The advantage in the use of the MEVD approach instead of the traditional one is limited to higher values of T/S when few peaks are selected, while it always outperforms the GEV distribution when a greater number of peaks is available. This is linked to the fact that, for the same T/S, having a small number of peaks is not adding much information to the distribution of maxima, like it does when the number of peaks increases. However, the robustness of the MEVD with respect to the GEV distribution is confirmed for high ratios of return period over sample size.

In the end, the focus is on answering the second question, i.e. whether it is beneficial to adopt a mixed-distribution MEVD approach accounting

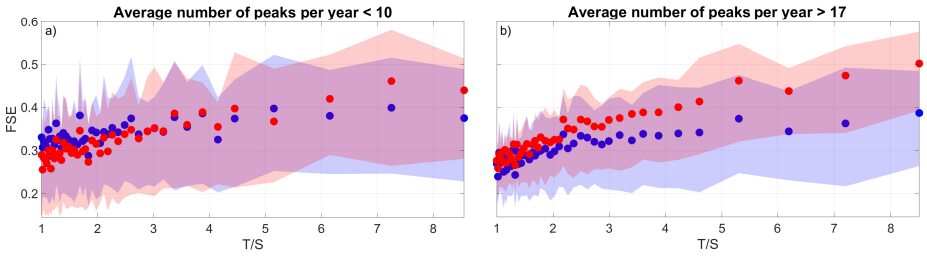


Figure 4.11: Fractional Standard Error as a function of T/S for the MEVD (blue) and GEV distribution (red). Dots represent the mean of the FSE in each bin (the width of the bins is chosen such that they contain at least 100 values). Shaded areas are limited by the 25^{th} and 75^{th} percentiles. Panel a refers to sites where, on average, less than 10 events/year occur. Panel b shows results for sites where the average number of yearly flood peaks is greater than 17.

for the different ENSO phases when estimating extreme flood magnitudes. Figure 4.12 shows the performance of the optimal MEVD when the single-component approach and the mixed one are compared, on the SS basis (the performance of the two MEVD approaches is presented as the relative difference between the SS from the single-component MEVD and the one from the mixed approach, divided by the absolute value of the SS for the mixed MEVD here assumed as a reference). It was found that including mixtures of distributions is not actually improving the estimations: given the highest SS, values are equally divided between mixed and single-population MEVD and in most of the cases there is no difference between them. Moreover, whenever the mixed MEVD is selected, the value of the SS is generally comparable to what obtained using a single MEVD. The signal that was detected in the ordinary distributions is confirmed by the use of a mixed distribution for the estimation of extremes only along the eastern and south-eastern United States, which are known to be more strongly affected by ENSO.

These conclusions are corroborated by the analyses of the FSE values obtained from multi-phase and single-phase MEVD approaches (Figure 4.13), which show negligible improvement in the estimation accuracy when adopting a mixed-distribution MEVD based on multiple ENSO phases.

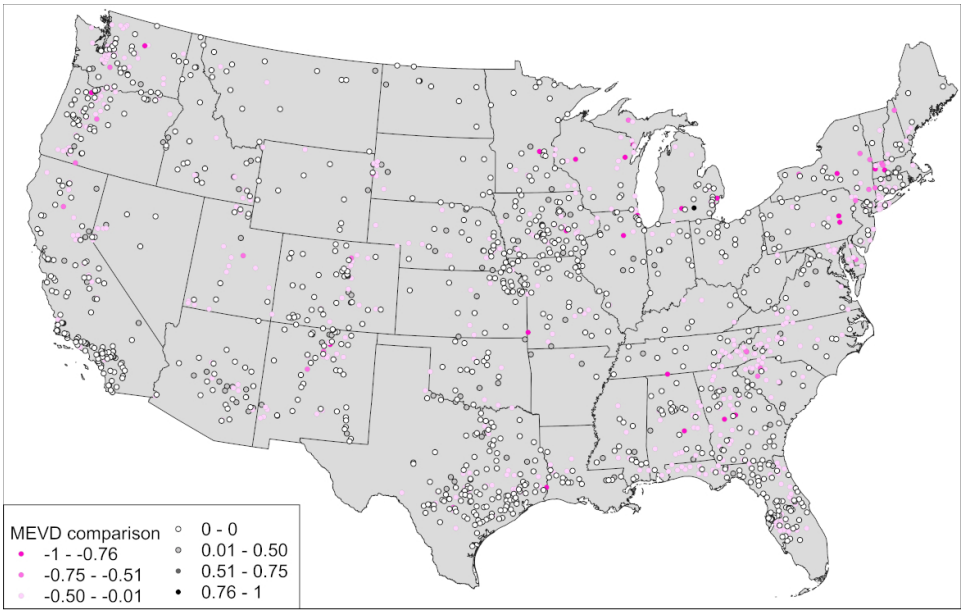


Figure 4.12: Map showing the optimal MEVD for those stations in which different phases have been detected by means of the Kolmogorov-Smirnov test. Magenta (black) dots represent stations for which the highest skill score is the one obtained (not) including the different phases in the MEVD. The relative performance of the two MEVD approaches has been evaluated through through the ratio: $(SS_{non-mixed-MEVD} - SS_{mixed-MEVD})/|SS_{mixed-MEVD}|$. White dots represent the stations for which this ratio is equal to 0.

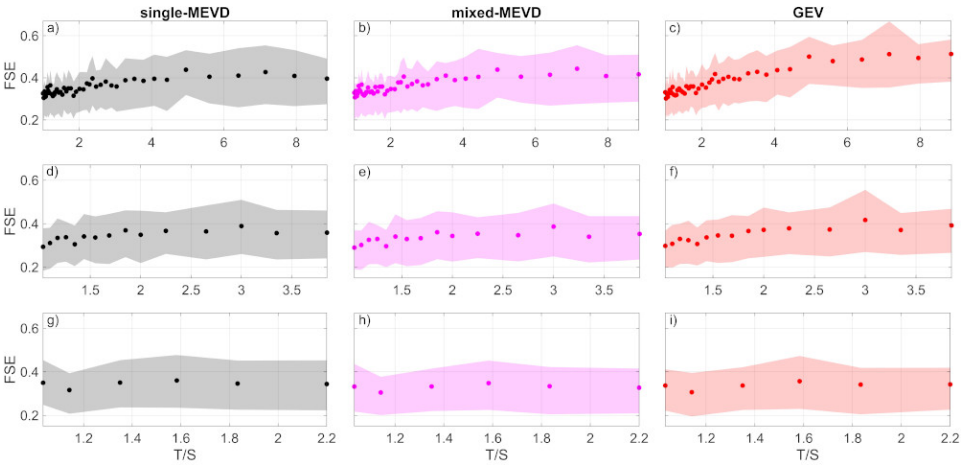


Figure 4.13: Fractional Standard Error as a function of the ratio return period (T) over calibration sample size (S) for those stations in which both the single-distribution and the mixed MEVD have been applied. The columns represent the different EV approaches (black is used for the single-component MEVD, magenta for the mixed approach and red for the GEV distribution), the rows the calibration sample sizes (S=10, 20 and 30 years) that have been evaluated. The dots represent the mean value, while the filled area is between the 25^{th} and 75^{th} percentile.

4.6 Discussion

This Chapter develops and applies for the first time an adapted formulation of the Metastatistical Extreme Value Distribution to flood peaks observed at more than 5,000 USGS stream gages across the continental United States. The key points can be summarized as follows.

1. The MEVD approach provides more accurate estimates of high flood peak discharge quantiles in about 76% of the stations, especially when short records are available to estimate the parameters of the distribution. The MEVD displays the smallest Fractional Standard Error for small calibration sample sizes and high return periods, the case of greatest practical interest.

2. When considering ENSO as the discriminating factor, the estimation of high return period flow values is not necessarily improved, even though the ENSO signature in the distributions of ordinary flood peaks were identified as statistically significant at a non-negligible number. Either the uncertainty intrinsic to extreme value estimation overwhelms the effects of ENSO phases or just one of the ENSO phases detected in the distribution of the ordinary events dominates the shape of the distributional tail. However, the introduction and formalization of this mixed-distribution MEVD for extreme value analysis remains important because it lends itself to applications in other contexts and where different physical drivers need to be considered, such as the North Atlantic Oscillation in western Europe (Marani and Zanetti, 2015) or the Arctic Oscillation in north eastern Europe (Bartolini et al., 2009).

3. ENSO signal is well recognisable in the distributions of the ordinary peaks; therefore, even if it was found that including mixtures of distributions in the MEVD approach brings to negligible improvement in the estimation accuracy of high return period quantiles, it can still provide valuable information. The index that distinguishes between an El Niño or La Niña year is forecasted in advance through climatic models. Knowing that a specific area needs to distinguish between two or three phases can support FFA in terms of which distribution should be use to make more accurate flow estimations.

Chapter 5

A mixed MEVD for different rainfall types: the case of tropical cyclones.

Extreme rainfall is a widely relevant topic, and much effort is devoted to the estimation of high return period rainfall events. Among the mechanisms that can generate extreme rainfall Tropical Cyclones (TCs) are an important example, with far-reaching socio-economic impacts. Here a mixed MEVD approach is explored in order to understand whether including the different rainfall-generating mechanisms (namely, stratiform and local convective systems as opposed to tropical cyclones) can reduce the estimation uncertainty.

5.1 Introduction

Tropical cyclones (TCs) cause loss of lives and economic damages every year (Pielke Jr. and Landsea (1998); Palmieri et al. (2006); Pielke Jr. (2007); Rappaport (2000); Rappaport (2014)). While at the global scale storms are the second most frequent natural disaster, in the United States they occupy first place in the ranking, and have costed the country 944.8 billion US$ in the period 1998-2017 (Wallemacq and House, 2018).
In addition to the well-known and most studied losses due to high wind speed and storm surges, TCs are responsible for widespread flooding (Villarini and Smith (2013); Rowe and Villarini (2013); Czajkowski et al. (2013); Villarini et al. (014a)) and flash flooding in urban areas and small basins (Hirschboeck, 1991), they trigger landslides (Bucknam et al., 2001). Czajkowski et al. (2017), for example, using insurance claims showed that the number of residential losses from TCs-caused flooding were twice as much as storm-surge losses in U.S. communities, and affected equally coastal and inland areas.
Besides the damages they cause, TCs can even have some positive side effects, for instance on groundwater recharge (Abdalla and Al-Abri, 2011) and on drought mitigation (Kam et al., 2013). It is therefore important to understand how streamflow response varies due to storms, for a more effective reservoir management and water resource allocation (Chen et al., 2015).

TCs have captured general attention especially after Hurricane Katrina (August 2005), which had devastating consequences (1,833 total fatalities and an estimated economic damage of 108 billion US$, as reported by Knabb et al. (2011)). Currently, there is no unequivocal agreement on whether climate change translated into an increased TC activity in recent years. Emanuel (2005) and Webster et al. (2005) underline the increase in tropical cyclone activity in the last decades; Pielke Jr. (2005) argued for the absence of trends in damages caused by hurricanes in the United States. Elsner et al. (2008), using satellite-derived tropical cyclone wind speeds (between 1981 and 2006), found significant trends for wind speed quantiles above the 70th percentile, which are consistent with the hypothesis that TC winds should increase with increasing ocean temperature (Emanuel, 1991). Landsea et al. (2006) were more cautious in the quantification of the anthropogenic impact on the frequency of extreme tropical cyclones, due to concerns about data quality, stating that additional efforts in the reanalysis of existing TC databases are necessary before performing reliable trend analyses. In a review on tropical cyclones and climate change and in more recent work, Knutson et al. (2010); Knutson et al. (2015) proposed projections according to which the frequency of the most intense cyclones will increase, with increases in the precipitation rate within 100 km of the storm center of the order of 20%. Moreover, damages on coastal zones due to storm surges associated with TCs are expected to increase.

With a focus on TC-induced rainfall, Prat and Nelson (2013) characterized the rainfall over land associated with TCs in several metropolitan areas around the world using TRMM precipitation data (three of the stations considered by Prat and Nelson (2013) are analyzed here: Charleston, Houston and New Orleans). For these stations, they found an average contribution of TCs to rainfall between 8.6 and 10.1% (maximum range: 18.2-25.7%).
Khouakhi et al. (2017) quantified the contribution of TCs to annual, seasonal and extreme rainfall, finding that along the U.S. East Coast the contribution of TCs to rainfall, despite being lower than in East Asia and Oceania and South East Africa, is still relevant (10-15% and up to 25% for annual and seasonal rainfall, respectively). These proportions increase when looking at extremes, reaching values higher than 30% when the variable of interest are TC-induced annual maxima. Consistently with the results by Khouakhi et al. (2017), Aryal et al. (2018) showed that rainfall annual maxima in large areas from the Gulf Coast to the U.S. North East are significantly influenced by TCs, with contributions from $\sim$20% (along the Gulf Coast) to $\sim$35% (in coastal Carolina/Virginia).
Furthermore, it is important to highlight that not only coastal areas are impacted by TC rainfall. TC remnants, indeed, can produce heavy rainfall further away from the coast (see Villarini et al. (2011); Khouakhi et al. (2017)): TC influence on rainfall can be detected in gauges located up to

400 km from the coastline.

The topic of TC rainfall contributions is of great relevance in engineering, for insurance companies and policy makers. In particular, the estimation of high return period quantiles from different sources of rainfall can provide precious support to the design of prevention and mitigation structures, as well as the basis for more accurate risk evaluation.
Several studies have been focused on the link between TCs and rainfall extremes at the regional scale: in the United States (Shepherd et al. (2007); Knight and Davis (2009); Kunkel et al. (2010)), Australia (Villarini and Denniston, 016b), on Asia (Kim et al. (2006); Ren et al. (2006)), North America (Barlow, 2011). Khouakhi et al. (2017) performed a global analysis.

Within the above framework, and leveraging the added value of incorporating the physical mechanisms into the statistical analysis of extreme events, the work presented here separates events associated with stratiform and local rainfall-generating systems from TC-induced rainfall, and proposes a mixed statistical approach to account for the different storm types. Marra et al. (2019) first explored the use of the MEVD to model extremes emerging from multiple underlying processes, focusing on the eastern Mediterranean region: two classes of synoptic systems that show different precipitation features were identified and distinguished, 1) Mediterranean lows and 2) active Red Sea troughs and other synoptic systems. Marra et al. (2019) proposed an appealing simplified MEVD approach (SMEV). SMEV ensures the sample size to be large enough for robustly fitting a distribution after the stratification in event types and reduces the effort in computing MEVD quantiles, yet with the drawback of neglecting the inter-annual variability among the distributions.

Traditionally, the EVT focuses on the analysis of annual maxima (AM) or on the POT approach (Coles (2001), see 2.1 for a detailed discussion). These methods are based on number of assumptions, as they either assume a large number of events/year or the validity of specific distributions (Generalize Pareto for exceedances over a high threshold and Poisson occurrence of events, see Section 2.1). Even more importantly, traditional EVT methods only use annual maxima or a few values over a high threshold, hence discarding most of the available information. Here, we propose the use of the MEVD (Marani and Ignaccolo (2015); Zorzetto et al. (2016)) and apply it to long series of daily rainfall in several American metropolitan areas, which have a high likelihood of being struck by a TC (see Section 5.2 for details about the case studies). Furthermore, in order to study the potential benefit of considering mixtures of distributions describing different rainfall types in the estimation of high return period quantiles, the performances of

the single-component MEVD and of its mixed version (which distinguishes between non-TC and TC-induced rainfall) are relatively compared.

The aims of the present Chapter are to understand 1) if including different distributions associated with different rainfall types (non-TC vs TC-induced) in the MEVD formulation improves the accuracy of the estimation of extremes, 2) what is the spatial variability, across the sites analyzed, of the TC influence on extreme rainfall, and 3) which is the temporal scale that better allows to identify TCs effects on extreme rainfall.
Section 5.2 describes the data analyzed and is followed by a Section summarizing the methodology adopted; Section 5.4 presents the results and a conclusions section highlights the key points.

5.2 Data

Long time series of daily rainfall from the Global Historical Climate Network (GHCN) are analyzed. The case studies are five metropolitan areas in the Continental United States (Charlotte, Charleston, Jacksonville, New Orleans and Houston) and Puerto Rico (Coloso station), which have a high probability to be impacted by a TC. Except for Charleston, the stations do not have a continuous record, due to missing or incomplete years (a year is considered as complete when it includes at least 330 values). To avoid gaps in the series and to extend the available records, when possible, they are complemented using nearby stations. The maximum distance between the "main" station and the complementary one is about 15 km. See Table 5.1 and Figure 5.1 about location and time coverage details for the selected stations.

To avoid mixing stations with significantly different rainfall observations, a two-sample Kolmogorov-Smirnov test at a significance level α=0.05 was applied to the records of the overlapping years between the main station and the one selected as candidate to complement the record. This check is important because, even if the stations considered are close enough, they might be located at different altitudes –as it happens in Puerto Rico- or they might be influenced by local phenomena to varying degrees (e.g., one on the coast with respect to one more inland). If the empirical distributions of rainfall are detected to be statistically different, the secondary station is discarded. Furthermore, in order to verify that a TC affecting one area was influencing rainfall to some extent the same day also 15 km farther, values of TC-induced rainfall occurred in the same day during the years shared by the "main" and the "complementary" station(s) were plotted one against the other (Figure 5.2). There is not a perfect match between these values, as may be expected. However, points are grouped around the 45 degrees line, suggesting a homogeneous behaviour across stations; moreover, when

Table 5.1: Name and identification number of the stations analyzed. Names in bold refer to the station with the longer record, while the others are the stations used to complement the main ones.

STATION	IDENTIFICATION NUMBER
Downtown Charleston, SC	USW00013782
Charlotte Douglas Airport, NC	USW00013881
Charlotte, NC	USC00311695
Houston William P Hobby Airport, TX	USW00012918
Houston Weather Bureau City, TX	USW00012945
Jacksonville, FL	USW00093852
Jacksonville International Airport, FL	USW00013889
New Orleans Audubon, LA	USW00012930
New Orleans Carrollton, LA	USC00166676
New Orleans Airport, LA	USC00166661
Coloso, PR	RQC00662801

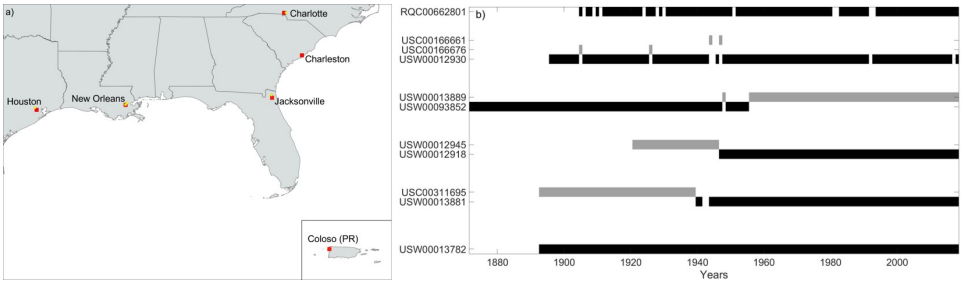


Figure 5.1: Panel a): Map showing the location of the stations analyzed, divided in "main" (red squares) and "secondary" (yellow dots). Puerto Rico is represented closer to the CONUS in order to reduce empty spaces in the map. Panel b): Years of record for the analyzed stations. Black bars refer to the main station, while gray bars to the secondary ones. See Table 1 for matching identification numbers and station names. The longest continuous record, for Charleston station, spans the period from 1893 to 2018.

the "main" station recorded a value of TC-induced rainfall, the same occurred also for the "secondary" station on the same day.
After these tests, no secondary station to fill the gaps in the record or to extend it was found for Puerto Rico, as may be seen in Figure 5.1b.

The trajectories of tropical storms are derived from the HURDAT2 database (Figures 5.3, 5.4), the revised version of the National Hurricane Center's (NHC) North Atlantic basin hurricane database, also known as "best track" database (Landsea and Franklin, 2013). Data go back to 1851, but observations before the 1970's, the pre-satellite era, are argued to display significant uncertainties, due to storm undercount (Vecchi and Knutson (2011); Landsea et al. (2012); Hagen et al. (2012)). For this reason analyses were performed both with and without the first part of the database (before 1970).

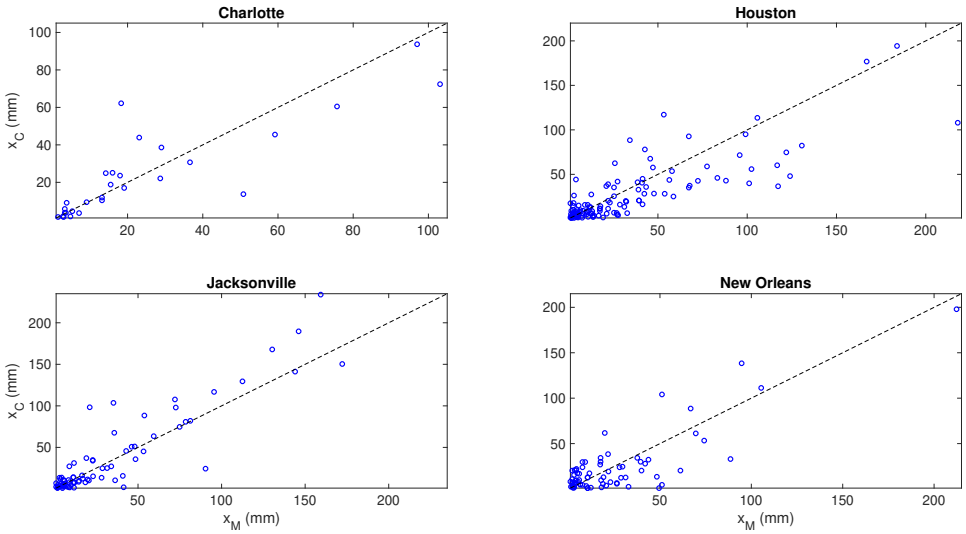
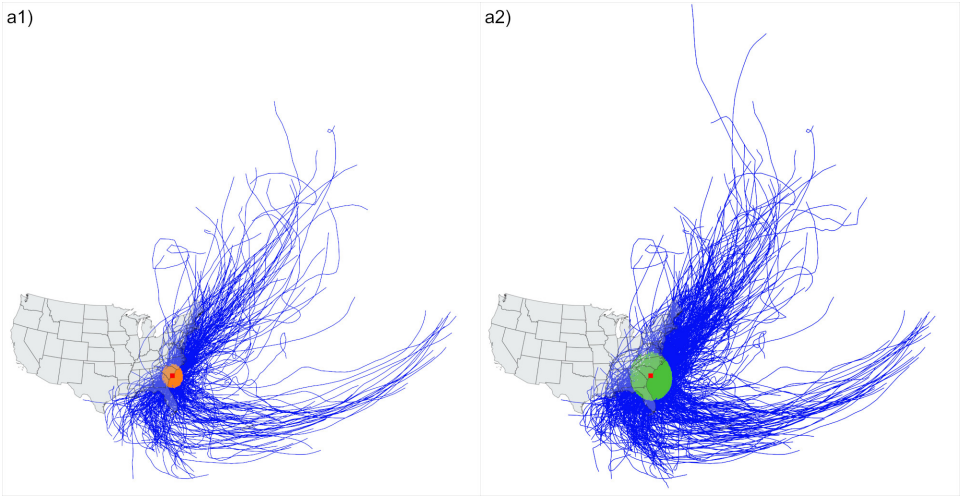
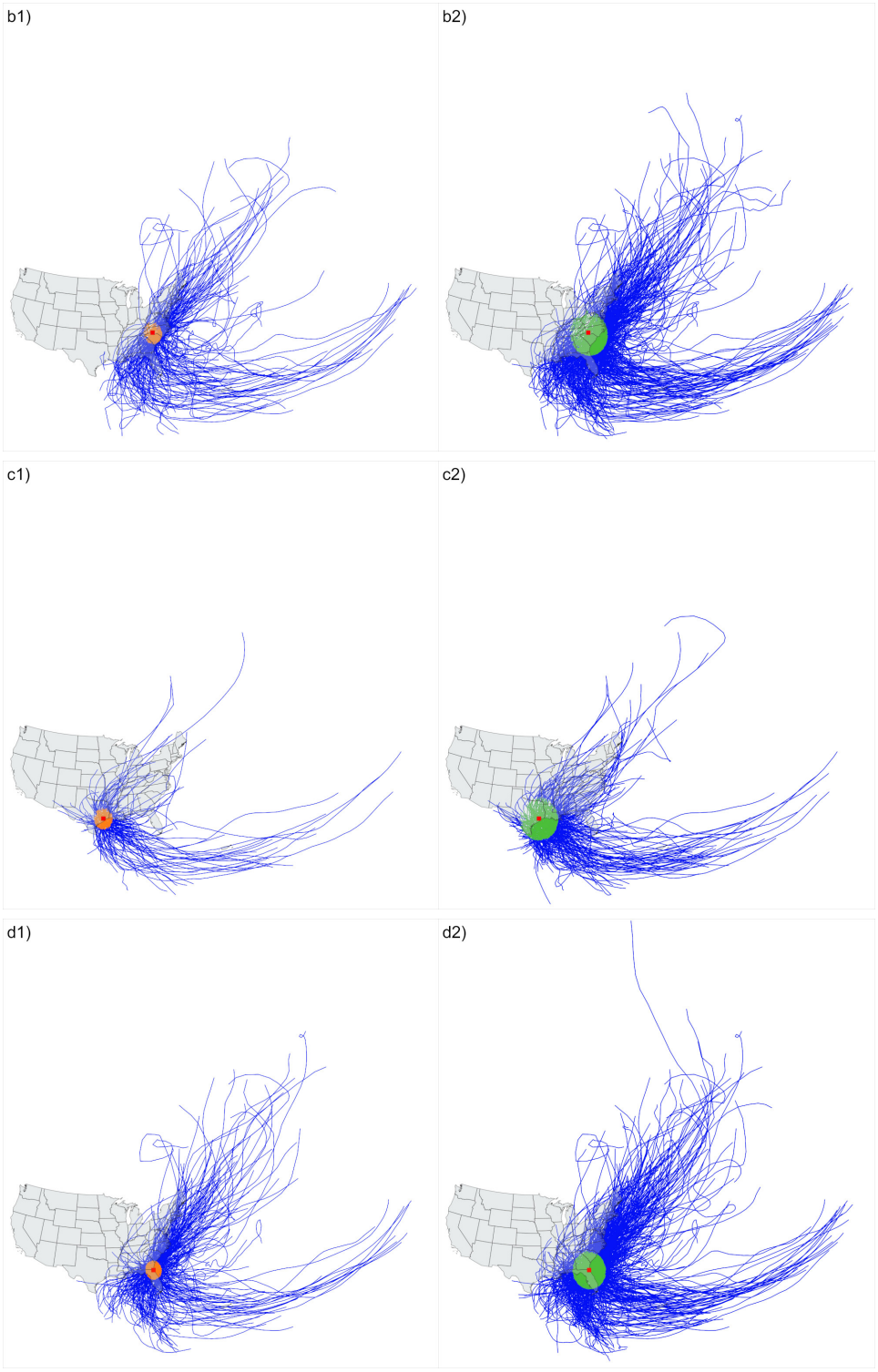


Figure 5.2: Scatter plot of TC-induced rainfall values recorded on the same day in the main station (x_M on the x-axis) and the secondary one(s) (x_C on the y-axis). Magenta dots (which are not present in this case) would represent values of rainfall that have been identified as TC-induced in one stations but not in the other.

While acknowledging that uncertainties characterizing the first part of the dataset might somewhat affect the estimation of high quantiles, using also the data from 1851-1969 can provide useful information on values corresponding to return periods much higher than the calibration sample size.




b1)
b2)
c1)
c2)
d1)
d2)

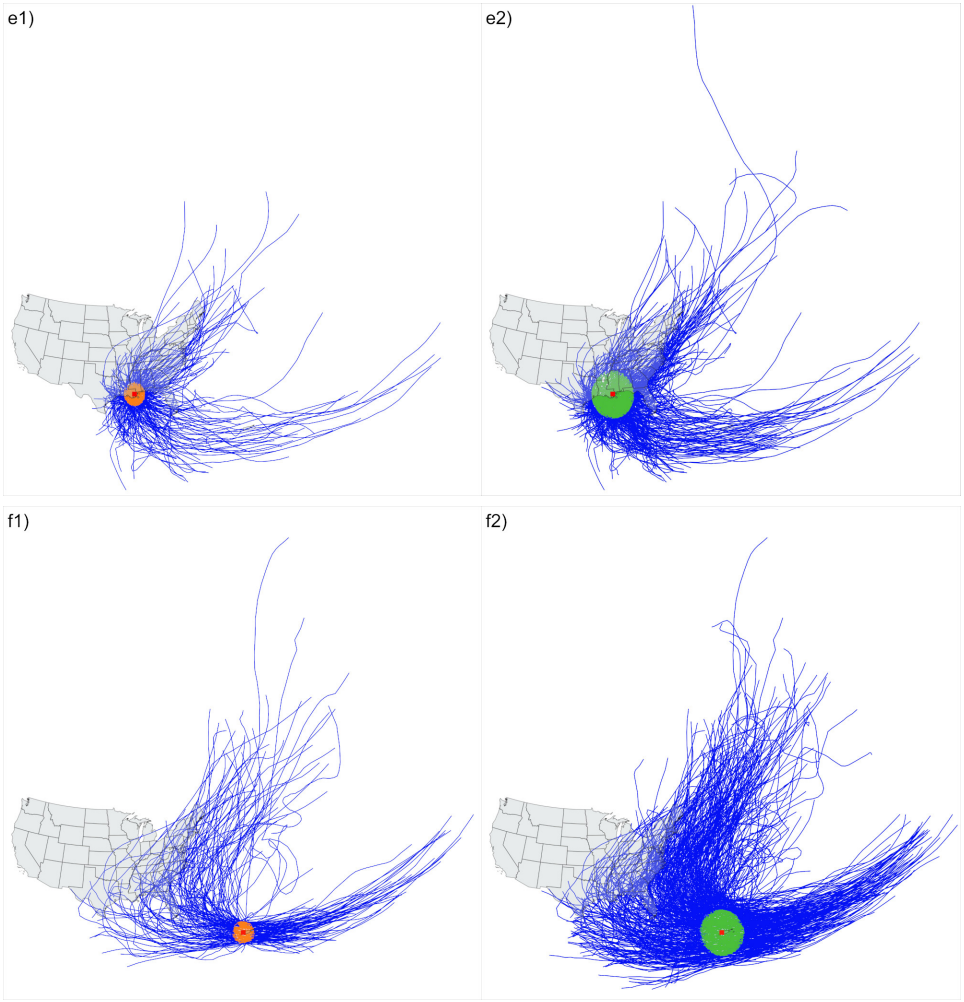


Figure 5.3: Trajectories of TCs influencing rainfall values in each station (blue lines). The trajectories within a buffer of 250 (500) km are represented by orange (green) lines and numbers 1 (2) in the panel name. Letters in the panel name indicate the different stations: a) Charleston, b) Charlotte, c) Houston, d) Jacksonville, e) New Orleans, f) Coloso (PR). The whole record available is considered here.

5.3 Methodology

In order to evaluate the possible advantage of including different rainfall generating mechanisms, the MEVD is applied here to daily rainfall records both in its single-component formulation (i.e., a MEVD that treats all the rainfall values as belonging to the same population, as in Eq.(2.6)) and in its mixed version (i.e., considering different distributions for non-TC rainfall and the TC-induced one).

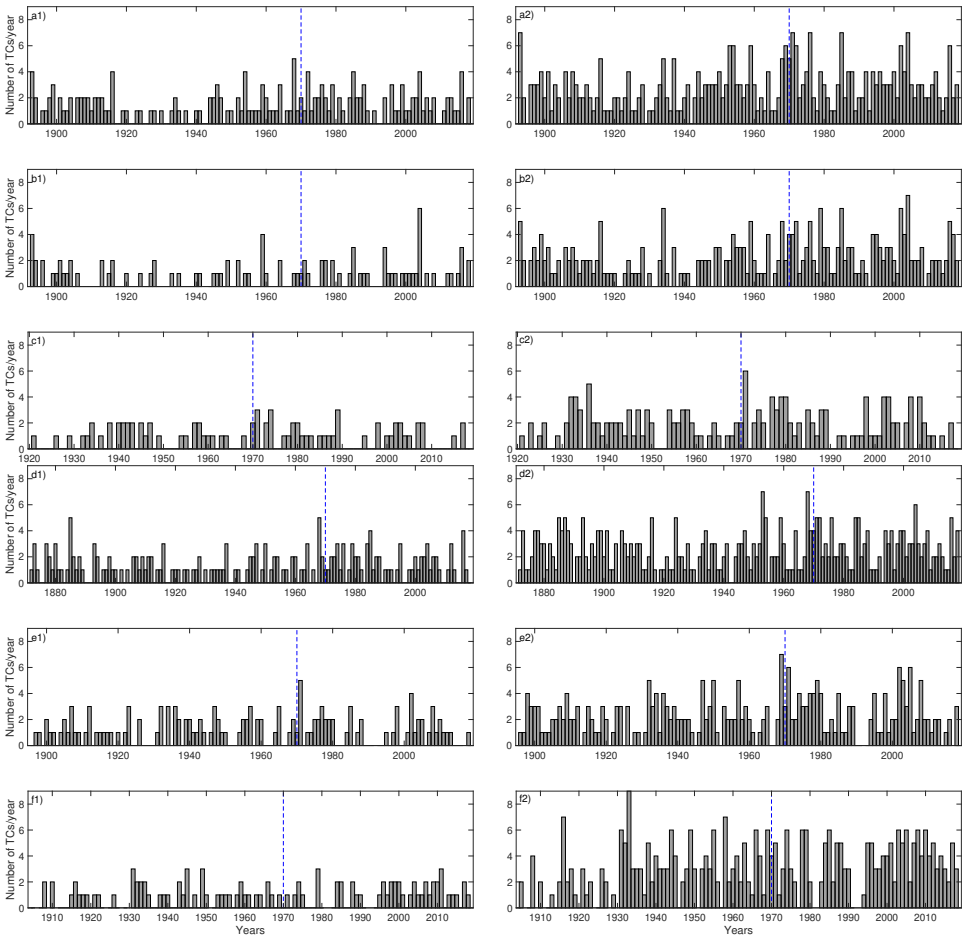


Figure 5.4: Number of Tropical Cyclones per year for the analyzed stations. The left column refers to TCs that enter the buffer of 250 km, while the right one to those entering the buffer of 500 km (number 1 and 2 in the panel name, respectively). Letters in the panel names indicate the stations: a) Charleston, b) Charlotte, c) Houston, d) Jacksonville, e) New Orleans and f) Coloso (PR). The blue dashed line indicates the beginning of the satellite era (1970).

The mixed approach is similar to the SMEV approach proposed by Marra et al. (2019) for rainfall from two synoptic systems, since they both include different distributions underlying the events. However, the approach proposed here differs from the SMEV formulation in the definition of the "ordinary distributions": the SMEV expression does not consider time-varying distributions (no dependence on j). In Chapter 3 the relevance of accounting for the inter-annual variability in the statistical properties of rainfall was emphasized by means of analyses on both historical and synthetic records of daily rainfall data. Those analyses indicate that it is useful to preserve the variability across the years when a sufficient number of events/year is available to estimate parameter values sufficiently accurately. Hence, parameter estimation is performed here on a yearly basis for non-TC rainfall, whereas only one distribution (resulting in one single pair of parameters) is fitted to the values of TC-induced rainfall. The use of the entire calibration sample in the second case becomes necessary since the number of rainfall events caused by TCs in any one year is not sufficient to ensure a large-enough sample on which to fit a distribution without introducing large uncertainties.

Leveraging the intrinsic ability of the MEVD to naturally account for different phenomena (in this case and TC and non-TC rainfall), a mixed version of the MEVD is applied as follows:

$$\zeta(x) = \frac{1}{M}\sum_{j=1}^{M}[F_{nTC}(x;\vec{\theta_j})]^{n_j^{nTC}} \cdot [F_{TC}(x;\vec{\theta})]^{n_j^{TC}} \tag{5.1}$$

where $F_{nTC}(x;\vec{\theta_j})$ is the distribution fitted to the non-TC ordinary events in year j, while $F_{TC}(x;\vec{\theta})$ is the distribution fitted on all the TC rainfall values in the calibration sample (and therefore independent on j); n_j^{nTC} (n_j^{TC}) is the number of non-TC (TC) events in any-one year.

Following the reasoning by Wilson and Toumi (2005); Marani and Ignaccolo (2015), the Weibull distribution is used to describe the ordinary values (daily rainfall values>1 mm).
Parameter estimation is performed by means of Probability Weighted Moments (PWM), since they are less sensitive to outliers and perform better for small samples if compared to Maximum Likelihood or the conventional method of moments (Greenwood et al., 1979).

5.3.1 Tropical Cyclone-Induced Rainfall

For the application of the mixed-MEVD, first the definition of TC-induced rainfall is needed. A rainfall value is attributed to a tropical storm if the value is measured in the time window ±1 day during which a storm trajectory enters a 250 or 500 km radius buffer around the rain gauge (see Figure

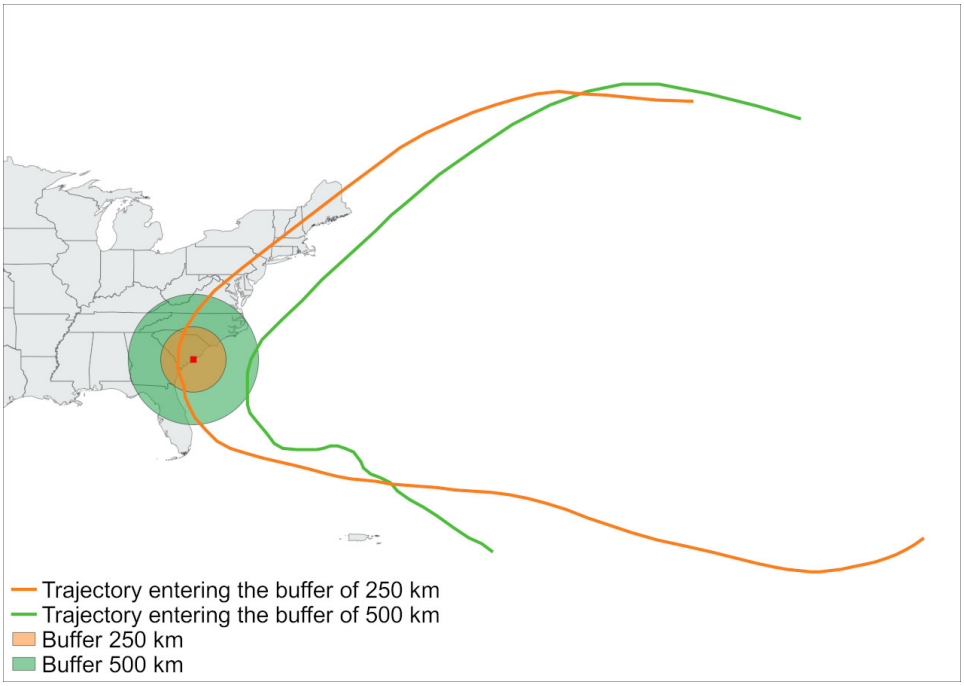


Figure 5.5: Example of two trajectories entering a buffer of 250 (orange) and 500 (green) km. If the station (red square) records a rainfall value in the time window (± 1day) in which the trajectory is within the buffer, that value is attributed to a TC.

5.5 for an example of TC influencing-rainfall trajectory identification). The choice of the buffer size is consistent with other studies in the literature (for example Villarini et al. (2014) indicate rainfall to be influenced by a hurricane up to a distance of 5^o, ∼500 km). Considering a time window widened to the previous day and the one after allows to capture both possible predecessor rainfall events (Galarneau et al., 2010) and storm remnants.
Consecutive wet days during a TC are highly correlated, which violates the independence hypothesis at the basis of the traditional EVT and of the MEVD. In order to only select independent events, the maximum daily rainfall depth associated with each TC is considered here.

Due to the extended influence that TCs have over time (if compared for example to their extratropical counterparts, Wallace and Hobbs (2006)), rainfall events with duration over multiple days are also studied. Hence, the cumulative rainfall in non-overlapping windows of different lengths, from 2 to 5 days, is also analyzed. In this case, the aggregated values of precipitation are considered as being TC-induced if at least one of the values in the aggregation time window has been influenced by a TC. The threshold for daily rainfall is consequently modified when considering rainfall over multiple days (thr_{agg}) and the selection criteria are as follows: $thr_{agg} = n^o aggr\ days \cdot 1mm$, where $n^o aggr\ days$ is the number of days on which rainfall is aggregated and

1 mm is the threshold set for daily rainfall.

5.3.2 Evaluation of estimation accuracy

A Monte Carlo experiment involving $N_r = 1000$ realizations was performed for each station and the uncertainty of MEVD-estimated quantiles (according to both the single and two-component versions) was quantified using a cross-validation procedure. In each realization j:

1. the observational record is randomly reshuffled (keeping all observations in their original year to preserve both their yearly frequency distributions and the distribution of the number of events/year) to generate a realization displaying no systematic variability;

2. the reshuffled record is divided into two independent sub-samples, randomly selecting S years for parameters calibration and using the remaining $L = L_{tot}–S$ years (where L_{tot} represents the length of the series) for testing the accuracy in the estimation of maxima;

3. the relative error ε_j is computed between the estimated maxima and the observed ones for each calibration sample size S and return period T as $\varepsilon_j(S,T) = (x_{est,j}(S,T) - x_{obs,j}(S,T))/x_{obs,j}(S,T)$;

4. the above process is repeated N_r=1000 times to obtain a full statistical description of error metrics. This whole process is repeated for different calibration sample sizes (S= 10, 20 and 30 years).

5. The Fractional Standard Error is computed over the 1000 realizations as:

$$FSE(S,T) = \frac{1}{N_r}\sum_{j=1}^{N_r}\varepsilon_j(S,T)^2]^{1/2} \tag{5.2}$$

5.4 Results

Berg et al. (2013), in a study on precipitation over Germany, showed that while stratiform precipitation has a power-law behavior, the curve of convective precipitation in a log-log plot is concave, that is, steeper at higher intensities. Here, the distinction is tested between non-TC and TC-induced rainfall; being the latter a sub-sample of convective rainfall, a similar behavior could be expected. As it can be noted in Figure 5.6, the distribution of TC-induced rainfall forms a more or less abrupt "elbow" bend for high values of rainfall depths.
A two-sample Kolmogorov-Smirnov test at the 0.05 significance level was applied in order to detect the statistical difference between the non-TC and the TC-rainfall; the empirical distributions resulted statistically different in

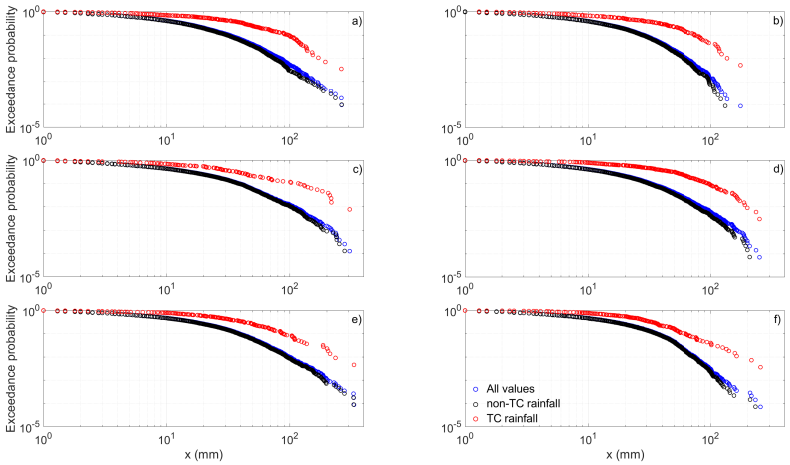


Figure 5.6: Log-log plot of the exceedance probability of ordinary event magnitude for the six stations analyzed: a) Charleston, b) Charlotte, c) Houston, d) Jacksonville, e) New Orleans and f) Coloso, PR. Blue dots indicate all the values of daily rainfall, irrespective of their generating mechanisms, black dots are the values not influenced by a TC and red dots represent TC-induced rainfall values.

all the analyzed cases.

After the detection of the statistical difference between the distributions of non-TC and TC-induced rainfall, the suitability of the Weibull distribution to describe rainfall values both in the single-type and the multiple-type approach was tested (see Quantile-Quantile plots in Figures 5.7,. . ., 5.12). The choice of the aggregation window of 3 days followed the evaluation of the frequency of the number of consecutive days affected by TCs in the analyzed stations (see Figure 5.13). In most of them, there is an important contribution of TCs on rainfall up to 2-3 consecutive days, therefore the use of 3-days non-overlapping windows allows capturing most of the impact of the TCs on rainfall.
In terms of the advantage introduced by explicitly considering different rainfall types, three cases are found, that may be illustrated by the Charleston, Houston, and Coloso cases. For the Charleston station (Figure 5.7), there is a slightly improvement in the accuracy of the estimations of TC-induced rainfall depths when moving from daily to 3-day cumulative rainfall, but values are well estimated also at the daily scale. In Houston (Figure 5.9), Hurricane Harvey caused more than 800 mm of rainfall in three days, which is more than two times higher than the next highest 3-day cumulative rainfall (estimated T∼2000 years, Emanuel (2017)). The fitted Weibull distribution tends to underestimate high return periods in the station of Coloso (Figure 5.12); Weibull estimates are less affected by this issue when considering 3-day cumulative rainfall.

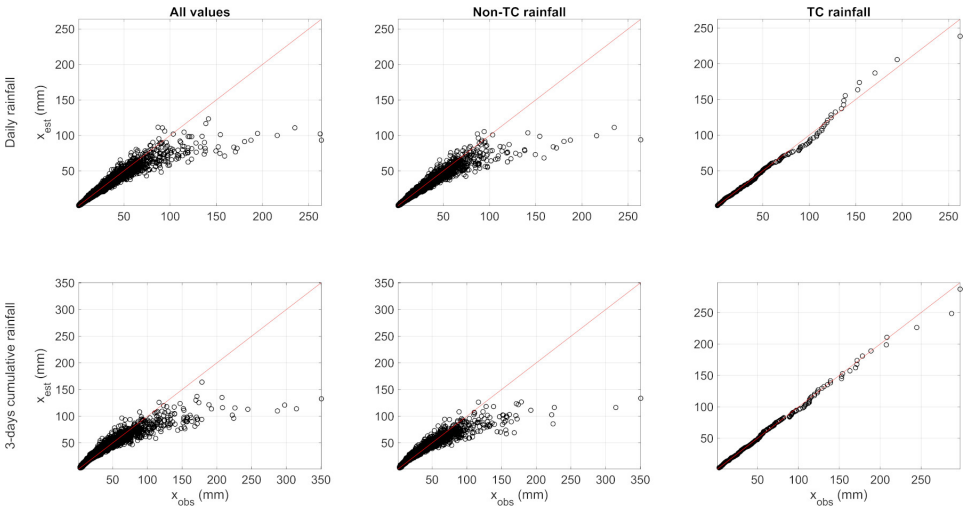


Figure 5.7: Quantile-Quantile (QQ) plots for Charleston station. The two rows represent daily and 3-day cumulative rainfall respectively. The three columns refer to: all the rainfall values, irrespective the generating mechanism (left), non-TC rainfall (center), and TC-induced rainfall (right). The data refer to the entire record available (1893-2018), and to a 500 km buffer. All-type rainfall values and non-TC rainfall values are fitted using yearly Weibull distributions, while TC-induced rainfall values are fitted with a single probability distribution.

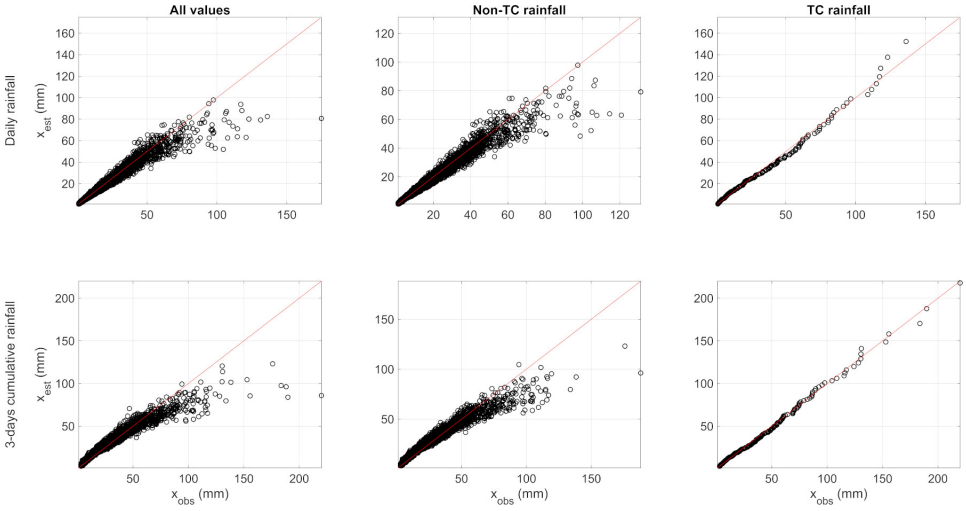


Figure 5.8: Quantile-Quantile (QQ) plots for Charlotte stations. The two rows represent daily and 3-day cumulative rainfall respectively. The three columns refer to: all the rainfall values, irrespective the generating mechanism (left), non-TC rainfall (center), and TC-induced rainfall (right). The data refer to the entire record available (1893-2018), and to a 500 km buffer. All-type rainfall values and non-TC rainfall values are fitted using yearly Weibull distributions, while TC-induced rainfall values are fitted with a single probability distribution.

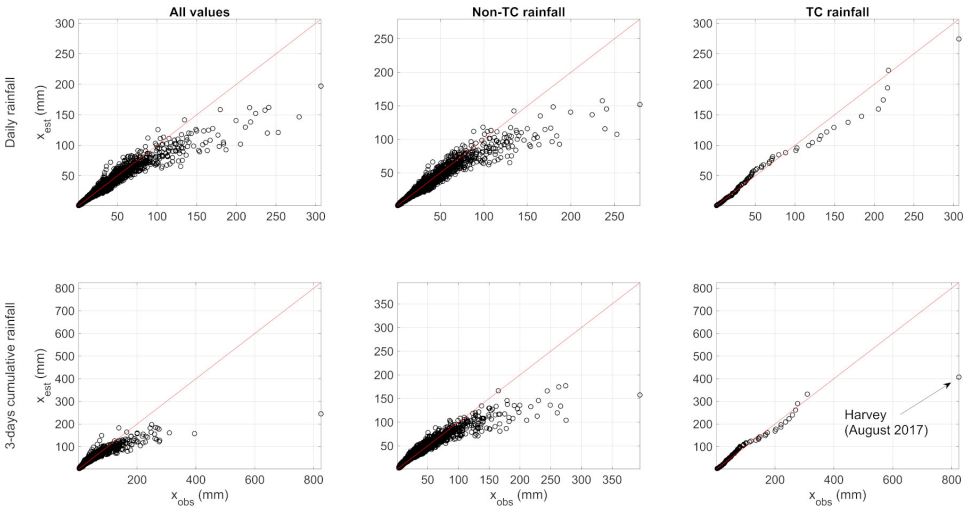


Figure 5.9: Quantile-Quantile (QQ) plots for Houston stations. The two rows represent daily and 3-day cumulative rainfall respectively. The three columns refer to: all the rainfall values, irrespective the generating mechanism (left), non-TC rainfall (center), and TC-induced rainfall (right). The data refer to the entire record available (1921-2018), and to a 500 km buffer. All-type rainfall values and non-TC rainfall values are fitted using yearly Weibull distributions, while TC-induced rainfall values are fitted with a single probability distribution.

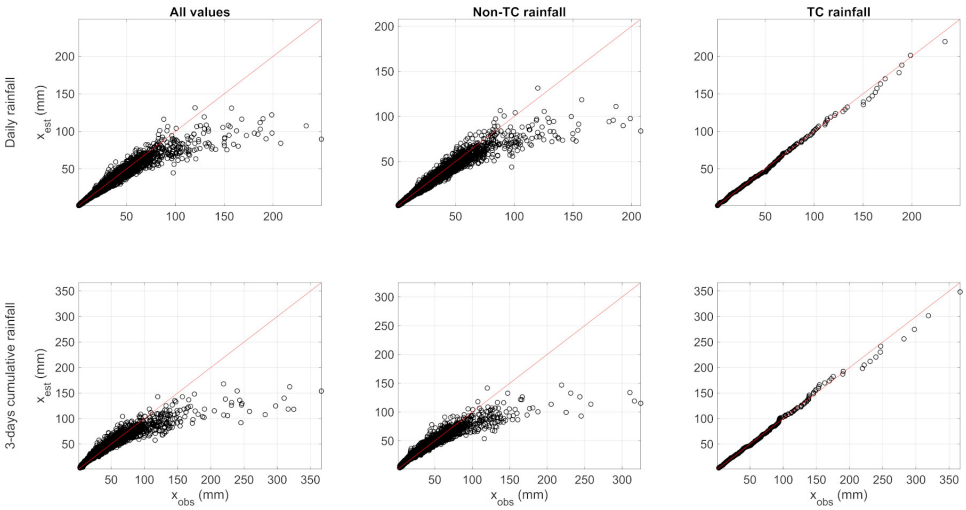


Figure 5.10: Quantile-Quantile (QQ) plots for Jacksonville stations. The two rows represent daily and 3-day cumulative rainfall respectively. The three columns refer to: all the rainfall values, irrespective the generating mechanism (left), non-TC rainfall (center), and TC-induced rainfall (right). The data refer to the entire record available (1872-2018), and to a 500 km buffer. All-type rainfall values and non-TC rainfall values are fitted using yearly Weibull distributions, while TC-induced rainfall values are fitted with a single probability distribution.

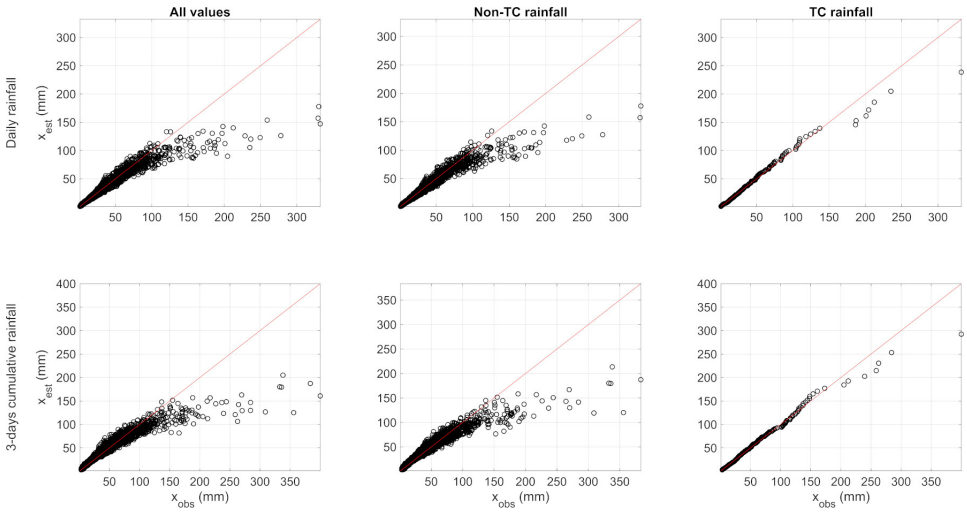


Figure 5.11: Quantile-Quantile (QQ) plots for New Orleans stations. The two rows represent daily and 3-day cumulative rainfall respectively. The three columns refer to: all the rainfall values, irrespective the generating mechanism (left), non-TC rainfall (center), and TC-induced rainfall (right). The data refer to the entire record available (1896-2018), and to a 500 km buffer. All-type rainfall values and non-TC rainfall values are fitted using yearly Weibull distributions, while TC-induced rainfall values are fitted with a single probability distribution.

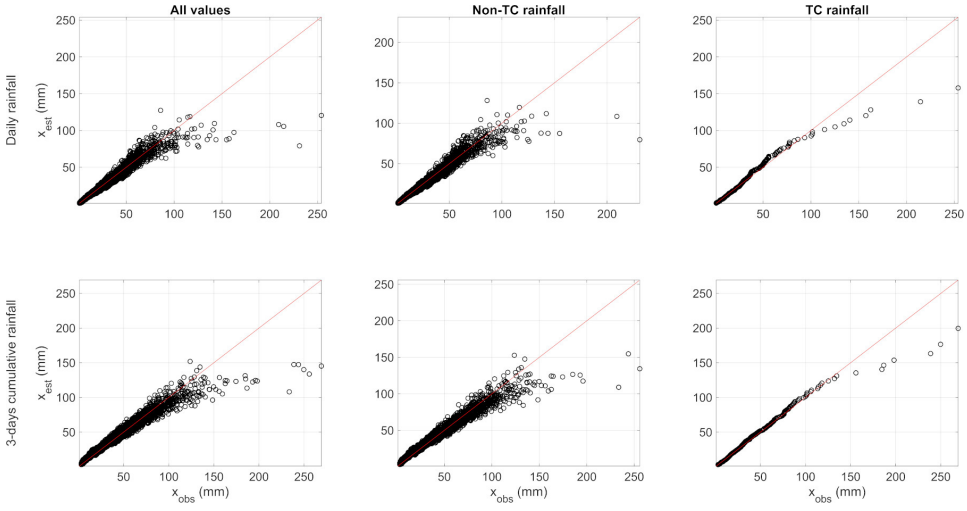


Figure 5.12: Quantile-Quantile (QQ) plots for Coloso (PR) station. The two rows represent daily and 3-day cumulative rainfall respectively. The three columns refer to: all the rainfall values, irrespective the generating mechanism (left), non-TC rainfall (center), and TC-induced rainfall (right). The data refer to the entire record available (1905-2018), and to a 500 km buffer. All-type rainfall values and non-TC rainfall values are fitted using yearly Weibull distributions, while TC-induced rainfall values are fitted with a single probability distribution..

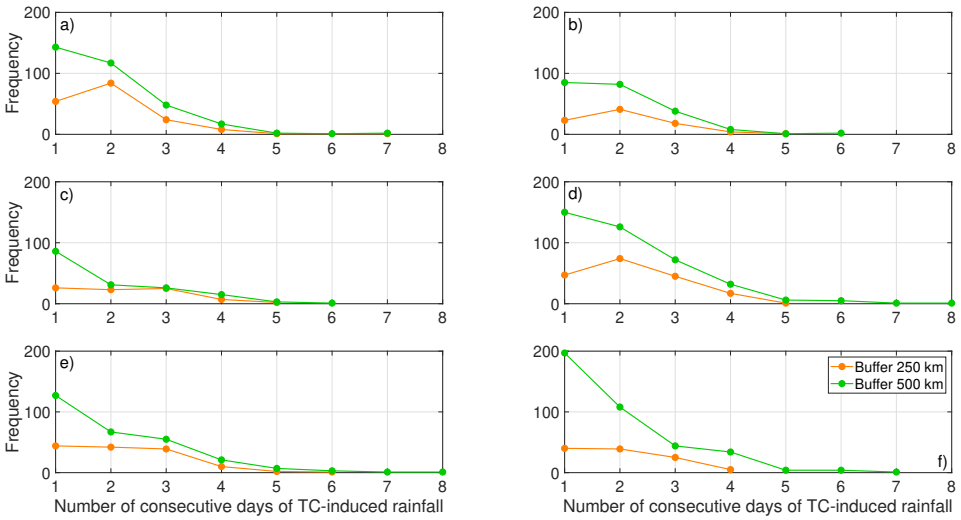


Figure 5.13: Frequency of the number of consecutive days of TC-induced rainfall for the 250 km buffer (orange) and the 500 km buffer (green) for the six stations analyzed: a) Charleston, b) Charlotte, c) Houston, d) Jacksonville, e) New Orleans, f) Coloso (PR).

After verifying the existence of a significant statistical difference between the distributions of non-TC and TC-induced rainfall values and the suitability of the Weibull distribution as a model to describe ordinary events, the potential advantage of using a mixed approach instead of the single-distribution MEVD was evaluated. Results are presented in terms of the ratio between the the value of the Fractional Standard Error for the mixed-MEVD and the one for the single-component MEVD as a function of the calibration sample size (S = 10, 20 and 30 years). This analysis allows to immediately appreciate which approach provides the most accurate estimations and, simultaneously, the effect of the sample size on the estimation error.

Through a qualitative assessment of the density of the TCs trajectories passing near the stations selected (Figure 5.3) and of the number of TCs entering the buffers every year (Figure 5.4), the estimation of high-return period quantiles at the Charleston, Jacksonville, New Orleans and Puerto Rico stations would be expected to benefit from the use of mixtures of distributions. Despite the results by Prat and Nelson (2013), who indicate a significant TC contribution on rainfall in the Houston area, the density of storms trajectories entering the buffers around the city does not seem to imply a significant influence of TCs on extreme rainfall. This conclusion is supported by the histograms in Figure 5.4 (panels c1 and c2), especially for the 250 km buffer.
Moreover, values show that Houston experiences very heavy non-TC rainfall. One explanation may be tornadoes generated by convective cells at the mesoscale, mainly in spring. Houston is indeed located in the so called

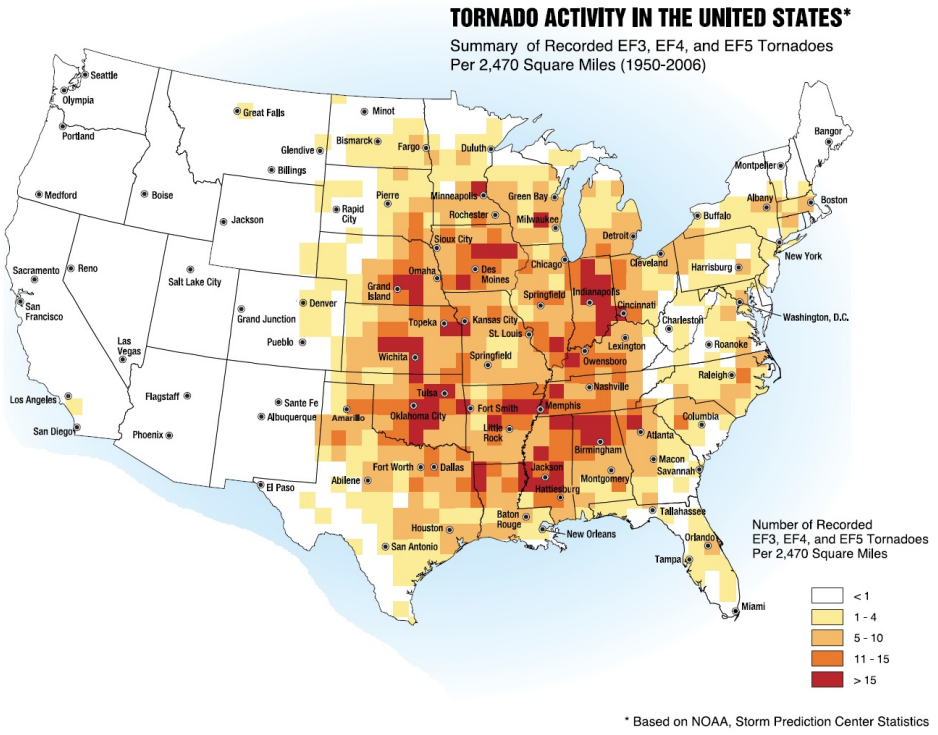


Figure 5.14: Number of tornadoes recorded per 2,470 mi^2. Modified from FEMA (2008)

"Tornado Alley", a term indicating U.S. areas that are often affected by tornadoes. According to the map that shows the tornado activity in the United States (page 3 in FEMA (2008), reported in Figure 5.14) based on NOAA Storm Prediction Center Statistics, Houston belongs to the third class (the highest among the stations analyzed here), where tornadoes are moderately frequent. For this class, the range is from 5 to 10 recorded EF3, EF4, and EF5 tornadoes per 2,470 square miles (EF stays for Enhanced Fujita Scale according to which tornadoes severity is categorized).
A comparison with the monthly reports available at `https://www.ncdc.noaa.gov/sotc/tornadoes`, which report occurrences dating back, at most, to 2006, shows that some of the highest rainfall values, if not associated to TCs, occurred in months in which tornadoes were reported in the Houston area. Therefore, even if a precise match with tornado occurrence cannot be tested due to the coarse monthly resolution of tornado records, most likely these heavy rainfall values were generated by these convective systems.

Considering the potential advantage of distinguishing separate distributions for non-TC and TC-induced daily rainfall, it can easily be seen that using mixtures of distributions is not beneficial for all the stations analyzed (Fig-

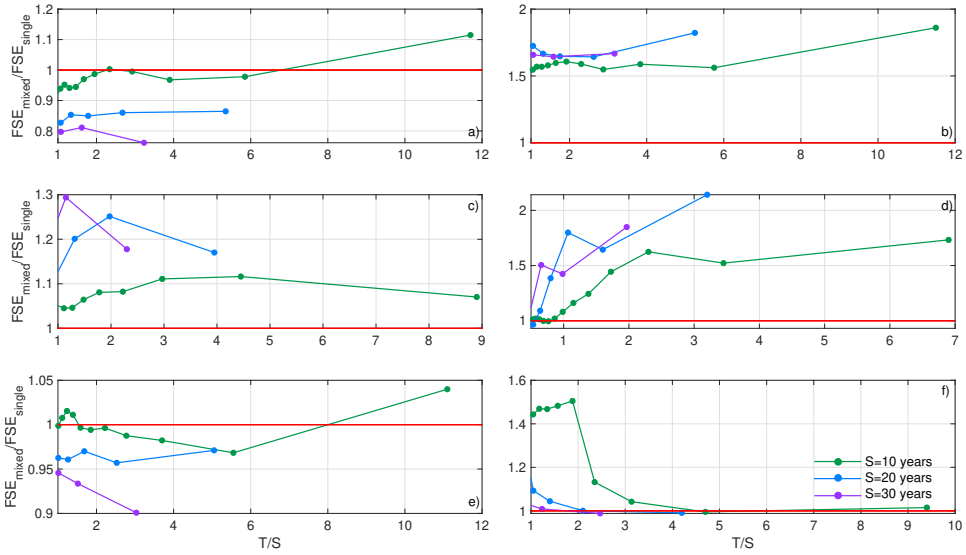


Figure 5.15: Ratio between the FSE computed with a mixed MEVD and the FSE obtained with the single-component MEVD, as a function of the ratio return period (T) to calibration sample size (S) for the stations analyzed. The whole hurricane record and a 500 km buffer is used. Colors refer to the different sample sizes: $S = 10$ (green), 20 (light-blue) and 30 (purple) years. Each panel shows one station: a) Charleston, b) Charlotte, c) Houston, d) Miami, e) New Orleans and f) Coloso (Puerto Rico). The red line sets the threshold between the performances of the two MEVD approaches: if dots are located below this line, the use of mixtures in the MEVD is beneficial, while if they are above the line, a single-component MEVD is providing more accurate estimations.

ures 5.15 and 5.16, for the whole record and the one from 1970, considering a 500 km buffer). With reference to Prat and Nelson (2013), distinguishing two types of distributions does not result in more accurate estimations for the Houston station, hence supporting the first hypothesis linked to the low frequency of TCs in this area.

For the Charleston station, characterized by the most uniform frequency of TCs over all the period of record, for both the record lengths a mixed MEVD outperforms the single-component one; not surprisingly, given the number and the uniformity over time of TCs entering the buffers around this station, increasing the calibration sample size from 10 to 20 and 30 years is still confirming the trend and the advantage of the mixed approach becomes higher. When TC-induced rainfall events are present throughout the whole record, indeed, using more years to estimate the parameters of the different (non-TC and TC rainfall) distributions includes more values in both the rainfall types, resulting in a higher accuracy in the parameters estimation for the distribution of TC rainfall.

Regarding Coloso station (PR), for which a strong signal due to the high density of the TCs trajectories would have been expected, including mixtures of distributions in the MEVD approach is not reducing the uncertainty

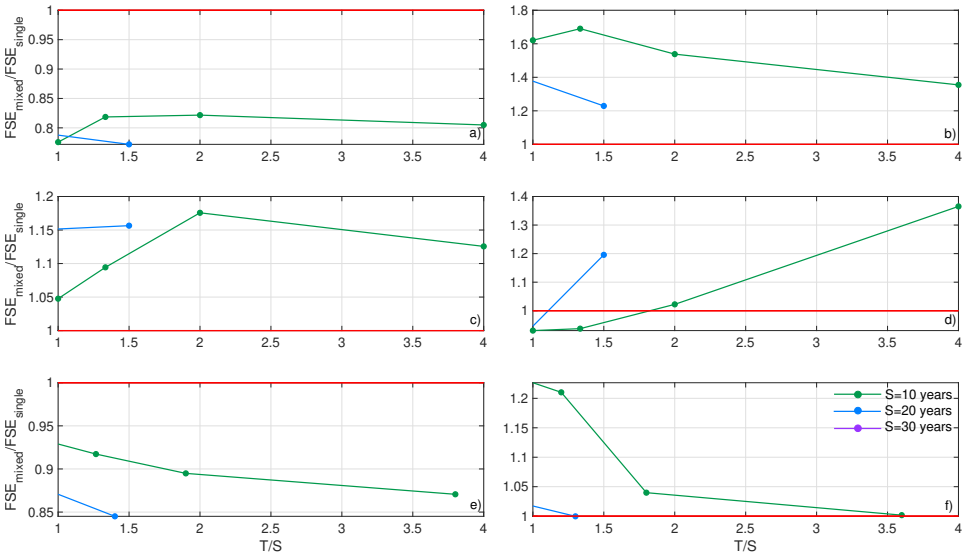


Figure 5.16: Ratio between the FSE computed with a mixed MEVD and the FSE obtained with the single-component MEVD, as a function of the ratio return period (T) to calibration sample size (s) for the stations analyzed. The hurricane record since 1970 and a 500 km buffer are used. Colors refer to the different sample sizes: $S = 10$ (green), 20 (light-blue) and 30 (purple) years. Each panel shows one station: a) Charleston, b) Charlotte, c) Houston, d) Miami, e) New Orleans and f) Coloso (Puerto Rico). The red line sets the threshold between the performances of the two MEVD approaches: if dots are located below this line, the use of mixtures in the MEVD is beneficial, while if they are above the line, a single-component MEVD is providing more accurate estimations. The purple line (S=30 years) is not present as the record is not long enough to allow an evaluation for T/S values greater than 1.

in the estimations (the ratio between FSE computed using the mixed approach and the FSE resulting from the application of the single-component MEVD is higher than one). When looking at high return period quantiles, the mixed and the single-component approaches provide the same uncertainty in the estimation. Also in this case, the calibration sample sizes of 20 and 30 years reduce the uncertainty in the estimations.

Following the reasoning about the impact of TCs on consecutive days and supported by the plots in Figure 5.13, the case of 3-days cumulative rainfall was also evaluated. In this case, a mixed MEVD is always outperforming the single-component one when considering aggregated values of rainfall over time windows longer than one day, especially for high T/S, with the exception of the Houston station. This would suggest that the effect of high intensity rainfall due to TCs is more evident when cumulative rainfall over time windows longer than one day. Figure 5.17 shows the results for the aggregated values of rainfall over a time window of 3 days: the FSE reduction associated with the inclusion of multiple distributions in the MEVD is generally higher than in the case of daily rainfall.
The Houston area, being impacted by a smaller number of TCs/year and possibly because it is affected also by tornadoes (that yet cause heavy rainfall for a shorter time, usually one day) and due to the heavy rainfall caused by Hurricane Harvey in August 2017 (more than 800 mm in three days) is not showing any advantage for the use of mixtures of distributions.

5.5 Discussion

This Chapter evaluates the potential improvement in the estimation of extremes including mixtures of distributions used to discriminate between non-Tropical Cyclones and Tropical Cyclones-induced rainfall. Understanding and differentiating rainfall-generating mechanisms is indeed relevant for a conceptual reasoning and it can be useful to more accurately estimate extremes. For this reason, the MEVD was applied to long series of daily rainfall in six U.S. metropolitan areas, considering non-TC and TC-induced rainfall both as coming from the same distribution and belonging to different ones.

On the basis of the non-parametric two-sample Kolmogorov-Smirnov test, the distributions of non TC and TC-induced ordinary rainfall events were found to be always statistically separable. A behavior similar to the one shown by Berg et al. (2013) in a study on stratiform and convective rainfall was highlighted here, as the exceedance probability of TC-induced rainfall generally presents a more pronounced concavity in a log-log plot.
When moving from the ordinary event distribution to the estimation of extreme values, it was shown that including mixtures of distributions describing non TC and TC-induced rainfall is reducing the estimation uncertainty

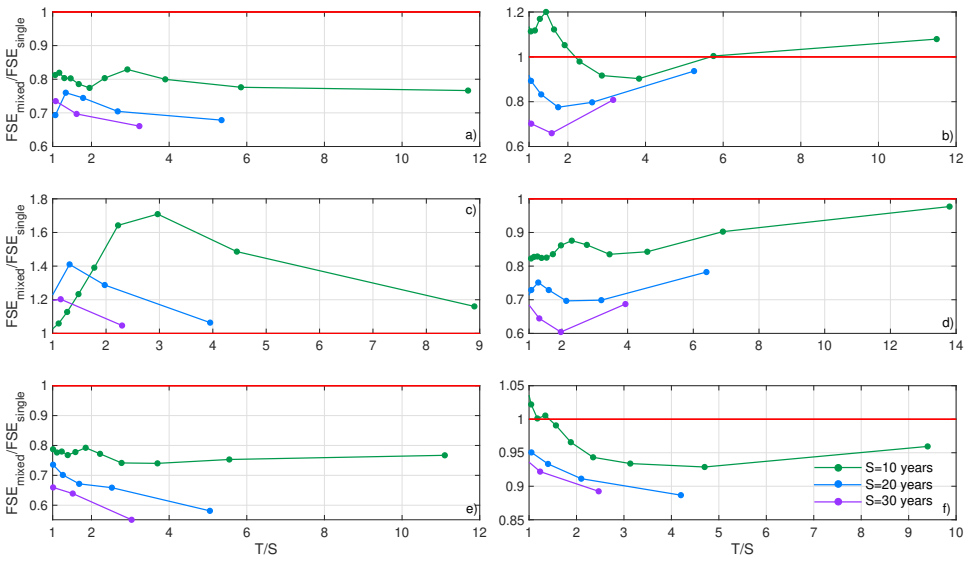


Figure 5.17: Ratio between the FSE computed with a mixed MEVD and the FSE obtained with the single-component MEVD, as a function of the ratio return period (T) over calibration sample size (S) for the stations analyzed, considering 3-day cumulative rainfall on the whole record and the 500 km buffer. The colors of the dots refer to the different sample sizes: $S = 10$ (green), 20 (light-blue) and 30 (purple) years. Each panel shows one station: a) Charleston, b) Charlotte, c) Houston, d) Miami, e) New Orleans and f) Coloso (Puerto Rico). The red line sets the threshold between the performances of the two MEVD approaches: if dots are located below this line, the use of mixtures in the MEVD is beneficial, while if they are above the line, a single-component MEVD is providing more accurate estimations

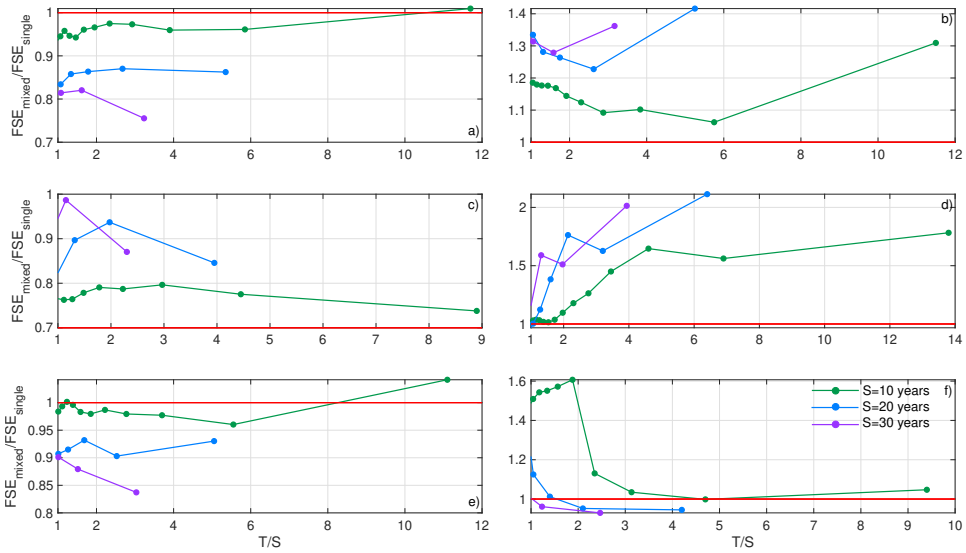


Figure 5.18: Ratio between the FSE computed with a mixed MEVD and the FSE obtained with the single-component MEVD, as a function of the ratio return period (T) over calibration sample size (s) for the stations analyzed, considering daily rainfall on the whole record and the buffer of 250 km. The colors of the dots refer to the different sample sizes: S = 10 (green), 20 (light-blue) and 30 (purple) years. Each panel shows one station: a) Charleston, b) Charlotte, c) Houston, d) Miami, e) New Orleans and f) Coloso (Puerto Rico). The red line sets the threshold between the performances of the two MEVD approaches: if dots are located below this line, the use of mixtures in the MEVD is beneficial, while if they are above the line, a single-component MEVD is providing more accurate estimations

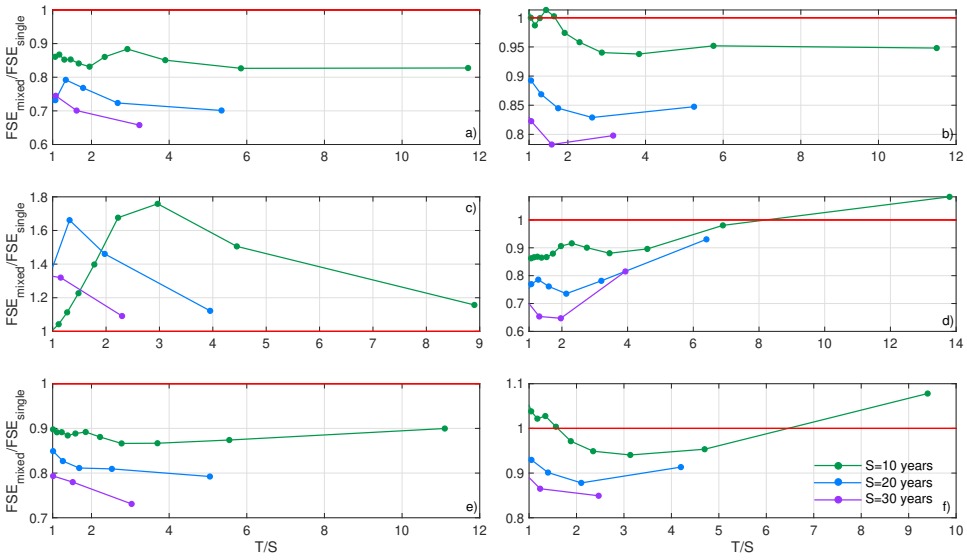


Figure 5.19: Ratio between the FSE computed with a mixed MEVD and the FSE obtained with the single-component MEVD, as a function of the ratio return period (T) to calibration sample size (S) for the stations analyzed, considering 3-days cumulative rainfall on the whole record and the 250 km buffer. The colors of the dots refer to the different sample sizes: $S = 10$ (green), 20 (light-blue) and 30 (purple) years. Each panel shows one station: a) Charleston, b) Charlotte, c) Houston, d) Miami, e) New Orleans and f) Coloso (Puerto Rico). The red line sets the threshold between the performances of the two MEVD approaches: if dots are located below this line, the use of mixtures in the MEVD is beneficial, while if they are above the line, a single-component MEVD is providing more accurate estimations

at several stations. Possible reasons for the mixed approach not being consistently superior to the single-component MEVD in the case of daily rainfall are the definition used for TC-induced daily rainfall and the temporal influence of TCs on rainfall fields. Considering only the maximum among the values of precipitation identified in a time window where a TC is present and attributing the others to non-TC rainfall is contaminating the two distributions, to variable degrees, depending on the features that TC-rainfall presents in each station (higher or lower duration of TC influence on rainfall in the area). Moreover, the life of a TC lasts much longer than one day (Wallace and Hobbs, 2006), and their impact on rainfall occurs on time scales longer than the daily one, hence a time window of 2-3 consecutive days better captures TC-induced rainfall.

The hypotheses above are corroborated by the results found when considering values of 3-day cumulative rainfall: in this case, the mixed approach outperforms the single-component MEVD because this window length allows a better description of the effect of a TC on rainfall. This emphasizes the importance of accurately defining and evaluating TC-induced rainfall on a time scale longer than the daily one, which would support water management and flood risk mitigation. The case of Houston remains an exception. Beside the low frequency of TCs in this are, other factors might play a role: a) the presence of tornadoes, which are categorized as generating non-TC rainfall but produce very heavy rainfall, and b) the amount of rainfall caused by Hurricane Harvey in August 2017, more than 800 mm/3 days, which affects the distributional parameters and quantile estimation, especially when non-TC and TC rainfall depths are kept separate.

Considering the different analyses performed by changing starting date, buffer size and calibration sample size, the main inferences are as follows:
a. results based on the whole available record and the subset starting in 1970 are found to be consistent;
b. the 500 km buffer shows a slight improvement in the use of the mixed-MEVD with respect to the single-component MEVD if compared to the 250 km buffer (compare Figures 5.15 and 5.18 with Figures 5.17 and 5.19 respectively); TCs signature on rainfall can be detected up to very large distances from the gauge (as indicated by Villarini et al. (2014));
c. increasing the calibration sample size from 10 to 30 years, especially in the case of cumulative rainfall, makes the advantage of using a mixed-MEVD generally higher; within a longer sample size it is more likely to have a sufficient number of TC-induced rainfall values, hence reducing the uncertainty in parameter estimation.

Chapter 6

Discussion and Conclusions

The work presented in this thesis is focused on the extension and optimization of the Metastatistical Extreme Value Distribution.

An important step forward has been made in understanding the main factors controlling the choice of the optimal estimation window used for parameter estimation. On the one hand, it is important to ensure that the size of the calibration sample is large enough to reduce uncertainty in parameter estimation, yet this is not the only factor that must be taken into account. The inter-annual variability of the distributions exerts a crucial role: provided that the number of values in the sample used for parameter estimation is large enough to allow robustly fitting a distribution, a yearly parameter estimation becomes optimal for lower values of average number of rainy days when the variability among the distributions is high.
When looking at trends, or performing non-stationary analyses, it is important to keep the estimation window as short as possible, in order to be able to follow climate variability. The MEVD in its optimized formulation (1 year estimation windows) is therefore an ideal candidate to perform trend analyses. The long series of daily rainfall at the Padova station, which allows analyses of long-term changes in rainfall regimes, was used to illustrate the benefits of the optimized MEVD approach in resolving the fluctuation and trends in the 100-year daily rainfall depth. It was shown how the MEVD minimizes the noise deriving from estimation uncertainty allowing a greater sensitivity in trend detection and an improved ability to resolve short time scale changes in extremes. The GEV distribution is indeed very sensitive to the possible presence of a few very large or very low maxima in the analysis window and it is largely affected by the use of short sliding windows, which are instead necessary to resolve the time variability of extremes.
Common methods used for trend analysis, e.g. quantile regression, are based on identifying a trend on high quantiles, and therefore are subject to the estimation uncertainty of the distribution chosen to describe them. Also in this sense, the MEVD can provide relevant support, being able to more accurate estimate the quantiles of interest.

This dissertation described the first application of the MEVD to flood frequency analysis. The main difference between discharge and rainfall records is that the former are characterized by a higher correlation; hence, an inhibition window to identify independent flood events is needed. A MEVD-Gamma was used to estimate extreme flow peaks over the CONUS, showing that it outperforms the traditional EVT in about 76% of the 5,311 USGS stations analyzed. The MEVD provided a reduction of about 30% in the estimation error for return periods that are higher than the sample used for parameter estimation.
Furthermore, given the growing interest in bringing physical mechanisms back into hydrological statistical analysis, a MEVD approach accounting for different flood generating mechanisms was developed, with particular focus on different ENSO phases. Differences in the probability distributions of ordinary peaks occurring during El Niño, La Niña or the neutral phase were demonstrated and a mixed MEVD ENSO-based was applied to streamflow records where the signature ENSO on flood events was identified. However, negligible or no improvement in the estimation accuracy of extreme values was found when including in the MEVD the different distributions based on phase detection.
Nonetheless, the ability of accounting for mixtures of distributions in the flood-peak MEVD formulation has significant practical potential: several flood drivers can be identified (e.g., snowmelt, rain-on-snow, ice jam, atmospheric rivers) whose role can be studied using the approach developed.
Regarding ENSO, its detection in the distributions of the ordinary peaks is also still valuable: the detection or prediction of the occurrence of different ENSO phases justifies the use of a phase-based MEVD with potential improvements for flood frequency analysis.

Leveraging the appealing property of the MEVD to naturally include the physical phenomena underlying hydrologic processes, the relevant topic of tropical cyclones-induced rainfall has been studied. Six metropolitan areas that are often impacted by TCs have been chosen as case studies. The ordinary distributions of non-TC and TC-induced daily rainfall were identified as statistically different, and distinguishing between the two event types was beneficial for extreme value analysis in some cases.
Due to the relatively long characteristic time scale associated with TCs effects, the daily time scale was found not to be the most appropriate scale for investigating this phenomenon. Hence, cumulative rainfall on time windows of 3 days was considered, and both the single-component MEVD and its mixed version were applied. The latter was found to yield a significant advantage.
The Houston station represented an exception, which spurred some interesting questions. In the Houston area case, the mixed MEVD approach did not provide improvements in estimation uncertainty at the daily scale, nor

for 3-days cumulative rainfall. The proposed explanation lies in the possible role of tornadoes, which cause heavy rainfall under non-TC conditions. Future work will need to extend the analysis based on a mixed MEVD which includes at least a third precipitation type, associated with the presence of tornadoes.

Chapter 7

Appendix

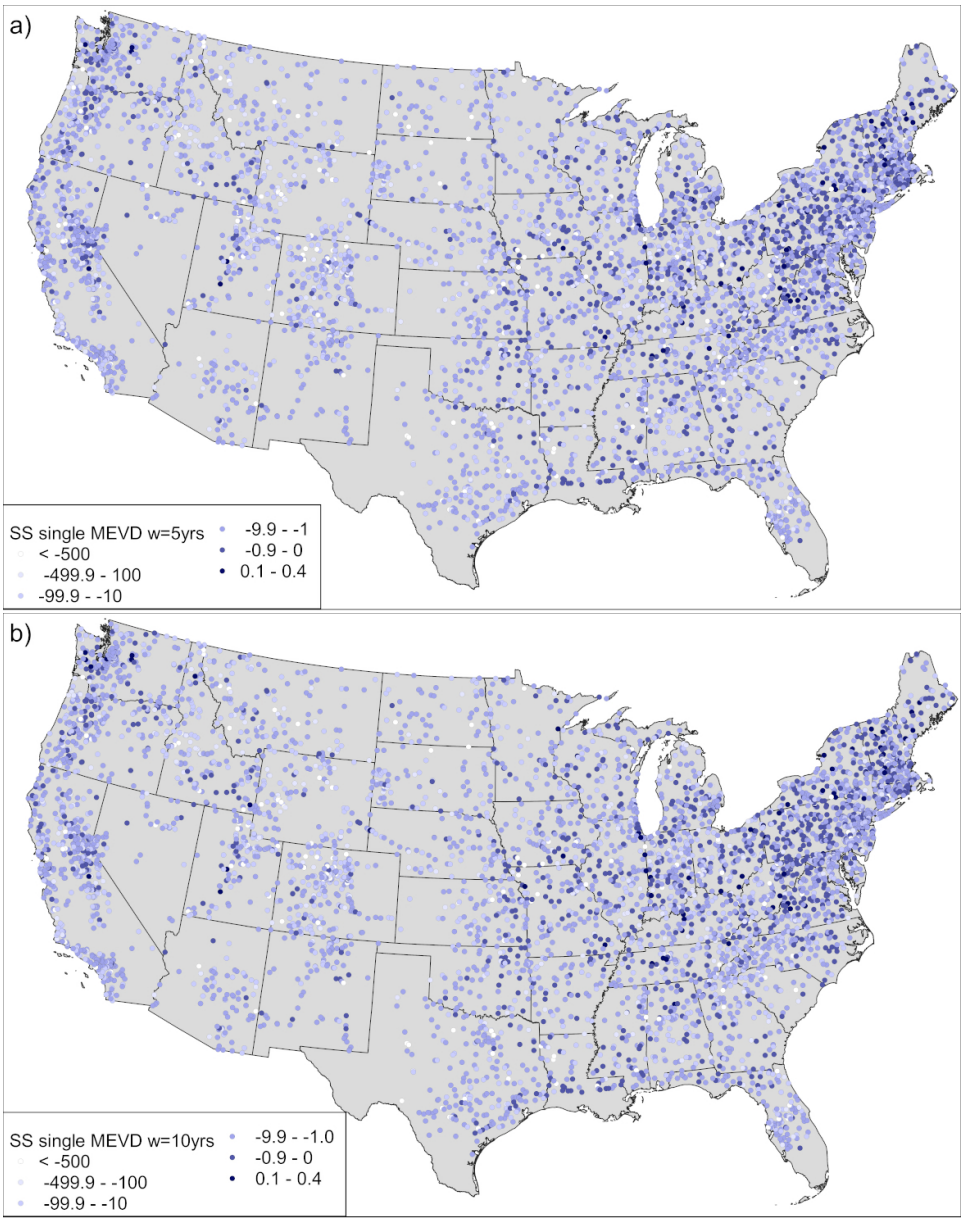


Figure 7.1: Skill score values for the single-distribution MEVD. Panel a (b) shows the skill score values for a MEVD with an estimation window of 5 (10) years.

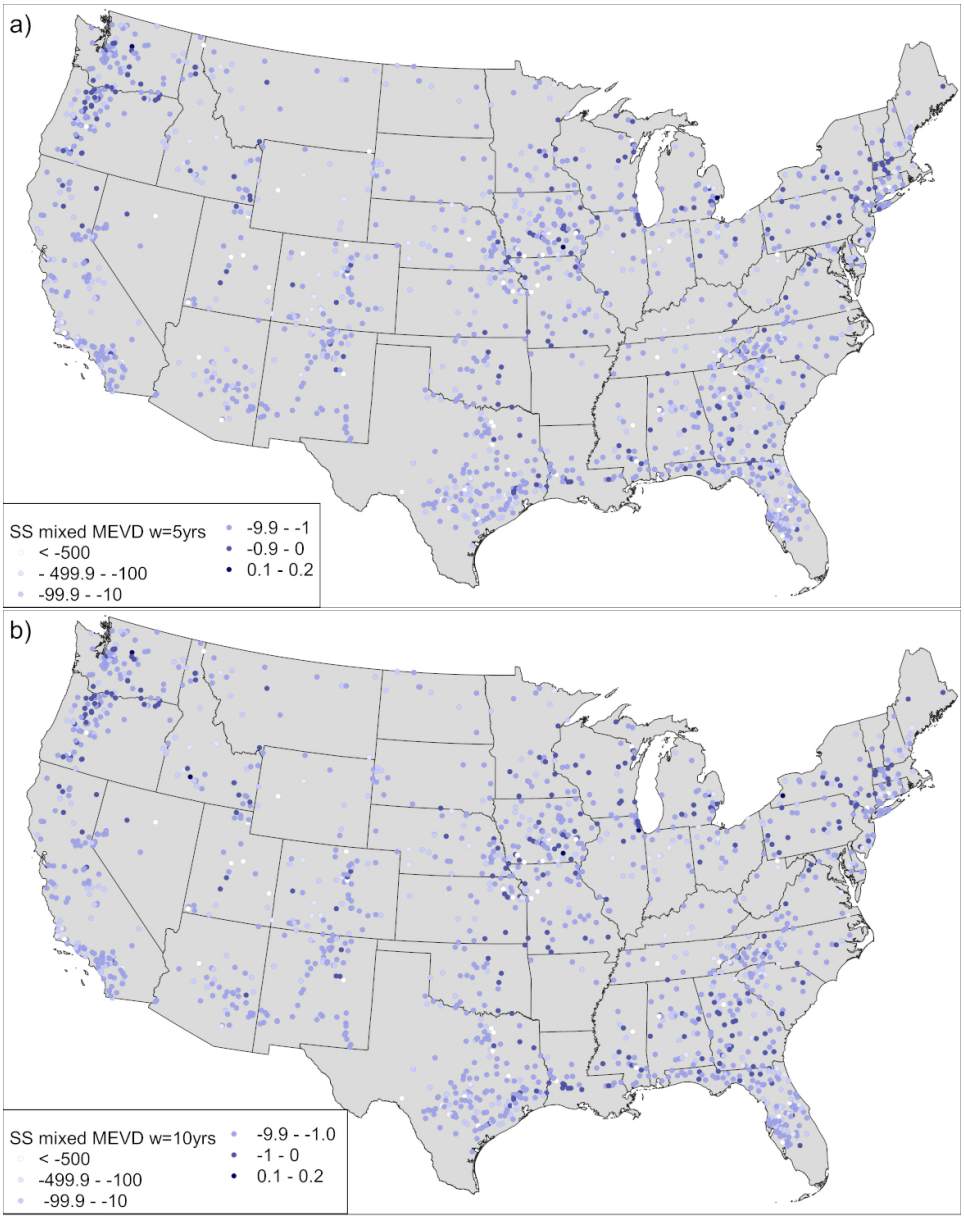


Figure 7.2: Skill score of the mixed MEVD. Panel a (b) shows the skill score values for a mixed MEVD with an estimation window of 5 (10) years.

Bibliography

Abdalla, O. and Al-Abri, R. (2011). Groundwater recharge in arid areas induced by tropical cyclones: lessons learned from Gonu 2007 in Sultanate of Oman. *Env. Earth Sci*, 63:229–239.

Alila, Y. and Mtiraoui, A. (2002). Implications of heterogeneous flood-frequency distributions on traditional stream-discharge prediction techniques. *Hydrological Processes*, 16(5):1065–1084.

Allamano, P., Croci, A., and Laio, F. (2015). Toward the camera gauge. *Water Resour. Res.*, 51:1744–1757.

Aryal, S., Bates, B., Campbell, E., Li, Y., Palmer, M., and Viney, N. (2009). Characterizing and modeling temporal and spatial trends in rainfall extremes. *J. Hydrometeorol.*, 10(1):241–253.

Aryal, Y., Villarini, G., Zhang, W., and Vecchi, G. (2018). Long term changes in flooding and heavy rainfall associated with North Atlantic tropical cyclones: Roles of the North Atlantic Oscillation and El Niño-Southern Oscillation. *J. Hydrol.*, 559:698–710.

Ashley, S. and Ashley, W. (2008). Flood fatalities in the United States. *J. Appl. Meteorol. Climatol.*, 47(3):805–818.

Balkema, A. and de Haan, L. (1974). Residual life time at great age. *Ann. Probbility*, 2(5):792–804.

Barlow, M. (2011). Influence of hurricane-related activity on North American extreme precipitation. *Geophys. Res. Lett.*, 38:L04705.

Barth, N., Villarini, G., Nayak, M., and White, K. (2017). Mixed populations and annual flood frequency estimates in the western United States: The role of atmospheric rivers. *Water Resour. Res.*, 53:257–269.

Barth, N., Villarini, G., and White, K. (2019). Accounting for Mixed Populations in Flood Frequency Analysis: Bulletin 17C Perspective. *J. Hydrol. Eng.*, 24(3):04019002.

Bartolini, E., Claps, P., and D'Odorico, P. (2009). Interannual variability of winter precipitation in the European Alps: Relations with the North Atlantic Oscillation. *Hydrol. Earth Syst. Sci.*, 13:17–25.

Beck, C. and Cohen, E. (2004). Superstatistics. *Phys. D Nonlinear Phenom.*, 322:267–275.

Beguería, S., Angulo-Martínez, M., Vicente-Serrano, S., López-Moreno, J., and El-Kenawy, A. (2011). Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: A case study in northeast Spain from 1930 to 2006. *Int. J. Climatol.*, 31(14):2102–2114.

Benson, M. (1962). Evolution of Methods for Evaluating the Occurrence of Floods. *U.S. Geol. Surv. Water Supply Pap.*, 1280-A.

Berg, P., Moseley, C., and Haerter, J. (2013). Strong increase in convective precipitation in response to higher temperatures. *Nat. Geosci.*, 6:181–185.

Berghuijs, W., Woods, R., Hutton, C., and Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. *Geophys. Res. Lett.*, 43(9):4382–4390.

Bonferroni, C. (1936). *Teoria statistica delle classi e calcolo delle probabilità*. Pubblicazioni del R. Istituto Superiore di Scienze Economiche e Commerciali di Firenze.

Botter, G., Basso, S., Rodriguez-Iturbe, I., and Rinaldo, A. (2013). Resilience of river flow regimes. *Proc. Natl. Acad. Sci.*, 110(32):12925–12930.

Bucknam, R., Coe, J., Chavarria, M., Godt, J., Tarr, A., Bradley, L.-A., Rafferty, S., Hancock, D., Dart, R., and Johnson, M. (2001). Landslides triggered by Hurricane Mitch in Guatemala - Inventory and discussion. Technical report, U.S. Geological Survey.

Cann, K., Thomas, D., Salmon, R., Wyn-Jones, A., and Kay, D. (2013). Extreme water-related weather events and waterborne disease. *Epidemiol. Infect.*, 141(4):671–686.

Castillo, E. (1988). *Extreme Value Theory in Engineering*. Academic Press Inc, San Diego, CA.

Chavez-Demoulin, V. and Davison, A. (2005). Generalized additive models for sample extremes. *Appl. Stat.*, 54:207–222.

Chen, X., Kumar, M., and Mcglynn, B. (2015). Variations in streamflow response to large hurricane-season storms in a southeastern U.S. watershed. *J. Hydrometeorol.*, 16:55–69.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London.

Coles, S. and Casson, E. (1998). Extreme value modelling of hurricane wind speeds. *Struct. Saf.*, 20(3):283–296.

Cook, N. and Harris, R. (2004). Exact and general FT1 penultimate distributions of extreme wind speeds drawn from tail-equivalent Weibull parents. *Struct. Saf.*, 26:391–420.

Curriero, F., Patz, J., Rose, J., and Lele, S. (2001). The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948-1994. *Am. J. Public Health*, 91(8):1194–1199.

Czajkowski, J., Villarini, G., Michel-Kerjan, E., and Smith, J. (2013). Determining tropical cyclone inland flooding loss on a large scale through a new flood peak ratio-based methodology. *Environ. Res. Lett.*, 8.

Czajkowski, J., Villarini, G., Montgomery, M., Michel-Kerjan, E., and Goska, R. (2017). Assessing Current and Future Freshwater Flood Risk from North Atlantic Tropical Cyclones via Insurance Claims. *Sci. Rep.*, 7:1–10.

Davison, A. and Smith, R. (1990). Models for Exceedances over High Thresholds. *J. R. Stat. Soc. Ser. B*, 52(3):393–442.

Dettinger, M., Ralph, F., Das, T., Neiman, P., and Cayan, D. (2011). Atmospheric Rivers, Floods and the Water Resources of California. *Water*, 3(2):445–478.

D'Odorico, P. and Fagherazzi, S. (2003). A probabilistic model of rainfall-triggered shallow landslides in hollows: A long-term analysis. *Water Resour. Res.*, 39(9):1–14.

Easterling, D., Meehl, G., Parmesan, C., Changnon, S., Karl, T., and Mearns, L. (2000). Climate Extremes: Observations, Modeling, and Impacts. *Science*, 289(5487):2068–2074.

Elsner, J. B., Kossin, J. P., and Jagger, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature*, 455:92–95.

Emanuel, K. (1991). The theory of hurricanes. *Annu. Rev. Fluid Mech.*, 23:179–196.

Emanuel, K. (2005). Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, 436:686–688.

Emanuel, K. (2017). Assessing the present and future probability of Hurricane Harvey's rainfall. *PNAS*, 114:12,681–12,684.

Emerton, R., Cloke, H., Stephens, E., Zsoter, E., Woolnough, S., and Pappenberger, F. (2017). Complex picture for likelihood of ENSO-driven flood hazard. *Nat. Commun.*, 8:1–9.

England, Jr., J., Cohn, T., Faber, B., Stedinger, J., Thomas, Jr., W., Veilleux, A., Kiang, J., and Mason, Jr, R. (2018). Guidelines for determining flood flow frequency – Bulletin 17C. Technical report, Reston,VA.

FEMA (2008). Taking shelter from the storm: building a safe room for your home or small business. Technical report.

Fisher, R. (1941). The negative binomial distribution. *Ann Eugen.*, 11(1):182–187.

Fisher, R. and Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math. Proc. Cambridge Philos. Soc.*, 24(2):180–190.

Fleig, A., Tallaksen, L., Hisdal, H., Demuth, S., Fleig, A., Tallaksen, L., Hisdal, H., and Demuth, S. (2006). A global evaluation of streamflow drought characteristics. *Hydrol. Earth Syst. Sci.*, 10(4):535–552.

Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Ann. Math.*, 44(3):423.

Greenwood, J., Landwehr, J., and Matalas, N. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water Resour. Res.*, 15(5):1049–1054.

Gumbel, E. J. (2004). *Statistics of extremes*. Dover Publications.

Hagen, A. B., Strahan-Sakoskie, D., and Luckett, C. (2012). A reanalysis of the 1944-53 atlantic hurricane seasons-the first decade of aircraft reconnaissance. *J. Clim.*, 25:4441–4460.

Haigh, I., Nicholls, R., and Wells, N. (2010). A comparison of the main methods for estimating probabilities of extreme still water levels. *Coast. Eng.*, 57(9):838–849.

Hann, C. (1977). *Statistical Methods in Hydrology*. The Iowa State University Press.

Hashino, T., Bradley, A. A., and Schwartz, S. S. (2006). Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrol. Earth Syst. Sci. Discuss.*, 3(2):561–594.

Heckert, N. A., Simiu, E., and Whalen, T. (1998). Estimates of hurricane wind speeds by "Peaks Over Threshold" method. *J. Struct. Eng.*, 124(4):445–449.

Hirschboeck, K. (1987). Hydroclimatically-defined mixed distributions in partial duration flood series. In *Hydrol. Freq. Model. Proc. Int. Symp. Flood Freq. Risk Anal.*, pages 199–212, Louisiana State University, Baton Rouge, U.S.A. Springer Netherlands, Dordrecht.

Hirschboeck, K. (1991). Hydrology of floods and droughts - climate and floods in "National water summary 1988-89: Hydrologic events and floods and droughts", U.S. Geological Survey Water-Supply Paper 2375. Technical report, U.S. Geological Survey.

Hosking, J. (1985). Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. *Technometrics*, 27(3).

Hosking, J. (1990). L-Moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. R. Stat. Soc. Ser. B*, 52(1):105–124.

IACWD (1982). Guidelines for determining flood flow frequency: Hydrology subcommittee bulletin 17B. Technical report, USGS, Reston,VA.

Kam, J., Sheffield, J., Yuan, X., and Wood, E. (2013). The influence of atlantic tropical cyclones on drought over the Eastern United States (1980-2007). *J. Clim.*, 26:3067–3086.

Katz, R., Parlange, M., and Naveau, P. (2002). Statistics of extremes in climatology and hydrology. *Adv. Water Resour.*, 25:1287–1304.

Kendall, M. (1975). *Rank Correlation Methods.* Charles Griffin, London.

Khaliq, M., Ouarda, T., Ondo, J., Gachon, P., and Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydrometeorological observations: A review. *J. Hydrol.*, 329(3-4):534–552.

Khouakhi, A., Villarini, G., and Vecchi, G. (2017). Contribution of tropical cyclones to rainfall at the global scale. *J. Clim.*, 30:359–372.

Kim, J., Ho, C., Lee, M., Jeong, J., and Chen, D. (2006). Large increase in heavy rainfall associated with tropical cyclone landfalls in Korea after the late 1970s. *Geophys. Res. Lett.*, 33:L18706.

Klein Tank, A. M. G. Zwiers, F. W. and Zhang, X. (2009). Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation, World Meteorological Organization. Technical report.

Klemeš, V. (1974). Some probems in pure and applied stochastic hydrology. In *Vol. 1275 Proc., Symp. Stat. Hydrol. Misc. Publ.*, pages 2–15, Washington, DC.

Klemeš, V. (1988). A hydrological perspective. *J. Hydrol.*, 100:3–28.

Klemeš, V. (1993). Probability of extreme hydrometeorological events; a different approach. In *Extrem. Hydrol. events Precipitation, Floods Droughts*, number 213, pages 167–176.

Knabb, R., Rhome, J., and Brown, D. (2011). Tropical Cyclone Report: Hurricane Katrina, 23-30 August 2005.

Knight, D. and Davis, R. (2009). Contribution of tropical cyclones to extreme rainfall events in the southeastern United States. *J. Geophys. Res.*, 114:D23102.

Knutson, T., McBride, J., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J., Srivastava, A., and Sugi, M. (2010). Tropical cyclones and climate change. *Nat. Geosci.*, 3(1):157–163.

Knutson, T., Sirutis, J., Zhao, M., Tuleya, R., Bender, M., Vecchi, G., Villarini, G., and Chavas, D. (2015). Global projections of intense tropical cyclone activity for the late twenty-first century from dynamical downscaling of CMIP5/RCP4.5 scenarios. *J. Clim.*, 28:7203–7224.

Koutsoyiannis, D. (2004a). Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. *Hydrol. Sci. J.*, 49(4):575–590.

Koutsoyiannis, D. (2004b). Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. *Hydrol. Sci. J.*, 49(4):591–610.

Kunkel, K.-E., Easterling, D., Kristovich, D., Gleason, B., Stoecker, L., and Smith, R. (2010). Recent increases in U.S. heavy precipitation associated with tropical cyclones. *Geophys. Res. Lett.*, 37:L24706.

Landsea, C., Feuer, S., Hagen, A., Glenn, D., Sims, J., Perez, R., Chenoweth, M., and Anderson, N. (2012). A reanalysis of the 1921-30 Atlantic hurricane database. *J. Clim.*, 25:865–885.

Landsea, C., Harper, B., Hoarau, K., and Knaff, J. (2006). Can we detect trends in extreme tropical cyclones? *Science*, 313:452–454.

Landsea, C. W. and Franklin, J. (2013). Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Weather Rev.*, 141:3576–3592.

Lang, M., Ouarda, T., and Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. *J. Hydrol.*, 225:103–117.

Mallakpour, I. and Villarini, G. (2017). Analysis of changes in the magnitude, frequency, and seasonality of heavy precipitation over the contiguous USA. *Theor Appl Clim.*, 130:345–363.

Mann, H. (1945). Nonparametric tests against trend. *Econometrica*, 13(3):245–259.

Marani, M. and Ignaccolo, M. (2015). A metastatistical approach to rainfall extremes. *Adv. Wat. Res.*, 79:121–126.

Marani, M. and Zanetti, S. (2015). Long-termoscillations in rainfall extremes in a 268 year daily time series. *Water Resour. Res.*, 51:639–647.

Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., and Morin, E. (2018). Metastatistical Extreme Value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors. *Adv. Wat. Res.*, 117:27–39.

Marra, F., Zoccatelli, D., Armon, M., and Morin, E. (2019). A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes. *Adv. Wat. Res.*, 127:280–290.

Martins, L., Wikuats, C. F. H., Capucim, M., de Almeida, D., da Costa, S., Albuquerque, T., Barreto Carvalho, V., de Freitas, E., de Fátima Andrade, M., and Martins, J. (2017). Extreme value analysis of air pollution data and their comparison between two large urban regions of South America. *Weather Clim. Extrem.*, 18:44–54.

Miller, S., Muir-Wood, R., and Boissonnade, A. (2008). An exploration of trends in normalized weather-related catastrophe losses. *Clim. Extrem. Soc.*, pages 225–247.

Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J., and Ward, P. J. (2016). A global reanalysis of storm surges and extreme sea levels. *Nat. Commun.*, 7:1–11.

Murphy, A. and Winkler, R. (1992). Diagnostic verification of probability forecasts. *J. Forecast.*, 7:435–455.

Nayman, E., Testut, N., Bothua, M., and Mayor, J. (2019). On the use of data mining and crowdsourcing to constrain historical seismicity. In *EVAN conference 2019*.

O'Connor, A. and O'Brien, E. (2005). Traffic load modelling and factors influencing the accuracy of predicted extremes. *Can. J. Civ. Eng.*, 32:270–278.

Ouarda, T. and El-Adlouni, S. (2011). Bayesian nonstationary frequency analysis of hydrological variables. *J. Am. Water Resour. Assoc.*, 47(3):496–505.

Palmieri, S., Teodonio, L., Siani, A., and Casale, G. (2006). Tropical storm impact in Central America. *Meteorol. Appl.*, 13:21–28.

Palynchuk, B. and Guo, Y. (2008). Threshold analysis of rainstorm depth and duration statistics at Toronto, Canada. *J. Hydrol.*, 348 (3-4):535–545.

Papalexiou, S. and Koutsoyiannis, D. (2013). Battle of extreme value distributions : A global survey on extreme daily rainfall. *Water Resour. Res.*, 49:187–201.

Papalexiou, S., Koutsoyiannis, D., and Makropoulos, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrol. Earth Syst. Sci.*, 17(2):851–862.

Papalexiou, S. and Montanari, A. (2019). Global and regional increase of precipitation extremes under global warming. *Water Resour. Res.*, pages 4901–4914.

Perry, C. (2000). Significant Floods in the United States During the 20th Century - USGS Measures a Century of Floods. *USGS Fact Sheet 024-00.*

Pickands, J. I. (1975). Statistical Inference Using Extreme Order Statistics. *Ann. Stat.*, 3:119–131.

Pielke Jr., R. (2005). Are there trends in hurricane destruction? *Nature*, 438:1.

Pielke Jr., R. (2007). Future economic damage from tropical cyclones: Sensitivities to societal and climate changes. *Phil. Trans. R. Soc. A*, 365:2717–2729.

Pielke Jr., R. and Landsea, C. (1998). Normalized hurricane damages in the United States: 1925-95. *Weather Forecast.*, 13:621–631.

Pisarenko, V., Sornette, A., Sornette, D., and Rodkin, M. (2014). Characterization of the tail of the distribution of earthquake magnitudes by combining the GEV and GPD descriptions of Extreme Value Theory. *Pure Appl. Geophys.*, 171:1599–1624.

Pisarenko, V. and Sornette, D. (2003). Characterization of the frequency of extreme earthquake events by the Generalized Pareto Distribution. *Pure Appl. Geophys.*, 160:2343–2364.

Porporato, A., Vico, G., and Fay, P. (2006). Superstatistics of hydro-climatic fluctuations and interannual ecosystem productivity. *Geophys. Res. Lett.*, 33(L15402).

Powell, J. and Reinhard, S. (2015). Measuring the effects of extreme weather events on yields. *Weather Clim. Extrem.*, 12:69–79.

Prat, O. and Nelson, B. (2013). Mapping the world's tropical cyclone rainfall contribution over land using the TRMM Multi-satellite Precipitation Analysis. *Water Resour. Res.*, 49:7236–7254.

Rappaport, E. (2000). Loss of life in the United States associated with recent atlantic tropical cyclones. *Bull. Amer. Meteor. Soc.*, 81(9):2065–2073.

Rappaport, E. (2014). Fatalities in the united states from atlantic tropical cyclones: new data and interpretation. *Bull. Amer. Meteor. Soc.*, 95:341–346.

Rebora, N., Molini, L., Casella, E., Comellas, A., Fiori, E., Pignone, F., Siccardi, F., Silvestro, F., Tanelli, S., and Parodi, A. (2013). Extreme rainfall in the mediterranean: What can we learn from observations. *J. Hydrometeorol.*, 14(3):906–922.

Ren, F., Wu, G., Dong, W., Wang, X., Wang, Y., Ai, W., and Li, W. (2006). Changes in tropical cyclone precipitation over China. *Geophys. Res. Lett.*, 33(L20702).

Roberts, E. (1979). Review of Statistics of Extreme Values with Applications to Air Quality Data: Part II. Applications. *J. Air Pollut. Control Assoc.*, 29(7):733–740.

Rosenberg, E., Keys, P., Booth, D., Hartley, D., Burkey, J., Steinemann, A., and Lettenmaier, D. (2010). Precipitation extremes and the impacts of climate change on stormwater infrastructure in Washington State. *Clim. Change*, 102(1-2):319–349.

Rosenzweig, C., Tubiello, F., Goldberg, R., Mills, E., and Bloomfield, J. (2002). Increased crop damage in the US from excess precipitation under climate change. *Glob. Environ. Chang.*, 12(3):197–202.

Rowe, S. and Villarini, G. (2013). Flooding associated with predecessor rain events over the Midwest United States. *Environ. Res. Lett.*, 8.

Schellander, H., Lieb, A., and Hell, T. (2019). Error structure of Metastatistical and Generalized Extreme Value Distributions for modeling extreme rainfall in Austria. *Earth Sp. Sci.*, page 2019EA000557.

Serinaldi, F. and Kilsby, C. (2014). Rainfall extremes: Toward reconciliation after the battle of distributions. *Water Resour. Res.*, 50(1):336–352.

Sheffer, N. A., Enzel, Y., Benito, G. andGrodek, T., Poart, N., Lang, M., and Naulet, R. oeur, D. (2003). Paleofloods and historical floods of the Ardèche River, France. *Water Resour. Res.*, 39:1–13.

Shepherd, J. M., Grundstein, A., and Mote, T. (2007). Quantifying the contribution of tropical cyclones to extreme rainfall along the coastal southeastern United States. *Geophys. Res. Lett.*, 34:L23810.

Slater, L., Singer, M., and Kirchner, J. (2015). Hydrologic versus geomorphic drivers of trends in flood hazard. *Geophys. Res. Lett.*, 42:370–376.

Slater, L. and Villarini, G. (2017). Evaluating the drivers of seasonal streamflow in the U.S. Midwest. *Water*, 9(9):1–22.

Smith, J., Villarini, G., and Baeck, M. (2011). Mixture distributions and the hydroclimatology of extreme rainfall and flooding in the eastern United States. *J. Hydrometeorol.*, 12:294–309.

Smith, L. and Baeck, M. L. (2015). "Prophetic vision, vivid imagination": The 1927 Mississippi River flood. *Water Resour. Res.*, 51:9964–9994.

Sousa, P., Trigo, R., Aizpurua, P., Nieto, R., Gimeno, L., and Garcia-Herrera, R. (2011). Trends and extremes of drought indices throughout the 20th century in the Mediterranean. *Nat. Hazards Earth Syst. Sci.*, 11(1):33–51.

Strömberg, D. (2007). Natural disasters, economic development, and humanitarian aid. *J. Econ. Perspect.*, 21(3):199–222.

Thayakaran, R. and Ramesh, N. (2017). Doubly stochastic Poisson pulse model for fine-scale rainfall. *Stoch. Environ. Res. Risk Assess.*, 31(3):705–724.

Tjøstheim, D. (1986). Some doubly stochastic time series models. *J. Time Ser. Anal.*, 7(1):51–72.

Tobías, A. and Scotto, M. (2005). Prediction of extreme ozone levels in Barcelona, Spain. *Environ. Monit. Assess.*, 100(1-3):23–32.

USWRC (1976). Guidelines for determining flood flow frequency. Technical report, United States Water Resources Council, Washington, DC.

van het Schip, T. I., Overeem, A., Leijnse, H., Uijlenhoet, R., Meirink, J. F., and van Delden, A. J. (2017). Rainfall measurement using cell phone links: classification of wet and dry periods using geostationary satellites. *Hydrological Sciences Journal*.

Vecchi, G. and Knutson, T. (2011). Estimating annual numbers of Atlantic hurricanes missing from the HURDAT database (1878-1965) using ship track density. *J. Clim.*, 24:1736–1746.

Villarini, G. (2016a). On the seasonality of flooding across the continental United States. *Adv. Water Resour.*, 87:80–91.

Villarini, G. and Denniston, R. (2016b). Contribution of tropical cyclones to extreme rainfall in Australia. *Int. J. Clim.*, 36:1019–1025.

Villarini, G., Goska, R., Smith, J., and Vecchi, G. (2014a). North atlantic tropical cyclones and U.S. flooding. *Bull. Amer. Meteor. Soc.*, 95:1381–1388.

Villarini, G., Lavers, D., Scoccimarro, E., Zhao, M., Wehner, M., Vecchi, G., Knutson, T., and Reed, K. (2014). Sensitivity of tropical cyclone rainfall to idealized global-scale forcings. *J. Clim.*, 27:4622–4641.

Villarini, G. and Slater, L. (2017). *Climatology of flooding in the United States*, volume 1.

Villarini, G. and Smith, J. (2013). Flooding in Texas: Examination of temporal changes and impacts of tropical cyclones. *J. Am. Water Resour. Assoc.*, 49(4):825–837.

Villarini, G., Smith, J., Baeck, M., Marchok, T., and Vecchi, G. (2011). Characterization of rainfall distribution and flooding associated with U.S. landfalling tropical cyclones: analyses of Hurricanes Frances, Ivan, and Jeanne (2004). *J. Clim.*, 116:D23116.

Villarini, G. and Strong, A. (2014c). Roles of climate and agricultural practices in discharge changes in an agricultural watershed in Iowa. *Agric. Ecosyst. Environ.*, 188:204–211.

Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., and Koutsoyiannis, D. (2015). One hundred years of return period: Strengths and limitations. *Wat. Resour. Res*, 51:8570–8585.

Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., and Koutsoyiannis, D. (2019). Save hydrological observations! Return period estimation without data decimation. *J. Hydrol.*, 571:782–792.

Von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Rev. Math. Union Interbalcanique*, pages 141–160.

Wallace, J. M. and Hobbs, P. V. (2006). *Atmospheric science. An introductory survey.* Elsevier, second edi edition.

Wallemacq, P. and House, R. (2018). Economic Losses, Poverty & Disasters 1998-2017. Technical report, United Nations Office for Disaster Risk Reduction (UNDRR) and Centre for Research on the Epidemiology of Disasters (CRED).

Webster, P., Holland, G., Curry, J., and Chang, H.-R. (2005). Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science*, 309:1844–1846.

Westra, S. and Sisson, S. (2011). Detection of non-stationarity in precipitation extremes using a max-stable process model. *J. Hydrol.*, 406(1-2):119–128.

Wilson, P. and Toumi, R. (2005). A fundamental probability distribution for heavy rainfall. *Geophys. Res. Lett.*, 32(14):1–4.

WMO (2009). Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation. Technical report.

Wright, D., Bosma, C., and Lopez-Cantu, T. (2019). U.S. hydrologic design standards insufficient due to large increases in frequency of rainfall extremes. *Geophys. Res. Lett.*, page GL083235.

Zhang, M., Chen, X., Kumar, M., Marani, M., and Goralczyk, M. (2017). Hurricanes and tropical storms: A necessary evil to ensure water supply? *Hydrol. Process.*, 31:4414–4428.

Zhang, W. and Villarini, G. (2017). On the unseasonal flooding over the Central United States during December 2015 and January 2016. *Atmos. Res.*, 196(November 2016):23–28.

Zhou, X.-Y., Schmidt, F., and Toutlemonde, F.and Jacob, B. (2016). A mixture peaks over threshold approach for predicting extreme bridge traffic load effects. *Probabilistic Eng. Mech.*, 43:121–131.

Zorzetto, E., Botter, G., and Marani, M. (2016). On the emergence of rainfall extremes from ordinary events. *Geophys. Res. Lett.*, 43:8076–8082.

Zorzetto, E. and Marani, M. (2019). Downscaling of rainfall extremes from satellite observations. *Wat. Resour. Res*, 55(1):156–174.