Improving Extreme Precipitation Frequency Analysis in Southwestern U.S. with Radar Product

by

Nehal Ansh Srivastava

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved June 2022 by the
Graduate Supervisory Committee:

Giuseppe Mascaro, Chair
Mikhail Chester
Margaret Garcia
Simon Michael Papalexiou

ARIZONA STATE UNIVERSITY

August 2022

ABSTRACT

Weather radars provide quantitative precipitation estimates (QPEs) with seamless spatial coverage that can complement limitations of sparse rain gage measurements, including those affecting intensity-duration-frequency (IDF) relations used for infrastructure design. The goal of this M.S. thesis is to assess the ability of 4-km, 1-h QPEs from the Stage IV analysis of the Next-Generation Radar (NEXRAD) network to reproduce the statistics of extreme precipitation (P) in central Arizona, USA, using a dense network of 257 rain gages as reference. The generalized extreme value (GEV) distribution is used to model the frequency of annual P maximum series observed at gages and radar pixels for durations, d, from 1 to 24 h. Estimates of P quantiles from radar QPEs are negatively biased (-20% – -30%) for d = 1 h. The bias tends to 0 and errors are small for $d \geq 6$ h, independently of the return period. The presence of scaling for the GEV location and scale parameters, needed to apply IDF scaling models, was found for both radar and gage products. Regional frequency analysis methods combined with bias correction of the GEV shape parameter allow reducing the statistical uncertainty and providing seamless spatial distribution of P quantiles at daily and subdaily durations that address limitations of current IDF relations in southwestern U.S. based on NOAA Atlas 14.

ACKNOWLEDGMENTS

This thesis would not have been possible without the support and encouragement of many people.

First, I would like to thank members of my graduate supervisory committee for taking out the time to supervise my work. I extend my gratitude to my supervisor, Prof. Giuseppe Mascaro for his constant support and guidance at each step of this process and for always being patient and encouraging. I further convey my regards to my fellow research group members for their continuous support during this period.

Finally, I would like to thank my family, specially my parents and grandparents for always believing in me.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

Extreme precipitation (P) is one the most impactful natural hazards causing property damages and loss of lives. For example, the National Centers for Environmental Information (NCEI) has reported an economic loss in 2021 of $20.8 billion related to severe storm events in the United States (U.S.) and a sharp increase in the number of these events from about <20 events observed prior to the 2000s per decade to 75 observed in the last decade alone (*NCEI, 2022*). Intense P storms are particularly impactful in urban regions, where population density and the concentration of economic activities are high. In cities, intense P storms could lead to flash floods and pluvial flooding conditions due to the high runoff coefficients and fast response times of urban basins (Rosenzweig et al., 2018). Ahmadalipour & Moradkhani (2019) summarized the spatial occurrence and impacts of flash floods that occurred during 1996-2017 in the U.S. and found that the southwestern U.S. to be a hot spot compared to other parts of the country.

The design and maintenance of infrastructure systems against extreme P events are based on the estimation of the so-called design storm, which is the P intensity, $q(T_R, d)$, over a duration, d, associated with a given frequency of occurrence quantified through the return period, T_R (in years). This information is synthetized by intensity-duration-frequency (IDF) curves (Burlando & Rosso, 1996b; Koutsoyiannis et al., 1998; Madsen et al., 2002; Requena et al., 2019; Tyralis & Langousis, 2018). In all U.S. states except ID, MT, OR, WA and WY, IDF curves are provided by the Volumes of the National Oceanic and Atmospheric Administration (NOAA) Atlas 14 (Bonnin et al.,

2019) for d ranging from 5 min to 60 day and T_R from 1 to 1000 years. The most common approach to generate IDF curves requires the (1) frequency analysis of P records observed at networks of rain gages, and (2) the application of techniques to spatially interpolate the point information derived at the gages over larger regions (Blanchet et al., 2016; Fitzgerald, 1989; Guttman et al., 1993; Madsen et al., 1997; Mascaro, 2020; Modarres & Sarhadi, 2011; Schaefer, 1990). The first task is usually conducted with the block maxima method, which involves fitting an appropriate probability distribution to annual precipitation maxima (APM) series. According to the extreme value theory (Coles, 2001), the asymptotic distribution of block maxima is the Generalized Extreme Value (GEV), and empirical evidence has shown that this distribution captures well the frequency of APM at different durations (Blanchet et al., 2016b; Coles et al., 2003; Coles & Dixon, 1999; Deidda et al., 2021; Gubareva & Gartsman, 2010; Koutsoyiannis, 2004a, 2004b; Koutsoyiannis & Langousis, 2011; Mascaro, 2020; Papalexiou & Koutsoyiannis, 2013).

The most popular method adopted to extrapolate information on extreme P frequency at un-gaged locations is based on the index-flood technique, which is commonly referred to as regional frequency analysis (Dalrymple, 1960; Hosking & Wallis, 1997). This method is based on the idea that hydrological records (in our case, APM) observed at different sites exhibit similar statistical properties if they belong to the same "homogeneous region" identified through proper statistical tests. APM series observed within a homogeneous region can then be pooled together and a single distribution (e.g., the GEV) can be used to characterize their frequency after they have been standardized by their mean or another index statistic. This allows reducing the

uncertainty of records with small sample size. Local estimates of $q(T_R, d)$ at un-gaged sites are then obtained by multiplying the quantiles of the single distribution by the local index statistic (e.g., the mean APM). IDF curves of NOAA Atlas 14 are obtained by applying the index-flood method with the GEV distribution.

A critical source of uncertainty of IDF curves is the density of the rain gage network. When the density is low, local features of extreme P may not be properly captured, leading to inaccurate estimates of $q(T_R, d)$ and, in turn, to the possible over- or under-sizing of infrastructure. Moreover, Deidda et al. (2021) have recently showed that regionalization techniques based on homogeneous regions have the drawback of causing abrupt shifts in the values of P quantiles along the boundaries of contiguous regions. These authors proposed a method involving the application of a hierarchical approach where geo-statistical interpolation is used to obtain continuous fields of parameter estimates. While promising, the efficacy of this technique relies on the resolution of the gage network.

An additional source of uncertainty of IDF curves is the lack of subdaily P observations. A strategy identified in literature to address this issue is the application of scaling models for IDF relations (Bara et al., 2010; Blanchet et al., 2016b; Borga et al., 2005; Mascaro, 2020; Sane et al., 2018; Yu et al., 2004). These models are based on the evidence of temporal scaling properties for P that extends to the parameters of the distribution characterizing the APM series. Scaling models can then be utilized to obtain P quantiles at subdaily resolution starting from information at daily scale where P is more widely available.

Quantitative precipitation estimates (QPEs) derived from weather radars characterize the spatial variability of P at resolutions of a few km and ≤1 h and, in certain regions of the world, they have been collected for more than a decade. Therefore, these areal P products have the potential to improve IDF curves by addressing the limitations caused by sparse rain gage networks and the lack of subdaily observations. Recent efforts have started to explore the utility of radar QPEs for the frequency analysis of extreme P. In a study in the Netherlands, Overeem et al. (2009) fitted the GEV distribution to records of 11 years of APM extracted from radar QPEs and found that the location parameter derived from the radar has a mean biases of -14% compared to estimates from the gages at d = 1 h. Marra & Morin (2015) developed IDF curves with radar products from 23 years data records in parts of Israel and compared the IDF curves from radar and gages and concluded that APM quantiles are less biased at d = 4 h compared to shorter durations of 20 min and 1 h. Studies have been also recently carried out in U.S., where the National Centers for Environmental Prediction (NCEP) Stage IV analyses provide gage-corrected QPEs at 4 km,1 h resolution in the Conterminous U.S. (CONUS) since 2002. Ghebreyesus & Sharif (2021) used 19 years of Stage IV radar observations to derive IDF curves in Texas and validated them using NOAA Atlas 14 precipitation frequency estimates (PFEs). They found bias in radar quantiles ranging within -/+27% at all durations with a lower range at higher d (24 h). In another study, McGraw et al. (2019) derived IDF relations using records of hourly P from (1) 539 rain gages covering the U.S., and (2) Stage IV QPEs in the co-located pixels. These authors findings grouped the rain gages based on the Köppen-Geiger climate classification and found that the discrepancies between two products is climate-dependent. While results of McGraw et al.

(2019) are very promising, they are based on single, at-site comparisons at each location which have low statistical significance given the short records of Stage IV QPEs. Additional studies based on a through comparison of IDF curves from radar and gages using dense network of rain gages in a region are then desirable to complement and expand these initial efforts.

The main goal of this M.S. thesis is to assess the ability of Stage IV radar QPEs to characterize the frequency of extreme P at different time aggregations to ultimately improve IDF relations in southwestern U.S. To address this goal, we use the high-density network of rain gages installed since the 1980s by the Flood Control District of Maricopa County (FCDMC) in central Arizona. We first show that the GEV is an appropriate distribution to model the frequency of APM series observed at 257 gages and radar pixels for d = 1 to 24 h. We then compute at-site estimates of GEV parameters and related P quantiles for the two products and compared them with values obtained through different regional frequency analysis methods based in part on Deidda et al. (2021). To further address the uncertainty due to the limited sample size (19 years), we bias correct the GEV shape parameter using recent empirical evidence by (Carney, 2016; Papalexiou & Koutsoyiannis, 2013). We also investigate the presence of scaling properties for the GEV location and scale parameters, needed to apply IDF scaling models. After gaining confidence on radar products in central Arizona, we show that the use of regional frequency analysis of extreme P and scaling IDF models based on Stage IV QPEs provide a seamless spatial distribution of P quantiles at daily and subdaily durations that address limitations of current IDF relations in southwestern U.S. based on NOAA Atlas 14.

CHAPTER 2

BACKGROUND

2.1. Study Area

The comparison of radar QPEs with gage records was performed in a 216 × 232 km^2 area in central Arizona centered around the Phoenix Metropolitan region. Figure 1a outlines the study area in southwestern U.S., with the background map showing the 2-year 1-hr PFEs from NOAA Atlas Volume 1. As shown in Figure 1b, the study domain includes a large area at low elevation (92–200 m asl), where Phoenix is located, and a mountainous region in the northwest, the Mogollon Rim, where elevation reaches 2577 m asl. Due to its arid/semi-arid climate, this desert area of southwestern U.S. is categorized as the hottest and driest in North America (Garfin et al., 2013; MacDonald, 2010). The average annual rainfall is low (189.5 mm at Phoenix Sky harbor) and temperature (min 15.2 ºC and max 30 ºC) is high. The presence of a quasi-permanent subtropical high pressure ridge guarantees dry and warm conditions for most of the year (Sheppard et al., 2002). Climate variability are largely attributed to the shift between the mid-latitude and subtropical atmospheric circulation regimes, with terrain exerting local controls (Sheppard et al., 2002).

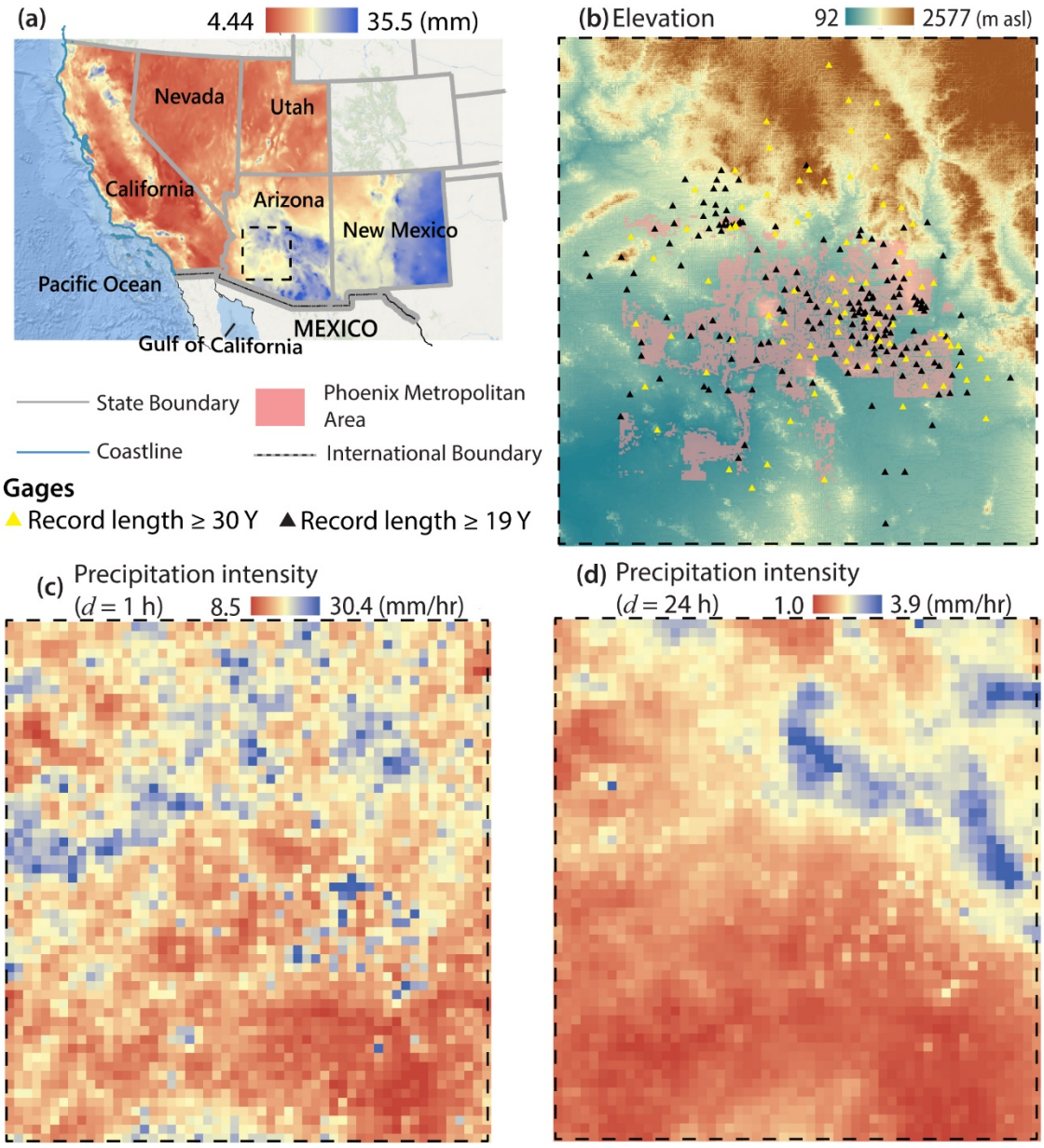

Figure 1. Location of the Study Area in (a) Southwestern U.S. With the Background Map Showing Precipitation Frequency Estimates from NOAA Atlas Volume 1 for d = 1 h and T_R = 2 Years. (b) Digital Elevation Model (from U.S. Geological Survey National Elevation Dataset) of the 216×232 km^2 Study Domain in Central Arizona Centered Around the Phoenix Metropolitan Area, along with the Rain Gage Network of the FCDMC. (c)-(d) Average Maximum Precipitation Intensity in the Study Domain for (c) d = 1 h and (d) d = 24 h from 19 Years of Stage IV QPEs. In (a) the States of NOAA Atlas 14 Volume 1 are Labeled.

The precipitation regime is characterized by two seasons with markedly different storm-generating mechanisms. From late fall to early summer (winter: November-

March), prolonged dry conditions are interrupted by storms due to cold fronts with moisture transported by westerly flows from the Pacific. Winter storms tend to last for a few days (Barry & Chorley, 1998), be widespread, and be characterized by relatively low precipitation intensity. These events are considered more important for water resources management due to their ability to recharge aquifers (Sheppard et al., 2002). Several studies have showed that winter storms could be caused by different mechanisms including Pacific/North American (PNA) circulations (Leathers & Palecki, 1992; Redmond & Koch, 1991; Simmons et al., 1983) and Southwestern troughing (Burnett, 2013; Sellers & Hill, 1974; Woodhouse & Meko, 1997). Sellers & Hill (1974) also highlighted the importance of localized orographic effects. Moreover, the magnitude and frequency winter storms has been linked to teleconnection patterns, including the El Niño Southern Oscillation (ENSO) (Cayan & Peterson, 1989; Wallace & Gutzler, 1981; Woodhouse & Meko, 1997; Yarnal & Diaz, 1986) and the Pacific Decadal Oscillation (Mantua et al., 1997; Zhang et al., 1997). While different mechanics are responsible for winter P in the Southwest, ENSO and PNA have been found to be the main cause (Blackmon et al., 1984; Dettinger et al., 1998; Shukla & Wallace, 1983; van Loon & Madden, 1981; van Loon & Rogers, 1981; Yarnal & Diaz, 1986). More recent findings from Svoma and Balling (2009) show that local factors such as vehicular emissions have started to impact and suppress winter precipitation formation.

In early July, the northward advection of water vapor from the Gulf of California and, to a lower extent, Gulf of Mexico (Favors & Abatzoglou, 2013; Sheppard et al., 2002) leads to the onset of the North American monsoon (NAM) (Adams & Comrie, 1997), which last until the end of September. The intrusion of these northward moisture

from the identified sources is driven by midlatitude (Anderson & Roads, 2002; Higgins et al., 2004) and tropical circulation patterns (Corbosiero et al., 2009; Fuller & Stensrud, 2000; Stensrud et al., 1995, 1997). During the NAM, convective thunderstorms with high intensity, short durations (< 1 h), and small spatial extent (a few km^2) occur according to a diurnally modulated cycle (Balling & Brazel, 1986). In particular, the mountainous regions in northern Arizona and southernmost parts of the state experience the maximum precipitation frequencies during daytime, while in the central portion of the state (including the Phoenix Metropolitan area) monsoonal storms are likely to occur at nighttime. These findings were originally identified by Balling & Brazel (1986) by analyzing a few gages in the entire state and were then further conformed and expanded by Svoma (2010) and Mascaro (2017) by analyzing the precipitation records of >200 gages of the FCDMC network. Finally, a prominent characteristic of the precipitation regimes in central Arizona is the dependence on orography, which affects annual and seasonal totals and, to a lower extent, extreme P (Mascaro, 2017, 2018).

2.2. Dataset

2.2.1. Radar

We used radar quantitative precipitation estimates (QPEs) at resolution of 1 h and 4 km from the National Centers for Environmental Prediction (NCEP) Stage IV analyses for the period 2002 to 2020. Stage IV QPEs are generated for the Conterminous United States (CONUS) by mosaicking reflectivity data from the NEXRAD network and adjusting rainfall rates with gauge and satellite observations and manually quality controlled (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 × 232 km^2 domain shown in Figure 1a, and projected into the Universal Transverse Mercator (UTM) Zone 12 N reference. The Stage IV documentation provides a detailed list of the missing data and its causes.

From the radar QPEs, we derived the records of annual P maxima (APM) (in mm/h), i_d, for durations d = 1, 2, 3, 6, 12, 24 and 48 h in each pixel of the study domain. For each year, we extracted the largest P intensity 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_d in the series extracted for the N complete years, and (2) retain (rejected) i_d if in case its rank is above (below) $f N$. As an example, Figures 1c,d show the mean APM for d = 1 and 24 h in the study region derived from Stage IV.

2.2.2. Gages and Operational Design Storms

The ALERT network of rain gages managed by the Flood Control District of Maricopa County (FCDMC) was used as reference to assess the radar QPEs' ability to characterize extreme P statistics. The network started operated in the early 1980s and currently includes 365 gages that monitor P in real-time over an area of about 29,600 km^2 centered around the Phoenix Metropolitan Area (Figure 1b). The gage elevation ranges from 220 to 2325 m MSL, although most (195) gages are installed below 800 m (Mascaro, 2020) the inter-gage distance varies between 0.5 km to 227 km with a median of 70 km. For our analyses, we used 257 gages that were active during the same 19 years

when Stage IV data are available. We also utilized a subset of 91 gages with long-term (≥30 years) observations to investigate the effect of sample size on the results. Figure 1b shows the location 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 in the original form of tipping instants in seconds, which were converted into signals at a given duration *d* according to the procedure suggested in Mascaro et al. (2013). The same procedure illustrated for the radar QPEs was applied to derive the series of APM while accounting for missing data.

To further validate the reliability of the radar-derived extreme P statistics, we also used the precipitation frequency estimates (PFEs) for different return periods and P durations available from the National Oceanic and Atmospheric Administration (NOAA) Atlas 14 (Bonnin et al., 2019). These design storms are utilized in most part of the U.S. as the basis of design and maintenance of infrastructure elements, such as stormwater management system, roadway drainage, culverts, and bridges. The NOAA Atlas 14 maps were obtained in the form of ASCII grids in units of thousands of inches referred to the chosen duration. The maps were then imported in ArcGIS, projected into UTM, and the values in each pixel was converted to mm/h.

CHAPTER 3

METHODOLOGY

3.1. The Generalized Extreme Value (GEV) Distribution

Extreme P was characterized statistically by fitting the generalized extreme value (GEV) distribution to the APM series for the different durations. This distribution was found to be appropriate for this purpose using gage observations in the region by (Mascaro, 2020) and is also used to derive the PFEs in NOAA Atlas 14 through a procedure based on parameter regionalization (Hosking & Wallis, 1997). Here, we tested its applicability to APM series derived from the radar QPEs. The cumulative distribution function (CDF) of the GEV distribution for the random variable $X_d \equiv$ "annual maximum P intensity for a given d" is defined as:

$$F(x|k_d,\mu_d,\sigma_d) = \Pr\{X_d \le x\} = \begin{cases} \exp\left\{-\left(1+k_d\frac{x-\mu_d}{\sigma_d}\right)^{-\frac{1}{k_d}}\right\} & k_d \neq 0 \\ \exp\left\{-\exp\left(-\frac{x-\mu_d}{\sigma_d}\right)\right\} & k_d = 0 \end{cases}, \qquad (1)$$

where $k_d \in (-\infty,+\infty)$ is the shape parameter, $\mu_d \in (-\infty,+\infty)$ is the location parameter, and $\sigma_d \in (0,+\infty)$ is the scale parameter. The location and the scale parameter represent the central tendency and the variance of the GEV distribution, whereas the shape parameter is indicative of the behavior of tails of the distribution. Depending on the value of the shape parameter, the distribution is further categorized as follows.

- *Type I or Gumbel.* This is defined when $k_d = 0$ with domain $-\infty < x < +\infty$. In this case, the right tail of the survival function, $\bar{F}(x) = 1 - F(x)$, exhibits an exponential behavior that corresponds to a decreasing linear relation in the log space.

- *Type II or Fréchet.* This is defined when $k_d > 0$ with domain $\mu_d - \frac{\sigma_d}{k_d} \leq x \leq \infty$. In this case, the tail of $\bar{F}(x)$ is subexponential, which, in the log space, implies that the exceedance probabilities of extreme quantiles show a lower rate of decline than the exponential case. This type of distributions is also denoted as "heavy tailed".
- *Type III or Weibull.* This is defined when $k_d < 0$ with an upper bound $-\infty < x \leq \mu_d - \frac{\sigma_d}{k_d}$ (S. Coles, 2001). The presence of a maximum possible value for x is unlikely when considering positive definite atmospheric variables.

The quantiles of the GEV associated with the annual return period $T_R = 1/[1 - F(x)]$ and duration d, $q(T_R, d)$, are computed by inverting equation (1) as:

$$q(T_R, d) = \begin{cases} \mu_d - \frac{\sigma_d}{k_d}\left\{1 - \left[-\log\left(1 - \frac{1}{T_R}\right)\right]^{-k_d}\right\} & k_d \neq 0 \\ \mu_d - \sigma_d \log\left[-\log\left(1 - \frac{1}{T_R}\right)\right] & k_d = 0 \end{cases} \quad (2)$$

Here, the suitability of the GEV distribution was first verified through the graphical methods of the L-moment ratio diagram (Hosking, 1990). Then, the distribution was fitted to the APM series extracted at the radar pixels and from the gage records using the method of probability weighted moments (PWM) (Hosking et al., 1985). The method of PWM was chosen since it is more robust to outliers when the sample size is small compared to techniques for parameter estimation based on simple moments and maximum likelihood (Hosking et al., 1985; Madsen et al., 1997; McGraw et al., 2019; Sankarasubramanian & Srinivasan, 1999; Vogel & Fennessey, 1993).

3.2. Spatial Frequency Analysis

We characterized the spatial variability of extreme P at a given duration *d* (for simplicity, we will omit the index *d* from the variables defined next) using gage records and radar QPEs through a stepwise technique based on the boundaryless approach of Deidda et al. (2021). This approach, in turn, relies on the index-flood method, whose main steps and equations are summarized next. Let $x^{(i)}$ be the AMP series at a given duration *d* in the *i*-th radar pixel or gage. The sample $x^{(i)}$ is first standardized as:

$$y^{(i)} = x^{(i)}/m^{(i)} \tag{3}$$

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

$$q(T_R)^{(i)} = m^{(i)} \cdot q_y(T_R). \tag{4}$$

If the GEV distribution is used to model the dimensionless variable *y* whose mean is 1, its dimensionless parameters, denoted as *k*, μ^*, and σ^*, are related to each other through the following relationship:

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

where Γ(.) represents the gamma function $\{\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt\}$ 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, instead of relying on homogenous zones where all $y^{(i)}$ samples are pooled together to estimate a single growth curve, it involves the spatial interpolation of the GEV parameters. This method has been adopted here with minor modifications, as summarized in the next steps. Additional methodological details of the steps are provided in the next subsections.

- *Step 1:* The at-site GEV parameters, k, μ^*, and σ^*, are estimated on each dimensionless sample $y^{(i)}$. k is then bias corrected through empirical relations to account for the short sample size.
- *Step 2:* The bias corrected shape parameter k is interpolated spatially into the 4-km radar grid with an appropriate technique. Here, we tested the performance of two methods: the kriging for uncertain data (KUD), used by Deidda et al. (2021), and the simpler moving average (MA).
- *Step 3:* The at-site scale parameter σ^* is re-estimated on the sample $y^{(i)}$, conditioned on the spatially interpolated, bias corrected k. KUD is used to spatially interpolate the re-estimated σ^*.
- *Step 4:* The at-site scale parameter μ^* is re-estimated on the sample $y^{(i)}$, conditioned on the spatially interpolated bias corrected k and σ^* obtained in step 3. KUD is used to spatially interpolate the re-estimated μ^*. At the end of this step, gridded estimates of the dimensionless GEV parameters k, μ^*, and σ^* are obtained.

- *Step 5:* The local mean $m^{(i)}$ is spatially interpolated with KUD and gridded estimates of the GEV parameters k, μ, and σ are obtained. These allow computing gridded P quantiles, $q(T_R)$.

3.2.1. Bias Correction of the GEV Shape Parameter

It is well known that estimates of the GEV shape parameter, k, from short APM records are subject to large uncertainty (Hosking & Wallis, 1997; Mascaro, 2020; Overeem et al., 2009; Papalexiou & Koutsoyiannis, 2013). Papalexiou & Koutsoyiannis (2013) investigated the relation between the variability of k and the sample size using long-term global records of daily APM from 15,137 rain gages of the Global Historical Climatology Network-Daily database. These authors found that, as the record length increases, the shape parameter converges asymptotically to a mean value of 0.114, suggesting that the distribution of daily APM is heavy tailed ($k > 0$), with a standard deviation of 0.045, which was attributed to the effect of local climate conditions. Moreover, Papalexiou & Koutsoyiannis (2013) proposed an empirical relation to correct the bias of k estimates as a function of its sample size. In a more recent study, Carney (2016) extended this analysis to the shape parameter of APM records at durations from 1 h to 60 days using APM records of the rain gages used in the NOAA Atlas 14 studies in the U.S. These authors proposed the relation:

$$\hat{k} = \frac{0.045}{0.045+1.27n^{-0.70}}\{\hat{k}(n) - [\mu_k + \beta/n]\} + \mu_k \qquad (6)$$

where $\hat{k}$ is the bias corrected value, $\hat{k}(n)$ is the PWM estimate of k from the sample, n is the sample size, and μ_k is the estimated unbiased average GEV shape parameter and β is a coefficient. The values of μ_k and β estimated for Volume 1 of NOAA Atlas 14 (Arid Southwest) and used here are reported in Table 1. Equation (6) was applied here to

correct the bias of the $\hat{k}(n)$'s estimated by fitting the GEV to the AMP series at each gage and radar pixel (Step 1 in Section 3.1). Note that the estimate of shape parameter does not change whether the original, $x^{(i)}$, or the standardized, $y^{(i)}$, variables are used.

Table 1. Value of Coefficients Used to Bias Correct k_d with Equation (6) for NOAA Atlas 14 – Volume 1.

Duration, *d* (h)	μ_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

3.2.2. Spatial Interpolation Techniques

We used two techniques to spatially interpolate the GEV parameters (Steps 2-5 in Section 3.2). The first is Kriging for Uncertain Data (KUD), which was recently adopted by Deidda et al. (2021) to interpolate the GEV parameters of daily APM recorded by 256 rain gages in the island of Sardinia, Italy. The KUD differs from ordinary kriging since it does not assume zero variance of the statistical parameter at the location where estimates are available. This is a desirable characteristic when the variable to be interpolated exhibits large uncertainty, as in the case of the GEV parameters and, particularly, the shape parameter. The technique was first developed by Marsily (1986) for homoscedastic fields (i.e., equal variance) and later expanded by Mazzetti & Todini (2008) to account for heteroscedasticity. As for ordinary kriging, the estimation of the target parameter θ at a given site with KUD is a linear combination of the θ estimates at N neighboring sites:

$$\hat{\theta} = \sum_{i=1}^{N} \lambda_i \, \theta_i \qquad (7)$$

where θ_i is the local parameter estimate at location i, and λ_i is the weight assigned to each of the sites. To obtain the weights at each location, the original kriging method is modified to account for the uncertainty in the following way:

$$\begin{cases} \gamma^*_{i,j} = \gamma_{i,j} + \frac{\sigma_i^2 + \sigma_j^2}{2} & i,j = 1,2 \dots N \text{ and } i \neq j \\ \gamma^*_{i,j} = \gamma_{i,j} & i = j \\ \gamma^*_{i,0} = \gamma_{i,0} + \frac{\sigma_i^2}{2} & i = 1,2,\dots N \end{cases} \quad (8)$$

Here, $\gamma_{(i,j)}$ refers to the variogram value between sites i and j, and $\gamma_{(i,0)}$ refers to the variogram value between the site i and the estimation site. The corresponding modified variogram values used in KUD are denoted by γ^*. σ_i^2 is the at-site measuring variance at site i, which is estimated through Monte Carlo simulations. For our application where we interpolated the GEV parameters, we performed 10,000 Monte Carlo simulations, each involving (1) the generation of n variates (n = 19) using the at-site GEV parameters, and (2) the estimation of the GEV parameters using these n variates. The simulations lead to 10,000 parameter values, whose sampling variance is used to estimate σ_i^2. The KUD technique was used to interpolate k in Step 2, σ in Step 3, μ in Step 4, and $m^{(i)}$ in Step 5.

As an alternative to KUD, we also evaluated the performance of a simple moving average (MA) to smooth the variability of the shape parameter. This method involves calculating, for each pixel or gage, the mean value of parameter estimates at sites located within a radius l^*. Here, we assumed l^* = 40 km, which is a distance within which the statistical properties of extreme P can be assumed not to exhibit marked changes Mascaro (2017). It is worth to highlight that KUD accounts for the spatial correlation of the targeted variable, while MA assumes independence among the values at all locations. The reasons why MA was tested will be better explained in the Results.

3.3. Error Metrics

The performance of Stage IV QPEs in characterizing EP was assessed through error metrics between the GEV quantiles obtained by different estimation methods. For the *j*-th gage and/or co-located radar pixel, we computed the relative bias (RB) as:

$$\mathrm{RB}_{j,REF} = \frac{q_j^{EST}(T_R,d) - q_j^{REF}(T_R,d)}{q_j^{REF}(T_R,d)} \times 100, \quad (9)$$

where $q_j^{EST}(T_R, d)$ and $q_j^{REF}(T_R, d)$ are the T_R-quantiles of EP at duration *d* estimated for a given estimation (*EST*) and reference (*REF*) method, respectively. The relative biases were then averaged across all *N* radar pixel-gage pairs as:

$$\mathrm{RB}_{REF} = \frac{1}{N}\sum_{j=1}^{N} \mathrm{RB}_{j,REF}. \quad (10)$$

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

$$\mathrm{RRMSE}_{REF} = \left[\frac{1}{N}\sum_{j=1}^{N} \mathrm{RB}_{j,REF}{}^2\right]^{1/2}. \quad (11)$$

The quantile estimation methods and correspondent reference datasets are summarized in Table 2, along with a rationale explaining why the comparison was made. We considered the following estimation methods:

- *At-site BC Gages (Radar)*: quantiles estimated at a given gage (radar pixel) using the GEV distribution with at-site bias corrected (BC) *k*, and μ and σ estimated conditioning the shape parameter to the bias corrected *k* value.
- *Gages (Radar) KUD*: quantiles estimated using spatially interpolated parameters based on the application of KUD for *k* in the steps of Section 3.1.
- *Gages (Radar) MA*: quantiles estimated using spatially interpolated parameters based on the application of MA for *k* (and KUD for μ and σ) in the steps of Section 3.1.

- *NOAA 14*: precipitation frequency estimates (PFEs) from NOAA 14 Volume 1 maps for given P durations and return periods.

Table 2. Combinations of Estimation Method (*EST*) and Corresponding Reference Method (*REF*) Used for the Determination of Error Metrics Identifying the Goal to be Achieved

EST	***REF***	**Goal**
At-site BC Radar	At-site BC Gages	Performance of best possible at-site estimates
Gages KUD	At-site BC Gages	Change in performance due to spatial interpolation
Gages MA	At-site BC Gages	
Radar KUD	At-site BC Radar	
Radar MA	At-site BC Radar	
Radar MA	At-site BC Gages	Performance of interpolated radar with best possible local estimate of gages
At-site BC Gages	NOAA 14	Performance of best possible at-site estimates with operational PFEs
Radar MA	NOAA 14	Performance of interpolated radar with operational PFEs

3.4. Scaling Models for Intensity-Duration-Frequency Analysis of Extreme Precipitation

Intensity-duration-frequency (IDF) curves of EP are empirical relations used to design and maintain infrastructure that provide the precipitation intensity, $i(T_R, d)$, for a given duration, d, and return period, T_R. A popular equation that well captures IDF curves has the general form (Koutsoyiannis et al., 1998):

$$i(T_R, d) = a(T_R)(d + \theta)^{-\eta}, \qquad (12)$$

where $a(T_R)$ depends on the probability distribution function of EP, and $\theta \geq 0$ and $0 < \eta \leq 1$ are site-dependent coefficients. If applied separately for two durations d and d_0, equation (9) leads to the convenient relationship between the corresponding P intensities:

$$i(d, T_R) = \frac{(d+\theta)^{-\eta}}{(d_0+\theta)^{-\eta}} i(d_0, T_R). \qquad (13)$$

When $\theta = 0$, equation (10) simplifies to a form that descends from the assumption of "strict sense simple scaling" of precipitation (Burlando & Rosso, 1996a; Gupta & Waymire, 1990). This can be written as:

$$I_d \overset{d}{=} \left(\frac{d}{d_0}\right)^{-\eta} I_{d_0}, \qquad (14)$$

with the symbol $\overset{d}{=}$ indicating equality in probability distributions and η being the scaling exponent. This implies that I_d and I_{d0} have the same distribution once the latter is rescaled by a factor $(d/d_0)^{-\eta}$. This property can be extended to the moments of order q as:

$$E\left(I_d^q\right) = \left(\frac{d}{d_0}\right)^{-q\eta} E\left(I_{d_0}^q\right), \qquad (15)$$

which has been also described as the "wide sense simple scaling" property (Gupta & Waymire, 1990).

The assumption of simple scaling has been used to derive an IDF model adopting the GEV distribution with constant shape parameter across durations. Under these conditions, equation (12) provides two scaling relations that link the scale and location parameter with the duration:

$$\mu_d = \left(\frac{d}{d_0}\right)^{-\eta} \mu_{d_0}, \qquad \sigma_d = \left(\frac{d}{d_0}\right)^{-\eta} \sigma_{d_0}. \qquad (16)$$

Equations (13) allow obtaining the GEV parameters at high temporal resolutions, where P data are less available (e.g., d = 1 h), from the values estimated from AMP records at coarser resolutions, which are more widely observed (e.g., d_0 = 24 h), through the scaling exponent η. The GEV simple scaling model has been applied in different regions of the

world (Blanchet et al., 2016a; Innocenti et al., 2017; Nguyen et al., 1998) including central Arizona (Mascaro, 2020) with relatively good accuracy

The assumption of simple scaling leads to a parsimonious IDF model that is operationally simple to apply. However, ample evidence has shown that P exhibits multiscaling properties (Burlando & Rosso, 1996b; Deidda et al., 1999; Mascaro et al., 2014), which can be defined according to Gupta & Waymire (1990) as:

$$E\left(I_d^q\right) = \left(\frac{d}{d_0}\right)^{-\alpha_q} E\left(I_{d_0}^q\right), \tag{17}$$

where α_q is a nonlinear function of q that accounts for the departure from simple scaling In recent work, Van de Vyver (2018) proposed a multiscaling IDF model based on the GEV distribution, where different scaling exponents are used for μ and σ:

$$\mu_d = \left(\frac{d}{d_0}\right)^{-\eta_\mu} \mu_{d_0}, \qquad \sigma_d = \left(\frac{d}{d_0}\right)^{-\eta_\sigma} \sigma_{d_0}. \tag{18}$$

This author has also proved that $\eta_\mu \leq \eta_\sigma$ and showed that the multiscaling GEV model reproduces better IDF relations in several gages in Belgium than the simple scaling model. Here, we (1) investigated whether the GEV parameters estimates from the radar QPEs at different durations exhibit scaling properties; (2) compared the radar-derived scaling exponents with those obtained for the gages, complementing the work of Mascaro (2020); and (3) used the large samples of scaling exponents estimated at the radar pixels to assess whether η_μ and η_σ are different in the region.

CHAPTER 4

RESULTS AND DISCUSSION

4.1. Evaluation of the GEV hypothesis

The suitability of the GEV distribution to model APM series observed at the gages and retried at the radar pixels was evaluated using the L-moment ratio diagram (Hosking, 1990, 1992; Hosking & Wallis, 1993; Peel et al., 2009). This is shown in Figure 2 for the AMP 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 91 gages with longer records of n =30 years, while those on the right report results at the same 91 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 (large symbols) are very close to or lie on the GEV line, indicating that this distribution well captures the AMP 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 large 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. The average 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 for both gages and radar allows capturing the average statistical properties in the region obtained from longer records; and (2) differences between the statistical properties of EP characterized

by gages and radar are larger at short durations. The analyses presented next will be based on the 257 gages with n = 19 years.

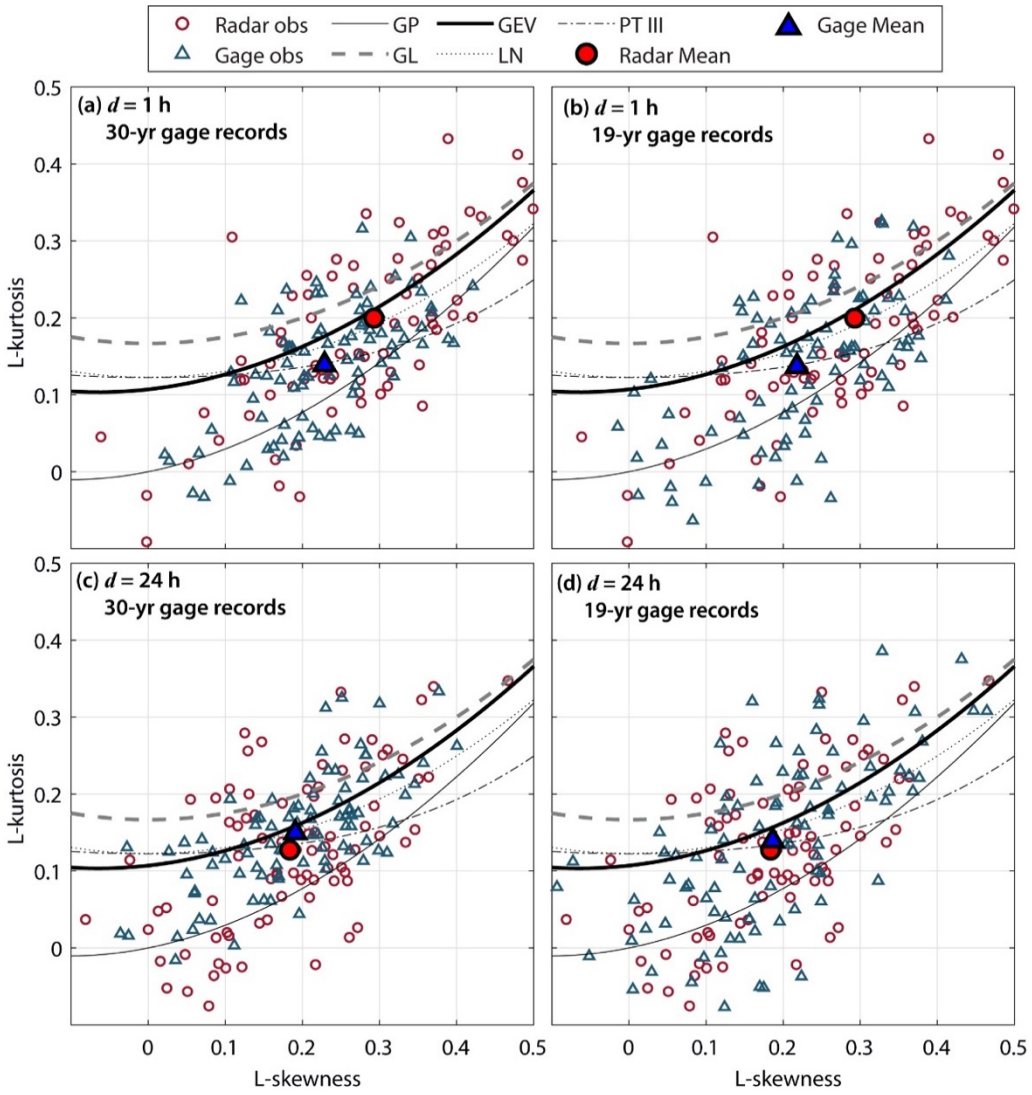

Figure 2. L-moment Ratio Diagram for Observed Records of AMP at 91 Gages and Co-located Radar Pixels for (a)-(b) d = 1 h and Sample Size n = 30 and 19 Years for the Gages, Respectively. (c) and (d) Are the Same as (a) and (b), but for d = 24 h. The Lines Show the Theoretical L-moment Combinations for the Generalized Pareto (GP), Generalized Logistic (GL), Generalized Extreme Value (GEV), Lognormal (LN), and Pearson Type III (PT3).

4.2. Comparison of At-Site and Spatially Interpolated GEV Parameter Estimates for Gages and Radar

As described in Section 3, at-site estimates and spatial interpolations were applied to obtain the shape, location, and scale parameters of the GEV distribution at each gage and co-located radar pixel. Figures 3-5 present the scatterplots between the parameter values at the 257 gages (x axis) and co-located radar pixels (y axis) under different methods for d = 1, 6, and 24 h, respectively (results for all durations are presented in Appendix A). The upper row in each figure (panels (a)-(c)) present the scatterplots with At-site estimates of k, μ and σ. For all durations, the relations between the k estimates for the two datasets is quite scattered with very low (<0.07) coefficient of determination, R^2. Figure 23 in Appendix A shows the scatterplot comparing k from radar and gage using At-site BC estimates with low correlation. This is an expected result given the large uncertainty of the shape parameter estimates for such a short sample size. The correspondence is instead stronger for at-site estimates of μ and σ, with R^2 increasing with d and reaching 0.50 (0.80) for σ (μ) for d = 24 h (Figures 5b,c). For both μ and σ, the radar estimates are negatively biased compared to gage estimates for the lower durations. The negative bias is reduced as d increases and becomes slightly positive for d = 24 h.

Panels (d)-(f) and (g)-(i) in Figures 3-5 present the scatterplots between the parameter estimates obtained through the spatial interpolation steps based on KUD and MA, respectively. Overall, the two interpolation techniques lead to very similar impacts on the GEV parameters. As described in Section 3.2, prior to performing the spatial interpolations, k was bias corrected to account for the uncertainty due to the low sample

size. This greatly reduces the scatter of the k estimates and leads this parameter to range between 0 and 0.2 for both KUD and MA, with larger (lower) values of d = 1 h (24 h). Interestingly, at lower durations, the R^2 of MA is higher than that of KUD. For both μ and σ, the spatial interpolation methods reduce the range of variability and lower the values for each product. The scatter between radar and gage values is importantly reduced compared to the at-site results. Consequently, the linear correlation increases with R^2 ranging from 0.20 (0.71) at d = 1 h to 0.70 (0.9) at d = 24 h for σ (μ). However, the negative bias at d = 1 h is still present and a slightly positive value is found at d = 24 h for σ > 0.6 mm/h and μ > 1.4 mm/h.

The impact of negatively biased scale and location parameters at smaller durations can be first understood in terms of the GEV distribution (probability distribution function) and then the quantiles estimated using them. A lower μ shifts the shape of the GEV distribution to the left compared to the true distribution and this in turn implies the underestimation in the probability of occurrence of a higher P intensity. A lower σ reduces the spread of the distribution and increases the concentration of probability around the mode of the AMP series. The negative bias of both σ and μ implies underestimation of P quantiles at shorter durations.

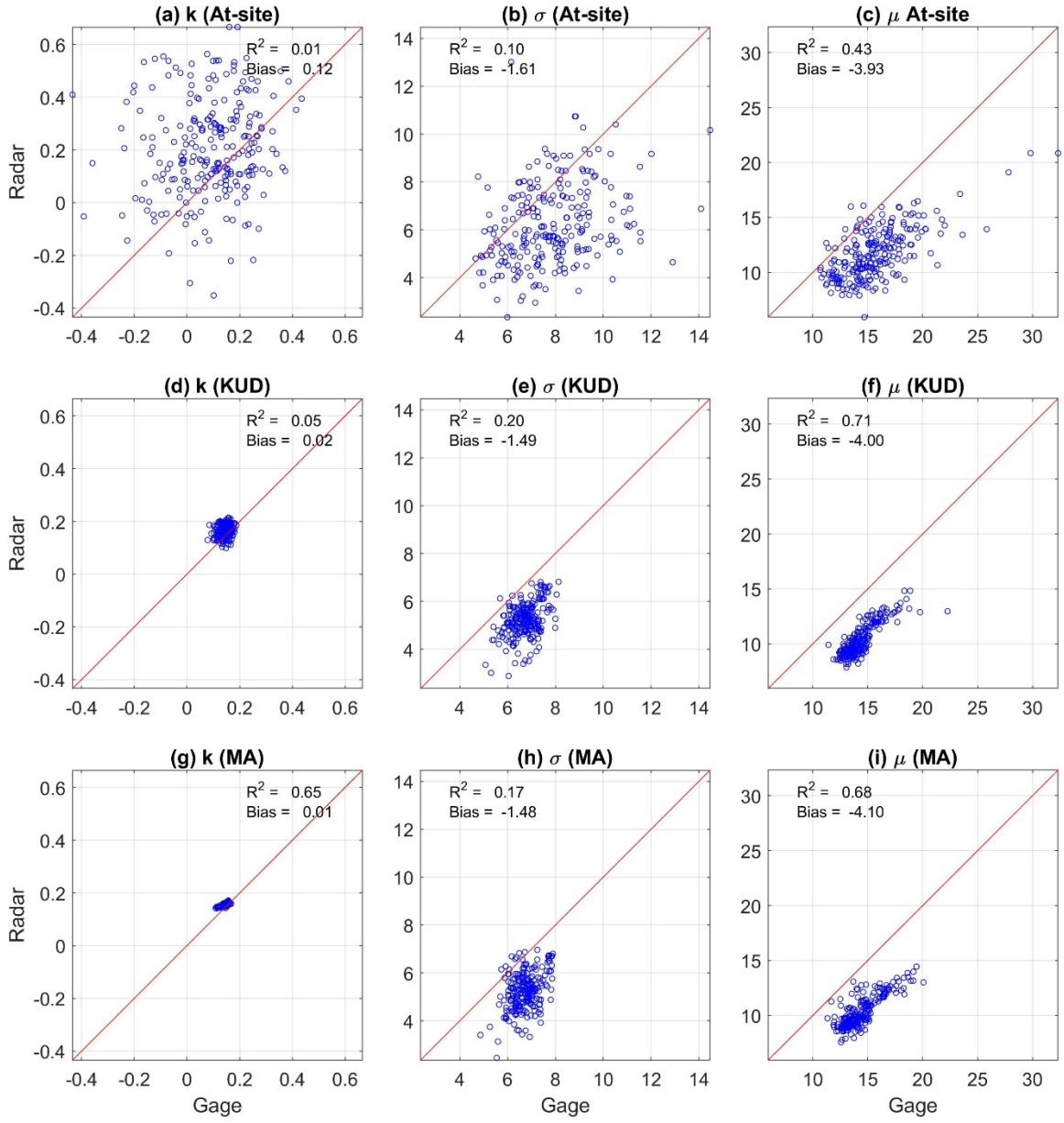

Figure 3. Scatterplots Comparing Radar and Gage Estimates for At-site, KUD, and MA Spatial Interpolation Methods When d = 1 h

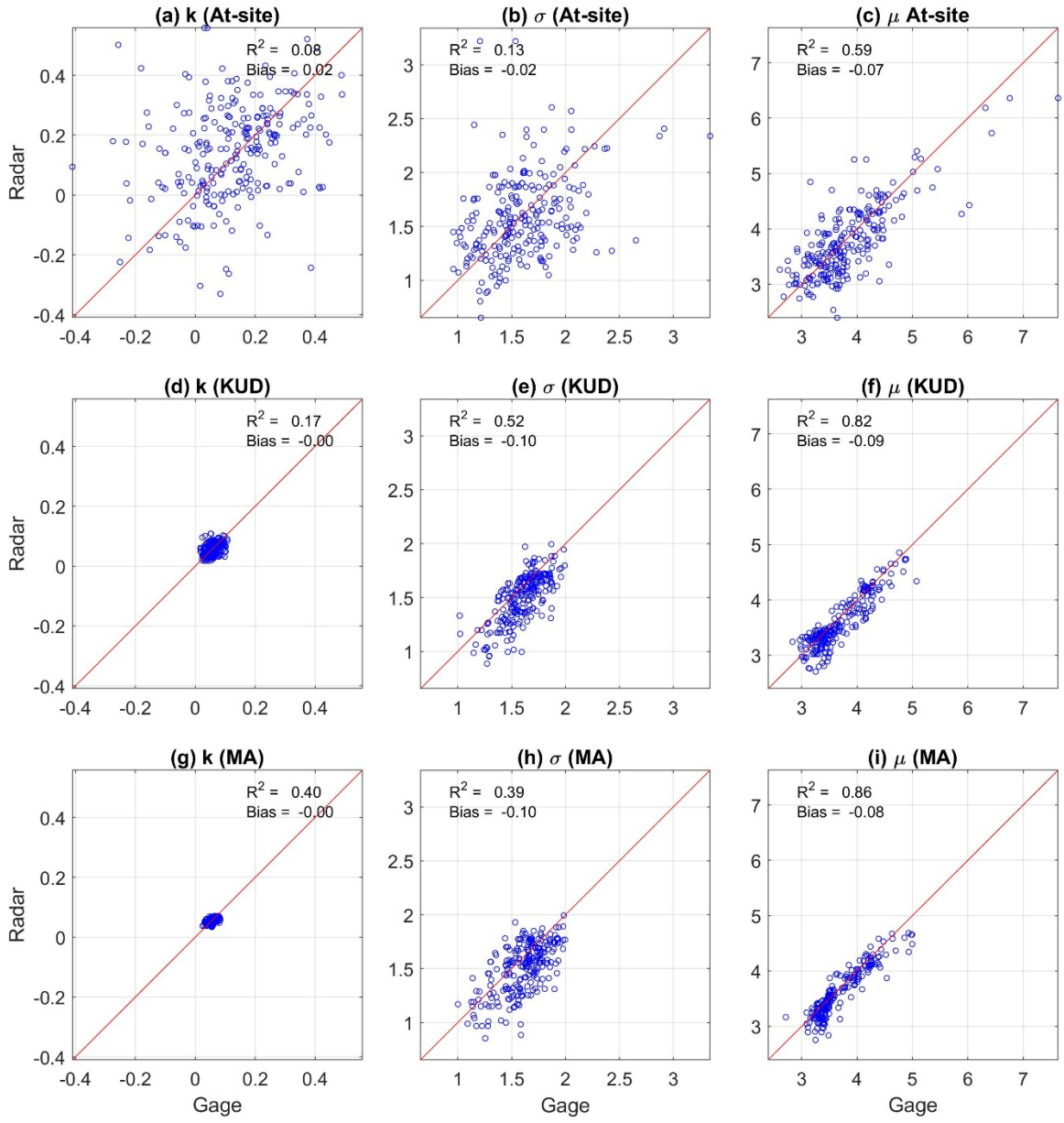

Figure 4. Scatterplots Comparing Radar and Gage Estimates for At-site, KUD, and MA Spatial Interpolation Methods When $d = 6$ h

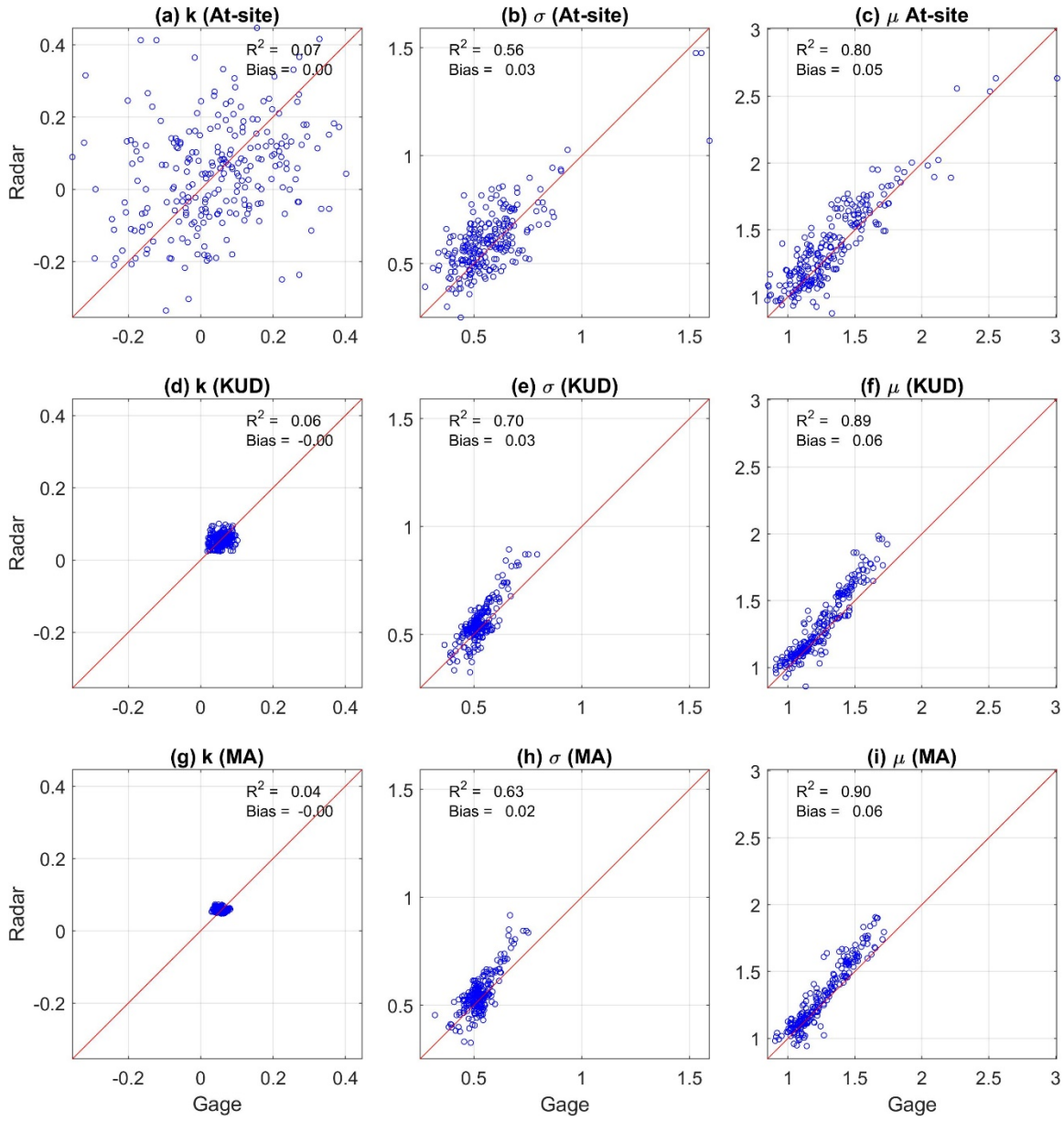

Figure 5. Scatterplots Comparing Radar and Gage Estimates for At-site, KUD, and MA Spatial Interpolation Methods When d = 24 h

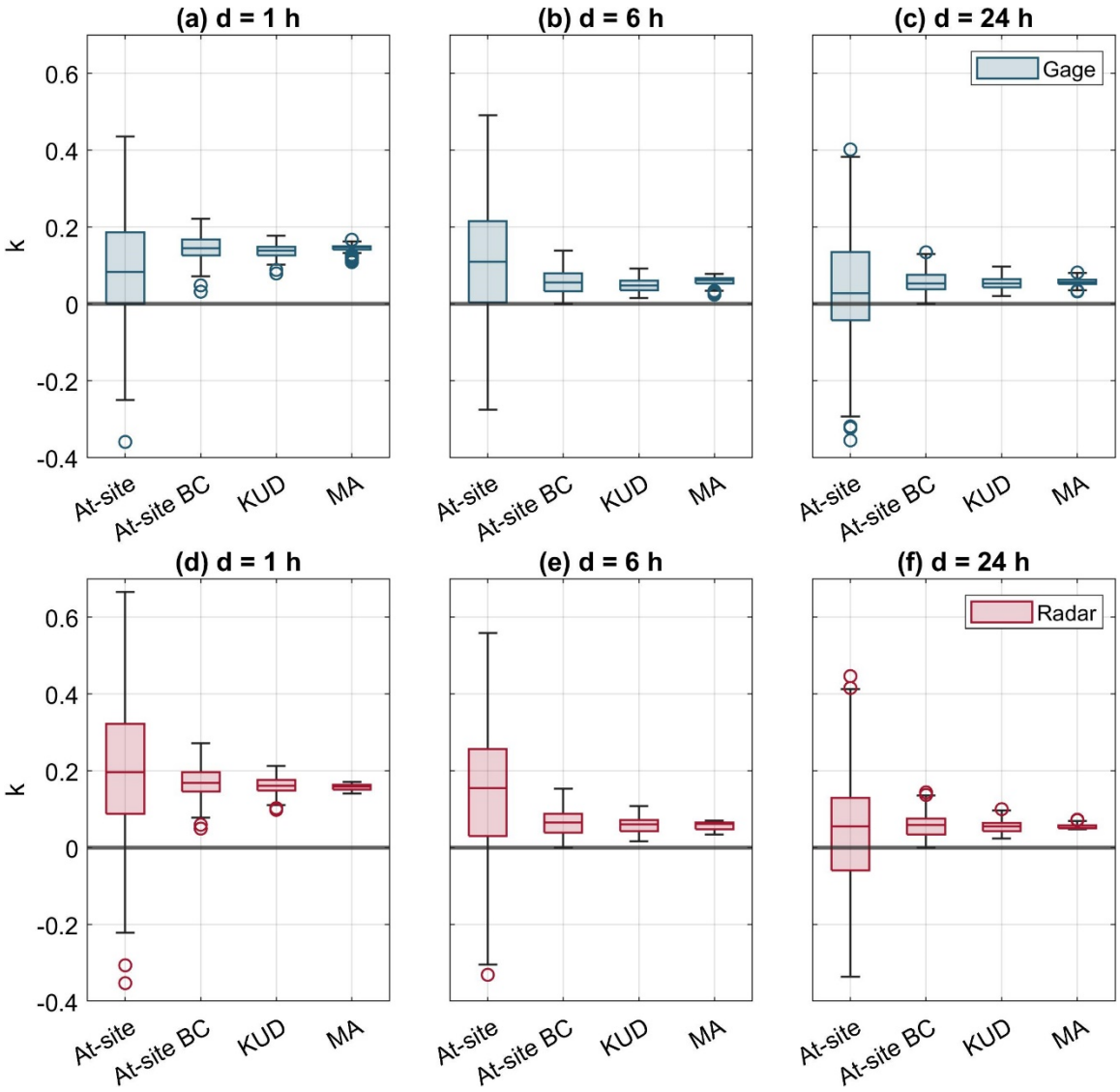

Figure 6. Boxplots Summarizing the Empirical Distributions of the Shape Parameter k Across (a)-(c) 257 Gages and (d)-(f) Co-located Radar Pixels for d = 1, 6 and 24 h and Different Estimation Methods, Including At-site, At-site BC, KUD and MA.

Results of the scatterplots are summarized in Figures 6-8 through boxplots that show the empirical distribution of the GEV parameters for the different estimation methods. As presented in Figure 6, the bias correction of k from At-site to At-site BC (now shown in scatterplots) greatly reduces the inter-site variability of this parameter. The use of interpolation methods leads to a further decrease of the variability for KUD

and, even more, for MA. The median value of k increases from At-site to the other methods for d = 1 h and d 24 h and decreases for d = 6 h. For all durations, the distribution is heavy tailed. All these considerations are true for both gage and radar records. The bias correction and spatial interpolations lead to practically the same empirical distributions of k for a given d obtained for gages and radar.

The boxplots for σ and μ are shown in Figures 7 and 8, respectively. It is important to note first that the range of values in units of mm/h for these parameters changes with the dependent. The figures reveal the following results. (1) The values and inter-site variability of both parameters do not change when k is bias corrected, i.e., comparing At-Site and At-Site BC. (2) When applying the spatial interpolation, the inter-site variability of both σ and μ decreases significantly, and their median values are also reduced. (3) No important difference emerges between KUD and MA. Considerations (1)-(3) apply to both gages and radar. (4) For each estimation method, the values of σ and μ estimated from the radar records are lower than those obtained for the gages for d = 1 h, and slightly lower for d = 6 h. The empirical distributions derived for the two products are practically the same for d = 24 h.

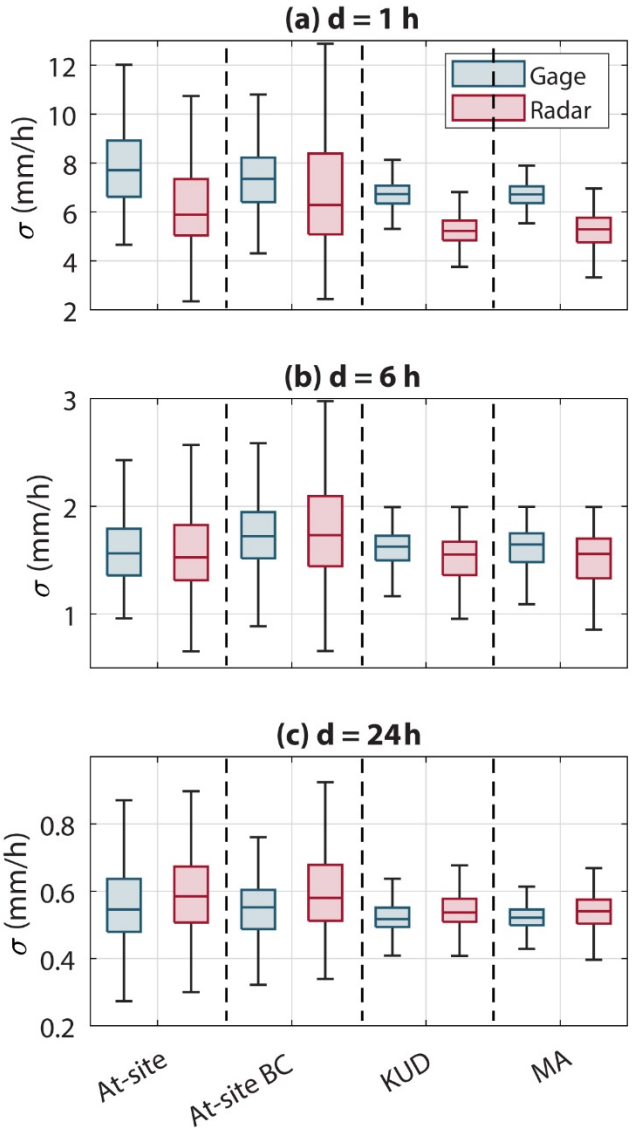

Figure 7. Boxplots Summarizing the Empirical Distributions of the Scale Parameter σ across 257 Gages and Co-located Radar Pixels for Different Estimation Methods, Including At-site, At-site BC, KUD and MA, for d = 1 h (a), 6 h (b), and (c) 24 h.

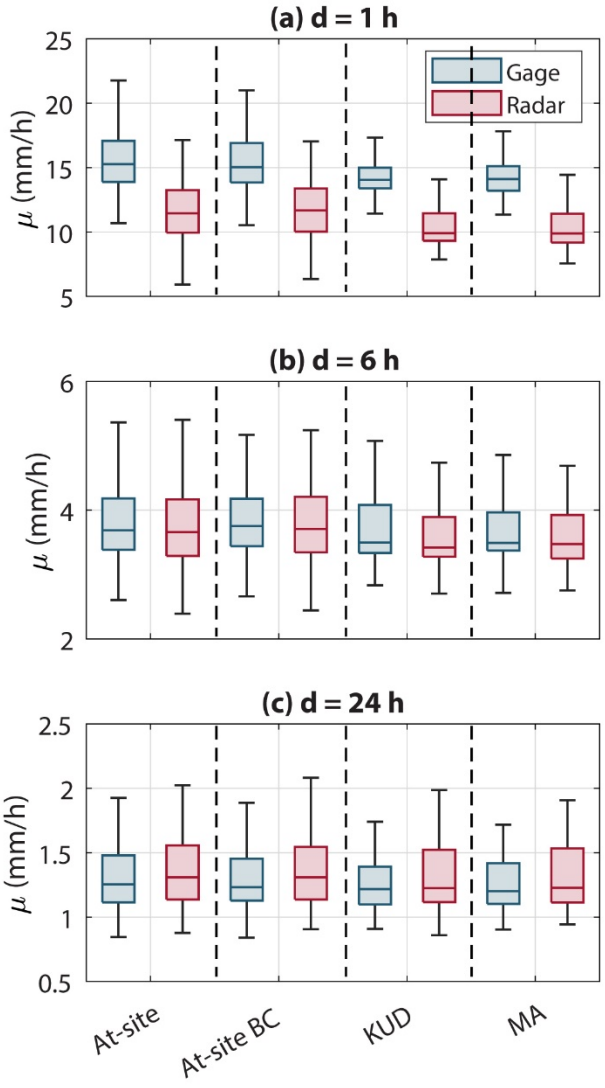

Figure 8. Same as Figure 6 but for Location Parameter μ.

4.3. Spatial Distribution of GEV Parameters

Arguably, the most valuable advantage of radar observations over gages is the availability of seamless spatial information. To prove this important point, Figures 9-11 show maps of the GEV parameters for d = 1, 6 and 24 h, respectively, derived from the radar QPEs using the different estimation methods. As already presented in the scatterplots and boxplots, the range of the At-site estimates of k is large with both high positive and negative values. The maps in panels (a) further show that the 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, as highlighted in the example of Figure 12a. The features of the spatial patterns of At-site k do not appear to be physical plausible, 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 (Figure 1b).

The major reason of these unrealistic spatial features is the large uncertainty in the estimation of the shape parameter using such a short sample, combined with the typical size of storms leading to extreme P. To better illustrate this, Figure 12b shows the year of occurrence of the largest APM (d = 1 h) using the same color in close pixels if the event was recorded within the same 24-hour period. It is clear that size and locations of the clusters in Figures 12a,b are similar, indicating that the spatial correlation of k is closely related to the spatial correlation of the storms causing the largest APM. This is further illustrated in Figures 12c,d, that demonstrate that the shape of the GEV distribution fitted

to the empirical CDF at pixels belonging to two clusters is severely affected by the largest APM values and that, for each cluster, these have occurred in the same year.

Here, to limit the impacts of this uncertainty, the At-site *k* estimates were bias corrected following Papalexiou & Koutsoyiannis (2013) and Carney (2016). As shown in the At-site BC maps in panels (d), the bias correction constrains the *k* estimates to positive values (i.e., heavy tailed distributions) and reduces the inter-site variability; however, the spatial correlation caused by the clusters is still present. Such correlation is somewhat reduced through the spatial interpolation with KUD (panels (g)), and mostly eliminated by MA (panels (j)). The spatial patterns of *k* obtained using this last interpolation technique exhibit instead large-scale spatial correlations that appear to be the most realistic.

As anticipated by the boxplots, the patterns of σ and μ are very similar for At-Site and At-Site BC, indicating that the bias correction of *k* does not significantly impact the estimates of scale and location parameters (compare panels (b)-(c) with panels (e)-(f)). The limited effect of using different *k* values when re-estimating σ and μ is confirmed by contrasting the spatial maps of these parameters obtained outcomes for KUD (panels (h)-(i)) and MA (panels (k)-(l)). The spatial patterns obtained for the two methods are in fact very similar, except for a slightly larger variability exhibited by μ in KUD. These considerations are true for all durations.

Finally, it was found that the spatial patterns of σ and μ are partially explained by elevation, with a dependence that increases with *d*, confirming results obtained by Mascaro (2020) analyzing gage records. The R^2 for σ and μ estimated using At-Site BC and MA chosen as representative methods are reported in Table 3 which indicate the

correspondence between the parameter estimates and the elevation at that pixel. Higher R^2 for μ confirms the effect of orography in the study domain, which implies that short duration convective storms are less reliant on topography compared to long lasting storms which have a higher dependence on terrain conditions. The dependence of σ on elevation is negligible for short duration storms, but for larger durations this relationship is somewhat existent explaining that scale parameter also varies with elevations. The use of spatial interpolation techniques reduces the overall noise associated with estimation of σ and μ which is also reflected through better R^2 values with the use of MA or KUD.

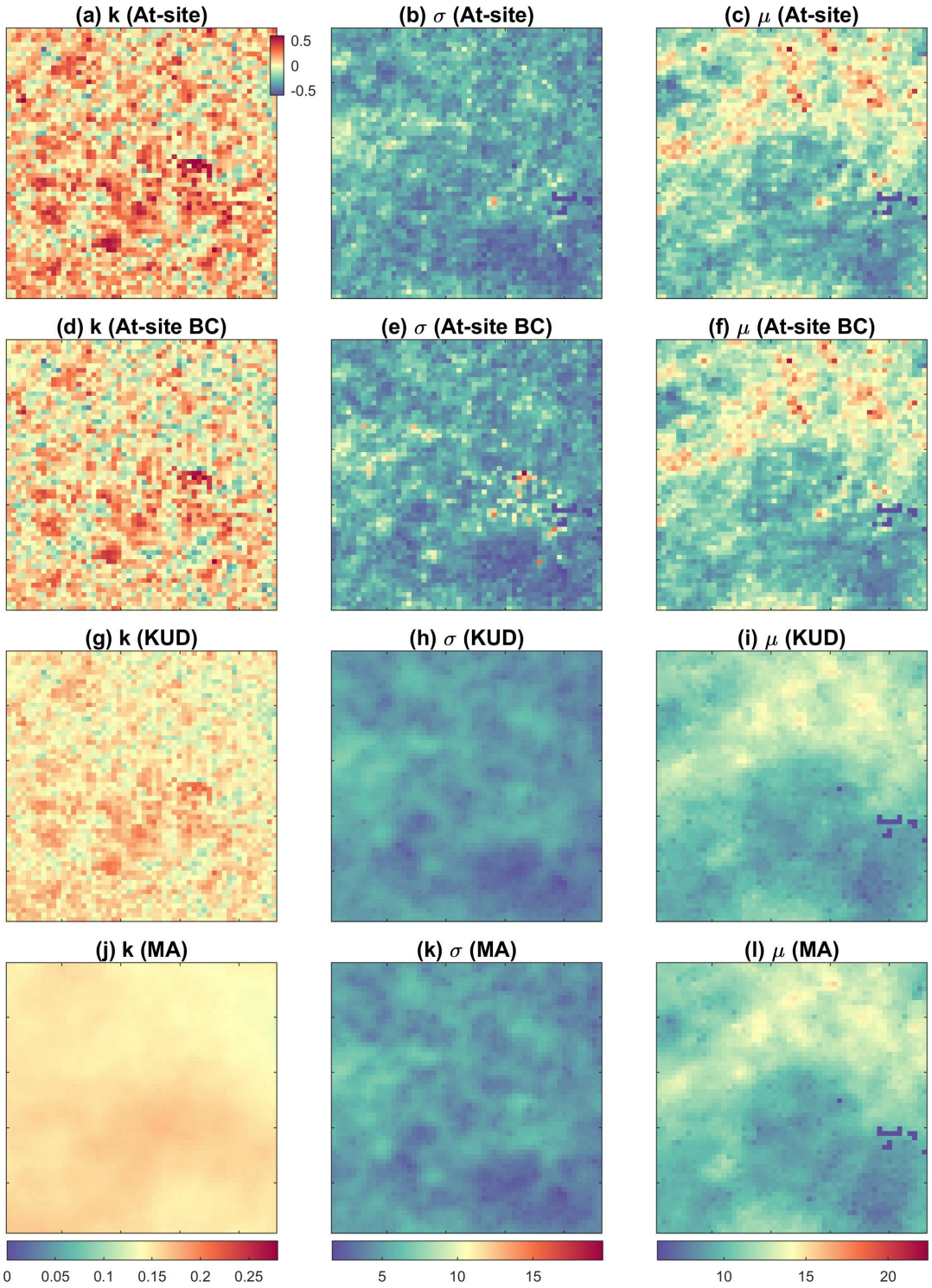

Figure 9. Spatial Variability of the GEV Parameters for d = 1 h Obtained with At-site, At-site BC, KUD and MA Estimation Methods. The Colormaps for Each Parameter are Shown at the Bottom of Each Column (Only Panel (a) has its Own Colormap)

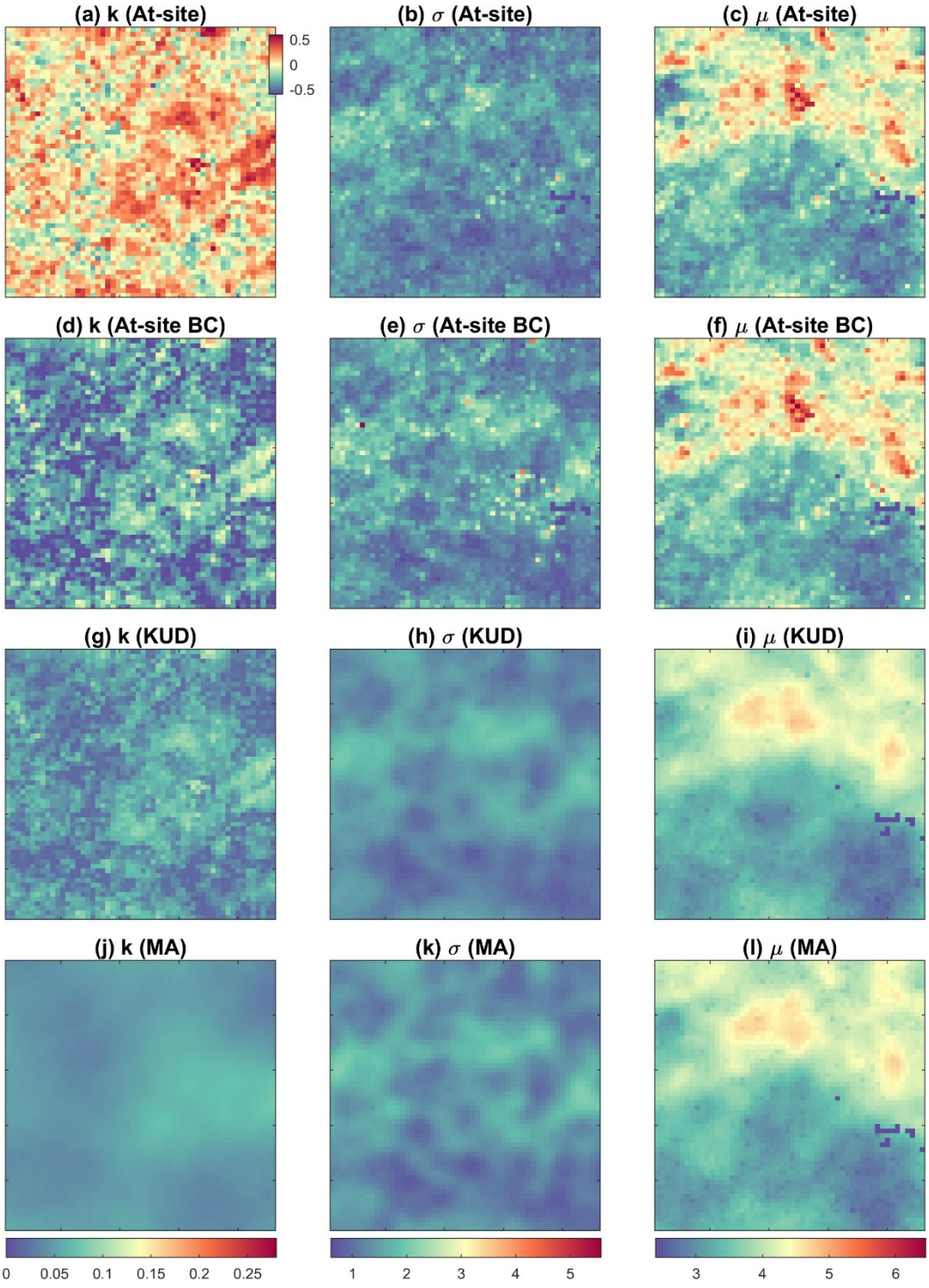

Figure 10. Same as Figure 9 but for d = 6 h.

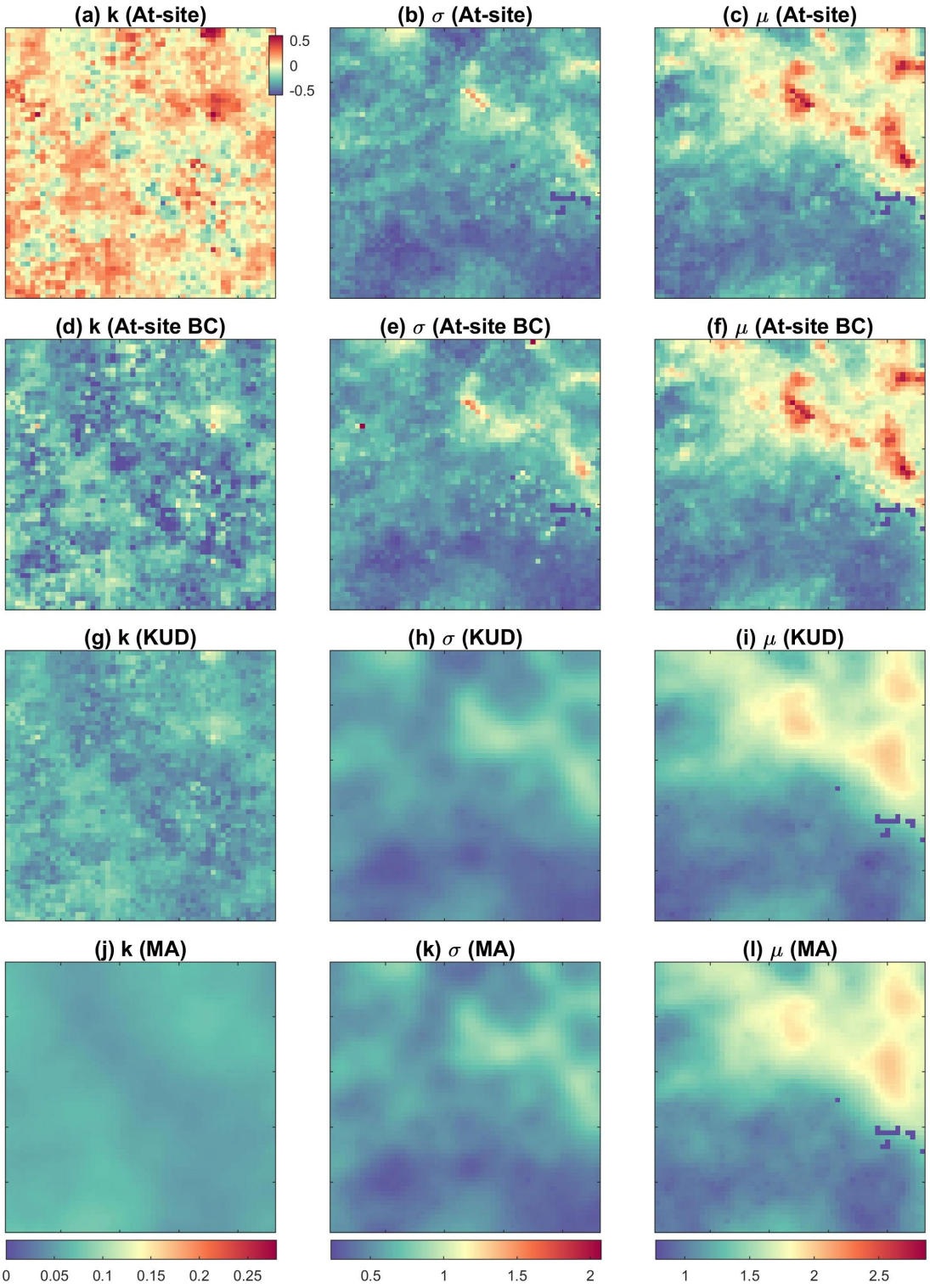

Figure 11. Same as Figure 9 but for d = 24 h.

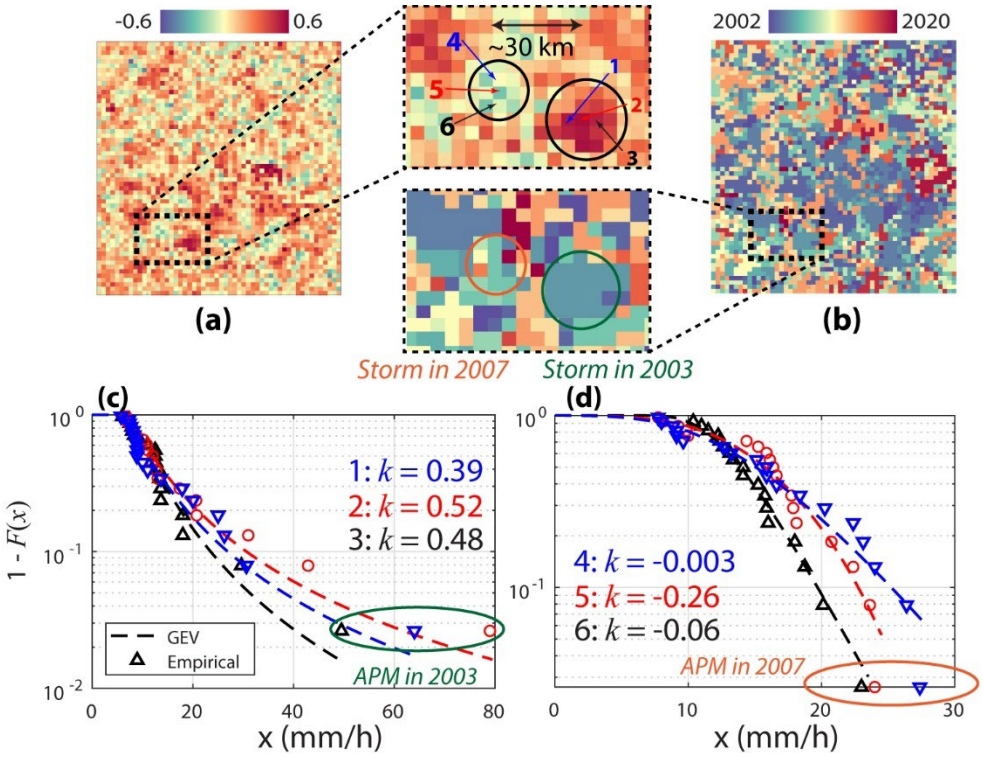

Figure 12. (a) Example of Two Clusters with High Positive and Low Values of At-site k for d = 1 h. (b) Year of Observation of Highest APM for d = 1 h. (c)-(d) Empirical and Fitted GEV Distributions of APM Records in Three Representative Pixels in Each of the Two Clusters.

Table 3. Estimated R^2 for a Linear Model Obtained by Fitting σ and μ to Elevation for a Given Estimation Method.

Duration	At-site BC		MA	
	σ	μ	σ	μ
1	0.0002	0.32	0.0003	0.49
2	0.0007	0.34	0.0024	0.52
3	0.002	0.37	0.003	0.53
6	0.017	0.47	0.019	0.58
12	0.14	0.58	0.19	0.63
24	0.27	0.63	0.30	0.66

4.4. Performance of Radar QPEs to Characterize Regional Extreme Precipitation Frequency

4.4.1. Performance of At-Site BC Radar Estimates

The ability of the radar QPEs to characterize the regional frequency of extreme P at different durations was tested through error metrics computed between the quantiles obtained for the estimation (*EST*) and reference (*REF*) methods summarized in Table 2. The first comparison was made between quantiles estimated with At-Site BC for both radar (*EST*) and gages (*REF*), in order to assess the capability of radar QPEs to capture local estimates of the gage quantiles. The mean relative bias (RB) and relative root-mean-square error (RRMSE) for T_R from 2 to 50 years and d = 1, 6, and 24 h are displayed in Figure 13. RB for d = 1 h is negative (i.e., the radar underestimates gage quantiles) and varies from around -20% for T_R = 2 years to -5% for T_R = 50 years. RB increases with the duration, being negligible for d = 6 h and slightly positive (up to ~+5%) for d = 24 h. For these two durations RB slightly increases with T_R. For all durations, the RRMSE doubles from T_R = 2 years to T_R = 50 years and is the largest for d = 1 h (maximum of 40%). The performances found at the different durations are expected given the biases of the GEV location and scale parameters shown in Figures 7 and 8 for At-Site BC.

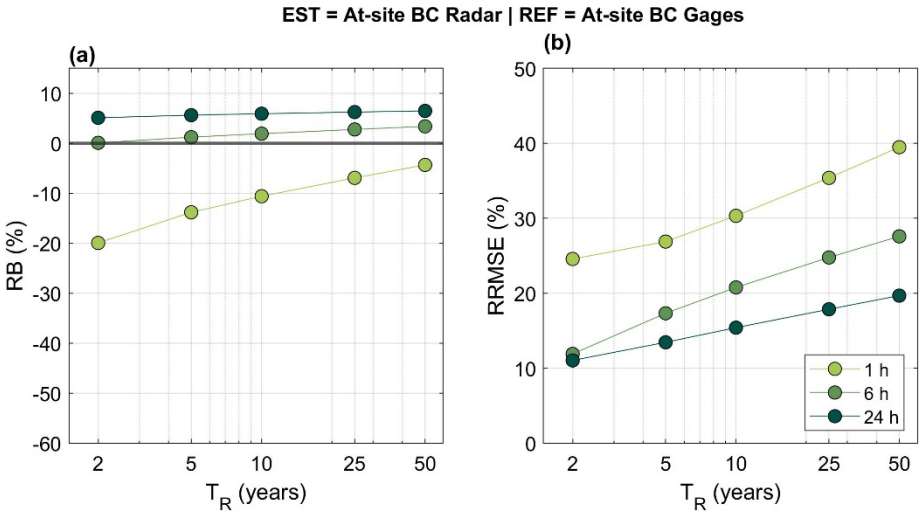

Figure 13. Error Metrics for *EST* = At-site BC Radar and *REF* = At-site BC Gage for *d* = 1 h, 6 h, and 24 h.

4.4.2. Performance of Spatial Interpolation Methods

To understand the impact due to the use of spatial interpolation techniques, we computed the error metrics for the Gage KUD and Gage MA (Radar KUD and Radar MA) estimation methods, respectively, assuming At-site BC Gage (At-site BC Radar) as reference. Results are shown in Figure 14. When comparing the two P products, the application of interpolation techniques leads to very similar trends in the error metrics against local quantile estimates, with (1) negative RB that decrease with T_R and is overall included between about -2% and -8%, (2) RRMSE that increases with T_R from roughly 7% to 15%, and (3) larger errors for lower durations. For the gages, RB is lower (i.e., less negative) for MA compared to KUD, while no important difference emerges between KUD and MA for the radar.

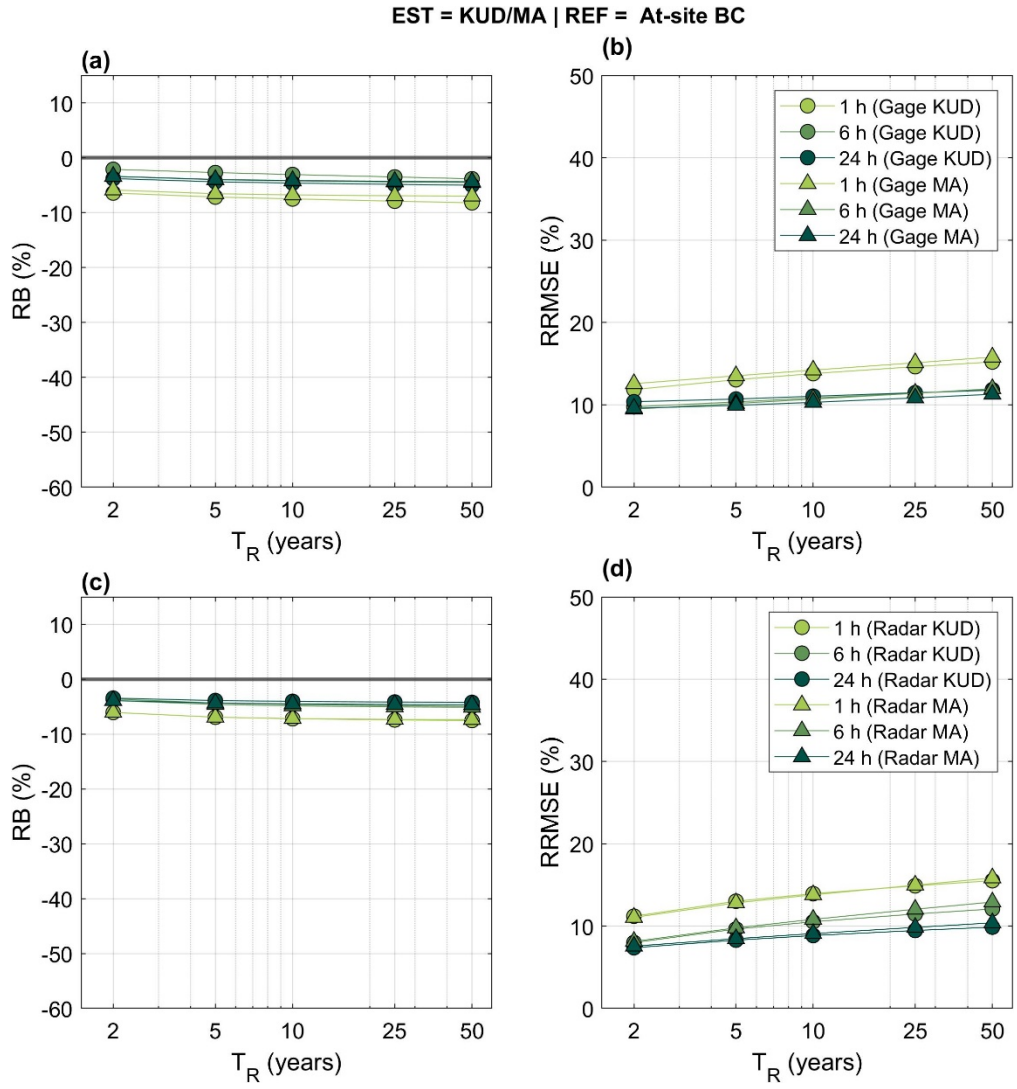

Figure 14. (a)-(b) Error Metrics for d = 1, 6, and 24 h for *EST* = Gages KUD or Gages MA and *REF* = At-site BC Gages. (c)-(d) Same as (a)-(b) but for *EST* = Radar KUD or Radar MA and *REF* = At-site BC Radar

As a next step, we tested the performance of the spatially interpolated radar quantiles with the best possible local estimates from gage records (i.e., At-site BC Gages). Results are displayed in Figure 16 only for Radar MA since the differences between the two interpolation techniques for the radar are very little, as shown in Figure 14. We first notice that the effect of T_R is quite small for all durations. RB for d = 24 h is

nearly 0% with a low RRMSE of less than 10%, implying that the spatial smoothing method is able to predict accurate estimates of extreme P quantiles at higher durations. Performances are fairly good also for intermediate durations (d = 6 h), where RB remains close to -10% and RRMSE is less than 20%. As found in previous cases, performances are the worst for d = 1 h with an RB of -30% and RRMSE of 30%. When compared to Figure 13, these findings reveal that the use of spatial interpolation methods with radar QPEs reproduces local gage estimates with only slightly larger errors than At-site BC Radar estimates. Moreover, since RB is not affected by T_R, bias correction techniques can be applied to remove the negative bias of the extreme P quantiles at lower durations, as discussed next.

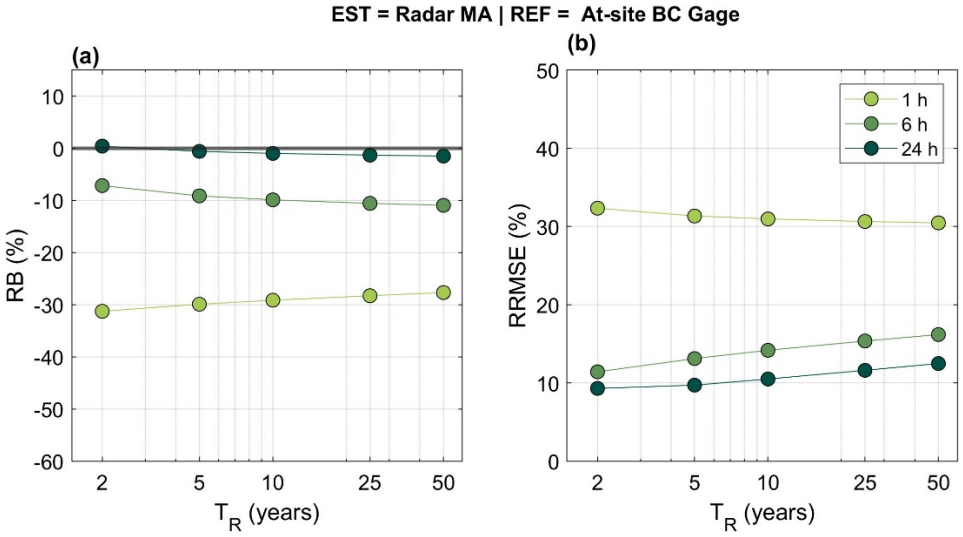

Figure 15. Same as Figure 14 but for *EST* = Radar MA and *REF* = At-site BC Gage.

4.4.3. Comparison with NOAA Atlas 14 PFEs

NOAA Atlas 14 provides PFEs that are routinely used in most of the U.S. to maintain and size infrastructure against extreme P. As previously mentioned, these PFEs are generated by applying regionalization techniques to rain gage records of different sample size using the GEV distribution. To complete the assessment of radar performance, we compared the quantiles generated with our gage and radar datasets in central Arizona with the PFEs from NOAA Atlas 14 Volume 1 that covers the semiarid Southwest (*REF*). The error metrics with *EST* = At-site BC Gages are displayed in Figures 16a,b, which indicate that some discrepancies exist between the two gage-based products. Local quantile estimates based on the 19-year-long gage records are negatively biased (about -20%) at T_R = 2 years for all durations. RB is gradually reduced as T_R increases and becomes slightly positive (up to +10%) for d = 1 years and T_R = 50 years. RRMSE varies between 10% and 15% with a minimum found at T_R = 5 years, with little differences across the durations.

For a more robust analysis, we also fit the GEV distribution to 30 years of gage records and compute P quantiles at each gage location. They are compared to PFEs where we find slightly lower errors. The RB at lower T_R remains similar as those from the 19 years, but at T_R = 50 years, the RB at all durations is ~0. The RRMSE also improves at higher recurrence intervals (< 20% for all durations at T_R = 50 years). This comparison here also gives an insight about the confidence in NOAA Atlas 14 estimates for different recurrence intervals. The presence of a negative bias at all durations even with the use of longer gage records located at a high spatial density suggest that there might be an overestimation of NOAA Atlas 14 quantiles at lower return periods.

The error metrics for *EST* = Radar MA are presented in Figures 16c,d. The values and trends of both error metrics for d = 24 h are very close to those found for At-site BC Gages, with the radar leading to lower RRMSE at higher T_R. The performance at d = 6 h is also comparable in terms of RB and better considering RRMSE. At d = 1 h, the quantiles derived from radar QPEs severely underestimate the NOAA 14 PFEs with RB increasing from about -45% at T_R = 2 years to -20% at T_R = 50 years. The RRMSE is also much larger than the values found at other durations. However, an important characteristic of the radar estimates is that their errors decrease significantly as T_R (and, thus, the statistical uncertainty) increases.

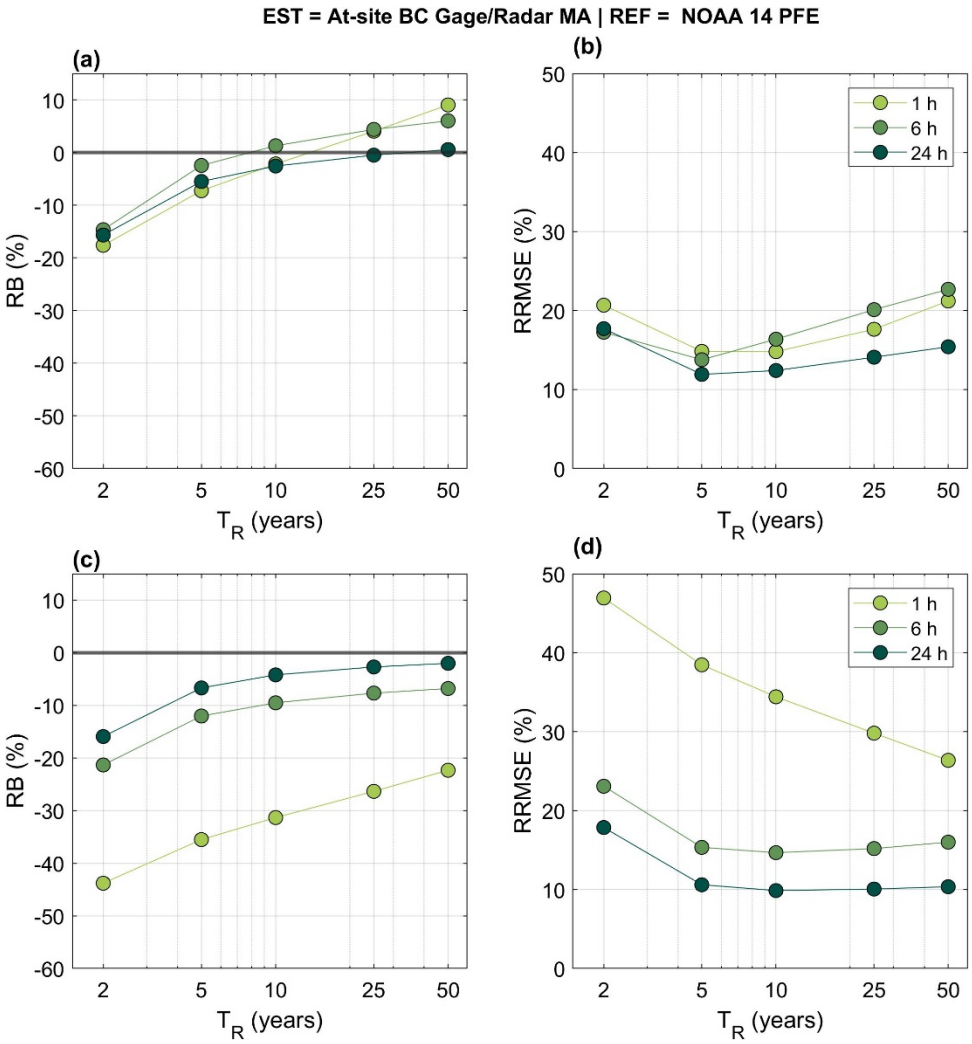

Figure 16. (a)-(b) Error Metrics for *d* = 1 h, 6 h, and 24 h for *EST* = At-site BC Gage and *REF* = NOAA 14. (c)-(d) Same as (a)-(b) but for *EST* = Radar MA and *REF* = NOAA 14.

4.4.4. Inter-Site Variability of Bias

As a final step of the analyses of the error metrics, Figure 17 presents the boxplots summarizing the empirical distribution of the bias at each gage vs. radar (or NOAA 14) pixel location, RB_j. In addition to confirming the outcomes of the previous figures based on RB, Figure 17 reveals the following two points: (1) The inter-site variability increases with T_R; and (2) The variability is the largest when comparing At-site BC Gage with NOAA 14 (Figure 17b) and it is quite similar when comparing Radar MA with At-site BC Gage (Figure 17a) or NOAA 14 (Figure 17c). We also investigated the existence of spatial patterns for RB_j and found this error metric to be largely randomly distributed in space (not shown).

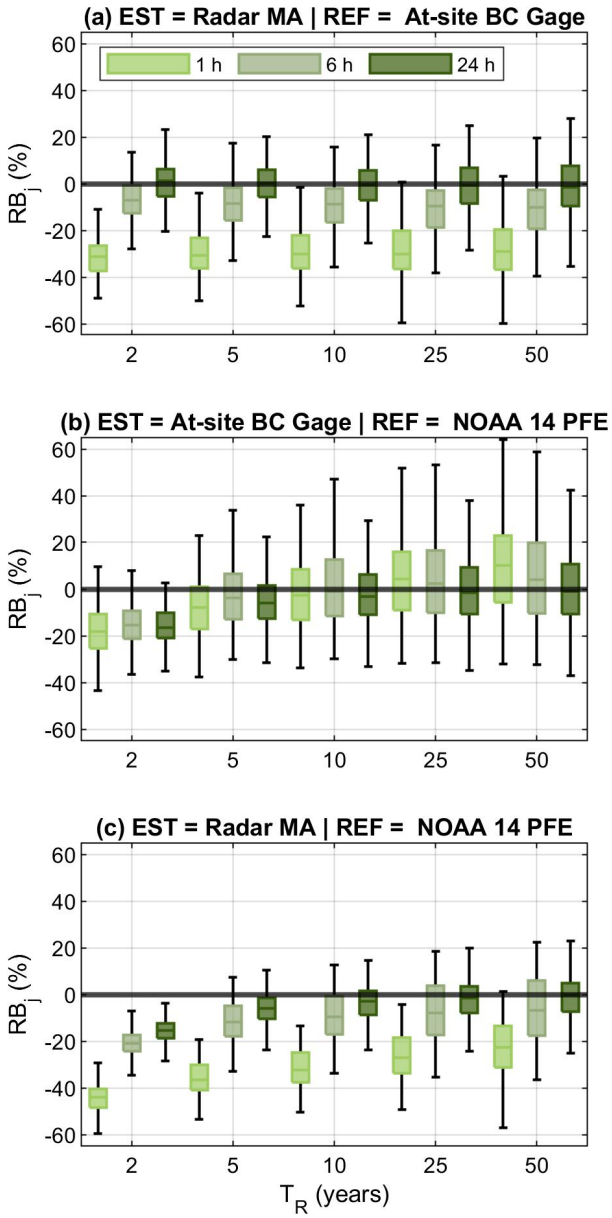

Figure 17. Boxplots of *RBj* for (a) *EST* = Radar MA and *REF* = At-site BC Gage, (b)*EST* = At-site BC Gage and *REF* = NOAA 14, and (c) *EST* = Radar MA and *REF* = NOAA 14.

4.6. Evidence of Scaling for the GEV Scale and Location Parameters

Mascaro (2020) found evidence of the scaling properties (equations (18)) for the At-site and At-site BC estimates of the location, μ, and scale, σ, GEV parameters using the same gage records analyzed here. This implies that a linear relation in the log-log plane was found between the parameter and the P duration, d, with the slope denoted as scaling exponent. Mascaro (2020) also found the scaling exponent for σ, η_σ, is slightly larger than that of μ, η_μ, suggesting that multiscaling GEV models should be used to develop more accurate IDF relations (Van de Vyver, 2018).

As a further way to assess the ability of radar QPEs to characterize the statistical properties of extreme P, here we verified whether the scaling relations are also exhibited by the GEV parameters estimated with the radar APM series using MA. We found this to be the case, but that the scaling exponents have a negative bias, which is a direct consequence of the bias between estimates of μ and σ from At-site gage BC and radar MA series found at lower P durations (see Figures 7-8). This is illustrated in Figure 18, where evidence of scaling for μ and σ is presented for a representative gage and co-located radar pixel (panels a-b), and results across all sites are summarized through boxplots and scatterplots (panels c-e). As found by Mascaro (2020), η_μ was found to be lower than η_σ, a condition that Van de Vyver (2018) proved to be needed to apply multiscaling IDF GEV models. Finally, Figure 18d,f presents the maps of η_μ and η_σ derived from the radar. Interesting patterns emerge with presence of correlations that are partially explained by elevation (R^2 of 0.25 of η_μ and 0.16 of η_σ). These maps indicate that regions in the southwestern portion of the domain have larger scaling exponents, which implies larger changes of APM intensities across durations values in these regions,

likely due to more localized and intense convective storms. In regions in the northeast with lower scaling exponents, APM intensities exhibit less variations across time aggregations.

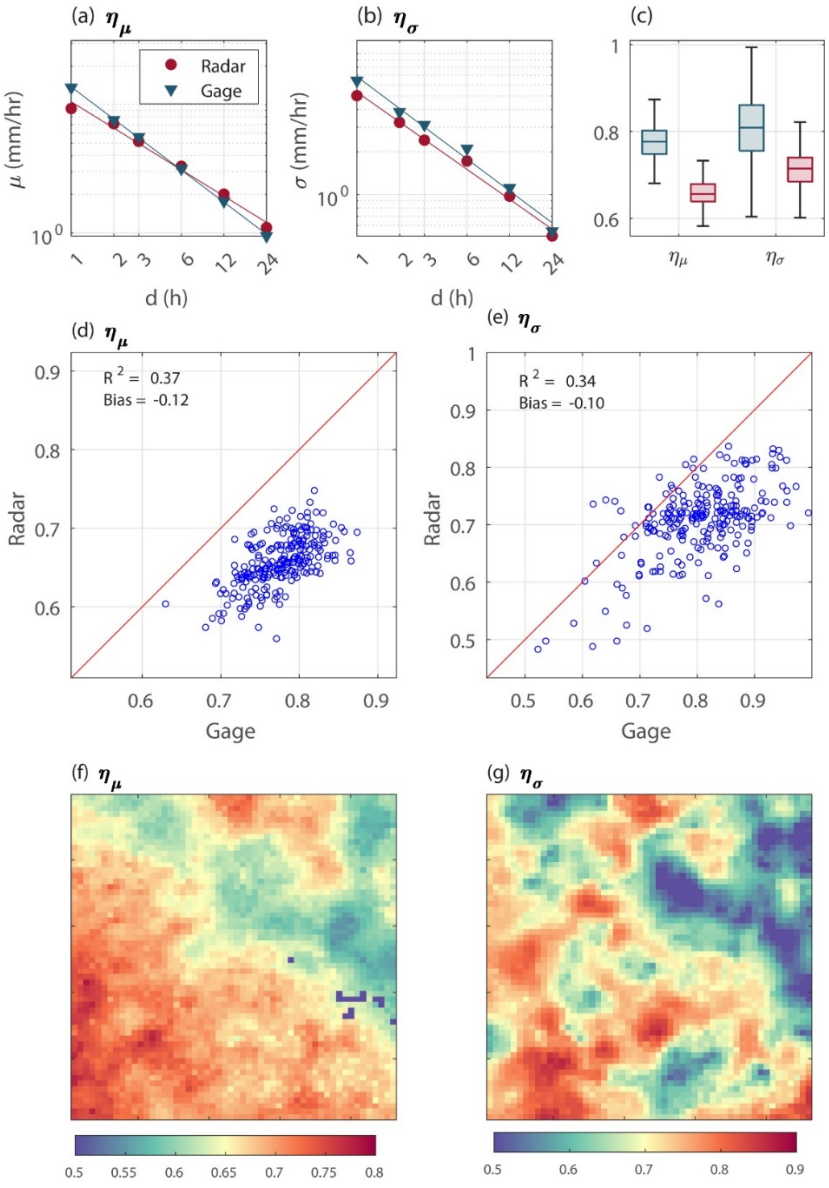

Figure 18. (a)-(b) Example of Linear Scaling Relationship in the Log-Log Plane between P Duration, d, and (a) μ or (b) σ Observed at a Representative Gage and Co-located Radar Pixel. The Scaling Exponents η_μ and η_σ Are the Slopes of the Lines. (c) Boxplots of η_μ and η_σ Estimated from 257 Gages (Blue) and Co-located Radar Pixels (Red). (d)-(e) Scatterplots Between (d) η_μ and (e) η_σ Estimated Using 257 Gages and Co-located Radar Pixels. (f)-(g) Maps of η_μ and η_σ Derived from the Radar MA in the Study Area.

4.7. Utility of Radar to Support NOAA Atlas 14

Our results indicate that extreme P quantiles generated with the spatial interpolation method applied to radar QPEs capture well the NOAA 14 PFEs at d = 24 h. In this section, we show that radar QPEs provide valuable information to address one of the limitations of NOAA 14, which is the presence of abrupt changes of quantile estimates within short distances due to the use of large homogenous regions. This problem has been recently illustrated by Deidda (2021) in Sardinia, Italy. Here, we estimated the 25-year quantile for extreme P at d = 6 h applying the Radar MA method in all State IV pixel in the states of NOAA Atlas 14 Volume 1 (see Figure 1a). Figure 19 shows the comparison of this map with the map of NOAA 14 PFEs. The values are from radar quantiles are smaller however the patterns are preserved as in NOAA 14 PFEs as can be observed from the quantiles observed in Arizona and New Mexico in Figure 19. However, there are areas NOAA 14 PFEs with sharp discontinuities as highlighted in the zoomed version of panel (a), which do not exist in Radar MA, as shown in panel (b).

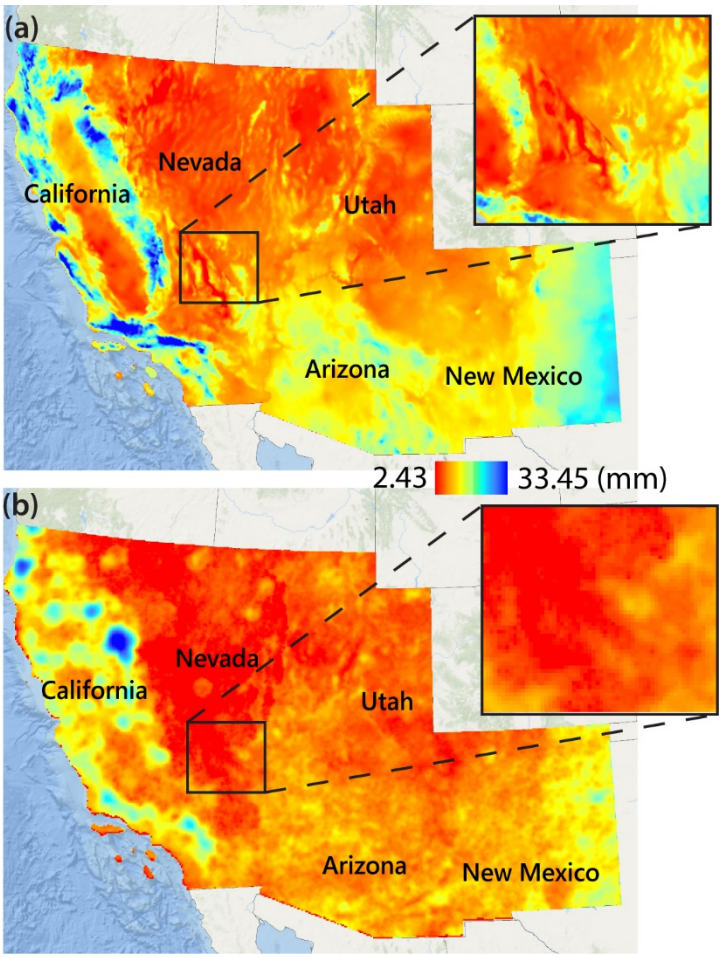

Figure 19. (a) PFE from NOAA Atlas 14 at T_R = 25 years and d = 6 h Emphasizing Non Continuous Values at the State Boundary of California and Nevada. (b) QPEs Estimated Using Stage IV in the States Covered by NOAA Atlas 14 Volume 1.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

The utility of 4 km, 1 h Stage IV radar QPEs to characterize extreme P statistics and improve IDF relations was assessed in central Arizona using a dense network of 257 rain gages as reference. The conclusions of this study are as follows:

1. The suitability of the GEV distribution to model annual P maxima (APM) series at durations, d, from 1 to 24 h obtained from radar and gages was confirmed using the L-moment ratio diagram.
2. At-site estimates of the GEV shape parameter, k, exhibited large variations due to the short sample size (19 years). This uncertainty was reduced using empirical relations based on the analysis of a large number of rain gage records covering the U.S. For all durations, the bias corrected k was found to be > 0, i.e., the distribution of APM is heavy tailed.
3. A hierarchal approach based on spatial interpolation methods (KUD and MA) was designed to further reduce the statistical uncertainty and obtain P quantile estimates that vary seamlessly in space. Estimates of P quantiles from radar QPEs are negatively biased (-20% – -30%) when compared to at-site gage estimates for d = 1 h. The bias tends to 0 and errors (RRMSE of ~10%) are small for d > 6 h, independently of the return period.
4. The use of Moving Average (MA) rather than Kriging for Uncertain Data (KUD) is advised to smoothen the spatial variability of k. The scale and location parameters of the GEV exhibited spatial patterns in part dependent on elevation.

5. Comparison with NOAA Atlas 14 precipitation frequency estimates (PFEs) yields to a further reduction of the bias (-40%) for P quantiles estimated from the radar QPEs at $d = 1$ h. Performance of gage and radars against NOAA Atlas 14 is instead similar for $d \geq 6$ h.
6. The presence of scaling for the location and scale GEV parameters, needed to apply scaling IDF models, was found for both radar and gages. The scaling exponents estimated for the radar were negatively biased compared to the gages, due to the bias identified for the P quantiles at $d = 1$ h. The spatial patterns of the scaling exponents are related to the impacts of localized convective storms, with flatter areas exhibiting larger exponents, i.e., increase in extreme P rates is larger as the time aggregation decreases.

A key idea for improving the analysis for more accurate results is to perform a localized bias correction using a relationship that is derived from rain gage records present in the specific area in which the comparison is performed rather than using one common to the entire US Southwest. Such an optimization will improve the parameter estimates used for deriving precipitation quantiles.

This M.S. thesis is part of the award “SCC: Community-Based Automated Information for Urban Flooding” (#1831475) funded by the National Science Foundation (NSF), which has the main goal of improving collection, analysis, and modeling of flooding in urban areas. The work conducted in this thesis contributes to the broader goals of the project by improving understanding of the accuracy of radar products which are crucial to improve the ability of modeling and forecasting urban flooding.

REFERENCES

Adams, D. K., & Comrie, A. C. (1997). The North American Monsoon. *Bulletin of the American Meteorological Society*, *78*(10), 2197–2213. https://doi.org/10.1175/1520-0477(1997)078<2197:TNAM>2.0.CO;2

Ahmadalipour, A., & Moradkhani, H. (2019). A data-driven analysis of flash flood hazard, fatalities, and damages over the CONUS during 1996–2017. *Journal of Hydrology*, *578*(August), 124106. https://doi.org/10.1016/j.jhydrol.2019.124106

Anderson, B. T., & Roads, J. O. (2002). Regional simulation of summertime precipitation over the southwestern United States. *Journal of Climate*, *15*(23), 3321–3342. https://doi.org/10.1175/1520-0442(2002)015<3321:RSOSPO>2.0.CO;2

Balling, R. C., & Brazel, S. W. (1986). Diurnal Variation in Arizona Monsoon Precipitation Frequencies. *American Meteorological Society*, *115*(Mon. Weather Rev.), 342–346. https://doi.org/https://doi.org/10.1175/1520-0493(1987)115<0342:DVIAMP>2.0.CO;2

Bara, M., Gaál, L., Kohnová, S., Szolgay, J., & Hlavčová, K. (2010). On the use of the simple scaling of heavy rainfall in a regional estimation of IDF curves in Slovakia. *Journal of Hydrology and Hydromechanics*, *58*(1), 49–63. https://doi.org/doi:10.2478/v10098-010-0006-0

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

Blackmon, M. L., Lee, Y.-H., Wallace, J. M., & Hsu, H.-H. (1984). Time Variation of 500 mb Height Fluctuations with Long, Intermediate and Short Time Scales as Deduced from Lag-Correlation Statistics. *Journal of Atmospheric Sciences*, *41*(6), 981–991. https://doi.org/10.1175/1520-0469(1984)041<0981:TVOMHF>2.0.CO;2

Blanchet, J., Ceresetti, D., Molinié, G., & Creutin, J.-D. (2016a). A regional GEV scale-invariant framework for Intensity–Duration–Frequency analysis. *Journal of Hydrology*, *540*, 82–95. https://doi.org/10.1016/j.jhydrol.2016.06.007

Blanchet, J., Ceresetti, D., Molinié, G., & Creutin, J. D. (2016b). A regional GEV scale-invariant framework for Intensity–Duration–Frequency analysis. *Journal of Hydrology*, *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).

Borga, M., Vezzani, C., & Dall Fontana, G. (2005). Regional Rainfall Depth–Duration–Frequency Equations for an Alpine Region. *Natural Hazards 2005 36:1*, *36*(1), 221–235. https://doi.org/10.1007/S11069-004-4550-Y

Burlando, P., & Rosso, R. (1996a). Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *Journal of Hydrology*, *187*(1–2), 45–64. https://doi.org/10.1016/S0022-1694(96)03086-7

Burlando, P., & Rosso, R. (1996b). Scaling and muitiscaling models of depth-duration-frequency curves for storm precipitation. *Journal of Hydrology*, *187*(1–2), 45–64. https://doi.org/10.1016/S0022-1694(96)03086-7

Burnett, A. W. (2013). REGIONAL-SCALE TROUGHING OVER THE SOUTHWESTERN UNITED STATES: TEMPORAL CLIMATOLOGY, TELECONNECTIONS, AND CLIMATIC IMPACT. *Http://Dx.Doi.Org/10.1080/02723646.1994.10642506*, *15*(1), 80–98. https://doi.org/10.1080/02723646.1994.10642506

Carney, M. C. (2016). Bias Correction to GEV Shape Parameters Used to Predict Precipitation Extremes. *Journal of Hydrologic Engineering*, *21*(10), 04016035. https://doi.org/10.1061/(asce)he.1943-5584.0001416

Cayan, D. R., & Peterson, D. H. (1989). *The influence of north Pacific atmospheric circulation on streamflow in the west* (pp. 375–397). https://doi.org/10.1029/GM055P0375

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. In *Springer* (Vol. 44, Issue 4). https://doi.org/10.1198/tech.2002.s73

Coles, S. G., & Dixon, M. J. (1999). Likelihood-Based Inference for Extreme Value Models. *Extremes 1999 2:1*, *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. *Journal of Hydrology*, *273*(1–4), 35–50. https://doi.org/10.1016/S0022-1694(02)00353-0

Corbosiero, K. L., Dickinson, M. J., & Bosart, L. F. (2009). The contribution of eastern North Pacific tropical cyclones to the rainfall climatology of the Southwest United States. *Monthly Weather Review*, *137*(8), 2415–2435. https://doi.org/10.1175/2009MWR2768.1

Dalrymple, T. (1960). Flood-frequency analyses, Manual of Hydrology: Part 3. *Water Supply Paper*. https://doi.org/10.3133/WSP1543A

Deidda, R., Benzi, R., & Siccardi, F. (1999). Multifractal modeling of anomalous scaling

laws in rainfall. *Water Resources Research*, *35*(6), 1853–1867. https://doi.org/10.1029/1999WR900036

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. *Stochastic Environmental Research and Risk Assessment*, *35*(12), 2605–2628. https://doi.org/10.1007/s00477-021-02008-x

Dettinger, M. D., Cayan, D. R., Diaz, H. F., & Meko, D. M. (1998). North–South Precipitation Patterns in Western North America on Interannual-to-Decadal Timescales. *Journal of Climate*, *11*(12), 3095–3111. https://doi.org/10.1175/1520-0442(1998)011<3095:NSPPIW>2.0.CO;2

Du, J. (2011). *EOL data archive -- NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data*. https://doi.org/https://doi.org/10.5065/D6PG1QDD.

Favors, J. E., & Abatzoglou, J. T. (2013). Regional surges of monsoonal moisture into the southwestern United States. *Monthly Weather Review*, *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 Hydrology and Hydraulics 1989 3:4*, *3*(4), 281–292. https://doi.org/10.1007/BF01543461

Fuller, R. D., & Stensrud, D. J. (2000). The relationship between tropical easterly waves and surges over the Gulf of California during the North American monsoon. *Monthly Weather Review*, *128*(8 II), 2983–2989. https://doi.org/10.1175/1520-0493(2000)128<2983:trbtew>2.0.co;2

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 Sensing*, *13*(15). https://doi.org/10.3390/rs13152890

Gubareva, T. S., & Gartsman, B. I. (2010). Estimating distribution parameters of extreme hydrometeorological characteristics by L-moments method. *Water Resources 2010 37:4*, *37*(4), 437–445. https://doi.org/10.1134/S0097807810040020

Gupta, V. K., & Waymire, E. (1990). Multiscaling properties of spatial rainfall and river

flow distributions. *Journal of Geophysical Research*, *95*(D3), 1999–2009. https://doi.org/10.1029/JD095iD03p01999

Guttman, N. B., Hosking, J. R. M., & Wallis, J. R. (1993). Regional Precipitation Quantile Values for the Continental United States Computed from L-Moments. *Journal of Climate*, *6*(12), 2326–2340. https://doi.org/10.1175/1520-0442(1993)006<2326:RPQVFT>2.0.CO;2

Higgins, R. W., Shi, W., & Hain, C. (2004). Relationships between Gulf of California moisture surges and precipitation in the southwestern United States. *Journal of Climate*, *17*(15), 2983–2997. https://doi.org/10.1175/1520-0442(2004)017<2983:RBGOCM>2.0.CO;2

Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, *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. *The American Statistician*, *46*(3), 186. https://doi.org/10.2307/2685210

Hosking, J. R. M., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, *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* (p. 242). Cambridge University Press. https://doi.org/10.1017/CBO9780511529443

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

Innocenti, S., Mailhot, A., & Frigon, A. (2017). Simple scaling of extreme precipitation in North America. *Hydrology and Earth System Sciences*, *21*(11), 5823–5846. https://doi.org/10.5194/hess-21-5823-2017

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. *Hydrological Sciences Journal*, *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. *Hydrological Sciences Journal*, *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 on Water Science*, *2*, 27–77. https://doi.org/10.1016/B978-0-444-53199-5.00027-0

Leathers, D. J., & Palecki, M. A. (1992). The Pacific/North American Teleconnection Pattern and United States Climate. Part II: Temporal Characteristics and Index Specification. *Journal of Climate*, *5*(7), 707–716. https://doi.org/10.1175/1520-0442(1992)005<0707:TPATPA>2.0.CO;2

MacDonald, G. M. (2010). Water, climate change, and sustainability in the Southwest. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21256–21262. https://doi.org/10.1073/pnas.0909651107

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 Resources Research*, *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 Resources Research*, *33*(4), 759–769. https://doi.org/10.1029/96WR03849

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production*. *Bulletin of the American Meteorological Society*, *78*(6), 1069–1080. https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2

Marra, F., & Morin, E. (2015). Use of radar QPE for the derivation of Intensity-Duration-Frequency curves in a range of climatic regimes. *Journal of Hydrology*, *531*, 427–440. https://doi.org/10.1016/j.jhydrol.2015.08.064

Marsily, G. de. (1986). *Quantitative hydrogeology : groundwater hydrology for engineers*. Academic Press.

Mascaro, G. (2017). Multiscale spatial and temporal statistical properties of rainfall in central Arizona. *Journal of Hydrometeorology*, *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. *Journal of Hydrology*, *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. *Journal of Applied Meteorology and Climatology*, *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. *Hydrology and Earth System Sciences*, *17*(1), 355–369. https://doi.org/10.5194/hess-17-355-2013

Mascaro, G., Vivoni, E. R., Gochis, D. J., Watts, C. J., & Rodriguez, J. C. (2014). Temporal downscaling and statistical analysis of rainfall across a topographic transect in northwest Mexico. *Journal of Applied Meteorology and Climatology*, *53*(4), 910–927. https://doi.org/10.1175/JAMC-D-13-0330.1

Mazzetti, C., & Todini, E. (2008). Combining weather radar and raingauge data for hydrologic applications. *Flood Risk Management: Research and Practice*, 1345–1348. https://doi.org/10.1201/9780203883020.CH159

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. *Journal of Hydrology*, *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 and Planetary 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 intercomparisons. *Weather and Forecasting*, *31*(2), 371–394. https://doi.org/10.1175/WAF-D-14-00112.1

Nguyen, V. T. V., Nguyen, T. D., & Wang, H. (1998). Regional estimation of short duration rainfall extremes. *Water Science and Technology*, *37*(11), 15–19. https://doi.org/10.2166/WST.1998.0425

Overeem, A., Buishand, T. A., & Holleman, I. (2009). Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resources Research*, *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 Resources Research*, *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*, *46*(1), 147–155. https://doi.org/10.1080/02626660109492806

Redmond, K. T., & Koch, R. W. (1991). Surface Climate and Streamflow Variability in the Western United States and Their Relationship to Large-Scale Circulation Indices. *Water Resources Research*, *27*(9), 2381–2399. https://doi.org/10.1029/91WR00690

Requena, A. I., Burn, D. H., & Coulibaly, P. (2019). Pooled frequency analysis for intensity–duration–frequency curve estimation. *Hydrological Processes*, *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 Interdisciplinary Reviews: Water*, *5*(6), e1302. https://doi.org/10.1002/WAT2.1302

Sane, Y., Panthou, G., Bodian, A., Vischel, T., Lebel, T., Dacosta, H., Quantin, G., Wilcox, C., Ndiaye, O., Diongue-Niang, A., & Diop Kane, M. (2018). Intensity-duration-frequency (IDF) rainfall curves in Senegal. *Natural Hazards and Earth System Sciences*, *18*(7), 1849–1866. https://doi.org/10.5194/NHESS-18-1849-2018

Sankarasubramanian, A., & Srinivasan, K. (1999). Investigation and comparison of sampling properties of L-moments and conventional moments. *Journal of Hydrology*, *218*(1–2), 13–34. https://doi.org/10.1016/S0022-1694(99)00018-9

Schaefer, M. G. (1990). Regional analyses of precipitation annual maxima in Washington State. *Water Resources Research*, *26*(1), 119–131. https://doi.org/10.1029/WR026I001P00119

Sellers, W. D., & Hill, R. H. (1974). *Arizona climate, 1931-1972* (W. D. Sellers & R. H. Hill (eds.); Rev., 2d ed.) [Book]. University of Arizona Press.

Sheppard, P. R., Comrie, A. C., Packin, G. D., Angersbach, K., & Hughes, M. K. (2002). The climate of the US Southwest. *Climate Research*, *21*(3), 219–238. https://doi.org/10.3354/cr021219

Shukla, J., & Wallace, J. M. (1983). Numerical Simulation of the Atmospheric Response to Equatorial Pacific Sea Surface Temperature Anomalies. *Journal of Atmospheric Sciences*, *40*(7), 1613–1630. https://doi.org/10.1175/1520-0469(1983)040<1613:NSOTAR>2.0.CO;2

Simmons, A. J., Wallace, J. M., & Branstator, G. W. (1983). Barotropic Wave Propagation and Instability, and Atmospheric Teleconnection Patterns. *Journal of Atmospheric Sciences*, *40*(6), 1363–1392. https://doi.org/https://doi.org/10.1175/1520-0469(1983)040<1363:BWPAIA>2.0.CO;2

Stensrud, D. J., Gall, R. L., Mullen, S. L., & Howard, K. W. (1995). Model Climatology of Mexican Monsoon. *Journal of Climate*, 1775–1794. https://doi.org/https://doi.org/10.1175/1520-0442(1995)008<1775:MCOTMM>2.0.CO;2

Stensrud, D. J., Gall, R. L., & Nordquist, M. K. (1997). Surges over the Gulf of California during the Mexican Monsoon. *Monthly Weather Review*, *125*(4), 417–437. https://doi.org/10.1175/1520-0493(1997)125<0417:SOTGOC>2.0.CO;2

Svoma, B. M. (2010). The influence of monsoonal gulf surges on precipitation and diurnal precipitation patterns in central Arizona. *Weather and Forecasting*, *25*(1), 281–289. https://doi.org/10.1175/2009WAF2222299.1

Svoma, B. M., & Balling, R. C. (2009). An anthropogenic signal in Phoenix, Arizona winter precipitation. *Theoretical and Applied Climatology*, *98*(3–4), 315–321. https://doi.org/10.1007/s00704-009-0121-1

Time Series | Billion-Dollar Weather and Climate Disasters | National Centers for Environmental Information (NCEI). (n.d.). Retrieved May 22, 2022, from https://www.ncei.noaa.gov/access/billions/time-series/US

Tyralis, H., & Langousis, A. (2018). Estimation of intensity–duration–frequency curves using max-stable processes. *Stochastic Environmental Research and Risk Assessment 2018 33:1*, *33*(1), 239–252. https://doi.org/10.1007/S00477-018-1577-2

Van de Vyver, H. (2018). A multiscaling-based intensity–duration–frequency model for extreme precipitation. *Hydrological Processes*, *32*(11), 1635–1647. https://doi.org/10.1002/hyp.11516

van Loon, H., & Madden, R. A. (1981). The Southern Oscillation. Part I: Global Associations with Pressure and Temperature in Northern Winter. *Monthly Weather Review*, *109*(6), 1150–1162. https://doi.org/10.1175/1520-0493(1981)109<1150:TSOPIG>2.0.CO;2

van Loon, H., & Rogers, J. C. (1981). The Southern Oscillation. Part II: Associations with Changes in the Middle Troposphere in the Northern Winter. *Monthly Weather Review*, *109*(6), 1163–1168. https://doi.org/10.1175/1520-0493(1981)109<1163:TSOPIA>2.0.CO;2

Vogel, R. M., & Fennessey, N. M. (1993). L moment diagrams should replace product moment diagrams. *Water Resources Research*, *29*(6), 1745–1752. https://doi.org/10.1029/93WR00341

Wallace, J. M., & Gutzler, D. S. (1981). Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. *Monthly Weather Review*, *109*(4), 784–812. https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2

Woodhouse, C. A., & Meko, D. (1997). Number of Winter Precipitation Days Reconstructed from Southwestern Tree Rings. *Journal of Climate*, *10*(10), 2663–2669. https://doi.org/10.1175/1520-0442(1997)010<2663:NOWPDR>2.0.CO;2

Yarnal, B., & Diaz, H. F. (1986). Relationships between extremes of the Southern oscillation and the winter climate of the Anglo-American Pacific Coast. *Journal of Climatology*, *6*(2), 197–219. https://doi.org/10.1002/JOC.3370060208

Yu, P. S., Yang, T. C., & Lin, C. S. (2004). Regional rainfall intensity formulas based on scaling property of rainfall. *Journal of Hydrology*, *295*(1–4), 108–123. https://doi.org/10.1016/J.JHYDROL.2004.03.003

Zhang, Y., Wallace, J. M., & Battisti, D. S. (1997). ENSO-like Interdecadal Variability: 1900–93. *Journal of Climate*, *10*(5), 1004–1020. https://doi.org/10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2

APPENDIX A

RESULTS FROM COMPARISON OF AT-SITE AND SPATIALLY INTERPOLATED GEV PARAMETERS ESTIMATED FOR GAGE AND RADAR

Figures 20 to 22 show scatterplots for different methods at d = 2, 3, and 12 h, respectively.

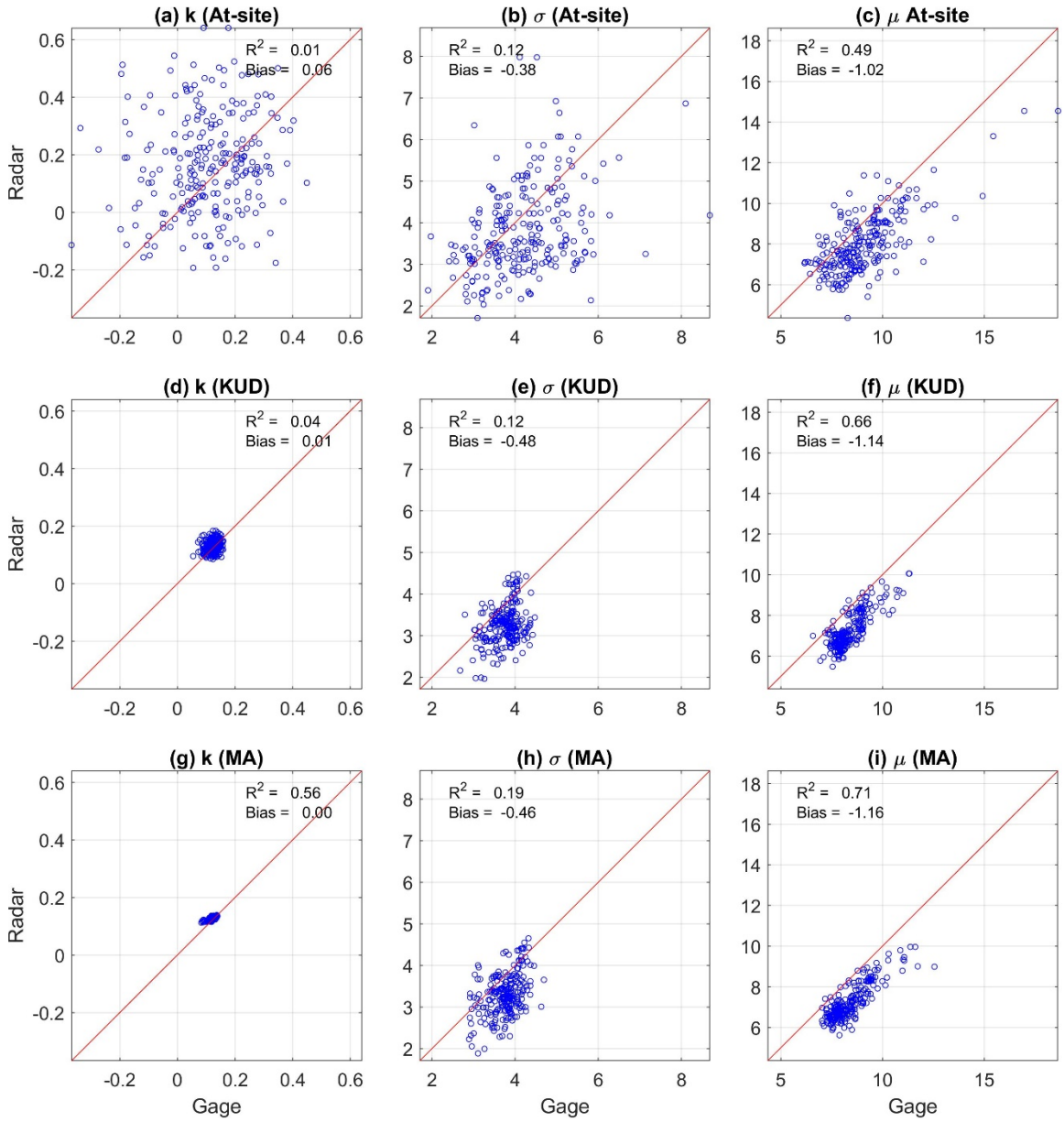

Figure 20. Scatterplots Comparing Radar and Gage Estimates for At-site, KUD, and MA Spatial Interpolation Methods When d = 2 h

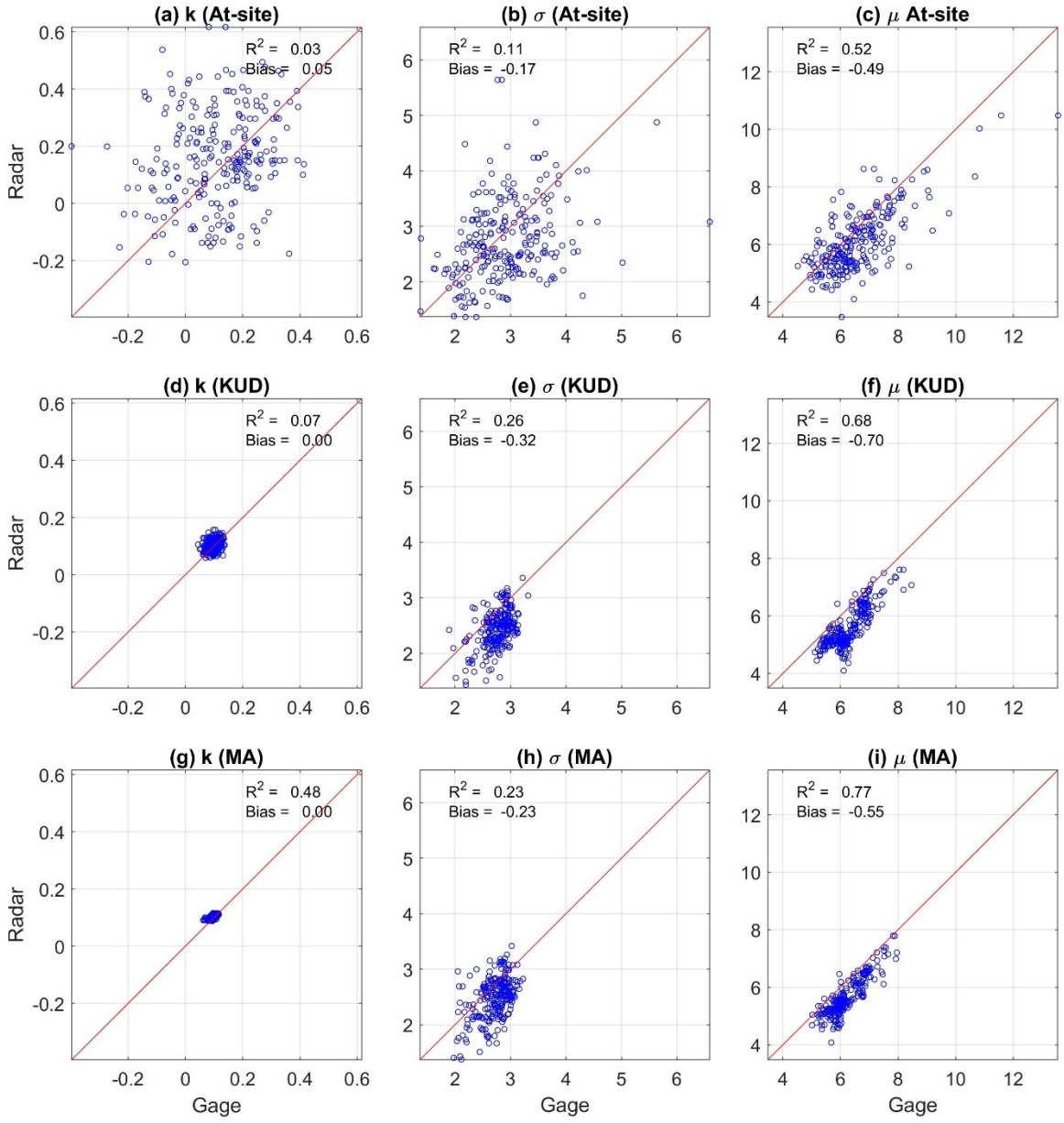

Figure 21. Same as Figure 20 but for d = 3 h

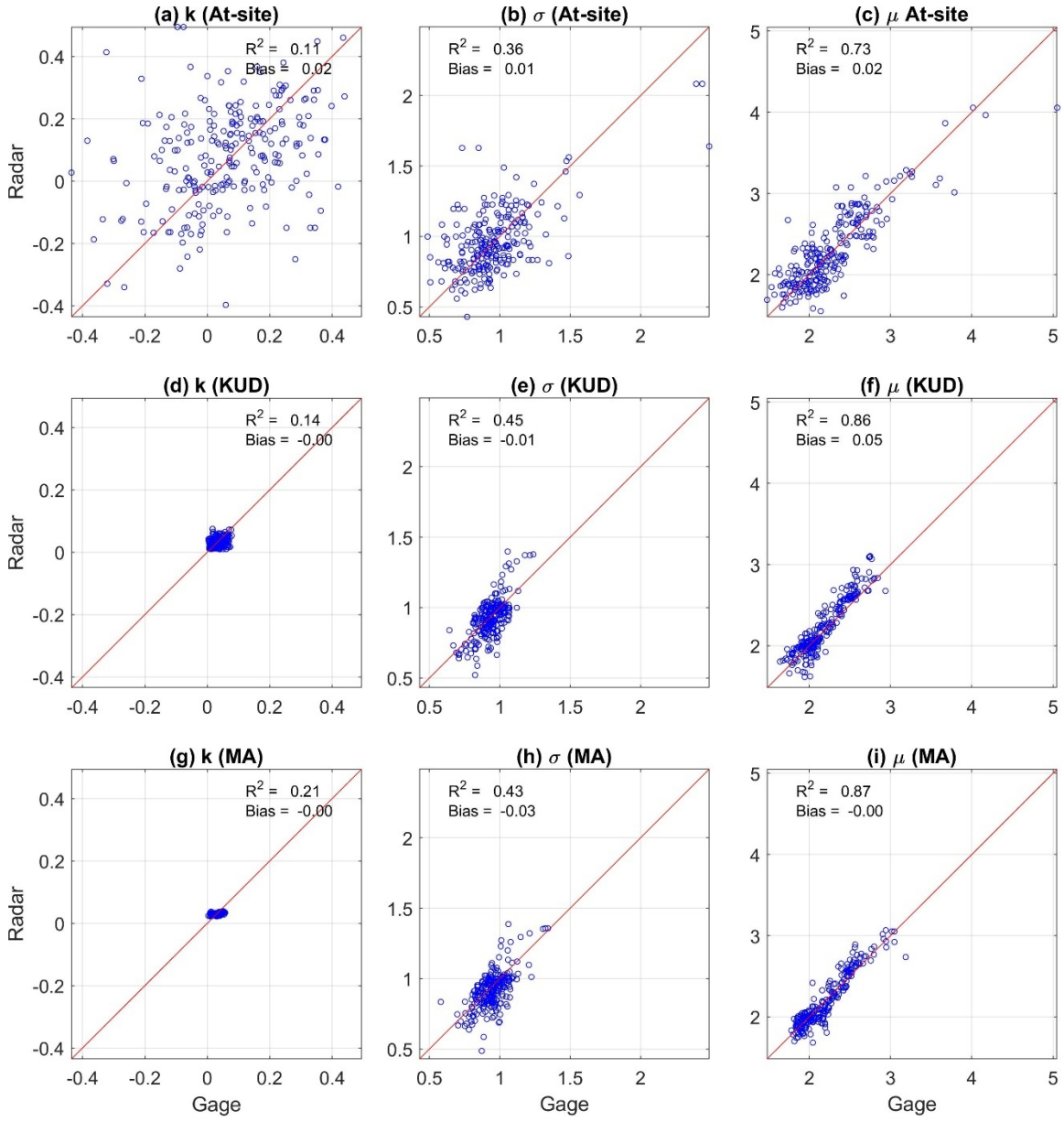

Figure 22. Same as Figure 20 but for d = 12 h

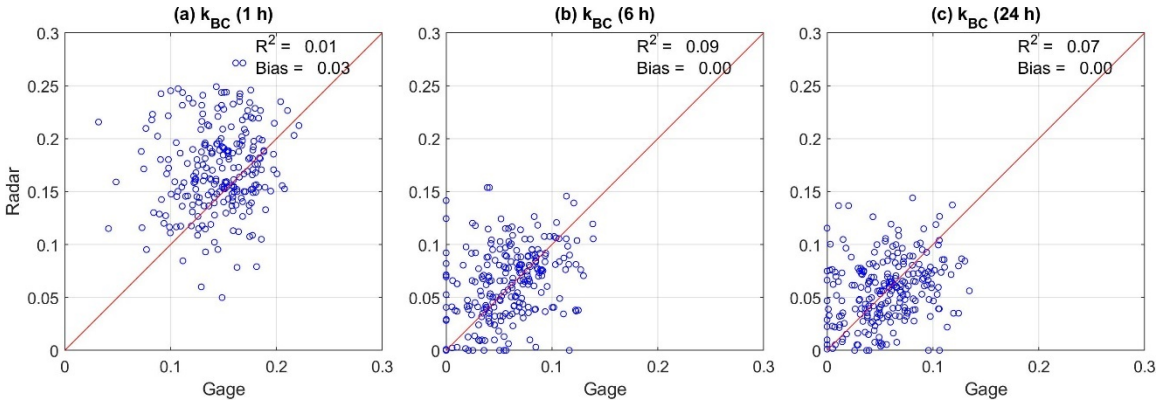

Figure 23. Scatterplots Comparing Radar and Gage Shape Parameter Estimates for At-site BC

APPENDIX B

RESULTS FROM SPATIAL DISTRIBUTION OF GEV PARAMETERS

Figures 26 to 28 show maps of GEV parameter obtained from gages comparing different spatial interpolation techniques at d = 2, 3, and 12 h, respectively.

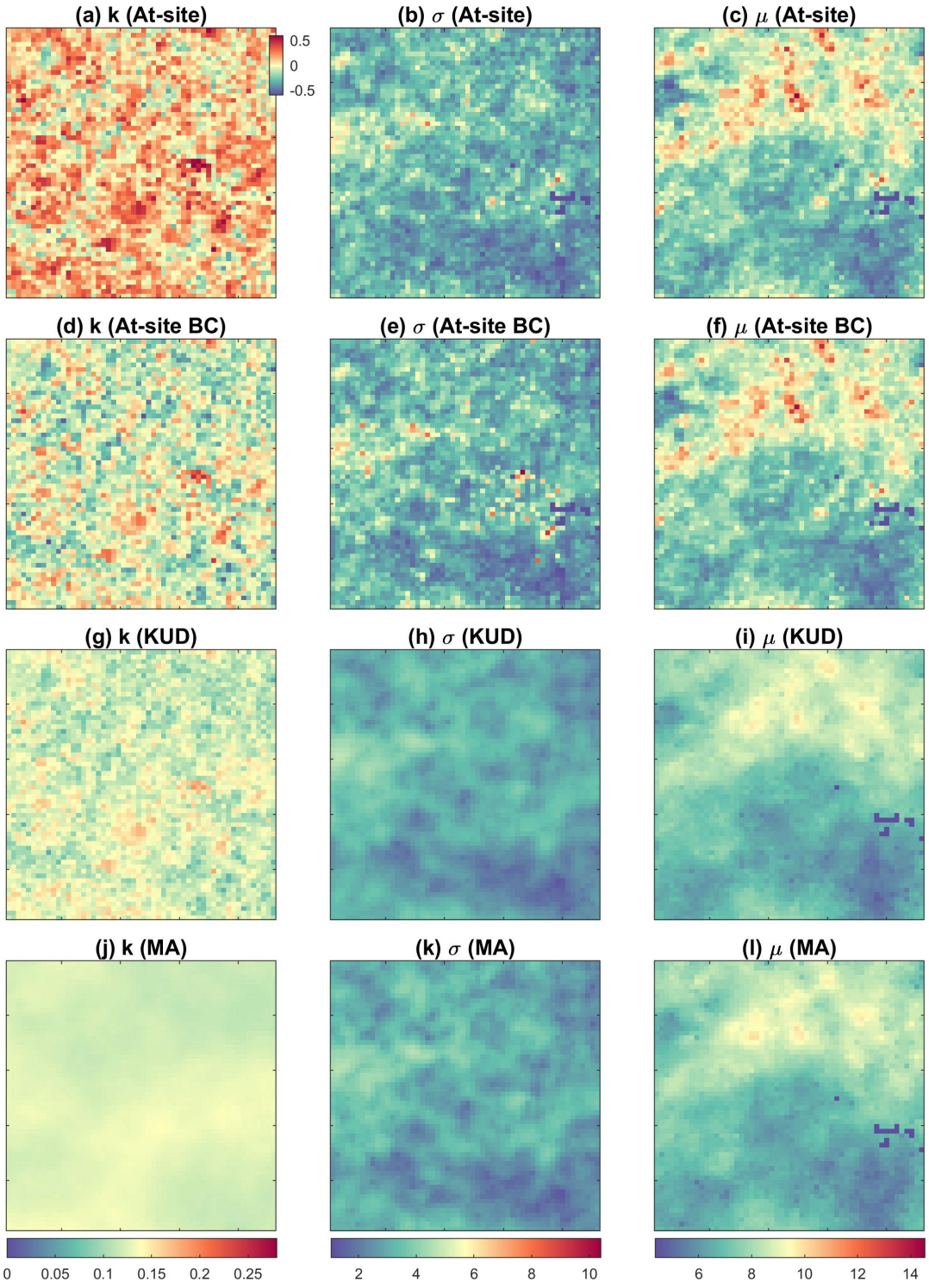

Figure 24. Spatial Variability of the GEV Parameters for d = 2 h Obtained with At-site, At-site BC, KUD and MA Estimation Methods. The Colormaps for Each Parameter are Shown at the Bottom of Each Column (Only Panel (a) Has Its Own Colormap)

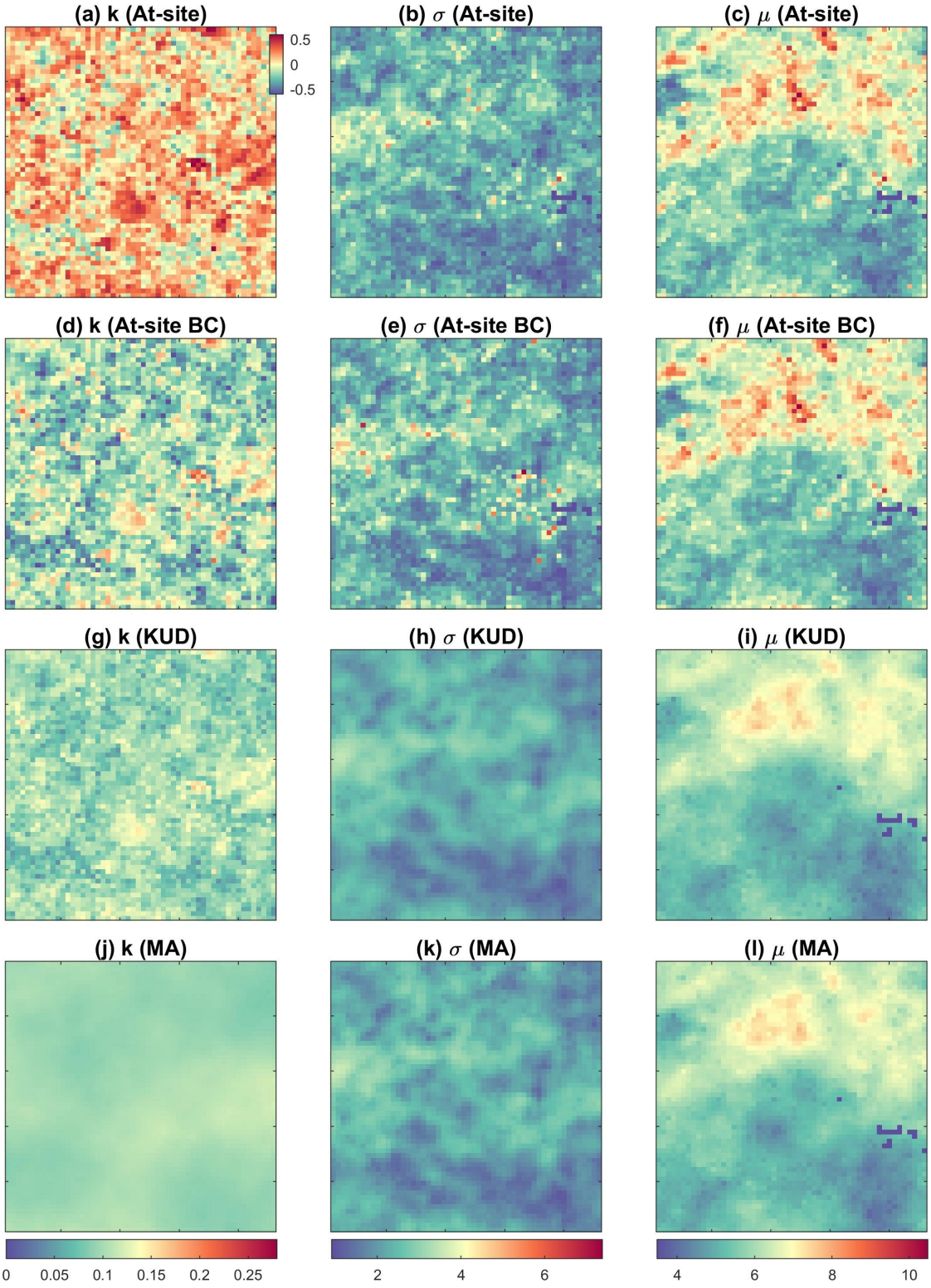

Figure 25. Same as Figure 24 but for d = 3 h.

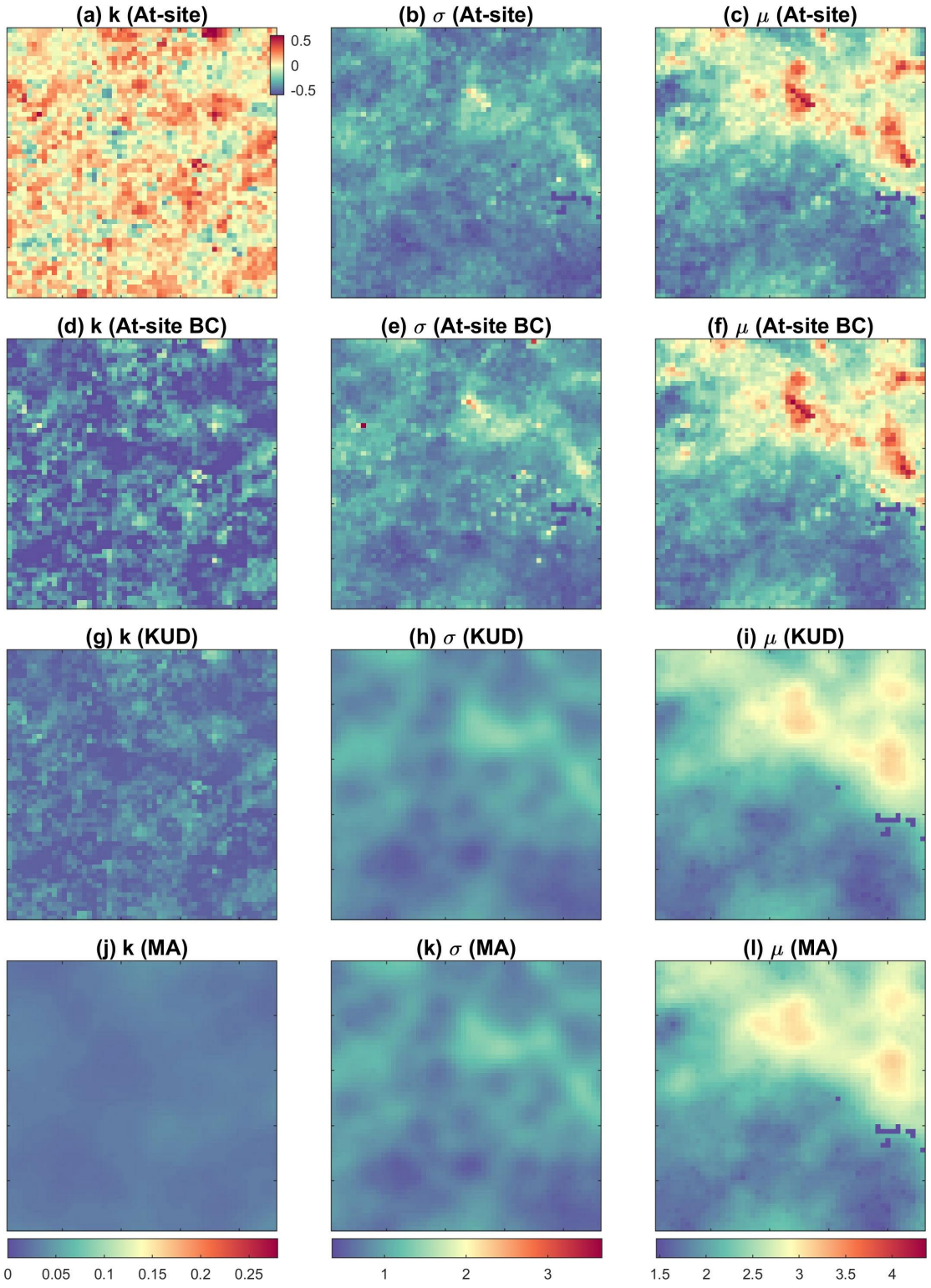

Figure 26. Same as Figure 24 but for d = 12 h.