

Uncertainties in Future U.S. Extreme Precipitation from Downscaled Climate Projections

Tania Lopez-Cantu¹, Andreas F. Prein², Constantine Samaras¹

¹Civil and Environmental Engineering, Carnegie Mellon University

²National Center for Atmospheric Research

Key Points:

- Extreme daily rainfall changes for the United States from five widely used public downscaled climate projection datasets are compared.
- Large differences in magnitude and spatial patterns of extreme precipitation changes exist between datasets.
- High-end extremes (i.e., 100-year event) increase 10-50 % faster than low-end extremes (i.e., 2-year event) at the continental scale.

Corresponding author: Tania Lopez-Cantu, tlopez@andrew.cmu.edu

Abstract

Impacts modelers and stakeholders use publicly available datasets of downscaled climate projections to assess and design infrastructure for changes in future rainfall extremes. If differences across datasets exist, infrastructure resilience decisions could change depending on which single dataset is used. We assess changes in U.S. rainfall extremes from 2044-2099 compared with 1951-2005 based on 227 projections under RCP4.5 and RCP8.5 from five widely used datasets. We show there are large differences in the change magnitude and its spatial structure between datasets. At the continental-scale, the datasets show between 10 % and 50 % larger increases in high-impact events than in more frequent extremes. These differences largely contribute to the overall uncertainty for short average recurrence intervals (ARI) extremes (2- to 10-year), while uncertainties due to short record length dominate long ARIs (25- to 100-year). The results indicate that robust infrastructure planning should consider these uncertainties to enable resilient infrastructure under climate change.

Plain Language Summary

Observed extreme rainfall magnitudes have increased since 1950, and climate model projections indicate that these increases will continue throughout the 21st century in many areas. Adapting and designing infrastructure for climate change requires future extreme rainfall projections at high resolutions. Global Climate Model (GCM) output resolution is generally much coarser, and methods have been developed to create relevant climate information at regional-scales. Today, multiple open datasets exist that provide downscaled projections of future rainfall extremes. We analyze changes in future daily extremes using five widely used datasets in climate change impacts assessments and decision-making. We found large differences between datasets in how much and where extreme events will intensify by the end of the 21st century. This and other sources of uncertainty need to be considered when designing resilient infrastructure.

1 Introduction

Extreme rainfall events have intensified in magnitude and become more frequent in many regions of the United States (U.S.) since 1950 (DeGaetano, 2009; Hoerling et al., 2016) and climate model simulations project these increases will continue throughout the 21st century (Prein et al., 2017). Observed and anticipated changes in rainfall extremes affect many sectors, including existing and proposed infrastructure such as stormwater systems. In the U.S., stormwater infrastructure is designed using federal, state and local design standards based on historical extreme rainfall probabilities, and most new infrastructure is similarly not designed to consider future climate change (Lopez-Cantu & Samaras, 2018; Wright et al., 2019). Making infrastructure systems resilient to future conditions involves identifying risks and vulnerabilities to inform climate resilient strategies (IPCC, 2014). However, it is not straightforward to develop these strategies because it requires making decisions under a deeply uncertain future climate (Hallegatte, 2009). Yet, taking no action is also a decision, and large economic damages can occur in the case of no adaptation (Martinich & Crimmins, 2019).

Many stakeholders incorporate future climate conditions into their planning and analyses by using downscaled global climate model (GCM) output to inform hydrologic models or update current engineering design standards (Forsee & Ahmad, 2011; Kuo et al., 2015). Although the downscaling process is necessary to match the spatial and temporal resolution to the required resolution by the specific application (Cook et al., 2017), different downscaling methods and resolutions can potentially give different results (Cook et al., 2020; Onof & Arnbjerg-Nielsen, 2009; Sunyer et al., 2012; Arnbjerg-Nielsen et al., 2015; Wu et al., 2019). For the Contiguous United States (CONUS), there are several downscaled climate projection datasets covering at least the late 20th century and the 21st century (Mearns et

Table 1. Some of the publicly available downscaled climate products for the United States and an overview of their characteristics.

Dataset	GCM models	Emissions Scenario	Downscaled Spatial Resolution (degrees)	Highest Temporal Resolution	Downscaling Technique
BCCAv2	21	2.6, 4.5, 6.0, 8.5	1/8	Daily	Statistical
MACA	20	4.5, 8.5	1/24	Daily	Statistical
LOCA	32	4.5, 8.5	1/16	Daily	Statistical
NA-CORDEX	6 ^a	4.5, 8.5	0.22, 0.44	Hourly	Dynamical

^a The number of GCM models dynamically downscaled depend on the target resolution and the emissions scenario.

al., 2013; Pierce et al., 2014; Abatzoglou & Brown, 2012; Maraun et al., 2010) (see Table 1 and Supplementary Information Fig. 2 and Tables 1-3). These datasets vary in downscaling method used, number of downscaled GCMs, emission scenarios, and horizontal resolution. The total number of available projections from the datasets in Table 1 is 103 for Representative Concentration Pathway (RCP) 4.5 and 124 for RCP 8.5. The massive amount of data, limited guidance or lack of compelling arguments to discard or adopt one or more datasets in Table 1, in addition to different ease of use among these datasets, are often reasons why many impact assessments only use a single dataset as input—neglecting possible differences across the datasets and their influence on the specific study area.

In this study, we assess the daily extreme rainfall climate change signal across the U.S. from five downscaled climate projection datasets (Table 1) for RCP 4.5 and 8.5. We use five downscaled climate projection datasets (hereafter, *datasets*) consisting of multiple simulations (hereafter, *members*), while the collection of available projections from all five datasets for each RCP scenario is referred to as *ensemble*. We focus on these datasets because they all downscaled GCMs from the Coupled Model Intercomparison Project Phase 5 (CMIP5), share a common temporal resolution and spatial coverage, comprise two RCP scenarios, are publicly available, and are frequently used as inputs in climate change impacts assessments for infrastructure (Cook et al., 2017; Alamdari et al., 2017; Kermanshah et al., 2017; Gelda et al., 2019).

We use the Generalized Extreme Value (GEV) distribution to describe the extreme rainfall data and quantify the climate change signal in terms of change factors corresponding to the ratio between the model future volume (estimated from annual rainfall series for the 2044-2099 period) and model historical volumes (1951-2006) of average recurrence interval (ARI) events ranging from 2- to 100-year. Additionally, because these datasets are used in a decision-making context that requires information about the relevant uncertainties (Mullan et al., 2018), we quantify the relative contribution of three sources of uncertainty; (1) the GEV distribution fit uncertainty related to the record length to fit the distribution and its accuracy to describe the extremes, (2) the dataset selection uncertainty from the different downscaling methods applied to a set of GCMs and (3) the uncertainty stemming from the emissions scenario (RCP 4.5 and RCP 8.5).

2 Materials and Methods

2.1 Publicly Available Downscaled U.S. Climate Projections Datasets

The datasets compared in this study can be categorized by the type of downscaling process used. The Bias Corrected Constructed Analogues (BCCA) version 2, Multivariate Adaptive Constructed Analogues (MACA) and the Localized Constructed Analogues (LOCA) datasets are products of statistically downscaling a subset of CMIP5 GCMs (Pierce

et al., 2014; Abatzoglou & Brown, 2012; Maraun et al., 2010). The constructed analog (CA) method constitutes the base of each statistical downscaling method. CA involves comparing the GCM control simulations to historical observations, traditionally in terms of anomalies, but all procedures above use absolute values instead. The historical observations are re-gridded to match the resolution of the GCM simulations. GCM simulations are typically bias-corrected using a quantile mapping (QM) approach against other source of historical observations (Pierce et al., 2014; Abatzoglou & Brown, 2012; Maraun et al., 2010). For a given GCM day, the goal is to find a set of days (situated within a maximum of 45 days centered at the GCM target day) in the re-gridded observations that when combined, they approximate the climate conditions on the day simulated by the GCM (Hidalgo et al., 2008).

The main difference between BCCAv2, MACA and LOCA lies in the searchable domain to find the analogue days. For BCCAv2 and MACA, the analogue days are selected from any grid within the domain (CONUS), while in LOCA, the analogues are identified at smaller climatically similar regions. Another difference lies in the number of variables that are jointly downscaled. In BCCAv2 and LOCA, variables are independently downscaled, while in MACA a multi-variate approach is used for variables other than precipitation (Abatzoglou & Brown, 2012). Also, while the BCCAv2 and MACA methods use a linear combination of 30 days to construct the CA at each point, a single day out of the 30 identified that is closest to GCM target day pattern across the neighboring region to each point is used in the LOCA procedure (Bracken, 2016; Pierce et al., 2014). Finally, the MACA method applies a final bias-correction step (once again QM) to the downscaled output, in contrast to the BCCA and LOCA methods.

The final dataset we assess is the North American branch of the Coordinated Regional Downscaling Experiment (NA-CORDEX), which is a dynamically downscaled dataset. There are projections available in two different resolutions (see Table 1), and since they include different members (see Supplementary Tables S2–S3), we treated each as an individual dataset. Dynamical downscaling techniques use physics-based models or regional climate models (RCM) constrained to a smaller region and are driven by GCM projections at the boundaries to produce higher resolution projections. The dataset corresponds to a collection of high-resolution projections from different GCM-RCM combinations and emissions scenarios (Mearns et al., 2014). Fig. S1 shows a detailed comparison of the methodological steps of the process followed to produce each dataset.

2.2 Estimation of Future Changes in Daily Rainfall Extremes

To compare the extreme rainfall climate change signal between datasets, we first re-gridded all ensemble members to the NA-CORDEX 0.44° grid (i.e. the coarsest grid spacing across datasets used in this study) using a conservative remapping approach (Jones, 1999) to minimize differences stemming from the datasets’ output spatial resolution.

For each ensemble member, extreme rainfall was characterized using a stationary GEV distribution fitted to series of annual maxima (Coles, 2001) for two periods, one future period between 2044 and 2099, and a historical period from 1950 to 2006. The GEV distribution is defined as,

$$F(x) = \begin{cases} e^{-\left(1 + \frac{\xi}{\sigma}(x-\mu)\right)^{-\frac{1}{\xi}}} & \xi \neq 0 \\ e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}} & \xi = 0 \end{cases}, \quad (1)$$

where μ , σ , ξ are the parameters of the GEV distribution defining its location, scale and shape, respectively. The best estimate and its 95 % confidence interval (CI) of the parameters (computed through a bootstrapping function implemented in the R package *extRemes* (Gilleland, 2019)) were estimated using the generalized maximum likelihood (GML) method, which extends the maximum likelihood (ML) method by adding a constraint that

restricts the shape parameter to take values within an physically coherent range (Adlouni et al., 2007; Martins & Stedinger, 2000). We used the same prior distribution for the shape parameter as defined in (Martins & Stedinger, 2000).

For each grid cell per dataset member, the $1/q$ -year ARI change factor (or climate signal) was defined as the ratio between the future precipitation volume and the historical volume for each ARI quantile, q , and was estimated as,

$$X_{q,fut} = \frac{F^{-1}(q; \mu_{fut}, \sigma_{fut}, \xi_{fut})}{F^{-1}(q; \mu_{ctrl}, \sigma_{ctrl}, \xi_{ctrl})}, \quad (2)$$

where μ_{fut} , σ_{fut} , ξ_{fut} are the GEV parameters of the distribution fitted to future period (2044-2099), μ_{ctrl} , σ_{ctrl} , ξ_{ctrl} are the GEV parameters of the distribution fitted to the historical (1950-2006) model period, and X_q the change factor computed for q equal to 0.5 (2-year), 0.2 (5-year), 0.1 (10-year), 0.04 (25-year), 0.02 (50-year), and 0.01 (100-year) ARI events.

For the BCCAv2 dataset that included multiple downscaled realizations for the same GCM, each individual run change factors were averaged. Also, because the number of members in each RCP scenario, members that were not common in both RCP scenarios where discarded (see Supplementary Tables S1-S3). Across datasets, the set of members is different, which makes a comparison of the impact of downscaling method on future extreme precipitation challenging. Although here we focus on the complete set of available members per dataset to stay consistent with how these datasets are used in the literature, we repeated the analysis by using the CMIP5 GCM model (CanESM2) that is downscaled in all five datasets to assess the effect of using different GCMs in each dataset.

Finally, the dataset mean climate change signal for the 2-, 5-, 10-, 25-, 50- and 100-year ARI was defined as the average of equally weighted member change factors in a dataset. We specifically chose these ARIs to produce actionable results, given that the selected ARIs are used as standards for designing water infrastructure in the U.S. (Lopez-Cantu & Samaras, 2018). To test the climate change signal's significance for each dataset and ARI, we first constructed a series of 100 artificial rainfall volumes for both the future and the historical period, where each volume was estimated through a GEV distribution whose parameters where randomly selected (with replacement) from the GEV parameters of any dataset member estimated in earlier steps. We then compare the future and historical artificial rainfall volumes, and detect statistically significant (at the 0.01 level) differences between both periods using the Mann-Whitney rank test.

2.3 Analysis of Uncertainty Sources and their Contribution to the Overall Uncertainty in the Projected Changes in Rainfall Extremes

To understand how important potential differences between datasets are in the overall uncertainty context, we identified other two main sources that contribute to uncertainty in the projected change of rainfall extremes. We define the uncertainty due to dataset selection (dataset variance), $D(r)$, uncertainty due to the GEV distribution fitting (distribution fit variance) and due to the emissions scenario (scenario variance), and contextualize these uncertainties in a similar way to (Hawkins & Sutton, 2009).

The dataset variance component by ARI was estimated as,

$$D(a) = \frac{1}{N_s} \sum_s \text{var}_d \left(\frac{1}{N_d} \sum_{m=1}^{N_d} x_{asdm} \right), \quad (3)$$

where $D(a)$ is the dataset variance component in the a -year event change factors. x_{asdm} is the a -year event change factor in RCP scenario s projected by member m in dataset d ,

and N_d is the total number members in dataset d , and N_s is the number of scenarios (in our case two, RCP 4.5 and 8.5). $D(a)$ measures the a -year change factor variability across datasets.

The distribution fit variance was estimated using the change factors computed by all combinations of dataset members, scenarios and GEV parameters estimates (mean and 95% CI). The distribution fit variance component by ARI was estimated as,

$$F(a) = \frac{1}{N_s} \sum_s \text{var}_d \left(\frac{\sum_{m=1}^{N_d} (x_{asdm}^U - x_{asdm}^L)}{N_d} \right), \quad (4)$$

where $F(a)$ is the distribution fit variance in the projected a -year event change factors. x_{asdm}^U is the a -year event change factor in scenario s computed using the GEV parameters of the 95% CI upper endpoint, and x_{asdm}^L , the a -year event change factor in scenario s projected by member m in dataset d computed using the GEV parameters of the 95% CI lower endpoint. The difference $x_{asdm}^U - x_{asdm}^L$ corresponds to the change factor spread resulting from the uncertainty in the GEV parameters and depends on the confidence level chosen. N_d is the total number of members in dataset d , and N_s is the total number of scenarios. $F(a)$ measures how much the a -year change factor varies due to the GEV fitting accuracy on average per member, dataset and scenario.

Finally, the scenario variance component by ARI was estimated as,

$$S(a) = \text{var}_s \left(\frac{1}{N_e} \sum_d \frac{1}{N_d} \sum_{m=1}^{N_d} x_{asdm} \right), \quad (5)$$

where $S(a)$ is the scenario variance in the projected a -year event change factors. x_{asdm} is the a -year event change factor in RCP scenario s projected by member m in dataset d , N_d is the total number members in dataset d , and N_e is the total number of datasets in the ensemble (i.e. five different datasets).

Assuming that the three sources of uncertainty are independent, the total variance in the projected change factors can be written as,

$$T(a) = D(a) + F(a) + S(a), \quad (6)$$

The contribution of each source of uncertainty can be found by dividing each by the total variance.

3 Results

3.1 Future Daily Rainfall Extremes Comparison Between Datasets

The datasets' multi-member means show increased extreme precipitation over nearly all CONUS. CONUS-averaged change factors generally increase with ARIs (see Fig.1a), indicating that the most untactful extreme events increase the most. Between datasets, the spread of the CONUS distribution of change factors is different, and, for each dataset, the spread increases with ARIs. For example, 5-year change factors from simulations included in the LOCA dataset ranges from 0.98 to 1.29 under RCP8.5, while the 50-year change factors range from 0.95 to 1.53. Despite exhibiting the least average bias compared to observed extremes (see Fig. S2–S4), the MACA multi-member mean change factors are considerably larger than those of the other datasets. Notably, BCCAv2 shows the lowest multi-member mean change factors which are mostly constant across ARIs (Fig. 1a and

Supplementary Figs 5b–10b). We found similar patterns when analyzing the change factors of the downscaled output of the common member across datasets (GCM: CanESM2) (Supplementary Fig. S11). A visual comparison between the datasets change factors and the native-resolution CanESM2 change factors (Supplementary Fig. S11) shows that the datasets preserve the pattern of change signal to some extent, but the magnitude of the signal is lower in BCCAv2, while it is higher in MACA. Supplementary Fig. S12–S13 show a quantitative comparison between the regionally averaged downscaled climate signals and the GCM climate signals for both RCP 4.5 and RCP 8.5 scenarios. For most of the U.S., the averaged climate signal is close to the GCM signal for short ARIs, while it gradually diverges with increasing ARI. For all regions and large ARIs, the MACA downscaled CanESM2 change signal is considerably higher than the original GCM signal, while the BCCAv.2 signal is mostly lower. Climate signals based on LOCA and NA-CORDEX are similar to the original GCM change signal, with largest differences (i.e. the West region) possibly due to better orographic representation in the dynamically downscaled NA-CORDEX datasets.

Figure 1a also shows the datasets agree that daily rainfall extremes intensify more under RCP 8.5 than under RCP 4.5 on continental scales, consistent with previous studies comparing both scenarios (Fix et al., 2018). However, the difference between RCP scenarios varies across datasets (see Fig. 1b). There exist little differences between the RCP 4.5 and 8.5 scenarios in the BCCAv2 dataset, while there is a larger difference in the low-resolution NA-CORDEX dataset.

In terms of the spatial structure of these changes, the datasets roughly agree on higher percent increases west of the continental divide and the Ohio River Basin and Northeast regions (see Fig. S12 for a graphical representation of the regions' location) for short return periods (2- to 10-year). In contrast, for larger return periods (25- to 100-year), there is no clear spatial agreement due to an increase in noisiness, yet the signal is statistically significant at the 0.01 level in all datasets for at least more than 40 % of CONUS. For the 25-year ARI, the change signal is statistically significant in more than 50 % of the grid cells in CONUS for RCP 4.5 and more than 73 % for RCP 8.5 (see Table S5 for statistical significance of other ARIs).

In general, there are noticeable differences in change factors of the dynamically downscaled NA-CORDEX datasets compared to the statistically downscaled methods, particularly over orographic regions (See SI Section S3).

Figures 1b–c shows the magnitude and the spatial structure of the 25-year ARI multi-member mean climate change signal which considerably differs across the five datasets (See Supplementary Figs 5–10 for other ARIs). For the 25-year ARI on average for CONUS, the multi-member mean change signal shows an increase of 11 %, 15 %, 36 %, 22 % and 15 % (BCCAv2, LOCA, MACA, NA-CORDEX high and NA-CORDEX low resolution, respectively) in the 25-year rainfall volume under RCP 4.5 emissions scenario, and 12 %, 18 %, 39 %, 27 % and 23 % under RCP 8.5 by the end of the century. We found negligible percent change differences between scenarios on average for CONUS for the 25-year event in BCCAv2, and between 3 % and 7 % for other datasets.

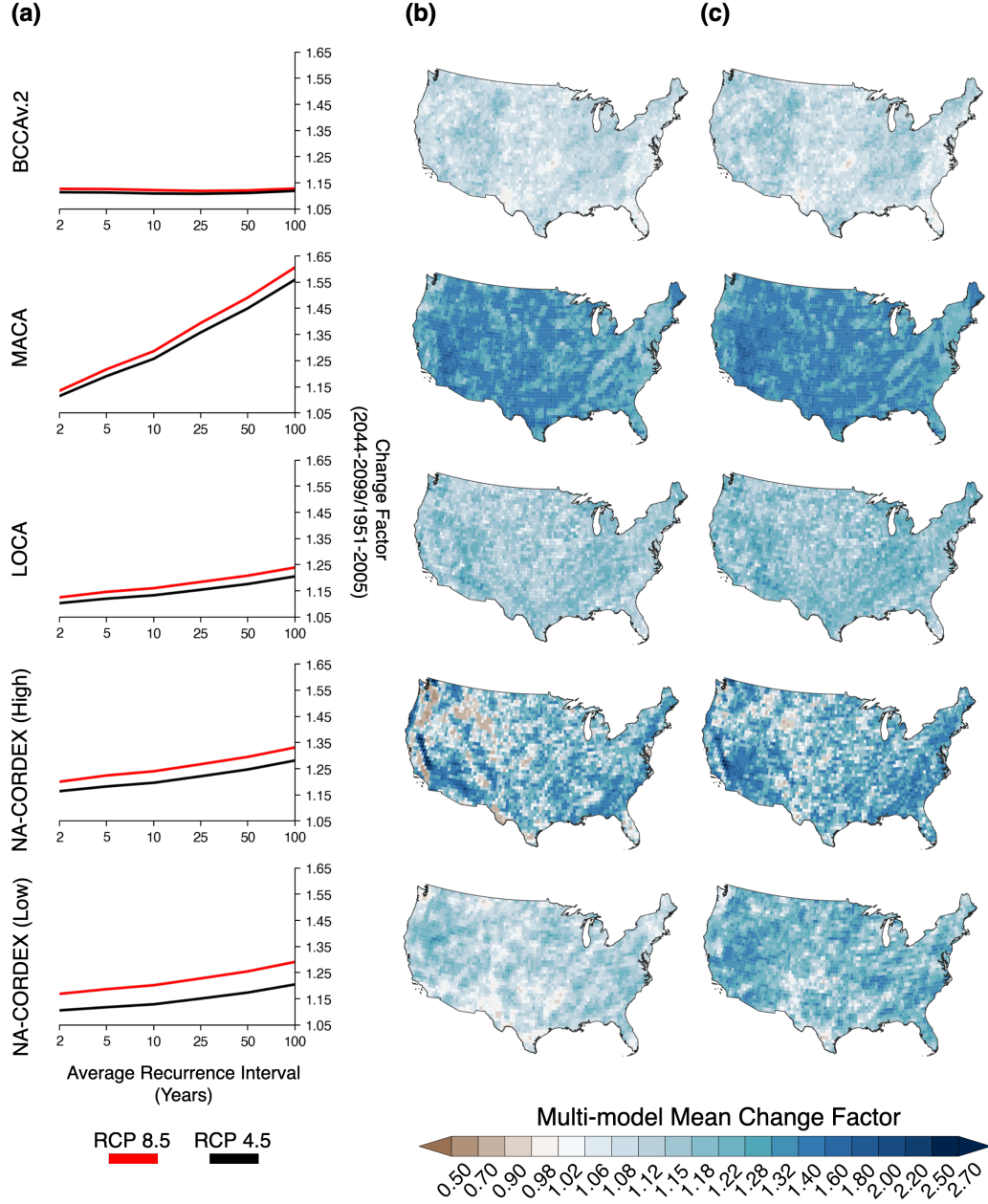


Figure 1. Future (2044-2099) extreme rainfall volumes are increasing in all datasets with larger increases for large ARI events and under RCP 8.5 compared to the past (1951-2005). Multi-member mean extreme rainfall signal averaged across the U.S. for different ARIs (a), spatial distribution of the 25-year event change factors under RCP 4.5 (b), and RCP 8.5 (c). Each dataset member was weighted equally in the multi-member mean computation.

Additional information about the distribution of change factors in each dataset (with respect to their individual members) can be obtained by examining the interquartile range (IQR), which is shown in Fig. S14–S19 for each ARI. For short ARIs, the IQR is generally small across CONUS and increases for larger ARI. The NA-CORDEX high-resolution IQR is consistently large across ARIs over the Sierra Nevada mountains. The large IQR might be affected by the smaller number of downscaled simulations in the NA-CORDEX project

at this resolution and the differences in RCM characteristics (i.e. dynamics, land surface model, etc).

3.2 Uncertainty in Future Rainfall Extremes Changes

In section 2.3, we distinguish three different sources of uncertainty in the projected change in rainfall extremes: dataset selection, RCP scenario, and distribution uncertainty, and we assess the individual contribution of uncertainty for each source relative to the other sources of uncertainty. Figure 2 shows the contribution of the three sources of uncertainty to the total variance in the projected percent change in rainfall volume for each grid cell.

High percentages indicate a larger difference in the projected change factors for a given source, compared to the difference of the other sources. There are marked differences in the dominant uncertainty with increasing ARI across the U.S. For the 2-year event, the dataset selection uncertainty is largest, reaching very large percentages ($\sim 60\%$) in several grid cells. The dataset uncertainty is especially high in coastal areas in the Northeast and the Southeast, as well as in the mountainous regions in the West (see Fig. S12 for a graphical representation of the regions' location). For the 100-year event, the dataset selection uncertainty is no longer prominent in comparison to the distribution fit uncertainty, which is now dominant (some cells reach about 70%). Larger contributions to the total variance from the scenario uncertainty occur at the 2-year than at the 100-year.

In most of the Ohio River Basin, the scenario uncertainty contribution is approximately between 40 and 50 %, but for the rest of the U.S., the percent is not higher than approximately 20 %. On average (Fig. 2b), we found that for short ARIs (< 5 -years), the dataset selection uncertainty is largest followed by the distribution fit and scenario uncertainty. The contribution of uncertainty from the distribution fit increases with ARI and becomes dominant for large ARIs (> 5 -year). Scenario uncertainty is largest ($\sim 20\%$) for short ARIs and is negligible for longer ARIs. Note that the scenario uncertainty would increase by including lower emission scenarios in the analysis (e.g., RCP2.6). However, this is currently not possible due to a lack of downscaled data products over the CONUS.

Across datasets, the difference in the change factors computed using the GEV 95% C.I. upper- and lower-end parameters was similar. For any given dataset and long ARIs, we found large variability in the projected change factors due to the extreme value distribution fit, which is consistent with previous studies that found that the record length strongly affects the estimate of the GEV parameters (Papalexiou & Koutsoyiannis, 2013).

4 Discussion and Conclusions

We showed that the magnitude and spatial pattern of the climate change signal in U.S. daily rainfall extremes greatly differs across the most commonly-used public down-scaled climate model output datasets for the second half of this century. While the set of CMIP5 GCMs downscaled in each dataset is different, we found that the differences between datasets are likely due to methodological choices within the downscaling framework rather than selected GCMs. We compared the climate change signal from the downscaled CanESM2 GCM simulations (common in all datasets), to the native resolution CanESM2 signal. In general, downscaling modifies the original GCM change signals with MACA simulations enhancing the original GCM trend for large ARIs (50-, 100-year). Conversely, the BCCAv.2 data lowers change signals. Previous studies have shown the original GCM trend can be modified by the downscaling method and/or bias correction steps due to improved surface representation and capability to solve physical processes at higher resolution in dynamical downscaling, while statistical downscaling can lead to an inflated signal artifact of the method because of correcting events that have occurred with very low frequency (Eum & Cannon, 2017). The magnitude of the difference in the projected change factors between datasets and RCP scenarios is in most cases non-negligible and might represent additional

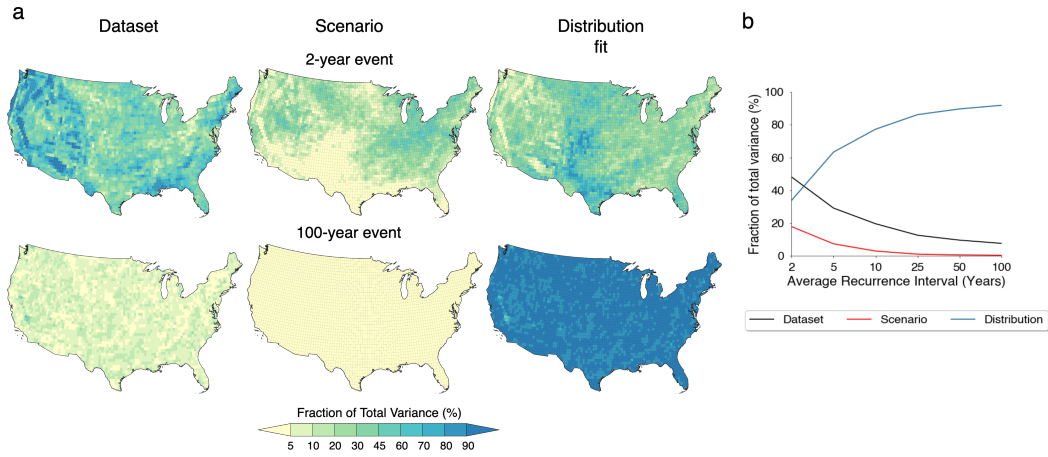


Figure 2. Dataset selection uncertainties dominate future extreme rainfall changes for short ARIs events while distribution fit uncertainties dominate long ARIs changes. Contribution of each uncertainty source to the total variance at the grid scale (a) and on average for the CONUS (b). High percentages indicate a larger difference in the projected change factors for a given source, compared to the difference per the other sources.

318 stress to infrastructure systems designed for historical conditions (Niemczynowicz, 1989;
 319 Guentchev et al., 2016).

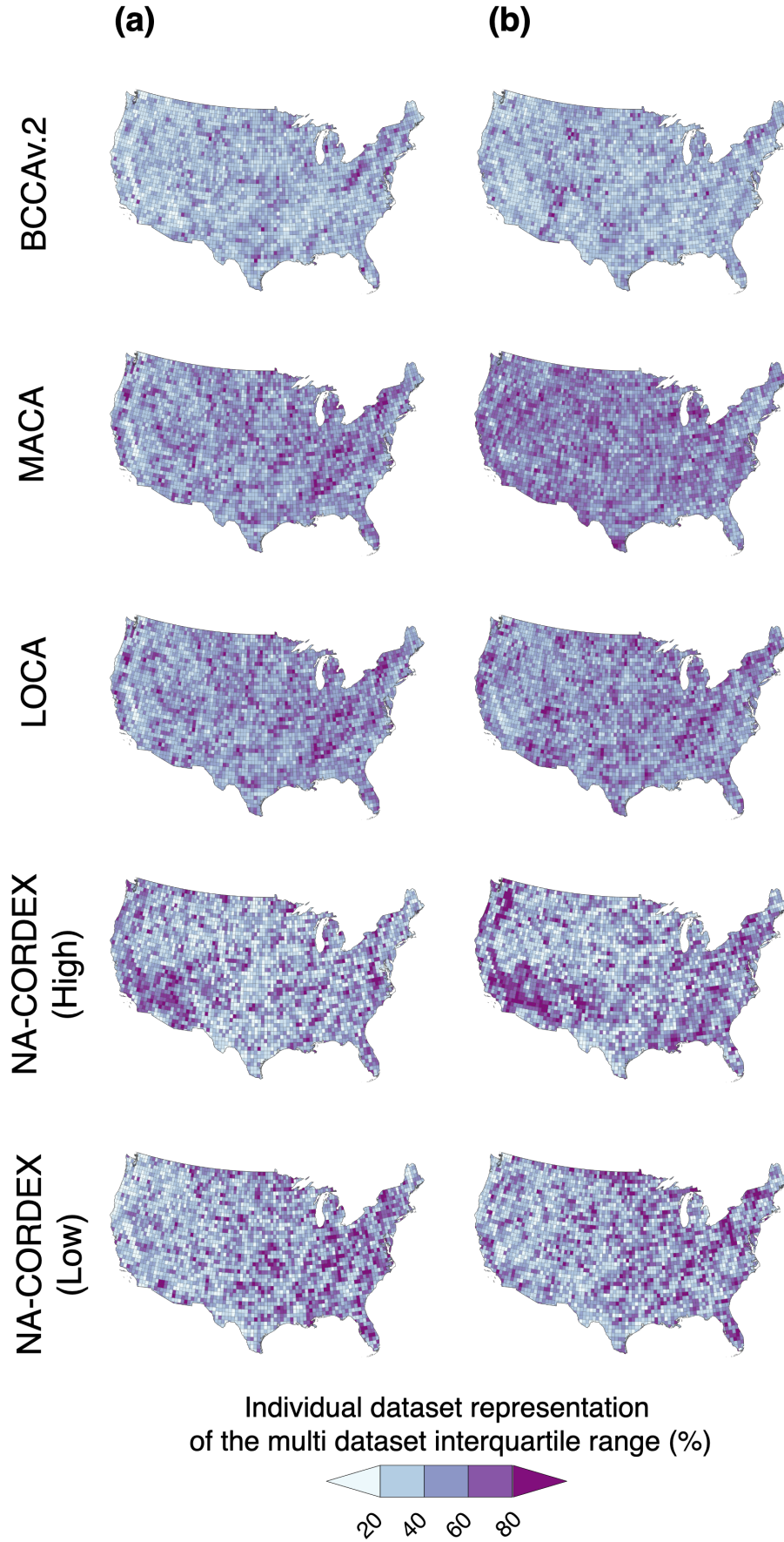


Figure 3. Percent covered by each individual dataset IQR of the ensemble IQR defined by the largest 75th and smallest 25th percentiles of all datasets for the 25-year event under (a) RCP 4.5 and (b) RCP 8.5. Darker purple means that the dataset IQR spans a large portion of the overall dataset uncertainty range defined above.

When modelers and stakeholders are evaluating critical areas and decisions in resilience assessments, it is prudent to acknowledge the differences between datasets in the projected changes in daily rainfall extremes, their multi-member distribution (i.e. the distribution of individual members' change factors), and the implications of using each dataset. Additionally, the varying magnitude of uncertainty sources with ARIs emphasizes the need to properly manage and propagate uncertainties in climate change impacts assessments and decision-making support (Arnbjerg-Nielsen et al., 2015).

Selecting a specific dataset implies constraining the range of future change factors to a range that might not be representative of the full spread of uncertainties. Figure 3 shows how much of the ensemble IQR range is represented by the members in each individual dataset IQR at the grid scale level for the 25-year ARI event (see Supplementary Figs. S20–S25 for other ARIs). A darker purple color means that the IQR from simulations within one dataset is similar to the projected IQR from the ensemble and vice versa. Some dataset representation percentages stand out in particular regions, for example NA-CORDEX high resolution downscaled output provides a better representation of the range of change factors overall (for all ARIs) at topographically complex regions. If computational resources are limited, and the decision-maker needs to select projections from one dataset, selecting the dataset with the darkest purple color for a given grid cell could be an option to at least partially represent signals from other datasets. However, this might not be the most robust decision, as Figure 3 also shows large spatial noise. Additionally, since the multi-member mean is different between datasets, the change factors of the dataset with the largest ensemble IQR representation might be considerably larger or smaller than other datasets' change factors. In the case of the former, it could be considered by the decision-maker as an added layer of safety or robustness, however it could be problematic in the latter case. Moreover, a modeler or decision-maker can interpret Fig. 1 and Fig. 3 in a way to understand the consequences of selecting a dataset based on different metrics. For instance, selecting the MACA dataset because of its low average bias, will lead to large climate change signals, and depending on the location of interest, might not be representative of other datasets since there is large spatial noise (see Fig. 3).

In summary, selecting a specific downscaled climate dataset for decision-making and neglecting others can result in potentially omitting considerable uncertainty in the range of rainfall extremes that inform resilience decisions. Given the differences we found, using more than a single dataset is preferable for robust decision-making. However, one limitation in these conclusions lies in the comparability of the downscaling method given that only a single GCM output (CanESM2) was downscaled across datasets. The differences found in dataset statistics (mean and IQR) highlight the need for conducting future downscaling efforts on a controlled set of GCM for ease of comparability.

We show that there is a need for longer records to facilitate more robust estimates of very large ARIs, and that the GEV distribution fit uncertainty should be considered by decision-makers and modelers when assessing impacts from large ARIs. Large ensemble datasets could be used to provide more robust estimates for very large ARIs (e.g. 100-year return levels) from model data to reduce the distribution fit uncertainty (Tandon et al., 2018). Another promising approach for reducing dataset uncertainties in future extreme precipitation estimates are kilometer-scale climate models that better simulate extreme precipitation than state-of-the-art coarser-resolution models (Prein et al., 2015).

In case the computational resources exist to perform impacts assessments under many scenarios based on the change factors from all datasets, we support and ease the process of generating future scenarios for analyses making available a grid-cell level dataset across CONUS for the climate change signal in each of the five datasets. For specific applications, decision-makers can use our data and results, coupled with their expertise and risk preferences, to inform resilience assessments and decisions. Additionally, it can be used to compare with forthcoming downscaled climate projections datasets and completed impacts

372 assessments or explore the impacts of extreme rainfall changes to a local system using a
373 large range of possible future changes in daily rainfall extremes.

Acknowledgments

This research was partially supported by the National Science Foundation (NSF Collaborative Award Number CMMI 1635638/1635686) and the UCAR Next Generation Fellowship. UCAR and NCAR are sponsored by the National Science Foundation. Student support was also provided by Consejo Nacional de Ciencia y Tecnología (CONACYT).

We acknowledge the World Climate Research Programmes Working Group on Regional Climate, and the Working Group on Coupled Modelling, former coordinating body of CORDEX and responsible panel for CMIP5. We are grateful to the climate modeling groups that produced and made available their downscaled climate projections listed in Table S1. We also acknowledge the U.S Department of Defense ESTCP for its support of the NA-CORDEX data archive. Computer resources were provided by the Computational and Information Systems Laboratory. 2017. Cheyenne: HPE/SGI ICE XA System (NCAR Community Computing). Boulder, CO: National Center for Atmospheric Research. doi:10.5065/D6RX99HX.

References

- Abatzoglou, J. T., & Brown, T. J. (2012). A comparison of statistical downscaling methods suited for wildfire applications. *International Journal of Climatology*, *32*(5), 772-780. doi: 10.1002/joc.2312
- Adlouni, S. E., Ouara, T. B. M. J., Zhang, X., Roy, R., & Bobée, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, *43*(3). doi: 10.1029/2005WR004545
- Alamdari, N., Sample, D. J., Steinberg, P., Ross, A. C., & Easton, Z. M. (2017, July). Assessing the Effects of Climate Change on Water Quantity and Quality in an Urban Watershed Using a Calibrated Stormwater Model. *Water*, *9*(7), 464. doi: 10.3390/w9070464
- Arnbjerg-Nielsen, K., Leonardsen, L., & Madsen, H. (2015). Evaluating adaptation options for urban flooding based on new high-end emission scenario regional climate model simulations. *Climate Research*, *64*(1), 73-84.
- Bracken, C. (2016). *Downscaled CMIP3 and CMIP5 Climate Projections - Addendum* (Tech. Rep.). Washington, D.C.: Reclamation.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values* (2001 edition ed.). London ; New York: Springer.
- Cook, L. M., Anderson, C. J., & Samaras, C. (2017, December). Framework for Incorporating Downscaled Climate Output into Existing Engineering Methods: Application to Precipitation Frequency Curves. *Journal of Infrastructure Systems*, *23*(4), 04017027. doi: 10.1061/(ASCE)IS.1943-555X.0000382
- Cook, L. M., McGinnis, S., & Samaras, C. (2020, January). The effect of modeling choices on updating intensity-duration-frequency curves and stormwater infrastructure designs for climate change. *Climatic Change*. doi: 10.1007/s10584-019-02649-6
- DeGaetano, A. T. (2009, October). Time-Dependent Changes in Extreme-Precipitation Return-Period Amounts in the Continental United States. *Journal of Applied Meteorology and Climatology*, *48*(10), 2086-2099. doi: 10.1175/2009JAMC2179.1
- Eum, H.-I., & Cannon, A. J. (2017). Intercomparison of projected changes in climate extremes for South Korea: Application of trend preserving statistical downscaling methods to the CMIP5 ensemble. *International Journal of Climatology*, *37*(8), 3381-3397. doi: 10.1002/joc.4924
- Fix, M. J., Cooley, D., Sain, S. R., & Tebaldi, C. (2018, February). A comparison of U.S. precipitation extremes under RCP8.5 and RCP4.5 with an application of pattern scaling. *Climatic Change*, *146*(3), 335-347. doi: 10.1007/s10584-016-1656-7
- Forsee, W. J., & Ahmad, S. (2011, November). Evaluating Urban Storm-Water Infrastructure Design in Response to Projected Climate Change. *Journal of Hydrologic*

- Engineering*, 16(11), 865-873. doi: 10.1061/(ASCE)HE.1943-5584.0000383
- Gelda, R. K., Mukundan, R., Owens, E. M., & Abatzoglou, J. T. (2019, February). A Practical Approach to Developing Climate Change Scenarios for Water Quality Models. *Journal of Hydrometeorology*. doi: 10.1175/JHM-D-18-0213.1
- Gilleland, E. (2019, November). *extRemes: Extreme Value Analysis*.
- Guentchev, G. S., Rood, R. B., Ammann, C. M., Barsugli, J. J., Ebi, K., Berrocal, V., ... Cinquini, L. (2016, March). Evaluating the Appropriateness of Downscaled Climate Information for Projecting Risks of Salmonella. *International Journal of Environmental Research and Public Health*, 13(3). doi: 10.3390/ijerph13030267
- Hallegatte, S. (2009, May). Strategies to adapt to an uncertain climate change. *Global Environmental Change*, 19(2), 240-247. doi: 10.1016/j.gloenvcha.2008.12.003
- Hawkins, E., & Sutton, R. (2009, August). The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society*, 90(8), 1095-1108. doi: 10.1175/2009BAMS2607.1
- Hidalgo, H. G., Dettinger, M. D., & Cayan, D. R. (2008). *Downscaling with Constructed Analogues: Daily Precipitation and Temperature Fields Over the United States* (Tech. Rep. No. CEC-500-2007-123). California Energy Commission.
- Hoerling, M., Eischeid, J., Perlwitz, J., Quan, X.-W., Wolter, K., & Cheng, L. (2016). Characterizing Recent Trends in U.S. Heavy Precipitation. *Journal of Climate*, 29(7), 2313-2332. Retrieved from `gr` doi: 10.1175/JCLI-D-15-0441.1
- IPCC. (2014). *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Summaries, Frequently Asked Questions, and Cross-Chapter Boxes. A Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (C. Field et al., Eds.). Geneva, Switzerland: World Meteorological Organization.
- Jones, P. (1999). First-and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review*, 127(9), 2204-2210.
- Kermanshah, A., Derrible, S., & Berkelhammer, M. (2017). Using climate models to estimate urban vulnerability to flash floods. *Journal of Applied Meteorology and Climatology*, 56(9), 2637-2650.
- Kuo, C.-C., Gan, T. Y., & Gizaw, M. (2015, May). Potential impact of climate change on intensity duration frequency curves of central Alberta. *Climatic Change*, 130(2), 115-129. doi: 10.1007/s10584-015-1347-9
- Lopez-Cantu, T., & Samaras, C. (2018). Temporal and spatial evaluation of stormwater engineering standards reveals risks and priorities across the United States. *Environmental Research Letters*, 13(7), 074006. doi: 10.1088/1748-9326/aac696
- Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., ... Thiele-Eich, I. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48(3). doi: 10.1029/2009RG000314
- Martinich, J., & Crimmins, A. (2019, May). Climate damages and adaptation potential across diverse sectors of the United States. *Nature Climate Change*, 9(5), 397-404. doi: 10.1038/s41558-019-0444-6
- Martins, E. S., & Stedinger, J. R. (2000, March). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3), 737-744. doi: 10.1029/1999WR900330
- Mearns, L. O., Bukovsky, M., Pryor, S. C., & Magaña, V. (2014). Downscaling of Climate Information. In G. Ohring (Ed.), *Climate Change in North America* (p. 201-250). Cham: Springer International Publishing. doi: 10.1007/978-3-319-03768-4_5
- Mearns, L. O., Sain, S., Leung, L. R., Bukovsky, M. S., McGinnis, S., Biner, S., ... Sloan, L. (2013, October). Climate change projections of the North American Regional Climate Change Assessment Program (NARCCAP). *Climatic Change*, 120(4), 965-975. doi: 10.1007/s10584-013-0831-3
- Mullan, M., Danielson, L., Lasgargues, B., Chrisna Morgado, N., & Perry, E. (2018). *Climate-resilient Infrastructure* (Tech. Rep. No. 14). OECD.

- Niemczynowicz, J. (1989, December). Impact of the greenhouse effect on sewerage systems—Lund case study. *Hydrological Sciences Journal*, *34*(6), 651-666. doi: 10.1080/02626668909491373
- Onof, C., & Arnbjerg-Nielsen, K. (2009, May). Quantification of anticipated future changes in high resolution design rainfall for urban areas. *Atmospheric Research*, *92*(3), 350-363. doi: 10.1016/j.atmosres.2009.01.014
- 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. doi: 10.1029/2012WR012557
- Pierce, D. W., Cayan, D. R., & Thrasher, B. L. (2014, September). Statistical Downscaling Using Localized Constructed Analogs (LOCA). *Journal of Hydrometeorology*, *15*(6), 2558-2585. doi: 10.1175/JHM-D-14-0082.1
- Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., ... others (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of geophysics*, *53*(2), 323-361.
- Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P., & Holland, G. J. (2017). The future intensification of hourly precipitation extremes. *Nature Climate Change*, *7*(1), 48-52. Retrieved 2018-07-13, from <https://www.nature.com/articles/nclimate3168> doi: 10.1038/nclimate3168
- Sunyer, M. A., Madsen, H., & Ang, P. H. (2012, January). A comparison of different regional climate models and statistical downscaling methods for extreme rainfall estimation under climate change. *Atmospheric Research*, *103*, 119-128. doi: 10.1016/j.atmosres.2011.06.011
- Tandon, N. F., Zhang, X., & Sobel, A. H. (2018). Understanding the Dynamics of Future Changes in Extreme Precipitation Intensity. *Geophysical Research Letters*, *45*(6), 2870-2878. (eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017GL076361>) doi: 10.1002/2017GL076361
- 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. *Geophysical Research Letters*, *46*(14), 8144-8153. doi: 10.1029/2019GL083235
- Wu, S., Markus, M., Lorenz, D., Angel, J. R., & Grady, K. (2019, June). A Comparative Analysis of the Historical Accuracy of the Point Precipitation Frequency Estimates of Four Data Sets and Their Projections for the Northeastern United States. *Water*, *11*(6), 1279. doi: 10.3390/w11061279