Comparison of Local, Regional, and Scaling Models for Rainfall Intensity–Duration–Frequency Analysis

GIUSEPPE MASCARO

School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, Arizona

(Manuscript received 20 April 2020, in final form 4 August 2020)

ABSTRACT: Intensity–duration–frequency (IDF) analyses of rainfall extremes provide critical information to mitigate, manage, and adapt to urban flooding. The accuracy and uncertainty of IDF analyses depend on the availability of historical rainfall records, which are more accessible at daily resolution and, quite often, are very sparse in developing countries. In this work, we quantify performances of different IDF models as a function of the number of available high-resolution (N_τ) and daily (N_{24h}) rain gauges. For this aim, we apply a cross-validation framework that is based on Monte Carlo bootstrapping experiments on records of 223 high-resolution gauges in central Arizona. We test five IDF models based on (two) local, (one) regional, and (two) scaling frequency analyses of annual rainfall maxima from 30-min to 24-h durations with the generalized extreme value (GEV) distribution. All models exhibit similar performances in simulating observed quantiles associated with return periods up to 30 years. When $N_\tau > 10$, local and regional models have the best accuracy; bias correcting the GEV shape parameter for record length is recommended to estimate quantiles for large return periods. The uncertainty of all models, evaluated via Monte Carlo experiments, is very large when $N_\tau \leq 5$; however, if $N_{24h} \geq 10$ additional daily gauges are available, the uncertainty is greatly reduced and accuracy is increased by applying simple scaling models, which infer estimates on subdaily rainfall statistics from information at daily scale. For all models, performances depend on the ability to capture the elevation control on their parameters. Although our work is site specific, its results provide insights to conduct future IDF analyses, especially in regions with sparse data.

KEYWORDS: Extreme events; Precipitation; Hydrology; Statistics

1. Introduction

Urban flooding is being increasingly recognized as a pressing global challenge because of growing population concentration in urban regions, the rapid conversion of natural landscapes into impervious surfaces, and climate change (National Academies of Sciences, Engineering, and Medicine 2019; Zhang et al. 2018). As a result of the complex combination of these factors, urban areas have been suffering significant socioeconomic impacts from flooding not only due to catastrophic events (e.g., hurricanes; NWS 2019), but also after more frequent and less intense storms (University of Maryland College Park and Texas A&M University 2018). A key task to mitigate, manage, and adapt to urban flooding is the characterization of rainfall extreme statistics at different durations, with focus on short time accumulations (<24 h) that—in several cases—represent the hydrologic response times of urban watersheds. In water resources engineering, this statistical information is synthesized through intensity–duration–frequency (IDF) curves, which provide an estimate of rainfall intensities $i(T_R, \tau)$ for different time durations τ and frequencies F, usually expressed as return periods $T_R = 1/(1 - F)$.

The generation of IDF curves has been the subject of several studies, mainly in the areas of technical and stochastic hydrology (e.g., Burlando and Rosso 1996; Koutsoyiannis et al. 1998; Madsen et al. 2002; Requena et al. 2019; Tyralis and Langousis 2019; and many others). When rainfall records are available at sufficiently fine temporal resolutions at a given site, the most common approach to conduct local IDF analyses consists of estimating $i(T_R, \tau)$ via frequency analysis of annual rainfall maxima at different τ with parametric statistical distributions. For example, as suggested by the extreme value theory (Coles 2001), the generalized extreme value (GEV) distribution has been shown to be able to model well annual rainfall maxima at several locations (e.g., Papalexiou and Koutsoyiannis 2013; Blanchet et al. 2016). If historical rainfall records are available at multiple sites, regional IDF curves are often generated by (i) spatially interpolating $i(T_R, \tau)$ or parameters of the statistical distributions from local or at-site estimations, or (ii) applying regionalization techniques that merge rain gauges into homogeneous regions to increase robustness in the estimate of the statistical distribution parameters (Hosking and Wallis 1997). As a notable example, the regionalization technique proposed by Hosking and Wallis (1997) was adopted in the National Oceanic and Atmospheric Administration (NOAA) Atlas 14 to generate IDF curves for most of the conterminous United States (CONUS; Bonnin et al. 2004).

The accuracy and uncertainty of IDF analyses are dependent on quality and availability of historical records. In developed countries, rainfall data are in general available; however, data availability is higher at daily scale compared to subdaily resolutions in terms of both spatial density and record length. As a result, the uncertainty of IDF estimates tends to be higher for subdaily durations. In several developing countries,

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JAMC-D-20-0094.s1.

Corresponding author: Giuseppe Mascaro, gmascaro@asu.edu

DOI: 10.1175/JAMC-D-20-0094.1

where urbanization proceeds at fast rates (Cohen 2006) and IDF curves are urgently needed, the problem of data availability is even more critical, since rainfall records are available at sparse locations (Sane et al. 2018) and almost exclusively at daily resolution (Lumbroso et al. 2011; Liew et al. 2014). These challenges have significantly limited the ability to carry out IDF analyses, so that recent studies have proposed the use of satellite (Ombadi et al. 2018), reanalysis products (Courty et al. 2019), and weather radars (Marra and Morin 2015) to complement observations from sparse rain gauges.

A strategy to address the lack of rain gauges at subdaily resolution is to infer estimates of $i(T_R, \tau)$ for $\tau < 24$ h from information on the rainfall statistics at the daily scale, which—as said—is more largely available (Koutsoyiannis and Foufoula-Georgiou 1993; Burlando and Rosso 1996; Menabde et al. 1999; Koutsoyiannis et al. 1998). A class of statistical models, known as scaling models, utilizes the evidence of temporal scaling properties of rainfall processes (Lovejoy and Schertzer 1985; Deidda et al. 1999; Venugopal et al. 1999; Mascaro et al. 2013, 2014) to derive simple relations linking parameters of the probability distributions of rainfall maxima at different subdaily durations to those estimated at daily scale. While the theoretical background of scaling models has been outlined more than two decades ago (e.g., Burlando and Rosso 1996), these techniques have been mainly applied for analyses at distinct locations, because of the lack of networks of long-term high-resolution rain gauges. Notable exceptions are the studies by Yu et al. (2004), Borga et al. (2005), Bara et al. (2010), and Blanchet et al. (2016) who used scaling models to derive regional IDF relationships in areas of northern Taiwan (~7400 km^2), northern Italy (~6400 km^2), western Slovakia (~14 000 km^2), and southern France (~38 000 km^2) with 46, 91, 19, and ~300 rain gauges, respectively. More recently, Sane et al. (2018) applied a scaling model with 14 rain gauges to derive IDF curves for the entire country of Senegal (~197 000 km^2).

The present study has the main goal of rigorously testing the accuracy of different models for IDF analysis as a function of the number of rain gauges available at different temporal resolutions. To pursue this goal, we use data from a high-resolution network of 223 gauges with record lengths from 20 to 40 years located in an area of ~29 600 km^2 in central Arizona including the Phoenix metropolitan region (Mascaro 2017, 2018). We apply existing IDF models based on local (or at-site), regional, and scaling frequency analyses with the GEV distribution. In addition, we (i) evaluate the effect of bias correcting local estimates of the GEV shape parameter for record length (Papalexiou and Koutsoyiannis 2013) and (ii) propose and test for the first time a simple IDF model that applies the scaling theory to the regional growth curves defined by Hosking and Wallis (1997). The IDF models vary in terms of number of regional and spatially interpolated parameters. While all models require high-resolution rainfall records for their calibration, scaling models also take advantage of the availability of additional daily gauges. We first quantify the performance of all methods using all available gauges. Performance is assessed through metrics measuring the ability to simulate empirical and at-site quantiles. We then apply a cross-validation framework based on Monte Carlo bootstrapping experiments to quantify the uncertainty of model performances under different combinations of available high-resolution and daily rain gauges. Finally, we discuss our findings in the context of the rainfall regime of our study region and investigate the physical factors affecting performances of the different models. Results of this work will be useful, in future applications, to guide the selection of the technique for IDF generation that provides the best accuracy and lowest uncertainty under a given combination of available daily and high-resolution rain gauges.

2. Study area and dataset

The study area is located in a region in central Arizona that includes the Phoenix metropolitan area (Fig. 1a). The majority of the region is part of the Sonoran Desert where elevation is low (from 200 to 600 m MSL), while a minor portion is located in the transition zone to the Colorado Plateau, where elevation reaches up to 2000 m MSL in the Mogollon Rim. Climate is arid and characterized by a strong seasonality affecting the rainfall regime. In the summer months from July to September, the occurrence of the North American monsoon leads to localized convective thunderstorms with high rainfall intensity over short durations (<1 h). From November to March (hereinafter called winter), prolonged dry periods are interrupted by cold fronts that cause widespread storms lasting for up to multiple days. As a result of these different rainfall-generating mechanisms, annual rainfall maxima occur mainly in winter for $\tau \geq 12$ h and almost exclusively in summer for $\tau \leq 2$ h (Fig. 1b). Mascaro (2017, 2018) found also that annual rainfall maxima tend to increase with elevation and that the orographic control is more significant for larger time accumulations.

We use records of the Automated Local Evaluation in Real Time (ALERT) rain gauge network of the Flood Control District of the Maricopa County (FCDMC), installed to monitor in real time regional and localized storms, and support flood and flash-flood forecasting. The gauges have been gradually deployed since the beginning of the 1980s; as of 2020, their number is 365. In this study, 223 gauges with more than 20 years of observation are selected for the analysis (Figs. 1c,d). These gauges cover an area of ~29 600 km^2, with a mean density of ~1 gauge every 100 km^2 that increases to 4.3 gauges in urban areas covering ~2000 km^2. The gauge elevations range from 220 to 2325 m MSL with the majority (195) installed below 800 m. All gauges are of the tipping-bucket type, with an accumulated rain depth of 1 mm. More details on this dataset and the characteristics of the study region are provided by Mascaro (2017, 2018), who used these rainfall records to investigate the spatiotemporal variability of rainfall statistical properties and the seasonality of daily extremes.

The FCDMC provides gauge data in raw format containing the tipping instants in seconds. We derive the cumulative rainfall signal for each year from the tipping instants with the technique described by Mascaro et al. (2013; see the appendix), which allows avoiding the discretization caused by the box counting method, that is often applied, and generating instead a smoother and more realistic signal. From this, the annual maximum rainfall intensities (mm h^{-1}) are extracted

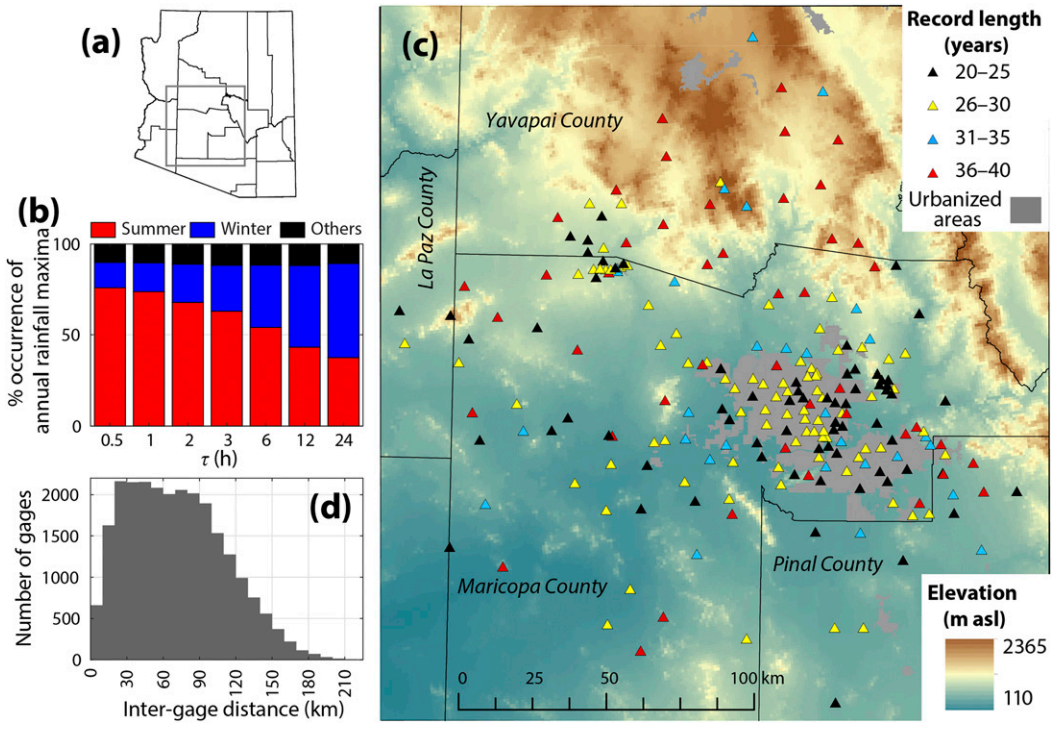

FIG. 1. (a) Study area within the county boundaries in Arizona. (b) Percent occurrence of annual rainfall maxima, averaged across the 223 gauges, for different time accumulations τ in summer (July–September), winter (November–March), and other months. (c) Digital elevation model (derived from the U.S. Geological Survey National Elevation Dataset) of the study region in Maricopa, La Paz, Yavapai, and Pinal Counties along with locations of the rain gauges of the FCDMC network, with colors indicating the corresponding record length. Urbanized areas are also shown. (d) Distribution of the intergauge distance.

for $\tau = 30$ min, 1, 2, 3, 6, 12, and 24 h through a moving window with duration τ. These durations are selected because a single scaling regime, which is needed to apply the IDF scaling models, emerges from 24 h to 30 min, consistent with previous applications (Menabde et al. 1999). We note that another regime appears in many stations for durations below 30 min, which are typical of single storms and cells. The existence of separate regimes at lower durations has been also reported by Innocenti et al. (2017) in an analysis of ~2700 stations in North America. In particular, these authors found a breaking point at ~1 h, which is higher than ours; this difference can be attributed to the diverse resolution of their datasets (15-min rainfall data with resolution of 2.54 mm). The approach proposed by Blanchet et al. (2016), in turn based on Papalexiou and Koutsoyiannis (2013), is adopted to retain or discard the annual rainfall maxima in years with missing data. The idea of this method is that, even if some data are missing in a given year, the annual maximum could have still been observed on that year and this value should be preserved. Assuming that N years without missing data are available and that a given year has a fraction f of missing data, the method involves the following steps: (i) the annual maxima of the N years are sorted, (ii) for a year with missing data, the annual maximum rainfall is extracted and its rank in the N sorted values is found, and (iii) this record is discarded or preserved if its rank is below or above fN, respectively. This procedure is applied for each duration τ. The resulting dataset includes a total of 44 781 annual rainfall maxima for seven durations over 223 gauges.

3. Methods

a. The GEV distribution

The GEV distribution is used to describe the frequency of annual rainfall maxima at different durations τ in all IDF models tested here. If I_τ is the random variable of the rainfall maxima, the cumulative distribution function (CDF) of the GEV is defined as

$$F(x;k_\tau,\mu_\tau,\sigma_\tau) = \Pr(I_\tau \le x) = \begin{cases} \exp\left[-\left(1+k_\tau\dfrac{x-\mu_\tau}{\sigma_\tau}\right)^{-1/k_\tau}\right] & k_\tau \neq 0 \\ \exp\left[-\exp\left(-\dfrac{x-\mu_\tau}{\sigma_\tau}\right)\right] & k_\tau = 0 \end{cases}, \quad (1)$$

where $k_\tau \in (-\infty, +\infty)$ is the shape parameter, $\mu_\tau \in (-\infty, +\infty)$ is the location parameter, and $\sigma_\tau > 0$ is the scale parameter. For $k_\tau = 0$, the GEV is of type I or Gumbel (exponential) and x is defined in the set $-\infty < x < \infty$. For $k_\tau > 0$, the GEV is of type II or Fréchet and is "heavy tailed" (subexponential); in this case,

TABLE 1. Overview of the IDF statistical models. Symbols are defined in the main text. On the basis of the categories described in section 3, At-site and At-siteBC are local models, Reg is a regional model, Sc is a local scaling model, and RegSc is a regional scaling model. The estimate of variables identified by $^{(+)}$ is improved when additional daily gauges are available.

Model	Description of parameter estimation	Data required for application at ungauged sites	No. of regional parameters (P) and maps (M)
At-site	For each τ, local estimation of k_τ, μ_τ, and σ_τ	Estimates of k_τ, μ_τ, and σ_τ from interpolated maps	$P = 0$; $M = 3D$
At-siteBC	For each τ, bias correction of k_τ and local reestimation of μ_τ and σ_τ	Estimates of k_τ, μ_τ, and σ_τ from interpolated maps	$P = 0$; $M = 3D$
Reg	For each τ, regional estimation of k_τ, μ_τ, and σ_τ of growth curves	Regional k_τ, μ_τ, and σ_τ of growth curves for each τ; estimates of $\bar{x}_\tau$ from interpolated maps	$P = 2D$; $M = D$
Sc	Local estimation of η, k^*, $\mu_{\tau 0}$, and $\sigma_{\tau 0}$ ($\tau_0 = 24$ h)	Estimates of $k^{*(+)}$, $\mu_{\tau 0}^{(+)}$, $\sigma_{\tau 0}^{(+)}$, and η from interpolated maps	$P = 0$; $M = 4$
RegSc	Regional estimation of k, μ, and σ of a single growth curve valid for all $\tau^{(*)}$; regional estimation of η for $\bar{x}_\tau$	Regional k, μ, and σ of the single growth curve; estimates of $\bar{x}_{\tau 0}^{(+)}$ and η from interpolated maps	$P = 2$; $M = 2$

$\mu_\tau - (\sigma_\tau/k_\tau) \le x \le \infty$. If $k_\tau < 0$, the GEV is of type III or Weibull and is bounded above so that $-\infty < x \le \mu_\tau - (\sigma_\tau/k_\tau)$. The quantile of the GEV for a given return period $T_R = 1/(1 - F)$ is computed as

$$i(T_R, \tau) = \begin{cases} \mu_\tau - \dfrac{\sigma_\tau}{k_\tau}\left\{1 - \left[-\log\left(1 - \dfrac{1}{T_R}\right)\right]^{-k_\tau}\right\} & k_\tau \neq 0 \\ \mu_\tau - \sigma_\tau \log\left[-\log\left(1 - \dfrac{1}{T_R}\right)\right] & k_\tau = 0 \end{cases}. \quad (2)$$

Prior to applying the IDF models, we assess the suitability of the GEV through L-moments ratio diagrams and goodness-of-fit (GOF) tests, including Lilliefors, Anderson–Darling, and Cramér–von Mises. The GOF tests are applied locally following the modification proposed by Deidda and Puliga (2006) to account for the data discretization at 1-mm resolution. To account for results of multiple tests, the false discovery ratio test of Wilks (2006) is then used to evaluate the field or global significance with the resulting p values. The false discovery ratio test is performed at a global significance level $\alpha_{\text{global}} = 0.05$; to consider the presence of spatial correlation of the rainfall records (Mascaro 2017), we conduct the local test at the significance level $\alpha_{\text{local}} = 2{\cdot}\alpha_{\text{global}}$, as suggested by the synthetic experiments of Wilks (2016; his Fig. 4). Since some of the IDF models tested here are based on regionalization techniques, the GOF metric Z proposed by Hosking and Wallis (1997) is also computed. This metric quantifies the degree to which the L-moments of the fitted GEV distribution match the regional averages of the sampling L-moments. The fitted GEV is considered adequate if $|Z| \le 1.64$.

b. Statistical models of rainfall IDF analysis

We apply two IDF models based on at-site estimations, one IDF model based on a regionalization technique, and two IDF simple scaling models. The models are described in the next sections, and their main features are summarized in Table 1.

1) IDF models based on local parameter estimation

The simplest IDF model (hereinafter labeled as At-site) consists of locally estimating the GEV parameters k_τ, μ_τ, and σ_τ for each τ, using the probability-weighted moments (PWM) method, which has been shown to perform better than alternative techniques for relatively short sample sizes (Hosking and Wallis 1997). In an analysis of long-term global daily rainfall, Papalexiou and Koutsoyiannis (2013) showed that the Fréchet distribution ($k_\tau > 0$) is the best suited to model annual maxima and proposed a set of analytical equations to bias correct k_τ as a function of the sample size. Here, to account for the sample size effect, an additional at-site IDF model is introduced where k_τ at each gauge is bias corrected using the relations proposed by Papalexiou and Koutsoyiannis (2013), and μ_τ, σ_τ are reestimated conditioned on the bias-corrected k_τ. This IDF model is labeled At-siteBC (or At-site bias corrected). For both At-site and At-siteBC models, estimates of $i(T_R, \tau)$ at ungauged locations can be obtained by spatially interpolating k_τ, μ_τ, and σ_τ and, then, applying Eq. (2). If D is the total number of durations, the application of these methods requires then $3D$ maps (Table 1).

2) IDF model based on regional parameter estimation

Regionalization techniques have been proposed to increase the robustness of parameter estimation by combining records of multiple stations located in a region where the underlying statistical distribution of the annual rainfall maxima can be considered the same except for the mean. Here, we use the regionalization technique based on the index flood procedure proposed by Hosking and Wallis (1997) and used in several studies (Fowler and Kilsby 2003; Bonnin et al. 2004; Liu et al. 2015; García-Marín et al. 2019; among others). This method requires the estimation of a regional dimensionless distribution, known as growth curve, using the weighted averages of the L-moments calculated from records of all stations in the region (the weights are the sample sizes at each station). Local estimates of $i(T_R, \tau)$ are then obtained by multiplying the T_R

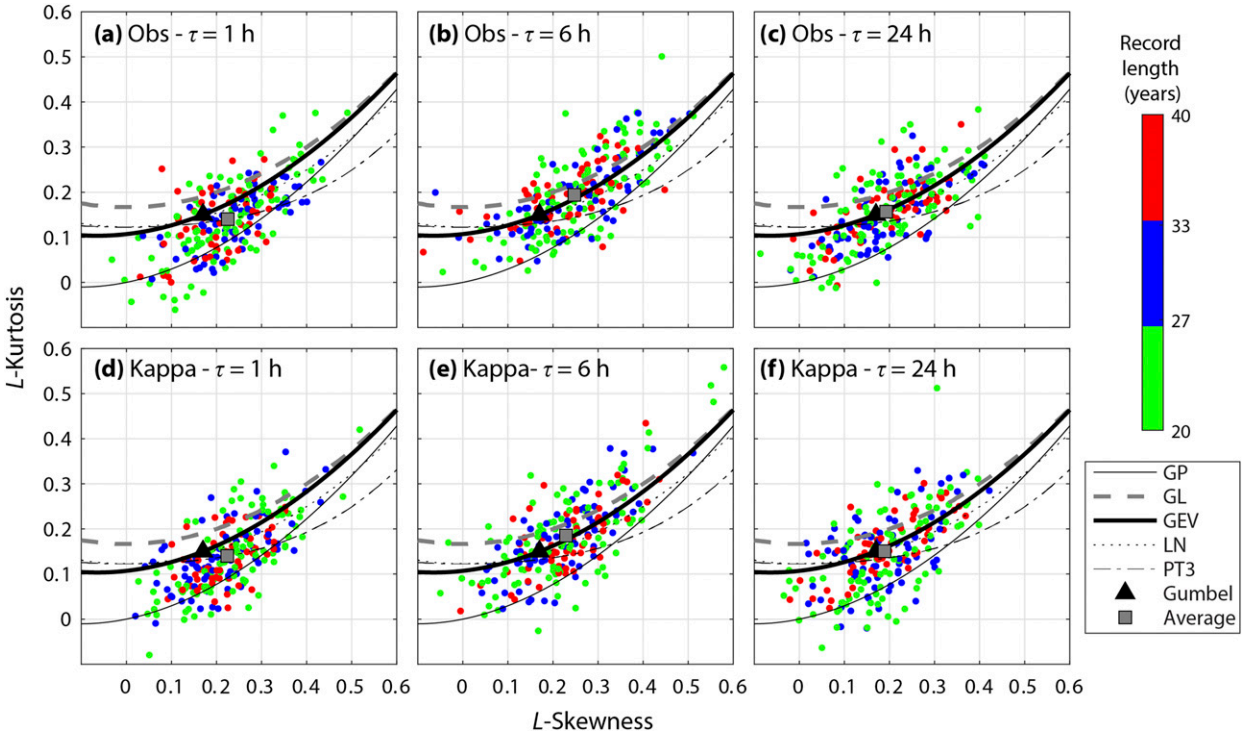

FIG. 2. The L-moment ratio diagrams (i.e., relation between L-skewness and L-kurtosis) for (a)–(c) observed records of annual rainfall maxima at durations $\tau = 1$, 6, and 24 h and (d)–(f) the corresponding synthetic records generated with the kappa distribution (see appendix for details). The sampling L-moment ratios are color coded on the basis of the sample length and are plotted along with the corresponding means (square). Theoretical relationships are plotted for the Gumbel (triangle), generalized Pareto (GP), generalized logistic (GL), generalized extreme value (GEV), lognormal (LN), and Pearson Type III (PT3).

quantiles of the growth curve, $i_{GC}(T_R, \tau)$, by the index flood, a variable representative of local rainfall characteristics. In this study, the GEV distribution is used to model the growth curve and the mean annual rainfall maxima at duration τ, $\bar{x}_\tau$, is utilized as index flood, leading to

$$i(T_R, \tau) = \bar{x}_\tau i_{GC}(T_R, \tau). \tag{3}$$

The application of the regionalization method at ungauged sites can be achieved by spatially interpolating the sample index floods at available stations. This IDF model is labeled as Reg; its application requires $2D$ regional parameters (two GEV parameters of the growth curves for D durations; note that, since the mean of the growth curve is 1, the third parameter is calculated as a function of the first two) and D maps of $\bar{x}_\tau$ (Table 1). Here, we apply the Reg IDF model using a single homogenous region, whose existence is tested by computing, for each considered duration, the heterogeneity measure H^* proposed by Hosking and Wallis (1997) and described in the appendix.

3) Local and regional IDF scaling models

A brief theoretical overview of simple scaling models is first provided. Based on empirical findings, Koutsoyiannis et al. (1998) proposed a mathematical formula to link $i(T_R, \tau)$ with τ and T_R:

$$i(T_R, \tau) = a(T_R)(\tau + \theta)^{-\eta}, \tag{4}$$

where $a(T_R)$ is determined from the probability distribution of the maximum rainfall intensity (e.g., the GEV), and $\theta \geq 0$ and $0 < \eta \leq 1$ are nonnegative coefficients to be estimated at the target site. By separating T_R from τ, an important advantage of Eq. (4) is that it allows relating rainfall intensities at different durations (e.g., τ and τ_0) for the same T_R:

$$i(T_R, \tau) = \frac{(\tau + \theta)^{-\eta}}{(\tau_0 + \theta)^{-\eta}} i(T_R, \tau_0). \tag{5}$$

For $\theta = 0$, it has been shown (e.g., Burlando and Rosso 1996; Blanchet et al. 2016) that, when applied to random variables, Eq. (5) corresponds to the hypothesis of simple scaling:

$$I_\tau \overset{d}{=} (\tau/\tau_0)^{-\eta} I_{\tau_0}, \tag{6}$$

where the symbol $\overset{d}{=}$ indicates equality in distribution and η is the scaling exponent. Equation (6), in turn, implies that moments of order q are related as

$$E(I_\tau^q) = (\tau/\tau_0)^{-q\eta} E(I_{\tau_0}^q). \tag{7}$$

A simple scaling model can be derived from Eq. (7) using the GEV distribution under the assumption of a constant shape parameter for all durations (i.e., $k_\tau = k^*$). In this

case, Eq. (7) leads to the following scaling relations for μ_τ and σ_τ:

$$\mu_\tau = (\tau/\tau_0)^{-\eta}\mu_{\tau_0} \quad \text{and} \tag{8a}$$

$$\sigma_\tau = (\tau/\tau_0)^{-\eta}\sigma_{\tau_0} \tag{8b}$$

Equations (8) imply that, if the GEV parameters are known at a coarse scale τ_0 (here, 24 h), then GEV parameters can be derived at smaller scales τ via the scaling exponent η. Recently, the accuracy of this scaling model was tested by Innocenti et al. (2017) on ~2700 gauges located in North America, showing adequate performances over different scaling regimes from 1 h to 7 days. In this study, we follow the application of the GEV scaling model of Blanchet et al. (2016), which requires the estimation of $\mu_{\tau 0}$, $\sigma_{\tau 0}$, k^*, and η at each gauge by maximizing the log-likelihood function that includes all durations at the same time (Blanchet et al. 2016). To apply this model (hereinafter labeled Sc) at ungauged sites, the four parameters have to be spatially interpolated; as a result, the model requires four spatial maps (Table 1). When available, records of additional daily gauges can be used to improve the accuracy of the spatial maps of $\mu_{\tau 0}$, $\sigma_{\tau 0}$, and k^*.

In addition to the at-site scaling model Sc, we develop here (to our knowledge, for the first time) a regional simple scaling model by combining the scaling relations in Eqs. (5) and (7) to Eq. (3) linking the quantiles of local distribution and growth curve. Using Eq. (3), the ratio between the quantiles of the local distributions for two durations τ and τ_0 can be expressed as

$$\frac{i(T_R,\tau)}{i(T_R,\tau_0)} = \frac{\overline{x}_\tau}{\overline{x}_{\tau_0}} \frac{i_{\mathrm{GC}}(T_R,\tau)}{i_{\mathrm{GC}}(T_R,\tau_0)}. \tag{9}$$

Per Eq. (5) with $\theta = 0$: $i(T_R, \tau)/i(T_R, \tau_0) = (\tau/\tau_0)^{-\eta}$. It is also easy to note that, for $q = 1$, Eq. (7) implies that

$$\overline{x}_\tau \big/ \overline{x}_{\tau_0} = (\tau/\tau_0)^{-\eta}. \tag{10}$$

As a result, Eq. (9) reduces to $i_{\mathrm{GC}}(T_R, \tau) = i_{\mathrm{GC}}(T_R, \tau_0)$, which shows that, under the simple scaling hypothesis, the regional growth curve is the same for all durations. This implies that rainfall intensity quantiles can be derived at each location and for any duration τ as a function of $\overline{x}_{\tau_0}$, the scaling exponent η of Eq. (10), and the GEV parameters of the single growth curve:

$$i(T_R,\tau) = \overline{x}_{\tau_0}(\tau/\tau_0)^{-\eta} i_{\mathrm{GC}}(T_R,\tau_0). \tag{11}$$

This IDF model, labeled RegSc, requires two regional parameters of the GEV distribution modeling the single growth curve and two spatial maps of $\overline{x}_{\tau_0}$ and η (Table 1). The availability of additional daily rain gauges allows improving the accuracy of the spatial maps of $\overline{x}_{\tau_0}$.

c. Monte Carlo bootstrapping simulations

The IDF models are assessed as a function of different numbers of available high-resolution and daily rain gauges through cross validation with Monte Carlo simulations based on bootstrap resampling. The simulations consist of the following steps:

1) First $N_{\mathrm{val}} = 100$ high-resolution gauges are randomly selected from the $N = 223$ available and are used as independent sites for validation.
2) From the $(N - N_{\mathrm{val}})$ remaining gauges, N_τ stations are randomly selected as high-resolution gauges and used to calibrate all models. Performance metrics (see section 3d) are computed at the N_{val} validation gauges.
3) From the $(N - N_{\mathrm{val}} - N_\tau)$ remaining gauges, $N_{24\mathrm{h}}$ gauges are randomly extracted and used as additional daily gauges. To quantify the value of increasing information available at daily scale, $N_{24\mathrm{h}} = 1$ gauge is initially selected and, then, other gauges are recursively added until $N_{24\mathrm{h}} = 50$, resulting in the tested values $N_{24\mathrm{h}} = 1, 2, \ldots, 10, 20, \ldots, 50$. This implies that, for example, the gauges selected for $N_{24\mathrm{h}} = 30$ include the stations extracted for $N_{24\mathrm{h}} = 20$. Note that daily records are simply obtained by aggregating the high-resolution records.
4) For a given combination of N_τ and $N_{24\mathrm{h}}$ stations, the Sc and RegSc scaling models are calibrated using the N_τ high-resolution and $N_{24\mathrm{h}}$ daily stations and performance metrics at the calibration gauges are quantified.
5) Step 1 is repeated 10 times to characterize the sampling variability of the validation gauges. For each N_{val}, step 2 is repeated 10 times for eight values of N_τ ranging from 1 to 50 to characterize the sampling variability of the high-resolution gauges once the validation stations are fixed. For given N_{val} and N_τ, steps 3 and 4 are repeated 10 times to characterize the sampling variability of the daily gauges. Overall, $10 \times 10 \times 8 \times 10 = 8000$ Monte Carlo iterations are performed.

The application of the IDF models at ungauged sites (i.e., the N_{val} validation gauges) involves the spatial interpolation of the

TABLE 2. Minimum and maximum values across the 223 gauges of the local and bias-corrected (in parentheses) estimates of the GEV parameters used for the At-site and At-siteBC models, respectively. The ranges reported in italics for $\tau = 24$ h are for parameters k^*, $\mu_{\tau 0}$, and $\sigma_{\tau 0}$ of the Sc scaling model.

τ	k_τ	μ_τ (mm h^{-1})	σ_τ (mm h^{-1})
30 min	−0.38–0.40	14.7–45.6	7.2–19.1
	(−0.01–0.20)	(14.7–45.4)	(7.6–18.3)
1 h	−0.27–0.41	8.5–28.2	4.0–12.3
	(0.02–0.19)	(8.6–27.9)	(4.5–11.7)
2 h	−0.26–0.37	5.2–16.6	1.9–7.5
	(0.03–0.19)	(5.3–16.1)	(2.0–6.7)
3 h	−0.35–0.39	4.0–11.7	1.4–5.5
	(0.00–0.19)	(4.1–11.4)	(1.4–4.9)
6 h	−0.39–0.44	2.5–6.8	0.9–2.8
	(−0.03–0.21)	(2.6–6.7)	(0.9–2.6)
12 h	−0.32–0.39	1.5–4.5	0.5–2.0
	(0.00–0.20)	(1.5–4.4)	(0.5–1.7)
24 h	−0.36–0.33	0.9–2.6	0.3–1.3
	(0.00–0.18)	(0.9–2.6)	(0.3–1.3)
	(−0.44–0.44)	*(0.9–2.6)*	*(0.3–1.1)*

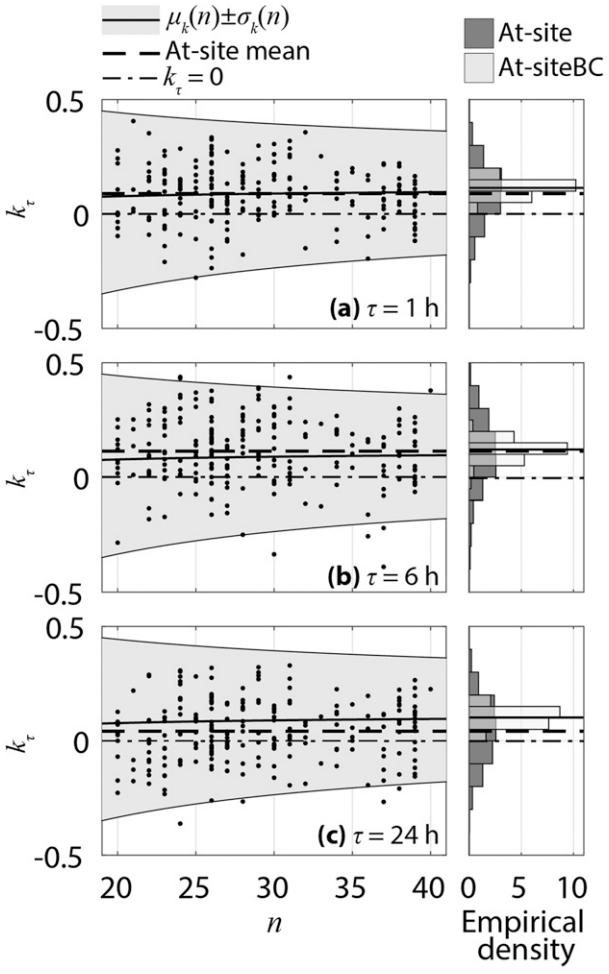

FIG. 3. Relation between shape parameter k_τ and sample size n for τ = 1, 6, and 24 h. The shaded area is the region of sampling variability for k_τ identified by Papalexiou and Koutsoyiannis (2013) as $\mu_k(n) \pm \sigma_k(n)$, with $\mu_k(n) = 0.114 - 0.69n^{-0.98}$ and $\sigma_k(n) = 0.045 + 1.27n^{-0.70}$. For each τ, the empirical densities are shown for k_τ estimated with PWM at each site and bias corrected as $k_{\tau,\mathrm{BC}}(n) = 0.114 + [0.045/\sigma_k(n)][k_\tau - \mu_k(n)]$, where $k_\tau(n)$ and $k_{\tau,\mathrm{BC}}(n)$ are the original and bias-corrected parameters, respectively (Papalexiou and Koutsoyiannis 2013). These values of k_τ are used in the At-site and At-siteBC IDF models.

model parameters (see Table 1 and section 3b), which can be achieved with several techniques (see, e.g., Watkins et al. 2005). Since evaluating the performance of the spatial interpolation method is out of the scope of this study, here we apply the inverse distance weighting interpolation method, which is widely known and easy to implement while being relatively accurate in our region (analyses not shown).

d. Model diagnostic

The statistical models are diagnosed through a set of metrics proposed by Hosking and Wallis (1997). For a given T_R, rainfall duration τ, and site j, the mth IDF model returns the quantile of annual maximum rainfall intensity $i_j^m(T_R, \tau)$.

TABLE 3. Parameters of the GEV growth curves used in the Reg IDF model and corresponding mean used for the single GEV growth curve in the RegSc IDF model.

τ	k_τ	μ_τ (-)	σ_τ (-)
30 min	0.04	0.77	0.38
1 h	0.08	0.75	0.38
2 h	0.10	0.76	0.35
3 h	0.11	0.77	0.33
6 h	0.12	0.78	0.32
12 h	0.06	0.79	0.32
24 h	0.04	0.79	0.34
Mean	0.08	0.77	0.34

The relative bias from a reference intensity $i_j^{\mathrm{REF}}(T_R, \tau)$ is defined as

$$\mathrm{RB}_j = \frac{i_j^m(T_R,\tau) - i_j^{\mathrm{REF}}(T_R,\tau)}{i_j^{\mathrm{REF}}(T_R,\tau)} \times 100. \quad (12)$$

To provide an overall picture of the model performance, the mean relative bias and relative root-mean-square error can be respectively computed as

$$\mathrm{RB} = \frac{1}{N}\sum_{j=1}^{N} \mathrm{RB}_j \quad \text{and} \quad (13)$$

$$\mathrm{RRMSE} = \left(\frac{1}{N}\sum_{j=1}^{N} \mathrm{RB}_j^2\right)^{1/2}, \quad (14)$$

where N is the total number of validation stations. Note that, for simplicity, the symbols m, τ, and T_R have been omitted in RB_j, RB, and RRMSE. Here, the performance metrics are computed using two references: (i) the quantiles of the empirical distributions for T_R up to 30 years and (ii) the quantiles of the At-site IDF model for T_R extending through 200 years. In such a way, we test the models' ability to simulate observed records and quantiles derived for large, nonobserved T_R, that are often needed for infrastructure design.

4. Results

a. Evaluation of the GEV hypothesis

Figures 2a–c show the L-moment ratio diagrams for records of annual rainfall maxima at τ = 1, 6, and 24 h, respectively, of the 223 rain gauges (results are similar at other durations). In all cases, the sample estimates are scattered around the GEV line, suggesting that this distribution appears to be best candidate to model annual rainfall maxima at all durations. This is corroborated by the facts that (i) stations with shorter (longer) records tend to be located farther from (closer to) the GEV line, and (ii) the sample means are located very close to or lie exactly on the GEV line. Results of the GOF tests confirm the suitability of the GEV distribution with at-site and bias-corrected parameters (see section 4b). For both estimation methods and all durations, the false discovery ratio test applied to all GOF tests indicates that the global null hypothesis (i.e., the sample is drawn from a GEV distribution) cannot be

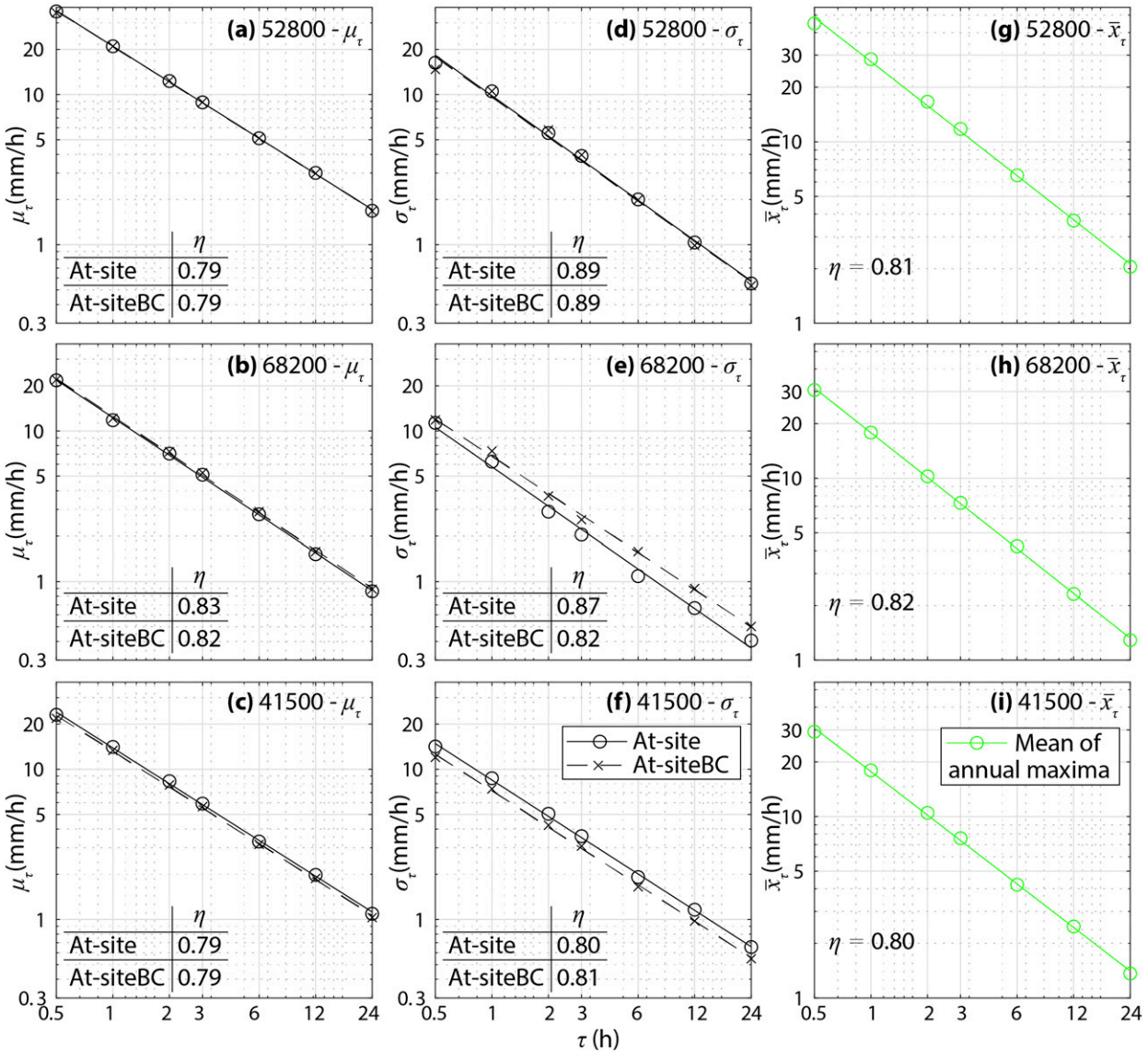

FIG. 4. Evidence of scaling for (a)–(c) GEV location (μ_τ) and (d)–(f) scale (σ_τ) parameters, and (g)–(i) for the mean of annual rainfall maxima ($\bar{x}_\tau$) in three representative gauges [identifiers (top) 52800, (middle) 68200, and (bottom) 41500]. The symbols are empirical estimates of the corresponding variable, and the line is the linear regression. For μ_τ and σ_τ, At-site and At-siteBC estimates are shown. The values of the scaling exponent η are also reported.

rejected at $\alpha_{\text{global}} = 0.05$. We also find that the regional GOF metric Z of Hosking and Wallis (1997) is always lower in absolute value than the suggested threshold of 1.64 ($Z \leq 0.81$ for $1 \leq \tau \leq 24$ h and $Z = 1.56$ for $\tau = 30$ min), thus indicating that the null hypothesis of a regional GEV cannot be rejected for any durations.

b. *Estimation of local and regional GEV parameters*

The local estimation of the GEV parameters conducted to apply the At-site and At-siteBC models is summarized in Table 2, which reports minimum and maximum parameter values for all τ (boxplots are also reported in Fig. S1 of the online supplemental material). Figure 3 also shows the relations between shape parameter k_τ and sample size for three representative durations. In all cases, the sample values of k_τ have mean larger than 0 and variability that is captured very well by the relation proposed by Papalexiou and Koutsoyiannis (2013), depicted by the shaded area. The bias correction of k_τ significantly reduces the variability of the shape parameter, as shown by the histograms and the values in Table 2.

To apply the Reg IDF model, we first compute the heterogeneity measure H^* and find this metric to range from 1.03 to 1.21 across the durations, suggesting that the hypothesis of a single operationally homogeneous region can be sustained. This is further supported by comparing visually the L-moment ratio diagrams of (i) the observed samples (Figs. 2a–c) and (ii)

synthetic samples generated by a kappa distribution fitted to the regional L-moments ratios (Figs. 2d–f; see appendix for details). The regional GEV parameters of the growth curves for the different durations are presented in Table 3, along with the means used for the single growth curve in the RegSc IDF model.

c. Evidence of simple scaling and estimation of GEV parameters for the scaling models

We find evidence of scaling in the range from $\tau = 30$ min to $\tau = 24$ h for μ_τ, σ_τ and mean of annual rainfall maxima $\bar{x}_\tau$. This is quantified by the coefficient of determination, R^2, of the linear regression between the log-transformed variables in Eqs. (8a) and (8b) and (10) being larger than 0.99 for μ_τ and $\bar{x}_\tau$ and 0.90 for σ_τ. As an example, Fig. 4 shows evidence of scaling for three representative gauges with L-moments ratios close to regional mean (identifier 52800) and in the upper-right (identifier 68200) and lower-left (identifier 41500) corners of the L-moments ratio diagram. The slope of the regression line provides an estimate of the scaling exponent η. The distributions of η computed for all gauges are presented as boxplots in Fig. 5. No evident differences exist between At-site and At-siteBC estimates. The η estimates for μ_τ are lower relative to those for σ_τ (mean η of 0.77 and 0.83 for μ_τ and σ_τ, respectively) and are characterized by less variability. While relatively small, these differences may affect performance of the simple scaling IDF models, which assume the same value of η for both location and scale parameters (see discussion in section 5). The distributions of η for $\bar{x}_\tau$ and μ_τ are very similar.

Based on the evidence of scaling, we estimate parameters of the Sc model at each gauge, including k^*, $\mu_{\tau 0}$, $\sigma_{\tau 0}$ ($\tau_0 = 24$ h), and the single scaling exponent, η. As reported in Table 2 (italic font), the ranges of $\mu_{\tau 0}$ and $\sigma_{\tau 0}$ are very close to those of the local estimates at the same duration. In contrast, the range of k^* spans the values of all local estimates of this parameter across all durations. This result is expected because the estimate of a constant shape parameter is affected by the variability of rain rates at all durations. Figure 5 shows that η for Sc displays the same intergauge variability of the η estimates for μ_τ and $\bar{x}_\tau$.

d. Performance of IDF models using all gauges

The IDF models are first applied using all $N = 223$ available gauges. Figure 6 presents examples of the comparison between the GEV parameterized with the five models and empirical CDFs for the same three gauges shown in Fig. 4. Results are shown for $\tau = 1$, 6, and 24 h. At gauge 52800 (Figs. 6a–c), whose L-moments are close to the regional means at all durations, all models capture well the observed records, apart from a slight underestimation of the right tail by the scaling models (Sc and Reg1Sc) for $\tau = 1$ h; that is, the lines are located to the left of the circles for larger values of the rainfall rate. At gauge 68200 (Figs. 6d–f), which is characterized by higher L-skewness and L-kurtosis than the regional means, all models capture well the bulk of the empirical distributions, but the regional models (Reg and RegSc) underestimate the right tails at all durations. At gauge 68200 (Figs. 6g–i), whose L-skewness and L-kurtosis are lower than the regional means, all models reproduce

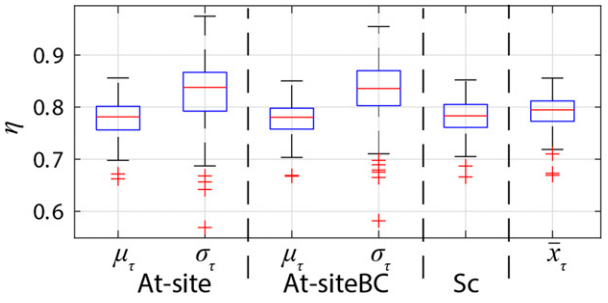

FIG. 5. Boxplots of η computed for (separated by the vertical dashed lines) At-site and At-siteBC estimates of μ_τ and σ_τ, the Sc model, and mean of annual rainfall maxima $\bar{x}_\tau$.

reasonably well all empirical distributions. Performance metrics computed on all gauges are summarized in Fig. 7 through heat maps showing RD and RRMSE for each model as a function of durations and return periods. Performances are relatively similar across all models when compared to the observed quantiles up to $T_R = 30$ years. When At-site quantiles are used as reference, the At-siteBC and Reg models exhibit the best performance with negligible (slightly positive) RB but higher (smaller) RRMSE for the Reg (At-siteBC) model. The scaling models (Sc and RegSc) have a larger range of RB with both positive and negative values, and RRMSE up to ~35% for $T_R = 200$ years. This lower accuracy using all available gauges is expected since these models depend on a lower number of parameters. These results provide an overview of model accuracy by quantifying their performance at the same sites used for model calibration. In the Monte Carlo bootstrapping experiments described next, model accuracy is tested using 100 randomly selected validation sites and varying the number of calibration gauges.

e. Performance of IDF models as a function of the number of gauges

For the sake of conciseness, we present results of the Monte Carlo simulations for selected values of τ, T_R, N_τ, and N_{24h} that are representative of the overall outcomes. In the following figures, the simulation uncertainty is visualized through the mean and 90% confidence intervals of the metrics and by reporting results for three randomly chosen simulations. We begin by showing in Fig. 8 how performances change with T_R for $\tau = 30$ min, $N_\tau = 5$ high-resolution gauges and no additional daily gauges. When compared to the observed quantiles (Fig. 8a), all models display similar results, with RB decreasing (i.e., the bias becomes more negative) and RRMSE increasing as T_R rises. Performances are worse for the scaling models that have more negative RB and larger uncertainty, that is, the width of the 90% confidence interval is comparatively larger. It is worth noting that the At-site and At-siteBC models have very similar accuracy, implying that bias correcting the shape parameter does not significantly impact the ability of simulating the observed quantiles for $T_R \leq 30$ years. Performances for larger T_R are explored using the At-site model as reference for the metrics (Fig. 8b). RRMSE increases significantly for $T_R >$ 50 years, with similar values across all models. The At-site and

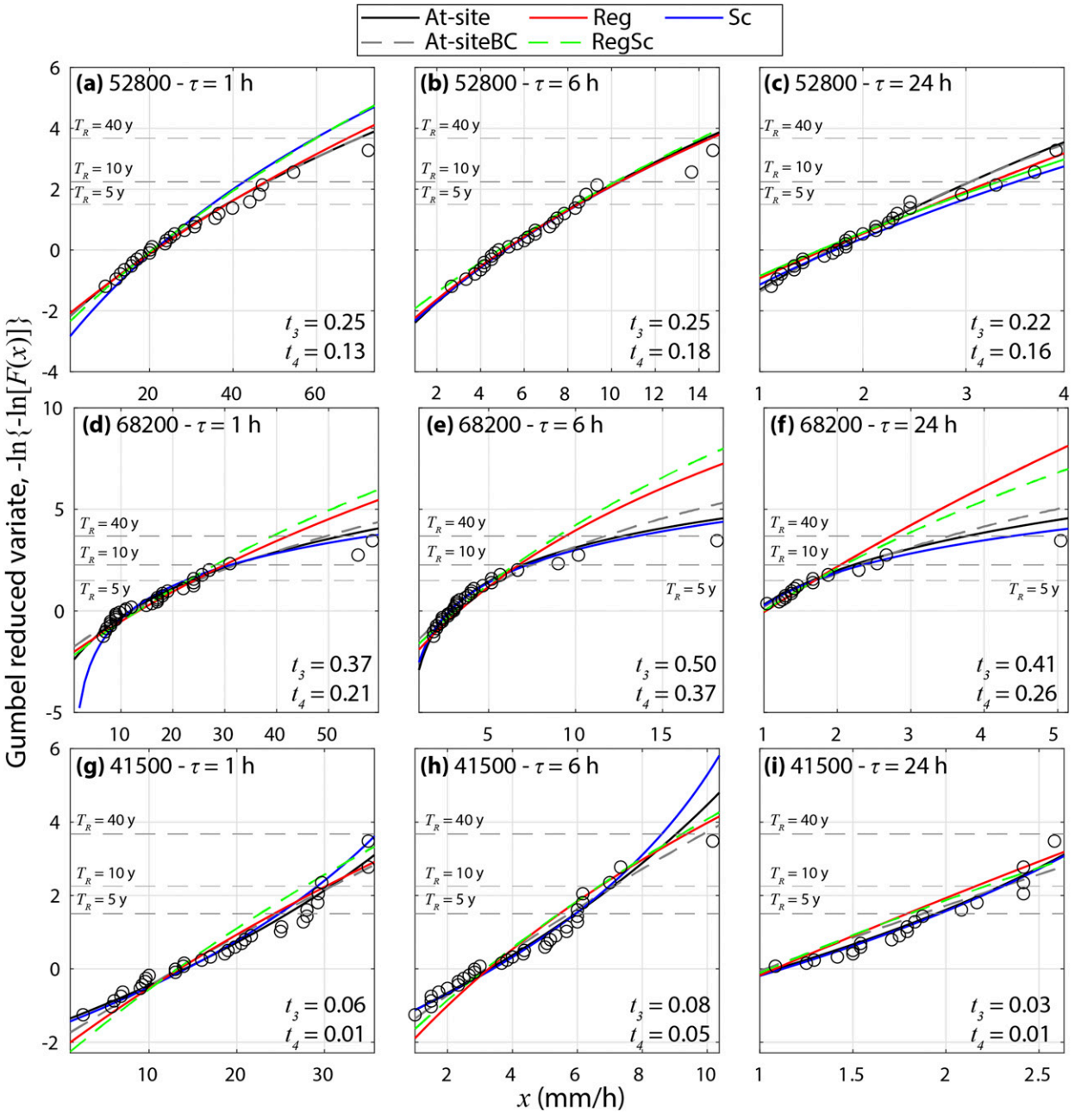

FIG. 6. Empirical CDFs compared with the GEV distribution parameterized with the IDF models for τ = (left) 1, (center) 6, and (right) 24 h for three representative gauges [identifiers (a)–(c) 52800, (d)–(f) 68200, and (g)–(i) 41500]. Frequencies are plotted as Gumbel reduced variate.

Reg models are, on average, unbiased. For the At-siteBC model, RB increases with T_R because bias correcting the shape parameter makes the tail of the GEV distributions heavier leading to larger quantiles at higher T_R. This also results in lower uncertainty. The scaling models exhibit instead negative mean RB for $T_R > 50$ years (more significantly, for the Sc model).

As a next step, we investigate how performances vary with τ, reporting results in Figs. 9a and 9b for $N_\tau = 5$ gauges, no additional daily gauges, and $T_R = 10$ and 100 years for the observed and At-site reference quantiles, respectively. When compared with the empirical quantiles, performances are similar for At-site, At-siteBC, and Reg models, with small variations of the metrics across the durations except for a slightly negative RB for $\tau < 12$ h. The scaling models (Sc and RegSc) exhibit instead significant negative biases for $\tau < 12$ h; their performances degrade particularly at $\tau = 1$ h where the mean RB and RRMSE reach about −15% and 22%, respectively. Interestingly, for all models the simulation uncertainty increases

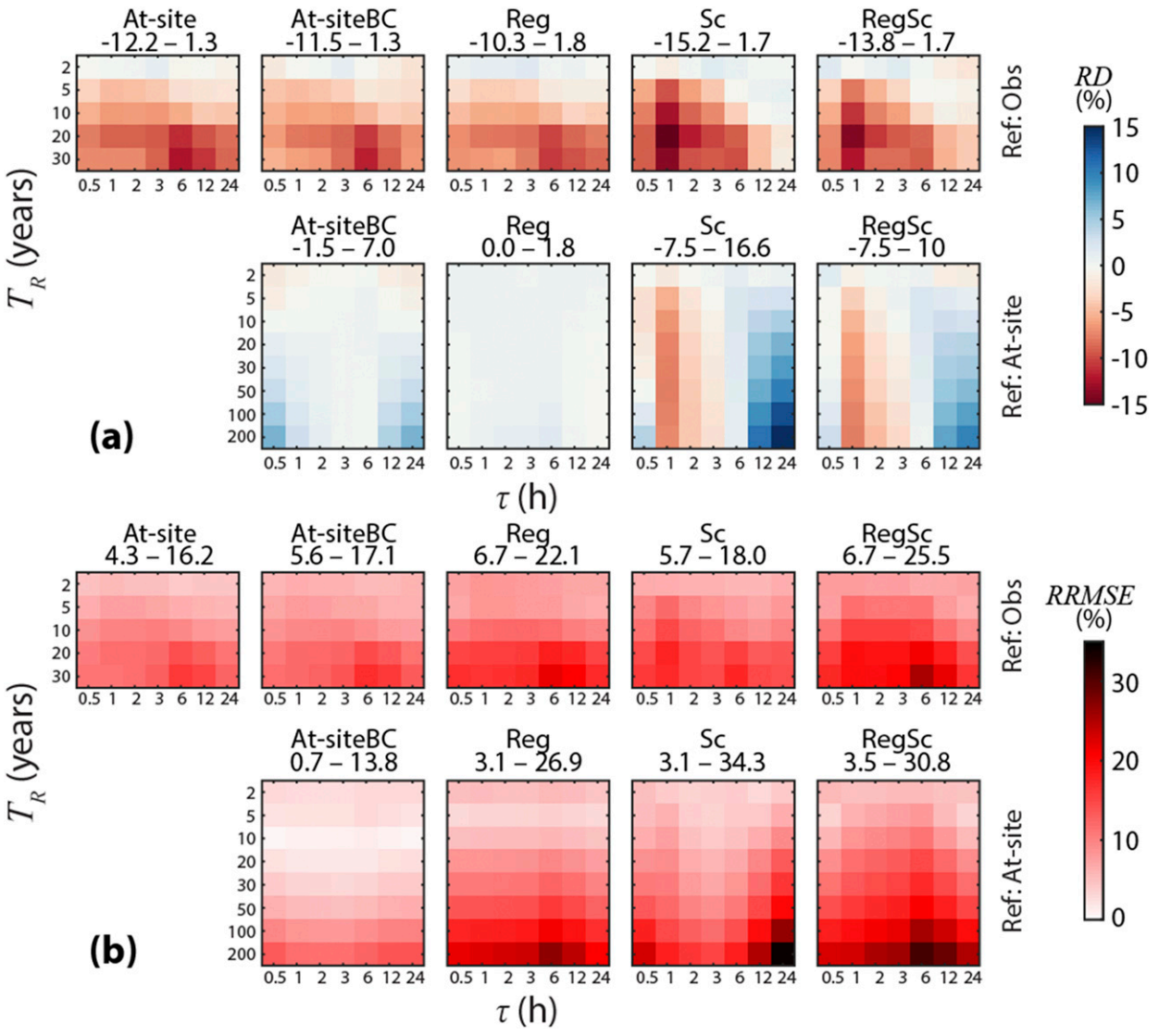

FIG. 7. Heat maps summarizing (a) RD and (b) RRMSE computed for each model using all gauges as a function of T_R and τ. For each of (a) and (b), the heat maps in the top row show the values calculated using observed quantiles as reference and those in the bottom row show the values calculated using At-site quantiles as reference. For each case, the minimum and maximum values of the performance metrics (in percent) computed across all T_R and τ are reported.

with τ. These findings are largely in line with results for the case of T_R = 100 years with At-site quantiles used as reference (Fig. 9b), with the exceptions that (i) the At-site and At-siteBC models are positively biased, (ii) the Reg model is unbiased, and (iii) for all models, RRMSE for τ = 6 h is the largest.

The effect of the number of available gauges on model performances is explored in Figs. 10–12 . We first evaluate how accuracy changes with N_τ when no additional daily gauges are used, showing in Figs. 10a and 10b results for τ = 30 min and T_R = 10 (100 years) for the observed (At-site) reference quantiles. In all cases, (i) the uncertainty across the simulations is very large for N_τ = 1 gauge and decreases dramatically up to N_τ = 5 and at a lower rate for $N_\tau > 5$, remaining constant after N_τ = 50 gauges (not shown); (ii) the average RB is constant across N_τ; and (iii) the mean RRMSE decreases as N_τ grows, reducing its value of ~30% when N_τ increases from 1 to 50. We subsequently investigate the value of additional information provided by daily rain gauges, considering the case of N_τ = 1 and N_{24h} = 10 that mimics the plausible situation of a single high-resolution gauge available (e.g., at the closest airport) plus additional daily gauges. Figure 11 displays the relations between metrics and T_R for At-Site and scaling models with N_{24h} = 0 and 10, using the At-Site quantiles as reference. It is apparent that adding ten daily gauges to the calibration of the scaling models results in a dramatic reduction of uncertainty, as quantified by the width of the 90% confidence interval decreasing by ~50%, and RRMSE, with reductions ranging

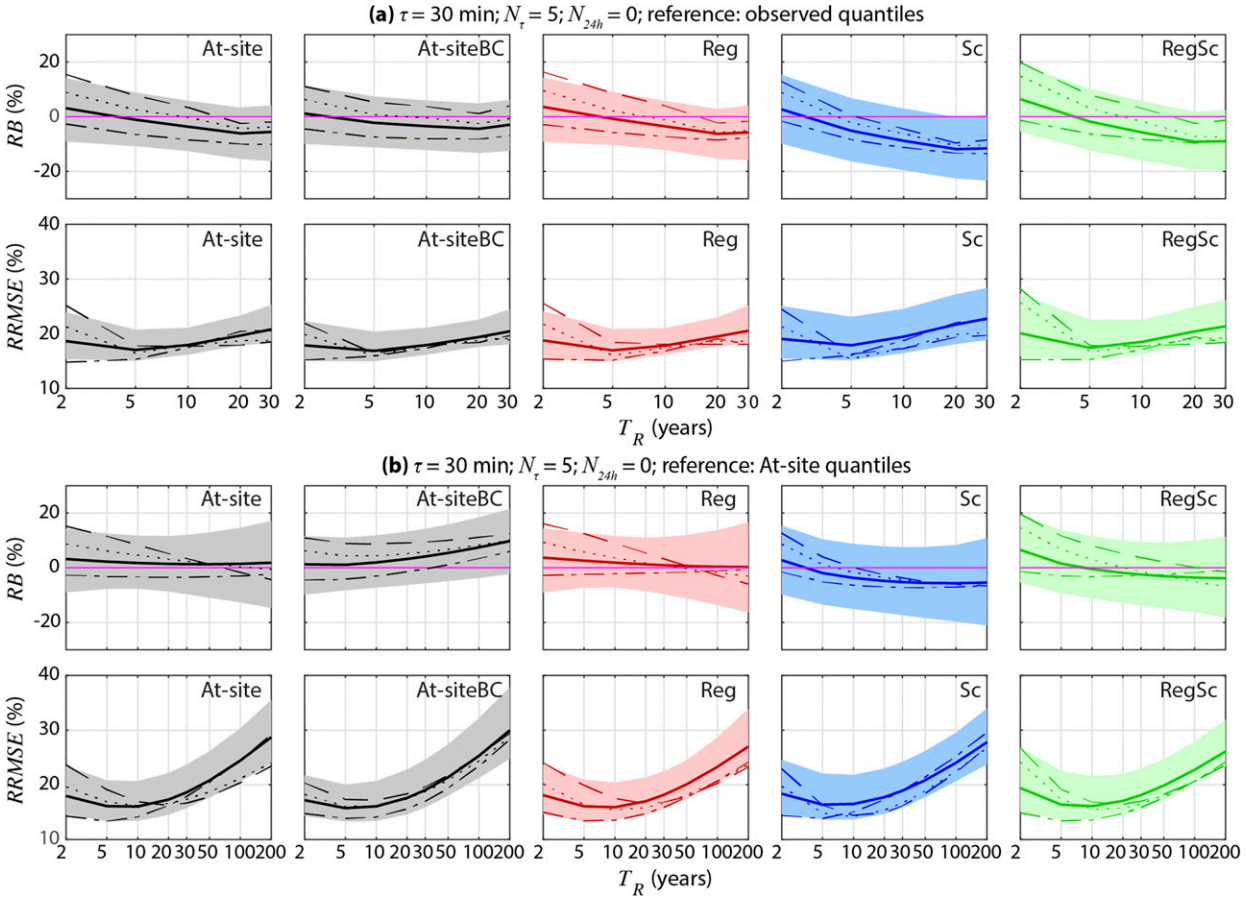

FIG. 8. Performance metrics of the IDF models as a function of T_R for $\tau = 30$ min, $N_\tau = 5$ high-resolution gauges, and $N_{24h} = 0$ additional daily gauges using (a) observed and (b) At-site quantiles as reference. The solid line is the metric mean across the Monte Carlo simulations, and the shaded area is the 90% confidence interval. Three randomly chosen simulations are also shown with dashed, dotted, and dot–dashed lines.

from −7.6% for $T_R = 10$ years to −22% for $T_R = 200$ years. To summarize results of the simulations, Fig. 12 presents heat maps of mean and width of the 90% confidence intervals of RB and RRMSE for all models as a function of N_τ and N_{24h}, for $\tau = 30$ min and $T_R = 100$ years. The key messages are (i) when $N_{24h} = 0$, results do not change significantly for $N_\tau \geq 10$; and (ii) the largest improvement introduced by the scaling models is for $N_\tau \leq 5$ and $N_{24h} \geq 10$.

5. Discussion

We first interpret our findings in the context of the rainfall climatology of the study region. Following previous studies indicating that elevation exerts a moderate to significant control on rainfall processes and statistical properties in this area (Mascaro 2017, 2018), we investigate the relations between IDF model parameters and gauge elevation, z. Figures 13a–c show these relations for the GEV parameters of the At-site and Sc models for $\tau = 24$ h, chosen as an example, while Fig. 13d presents the corresponding correlation coefficient (CC) as a function of τ (results are similar for the other models). The spatial maps of the parameters for two representative durations are also reported in Fig. S2 of the online supplemental material. The effect of elevation is important for μ_τ ($0.67 \leq CC \leq 0.80$) and, to a lesser extent, σ_τ ($0.26 \leq CC \leq 0.65$); for both parameters, the p values of the linear regressions are always lower than 0.01. The shape parameter k_τ does not exhibit instead significant orographic or geographic control. This is also true for the scaling exponent η as displayed in Fig. 13e and quantified by the high p values of the linear regressions with z (not shown). These findings suggest that terrain affects mainly the mean of the distributions of annual rainfall maxima, but it does not induce important effects on the shape of the distributions (i.e., the right tail) and the rainfall scaling properties. This result is consistent with Blanchet et al. (2016) for location and scale parameters, while it differs for shape parameter and scaling exponent that, in their study region, are affected by topography and distance from the coast.

The different elevation controls on GEV parameters and scaling exponent are a main factor impacting IDF model performances. In general, since all IDF models require estimates of μ_τ and σ_τ that are affected by terrain, the ability to reproduce extreme rainfall in this region depends on the elevation range of the available high-resolution gauges. In particular,

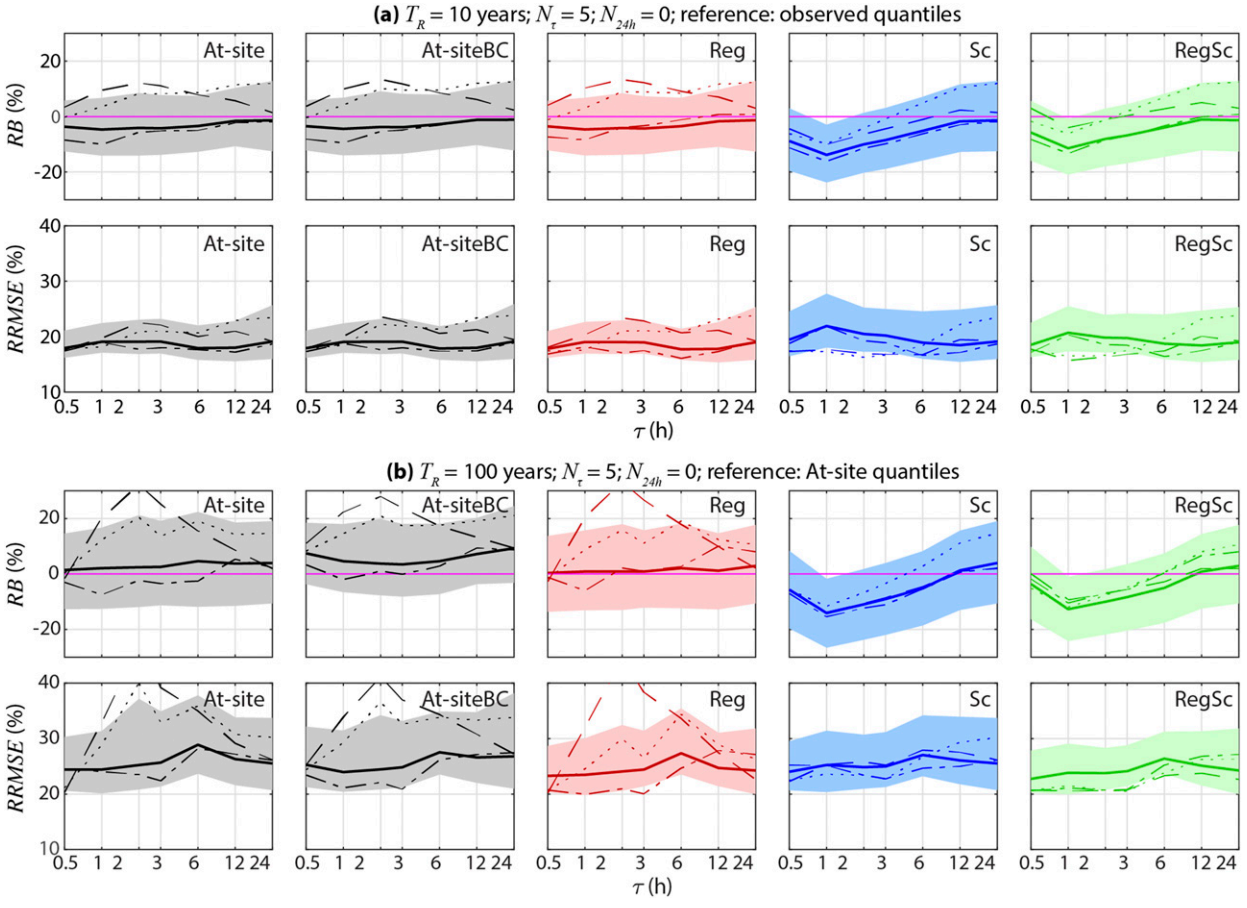

FIG. 9. As in Fig. 8, but for the performance metrics plotted as a function of τ for $N_\tau = 5$, $N_{24h} = 0$, and $T_R = 10$ or 100 years respectively using observed or At-site reference quantiles.

At-site, At-siteBC and Reg models use information on parameter variability based on gauge elevation to estimate quantiles at each duration. Scaling models derive instead quantiles at the different durations through the scaling exponent η, which exhibits spatial random variations. As a result, when a sufficient number of high-resolution gauges is available to capture the orographic control on μ_τ and σ_τ (e.g., $N_\tau \geq 10$), At-site, At-siteBC and Reg models overperform scaling models because, for these, that number of gauges does not allow describing in detail the spatial variability of η. An additional factor conditioning the accuracy of simple scaling models is the adoption of a single scaling exponent, which may not hold at all gauges, as shown in Fig. 5.

6. Conclusions

The comparison of local, regional, and simple scaling models for rainfall IDF analysis conducted in this study using 223 rain gauges in central Arizona makes it possible to draw the following recommendations supporting the generation of IDF curves in other regions:

1) Performances of IDF models are impacted by the control of rainfall regime and local geographic features on the spatial variability of model parameters. In our study region, elevation controls the mean of the annual rainfall maxima distributions, which is related to the GEV scale and location parameters. The shape parameter and scaling exponent do not exhibit instead clear spatial patterns. These features lead to distinct impacts on model performance that may vary at other sites with different rainfall regimes. Future work based on synthetic experiments should investigate the relative importance of each model parameter under different scenarios where the corresponding spatial variability is assumed to be random or linked to geographic features (e.g., latitude, longitude, elevation, and distance from the coast).
2) As expected, model uncertainty and RRMSE decrease with the number of high-resolution gauges N_τ used for calibration. This reduction is dramatic from $N_\tau = 1$ to $N_\tau = 5$, less significant for $N_\tau > 5$, and negligible after $N_\tau = 50$. The mean bias is mostly constant with N_τ.
3) All models exhibit similar performances in the ability to simulate observed quantiles of relatively short records ($\leq$30 years), apart from the scaling models that display slightly negative biases at shorter durations.
4) If N_τ is large enough to allow capturing with sufficient details the spatial variability of the GEV scale and location

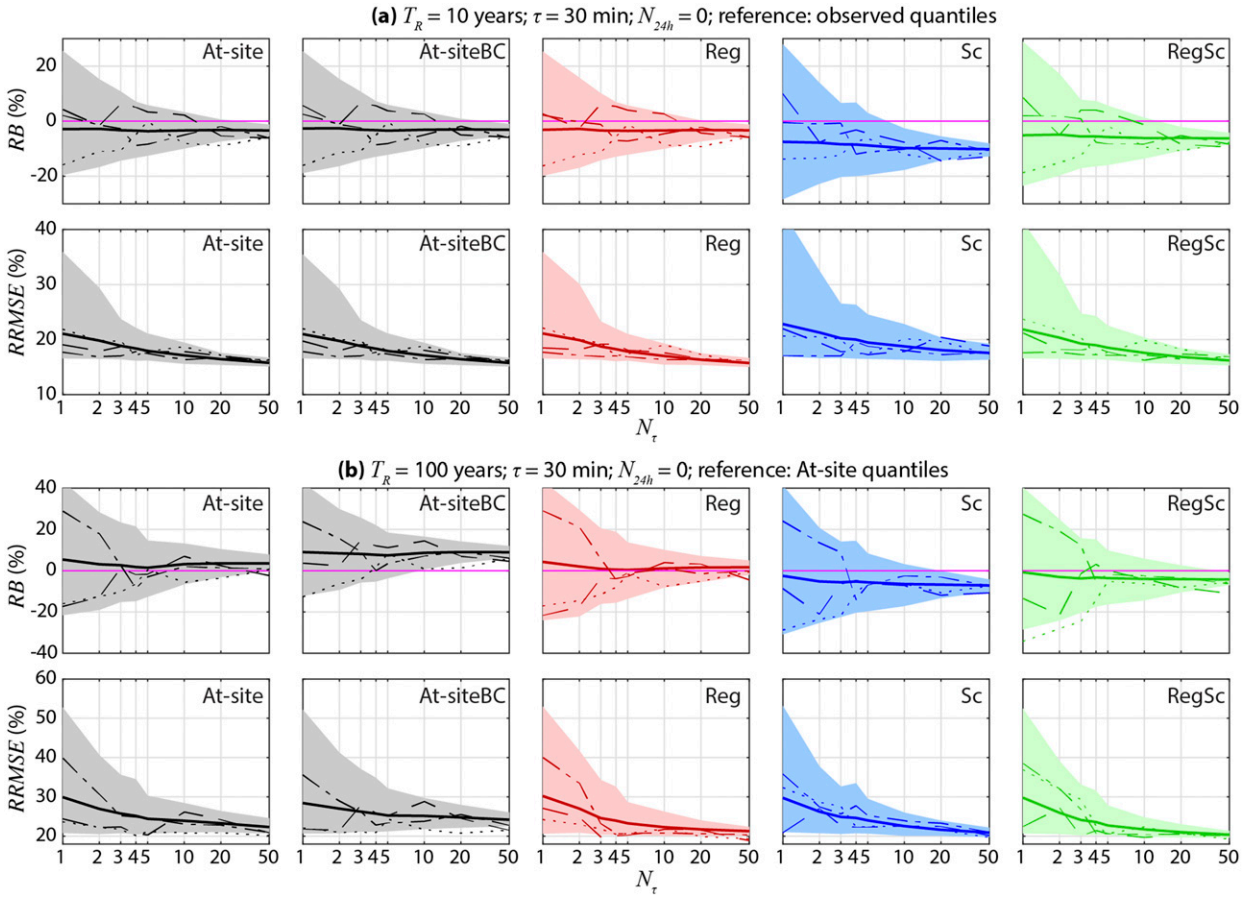

FIG. 10. As in Fig. 8, but for the performance metrics plotted as a function of N_τ for τ = 30 min, N_{24h} = 0, and T_R = 10 or 100 years respectively using observed or At-site reference quantiles.

parameters at different durations, the At-site, At-siteBC and Reg models have the best performances when compared to quantiles of the GEV distributions estimated locally (the At-site quantiles). In our study region, this occurs when $N_\tau \geq 10$. The Reg and At-site model are unbiased and slightly positively biased, respectively. The bias correction of the shape parameter based on global rainfall records performed in the At-siteBC model leads to

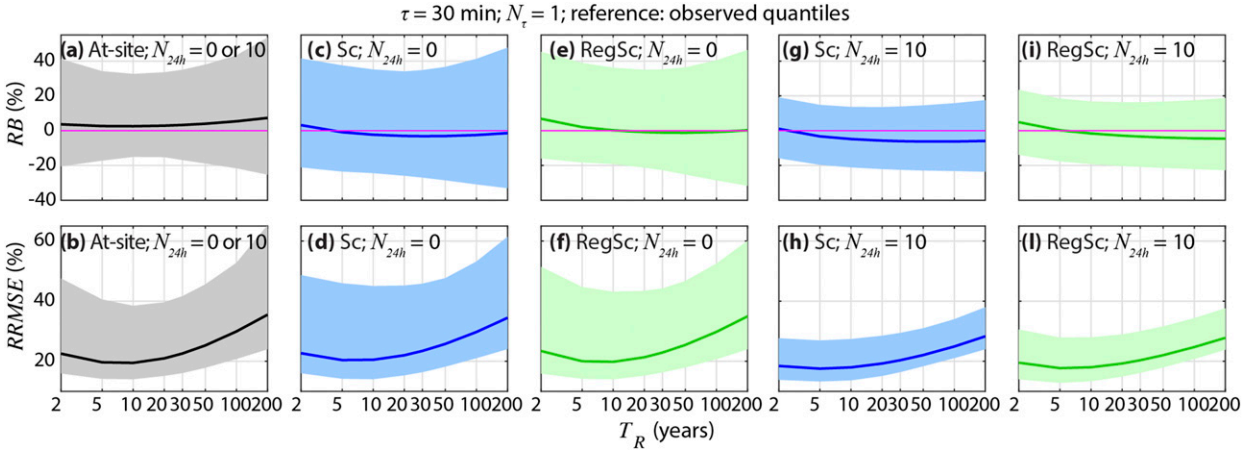

FIG. 11. Mean and 90% confidence interval of performance metrics of the At-Site, Sc, and RegSc IDF models as a function of T_R for τ = 30 min, N_τ = 1 high-resolution gauge, and N_{24h} = 0 and 10 additional daily gauges using At-site quantiles as reference.

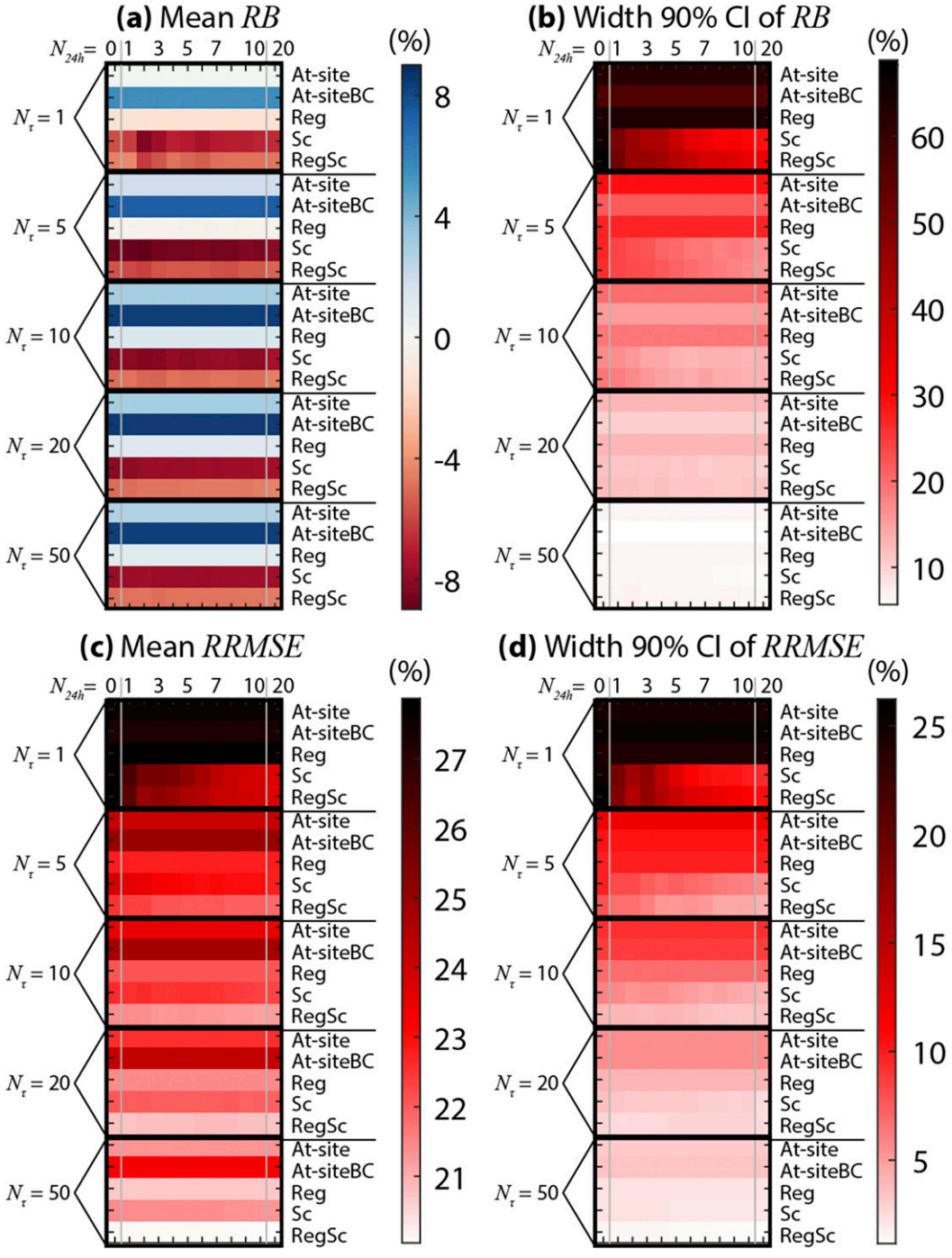

FIG. 12. Heat maps showing (left) mean and (right) width of the 90% confidence interval (CI) of (a), (b) RB and (c), (d) RRMSE for all models as a function of N_τ (rows) and N_{24h} (columns) for $\tau = 30$ min, $T_R = 100$ years, and At-site quantiles used as reference. The vertical gray lines separate the regions with $N_{24h} = 0$, $1 \leq N_{24h} \leq 10$, and $N_{24h} = 20$.

positive biases at larger T_R. Since these return periods are used for design of critical infrastructure, the use of the At-siteBC model is recommended. Future work should focus on (i) investigating whether the relations for correcting the bias of shape parameter estimates change for durations below 24 h and (ii) incorporating the shape parameter bias correction in regionalization techniques.

5) The accuracy of scaling models is affected by two main factors, including the ability to capture the spatial variability of the scaling exponent, and the possibility that the

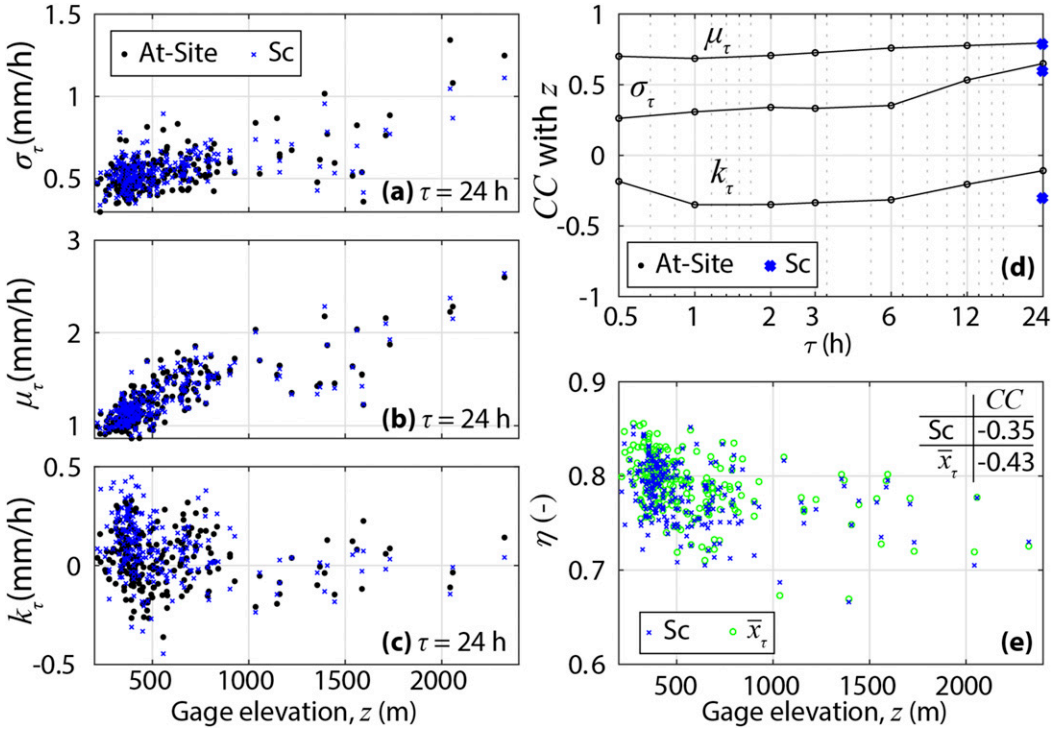

FIG. 13. (a)–(c) Relation between GEV parameters and gauge elevation z for $\tau = 24$ h and At-Site and Sc models, (d) CC between GEV parameters and z as a function of τ for At-Site and Sc (available only for $\tau = 24$ h) models, and (e) relation between scaling exponent and z for Sc model and mean of annual rainfall maxima $\bar{x}_\tau$ used in RegSc.

assumption of a single scaling exponent may not hold. This may require the use of more complicated multiple scaling models (Burlando and Rosso 1996; Van de Vyver 2018). In our study site, simple scaling IDF models overperform local and regional models when $N_\tau \leq 5$ and more than 10 additional daily gauges are available. These conditions can be very common in developing countries.

Acknowledgments. The author thanks Dr. Juliette Blanchet, two anonymous reviewers, and the peer-review editor for their constructive comments that helped to improve the quality of the paper. This work has been supported by the National Science Foundation (NSF) Award "SCC: Community-Based Automated Information for Urban Flooding" (Award 1831475). The support by NSF (Grant EAR-1928724) and NASA (Grant 80NSSC19K0726) to organize the 12th International Precipitation Conference (IPC12), Irvine, California, June 2019, and produce the IPC12 special collection of papers is also gratefully acknowledged. The author thanks Stephen D. Waters from the Flood Control District of Maricopa County for providing the rainfall data of the network. FCDMC data are also available online (https://www.fcd.maricopa.gov/625/Rainfall-Data).

APPENDIX

Heterogeneity Measure

Following Hosking and Wallis (1997), let t^R, t_3^R, and t_4^R be the regional L-coefficient of variation (L-CV), L-skewness, and L-kurtosis, computed by averaging the sampling L-moments ratios of records of N gauges with weights given by the sample size. Let also V be the weighted standard deviation of the sample L-CVs (labeled V_{sam}). The computation of the heterogeneity measure requires the following steps:

1) The kappa distribution is fitted to the regional L-moment ratios. This four-parameter distribution is chosen for its ability to represent many distributions used for rainfall extremes. In such a way, no commitment is made on the underlying distribution.
2) Then $N_{\text{ens}} = 2000$ statistical simulations are conducted. In each simulation, N variates are generated from the kappa distribution using the same record lengths of the original gauges. Figures 2d–f show examples of scatterplots of L-skewness versus L-kurtosis for one simulation of kappa covariates for three durations.
3) Calculate V for each simulation, and from these 2000 values compute the mean, μ_V.
4) Compute the heterogeneity measure $H^* = V_{\text{sam}}/\mu_V$, which allows detection of operationally homogeneous regions when a large number of gauges is used, as in this case where $N = 223$.

Values of H^* close to 1 indicate that the sample L-CVs exhibit similar variability to those generated by a single homogenous region.

REFERENCES

Bara, M., L. Gaál, S. Kohnová, J. Szolgay, and K. Hlavčová, 2010: On the use of the simple scaling of heavy rainfall in a regional estimation of IDF curves in Slovakia. *J. Hydrol. Hydromech.*, **58**, 49–63, https://doi.org/10.2478/v10098-010-0006-0.

Blanchet, J., D. Ceresetti, G. Molinié, and J. D. Creutin, 2016: A regional GEV scale-invariant framework for intensity–duration–frequency analysis. *J. Hydrol.*, **540**, 82–95, https://doi.org/10.1016/j.jhydrol.2016.06.007.

Bonnin, G. M., D. Martin, B. Lin, T. Parzybok, M. Yekta, and D. Riley, 2004: *Semiarid Southwest (Arizona, Southeast California, Nevada, New Mexico, Utah).* Vol. 1, *Precipitation-Frequency Atlas of the United States*, NOAA Atlas 14, 271 pp., https://www.nws.noaa.gov/oh/hdsc/PF_documents/Atlas14_Volume1.pdf.

Borga, M., C. Vezzani, and G. Dall Fontana, 2005: Regional rainfall depth-duration-frequency equations for an alpine region. *Nat. Hazards*, **36**, 221–235, https://doi.org/10.1007/s11069-004-4550-y.

Burlando, P., and R. Rosso, 1996: Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *J. Hydrol.*, **187**, 45–64, https://doi.org/10.1016/S0022-1694(96)03086-7.

Cohen, B., 2006: Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. *Technol. Soc.*, **28**, 63–80, https://doi.org/10.1016/j.techsoc.2005.10.005.

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer, 208 pp.

Courty, L. G., R. L. Wilby, J. K. Hillier, and L. J. Slater, 2019: Intensity-duration-frequency curves at the global scale. *Environ. Res. Lett.*, **14**, 084045, https://doi.org/10.1088/1748-9326/ab370a.

Deidda, R., and M. Puliga, 2006: Sensitivity of goodness-of-fit statistics to rainfall data rounding off. *Phys. Chem. Earth*, **31**, 1240–1251, https://doi.org/10.1016/j.pce.2006.04.041.

——, R. Benzi, and F. Siccardi, 1999: Multifractal modeling of anomalous scaling laws in rainfall. *Water Resour. Res.*, **35**, 1853–1867, https://doi.org/10.1029/1999WR900036.

Fowler, H. J., and C. G. Kilsby, 2003: A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *Int. J. Climatol.*, **23**, 1313–1334, https://doi.org/10.1002/joc.943.

García-Marín, A. P., R. Morbidelli, C. Saltalippi, M. Cifrodelli, J. Estévez, and A. Flammini, 2019: On the choice of the optimal frequency analysis of annual extreme rainfall by multifractal approach. *J. Hydrol.*, **575**, 1267–1279, https://doi.org/10.1016/j.jhydrol.2019.06.013.

Hosking, J. R. M., and J. R. Wallis, 1997: *Regional Frequency Analysis*. Cambridge University Press, 242 pp.

Innocenti, S., A. Mailhot, and A. Frigon, 2017: Simple scaling of extreme precipitation in North America. *Hydrol. Earth Syst. Sci.*, **21**, 5823–5846, https://doi.org/10.5194/hess-21-5823-2017.

Koutsoyiannis, D., and E. Foufoula-Georgiou, 1993: A scaling model of a storm hyetograph. *Water Resour. Res.*, **29**, 2345–2361, https://doi.org/10.1029/93WR00395.

——, D. Kozonis, and A. Manetas, 1998: A mathematical framework for studying rainfall intensity-duration-frequency relationships. *J. Hydrol.*, **206**, 118–135, https://doi.org/10.1016/S0022-1694(98)00097-3.

Liew, S. C., S. V. Raghavan, and S. Y. Liong, 2014: Development of intensity-duration-frequency curves at ungauged sites: Risk management under changing climate. *Geosci. Lett.*, **1**, 8, https://doi.org/10.1186/2196-4092-1-8.

Liu, J., C. D. Doan, S. Y. Liong, R. Sanders, A. T. Dao, and T. Fewtrell, 2015: Regional frequency analysis of extreme rainfall events in Jakarta. *Nat. Hazards*, **75**, 1075–1104, https://doi.org/10.1007/s11069-014-1363-5.

Lovejoy, S., and D. Schertzer, 1985: Generalized scale invariance in the atmosphere and fractal models of rain. *Water Resour. Res.*, **21**, 1233–1250, https://doi.org/10.1029/WR021i008p01233.

Lumbroso, D. M., S. Boyce, H. Bast, and N. Walmsley, 2011: The challenges of developing rainfall intensity-duration-frequency curves and national flood hazard maps for the Caribbean. *J. Flood Risk Manage.*, **4**, 42–52, https://doi.org/10.1111/j.1753-318X.2010.01088.x.

Madsen, H., P. S. Mikkelsen, D. Rosbjerg, and P. Harremoës, 2002: Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series statistics. *Water Resour. Res.*, **38**, 1239, https://doi.org/10.1029/2001wr001125.

Marra, F., and E. Morin, 2015: Use of radar QPE for the derivation of intensity-duration-frequency curves in a range of climatic regimes. *J. Hydrol.*, **531**, 427–440, https://doi.org/10.1016/j.jhydrol.2015.08.064.

Mascaro, G., 2017: Multiscale spatial and temporal statistical properties of rainfall in central Arizona. *J. Hydrometeor.*, **18**, 227–245, https://doi.org/10.1175/JHM-D-16-0167.1.

——, 2018: On the distributions of annual and seasonal daily rainfall extremes in central Arizona and their spatial variability. *J. Hydrol.*, **559**, 266–281, https://doi.org/10.1016/j.jhydrol.2018.02.011.

——, R. Deidda, and M. Hellies, 2013: On the nature of rainfall intermittency as revealed by different metrics and sampling approaches. *Hydrol. Earth Syst. Sci.*, **17**, 355–369, https://doi.org/10.5194/hess-17-355-2013.

——, E. R. Vivoni, D. J. Gochis, C. J. Watts, and J. C. Rodríguez, 2014: Temporal downscaling and statistical analysis of rainfall across a topographic transect in northwest Mexico. *J. Appl. Meteor. Climatol.*, **53**, 910–927, https://doi.org/10.1175/JAMC-D-13-0330.1.

Menabde, M., A. Seed, and G. Pegram, 1999: A simple scaling model for extreme rainfall. *Water Resour. Res.*, **35**, 335–339, https://doi.org/10.1029/1998WR900012.

National Academies of Sciences, Engineering, and Medicine, 2019: *Framing the Challenge of Urban Flooding in the United States*. National Academies Press, 100 pp., https://doi.org/10.17226/25381.

NWS, 2019: Weather related fatality and injury statistics. NOAA, https://www.weather.gov/hazstat/.

Ombadi, M., P. Nguyen, S. Sorooshian, and K. Hsu, 2018: Developing intensity-duration-frequency (IDF) curves from satellite-based precipitation: Methodology and evaluation. *Water Resour. Res.*, **54**, 7752–7766, https://doi.org/10.1029/2018WR022929.

Papalexiou, S. M., and D. Koutsoyiannis, 2013: Battle of extreme value distributions : A global survey on extreme daily rainfall. *Water Resour. Res.*, **49**, 187–201, https://doi.org/10.1029/2012WR012557.

Requena, A. I., D. H. Burn, and P. Coulibaly, 2019: Pooled frequency analysis for intensity–duration–frequency curve estimation. *Hydrol. Processes*, **33**, 2080–2094, https://doi.org/10.1002/hyp.13456.

Sane, Y., and Coauthors, 2018: Intensity-duration-frequency (IDF) rainfall curves in Senegal. *Hazards Earth Syst. Sci*, **18**, 1849–1866, https://doi.org/10.5194/nhess-18-1849-2018.

Tyralis, H., and A. Langousis, 2019: Estimation of intensity–duration–frequency curves using max-stable processes. *Stochastic Environ. Res. Risk Assess.*, **33**, 239–252, https://doi.org/10.1007/s00477-018-1577-2.

University of Maryland College Park and Texas A&M University, 2018: The growing threat of urban flooding: A national challenge—2018. TAMU Rep., 44 pp., https://today.tamu.edu/wp-content/uploads/sites/4/2018/11/Urban-flooding-report-online.pdf.

Van de Vyver, H., 2018: A multiscaling-based intensity-duration-frequency model for extreme precipitation. *Hydrol. Processes*, **32**, 1635–1647, https://doi.org/10.1002/hyp.11516.

Venugopal, V., E. Foufoula-Georgiou, and V. Sapozhnikov, 1999: A space-time downscaling model for rainfall. *J. Geophys. Res.*, **104**, 19 705–19 721, https://doi.org/10.1029/1999JD900338.

Watkins, D. W., G. A. Link, and D. Johnson, 2005: Mapping regional precipitation intensity duration frequency estimates. *J. Amer. Water Resour. Assoc.*, **41**, 157–170, https://doi.org/10.1111/j.1752-1688.2005.tb03725.x.

Wilks, D. S., 2006: On ''field significance'' and the false discovery rate. *J. Appl. Meteor. Climatol.*, **45**, 1181–1189, https://doi.org/10.1175/JAM2404.1.

——, 2016: ''The stippling shows statistically significant grid points'': How research results are routinely overstated and overinterpreted, and what to do about it. *Bull. Amer. Meteor. Soc.*, **97**, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1.

Yu, P. S., T. C. Yang, and C. S. Lin, 2004: Regional rainfall intensity formulas based on scaling property of rainfall. *J. Hydrol.*, **295**, 108–123, https://doi.org/10.1016/j.jhydrol.2004.03.003.

Zhang, W., G. Villarini, G. A. Vecchi, and J. A. Smith, 2018: Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston. *Nature*, **563**, 384–388, https://doi.org/10.1038/s41586-018-0676-z.