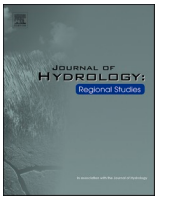
Journal of Hydrology: Regional Studies 44 (2022) 101276

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

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



Assessment of the standard precipitation frequency estimates in the United States

Jungho Kim *, Evelyn Shu, Kelvin Lai, Mike Amodeo, Jeremy Porter, Ed Kearns

First Street Foundation, Brooklyn, NY, USA

ARTICLE INFO

Keywords:
Flood
Extreme precipitation
Precipitation frequency estimate
NOAA Atlas
NEXRAD Stage-IV
Radar-based IDF curve

ABSTRACT

Study region: The Conterminous United States
Study focus: The NOAA Atlases have provided the standard precipitation frequency estimates (PFEs) for over two decades in the United States, but they are losing that status due to climate change. This study evaluates the Atlases compared to new PFEs developed based on the Automated Surface Observing System and Regional Frequency Analysis (ASOS-RFA) as a benchmark and examines a radar-based precipitation product as a data source eligible for developing the post-Atlas.
New hydrological insights for the region: The stationarity assumption of precipitation was highlighted as a significant factor in increasing the uncertainties in the Atlases. The Atlases explicitly tended to underestimate PFEs compared to the ASOS-RFA. The difference between the Atlases and the ASOS-RFA increases as the return period increases and the duration decreases. The age of the Atlases correlated with the difference, suggesting that recent observations diverge from older estimates. The radar-based PFEs are well-matched with the ASOS-RFA at frequent return periods, but the uncertainties increased as the return period increased. This is because the conventional bias correction was limited in improving the annual maximum series of precipitation. Lastly, we need to pay careful attention to an increase in precipitation frequency estimates shorter than 12-hr durations.

1. Introduction

The intensity and frequency of extreme precipitation events have rapidly grown within the majority of the United States in the last two decades (NOAA. U.S. Climate Extremes Index, 2016; Wright et al., 2019). The continuity of this trend is expected to continue and has been considered in projecting dynamic atmospheric phenomenons with many global climate models (Sillmann et al., 2013; Mallakpour and Villarini, 2015). There is a strong argument that flooding driven by extreme precipitation has already exceeded infrastructure design criteria primarily based on Intensity-Duration-Frequency (IDF) curves (Kunkel et al., 2003; Vu and Mishra, 2019; Swain et al., 2020). Therefore, consistently updating IDF curves following an informed evaluation of the risk associated with a change in the trend of extreme weather events, is essential to prevent substantial loss of life and property from catastrophic floods (Keifer and Chu, 1957; Kidd et al., 2017; DeGaetano and Castellano, 2018; Sun et al., 2019). This study aims to assess the standard IDF curves and evaluate radar-based precipitation data as the primary data for developing the post-Atlas in the United States.

An IDF curve describes precipitation intensity (e.g., precipitation frequency estimates, hereafter, PFE) as a function of duration for a

* Corresponding author.
E-mail address: Jungho@firststreet.org (J. Kim).

https://doi.org/10.1016/j.ejrh.2022.101276
Received 28 July 2022; Received in revised form 15 November 2022; Accepted 19 November 2022
Available online 23 November 2022

given return period, which can be converted to a depth-duration-frequency (DDF) curve. IDF curves help to understand the tendency of extreme precipitation in a return period. A map of PFEs from IDF curves is a tool to capture a change in the spatial distribution of extreme precipitation in both gauged and ungauged areas. IDF curves are commonly used for several purposes, including the development of flood maps by representing occurrence probabilities of extreme scenarios (Dupont and Allen, 2000; Elsebaie, 2012), flood design, and hydraulic infrastructure assessment (Mamo, 2015; Yan et al., 2019), stormwater management (Chin, 2004), engineering design and standard development, resiliency evaluation, and vulnerability assessment (Keifer and Chu, 1957; Kidd et al., 2017; Sun et al., 2019). The development of IDF curves typically involves fitting historical precipitation observations to theoretical extreme value distributions and is heavily dependent on the length and quality of the record (Ghebreyesus et al., 2021).

Many countries and local governments have developed standard IDF curve data for their territories to support engineering and scientific fields (Akan, 1993; Martel et al., 2021). The National Weather Service's Hydrometeorological Design Studies Center, National Oceanic and Atmospheric Administration (NOAA) developed a set of stationary IDF curve maps under the project NOAA Atlas 14: Precipitation Frequency Atlas of the United States (hereafter, Atlas 14) for the conterminous United States (CONUS; Wright et al., 2019). Atlas 14 serves as the nation's standard for understanding site-specific precipitation events and provides precipitation depth in various durations and return periods for a unit grid (approx. 1 km $\times$ 1 km), which are statistically derived from measurements collected at ground weather stations, part of the National Weather Service (NWS) Cooperative Observer Program (Leathers et al., 2020). Eleven volumes of Atlas 14 have been developed to aid in better-informed decision-making regarding infrastructure investment, design, and managing flood risks and water resources. Before Atlas 14, Atlas 2 was developed using observation records prior to 1970 for the northwestern U.S. region, including Washington, Oregon, Idaho, Montana, and Wyoming, which Atlas 14 does not cover.

Atlases were developed based on the temporal stationarity of precipitation, which assumes that the occurrence probability of extreme precipitations is not expected to change over time (Cheng and AghaKouchak, 2014). However, climate change has altered the intensity and frequency of extreme precipitation over time, and most climate models project that the features of extreme precipitation will continue to grow throughout the twenty-first century (IPCC, 2014; Donat et al., 2016; Ning et al., 2015). Thus, this has raised practical concerns about assuming stationarity of the annual maximum series (AMS) of precipitation to design infrastructure with a life that extends well into the future (DeGaetano and Castellano, 2018). There were many efforts to consider the nonstationarity of precipitation by adopting and developing various methods to address the expected change in extreme precipitation enhanced by climate change (Katz et al., 2002; Rodríguez et al., 2014; DeGaetano and Castellano, 2017).

Data used in the development of Atlases are also disputed over time. Firstly, the outdated climatic records of Atlases are limited in their validity for application to future climate conditions unless updated to current and future trends (Guo, 2006; Mailhot and Duchesne, 2010; Shrestha et al., 2017). DeGaetano and Castellano (2018) pointed out that the practice of including stations with a record of limited length that end several decades before the present should be avoided. Secondly, the heterogeneity of the gauge data features is another concern since gauge data vary in the quality and quantity of the measurements. The length of a gauge's record is significantly different from another and may be reported at varying time intervals (Leathers et al., 2020). Also, there is inconsistency in the number of gauges used in developing IDF curves for different durations (e.g., 1-min, sub-daily, and daily). The locational distribution of gauges also varies considerably, lacking reliable spatial coverage (Leathers et al., 2020). Sparse gauge coverage thus requires a large portion of data to rely on interpolation and extrapolation (Ghebreyesus et al., 2021), which is problematic provided that the spatial and temporal variability of precipitation events and that the gauges are only small point measurements (Kidd et al., 2017; Gao et al., 2018; Kim et al., 2020). Even in some better situations, adjacent gauges have surprisingly limited correlational ability (Habib et al., 2001).

Many studies have attempted various methods by employing remote sensing data such as high-resolution precipitation data from weather radars (e.g., Next Generation Weather Radar, hereafter NEXRAD, a network of 160 high-resolution S-band Doppler weather radars operated by NWS) and satellites to address the limitations of IDF curves developed based on gauge data. Radar-based precipitation data (i.e., quantitative precipitation estimation, hereafter QPE, derived from a radar reflectivity-rain rate relationship) have been utilized to develop a map of PFE using its ability to represent the spatial distribution of precipitation (Marra et al., 2019; Ghebreyesus et al., 2021). Wright et al. (2020) demonstrated the utility of a QPE product in building storm catalogs as part of the stochastic storm transposition method for producing IDF estimates. A high-resolution spatio-temporal precipitation pattern of remote sensing data is used to improve estimates of areal reduction factors for transforming IDF values from point-based gauge data to areal estimates (Omranian et al., 2018). Transformations are based on the structure of precipitation's spatial correlation (Qiao et al., 2014) and the development of a flood frequency analysis framework reliant on stochastic storm transposition (Wright et al., 2013). The gridded format of these remote-sensing data reduces the uncertainty introduced by interpolation methods required to develop a map of PFEs from gauge data. The common idea in the aforementioned studies employing remote sensing data is taking advantage of high resolutions representing better spatial patterns of precipitation than ground observations. Depending on the quality of remote sensing data, they directly or indirectly used the remote sensing data to develop new PFEs.

While there is ongoing advancement in sensors and algorithms used, there is concern about the accuracy of the data as several validation efforts have noted their limitations in capturing heavy storm events (Prat and Nelson, 2015). A radar observation system is affected by geological factors that block radar reflectivities, such as mountains and earth curvature. Also, there are well-known challenges (e.g., precipitation classification, vertical reflectivity profile correction, and bias correction) to improving QPE at the ground level (Wesson, 2014; Kim et al., 2015; Yoo et al., 2016). Satellite data typically can provide a longer record of data and at a larger scale than radar data. However, satellite data has a lower spatial resolution than radar data and is typically less accurate and sensitive to weather conditions (Kim et al., 2020; Kim and Han, 2021).

Employing the Atlas may not be appropriate in some U.S. regions due to the limitations mentioned above and the current climate circumstances. A few studies assessed the Atlas only for selected states against the developed PFEs, and the studies evaluating the Atlas

over the CONUS with decent data are barely found (Wallis et al., 2007; Ghebreyesus et al., 2021). Since the ease of accessing remote sensing data has been approximately equivalent to that of collecting gauge data, the use of remote sensing data is a technical advancement that should be further developed and it has various other benefits compared to gauge data, such as the capability in representing spatial distributions of precipitation.

This study aims to answer questions that require explicit description before developing a remote sensing data-based IDF map at a continental scale. The two questions are as follows:

(1) Are the Atlases able to represent the precipitation intensity of extreme storms under the current climate conditions?
(2) Could the QPE-driven IDF curves be an appropriate alternative to the Atlases?

The rest of this paper is organized as follows: Section 2 introduces application data, methods, and research strategy. Section 3 presents mixed results that could provide insights and preliminary answers to the questions above, including the Atlases evaluation, the radar QPE evaluation and correction, and a comparison of IDF curves from different sources. Section 4 discusses the potential uncertainties which may arise from the applied procedures and addresses the conclusion of the study. All terminologies with its acronym mentioned in this study are listed in Appendix A.

2. Material and method

2.1. Data

2.1.1. Study area and gauge data

Fig. 1 shows the entire study domain, the CONUS, with the Automated Surface Observing Systems (ASOS) and the gauge locations (hereafter, Atlas gauges) used in the development of the Atlas 14.

ASOS is a well-known weather observation system with more than 913 operational observation stations in the U.S. that provide high temporal resolution data. A 60-minute precipitation measurement of ASOS is used as the primary data to develop the benchmark PFEs, evaluate Atlases, and correct biases in Stage-IV. ASOS data from 2002 to 2021 are used; we believe the period represents the current climate circumstances. The ASOS data is available from the online site (https://mesonet.agron.iastate.edu/request/asos/1 minphtml).

Atlas gauges consist of various ground-based observation systems operated by various agencies. According to Atlas volume 1: Semiarid Southwest, the primary source of precipitation data, is the National Weather Service Cooperative Observer Program's daily and hourly stations. As for the other sources, precipitation data from county flood control districts, the state department of water resources, and the U.S. Geological Survey are employed. For more detail, the Atlas public reports are available for each region (https://www.weather.gov/owp/hdsc_currentpf). This study used the AMS data in the Atlas gauges that the NOAA provides. The AMS data is used to replicate the PFEs under various conditions to validate the Atlases. Also, the locations of the Atlas gauges are used as the sampling points to evaluate the Atlases in comparison to ASOS-RFA quantitatively.

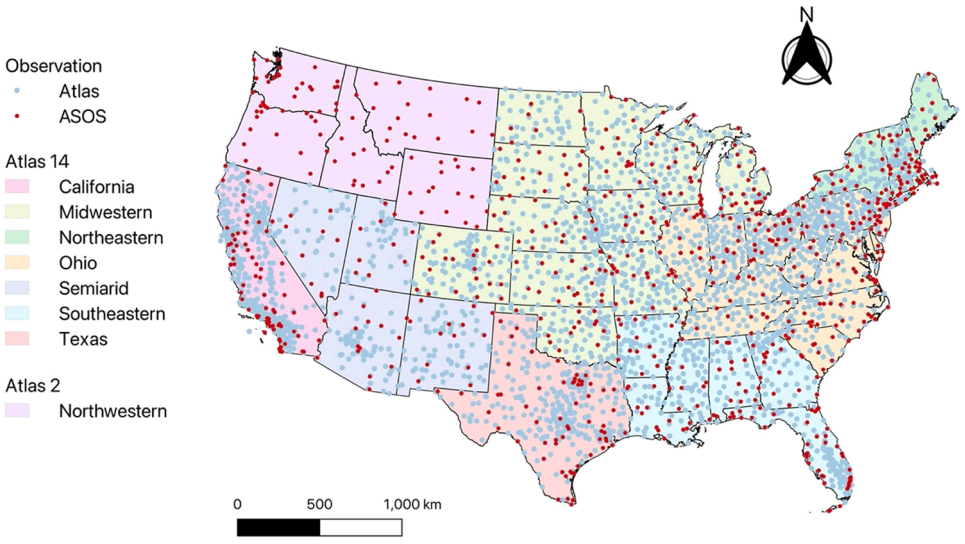


Fig. 1. Application domain and precipitation gauge locations used in the analysis: Atlas gauge indicates weather stations used in the development of NOAA Atlas: California (CA) developed in 2011, Midwestern states (MW) in 2013, Northeastern states (NE) in 2015, Northwestern states (NW) in 1973, Ohio River Basins and surrounding states (OH) in 2004, Semiarid states (SA) in 2004, Southeastern states (SE) in 2013, and Texas (TX) in 2018 are targeted.

2.1.2. NOAA Atlas

NOAA developed a range of Atlas PFEs representing frequent and extreme precipitation scenarios in the United States. The PFEs for durations from 5 min to 60-day and annual exceedance probabilities from ½ to 1/1000, corresponding to 2- to 1,000-year annual recurrence intervals, are provided. NOAA Atlas exists as either Atlas 2 or 14, depending on the development years and regions. Atlas 2, the first version of NOAA Atlas, was developed for 11 states (Montana, Wyoming, Colorado, New Mexico, Idaho, Utah, Nevada, Arizona, Washington, Oregon, and California) located west of 100 W longitude.

In Atlas 2, NOAA provides the gridded IDF curve based on the AMS approach for the northwestern U.S. region covering the states of Washington, Oregon, Idaho, Montana, and Wyoming. The primary data sources were Climatological Data for the U.S. by sections and hourly precipitation data. The precipitation record used in the development of Atlas 2 is outdated by over 50 years, suggesting that Atlas 2 might not be eligible to represent extreme storm events under current climate conditions due to the accelerated climate change. The gridded IDF curve data is available online (https://www.weather.gov/owp/hdsc_noaa_atlas2). The data is limited to two storm durations (6- and 24-h) and two return periods (2- and 100-years).

Since the early 2000s, the NOAA has been updating PFEs for various regions of the United States and affiliated territories. Updated estimates with relevant supplementary information are published in NOAA Atlas 14 "Precipitation-Frequency Atlas of the United States." Atlas 14 updates are done in stages based on state boundaries. The first volume of Atlas 14 was released in 2004, and the latest volume for Texas was finished in 2018. As it took around 20 years to develop Atlas 14 for various regions, the period of precipitation data used in each region is different. Volume 1 for the Semiarid Southwest region used precipitation records up to 2000, and Volume 11 for the state of Texas used up to 2017. This is a point that many previous studies are concerned about regarding the inconsistency of considered years and the absence of recent years (except for Texas) of used precipitation records (Lopez-Cantu and Samaras, 2018; Wright et al., 2019). The cell size of Atlas 14 is approximately 1.0 km (0.00833 degrees). In the latest volume of Atlas 14 for Texas, PFEs have been derived for a range of frequencies and durations using a regional frequency analysis approach based on L-moment statistics calculated from the AMS.

In the development of Atlas 14, the GEV was selected for all regions in the CONUS. The volumes 1 and 2 of the Atlas provide the goodness-of-fit test result, which generally recommends the GEV as it showed the best-fit to data for most regions than any other distributions. Since the volume 6 (the third volume for the region in CONUS), the GEV was not the best-fit distribution, but acceptable and selected for consistency in frequency estimates over the CONUS, including Texas.

2.1.3. NEXRAD Stage-IV

The NWS/National Centers for Environmental Prediction (NCEP) Stage-IV is the QPE product used in this study. Stage-IV is a near-real-time product that is generated at NCEP separately based on the NEXRAD precipitation processing system (Fulton et al., 1998) and the NWS River Forecast Center (RFC) precipitation processing (Nelson et al., 2016). The Stage-IV used in this study is based on the multi-sensor hourly Stage-III analyses (on local 4 km polar-stereographic grids) produced by the 12 RFCs in CONUS. NCEP mosaics Stage-III into a national product (i.e., Stage-IV). In addition, Stage-IV differs from the NCEP Stage-II chiefly in that the NCEP Stage-II contains no manual quality control, while Stage-IV benefits from manual quality control performed on the Stage-III data at the RFCs. For more information, the Stage-IV precipitation processing details are well documented in Nelson et al. (2016).

Many studies have used Stage-IV as a reference and benchmark for analysis and comparison with comparable precipitation products based on remote sensing observation systems, suggesting that the overall quality and accuracy of Stage-IV are acceptable for hydrological purposes. Especially, Stage-IV showed a good performance at a high rain rate (>90th percentile) (Nelson et al., 2016). However, since the performance of Stage-IV varies with regions, a part of this study includes the evaluation of Stage-IV against gauge observations.

2.2. Methodology

2.2.1. Framework

Three analyses were mainly implemented to answer the two questions raised in the introduction: (1) Validation of the Atlases from the data and methodology perspectives, (2) Evaluation of the Atlases using a benchmark, and (3) Evaluation of the Stage-IV IDF curves. Before the primary analyses, the pre-processing to fill missing data and detect/remove outliers was implemented for ASOS and Stage-IV data. A simple interpolation method using a time window as a parameter was employed to fill missing data. Outliers were detected using the inter-quartile range (IQR) method to define outliers' upper and lower bounds. The IQR method has a scale factor to adjust the bounds and is 1.5 in general. This study considered the scale factor for only the upper bound and determined it based on a sensitivity analysis (scale factor = 5.0). Detected outliers were removed and replaced with the interpolated value. 1-min precipitation of ASOS was accumulated and converted to 60-min duration precipitation to match the temporal resolution of Stage-IV.

In the validation of the Atlases, the AMS used in the development of the Atlas 14 in Texas was employed since the latest period of AMS data is available up to 2017 (the last year of the data is 2018, but it is excluded due to the relatively few numbers of gauges). Fundamental analyses to understand a trend of the AMS, validate the selected probability distribution, and examine the effect of precipitation nonstationarity on developing IDF curves were mainly used. The Automated Surface Observing System and Regional Frequency Analysis (ASOS-RFA) was developed based on the recent 20 years of ASOS data from 796 gauge locations across the CONUS to evaluate the Atlases. ASOS was selected since the density as a single gauge observation system is comparable with the Atlas gauge network and able to provide the latest precipitation data. In the evaluation process, the difference defined as the difference percentage between PFEs from ASOS-RFA and Atlas was used as a criterion for measuring uncertainties in the Atlas. The Atlas was paired with ASOS-RFA by sampling PFEs, collocated with the selected ASOS gauges.

To examine the eligibility of Stage-IV for developing IDF curves, hourly Stage-IV precipitation data was evaluated against ASOS data using the bias correction factor (BCF). BCF was selected as it measures systematic errors against gauge data and can directly correct biases in the Stage-IV to improve the accuracy of AMS. Bias correction was implemented to improve the accuracy of hourly precipitation data from Stage-IV using 60-min ASOS precipitation data. To understand the accuracy of extreme precipitation estimates of Stage-IV, the AMS of the corrected and uncorrected Stage-IV were evaluated against ASOS data. The corrected data were used to develop the Stage-IV-based PFEs.

The CONUS is the primary domain to develop new PFE, assess the NOAA Atlas, and answer the questions raised in the introduction. Texas is used to review and validate the methodology developed in the NOAA Atlas and this study as the state is the latest development (i.e., the last volume of the NOAA Atlas 14) based on the recent period of record.

2.2.2. Frequency analysis

Regional frequency analysis (RFA) based on the index-flood procedures was employed to develop new PFEs from ASOS and Stage-IV precipitation data. The RFA was developed to derive more reasonable IDF curves by analyzing multi-site sampling data in a homogeneous region rather than using single-site data (Hosking and Wallis, 2002). The RFA is similar to the at-site frequency analysis, mainly consisting of the four steps: the AMS preprocessing, fitting AMS to theoretical distributions, selecting a distribution, and deriving PFEs for a specific scenario of a storm (e.g., duration and return period). However, the RFA requires a regionalization process by categorizing groups of gauges having statistical similarities. In the general RFA process, the quantile function of the frequency distribution at site i is defined as follows:

$$Q_i(F) = \mu_i q(F), i = 1 \ldots N \tag{1}$$

where, (a) μ_i is a scaling factor defined as the index-flood, (b) $q(F)$ is the regional growth curve as a function of F ranging from 0 to 1, (c) F is an annual exceedance probability in precipitation frequency analysis and (d) N is the number of gauges in a homogeneous region.

A homogeneous region, in this context, is a group of gauges with similar statistics of selected variables such as L-moments. The L-moments are an alternative system of representing the shape of distributions, and they historically arose as modifications of the conventional probability-weighted moments (PWM) (Hosking and Wallis, 2002). In the index-flood procedures, a regional growth curve was derived from multiple gauge data in a homogeneous region under the assumption that the probability distributions of N-gauges are identical except for the scaling factor of each gauge. Hence, this method is beneficial for reducing uncertainties arising from a short period of record and the shortage of various extreme cases. To identify homogeneous regions, K-means was employed as a clustering method. The clustering method seeks gauges with similar characteristics and groups them as homogeneous regions. Coordinates, gauge altitude, mean annual maximum precipitation, and L-moments ratios (e.g., L-CV and L-skewness) were used as variables to cluster groups.

A goodness-fit-test was implemented to choose the optimal probability distribution using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for various durations. Parameters of the selected probability distribution were estimated using the L-moment and PWM methods. For this process, the L-moments3 library in Python was used. In this study, the four 3-parameters distributions (GEV, GLO, GNO, and GPA) were considered to validate the methodology of the NOAA Atlas. For more details, readers may care to refer to Hosking and Wallis (2002) as the RFA and the associated index-procedures in this study are developed based on that reference.

2.2.3. Bias correction

Correcting biases of quantitative precipitation estimation from a weather radar is one of the most critical problems regarding radar precipitation products. A bias correction factor (BCF) was used to evaluate Stage-IV and reduce biases in the Stage-IV quantitative precipitation estimates. The G/R ratio approach, the ratio between a gauge measurement and radar precipitation estimate, is the most typical form of the BCF, indicating the extent of systematic errors. However, bias is intrinsically defined as the difference between the actual and sample means. The fact that the BCF takes the form of the ratio between the gauge rate and the radar precipitation estimate instead of the difference between them indicates that both precipitation intensities follow the log-normal distribution. The logarithmic mean difference between the gauge and radar precipitation estimate is converted into the ratio between means (Yoo et al., 2014).

The other way of explaining the BCF is the regression analysis instead of employing the G/R ratio. The linear regression line can also be used as the BCF because the G/R ratio can be interpreted as the slope of the line connecting the origin and the mass center of measurement pairs (Seo et al., 1996). This approach is straightforward and has adequate performance for correcting bias in a time series of Stage-IV precipitation. Thus, this study used a linear regression approach to estimate BCF. BCF was estimated by calculating the slope of the well-known linear regression line, determined as follows:

$$\beta = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \Big/ \sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{2}$$

where the precipitation values are recorded at the time i and at gauge x_i with the collocated Stage-IV grid point y_i, and n is the length of the designed time window.

Bias correction was implemented for individual storm events as it varies with the type of storm. Individual storm events used Interstorm Event Duration Time (IETD) to classify an individual storm event from a time series of precipitation (Yoo et al., 2016). In this study, 6 hrs of IETD were selected through a sensitivity analysis.

Instead of considering all storm events with various durations, a storm event of greater than 6-hr duration was considered for estimation, as BCFs for storm events of a shorter duration (especially in a case of storm events with less than 6-hr duration) are abnormally high or low compared to its typical range (e.g., $0.5 < BCF < 2.0$). Bias correction was implemented by pairing the extracted Stage-IV precipitation data with ASOS gauge data at the collocated sites. The designed time window indicates a storm duration as BCF is calculated for each storm event.

3. Results

3.1. Validation of the Atlas in Texas

A fundamental analysis of the AMS used in the development of Atlas in Texas is implemented. Fig. 2 shows the analysis results of the spatial and temporal features of the AMS. The observation period is from 1887 to 2017, but considering variations in time series of precipitation and the number of gauges in Fig. 2(a), a period of data since 1950 is considered for further analyses. Overall, the nonstationarity of precipitation is observed from various perspectives. In Fig. 2(b), an increase in the mean value and variation of the AMS is consistently observed until 2017 across all percentiles. In Fig. 2(c), the increase in the mean value was 22 % at the 50th

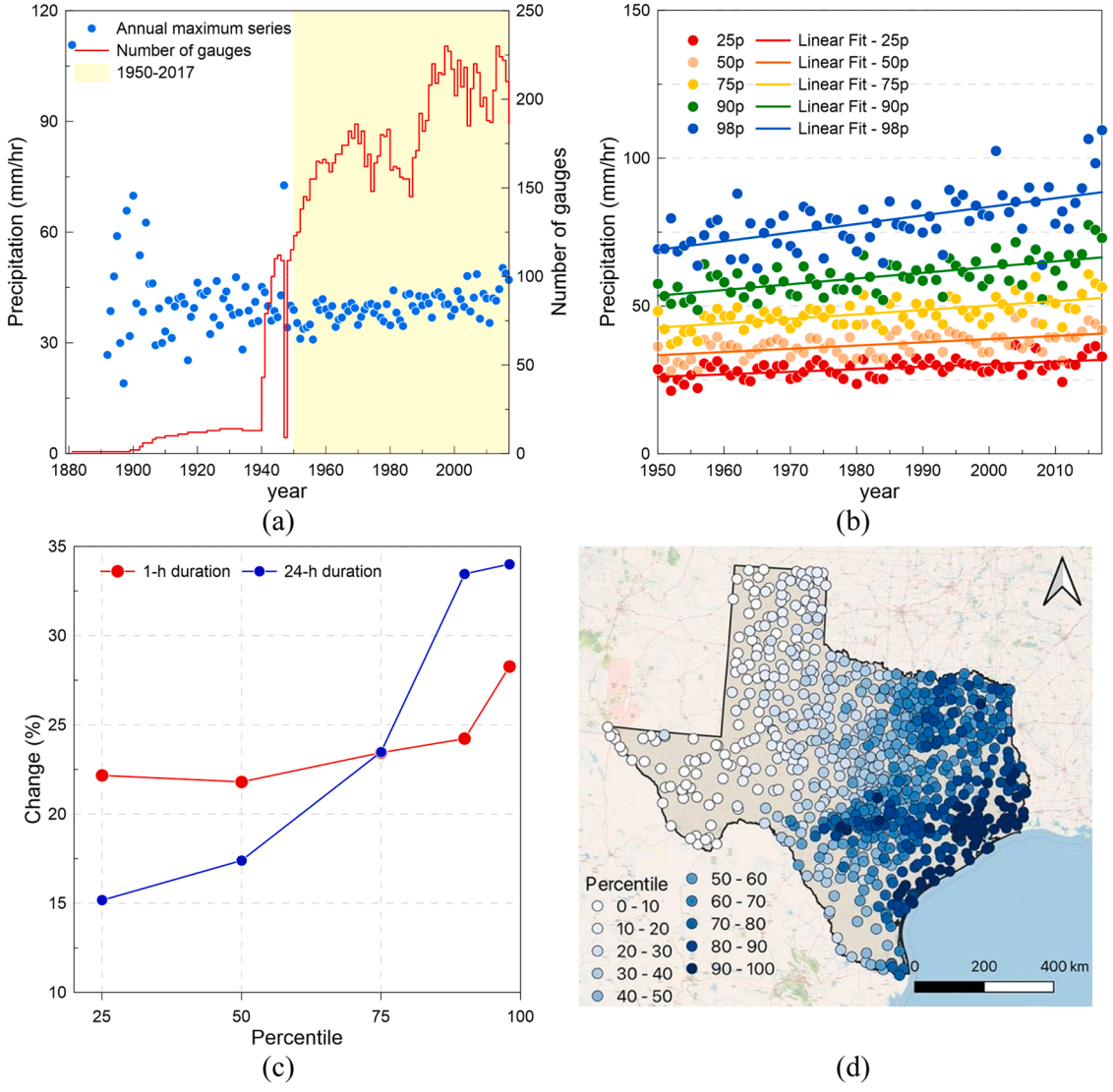


Fig. 2. Spatial and temporal characteristics of the AMS in Texas: (a) shows time series of the mean 1-hr AMS and the number of gauges used to calculate the AMS. (b) shows time series of mean 1-hr AMSs for five percentiles, 25, 50, 75, 90, and 98. 'Change (%)' in (c) represents an increment in the AMS between 1950 and 2017 for the five percentiles. The changes are derived from the regression lines in (b). (d) is a map to represent the spatial distribution of percentiles of the mean AMS for the 24-hr duration.

percentile and 28 % at the 98th percentile for 1-h precipitation. The increasing trend was dramatic in 24-h precipitation. The increment in the 98th percentile was two times larger than that in the 50th percentile. In Fig. 2(d), the spatial distribution of the mean AMS varied with the geographical locations of the gauge. Higher percentile groups are located in coastal areas, while lower percentile groups are populated in the northwestern inland area. Collecting altogether, the increment in the AMS was proportional to its magnitude across the state and the AMS in the coastal areas has been more developed than in the inland areas.

The goodness-of-fit test is implemented to validate the probability distribution selected in the Atlas. The four 3-parameters distributions (GEV, GLO, GNO, and GPA), which presented acceptable goodness-of-fit results over the state in the NOAA Atlas report, were considered. Fig. 3 shows the comparison result of the four probability distributions. In Fig. 3(a), GNO has the most significant fraction (50 % on average of the six durations from 1- to 48-h) among the four probability distributions, and then GEV (31 %) and GLO (11 %) are high in the order. In Fig. 3(b), GNO is uniformly distributed while GLO and GEV are limited in some locations relatively.

The primary driver of uncertainty in regional frequency analysis is the choice of a probability distribution (Al Kajbaf and Bensi., 2021). Particularly, it is significant for PFEs with rare return periods, which have a high possibility to occur flooding. Fig. 3(c) compares the PFEs as a function of AEPs derived from the four distributions. All probability distributions provide quite similar PFEs for the AEPs of greater than 2 %, meaning that PFEs for under 50-year return periods might have no significant differences depending on the distributions. After that, it was obvious that the difference increased as the AEP decreased. Considering a period of the Atlas gauge record shorter than 50 years (41 years on average in Texas), the uncertainty for over 50-year return periods is inevitable and understandable. These results suggest that the choice of a probability distribution should be implemented from the technical and objective perspectives aiming to develop optimal PFEs for extreme events. Lastly, there is a possibility that the Atlas in Texas has provided overestimated PFEs for the high return periods because the GNO, the best-fit distribution, presents smaller PFEs for over a 50-year return period than that from the GEV.

To examine the effect of a time window of record on deriving PFEs under climate change, the PFEs derived from the three time windows (1958–1977, 1978–1997, and 1998–2017) are compared in Fig. 4. From Fig. 4(a)-(b), the overall spatial patterns are fairly similar except for the magnitude of the PFEs in the coastal area. However, the PFEs derived from 1998 to 2017 in Fig. 4(c) presented different features, especially in 'A' and 'B' areas, the cities of Austin and Houston, respectively. In Fig. 4(d), comparing the median values, the period of 1998–2017 is 25.5 % and 21.1 % higher than the two former time windows, respectively. In Fig. 4(e) and (f), the PFEs from the different time windows strongly correlated with each other as the correlation coefficients (CC) for both are above 0.95.

The difference arising from the different time windows is because of the nonstationarity of precipitation. It influences determining two factors in RFA: the scale factor (i.e., index flood) and the shape of the probability distribution (e.g., tails). Most previous studies, including the Atlas, employed the mean value of the AMS as the scale factor, which varies with a time window. As shown in Fig. 2(b), the increasing trend of the AMS over time is observed. It is confirmed that 5.0 % and 6.5 % of the mean AMS increased in 1978–1997 and 1998–2017, respectively, compared to the former time window.

The right tail of the probability distribution is the most sensitive portion as it determines the PFEs in lower AEPs (i.e., irregular events). An increase in variation of AMS is important since significant variation influences estimating parameters of the selected probability distribution. In Fig. 5, Intensity-frequency (IF) curves from different time windows are compared to examine the effect of changes in precipitation variation on PEFs. To minimize the effect of the difference in the mean AMSs on the analysis, PFEs are normalized with the mean AMSs for each time window. The mean AMSs were bigger from 1998 to 2017, 1958–2017, 1978–2977, and 1958–1977 in order. The difference increases as AEP decreases, meaning that the PFEs at the high return periods have more uncertainty due to changes in precipitation variability. In the extreme case corresponding to a 1,000-year return period (i.e., 0.1 AEP), the

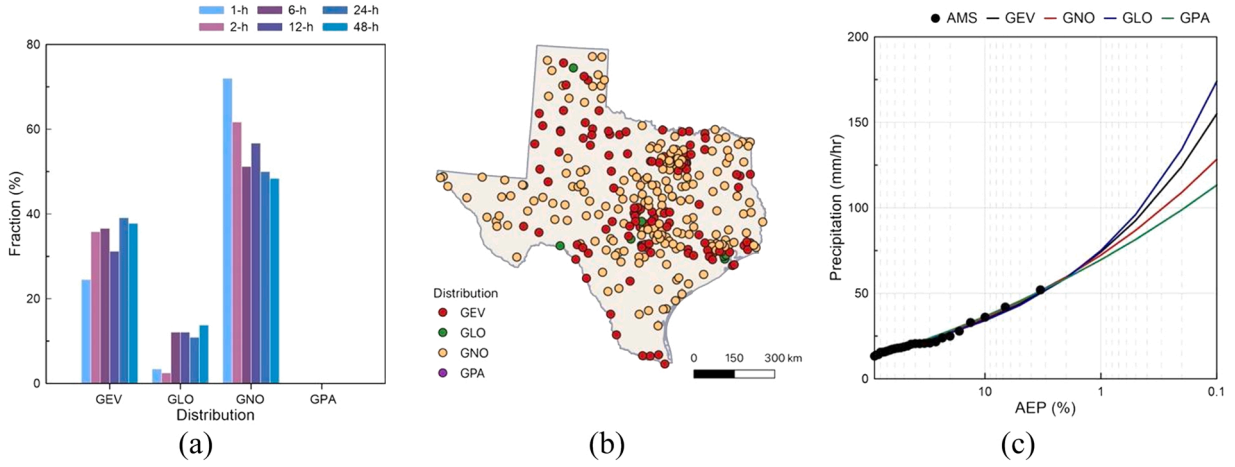


Fig. 3. Comparison of the probability distributions (GEV, GLO, GNO, and GPA) for the regional frequency analysis in Texas: (a) is 'Fraction (%)' of the best-fit probability distributions depending on durations. (b) is the spatial distribution of the best-fit probability distribution based on the goodness-fit test for 1-hr duration. (c) is a comparison result of the PFE and annual exceedance probability (%) curves derived from the four distributions for a scenario of a 100-yr return period and 6-hr duration. The result is an example from a gauge station (site#: 60–0011).

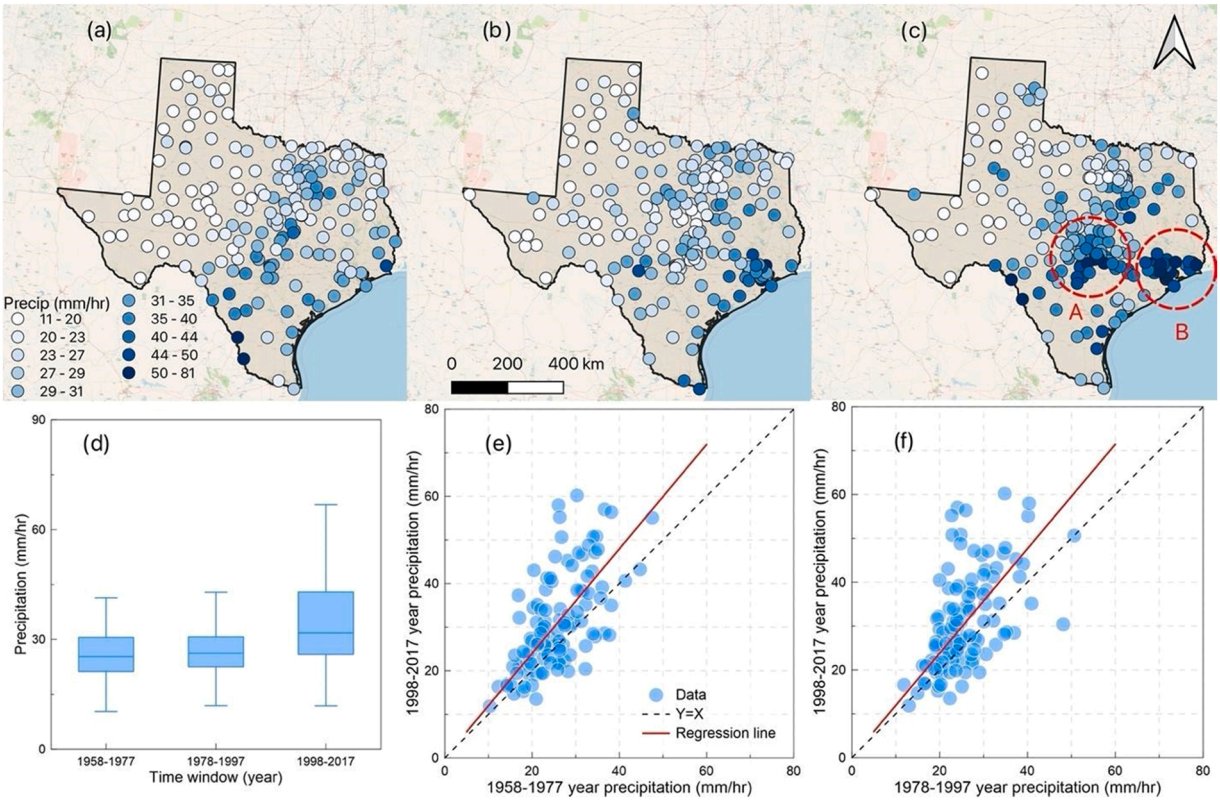


Fig. 4. Comparison of PFEs in different time windows: (a), (b), and (c) present the maps of the spatial distribution of PFEs (intensity, mm/hr) for a scenario of a 100-yr return period and 6-h duration for three time-windows of 20 years: (a) 1958–1977, (b) 1978–1997, and (c) 1998–2017. In (c), the red dotted circles, A and B, indicate the highlighted areas, the city of Austin and Houston, respectively. (d) presents a range of PFEs in the three time-windows. (e) and (f) show a relation between PFEs from a 1998–2017 period and the other two time-windows.

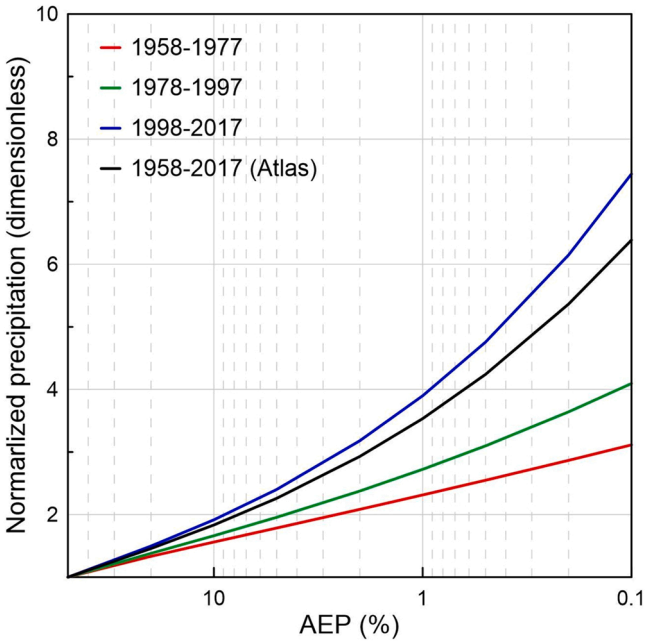


Fig. 5. Comparison of intensity-frequency curves from the four time windows: The IF curves describes PFEs as a function of the AEP for a 6-hr duration case. Y-axis is precipitation intensity normalized by the mean AMS. The results are from an Atlas gauge station (site#: 41–3640).

PFE from 1998 to 2017 is over two times larger than that from older time windows. Also, the Atlas was smaller than PFEs from 1998 to 2017, and it is suspected that it is due to a more extended period of data with minor precipitation variation compared to the recent 20 years.

The effect of interpolation on PFEs is examined to understand the difference between PFEs at a point and area. Fig. 6 compares the 2-yr return period (corresponding to 50 % AEP) PFEs and the mean AMSs. In RFA, the mean AMS is employed as a scale factor representing the PFEs corresponding to a 2-yr return period (i.e., 50 % AEP). Hence, the mean AMS should significantly correlate with the PFE of a 2-yr return period. Overall, they have a strong spatial correlation, but some areas differ in magnitudes depending on the locations. The spatial distribution of the Atlas is relatively homogeneous and regionalized by an interpolation process, while the mean AMS has more spatial variation. Most areas with a difference between them are located around a high observation density and near the coast (the dotted red circles in the figure). Considering significant differences associated with the coastal area, the effect of interpolation on the Atlas is substantial to the coastal area.

The scatter plot shows a relationship between the mean AMS and PFEs from Atlas and ASOS-RFA. Both PFEs correlate highly with the mean AMS as the C.C.s are over 0.95. Atlas has an 11 % bias against the mean AMS, regarded as the acceptable point-to-area reduction by regionalization. The difference between the two datasets is systematic and consistent. It is interpreted as the difference resulting from a point-to-area transferring process by interpolation. The higher difference (around 20 mm/h of the mean AMS) was confirmed near the coastal area, where the AMS is higher than in the other regions. The difference is over 11 %, suggesting that Atlas in the coastal area provides more underestimated PFEs than it should. Additionally, RFA has no bias and a minor variation against the mean AMS, suggesting that the RFA method in this study is well developed and can create PFEs very similar to the Atlas PFEs before the interpolation process. In the evaluation section, this bias analysis is implemented over the CONUS and considered in the comparison with ASOS-RFA.

3.2. *Evaluation of the Atlas from a benchmark perspective*

A benchmark of PFEs, ASOS-RFA, was developed using the RFA method and ASOS data to evaluate the Atlases. A goodness-fit-test is implemented first to select the optimal probability distribution. Fig. 7 compares the fractions of the best-fit probability distributions for six durations. GLO was the highest fraction, and GNO, GEV, and GPA increased in order. Considering the goodness-fit-test result and the trend of the increasing AMS, GLO is selected as the best-fit probability distribution for creating the benchmark PFEs. Thus, the ASOS-RFA was developed across the CONUS, based on that.

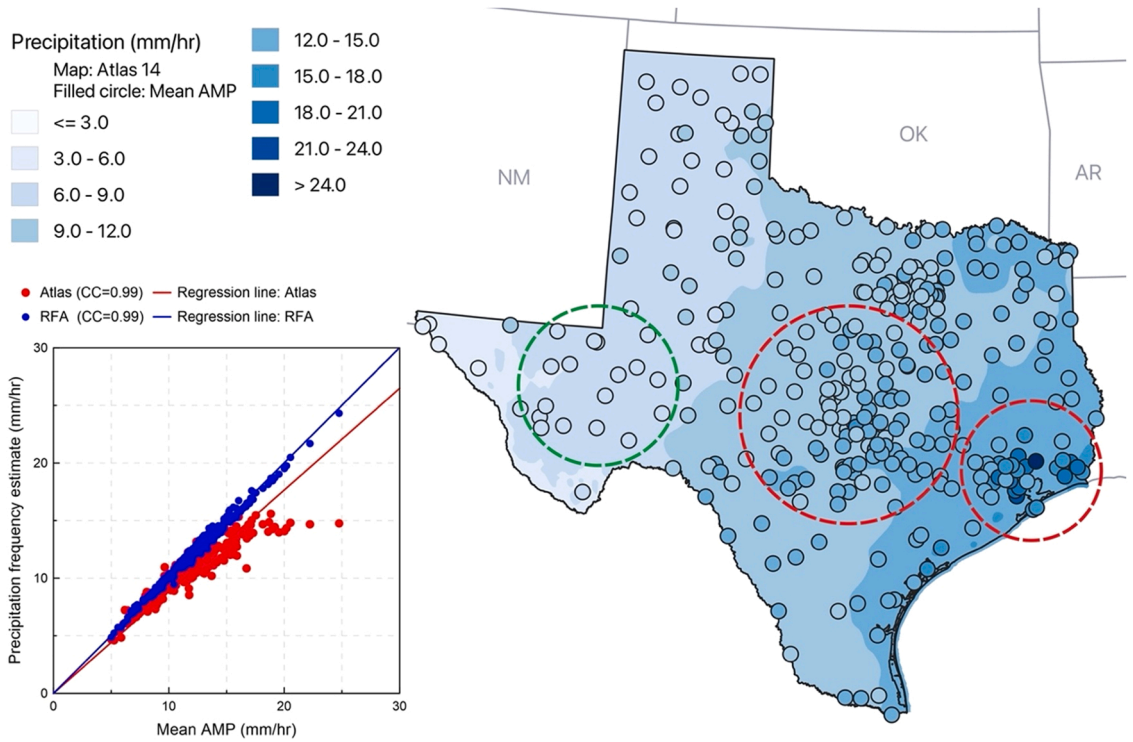


Fig. 6. Comparison of the mean values of the AMP and PFEs from Atlas 14 for a scenario of a 2-yr return period and 6-hr duration. In the scatter plot, the regression lines of Atlas (Y=0.894X) and RFA (Y=1.00X) are presented. In the map, the green circle indicates where the spatially interpolated PFEs have a strong correlation with the point PFEs, and the red circles are where the interpolated PFEs are under-predicted compared to the point PFEs.

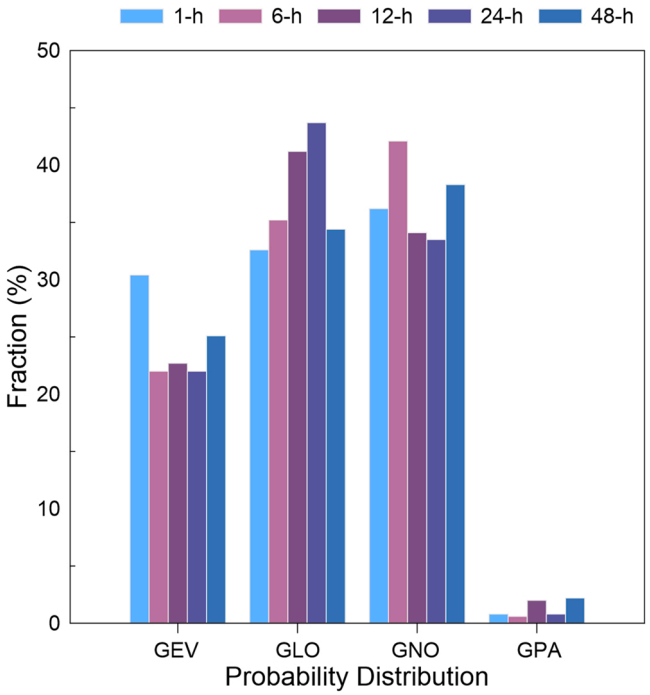


Fig. 7. 'Fraction (%)' of the best-fit probability distributions depending on durations: The fraction is calculated from all ASOS gauges in the CONUS.

Fig. 8 compares the two PFEs, the Atlas and ASOS-RFA, for various scenarios using the quantitative difference. For the 1-hr duration scenarios, the negative difference, indicating that the Atlas is smaller than ASOS-RFA, was spatially dominant over the CONUS. As the return period increases, the extent of negative differences increases, especially in California, Houston in Texas, Ohio River Basin and surrounding states, and most states in the northeastern region. Among the areas, the significant negative differences (less than 40 %) in the Ohio River Basin and surrounding states are consistently observed for higher scenarios corresponding to over a 100-yr return period, suggesting that the regions have a notable change in extreme precipitation occurrences.

The results for a 24-h duration were quite different from the 1-hr duration cases. In the 5-year return period, a fraction of the moderate differences in $+/-$ 10 % was dominant, and significant differences were not found. In 100- and 500-yr return periods, significant negative differences were limitedly observed in some specific regions in West Virginia, Tennessee, and the east coastline from North Carolina to New York, compared to the 1-hr duration cases. In the histograms, variation increases, as does the return period, suggesting that moderate differences turn into significant differences, which is the same pattern as the 1-hr duration results.

In Fig. 9, changes in the difference depending on the duration and return periods are compared. A pattern of the change by a return period was consistent in all duration results from (a) to (d). The change in the return periods under 10 years was relatively constant and increased in the negative direction as the return period increased. In all regions, the same pattern was observed even if the level of the difference was slightly different. The increasing trend was significant in Ohio. The results in the northwestern region associated with Atlas 2 were similar to the Ohio results, which is understandable when considering the age of Atlas 2 as developed in the 1970s. The change by duration was consistent in all return periods and regions. A change in differences from 1- to 12-hr durations was dramatic, suggesting a high possibility that the Atlases have relatively high uncertainties in PFEs for the durations. For the longer durations, the differences converged to a specific level of difference depending on the regions, as shown in Fig. 9.

A relationship of the differences depending on the return periods is compared to examine a correlation of the PFEs in different return periods in Fig. 10. As a result, the PFEs in different return periods are strongly correlated. The relationship between 5- and 100-yr return periods has a CC of 0.80, and the direction of differences (e.g., negative and positive) are matched by 82 %. As expected, the relationship between 100- and 500-yr return periods is more strongly correlated (CC = 0.96) since their annual exceedance probabilities are close. The same results are also observed in the other cases with extreme return periods and longer durations. Collecting all together, the findings suggest that the different levels and directions are highly consistent across the scenarios in a specific gauge location.

We have observed a significant difference in the PFEs for a 1-h duration from Fig. 9. Fig. 11 compares the mean values of the 1-h annual precipitation series to analyze the difference. Since the observation networks used in Atlas and ASOS are different, gauges located within 2 km of each other are selected to reduce the sampling error resulting from gauges not collocated. The mean value is used in the regional frequency analysis as the index flood (i.e., scale factor) that determines a scale of the PFEs based on the selected probability distribution. The mean value of Atlas is calculated from a long-term period of data depending on the region (The mean period of record in Atlas is over 60 years, while ASOS is 16 years on average).

Overall, the mean value of the Atlas was 22% less than that of ASOS across the CONUS. Notably, the extent of the difference was

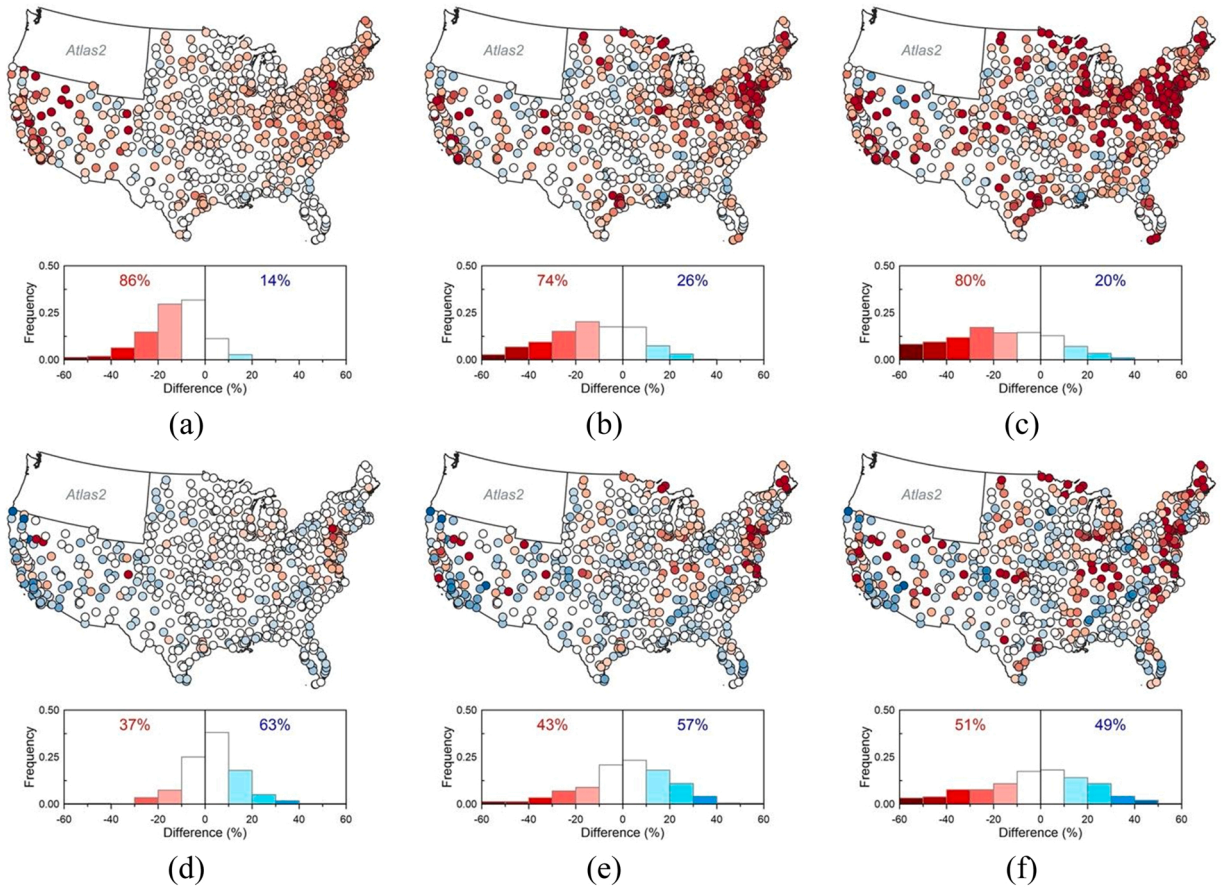


Fig. 8. Comparison of the PFEs from the Atlas 14 and ASOS-RFA: 'Difference (%)' between the two estimates is calculated from (Atlas–ASOS-RFA/Atlas x 100%). Negative values of the difference (shown in red) indicates that the PFEs from Atlas 14 are smaller than that from the ASOS-RFA curves. Positive values in blue are the opposite case of that. (a)-(c) are the results for the scenarios of 1-hr duration and 5, 100, and 500-yr return periods from left to right. (d)-(e) are the same results, but for the 24-hr duration. Each scenario result consists of a map and histogram, and in the histogram, the two numerical values indicate the total fraction of the negative and positive cases, respectively. The maps use the same color codes as the histogram. Data for the Atlas 2 area (i.e., the northwestern region) is not presented due to data availability for the scenarios.

distinct by region. In the map, the significant differences (greater than +/−50 %) between the mean intensities from the NOAA Atlas and ASOS are confirmed in California, Semiarid, and Ohio regions. The locations with the moderate difference (within +/−10 %) are in Midwestern and Southeastern states. In the region with the significant difference, the mean value of ASOS was 66 % higher than that of Atlas. These results are interpreted as the difference between Atlas and ASOS arises from the different time windows. The time window represents how long a period of records is used and under what climate conditions RFA is implemented.

A period of records is related to determining the extent of the mean values of the AMS. In developing Atlas, on average, 40–50 years of records were used, and some gauges provided over 100 years of records were used to calculate the mean value. However, as shown in Fig. 2 for Texas, for 68 years (1950–2017), the AMS increased at least 20%. Thus, using a long period of records enough to include the status of different climate change underestimates the mean value under the current acceleration of climate change, and the fact directly links to underestimating the PFEs.

Notably, the fraction of the latest records (i.e., the recent 20 years) used in developing Atlas is associated with the extent of the differences. Fig. 12 shows a relationship between the Atlas development year and the difference between the AMSs from Atlas and ASOS. In the case of a 1-h duration, the difference explicitly increased as the development year is outdated, suggesting that the Atlas age correlates with and contributes to the difference. In the case of a 24-h duration, the same trend was observed as the 1-hr duration, but it was not significant.

3.3. Evaluation of the Stage-IV-based precipitation frequency estimates

Stage-IV is evaluated using the BCF. Fig. 13 shows a map of the median value of the event-based BCFs estimated using 1-h quantitative precipitation estimates from Stage-IV and 1-hr precipitation intensity from ASOS. Overall, Stage-IV tended to underestimate quantitative precipitation estimates, and the level of bias was within 20 % (see the probability distribution in the figure).

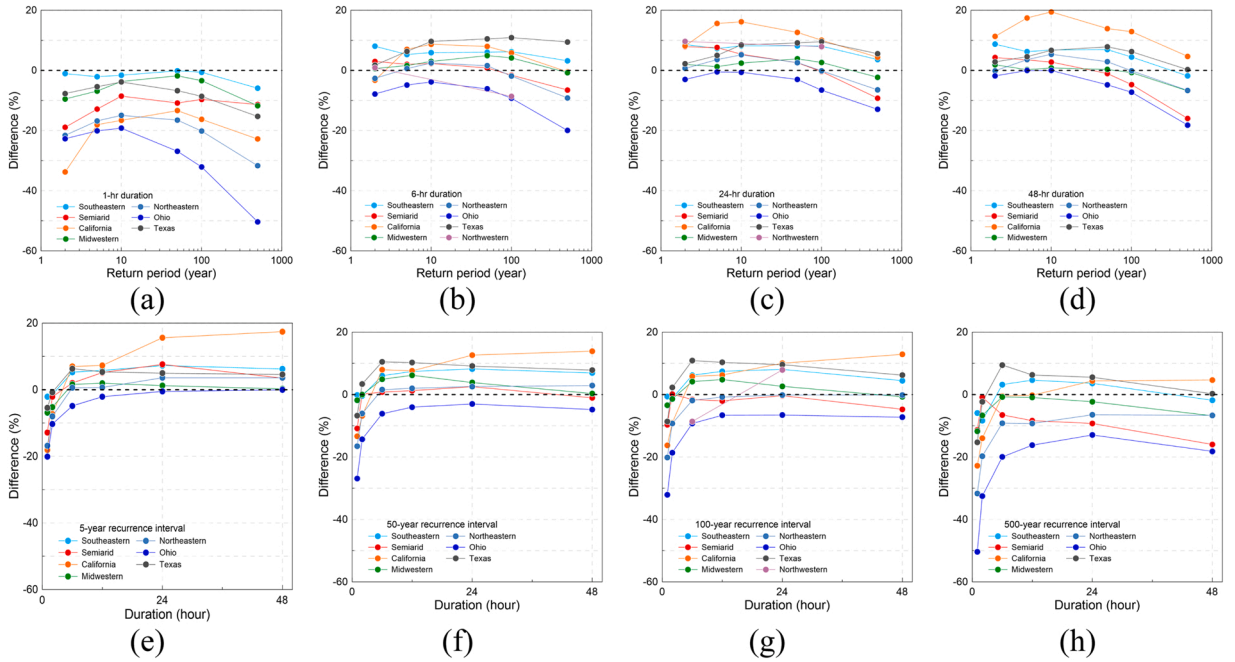


Fig. 9. A change of difference (%) between Atlas 14 and ASOS-RFA depending on durations and return periods. (a)-(d) represent the results for 1- to 48-h durations in the eight regions (Atlas 14: Southeastern, Semiarid, California, Midwestern, Northeastern, and Ohio, Texas and Atlas 2: Northwestern). Results for the northwestern region are presented for 6- and 24-h durations and 5- and 100-yr return periods as Atlas 2 (for the northwestern states) is only available for the scenarios. (e)-(f) are the same as (a)-(d), but depending on durations from 5- to 500-yr.

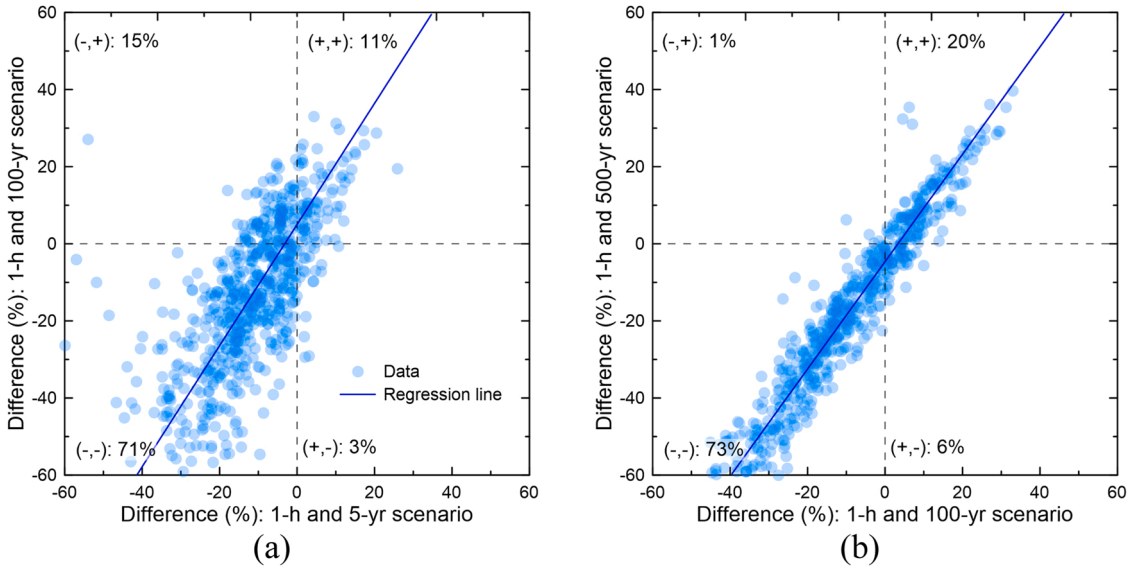


Fig. 10. A relationship between the PFEs of low (1- and 5-yr)- and high (100- and 500-yr) return periods: Difference (%) between Atlas and ASOS-RFA is used as the measurement. Each plot is divided into four areas representing the difference's directions (+ positive or - negative). For example, (-,+) in (a) is the negative difference between a 5-yr return period and a 1-hr duration scenario and the positive difference between a 100-yr return period and a 1-hr duration scenario. The numerical value indicates a fraction (%) of each direction.

California and the Northwestern regions with complicated geomorphological characteristics presented higher bias than the other regions. Also, numerous missing data from Stage-IV are identified in the states of California and Washington. Thus, some sampling points in both states are excluded from developing a new IDF curve based on Stage-IV.

Correcting biases in Stage-IV is implemented as a pre-processing of the IDF curve development. Fig. 14 compares 1-hr AMSs from ASOS, uncorrected, and corrected Stage-IV. Among all regions, the Northwestern region, where the quality was relatively poor, and the

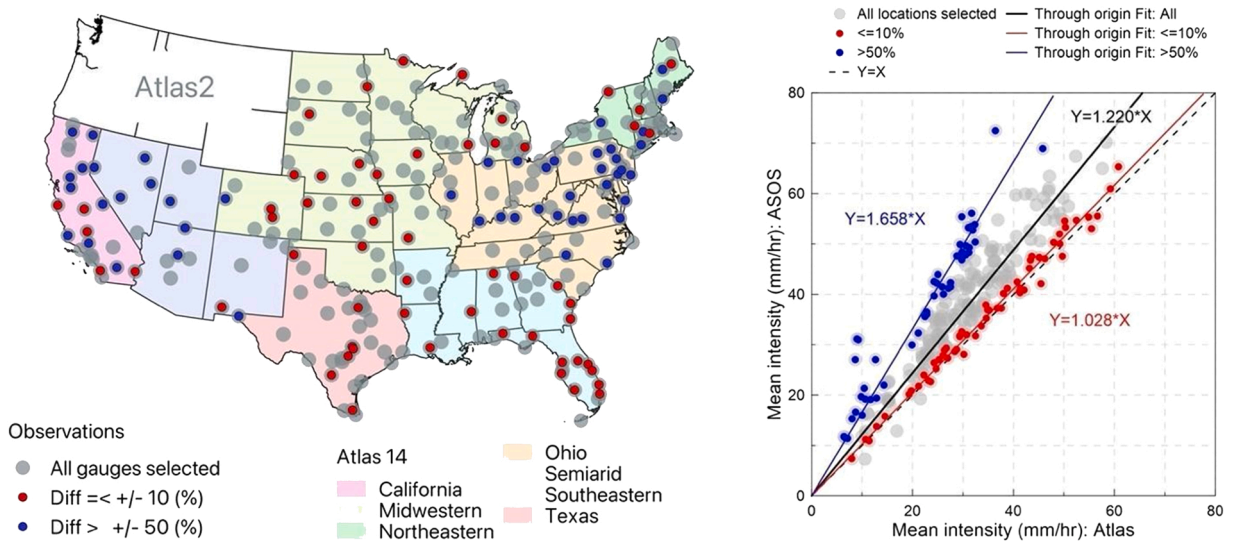


Fig. 11. Analysis results of the mean value of 1-hr AMS from the Atlas and ASOS to examine the significant difference in the PFEs from Fig. 9: On the left side, the map presents the locations where Atlas and ASOS gauges (total 249 stations) are located 2 km from each other. 'Diff' indicates a difference (%) between 1-hr mean AMS from the Atlas and ASOS. Points in red and blue represent the low difference case (less than +/- 10%) and the high difference case (over +/- 50 %), respectively. On the right side, the scatter plot presents a relationship between the mean values from the Atlas and ASOS, and it follows the same color code as the map on the left side.

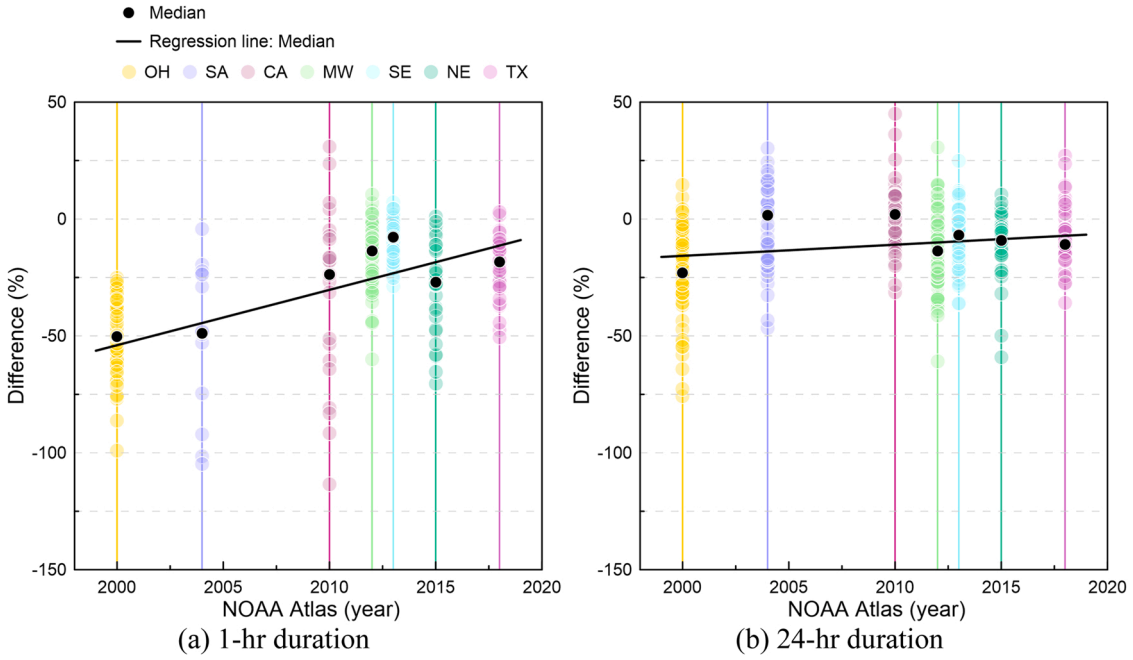


Fig. 12. A relationship between the NOAA Atlas development year and difference (%) between the mean values of the AMS from Atlas and ASOS: (a) and (b) represent the result for 1- and 24-hr duration, respectively.

Ohio region, where the quality was fairly good, are selected to show the comparison. In the case of the Northwestern region, the effect of bias correction on improving AMS was significant. Especially, the uncorrected Stage-IV's AMS, mostly less than ASOS' AMS, was improved, and the improvement is confirmed from the probability distribution comparison, and 1:1 scatter plot. In Ohio, bias correction's effect was insignificant as Stage-IV's AMS was relatively moderate, compared to ASOS's AMS. This result was confirmed not only in Ohio but also in all regions except California and the Northwestern states.

IDF curves based on the corrected Stage-IV were developed, and PFEs compared with ASOS-RFA in Fig. 15. Regardless of the duration, both PFEs are highly correlated with each other in a high-frequency scenario (i.e., 5-yr return period), and the bias was small

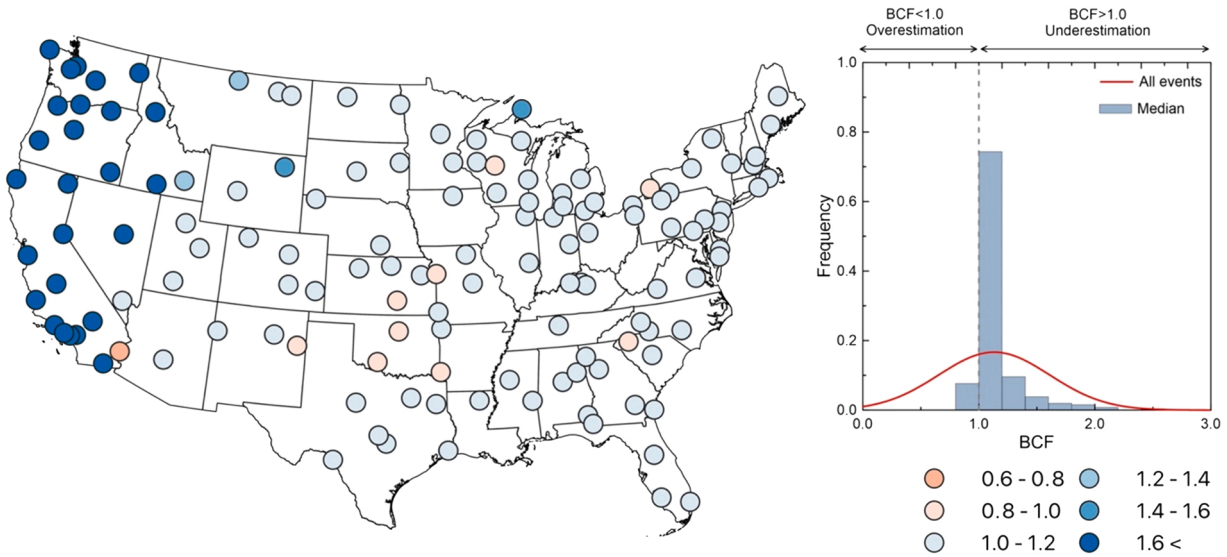


Fig. 13. Distributions of the BCF in CONUS: The map presents the spatial distribution and probability distribution of BCF. BCF > 1.0, $= 1.0$, and < 1.0 indicate underestimation, no bias, and overestimation of Stage-IV against ASOS. The histogram shows the probability distribution of BCF in CONUS.

enough to ignore. This is because the AMS's mean value from Stage-IV is close to that of ASOS using a bias correction process. From all results, as the duration increased, it is clear that the two estimates tended to agree well, and the variation decreased.

As the return period increases, the variation of the scatter gradually increases due to significant differences identified in some locations, but a high correlation was maintained even in the case of a 500-year return period. Bias increases as the return period increases, but we confirmed that it is derived from the significant differences identified in a portion of locations instead of from systematic differences in all locations. However, the variation tended to decrease at longer durations, and this is the typical accumulation effect that an error decreases as temporal resolution increases, in quantitative precipitation estimation. It appears that Stage-IV could provide relatively better PFEs in longer durations.

Based on the BCF and Stage-IV-based PFE results, the meaningful results are found in TX, MW, SE, and OH, suggesting that the Stage-IV is eligible to develop the IDF curves in the regions. In CA and NW, where the Stage-IV performance was not as good as the other regions, the conventional bias correction method is limited to improving Stage-IV to develop the IDF curves comparable with the ASOS-RFA. In SA, the correlation between the Stage-IV-based PFEs and ASOS-RFA is high enough to conclude that they are strongly correlated, but the Stage-IV tended to underestimate the PFEs more than the ASOS-RFA.

4. Discussion

There are a few points that should be discussed as they might have effects on the results. The first is the sampling errors from using two different ground observation systems, the Atlas gauge and ASOS. The primary features occurring in the sampling errors are the gauge location and density depending on the regions. Precipitation measurement varies with gauge locations; even in better situations, adjacent gauges have surprisingly limited correlational ability (Kidd et al., 2017). For example, the correlation coefficient was less than 0.5 between adjacent gauges 4 km apart (Habib et al., 2001). Also, the gauge density is related to the data population for calculating statistics such as the mean value. From this perspective, there is a distinct possibility that both features slightly affect estimating an index-flood in RFA. Since the numbers of gauges in both observation systems are large enough to calculate reasonable and comparable statistics, we assume that the effect of the sampling error on the results was not enough to change the conclusions.

The recent 20 years of data were considered in developing the new IDF curves in this study. The length of the record is shorter than what typical studies used in frequency analysis, generally a 30-year of record. Thus, there are several expected effects arising from the use of the shorter length of a record. The first one is that an index-flood could be estimated higher than the case from longer records. This is because an index-flood decreases as the fraction of older records increases, due to nonstationarity of precipitation.

The second one is associated with building the probability distribution. The record length indicates the maximum that the AEP data could present and influences the fitting of a probability distribution to data. For example, ASOS could represent 5 % as the maximum AEP (based on 20 years of a record), and Atlas gives 2.5 % (based on 40 years of a record on average). However, by employing the RFA in this study, we believe the effect on the developed IDF curves is minimized. Using randomly selected two stations in Texas, the PFEs from 20 and 30 years of a record are compared in Fig. 16. The AMP data provided by the NOAA Atlas 14 are used. As expected, the difference between the PFEs from different record lengths increases as the return period increases, suggesting that a PFE for an extreme catastrophe event (e.g., over a 100-yr return period) could have a significant uncertainty depending on data length. However, overall differences were not remarkable at the stations. Also, many studies have used around 20 years of a record (some cases considered

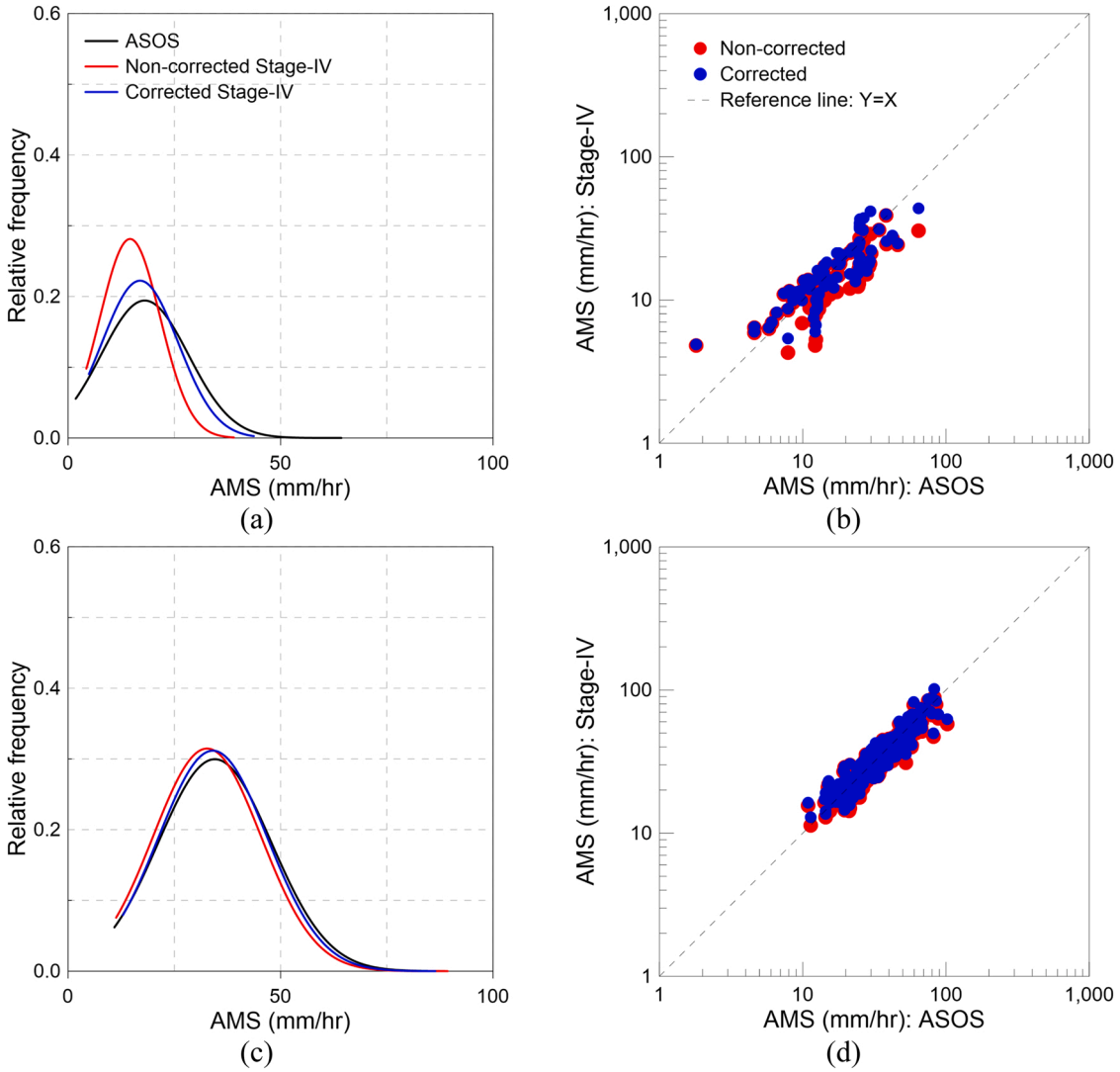


Fig. 14. Comparison of 1-hr AMS from uncorrected and corrected Stage-IV data: (a) and (c) represent a comparison of the PDFs of AMS with ASOS as reference data. (b) and (d) describe a scatter plot of AMS of corrected and uncorrected Stage-IV data against ASOS. (a) and (b) are for 'Northwestern States', where relatively higher bias correction factors were identified. (c) and (d) is for 'Ohio River Basin and Surrounding States' where lower bias correction factors are identified.

shorter periods) to establish reasonable IDF curves, and the fact suggests that the use of 20 years of record is reasonable and usable for implementing frequency analysis from various perspectives (Ombadi et al., 2018; McGraw et al., 2019; Ghebreyesus and Sharif, 2021).

The effort to develop the post-Atlas is urgently needed. Using consistent data features, with a reasonable period of data well representing the current climate circumstances, and selecting an appropriate probability distribution that could reflect the acceleration of climate change and associated changes in an increasing trend of the AMS should be high priorities in the development. Lastly, considering a growing trend in the AMS for an interval of less than 12 hrs, an increase of interest in flash flooding and pluvial floods is projected.

5. Conclusions

The NOAA Atlases and the Stage-IV were evaluated and validated based on the recent 20 years of the gauge record. Nonstationarity of precipitation resulting in an increase in the mean AMS and a change in precipitation frequencies was the primary reason that the older Atlases diverged from recent observations. The nonstationarity influenced the choice of an appropriate probability distribution. The GEV used in developing the Atlases might not be the best option for the last two decades of precipitation record, especially for deriving the PFEs at higher return periods. Instead, considering GLO, which has similar characteristics to the PFEs under 100-year return periods and provides slightly higher values for return periods greater than 100-year, could be a way to reflect the

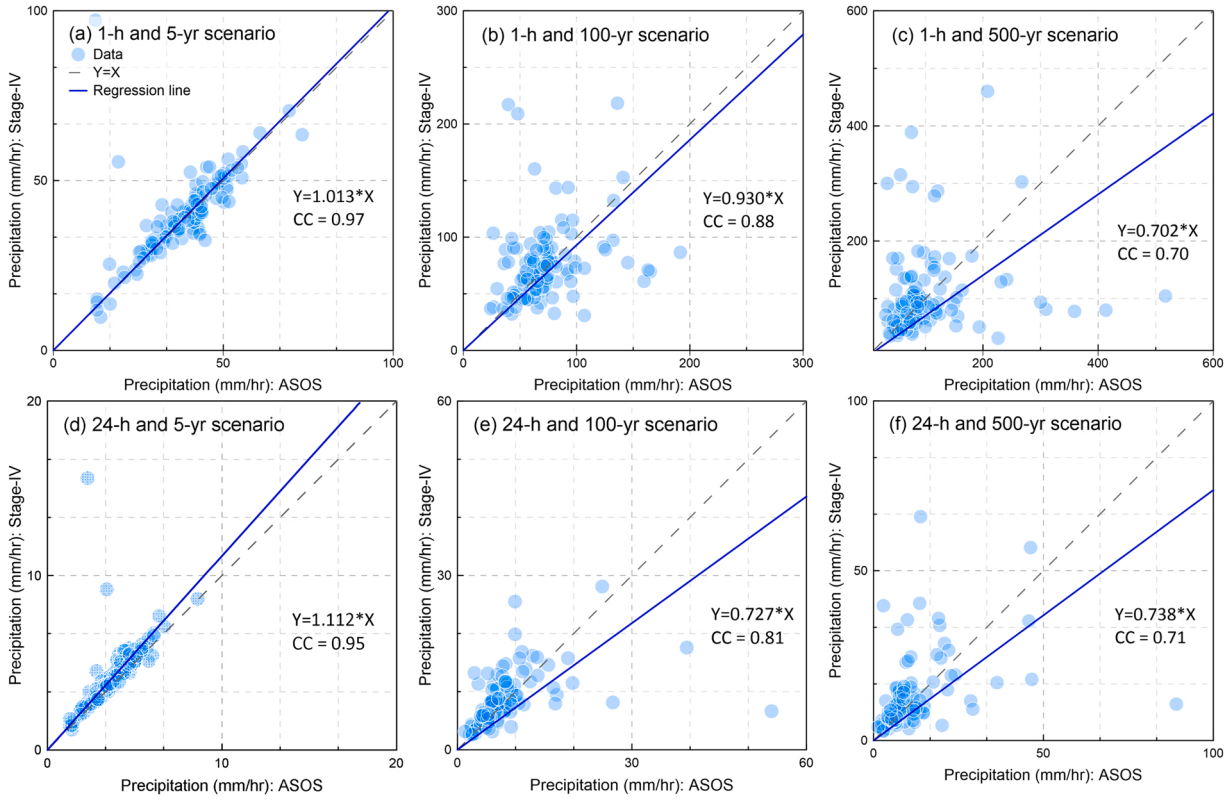


Fig. 15. Comparison of the PFEs from ASOS and Stage-IV. The regression equation and CC are presented.

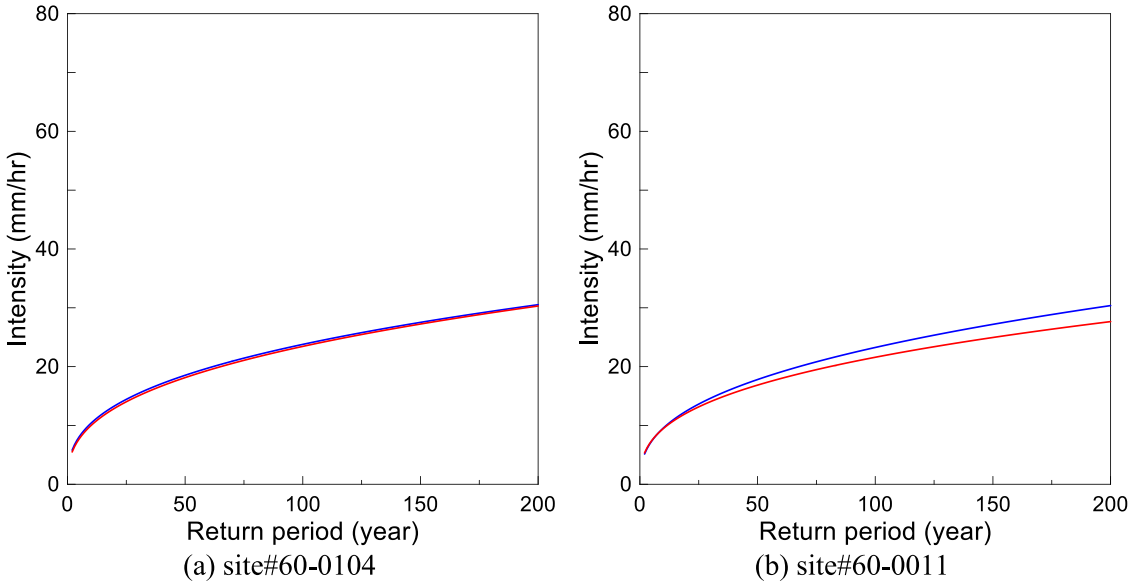


Fig. 16. Comparison of intensity-frequency curves for a 24-hr duration from 20 (blue) and 30 (red) years of data.

nonstationarity in an alternative way. In addition, using a recent period of data that well represents the current climate circumstance could reduce uncertainties arising from the effect of nonstationarity, especially for calculating the index-flood in RFA.

Overall, Atlases underestimated the PFEs against the benchmark over the CONUS, under the current climate condition. Notably, the trend was significant in the Ohio River Basin and the surrounding states and coastal areas in Texas and California. The difference between the Atlases and benchmark increased at higher return periods and decreased at longer durations. Also, the difference was

highly correlated with the age of Atlas, suggesting that the outdated data is the primary source that generates uncertainties in developing the PFEs.

Stage-IV can potentially be a data source for developing the post-Atlas period. That was shown by the PFEs at high-frequency return periods by bias correction, resulting in the well-matched mean AMS with the benchmark. However, due to a difference between the AMS and the exceedance probabilities, the uncertainty in the PFEs for return periods greater than 100 years increased more than did that of the lower return period's uncertainty, as expected. Also, numerous missing data from Stage-IV were identified in the states of California and Washington and excluded from developing the IDF curves. This fact is not only the issue found in the radar QPE data but also in most remote sensing data, suggesting that measuring the eligibility of remote sensing data for developing IDF curves should be implemented by preliminarily detecting missing data.

Answering the first question raised in the introduction, the NOAA Atlas is diverging from the eligibility to represent standard PFEs under the current climate conditions and the acceleration of climate change over the CONUS. The uncertainty of Altas will keep increasing due to the nonstationarity of the AMS, enhanced by climate change and older records used in its development.

For the second question, the Stage-IV has a high potential to be one of the data sources for developing a seamless map of the PFEs, but several issues should be resolved, including the nonhomogeneous quality of the Stage-IV before accounting it for frequency analysis. Objectively, the bias correction method employed in this study is to remove systematic errors against reference data (i.g., gauge) instead of designed to improve the accuracy of the AMS. It was an essential process, but for deriving the PFEs comparable to the benchmark, an additional process that could improve the AMS considering annual exceedance probabilities is recommended. In addition, the Multi/Radar and Multi/Sensor (MRMS), well-known as the advanced version of a radar quantitative precipitation estimate product, is a reasonable option for the post-Atlas.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgments

The First Street Foundation supported this work.

Appendix A. List of terminologies

AIC Akaike Information Criterion.
AEP Annual Exceedance Probability.
AMP Annual Maximum Precipitation.
AMS Annual Maximum Series.
ASOS Automated Surface Observing System.
ASOS-RFA Automated Surface Observing System and Regional Frequency Analysis.
BIC Bayesian Information Criterion.
BCF Bias Correction Factor.
CONUS Conterminous United States.
CC Correlation Coefficient.
DDF Depth-Duration-Frequency.
IDF Intensity-Duration-Frequency.
IQR Inter-Quartile Range.
IETD Inter-Storm Event Duration Time.
MRMS Multi/Radar and Multi/Sensor.
NCEP National Centers for Environmental Prediction.
NOAA National Oceanic and Atmospheric Administration.
NWS National Weather Service.
Stage-IV NEXRAD Stage-IV.
NEXRAD Next Generation Weather Radar.
PFE Precipitation Frequency Estimate.
PWM Probability-Weighted Moments.
QPE Quantitative Precipitation Estimation.
RFA Regional Frequency Analysis.
RFC River Forecast Center.

CRediT authorship contribution statement

Jungho Kim, Mike Amodeo, Jeremy Porter, Ed Kearns: Conceptualization, Methodology, Software. **Jungho Kim, Evelyn Shu**: Data curation, Writing- Original draft preparation. **Jungho Kim**: Visualization, Investigation. **Mike Amodeo, Jeremy Porter, Ed Kearns**: Supervision. **Jungho Kim, Kelvin Lai**: Data acquisition. **Jungho Kim, Evelyn Shu, Mike Amodeo, Jeremy Porter, Ed Kearns:** Writing- Reviewing and Editing,

References

Akan, O.A., 1993. Urban Stormwater Hydrology: A Guide to Engineering Calculations. CRC Press.

Al Kajbaf, A., Bensi, M., 2021. Assessment of uncertainty in regional and at-site precipitation frequency analysis for the localized region of Ellicott City, Maryland. Natural Hazards 108 (3), 2513–2541.

Cheng, L., AghaKouchak, A., 2014. Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. Sci. Rep. 4, 7093. https://doi.org/10.1038/srep07093.

Chin, D.A., 2004. An overview of urban stormwater management practices in Miami-dade County, Florida: U.S. Geological Survey Open-File Report 2004–1346, 1–17.

DeGaetano, A., Castellano, C., 2018. Selecting time series length to moderate the impact of nonstationarity in extreme rainfall analyses. J. Appl. Meteorol. Climatol. 57 (10), 2285–2296.

DeGaetano, A.T., Castellano, C.M., 2017. Future projections of extreme precipitation intensity-duration-frequency curves for climate adaptation planning in New York State. Clim. Serv. 5, 23–35. https://doi.org/10.1016/j.cliser.2017.03.003.

Donat, M.G., Lowry, A.L., Alexander, L.V., O'Gorman, P.A., Maher, N., 2016. More extreme precipitation in the world's dry and wet regions. Nat. Clim. Change 6 (5), 508–513.

Dupont, B., Allen, D.L., 2000. Revision of the Rainfall-intensity Duration Curves for the commonwealth of Kentucky (No. KTC-00–18). University of Kentucky Transportation Center.

Elsebaie, I.H., 2012. Developing rainfall intensity–duration–frequency relationship for two regions in Saudi Arabia. J. King Saud. Univ. -Eng. Sci. 24 (2), 131–140.

Fulton, R.A., Breidenbach, J.P., Seo, D.J., Miller, D.A., O'Bannon, T., 1998. The WSR-88D rainfall algorithm. Weather Forecast. 13 (2), 377–395.

Gao, P., Carbone, G.J., Lu, J., Guo, D., 2018. An area-based approach for estimating extreme precipitation probability. Geogr. Anal. 50 (3), 314–333.

Ghebreyesus, D.T., Sharif, H.O., 2021. Development and assessment of high-resolution radar-based precipitation intensity-duration-curve (IDF) curves for the State of Texas. Remote Sens. 13 (15), 2890.

Guo, Y., 2006. Updating rainfall IDF relationships to maintain urban drainage design standards. J. Hydrol. Eng. 11 (5), 506–509.

Habib, E., Krajewski, W.F., Ciach, G.J., 2001. Estimation of rainfall interstation correlation. J. Hydrometeorol. 2 (6), 621–629.

Hosking, J.R.M., Wallis, J.R., 2002. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge Univ. Press,.

IPCC, 2014. Climate change 2014: synthesis report. contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 1–151.

Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. Adv. Water Resour. 25, 1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8.

Keifer, C.J., Chu, H.H., 1957. Synthetic storm pattern for drainage design. J. Hydraul. Div. 83 (4), 1332-1.

Kidd, C., Becker, A., Huffman, G.J., Muller, C.L., Joe, P., Skofronick-Jackson, G., Kirschbaum, D.B., 2017. So, how much of the Earth's surface is covered by rain gauges? Bull. Am. Meteorol. Soc. 98 (1), 69–78.

Kim, J., Han, H., 2021. Evaluation of the CMORPH high-resolution precipitation product for hydrological applications over South Korea. Atmos. Res. 258.

Kim, J., Yoo, C., 2014. Use of a dual Kalman filter for real-time correction of mean field bias of radar rain rate. J. Hydrol. 519, 2785–2796.

Kim, J., Yoo, C., Lim, S., Choi, J., 2015. Usefulness of relay-information-transfer for Radar QPE. J. Hydrol. Vol. 531, 308–319.

Kim, J., Han, H., Kim, B., Chen, H., Lee, J.H., 2020. Use of a high-resolution-satellite-based precipitation product in mapping continental-scale rainfall erosivity: a case study of the United States. Catena 193, 104602.

Kunkel, E.K., Easterling, D.R., Redmond, K., Hubbard, K., 2003. Temporal variations of extreme precipitation events in the United States: 1895-2000. Geophys. Res. Lett. 30 (17).

Leathers, D.J., Brasher, S.E., Brinson, K.R., Hughes, C., Weiskopf, S., 2020. A comparison of extreme precipitation event frequency and magnitude using a high-resolution rain gage network and NOAA Atlas 14 across Delaware. Int. J. Climatol. 40 (8), 3748–3756.

Lopez-Cantu, T., Samaras, C., 2018. Temporal and spatial evaluation of stormwater engineering standards reveals risks and priorities across the United States. Environ. Res. Lett. 13 (7), 074006.

Mailhot, A., Duchesne, S., 2010. Design criteria of urban drainage infrastructures under climate change. J. Water Resour. Plan. Manag. 136 (2), 201–208.

Mallakpour, I., Villarini, G., 2015. The changing nature of flooding across the central United States. Nat. Clim. Change 5 (3), 250–254.

Mamo, T.M., 2015. Evaluation of the potential impact of rainfall intensity variation due to climate change on existing drainage infrastructure. J. Hydrol. Eng. 141 (10), 05015002.

Marra, F., Nikolopoulos, E.I., Anagnostou, E.N., Bárdossy, A., Morin, E., 2019. Precipitation frequency analysis from remotely sensed datasets: a focused review. J. Hydrol. 574, 699–705.

Martel, J.L., Brissette, F.P., Lucas-Picher, P., Troin, M., Arsenault, R., 2021. Climate change and rainfall intensity–duration–frequency curves: overview of science and guidelines for adaptation. J. Hydrol. Eng. 26 (10), 03121001.

McGraw, D., Nikolopoulos, E.I., Marra, F., Anagnostou, E.N., 2019. Precipitation frequency analyses based on radar estimates: an evaluation over the contiguous United States. J. Hydrol. 573, 299–310.

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 Forecast. 31 (2), 371–394.

Ning, L., Riddle, E.E., Bradley, R.S., 2015. Projected changes in climate extremes over the northeastern United States. J. Clim. 28 (8), 3289–3310.

NOAA. U.S. Climate Extremes Index. 2016. Available online: www.ncdc.noaa.gov/extremes/cei.

Ombadi, M., Nguyen, P., Sorooshian, S., Hsu, K.L., 2018. Developing intensity-duration-frequency (IDF) curves from satellite-based precipitation: methodology and evaluation. Water Resour. Res. 54 (10), 7752–7766.

Omranian, E., Sharif, H.O., Tavakoly, A.A., 2018. How well can global precipitation measurement (GPM) capture hurricanes? Case study: hurricane Harvey. Remote Sens. 10 (7), 1150.

Prat, O.P., Nelson, B.R., 2015. Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain gauge data sets at daily to annual scales (2002–2012). Hydrol. Earth Syst. Sci. 19 (4), 2037–2056.

Qiao, L., Hong, Y., Chen, S., Zou, C.B., Gourley, J.J., Yong, B., 2014. Performance assessment of the successive Version 6 and Version 7 TMPA products over the climate-transitional zone in the southern Great Plains, USA. J. Hydrol. 513, 446–456.

Rodríguez, R., Navarro, X., Casas, M.C., Ribalaygua, J., Russo, B., Pouget, L., Redaño, A., 2014. Influence of climate change on IDF curves for the metropolitan area of Barcelona (Spain). Int. J. Climatol. 34, 643–654. https://doi.org/10.1002/joc.3712.

Seo, D.J., Fulton, R.A., Breidenbach, J.P., Taylor, M., Miller, D.A., 1996. Interagency Memorandum of Understanding Among the NEXRAD Program. WSR-88D Operational Support Facility. the National Weather Service Office of Hydrology, Hydrologic Research Laboratory, National Weather Service, Silver Spring, Maryland.
Shrestha, A., Babel, M.S., Weesakul, S., Vojinovic, Z., 2017. Developing intensity–duration–frequency (IDF) curves under climate change uncertainty: The case of Bangkok, Thailand. Water 9 (2), 145.
Sillmann, J., Kharin, V.V., Zwiers, F.W., Zhang, X., Bronaugh, D., 2013. Climate extremes indices in the CMIP5 multimodel ensemble: part 2. Future climate projections. J. Geophys. Res.: Atmos. 118 (6), 2473–2493.
Sun, Y., Wendi, D., Kim, D.E., Liong, S.Y., 2019. Deriving intensity–duration–frequency (IDF) curves using downscaled in situ rainfall assimilated with remote sensing data. Geosci. Lett. 6 (1), 1–12.
Swain, D.L., Wing, O.E., Bates, P.D., Done, J.M., Johnson, K.A., Cameron, D.R., 2020. Increased flood exposure due to climate change and population growth in the United States. Earth's Future 8 (11).
Vu, T.M., Mishra, A.K., 2019. Nonstationary frequency analysis on the recent extreme precipitation events in the United States. J. Hydrol. 575, 999–1010.
Wallis, J.R., Schaefer, M.G., Barker, B.L., Taylor, G.H., 2007. Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. Hydrol. Earth Syst. Sci. 11 (1), 415–442.
Wesson, S.M., Pegram, G.G.S., 2006. Improved radar rainfall estimation at ground level. Nat. Hazards Earth Syst. Sci. 6, 323–342.
Wright, D., Li, Z., Booth, E., 2020. Using stochastic storm transposition to update rainfall intensity-duration-frequency (IDF) curves for the Coon Creek and West Fork Kickapoo Watersheds.
Wright, D.B., Smith, J.A., Villarini, G., Baeck, M.L., 2013. Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition. J. Hydrol. 488, 150–165.
Wright, D.B., Bosma, C.D., Lopez-Cantu, T., 2019. U.S. hydrologic design standards insufficient due to large increases in frequency of rainfall extremes. Geophys. Res. Lett. 46, 8144–8153.
Yan, H., Sun, N., Wigmosta, M., Skaggs, R., 2019. Next-generation intensity-duration-frequency curves to reduce errors in peak flood design. J. Hydrol. Eng. 24 (7).
Yoo, C., Park, C., Yoon, J., Kim, J., 2014. Interpretation of mean-feild bias correction of radar rain rate using the concept of linear regression. Hydrol. Process. 28, 5081–5092.
Yoo, C., Park, C., Jun, C., 2016. Evaluation of the concept of critical rainfall duration by bivariate frequency analysis of annual maximum independent rainfall event series in Seoul, Korea. J. Hydrol. Eng. 21 (1).