

Research papers

Improving the utility of weather radar for the spatial frequency analysis of extreme precipitation

Nehal Ansh Srivastava, Giuseppe Mascaro^{*}

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



ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Uwe Haberlandt, Associate Editor

Keywords:

Extreme precipitation
Intensity–duration–frequency curves
Weather radar
Regional frequency analysis

ABSTRACT

Spatially seamless quantitative precipitation estimates (QPEs) from weather radars have the potential to address key limitations of intensity–duration–frequency (IDF) relations derived from sparse rain gage measurements. However, this potential has not been yet fully explored. Here, a methodological framework is designed for the spatial frequency analysis of extreme precipitation (P) with radar QPEs that leads to realistic quantile patterns while reducing the sampling uncertainty. The framework was applied with 19 years of QPEs from 1-h, 4-km Stage IV reanalysis from the Next Generation Weather Radar (NEXRAD) network and robustly tested against (1) a network of 204 high-resolution rain gages in central Arizona with one of the largest densities and spatial coverages in the world, and (2) extreme P quantiles from NOAA Atlas 14. It was first showed that (1) the generalized extreme value (GEV) is a suitable distribution to model the series of annual P maxima of gage records and radar QPEs across multiple durations from 1 h to 24 h, and (2) correcting the bias of the GEV shape parameter estimates due to the short sample size is a critical step. Spatial estimates of extreme P quantiles were then obtained through a hierarchical approach based on the index-flood method and the spatial smoothening (interpolation) of the GEV parameters estimated from radar QPEs (gage records). For each parameter, the most effective interpolation method was identified that limits the uncertainty caused by the short sample size and captures the local variability of extreme P. The extreme P quantiles generated from radar QPEs exhibited similar or, in some cases, higher accuracy than those generated by interpolating sparse gage information and exhibit more realistic patterns. While derived in central Arizona, the insights of this work are useful to incorporate radar QPEs into operational IDF curves in any region of the world monitored by weather radars.

1. Introduction

Extreme precipitation (P) is a natural hazard responsible for significant property damage and loss of lives. In the U.S., the National Centers for Environmental Information (NCEI) reported an average occurrence of 1.15 storm events/year between 1980 and 2000 that caused economic losses of \$2.4 billion/year, and that these numbers significantly increased to 6.61 events/year and \$15.87 billion/year between 2001 and 2022 (NOAA, 2023). Extreme P is the primary input of flooding, which NCEI reports to have led to damages quantified in \$4.26 billion/year from 1980 to 2022. Moreover, in cities, the combined effect of extreme P and high runoff coefficients of impervious urban basins could result in pluvial flooding that, in addition to harming properties, may impact importantly traffic and the operation of other urban infrastructure (Hjelmstad et al., 2021; Rosenzweig et al., 2018).

A key piece of information used to mitigate the negative impacts of

extreme P is provided by intensity–duration–frequency (IDF) curves. These synthesize the relationships between P intensity over a given duration, d , and the associated frequency of occurrence quantified through the return period, T_R (in years) (Burlando and Rosso, 1996; Koutsoyiannis et al., 1998; Madsen et al., 2002; Requena et al., 2019; Tyrallis and Langousis, 2019; among many others). IDF curves are routinely used by civil engineers to design and retrofit stormwater infrastructure, culverts, and bridges, among other goals. The most common approach to generate IDF curves over a region requires the (1) frequency analysis of P records observed at several rain gages, and (2) the application of techniques to spatially interpolate the point information derived at the gages to unmonitored sites (Blanchet et al., 2016; Fitzgerald, 1989; Guttman et al., 1993; Madsen et al., 1997; Mascaro, 2020; Modarres and Sarhadi, 2011; Schaefer, 1990).

The first task has been usually conducted with the block maxima method, which involves fitting an appropriate probability distribution to

^{*} Corresponding author at: School of Sustainable Engineering and the Built Environment, Arizona State University, 777 E. University Dr., Tempe, AZ 85287, USA.
E-mail address: gmascaro@asu.edu (G. Mascaro).

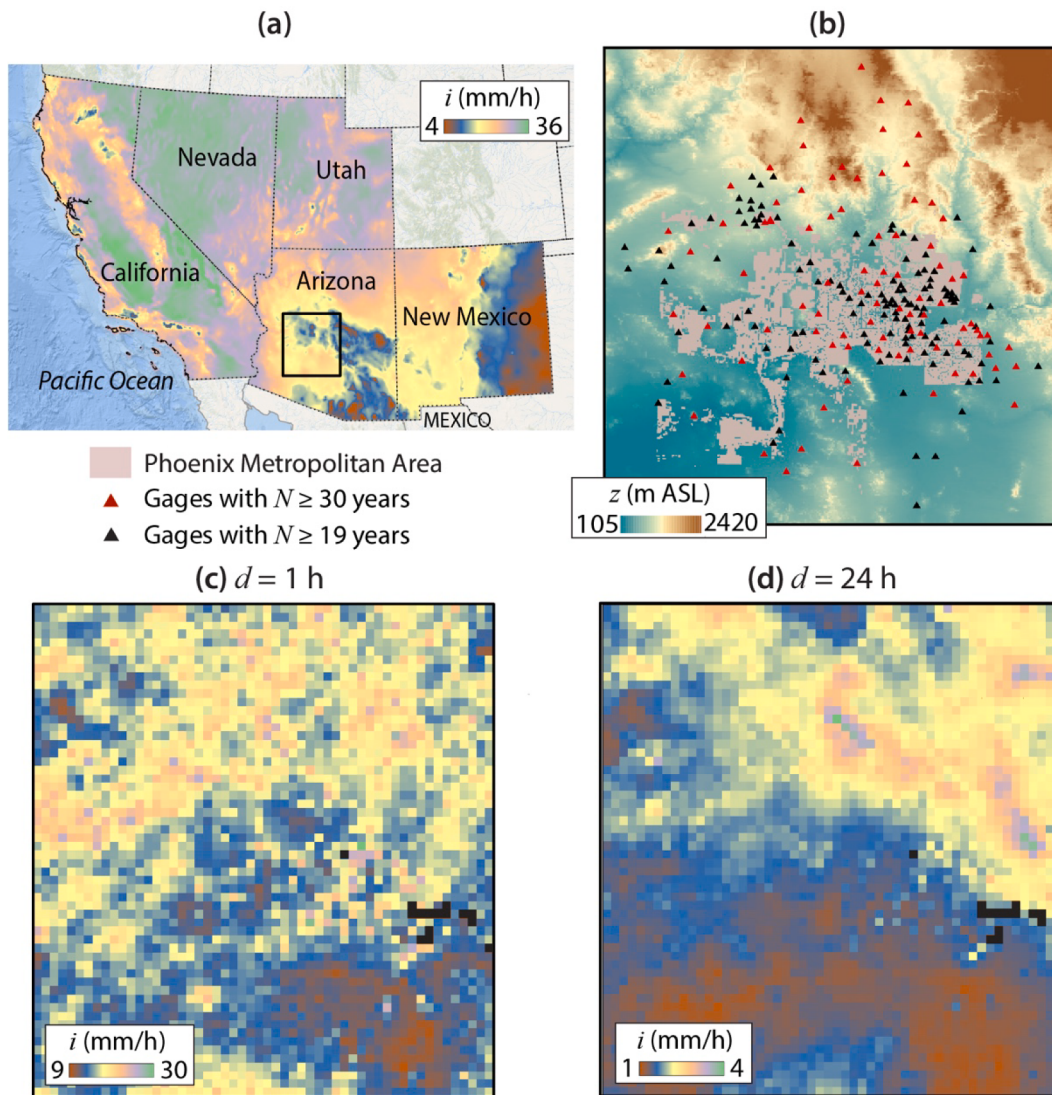


Fig. 1. Location of the study area in (a) southwestern U.S. with the background map showing PFEs for $d = 1$ h and $T_R = 2$ years in the states of Volume 1 of NOAA Atlas 14. (b) Digital elevation model from the U.S. Geological Survey National Elevation Dataset for the 216×232 km² study domain in central Arizona centered around the Phoenix Metropolitan area, along with the rain gage network of the FCDMC (z is elevation and N is the number of years with full records). (c)-(d) Average maximum precipitation intensity, i , in the study domain for (c) $d = 1$ h and (d) $d = 24$ h computed from 19 years of Stage IV QPEs.

the series of annual precipitation maxima (APM). According to the extreme value theory (Smith, 2002), the asymptotic distribution of block maxima is the generalized extreme value (GEV), which empirical evidence has shown to capture well the frequency of APM at different durations (Blanchet et al., 2016; Coles et al., 2003; Coles and Dixon, 1999; Deidda et al., 2021; Gubareva and Gartsman, 2010; Koutsoyiannis, 2004a, 2004b; Koutsoyiannis and Langousis, 2011; Mascaro, 2020; Papalexiou and Koutsoyiannis, 2013). The most popular method adopted to extrapolate information on extreme P frequency at ungauged locations is based on the index-flood technique, commonly referred to as regional frequency analysis (Dalrymple, 1960; Hosking and Wallis, 1997). This method involves grouping the gages into homogenous regions within which the probability distribution of the standardized samples of APM (e.g., divided by the sample mean) is assumed to be the same. In this way, it is possible to pool together records of multiple gages and reduce the parameter estimation uncertainty, which is usually large for records of extreme events. Precipitation frequency estimates (PFEs) at ungauged sites are then obtained by multiplying the quantiles of the single distribution of the homogenous region by the local index statistic (e.g., the local mean APM). As an example, the National Oceanic and Atmospheric Administration Atlas 14 (NOAA 14) provides IDF relations

for most of the U.S. states by applying the index-flood method with the GEV as the theoretical distribution model in most regions (Bonnin et al., 2019).

The reliability of IDF curves is critically dependent on the rain gage network density: fewer gages in a region can lead to inaccurate PFEs and, in turn, to possibly over- or under-sizing of infrastructure. Moreover, Deidda et al. (2021) recently showed that regionalization techniques based on homogeneous regions have the drawback of causing abrupt shifts in PFEs along the boundaries of contiguous regions, which are not physically plausible. To address this limitation, these authors proposed a boundaryless approach where at-site estimates of the GEV distribution parameters are interpolated in space through geostatistical methods. The proposed method was found to be effective in the case study of the island of Sardinia, Italy, based on 256 daily P gages; however, its accuracy should be further tested since it still relies on the resolution of the gage network, like the regional analyses based on homogeneous regions.

The limitations caused by sparse rain gage networks on the reliability of regional IDF curves can be addressed by using quantitative precipitation estimates (QPEs) derived from weather radars, which characterize the spatial variability of P at resolutions of up to a few km and ≤ 1

h. Research efforts have started to explore the utility of radar QPEs. [Overeem et al. \(2009\)](#) derived IDF curves for the Netherlands for d from 15 min to 24 h through the spatial frequency analysis of 11 years of radar QPEs. The authors applied the GEV distribution with regional shape and scale parameters estimated with the index-flood method, and spatially variable location parameter, and found the latter one to be negatively biased (mean of -14%) compared to estimates from gages at $d = 1$ h, likely due to the areal reduction effect, i.e., the scale mismatch between radar pixel of 6 km^2 and gage. [Marra and Morin \(2015\)](#) compared IDF relations derived from both 23 years of radar QPEs and gage records in 14 regions of Israel with different climatic features, finding that the radar overestimates the gage quantiles especially for higher return periods and in arid climates; on the positive side, the radar IDF curves are within the statistical uncertainty of the gage IDF relations in 70% of the cases.

Studies have been also recently carried out in the U.S., where the National Centers for Environmental Prediction (NCEP) Stage IV analyses provide gage-corrected QPEs at 4-km, 1-h resolution for the conterminous U.S. (CONUS) since 2002. [Ghebreyesus and Sharif \(2021\)](#) used 19 years of Stage IV QPEs to derive IDF curves for the state of Texas and validated them against NOAA 14 PFEs; they found that radar-derived PFEs have a bias within $\pm 27\%$ that is larger at $d = 1$ h, decreases with d , and becomes negligible for $d \geq 24$ h. In another effort, [McGraw et al. \(2019\)](#) compared at-site IDF relations derived from 50 years of hourly records at 539 gages covering the U.S. and 16 years of Stage IV QPEs at the co-located pixels. These analyses revealed that the radar tends to overestimate (underestimate) quantiles at $d \leq 3$ h and low T_R ($d = 24$ h and high T_R), and that performance exhibits geographic patterns explained by climate.

The findings of past studies have demonstrated the potential of radar QPEs to enhance the spatial frequency analysis of extreme P. However, radar products have not been yet incorporated into operational IDF relations, highlighting the need to further investigate their utility ([Claps et al., 2022](#)). This study contributes to addressing such a need by (1) developing a methodological framework to generate realistic patterns of extreme P quantiles from radar QPEs while reducing the statistical uncertainty due to the short sample size, and (2) robustly assessing radar-derived regional IDF relations against multiple gages, thus limiting the uncertainty of at-site comparisons. These objectives were pursued using Stage IV radar QPEs in central Arizona, which is monitored by a network of rain gages with one of the largest densities and spatial coverages in the world. An approach for the spatial frequency analysis of extreme P was designed based in part on [Deidda et al. \(2021\)](#) which relies on the index-flood method and the interpolation (for the gages) and smoothing (for the radar) of the GEV parameters. To address the uncertainty due to the limited sample size of radar QPEs (19 years), the GEV shape parameter was bias corrected through the empirical relations of [Carney \(2016\)](#). The patterns of extreme P quantiles derived from radar QPEs, gage records, and NOAA 14 were then compared in terms of accuracy and physical reliability. This work provides methodological and practical insights that are useful to improve IDF relations in the U.S. and other regions of the world by incorporating the seamless spatial information of radar P products.

2. Study area

The comparison of radar QPEs with gage records was performed in a $216 \times 232 \text{ km}^2$ area in central Arizona centered around the Phoenix Metropolitan region. [Fig. 1a](#) outlines the study area in the southwestern U.S. within the map of the 2-year, 1-hr PFEs from NOAA 14. As shown in [Fig. 1b](#), the study domain includes a large area at a low elevation in the Sonoran Desert (92–200 m above the seal level or ASL), where Phoenix is located, and the Mogollon Rim mountainous region in the northwest, where elevation reaches 2420 m ASL. Due to its arid/semi-arid climate, this desert area is categorized as the hottest and driest in North America ([Garfin et al., 2013; MacDonald, 2010](#)). For example, the average annual

P, and minimum and maximum temperature at the Phoenix Sky Harbor airport are 190 mm, 15.2°C , and 30.0°C , respectively. The P regime is characterized by two seasons with markedly different storm-generating mechanisms. From late fall to early summer, prolonged dry conditions are interrupted by storms due to cold fronts with moisture transported by westerly flows from the Pacific. These winter storms tend to last for a few days ([Barry and Chorley, 1998](#)) and be widespread with relatively low P intensity. In early July, the northward advection of water vapor from the Gulf of California and, to a lower extent, the Gulf of Mexico ([Favors and Abatzoglou, 2013; Sheppard et al., 2002](#)) leads to the onset of the North American monsoon (NAM; [Adams and Comrie, 1997](#)), which lasts until the end of September. During the NAM, convective thunderstorms with high intensity, short durations ($<1\text{h}$), and small spatial extent (a few km^2) occur according to a diurnally modulated cycle ([Balling and Brazel, 1986](#)). The different rainfall-generating mechanisms of the seasons affect the occurrence of APM at different durations with extreme events mainly happening in winter for $d \geq 12$ h and almost only in summer for $d \leq 2$ h ([Mascaro, 2020](#)). Elevation was found to be moderately correlated with extreme P by [Mascaro \(2017, 2018\)](#) with an increasing magnitude as d increases.

3. Dataset

3.1. Radar quantitative precipitation estimates

We used radar 1-h, 4-km QPEs from the NCEP Stage IV analyses for the period 2002 to 2020. Stage IV QPEs are generated for the CONUS by mosaicking reflectivity data from the Next Generation Weather Radar (NEXRAD) network, adjusting rainfall rates with gage and satellite observations, and performing manual quality control ([Nelson et al., 2016](#)). Data were acquired from the Earth Observing Laboratory (EOL) data archive ([Du, 2011](#)) in polar-stereographic coordinates for the CONUS, clipped to the $216 \times 232 \text{ km}^2$ domain shown in [Fig. 1b](#), and projected into the Universal Transverse Mercator (UTM) Zone 12 N reference leading to 58×54 pixels. From the radar QPEs, we derived the records of APM for durations $d = 1, 2, 3, 6, 12$, and 24 h in each pixel of the study domain. For each year, we extracted the largest P intensity, i (in mm/h), over d -long moving windows independently of the presence of missing data. In years with missing values, we used the method of [Papalexiou and Koutsoyiannis \(2013\)](#) and [Blanchet et al. \(2016\)](#) to decide whether there are enough observations to retain the annual maxima. If N is the number of years with no missing data, we first sorted the associated N APM. For a year with a fraction f of missing data, we (1) computed the rank of its i in the series extracted for the N complete years, and (2) retained (rejected) i in case its rank is above (below) $f \cdot N$. As an example, [Fig. 1c,d](#) show the mean APM for $d = 1$ and 24 h in the study region derived from Stage IV. For the analyses based on at-site estimates, we excluded 15 radar pixels with a few extremely high values of APM, which are likely due to errors in the reflectivity-rain rate conversion algorithm. Extreme P statistics were subsequently estimated at these locations when the spatial interpolation techniques were applied.

3.2. Rain gages

The ALERT network of rain gages managed by the Flood Control District of Maricopa County (FCDMC) was used as a reference to assess the ability of radar QPEs to characterize extreme P statistics and support the generation of IDF curves. The network started operating in the early 1980s and currently includes 365 gages that monitor P in real time over a region of about $29,600 \text{ km}^2$ centered around the Phoenix Metropolitan area ([Fig. 1b](#)). The gage elevation ranges from 220 to 2325 m ASL, although most (195) gages are installed below 800 m ([Mascaro, 2020](#)); the inter-gage distance varies between 0.5 km and 227 km with a median of 70 km. For our analyses, we used 204 gages without missing data across all d 's during the same 19 years when Stage IV data are available. We also utilized a subset of 87 gages with long-term (≥ 30 years)

observations to investigate the effect of sample size on the results and generate robust at-site PFEs used as reference when testing the accuracy of different gage- and radar-derived regional IDF relations. Fig. 1b shows the location of the gages with the two record lengths. All gages are of the tipping bucket type with a tipping depth of 1 mm. Records were obtained from the FCDMC in the original form of tipping instants in seconds; these were then converted into signals at a given duration d according to the procedure described in Mascaro et al. (2013). The same method illustrated for the radar QPEs was applied to derive the series of APM while accounting for missing data.

3.3. Precipitation frequency estimates from NOAA Atlas 14

To further validate the reliability of the radar-derived extreme P statistics, we also used gridded PFEs from NOAA 14 (Bonnin et al., 2019). These are released in different Volumes for most U.S. states, with Volume 1 covering the southwestern region (Fig. 1a). As previously mentioned, the methodology used in NOAA 14 to generate PFEs is based on the regional frequency analysis of gage records of APM with the index-flood method (Hosking and Wallis, 1997), which is briefly described in section 4.2. A key step of the procedure needed to generate PFEs at ungaged locations involves the computation of the local mean APM at a generic site j , indicated as $m^{(j)}$. In NOAA 14, grids of $m^{(j)}$ at ~ 800 -m resolution were produced for the different d 's through cell-specific regressions based on multiple predictors, including the mean annual precipitation from the Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly et al., 1997), distance from the coast, terrain, and geographic features, as well as on user knowledge (Bonnin et al., 2019). It is worth noting that the number of hourly gages used in Volume 1 of NOAA 14 is significantly lower than that of the daily gages, implying that the grids of $m^{(j)}$ and PFEs for $d < 24$ h have higher uncertainty compared to those for $d \geq 24$ h. For our analyses, we downloaded the PFE grids for all durations and return periods from the NOAA website. Since the $m^{(j)}$ grids were not available on the website, we contacted the NOAA staff that sent us the digital maps for $d = 1, 6$, and 24 h.

4. Methods

4.1. The generalized extreme value (GEV) distribution

The frequency of extreme P was modeled by fitting the GEV distribution to the APM series for the different durations, d . This distribution was found by Mascaro (2020) to be appropriate in the region using gage observations through L-moments ratio diagrams and goodness-of-fit (GOF) tests, including Lilliefors, Anderson–Darling, and Cramér–von Mises; the GEV is also used to obtain PFEs in NOAA 14 (Bonnin et al., 2019). Here, we tested its applicability to the APM series derived from the radar QPEs. The cumulative distribution function (CDF) of the GEV distribution for the random variable $I \equiv$ “annual maximum P intensity for a given d ” is defined as:

$$F(x) = F(x|k, \mu, \sigma) = \Pr\{I \leq x\} = \begin{cases} \exp\left\{-\left(1 + k\frac{x-\mu}{\sigma}\right)^{-\frac{1}{k}}\right\} & k \neq 0 \\ \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\} & k = 0 \end{cases} \quad (1)$$

where $k \in (-\infty, +\infty)$ is the shape parameter, $\mu \in (-\infty, +\infty)$ is the location parameter, and $\sigma \in (0, +\infty)$ is the scale parameter. The GEV distribution is categorized as Type I or Gumbel if $k = 0$ with support $-\infty < x < +\infty$, Type II or Fréchet if $k > 0$ (heavy tail) with support $\mu - \frac{\sigma}{k} \leq x \leq \infty$, and Type III or Weibull if $k < 0$ with an upper bounded support $-\infty < x \leq \mu - \frac{\sigma}{k}$ (Smith, 2002). The quantiles associated with the annual return period $T_R = 1/[1 - F(x)]$ are computed by inverting equation (1).

Table 1

Coefficients used to bias correct k with Eq. (2) for Volume 1 of NOAA Atlas 14 from Carney (2016).

d (h)	α_k	β
1	−0.149	0.8726
2	−0.117	0.7172
3	−0.087	0.5945
6	−0.035	0.4231
12	−0.013	0.3662
24	−0.053	0.4774

The suitability of the GEV distribution was verified through the L-moment ratios diagram (Hosking, 1990). The distribution was then fitted to the APM series using the method of probability weighted moments (PWM; Hosking et al., 1985), which was chosen since it is more robust to outliers than other techniques when the sample size is small (Hosking et al., 1985; Vogel and Fennessey, 1993, among others). Since estimates of the shape parameter from short records are highly uncertain (Hosking and Wallis, 1997; Mascaro, 2020; Overeem et al., 2009; Papalexiou and Koutsoyiannis, 2013), k was bias corrected to account for the sample size using the empirical relations suggested by Carney (2016). Based on the global study of Papalexiou and Koutsoyiannis (2013) with daily rain gages, Carney (2016) obtained the following relation, available for d from 1 h to 60 days in different U.S. regions, using hourly and daily gages of NOAA 14:

$$\hat{k} = \frac{0.045}{0.045 + 1.27N^{-0.70}} \{\hat{k}(N) - [\alpha_k + \beta/N]\} + \alpha_k \quad (2)$$

In Eq. (2), $\hat{k}$ is the bias corrected value, $\hat{k}(N)$ is the PWM sample estimate of k for the sample size N , α_k is the unbiased average of k , and β is a coefficient. The values of α_k and β for the durations analyzed here are reported in Table 1.

4.2. Spatial frequency analysis

The spatial frequency analysis of extreme P at a given d was carried out for both radar and gages through a stepwise technique based in part on the boundaryless approach of Deidda et al. (2021). This, in turn, relies on the index-flood method, whose main steps and equations are summarized next. Let $\mathbf{x}^{(j)}$ be the APM series at a given d in the j -th radar pixel or gage. The sample $\mathbf{x}^{(j)}$ is first standardized as:

$$\mathbf{y}^{(j)} = \mathbf{x}^{(j)} / m^{(j)} \quad (3)$$

where $m^{(j)}$ is the sample mean (i.e., the index-flood, here denoted as index-rainfall). In the index-flood method, statistical tests are applied to investigate the hypothesis that the distribution of $\mathbf{y}^{(j)}$ is the same at all sites. If confirmed, the standardized records at all sites are pooled together and their distribution is characterized by a proper parametric model (e.g., the GEV). Its quantile function, $i_y(T_R)$, is known as growth curve and is used to obtain the quantiles at any location j of the homogeneous region, $i(T_R)^{(j)}$, as:

$$i(T_R)^{(j)} = m^{(j)} \cdot i_y(T_R) \quad (4)$$

If the GEV distribution is used to model the dimensionless variable y whose mean is 1, its dimensionless parameters $[k, \mu^*, \sigma^*]$ are related to each other via the relationship:

$$\mu^* = \begin{cases} 1 + \frac{\sigma^*}{k} \{1 - \Gamma(1 - k)\} & k \neq 0 \\ 1 - \gamma\sigma^* & k = 0 \end{cases} \quad (5)$$

where $\Gamma(\cdot)$ represents the gamma function and γ is the Euler's constant.

The boundaryless approach of Deidda et al. (2021) is based on a hierarchical parameter estimation that relies on the dimensionless variables and GEV parameters, as in the index-flood method. However,

Table 2

Techniques used to interpolate (smoothen) the index-rainfall, $m^{(j)}$, the bias corrected k , and the dimensionless parameters σ^* and μ^* for the gages (radar) P products in the hierarchal approach described in section 4.2. KUD = kriging under uncertainty, KED = kriging with external drift, and MA = moving average (see Appendix for details).

Precipitation product	$m^{(j)}$	k	σ^*	μ^*
Gages	KUD and KED	KUD	KUD	Eq. (5)
Radar	KUD	MA	MA	Eq. (5)

instead of using homogenous zones where all $y^{(j)}$ samples are pooled together to estimate a single growth curve, it involves the spatial interpolation of the at-site dimensionless GEV parameters. This method was adopted here for both gages and radar with some modifications, as summarized in the next steps:

- **Step 1:** The at-site GEV parameters $[k, \mu^*, \sigma^*]$ were estimated from each dimensionless sample $y^{(j)}$ at the individual gages and radar pixels. The shape parameter k was then bias corrected through the empirical relation (2) to account for the short sample size; negative estimates were set to zero for the reasons explained in section 5.2.
- **Step 2:** For the gages, the bias corrected k was spatially interpolated into the 58×54 radar grid at 4-km resolution with kriging for uncertain data (KUD; Mazzetti and Todini, 2008), as in Deidda et al. (2021). For the radar, the already gridded bias corrected k was smoothened with a simpler moving average (MA).
- **Step 3:** The at-site scale parameter σ^* was re-estimated conditioned on the gridded and bias corrected k from step 2. KUD (MA) was then

used to spatially interpolate (smoothen) the re-estimated σ^* for the gages (radar).

- **Step 4:** The at-site scale parameter μ^* was computed using equation (5) with the gridded k and σ^* from steps 2 and 3. At the end of this step, gridded estimates of the dimensionless GEV parameters k , σ^* , and μ^* were obtained.
- **Step 5:** Maps of the index-rainfall $m^{(j)}$ were obtained for the 58×54 radar grid with both P products. For the gages, two methods were tested to interpolate the APM, including KUD and kriging with external drift (KED; Goovaerts, 2000a) to account for the effect of elevation (see section 5.3). For the radar, the mean APM was smoothened using KUD to reduce the small-scale variability of this metric that could be large given the short sample size.
- **Step 6:** For each P product, gridded estimates of the dimensional GEV parameters k , μ , and σ were then obtained and, from these, of the P quantiles, $i(T_R)^{(j)}$. Note that (1) k does not change from the dimensionless value, and (2) μ and σ were obtained by multiplying the dimensionless parameters by $m^{(j)}$.

A brief description of the KUD, MA, and KED interpolation/smoothening techniques and how they were implemented here is provided in the Appendix A, while Table 2 summarizes the individual techniques used to interpolate or smoothen the index-rainfall and the GEV parameters for the two P products.

4.3. Error metrics

The performance of the spatial frequency analysis based on Stage IV QPEs, gage records, and NOAA 14 was assessed through error metrics

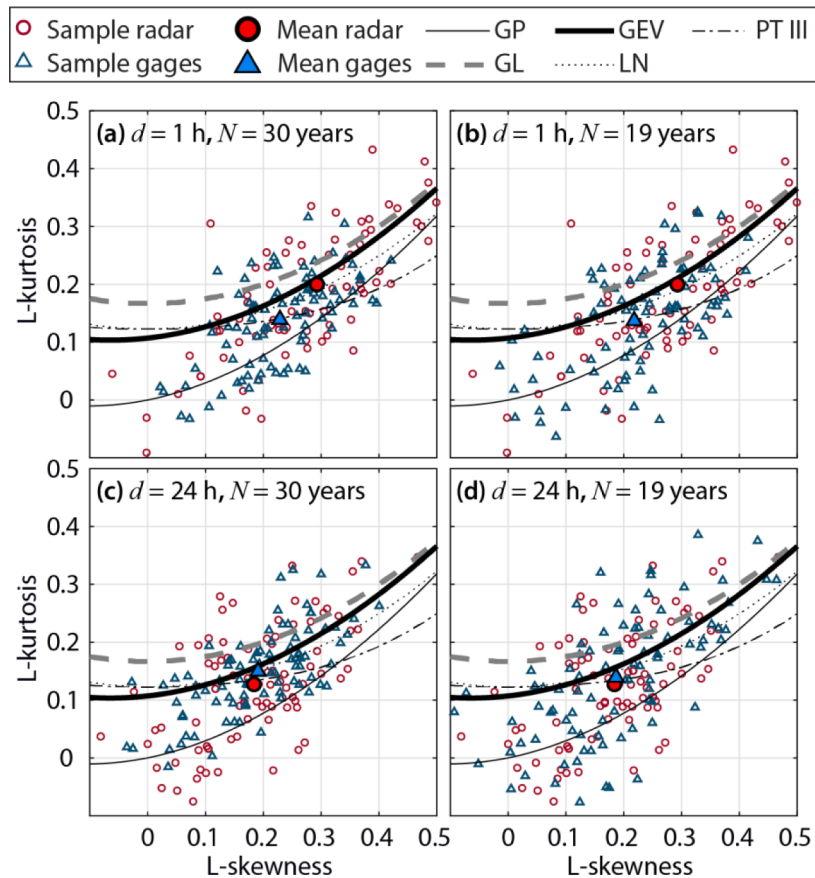


Fig. 2. L-moment ratios diagram for records of AMP at 87 gages and co-located radar pixels for (a)–(b) $d = 1$ h and $N = 30$ and 19 years records for the gages, respectively. (c) and (d) are as (a) and (b), but for $d = 24$ h. The lines show the theoretical L-moment combinations for the generalized Pareto (GP), generalized logistic (GL), generalized extreme value (GEV), lognormal (LN), and Pearson Type III (PT3). The means of the observed samples are also reported.

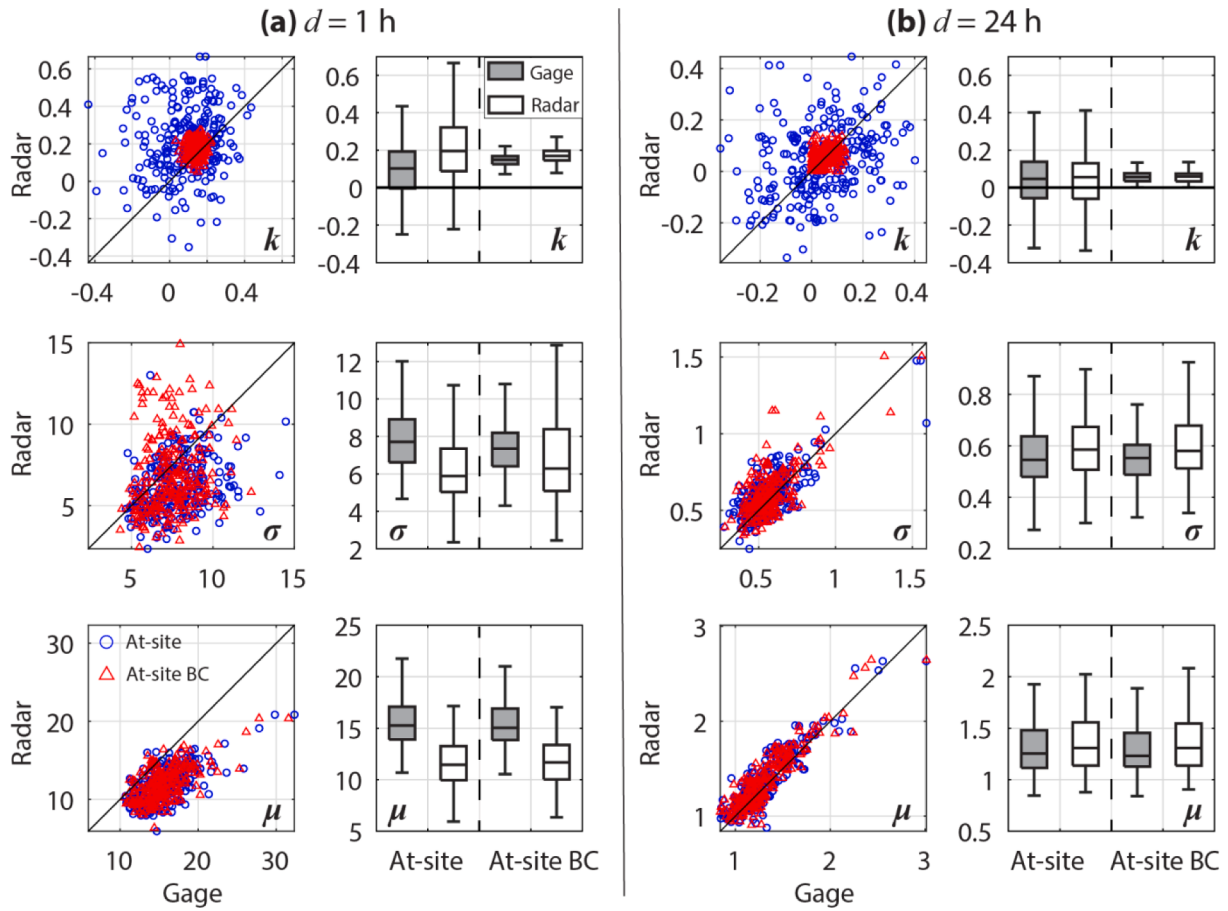


Fig. 3. Comparison between the GEV parameters estimated at the 204 gages and co-located radar pixels under At-site (red) and At-site BC (blue) for (a) $d = 1$ h and (b) $d = 24$ h, shown through scatterplots (left) and boxplots (right). Results for k , σ and μ are reported in the top, middle, and bottom rows, respectively. Units of σ and μ are mm/h. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between the quantiles of each product against those derived from the at-site application of the GEV distribution at the 87 gages with long-term (≥ 30 years) observations. For the j -th gage and/or co-located radar pixel, the relative bias (RB) for a given d was computed as:

$$RB^{(j)} = \frac{i_{EST}(T_R)^{(j)} - i_{REF}(T_R)^{(j)}}{i_{REF}(T_R)^{(j)}} \times 100 \quad (6)$$

where $i_{EST}(T_R)^{(j)}$ and $i_{REF}(T_R)^{(j)}$ are the T_R -quantiles calculated for the estimation (EST) and reference (REF) methods, respectively. The relative biases were then averaged across all J sites as:

$$RB = \frac{1}{J} \sum_{j=1}^J RB^{(j)} \quad (7)$$

We also computed the relative root-mean-square error (RRMSE) as:

$$RRMSE = \left[\frac{1}{J} \sum_{j=1}^J (RB^{(j)})^2 \right]^{1/2} \quad (8)$$

5. Results and discussion

5.1. Evaluation of the GEV hypothesis

The suitability of the GEV distribution to model APM series observed at gages and radar pixels was evaluated using the L-moment ratios diagram (Hosking, 1990, 1992; Hosking and Wallis, 1993; Peel et al., 2009). This is shown in Fig. 2 for the APM series at $d = 1$ and 24 h at the

gages and co-located radar pixels. To investigate the effect of sample size, the panels on the left show results for the 87 gages with longer records of $N \geq 30$ years, while those on the right report results at the same 87 gages for the $N = 19$ years where radar QPEs are available. For all durations and products, the sample estimates are scattered around the theoretical GEV curve, and the mean L-skewness and L-kurtosis are very close to or lie on the GEV line, indicating that this distribution captures well the APM series of both gages (as also found by Mascaro, 2020) and radar. As expected, the scatter of the gage sample estimates is larger for $N = 19$ years; however, the mean L-skewness and L-kurtosis for the gages do not change significantly with N (i.e., the position of the filled blue triangle is practically the same in the left and right panels for the same d). The scatter of the radar samples is very similar to that of the gages with $N = 19$ years, while the averaged L-moments for the radar are slightly higher than those of the gages for $d = 1$ h and practically the same for $d = 24$ h. These findings suggest that (1) the use of 19-year-long records allows capturing the average statistical properties in the region obtained from longer records; and (2) the L-moments of APM series from gages and radar are slightly different at lower N , but these differences do not depend on N . The at-site comparisons and the generation of regional IDF relations for the gages presented next were based on the 204 gages with the same $N = 19$ years of the radar records, while the 87 long-term gages were used as the reference for the error metrics.

5.2. Comparison of at-site GEV parameters from gages and radar

As a first-level assessment of radar QPEs' ability to characterize extreme P, we compared at-site estimates of the GEV parameters at the

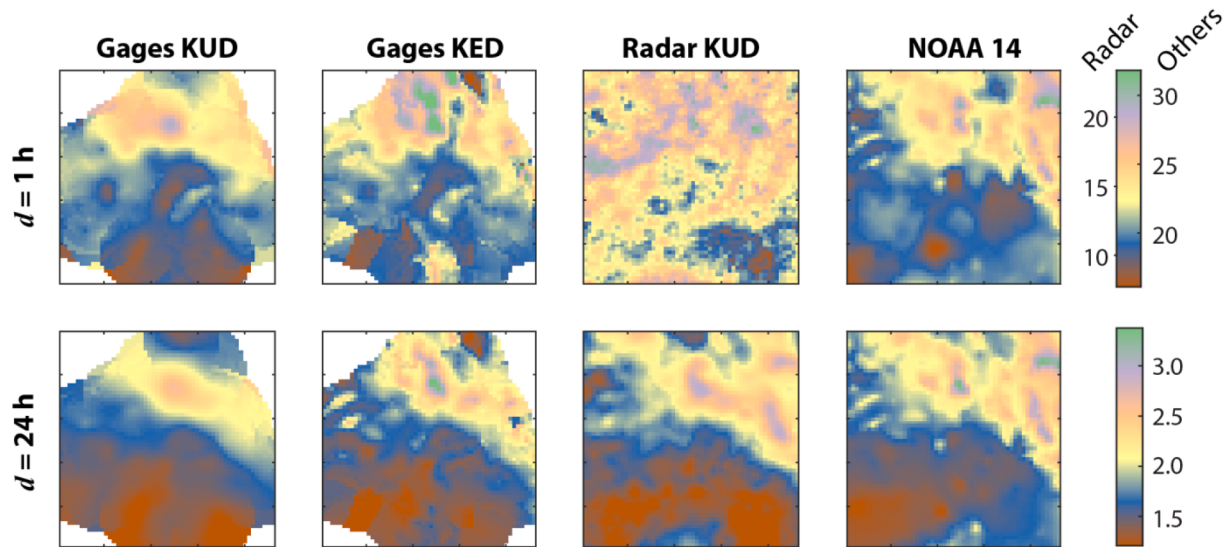


Fig. 4. Maps of $m^{(j)}$ (mm/h) for $d = 1$ h (top row) and $d = 24$ h (bottom row) derived by interpolating the gage records with KUD and KED, smoothing the radar QPEs with KUD, and from NOAA 14. The limits of the colormap for the radar-derived map for $d = 1$ h are different from the other products due to a negative bias. The white color in the gage maps indicates missing values (see Appendix).

204 gages and co-located radar pixels, as in Marra and Morin (2015), McGraw et al. (2019), and Overeem et al. (2009). Results are presented in Fig. 3 via scatterplots and boxplots for $d = 1$ and 24 h (the other durations are reported in Fig. S1 of the Supplementary Material), while error metrics measuring linear correlation (coefficient of determination, R^2), scatter (RRMSE or RMSE), and bias are reported in Table S1. The scatterplots show that the relation between the local unconditioned estimates (labeled At-site) of k for the two datasets exhibits a large scatter and low correlation for all d 's because the estimation of this parameter is highly uncertain with such a short sample size. The correspondence is instead stronger for At-site estimates of σ and, even more, of μ , with R^2 (RRMSE) increasing (decreasing) with d . Moreover, as better visualized through the boxplots, the radar estimates of both σ and μ are negatively biased compared to the gage values for $d = 1$ h, and the negative bias is reduced as d increases (Fig. S1) becoming slightly positive for $d = 24$ h.

The bias correction of the at-site estimates of k (labeled At-site BC) significantly reduces the scatter (although R^2 is still low) and results in positive values (i.e., heavy tailed distributions) at almost all sites. The cases with negative k (none for $d \leq 3$ h and as high as 4% for $d = 24$ h) were found to be placed at random locations (not shown) and attributed to sampling variability. At these sites, the k values were replaced with 0 to avoid the existence of an upper limit for P , which was not considered physically possible. The bias corrected k decreases with d from a median of 0.15 at $d = 1$ h to 0.05 for $d \geq 12$ h (see Figs. 3 and S2). Despite the dramatic reduction of the variability of k , the estimates of σ and μ conditioned on the bias corrected k are very similar to the At-site values, indicating that the body of the distribution is robustly characterized despite the short record. The bias of σ and μ at lower durations is also still present. As shown in the next section, a negative bias at short durations was also found for the index-rainfall. This bias is most likely due to the discrepancy between the radar pixel area (16 km²) and the point information at the gages, which, in turn, is relevant at short durations that are dominated by spatially isolated monsoonal thunderstorms and becomes negligible at larger durations that are mainly affected by more widespread frontal systems (Mascaro, 2020). A similar result was also found by Overeem et al. (2009) and Ghebreyesus and Sharif (2021).

Table 3

R^2 of the linear regression between elevation and $m^{(j)}$ at the 204 gage locations derived from gage measurements (Gages At-site), estimated by interpolating the gage records with KED (Gages KED), computed by smoothing the radar QPEs with KUD (Radar KUD), and extracted from NOAA 14.

Duration (h)	1	2	3	6	12	24
Gages At-site	0.36	0.37	0.38	0.42	0.55	0.61
Gages KED	0.65	0.63	0.64	0.68	0.71	0.69
Radar KUD	0.32	0.37	0.41	0.50	0.64	0.68
NOAA 14	0.70	–	–	0.76	–	0.81

5.3. Comparison of index-rainfall maps from gages, radar, and NOAA 14

Regional IDF relations were then derived through the proposed framework. The generation of $m^{(j)}$ grids is presented first since it is an independent step that allows discussing some details useful to better interpret the other steps. Fig. 4 shows the grids of the index-rainfall, $m^{(j)}$, for $d = 1$ h and 24 h generated from gage and radar P records, and NOAA 14. To facilitate the comparison against the radar products, the $m^{(j)}$ grids for the gages (NOAA 14) were derived (aggregated) at the same resolution as the radar. The interpolation of gage observations with KUD leads to fields that capture the overall south-to-north increase of the mean APM, but that are rather smooth with artificial “islands” of high or low values dependent on local observations. To address these limitations, following Mascaro (2017, 2018, 2020), we considered elevation as an ancillary predictor to better capture the spatial variability of $m^{(j)}$. Table 3 reports the R^2 between at-site estimates of $m^{(j)}$ at the gages and the corresponding elevation (top row: Gages At-site), which suggests that the effect of elevation is moderate at lower d (e.g., $R^2 = 0.36$ for $d = 1$ h) and becomes stronger as d increases (e.g., $R^2 = 0.61$ for $d = 24$ h). Previous studies in other regions have also found that the elevation control on extreme P statistics varies in terms of strength and sign depending on the P duration (Avanzi et al., 2015; Formetta et al., 2022; Mazzoglio et al., 2022; Rossi et al., 2020). We then interpolated $m^{(j)}$ for the gage records with KED using elevation as an ancillary predictor for all d 's. The resulting maps in Fig. 4 show that small-scale terrain features of $m^{(j)}$ are now represented with a similar level of detail as NOAA 14. However, the patterns of both Gages KED and NOAA 14 are quite similar across durations and highly linked to elevation ($R^2 \geq 0.63$ across all d 's; see Table 3), even at smaller d where local gage observations indicate a lower orographic control. The discrepancies between local gage

Table 4

Mean across 204 gage sites of the bias between $m^{(j)}$ (i.e., mean APM) of the radar QPEs at the co-located pixel and the gage records.

d (h)	Mean bias (mm/h)
1	-4.0
2	-1.0
3	-0.5
6	-0.07
12	0.02
24	0.07

estimates and NOAA 14 can be ascribed to (1) the different approach adopted in NOAA 14 to compute the index-rainfall as a function of mean annual precipitation, and (2) the lower number of gages used in NOAA 14 (33 hourly and 75 daily) compared to our network (204 for all durations).

The $m^{(j)}$ grids for the radar were obtained by applying KUD to the mean APM at the pixels to smoothen the sampling variability of this P statistic. At $d = 1$ h, the $m^{(j)}$ estimates are negatively biased compared to the gages (note the different legend used in Fig. 4; mean bias values reported in Table 4). The bias is reduced at larger durations and becomes negligible for $d \geq 12$ h, a result directly related to the bias in σ and μ displayed in Fig. 3. The radar maps exhibit a small-scale variability with a similar level of detail of Gages KED and NOAA 14. However, differently from these two products that are highly related to elevation across

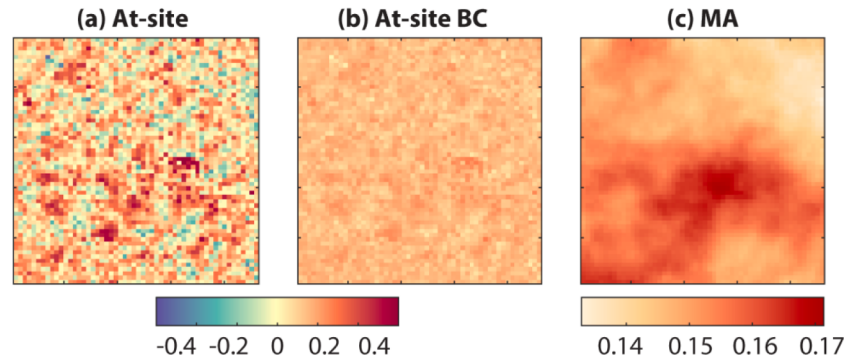


Fig. 5. Maps of the GEV shape parameter, k , for $d = 1$ h obtained from the radar QPEs with At-site, At-site BC, and MA methods. The same colormap with a large range of negative and positive values is used for At-site and At-site BC to highlight the clustering of k and the effect of bias correction. A colormap with a smaller range of positive values is instead used for MA to better visualize the spatial variability.

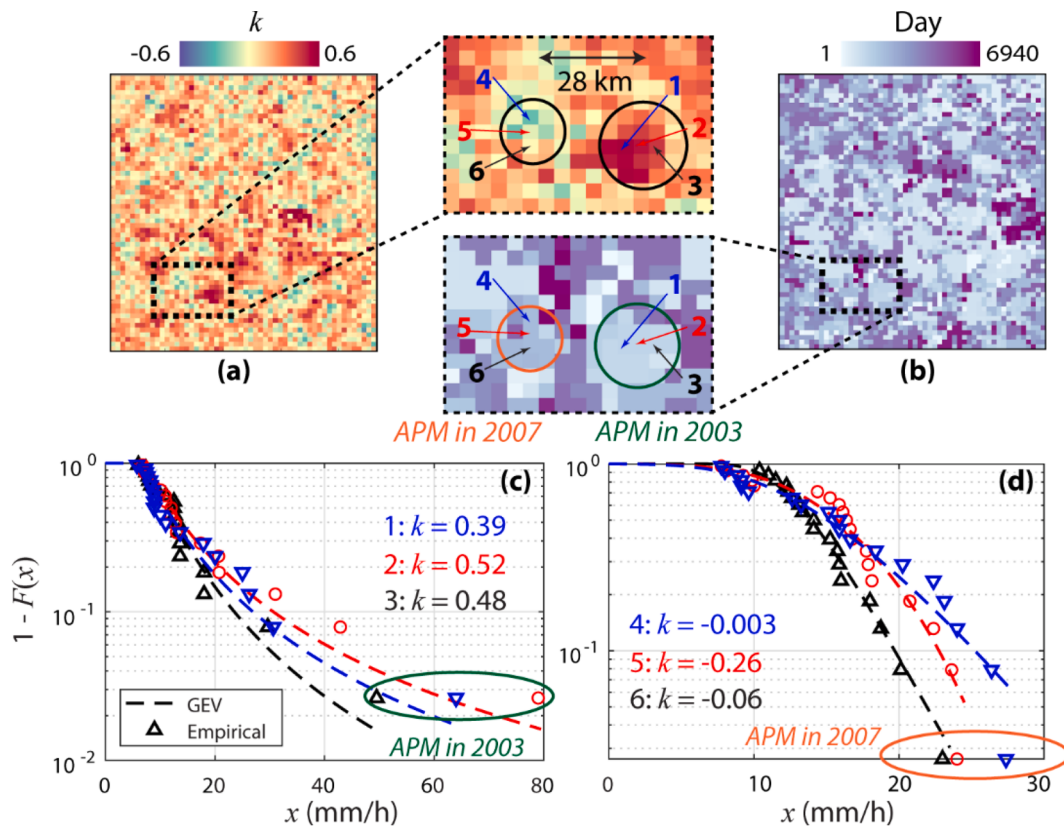


Fig. 6. Maps of (a) At-site k from radar QPEs for $d = 1$ h and (b) day of observation ($1 = 1/1/2002$; $6940 = 12/31/2020$) of the highest APM at $d = 1$ h, with zooms on two clusters with high positive and negative values of k . (c)-(d) Empirical CDF and fitted GEV distributions of APM records in three representative pixels in each of the two clusters.

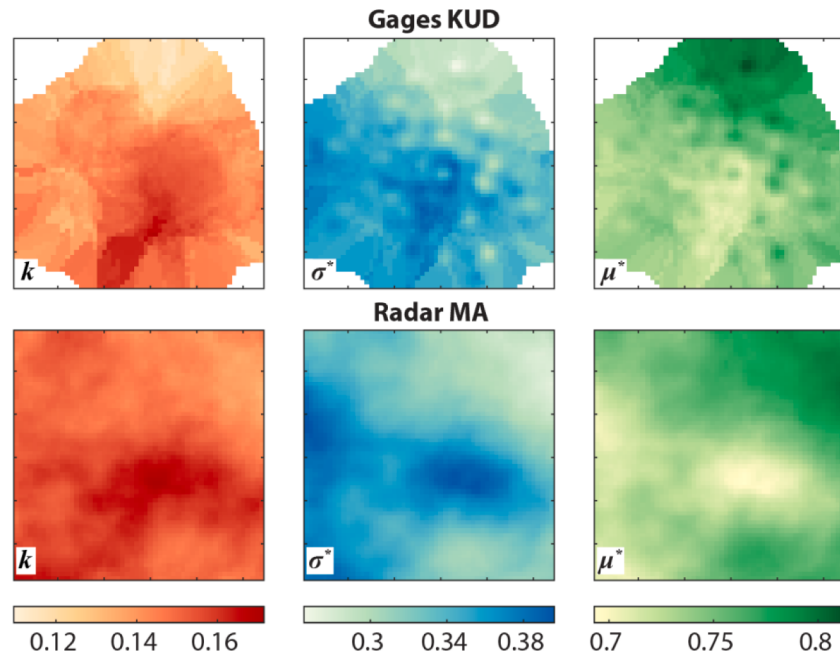


Fig. 7. Maps of k and the dimensionless parameters σ^* and μ^* for $d = 1$ h obtained from the gages with KUD (top row) and the radar with MA (bottom row).

all d 's, the patterns of Radar KUD change with d and have a very similar R^2 with elevation to the at-site gage observations (compare values for Gages At-Site and Radar KUD in Table 3). This indicates that the radar-derived maps can capture not only elevation-dependent trends but also a portion of the spatial variability of $m^{(j)}$ that is not controlled by orography (Mazzoglio et al., 2022). This is well visualized by comparing the maps for $d = 1$ h of all products. To conclude, these analyses suggest that the spatial smoothing of the radar mean APM with KUD reveals key features of the index-rainfall $m^{(j)}$ that the spatial interpolation of gage information (here, Gages KED and NOAA 14) is not able to capture.

5.4. Generation of maps for the GEV shape parameter with radar QPEs

The first step of the hierarchical approach adopted for the spatial frequency analysis of extreme P is the generation of maps for the GEV shape parameter, k . The maps derived from radar QPEs using different estimation methods are presented in Fig. 5. Results are shown only for $d = 1$ h because they are qualitatively similar for other durations (see Fig. S2). Since the radar provides seamless spatial QPEs, the easiest approach to generate spatial maps is via the grid of At-site estimates (Fig. 5a). However, as already illustrated in Fig. 3, the range of these local estimates is quite large with both high positive and negative values due to the short sample size of the APM records. The map further reveals that At-site k estimates could (1) vary significantly within relatively small distances, and (2) exhibit clusters of ~ 10 – 20 pixels with very similar high or low values. The presence of an organized spatial structure for k does not appear to be the result of the physical processes affecting P in the region, since it is very unlikely that the right tail of the extreme P distribution (1) is bounded (negative k) only in limited spatial areas, and (2) changes so abruptly within short distances (8–40 km), especially in the rather flat Phoenix Metropolitan region (Fig. 1b).

A plausible explanation of these unrealistic spatial features is the large uncertainty in the estimation of the shape parameter combined with the typical size of storms leading to extreme P. To better illustrate this, we compared in Fig. 6a,b the maps of At-Site k and the day of occurrence of the largest APM. The rationale of this analysis is as follows. When the sample size is short, the largest APM might significantly affect the shape of the distribution tail, i.e., the estimate of k . Thus, if the dates are randomly distributed in space, then the largest APM values at

neighboring pixels are originated from different storms and the corresponding k 's are likely different. If the same date instead occurs in connected pixels, then the largest APM in this region is caused by the same storm and the chance of observing a cluster of similar k 's is high. As shown in Fig. 6a,b, the maps of k and dates exhibit clusters with similar sizes and locations, indicating that the spatial correlation of k is closely related to the size of the storms causing the largest APM. This is further illustrated by reporting in Fig. 6c,d the GEV distribution for some pixels belonging to two spatial clusters with high and low k values, respectively, placed ~ 28 km apart (zoomed areas in separate panels). For the pixels of each cluster, the shape of the distribution is severely affected by the largest APM values which occurred on the same day (4/4/2003 for pixels 1, 2, and 3; and 7/24/2007 for pixels 4, 5, and 6).

Turning our attention back to Fig. 5, we can notice that the large variability of the At-site map of k was greatly reduced through the bias correction of k in At-site BC (Fig. 5b), which led to non-negative estimates. However, the use of equation (2) to correct the bias still preserved the spatial correlation of k caused by the clusters of the most intense storms. To eliminate this physically unrealistic feature from the map, the At-site BC grid was smoothed through MA (Fig. 5c). This resulted in a pattern where k varies gradually without abrupt changes or clusters, exhibiting relatively higher (smaller) values in the southwestern and central portion (northeastern) of the domain. Note that we also tested the accuracy of KUD finding that this technique did not eliminate the unrealistic clusters of k (not shown). In conclusion, these findings indicate that the use of MA on the bias corrected k estimates from radar QPEs leads to the most realistic patterns and avoids the spatial discontinuities of precipitation frequency estimates that could arise if a single regional value is assumed as done in the traditional regional frequency analysis (Deidda et al., 2021). The MA maps were then used for the subsequent analyses.

5.5. Comparison of maps for the dimensionless GEV parameters from radar and gages

The next steps of the hierarchical approach involved the computation of maps for the GEV dimensionless parameters σ^* and μ^* conditioned on k . This was done for both radar QPEs using MA and gage records using KUD (Table 2). Results are shown in Fig. 7 for $d = 1$ h and

Table 5

R^2 of the linear regression between elevation and GEV dimensionless parameters estimated from the gages with KUD and from the radar with MA.

d (h)	Gages KUD			Radar MA		
	k	σ^*	μ^*	k	σ^*	μ^*
1	0.48	0.65	0.67	0.57	0.54	0.56
2	0.47	0.44	0.50	0.32	0.46	0.48
3	0.39	0.37	0.42	0.07	0.46	0.44
6	0.13	0.41	0.39	0.003	0.45	0.39
12	0.27	0.37	0.38	0.04	0.37	0.34
24	0.14	0.32	0.33	0.007	0.27	0.24

Fig. S3 for the other durations. For both products, each parameter is included within similar ranges and exhibits large-scale variations with larger (smaller) values of k and σ^* (μ^*) in the southwestern and central parts of the domain at lower elevations, and smaller (larger) values in the northeastern part at higher altitude. This is quantified by similar values of R^2 between parameters and elevation for the two products reported in Table 5. Interestingly, contrary to what was found for $m^{(j)}$, the link between the dimensionless GEV parameters and elevation is stronger for $d = 1$ h and weakens at larger durations. Despite such similarities between the two products, the patterns obtained for the gages using KUD exhibit artifacts caused by the variable spatial density of the point observations. The patterns derived from the seamless radar QPEs are instead smoother and appear more physically plausible.

5.6. Performance of extreme precipitation quantiles from radar, gages, and NOAA 14

The dimensionless scale and location GEV parameters were multiplied by the index rainfall, $m^{(j)}$, to obtain the dimensional values that, along with the shape parameter, were used to compute extreme P quantiles for all d 's and $T_R = 2, 5, 10, 25$ and 50 years, thus characterizing a wide range of IDF relations. The error metrics, RB and RRMSE, for the different products against at-site GEV quantile estimates at 87 gages with long-term P records are summarized via the heatmaps in Fig. 8. The spatial frequency analysis of 19 years of P data at 204 gages (Gages KUD) produced IDF relations with slightly positive RB between

2.4% and 5.4% that have no evident link with d and T_R , and RRMSE between 7.5% and 13.7% which is slightly larger for lower d and higher T_R . The IDF relations obtained with the radar QPEs (Radar MA) are negatively biased for small d (e.g., RB is $\sim -20\%$ and RRMSE is $\sim 24\%$ for $d = 1$ h) and have only slightly worse performance than Gages KUD for $d \geq 6$ h. The negative bias is caused by the bias in $m^{(j)}$ that, as previously noted, is most likely due to the scale discrepancy between radar pixel and gage. The extreme P quantiles for the radar were then recalculated for $d \leq 6$ h by removing the bias in $m^{(j)}$ via the simple subtraction of the mean bias reported in Table 4. As shown in Fig. 8 (Radar MA with bias-corrected (BC) $m^{(j)}$), this adjustment dramatically improved the error metrics that became very close to and, in some cases, better than those for Gages KUD. Finally, the performances of NOAA 14 PFEs were the worst in terms of RRMSE for all d 's and T_R 's as well as of RB for $T_R \leq 10$ years, while they were comparable to the other products when considering RB for $T_R > 10$ years. The possible reasons of the lower performance are described in section 5.3.

Examples of maps of extreme P quantiles for the different products are presented in Fig. 9. For $d = 1$ h, while all products show a general south-to-north increasing trend, the small-scale variability is quite different, especially for larger T_R . The spatial variability of the quantiles for the bias-corrected Radar MA appears the most realistic, while Gages KUD and NOAA 14 exhibit the typical artifacts caused by the interpolation of point information, including the presence of "islands" with larger or higher values. At the largest $d = 24$ h, the patterns are quite similar across the products, especially when comparing Gages KUD and Radar MA; this latter product and NOAA 14 also represent in a remarkably similar way the variability in the northeastern portion of the domain. Moreover, when considering a given product, the relative spatial variability is quite similar across T_R .

6. Conclusions

This study proposed and tested a framework to perform the spatial frequency analysis of extreme precipitation (P) with radar QPEs that generates realistic quantile patterns while reducing the sampling uncertainty. The framework was applied using 19 years of Stage IV radar QPEs in central Arizona and thoroughly tested against a network of high-resolution rain gages with one of the largest densities and spatial

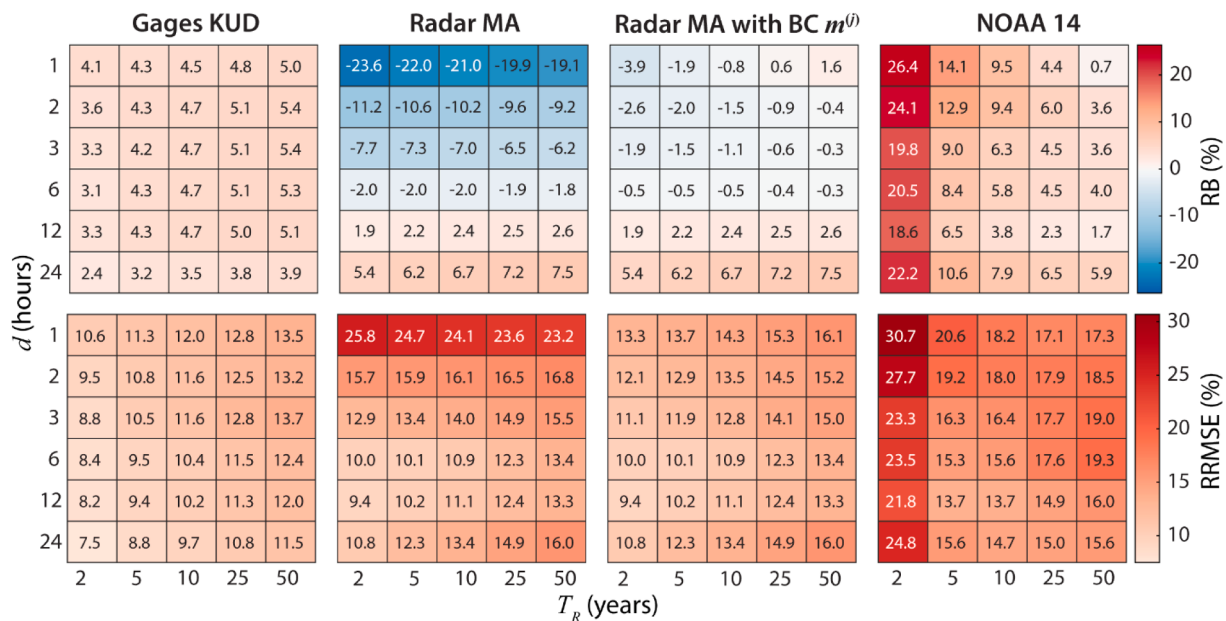


Fig. 8. Heatmaps of error metrics (RB, top; RRMSE, bottom) between extreme P quantiles computed with different estimation methods (see text for details) and at-site estimates with the GEV distribution at 87 gages with ≥ 30 years of observations. The error metrics are reported as a function of P duration, d , and return period, T_R .

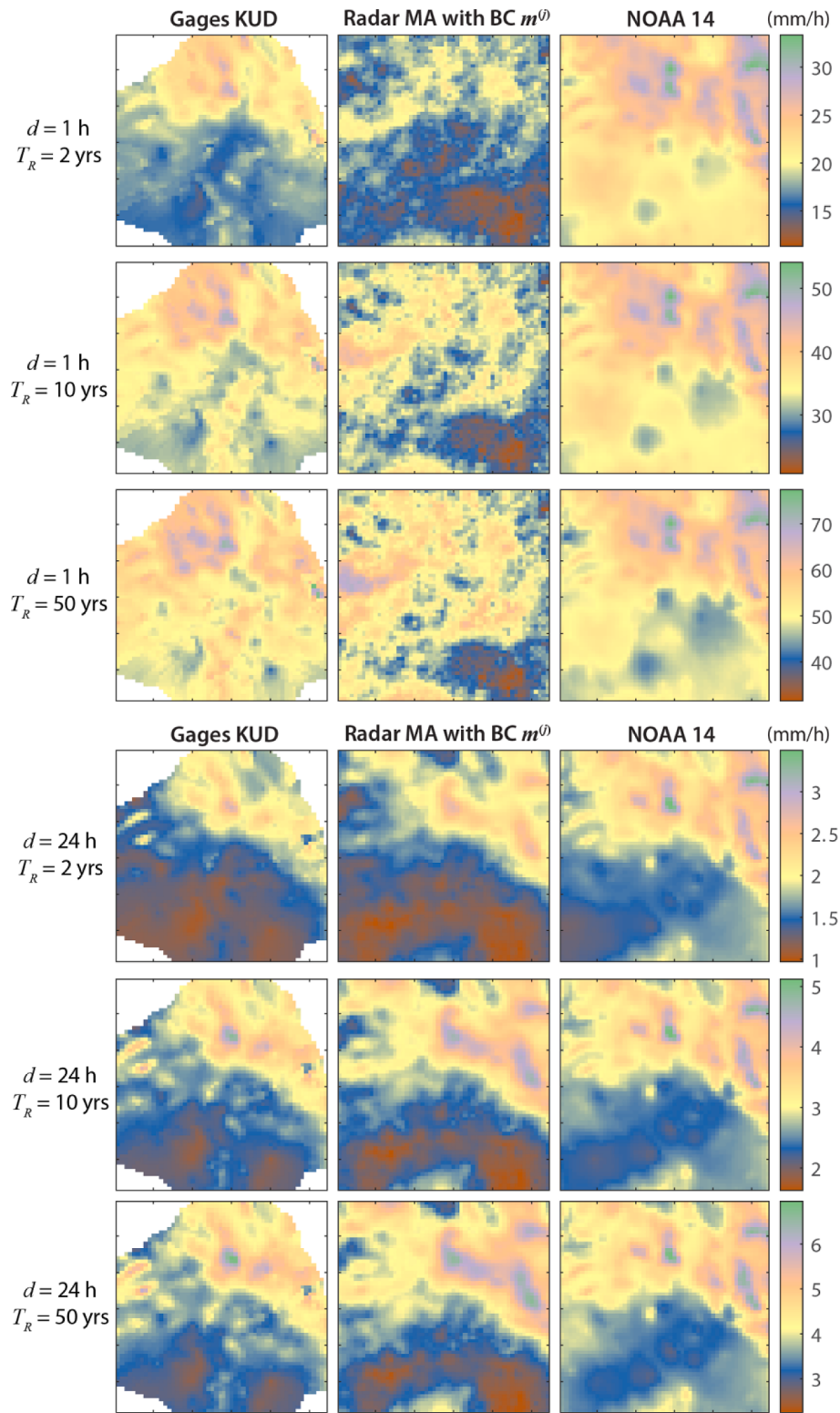


Fig. 9. Maps of extreme P quantiles for $T_R = 2, 10$, and 50 years at $d = 1$ h and 24 h obtained from gage records (Gage KUD), radar QPEs (Radar MA with BC $m^{(j)}$), and NOAA14.

coverages in the world, as well as PFEs from NOAA 14. The conclusions of the study are as follows:

1. The GEV was found to be a suitable distribution to model APM series from gage records and radar QPEs across multiple durations, d , from 1 h to 24 h.
2. Due to the short sample size of 19 years, at-site estimates of the GEV shape parameter, k , from both gage and radar records exhibited large variations that are physically nonrealistic. This uncertainty was greatly reduced by bias correcting k as a function of the sample size through the empirical relations proposed for the U.S. by [Carney \(2016\)](#). For all durations, the bias corrected k was found to be

nonnegative in the great majority of the cases, i.e., the distribution of APM is heavy tailed. Since the locations with negative k were randomly placed, k was set equal to 0 to avoid the existence of an unrealistic upper limit for P .

3. An approach for the spatial frequency analysis of extreme P was designed that combines the index-flood (or index-rainfall) method and techniques for spatially interpolating (for the gages) and smoothening (for the radar) the GEV parameters. This approach allowed further reducing the statistical uncertainty and obtaining P quantile estimates that vary seamlessly in space.
4. The use of the moving average (MA) permitted smoothening of the spatial variability and possible artificial clusters found for k estimated from the radar. For consistency, this method was also used to smoothen the radar-derived scale and location GEV parameters; however, for these two parameters, the choice of the interpolation technique had fewer impacts on the outcome. Kriging for uncertain data (KUD) was instead used for the interpolation of all GEV parameters from the gages to better account for their high sampling uncertainty; MA was also applied to interpolate the gage-derived parameters, but the resulting patterns were found to be unrealistic (not shown).
5. In this study region, the index-rainfall, $m^{(j)}$, is moderately to strongly linked to elevation as d increases. Kriging with external drift (KED) and KUD were the best interpolation/smoothening techniques for capturing this effect when generating maps from gages and radar, respectively; the stronger smoothing effect of MA for the radar resulted instead in fields that did not exhibit significant small-scale variability and the signature of orography (not shown). The values of $m^{(j)}$ from the radar were negatively biased compared to the gages at low d 's, which are dominated by localized storms that have smaller rain rates when averaged over the 16-km² radar pixels. However, the patterns of $m^{(j)}$ generated from the radar appeared the most realistic, addressing the limitations of those obtained from sparse gage information, including NOAA 14 which relies on a much lower number of gages at sub-daily durations.
6. After removing the bias in $m^{(j)}$ for low d 's, the spatial frequency analysis of radar QPEs reproduced at-site extreme P quantiles from long-term gage records with comparable or, in some cases, better performance than the spatial analysis of gage records with the sample size of 19 years. The performance was also better than that of PFEs from NOAA 14 which are routinely used for infrastructure design. Moreover, the patterns of extreme P quantiles generated from radar QPEs were more realistic than those generated by interpolating sparse gage information.

While focused on central Arizona, the analyses presented in this work provide useful methodological and practical insights to incorporate radar QPEs into the generation of improved IDF relations in the U.S. and other places of the world where radar products are available. To bias

correct the GEV shape parameter, the global empirical relation of Papalexiou and Koutsoyiannis (2013) could be used for the daily duration, while analyses of rain gage records at subdaily resolutions will be needed to refine the parameterization of the relations proposed by Carney (2016) and valid for $d < 24$ h. Future work will be devoted to designing a methodology of spatial frequency analysis that merges information from radar QPEs and gage records, building upon recent relevant studies (e.g., Benoit, 2021; Cuccoli et al., 2020; Ochoa-Rodriguez et al., 2019). Moreover, it will be interesting from both the scientific and practical perspectives to compare the performance of this approach based on the extreme value theory with novel alternative methods that use non-asymptotic extreme value models (see, e.g., Marra et al. (2022)).

CRediT authorship contribution statement

Nehal Ansh Srivastava: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Giuseppe Mascaro:** Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing – original draft, Visualization, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Nehal Ansh Srivastava reports financial support was provided by National Science Foundation.

Data availability

I have shared the link to the data used in the paper in the Acknowledgments

Acknowledgments

We thank the eponymous reviewer, Dr. Francesco Marra, and an anonymous reviewer for their comments that helped to improve the quality of the manuscript. This work has been supported by the National Science Foundation (NSF) award: “SCC: Community-Based Automated Information for Urban Flooding” (award #1831475) and by the National Institute of Standards and Technology (NIST) award “Assessing the Utility of Safe-to-Fail Design to Improve Climate Hazards Resilience of Interdependent Infrastructure Systems” (award #70NANB22H056). The authors thank Stephen D. Waters from the Flood Control District of Maricopa County for providing the rainfall data of the network. FCDMC data are also available at the website <https://www.fcd.maricopa.gov/625/Rainfall-Data>.

Appendix A

The spatial interpolation techniques used in this study are based on the well-known method of ordinary Kriging (OK), which is briefly illustrated next. Let θ be the variable to be interpolated. OK provides a best linear unbiased estimation (BLUE) of θ at a given point based on observations at close locations. The method quantifies the spatial dependence of θ through the semivariogram, $\gamma(h)$, defined as:

$$\gamma(h) = \gamma_{ij} = \frac{1}{2} E[\theta_i - \theta_j]^2 \quad (A1)$$

where θ_i and θ_j are the observations at two locations i and j at a distance h . To apply OK, the empirical semivariogram is fitted to an analytical model, such as the exponential, spherical, Gaussian, logarithmic, and power models, among others (Goovaerts, 2000b). The estimation at an unmonitored location, $\hat{\theta}$, is a linear combination of the observations at M neighboring sites:

$$\hat{\theta} = \sum_{k=1}^M \lambda_k \cdot \theta_k \quad (A2)$$

where λ_k is a site-specific weight. The weights λ_k are obtained by solving the system of $(M + 1)$ equations:

$$\begin{cases} \sum_{j=1}^M \lambda_j \gamma_{ij} - \mu_0 = \gamma_{i,0} & i = 1, 2, \dots, M \\ \sum_{j=1}^M \lambda_j = 1 \end{cases} \quad (\text{A3})$$

where μ_0 is the Lagrange parameter accounting for the constraint on the weights, γ_{ij} is the semivariogram value between sites i and j , and $\gamma_{i,0}$ is the semivariogram value between site i and the location where the estimate is made.

By solving the system of equations (A3), the empirical estimates of OK at the locations with available observations are exactly equal to the measured values. This implies assuming zero variance for the target variable, which is an undesirable property when the variable exhibits large uncertainty, as in the case of the index-rainfall and the GEV parameters (particularly, the shape parameter). We addressed this limitation of OK by adopting Kriging with uncertain data (KUD), a technique that introduces nonzero variance at the measuring sites. KUD was first developed by de Marsily (1986) for homoscedastic fields and later expanded by Mazzetti and Todini (2008) to account for heteroscedasticity. This method was recently used by Deidda et al. (2021) to interpolate the GEV parameters from gage observations; for simplicity, the same notation of these authors is adopted here. As for OK, equation (A2) is used to estimate θ at a given site; however, the weights λ_k are obtained by solving the system of equations (A3) where the semivariogram values are modified to account for the uncertainty as follows:

$$\begin{cases} \gamma_{ij}^* = \gamma_{ij} + \frac{\sigma_i^2 + \sigma_j^2}{2} & i, j = 1, 2, \dots, M \text{ and } i \neq j \\ \gamma_{ij}^* = \gamma_{ij} & i = j \\ \gamma_{i,0}^* = \gamma_{i,0} + \frac{\sigma_i^2}{2} & i = 1, 2, \dots, M \end{cases} \quad (\text{A4})$$

In Eq. (A4), γ^* is the modified semivariogram under KUD, σ_i^2 is the measuring variance of θ at location i , while the other symbols have been previously defined. In our study, the target variable θ is either the shape (k) or scale (σ) GEV parameters or the index-rainfall ($m^{(j)}$) (see Table 2). To estimate σ_i^2 for each of these parameters, we performed Monte Carlo simulations where (1) 1000 GEV variates were randomly generated at the measuring sites using at-site parameter estimates and the same sample size of the APM observations ($N = 19$), (2) the GEV parameters were re-estimated on each synthetic sample using the method of L-moments, and (3) the variance of the re-estimated parameters (k , σ , or $m^{(j)}$) was used to estimate σ_i^2 .

As shown in the main text, the observations of $m^{(j)}$ from the rain gage records have moderate-to-strong linear correlations with elevation. To incorporate this information into the spatial interpolation of $m^{(j)}$ from the gages, we used Kriging with external drift (KED; Goovaerts, 2000b). In this case, equation (A2) is used again to estimate the target variable at a given point, but the weights λ_k are estimated by solving the system of $(M + 2)$ linear equations:

$$\begin{cases} \sum_{j=1}^M \lambda_j \gamma_{ij} + \mu_0 + \mu_1 \cdot z_i = \gamma_{i,0} & i = 1, 2, \dots, M \\ \sum_{j=1}^M \lambda_j = 1 \\ \sum_{j=1}^M \lambda_j \cdot z_j = 1 \end{cases} \quad (\text{A5})$$

where μ_0 and μ_1 are Lagrange parameters accounting for weight constraints, z_i is the elevation at site i , and the other symbols have been previously defined.

The logarithmic, Gaussian, and power analytical models were found to best fit the empirical semivariograms for k , σ , and $m^{(j)}$, respectively. The number of neighboring sites used for the interpolation was determined through leave-one-out cross validation. For each of the 204 gages, we applied the interpolation method using the other 203 gages and estimated the parameter by selecting: (1) the gages within a radius R from 5 to 60 km, and (2) the closest M gages with M ranging from 1 to 20. We then computed the RRMSE between observed and estimated parameter values across all gages. Results for $d = 1$ and 24 h are reported in Fig. S4 of the Supporting Material, which shows that the mean RRMSE becomes constant for $R \geq 40$ km and $M \geq 12$ gages. Therefore, we performed the interpolation using all gages within $R = 40$ km if their number was ≥ 12 . If this number was instead lower than 12, we used the closest 12 gages. If no gage was found within the 40-km radius, the interpolation was not performed; this occurred at the corners of the domain (see Figs. 4 and 7).

The spatial smoothening of the GEV parameters estimated from the radar QPEs was conducted by computing the moving average (MA), which is based on equation (A2) where the weights are the same for all sites, i.e., $\lambda_k = 1/M$. In this case, we included all pixels within a radius of 40 km.

Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhydrol.2023.129902>.

References

- Adams, D.K., Comrie, A.C., 1997. The North American Monsoon. *Bull. Am. Meteorol. Soc.* 78 (10), 2197–2213. [https://doi.org/10.1175/1520-0477\(1997\)078<2197:TNAM>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2197:TNAM>2.0.CO;2).
- Avanzi, F., De Michele, C., Gabriele, S., Ghezzi, A., Rosso, R., 2015. Orographic signature on extreme precipitation of short durations. *J. Hydrometeorol.* 16 (1), 278–294. <https://doi.org/10.1175/JHM-D-14-0063.1>.
- Balling, R.C., Brazel, S.W., 1986. Diurnal variation in Arizona Monsoon precipitation frequencies. *Am. Meteorol. Soc.*, 115(Mon. Weather Rev.), 342–346. [https://doi.org/https://doi.org/10.1175/1520-0493\(1987\)115<0342:DVIAMP>2.0.CO;2](https://doi.org/https://doi.org/10.1175/1520-0493(1987)115<0342:DVIAMP>2.0.CO;2).

- Barry, R.G., Chorley, R.J., 1998. Atmosphere, weather, and climate. 409. https://books.google.com/books/about/Atmosphere_Weather_and_Climate.html?id=p-5XswEACAAJ.
- Benoit, L., 2021. Radar and rain gauge data fusion based on disaggregation of radar imagery. *Water Res.* 57 (2).
- Blanchet, J., Ceresetti, D., Molinié, G., Creutin, J.D., 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., Martin, D., Lin, B., Parzybok, T., Yekta, M., Riley, D., 2019. NOAA Atlas 14 Volume 1: precipitation-frequency Atlas of the United States. In: NOAA, National Weather Service, Silver Spring, MD, Vol. 1.
- Burlando, P., Rosso, R., 1996. Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *J. Hydrol.* 187 (1–2), 45–64. [https://doi.org/10.1016/S0022-1694\(96\)03086-7](https://doi.org/10.1016/S0022-1694(96)03086-7).
- Carney, M.C., 2016. Bias correction to GEV shape parameters used to predict precipitation extremes. *J. Hydrol. Eng.* 21 (10), 04016035. [https://doi.org/10.1061/\(asce\)jhe.1943-5584.0001416](https://doi.org/10.1061/(asce)jhe.1943-5584.0001416).
- Claps, P., Ganora, D., Mazzoglio, P., 2022. Chapter 11 - Rainfall regionalization techniques, pp. 327–350.
- Coles, S.G., Dixon, M.J., 1999. Likelihood-based inference for extreme value models. *Extremes* 2(1), 5–23. <https://doi.org/10.1023/A:1009905222644>.
- Coles, S., Pericchi, L.R., Sisson, S., 2003. A fully probabilistic approach to extreme rainfall modeling. *J. Hydrol.* 273 (1–4), 35–50. [https://doi.org/10.1016/S0022-1694\(02\)00353-0](https://doi.org/10.1016/S0022-1694(02)00353-0).
- Cuccoli, F., Facheris, L., Antonini, A., Melani, S., Baldini, L., Member, S., 2020. Weather radar and rain-gauge data fusion for quantitative precipitation estimation: two case studies 58 (9), 6639–6649.
- Dalrymple, T., 1960. Flood-frequency analyses, Manual of Hydrology: Part 3. Water Supply Paper. <https://doi.org/10.3133/WSP1543A>.
- Daly, C., Taylor, G., Gibson, W., 1997. The Prism approach to mapping precipitation and emperature. In: 10th AMS Conference on Applied Climatology, 1, 1–4. <http://citeserx.ist.psu.edu/viewdoc/download?doi=10.1.1.730.5725&rep=rep1&type=pdf>.
- de Marsily, G., 1986. Quantitative hydrogeology : groundwater hydrology for engineers. Academic Press.
- Deidda, R., Hellies, M., Langousis, A., 2021. A critical analysis of the shortcomings in spatial frequency analysis of rainfall extremes based on homogeneous regions and a comparison with a hierarchical boundaryless approach. *Stoch. Env. Res. Risk A.* 35 (12), 2605–2628. <https://doi.org/10.1007/s00477-021-02008-x>.
- Du, J., 2011. EOL data archive – NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. <https://doi.org/10.5065/D6PGIQQD>.
- Favors, J.E., Abatzoglou, J.T., 2013. Regional surges of monsoonal moisture into the southwestern United States. *Mon. Weather Rev.* 141 (1), 182–191. <https://doi.org/10.1175/MWR-D-12-00037.1>.
- Fitzgerald, D.L., 1989. Single station and regional analysis of daily rainfall extremes. *Stochastic Hydrol. Hydraul.* 3 (4), 281–292.
- Formetta, G., Marra, F., Dallan, E., Zaramella, M., Borga, M., 2022. Differential orographic impact on sub-hourly, hourly, and daily extreme precipitation. *Adv. Water Resour.* 159, 104085. <https://doi.org/10.1016/J.ADVWATRES.2021.104085>.
- Garfin, G., Jardine, A., Merideth, R., Black, M., LeRoy, S., 2013. Assessment of climate change in the Southwest United States: A report prepared for the national climate assessment. In: Assessment of Climate Change in the Southwest United States: A Report Prepared for the National Climate Assessment. <https://doi.org/10.5822/978-1-61091-484-0>.
- Ghebreyesus, D.T., Sharif, H.O., 2021. Development and assessment of high-resolution radar-based precipitation Intensity–Duration–Curve (IDF) curves for the state of Texas. *Remote Sens. (Basel)* 13 (15). <https://doi.org/10.3390/rs13152890>.
- Goovaerts, P., 2000a. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* 228 (3), 442–445.
- Goovaerts, P., 2000b. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* 228 (1–2), 113–129. [https://doi.org/10.1016/S0022-1694\(00\)00144-X](https://doi.org/10.1016/S0022-1694(00)00144-X).
- Gubareva, T.S., Gartsman, B.I., 2010. Estimating distribution parameters of extreme hydrometeorological characteristics by L-moments method. *Water Resour.* 37 (4), 437–445.
- Guttman, N.B., Hosking, J.R.M., Wallis, J.R., 1993. Regional precipitation quantile values for the continental United States computed from L-moments. *J. Clim.* 6 (12), 2326–2340. [https://doi.org/10.1175/1520-0442\(1993\)006<2326:RPQVFT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<2326:RPQVFT>2.0.CO;2).
- Hjelmstad, A., Shrestha, A., Garcia, M., Mascaro, G., 2021. Propagation of radar rainfall uncertainties into urban pluvial flood modeling during the North American monsoon. *Hydrol. Sci. J.* 66 (15), 2232–2248. <https://doi.org/10.1080/02626667.2021.1980216>.
- Hosking, J.R.M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* 52 (1), 105–124. <https://doi.org/10.1111/J.2517-6161.1990.TB01775.X>.
- Hosking, J.R.M., 1992. Moments or L moments? An example comparing two measures of distributional shape. *Am. Stat.* 46 (3), 186. <https://doi.org/10.2307/2685210>.
- Hosking, J.R.M., Wallis, J.R., Wood, E.F., 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27 (3), 251–261. <https://doi.org/10.1080/00401706.1985.10488049>.
- Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. *Water Resour. Res.* 29 (2), 271–281. <https://doi.org/10.1029/92WR01980>.
- Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis: An Approach Based on L-Moments. In: Regional Frequency Analysis. Cambridge University Press, p. 242. <https://doi.org/10.1017/CBO9780511529443>.
- Koutsoyiannis, D., 2004a. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation / Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: I. Recherche théorique. *Hydrol. Sci. J.* 49 (4) <https://doi.org/10.1623/hysj.49.4.575.54430>.
- Koutsoyiannis, D., 2004b. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records / Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations. *Hydrol. Sci. J.* 49 (4) <https://doi.org/10.1623/hysj.49.4.591.54424>.
- Koutsoyiannis, D., Kozonis, D., Manetas, A., 1998. Intensity–Duration–Frequency Relationships 206, 118–135.
- Koutsoyiannis, D., Langousis, A., 2011. Precipitation. *Treatise Water Sci.* 2, 27–77. <https://doi.org/10.1016/B978-0-444-53199-5.00027-0>.
- MacDonald, G.M., 2010. Water, climate change, and sustainability in the southwest. *Proc. Natl. Acad. Sci. U.S.A.* 107 (50), 21256–21262.
- Madsen, H., Mikkelsen, P.S., Rosbjerg, D., Harremoës, P., 2002. Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series statistics. *Water Resour. Res.* 38(11), 21–1. <https://doi.org/10.1029/2001WR001125>.
- Madsen, H., Pearson, C.P., Rosbjerg, D., 1997. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional modeling. *Water Resour. Res.* 33 (4), 759–769. <https://doi.org/10.1029/96WR03849>.
- Marra, F., Armon, M., Morin, E., 2022. Coastal and orographic effects on extreme precipitation revealed by weather radar observations. *Hydrol. Earth Syst. Sci.* 26, 1439–1458. <https://doi.org/10.5194/hess-26-1439-2022>.
- Marra, F., Morin, E., 2015. Use of radar QPE for the derivation of Intensity–Duration–Frequency curves in a range of climatic regimes. *J. Hydrol.* 531, 427–440. <https://doi.org/10.1016/j.jhydrol.2015.08.064>.
- Mascaro, G., 2017. Multiscale spatial and temporal statistical properties of rainfall in central Arizona. *J. Hydrometeorol.* 18 (1), 227–245. <https://doi.org/10.1175/JHM-D-16-0167.1>.
- Mascaro, G., 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>.
- Mascaro, G., 2020. Comparison of local, regional, and scaling models for rainfall intensity–duration–frequency analysis. *J. Appl. Meteorol. Climatol.* 59 (9), 1519–1536. <https://doi.org/10.1175/JAMC-D-20-0094.1>.
- Mascaro, G., Deidda, R., Hellies, M., 2013. On the nature of rainfall intermittency as revealed by different metrics and sampling approaches. *Hydrol. Earth Syst. Sci.* 17 (1), 355–369. <https://doi.org/10.5194/hess-17-355-2013>.
- Mazzetti, C., Todini, E., 2008. Combining weather radar and raingauge data for hydrologic applications. *Flood Risk Manag.: Res. Pract.* 1345–1348 <https://doi.org/10.1201/9780203883020.CH159>.
- Mazzoglio, P., Butera, I., Alvioli, M., Claps, P., 2022. The role of morphology in the spatial distribution of short-duration rainfall extremes in Italy. *Hydrol. Earth Syst. Sci.* 26, 1659–1672. <https://doi.org/10.5194/hess-26-1659-2022>.
- McGraw, D., Nikolopoulos, E.I., Marra, F., Anagnostou, E.N., 2019. Precipitation frequency analyses based on radar estimates: an evaluation over the contiguous United States. *J. Hydrol.* 573 (November 2018), 299–310. <https://doi.org/10.1016/j.jhydrol.2019.03.032>.
- Modarres, R., Sarhadi, A., 2011. Statistically-based regionalization of rainfall climates of Iran. *Global Planet. Change* 75 (1–2), 67–75. <https://doi.org/10.1016/J.GLOPLACHA.2010.10.009>.
- Nelson, B.R., Prat, O.P., Seo, D.J., Habib, E., 2016. Assessment and implications of NCEP stage IV quantitative precipitation estimates for product inter-comparisons. *Weather Forecast.* 31 (2), 371–394. <https://doi.org/10.1175/WAF-D-14-00112.1>.
- NOAA, 2023. NOAA National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate. Disasters. <https://doi.org/10.25921/stkw-7w73>.
- Ochoa-Rodriguez, S., Wang, L.-P., Willems, P., Onof, C., 2019. A review of radar-rain gauge data merging methods and their potential for urban hydrological applications. *Water Resour. Res.* 55 (8), 6356–6391.
- Overeem, A., Buishand, T.A., Holleman, I., 2009. Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resour. Res.* 45 (10), 1–15. <https://doi.org/10.1029/2009WR007869>.
- Papalexiou, S.M., Koutsoyiannis, D., 2013. Battle of extreme value distributions: a global survey on extreme daily rainfall. *Water Resour. Res.* 49 (1), 187–201. <https://doi.org/10.1029/2012WR012557>.
- Peel, M.C., Wang, Q.J., Vogel, R.M., McMahon, T.A., 2009. The utility of L-moment ratio diagrams for selecting a regional probability distribution. <https://doi.org/10.1080/02626660109492806>.
- Requena, A.I., Burn, D.H., Coulibaly, P., 2019. Pooled frequency analysis for intensity–duration–frequency curve estimation. *Hydrol. Process.* 33 (15), 2080–2094. <https://doi.org/10.1002/HYP.13456>.
- Rosenzweig, B.R., McPhillips, L., Chang, H., Cheng, C., Welty, C., Matsler, M., Iwaniec, D., Davidson, C.I., 2018. Pluvial flood risk and opportunities for resilience. *Wiley Interdiscip. Rev. Water* 5 (6), e1302.
- Rossi, M.W., Anderson, R.S., Anderson, S.P., Tucker, G.E., 2020. Orographic controls on subdaily rainfall statistics and flood frequency in the Colorado Front Range, USA. *Geophys. Res. Lett.* 47 (4) <https://doi.org/10.1029/2019GL085086>.
- Schaefer, M.G., 1990. Regional analyses of precipitation annual maxima in Washington State. *Water Resour. Res.* 26 (1), 119–131. <https://doi.org/10.1029/WR0261001P00119>.

- Sheppard, P.R., Comrie, A.C., Packin, G.D., Angersbach, K., Hughes, M.K., 2002. The climate of the US Southwest. *Climate Res.* 21 (3), 219–238. <https://doi.org/10.3354/cr021219>.
- Smith, E.P., 2002. An Introduction to Statistical Modeling of Extreme Values. *Technometrics* 44 (4), 397.
- Tyralis, H., Langousis, A., 2019. Estimation of intensity–duration–frequency curves using max-stable processes. *Stoch Environ Res Risk Assess* 33 (1), 239–252.
- Vogel, R.M., Fennessey, N.M., 1993. L moment diagrams should replace product moment diagrams. *Water Resour. Res.* 29 (6), 1745–1752. <https://doi.org/10.1029/93WR00341>.