

ANALYSIS OF IMPACT OF NONSTATIONARY CLIMATE ON NOAA ATLAS 14 ESTIMATES

Assessment Report

National Weather Service

Office of Water Prediction

January 31, 2022



DISCLAIMER

This report and research was prepared under contract between the Department of Transportation (DOT) Federal Highway Transportation Administration (FHWA) and Office of Water Prediction (OWP) National Oceanic and Atmospheric Administration (NOAA). The results presented in this report and its publication do not indicate endorsement by NOAA or FHWA.

The data and information presented in this report are provided to demonstrate research results associated with this project. Values presented herein are not intended for any other use beyond considerations within the scope of this assessment report. Use of any data or information presented in this report for any other purpose is not endorsed by NOAA.

Table of Contents

Acknowledgments	3
1. Introduction	6
1.1. Stationary NOAA Atlas 14 methodology overview and related terminology	7
1.2. Nonstationary NOAA Atlas 14 needs and related terminology	9
2. Phase I: Local likelihood nonstationary precipitation frequency method	11
2.1. Overview of the Phase I research project	11
3. Phase II: Climate model datasets for the calculation of future precipitation frequency estimates	18
3.1. Overview of the Phase II research project	18
3.2. One-day and ten-day durations	22
3.3. One-hour duration	26
4. Complementary analyses	28
5. Conclusions and recommendations	39
References	43
Appendix A-1. The Pennsylvania State University Final Report	A-1.1–A-1.76
Appendix A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report	A-2.1–A-2.160

Acknowledgements

This project was funded by the Office of Bridges and Structures and the Office of Sustainable Transportation Resilience within the Department of Transportation Federal Highway Administration (FHWA). We extend our thanks to Robert Kafalenos, Michael Culp, Joseph Krolak, Brian Beucler, Cynthia Nurmi, and Kornel Kerenyi for key activities that made this research possible.

This project was carried out in collaboration with researchers from Pennsylvania State University, University of Illinois Urbana-Champaign and University of Wisconsin-Madison. We want to thank our collaborators: Dr. Benjamin Shaby, Dr. Gregory Bopp, Dr. Alfonso Mejia and Dr. Sanjib Sharma from the Pennsylvania State University; Dr. Momcilo Markus, Dr. Kevin Grady, Dr. James Angel, and Dr. Seid Koric from the University of Illinois at Urbana-Champaign; Dr. Shu Wu, Dr. Fuyao Wang, and Dr. David Lorenz from the University of Wisconsin-Madison. The analysis and work performed by the two research teams is included as appendices in this report.

The NOAA National Weather Service, Office of Water Prediction, Hydrometeorological Design Study Center (HDSC) members Dr. Sanja Perica, Sandra Pavlovic, Michael St.Laurent, Carl Trypaluk, Dale Unruh, Mark Glaudemans, provided research guidance, review, analysis, and coding for this work and authored this assessment report. Special thank you to Dr. Sanja Perica, HDSC Director for many years, who retired in 2020.

1. Introduction

The Hydrometeorological Design Studies Center (HDSC), within the Office of Water Prediction (OWP) of the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS), has been updating the precipitation frequency estimates for portions of the United States and affiliated territories. Updated precipitation frequency estimates, accompanied by additional relevant information, are published in NOAA Atlas 14. All NOAA Atlas 14 products and documents are available for download from the [Precipitation Frequency Data Server \(PFDS\)](#).

The current NOAA Atlas 14 frequency analysis methodology is based on the concept of temporal stationarity. The temporal stationarity assumes that the extreme precipitation events do not change significantly over time, and that future climate conditions can be represented by the past observed precipitation (Cheng and AghaKouchak, 2014). However, historical assumptions about the magnitude and frequency of extreme future events are not appropriate in the presence of nonstationary climate (USGCRP, 2018). Considering that many climate models indicate that the increasing trend in intensity and frequency of precipitation will likely continue in the future (Ragno, et al., 2018), using only statistics from the past observations could underestimate the precipitation frequency quantiles as well as their confidence limits, which could result in undersized civil engineering water resource infrastructure.

To understand the potential impact of nonstationary climate conditions on precipitation frequency estimates, in 2016 the Federal Highway Administration tasked the HDSC with examining this issue with the help of academia. This collaborative effort between the federal government and academia aimed to develop a method that would investigate and allow nonstationary climate effects to be integrated into the NOAA Atlas 14 process, make use of the climate models for modeling future conditions, and understand the limitations and applications of different techniques for developing the future nonstationary precipitation frequency estimates.

The project "Analysis of Impact of Nonstationary Climate on NOAA Atlas 14 Estimates," carried out between June 2016 and June 2020, was divided into two phases. The objective of the project's first phase was to assess the suitability of state-of-the-science methodologies for nonstationary precipitation frequency analysis, identify the most suitable method relevant to NOAA Atlas 14, and test the proposed methodology on a pilot project area. For this phase, the HDSC partnered with the Statistics and Civil Engineering departments at the Pennsylvania State University (Penn State), a group with expertise in extreme value statistical theory and modeling. The work performed and recommendations made by the Penn State collaborators are appended to this report as Appendix A-1.

The objective of the second phase of the project was to evaluate the feasibility of incorporating downscaled climate projections into frequency analysis and, if applicable, to recommend which downscaled climate projections could be used for future products. In this phase, we partnered with the University of Illinois at Urbana-Champaign and the University of Wisconsin-Madison, which have expertise in climate projection modeling and downscaling. Between 2018 and 2020, researchers from these two universities worked on evaluating four Coupled Model Intercomparison Project Phase 5 (CMIP5) downscaled datasets to assess their ability to reproduce observed precipitation frequencies from NOAA Atlas 14 Volume 10 (Northeastern

states) and incorporated and tested emission scenarios for the selected datasets using the methodology from the first project phase. The final report by this research team is appended to this report in Appendix A-2.

The Northeast project area was selected for the pilot study because several studies have shown that the Northeast is experiencing warming temperatures and a large increase in the amount of rainfall measured during heavy precipitation events (Kunkel et al., 2017; Howarth et al., 2019). Furthermore, the HDSC compiled and quality controlled a historic precipitation record during the development of the NOAA Atlas 14 Volume 10 project and provided the data to collaborators for this study. In the second phase of the project, the state of Texas project area was added to the analysis to test the downscaled climate model data in comparison to the observations on an additional project area. For more information on the observed data used, please refer to the NOAA Atlas 14 Volume 10 and Volume 11 documentation (Perica et al., 2015; Perica et al., 2018).

In this report, we summarize the work performed in the two project phases and evaluate the work in the context of engineering design applications and the precipitation frequency analysis by HDSC. The intent is that the work performed by both research teams be integrated into one methodology. In this report, we also propose several changes to the original recommendations made by the research teams and discuss further research needed on the topic of the development of future precipitation frequency estimates. The proposed changes and further research recommendations are necessary to make future precipitation estimates more robust and applicable to engineering design and to streamline the future production of estimates for regions across the United States.

1.1. Stationary NOAA Atlas 14 methodology overview and related terminology

The current NOAA Atlas 14 approach provides precipitation magnitude–frequency relationships at individual stations, computed using a regional frequency analysis approach based on L-moment statistics developed by Hosking and Wallis (1997). The L-moment statistics, used for calculating the distribution parameters, are suitable and robust when estimates are calculated using a small sample size and when high outliers are present in the small sample. Other distribution parameter estimators such as the maximum likelihood estimation (MLE) approach or the generalized maximum likelihood estimation (GMLE) approach, may be more sensitive to outliers or small sample sizes, but are more easily adaptable for modeling nonstationary climate. The MLE or GMLE techniques have been used extensively in research and practice for modeling extremes and nonstationary climate, because they easily incorporate independent variables called covariates (to account for temporal nonstationarity and incorporate the projections of future climate) into the calculation of the distribution parameters (Coles and Dixon, 1999; Coles, 2001; Salas et al., 2018).

The precipitation frequency analysis approach used in NOAA Atlas 14 is based on an analysis of annual maximum series (AMS) across a range of fixed durations. The AMS for each station are obtained by extracting the highest precipitation amount for a particular duration in each successive year. An alternative time series dataset to AMS is the partial duration series (PDS).

PDS are capable of making more efficient use of data by considering more than a single observation per year, but they add other sources of uncertainty into the precipitation frequency process, such as a selection of thresholds for extracting extreme events, selection of appropriate declustering methods needed for handling temporal dependence in this time series, and additional modeling complexity (Madsen et al., 1997).

Both AMS-based and PDS-based precipitation frequency estimates are provided in NOAA Atlas 14. The AMS-based estimates are calculated directly from the AMS data, while the PDS-based precipitation frequency estimates are calculated indirectly from a widely used empirical formula (based on Langbein, 1949). Differences in the magnitudes of the corresponding PDS-based and AMS-based estimates, based on historical observations, are negligible for average recurrence intervals (ARIs) of about 20 years and above, but notable at smaller ARIs (especially below five years). Because the PDS can include more than one event in any particular year and provide more conservative estimates, the results from a PDS analysis are considered more reliable for designs based on frequent events.

The Langbein empirical relationship or any other empirical relations based on historical observations can be used only for analysis under the assumption of stationary climate. For the NOAA Atlas 14 nonstationary framework, the AMS-based and PDS-based nonstationary methodologies would have to rely on appropriate analysis of extreme precipitation in the presence of nonstationary climate. These two approaches (AMS-based and PDS-based) are conceptually similar for both stationary and nonstationary modeling. The main differences between the AMS-based approach and the PDS-based one, aside from the data, are the different types of distributions used to approximate the data, and how the regional parameters are calculated. The generalized extreme value (GEV) distribution is used to approximate the AMS, while, most often, the generalized Pareto distribution (GPD) is used to fit the PDS.

One of the primary problems in precipitation frequency analysis is the need to provide estimates for ARIs that are significantly longer than the available records. Regional approaches, where data from stations with similar frequency distributions are grouped, have been shown to provide more accurate estimates of extreme quantiles than approaches that use only data from a single station (Burn, 1990; Hosking and Wallis, 1997). The number of stations used to define a region should be large enough to smooth the variability of at-station estimates, and also small enough for regional estimates to represent local conditions adequately. In the most recent volumes, the regional approach used in NOAA Atlas 14 is based on the region of influence method, where estimates from each observed location (i.e. precipitation gauge station) are calculated by pooling information from a unique combination of nearby stations. These stations are selected based on an examination of their distance from a target station, elevation difference, and geography (e.g., distance from the coast), and an assessment of similarities in the progression of relevant L-moment statistics across durations (Perica et al., 2018).

Once the regional estimates are calculated at each gauged location, the ungauged locations are calculated using gauged regional estimates and an interpolation technique based on grids of the mean annual maxima (MAM). The MAM grids for each duration were developed by Oregon State University's PRISM Climate Group (e.g., Daly et al., 2002) using a hybrid statistical-geographic approach for mapping climate data, named PRISM (Parameter-Elevation Regressions on Independent Slopes Model), and gauged MAM data. The MAM grids reflect the

terrain's ability to affect precipitation patterns, because both topography and coastal effects are accounted for in the PRISM product development (Perica et al., 2018). The NOAA Atlas 14 spatial interpolation procedure maintains the extreme rainfall characteristics developed at the gauge locations, while introducing the MAM dataset's spatial variability into the final gridded NOAA Atlas 14 product.

In NOAA Atlas 14, precipitation frequency estimates are calculated independently for the following 17 durations: 15 minutes, 30 minutes, 1 hour, 2 hours, 3 hours, 6 hours, 12 hours, 1 day, 2 days, 3 days, 4 days, 7 days, 10 days, 20 days, 30 days, 45 days, and 60 days. Because the estimates are developed independently, inconsistencies between durations (e.g., a 24-hour estimate has to be at least equal to or greater than a 12-hour estimate) can emerge. Duration-based consistency checks and adjustments between all grid cell values across all durations and frequencies are performed to ensure consistency. Furthermore, the final intensity–depth–frequency (IDF) curves are smoothed using cubic spline functions for use in the engineering design. For a detailed description of any of the processes used to develop the stationary NOAA Atlas 14 estimates and the quality control procedures performed on AMS data, please refer to the NOAA Atlas 14 Volume documentation (Perica et al., 2018).

1.2. Nonstationary NOAA Atlas 14 needs and related terminology

The current NOAA Atlas 14 development, production, and publication processes require reexamination and alterations to account for changing precipitation regimes resulting from effects of nonstationary climate. The aim of this research is to reexamine the current NOAA Atlas 14 development process, while considering the production process.

For this study, the following nonstationary needs were considered regarding the development of the nonstationary NOAA Atlas 14 estimates:

- Developing a nonstationary NOAA Atlas 14 process that has the ability to account modeling the temporal nonstationarity and integrate future climate projection into precipitation frequency analysis
- Evaluating the feasibility of incorporating downscaled climate projections into frequency analysis and, if applicable, recommending what downscaled climate projections could be used for this product
- Altering regional and interpolation stationary techniques to work in a nonstationary framework
- Replacing the L-moment method with the GMLE parameterization method
- Allowing the distribution parameters to vary in time, where a type of trend and amount of variation are determined by products developed from CMIP5 datasets

- Modeling a PDS dataset as an alternative to an AMS dataset and abandoning conversion from AMS-based to PDS-based precipitation frequency estimates using a historical empirical relationship
- Characterize uncertainty in the nonstationary estimates

There is substantial uncertainty associated with modeling precipitation frequency estimates (Hosking and Wallis, 1997). In the current NOAA Atlas 14 process, the uncertainty of stationary estimates is characterized by providing a 90% confidence interval for the estimates. The purpose of publishing confidence limits is to help users recognize that the actual value might be different from the “average” value and to encourage them to look at a range of possible scenarios in their designs. The published NOAA Atlas 14 confidence intervals are only associated with sample variation. There is additional uncertainty not accounted for in the currently published confidence intervals, such as selection of the parameter estimation, regionalization, interpolation techniques, distribution selection, data quality, and so on.

With the inclusion of a nonstationary approach there are additional uncertainties, including the ability of climate models and future scenarios to predict future extreme precipitation, the variability of climate models and downscaling techniques, and the large range of potential methodologies. The most notable comes from the ability of the climate models to predict future extreme precipitation under assumed scenarios. Furthermore, the climate models, scenarios, and downscaling techniques are continually evolving, and these advancements in science need to be integrated into the nonstationary NOAA Atlas 14 process.

In this study, the most recent coordinated climate model experiments, the World Climate Research Programme's CMIP5, was used. The CMIP5 was developed by the Intergovernmental Panel on Climate Change (IPCC) in 2013 and is based on four representative constriction pathway (RCP) emission scenarios. The four RCPs, RCP2.6, RCP4.5, RCP6, and RCP8.5, represent a possible increase in radiative forcing amounts (+2.6, +4.5, +6.0, and +8.5 W/m², respectively) at the end of the century with respect to preindustrial values. The increase of +4.5 W/m² accounts for corrective greenhouse gas emission measures, while +8.5 W/m² considers continuous high greenhouse gas emissions. For this research study, both the medium (RCP4.5) and high (RCP8.5) greenhouse gas emission scenarios have been selected for assessment.

During this study, the next generation of the climate model was being developed. The Coupled Model Intercomparison Project Phase 6 (CMIP6) climate projection project started in 2013 and will fully be completed in 2022 (Eyring et al., 2016; ESGF CMIP6 Data Holdings, 2021). The critical difference between CMIP5 and CMIP6 is in the development of future scenarios. In CMIP5, future scenarios are based on the increase in radiative forces (RCPs), while CMIP6 uses socioeconomic pathways (SSPs) and integrates them with the CMIP5 RCP (O'Neill et al., 2016). Further advancement of the climate models is planned for the next couple of decades, with Phase 7 and Phase 8 of the Coupled Model Intercomparison Project. For this reason, the development and production processes of the nonstationary NOAA Atlas 14 should be designed to easily integrate and replace the climate model projections as they become available.

2. Phase I: Local likelihood nonstationary precipitation frequency method

During the period 2016–2018, a team of scientists from the Pennsylvania State University, led by Dr. Benjamin Shaby and Dr. Alfonso Mejia, worked on the development of nonstationary methods for NOAA Atlas 14 precipitation frequency estimates. The objectives of this study were to assess the suitability of state-of-the-science methodologies for nonstationary precipitation frequency analysis, to identify the most suitable method for NOAA Atlas 14, and to test the proposed method on the Volume 10 Northeast project area.

The final deliverables from the Penn State group included the proposed methodology for nonstationary precipitation frequency analysis, and the final report (see Appendix A-1). In this section, the HDSC summarizes the proposed Penn State method. Additional work performed as part of Phase I research (e.g. trend detection), not summarized in this section, is presented in Appendix A-1.

2.1. Overview of the Phase I research project

The Penn State research group proposed a nonstationary method based on a local likelihood approach, where precipitation estimates at both observed and unobserved locations are calculated by pooling information from nearby gauge locations, and weighted using a kernel function and a bandwidth. This method also accommodates any list of covariates, both in the calculation of the distribution parameters and of the weights that determine a contribution of pooled information at the predicted location. Incorporating the climate model as a covariate in calculating the distribution parameters allows this model to account for the temporal nonstationarity in data (e.g. trends) and compute precipitation frequency estimates for future dates, while incorporating the spatial covariates (elevation, mean annual precipitation, etc.) in the weighting procedure allows this model to explain differences in extreme precipitation spatially without the use of additional interpolation techniques. Alternatively, the spatial covariates also can be incorporated in calculating the distribution parameters.

This methodology can be applied to both AMS and PDS data. As with the current NOAA Atlas 14 model, AMS and PDS time series are needed at gauge locations. The AMS extraction process from each historical precipitation record is simple, given that a maximum of only one event a year over the observation period is needed. The PDS extraction process is more complex, because the choice of threshold can alter the final estimates and the threshold selection changes with climate regions and conditions. The choice of threshold for the PDS data should be high enough to satisfy the assumptions of the extreme value theory and to reduce the bias in the parameter estimates. Simultaneously, it should be low enough to increase the precision of the estimates by including enough data.

Penn State researchers reviewed the literature on the threshold selection and applied several threshold selection techniques (mean residual life, parameter stability, and quantile–quantile plots) to daily and hourly observed rainfall data in the pilot project area. Based on analysis of stations in the pilot region (see Appendix A-1, Section 3.1), the Penn State research team concluded that the 98th-percentile threshold of nonzero observations was an adequate

threshold for the Northeastern pilot project area. For the Northeast, the 98th-percentile nonzero threshold, on average, provides two exceedances in any given year. The 98th-percentile nonzero observations were calculated by removing all days with no precipitation from historic observations at the station, sorting the sample, and calculating the 98th-percentile value of that sample. Throughout this project, the sensitivity of the PDS threshold selection was tested, and this study found that threshold selection between 2 to 3 exceedances per year provides similar estimates for the Northeastern project area. However, the PDS selection should be tested for the different climate regions before national implementation.

The extreme value theory assumes that the events in the time series are independent. In the AMS only the largest event is extracted within each year. The largest event from one year is almost always independent from the largest event from another year. In the PDS, more than one event may be extracted during a year, causing the possibility of strong dependencies between events. Moreover, it is common for more than one extreme to be extracted in a year in which an extreme multi-day storm occurred. For example, if an extreme storm lasts three days, it is possible that two consecutive 1-day events from the storm will pass the PDS threshold. Because these two events come from the same storm, they are not independent and are strongly correlated. The Penn State research team analysis (Appendix A-1, Section 3) revealed that dependent events in the time series do not impact the precipitation frequency estimates significantly and that declustering the 98th-percentile nonzero PDS for 1 day using a 1-day separation period between events would be sufficient for the NOAA Atlas 14 application. They also recommended that 1-hour 98th-percentile nonzero PDS be declustered using a 1-day separation period, which agrees with the literature (Xuereb and Green, 2012).

The Penn State team performed a simulation test to evaluate how three distribution parameter estimators (L-moments, MLE, and GMLE) perform under varying sample sizes and for a range of shape parameters (Appendix A-1, Section 4). The simulation analysis was performed for both the GEV and GPD parameters. The sample size criteria are essential because most of the extreme precipitation observation sample sizes are, statistically speaking, small. The average sample size at the gauge locations in the NOAA Atlas 14 volumes is approximately 60 years for daily durations and 40 years for hourly durations. The sample size for the nonstationary framework might be even smaller because the research community recommends using only observations after the 1960s, when the change in climate temperature was observed (Kilgore et al., 2016). Thus, it was important to evaluate the performance of all distribution parameter estimator models for the small sample, given the limited historical observation record.

The other criterion in the simulation test was the distribution shape parameter. The shape parameter in modeling extremes defines the prediction of rare quantiles. Given the limited observation sample, the shape parameter using a small sample size is rather uncertain (Martins and Stedinger, 2000). It is subject to randomly high and low values resulting from the presence of outliers in the small sample. Shape parameter values that are too high or too low can generate unreasonable quantiles, particularly for the longer ARIs (>100-year). In the current L-moment-based NOAA Atlas 14 approach, the shape parameters of the GEV and GPD distributions are implicitly constrained. This constraint makes the L-moment estimator robust for small sample sizes. The difference between the MLE and GMLE approaches is only in the way the shape parameter is restricted. MLE does not impose such a restriction, making it more

sensitive when a small sample is used, while the GMLE method explicitly constrains the shape parameter, making it more applicable for engineering applications.

Based on the analysis, the Penn State team recommends the GMLE approach, given that it performs similarly to the other two methods, can easily be adapted to incorporate covariates (e.g., to capture time trends), and has the ability to restrict the shape parameter. Penn State recommended a beta distribution probability density function for limiting the shape parameter in the GMLE approach, where the beta density function proposal is in the following form and is centered at the shape parameter of 0 (Eq. 1):

$$\pi(\xi) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} (0.5+\xi)^{p-1} (0.5-\xi)^{q-1}, \quad \xi \in [-0.5, 0.5] \quad [\text{Eq. 1}]$$

where ξ is the shape parameter, and p and q are beta distribution coefficients. The Penn State researchers recommended penalized coefficients range between -0.2 and $+0.2$ (see Appendix A-1, Section 8.1). However, for the application of the NOAA Atlas 14 for all the regions of the contiguous United States, a penalized coefficient range of -0.5 to 0.5 is more adequate and less constraining. Additionally, other studies recommend using regional historical information to precondition the shape parameter (Martins and Stedinger, 2000).

Nonstationary extreme conditions are modeled similarly to stationary conditions, except for how distribution parameters are calculated. In nonstationary analysis, one or more distribution parameters are assumed to be a linear or nonlinear function of one or more covariates. There are many possibilities for covariate selection in nonstationary analysis. One of the options is the simple year as a covariate, but this covariate does not physically impact precipitation. Instead, covariates that depend on the underlying physical conditions and describe the change in extreme rainfall over time may be a better covariate of choice than time. Nonetheless, the covariate selection is complex (since storm-type and region dependent) and somehow arbitrary, given that there are many possibilities (Nielsen-Gammon, 2020; Agilan and Umamahesh, 2017; El Adlouni et al., 2007).

The goal of the Penn State research was to identify the most suitable method relevant to NOAA Atlas 14 that would be flexible enough to be adjusted and to include various covariates that would be investigated by the second phase of this project. For the identified model, the Penn State research recommended using the linear function of one or more covariates. In their final report, they used CMIP5 (temperature) datasets to account for the temporal changes and the digital elevation model (DEM) elevation covariate to add the spatial component into modeling the parameters. However, these two covariates were used only for testing the model. The second phase of the project evaluated the precipitation element from the climate models as a covariate.

Although the Penn State research did not evaluate covariates for NOAA Atlas 14 use, it did test different relationship functions between the parameters and covariates, and found that the linear (and log-linear for the scale parameter) relationship provides the most reasonable estimates. The research team also looked at how these final estimates were impacted by adding the covariates into location, scale, and shape, and all three distribution parameters. Based on their analysis (see Appendix A-1, Section 8.2), the final recommendation was to keep the shape

parameter constant over time while modeling both the location and scale parameters as a linear and log-linear combination of covariates. The final recommendation is based on the following model (and example with the two covariates):

$$\text{location} (\mu_{ns}) = a_0 + a_1 \times cov_1 + a_2 \times cov_2 \quad [\text{Eq. 2}]$$

$$\ln(\text{scale}(\sigma_{ns})) = b_0 + b_1 \times cov_1 + b_2 \times cov_2 \quad [\text{Eq. 3}]$$

$$\text{shape} (\xi_{ns}) = c_0 \quad [\text{Eq. 4}]$$

where μ_{ns} , σ_{ns} , and ξ_{ns} are the location, scale, and shape nonstationary parameters, and a_0 , a_1 , a_2 , b_0 , b_1 , b_2 , and c_0 are optimized parameter coefficients. The covariate location and shape parameter contributions are designated with the subscript *cov*.

In this proposed model, the GEV distribution is used to approximate the AMS, while the Poisson point process (PPP) parameterization of the GPD is used to fit the PDS. The advantage of the PPP characterization model is that it unifies the GEV and GPD models (easy transformation between the GEV and GPD) and provides a natural way to model nonstationarity in threshold exceedances. In the PPP model, the rate of exceedances and the distribution of independent exceedances are modeled separately (Roth et al., 2012). The rate of exceedances can be modeled using the homogeneous or nonhomogeneous PPP. The homogenous PPP is assumed in the case of a stationary climate, whereas the nonhomogenous PPP is used to model nonstationary climate (Coles, 2001).

Figure 1 shows the simplified schematic of the proposed Penn State methodology framework. The framework allows for calculation of precipitation frequency estimates at the gauge or grid locations, and can use gauge-based or the climate-based time series data. It is designed to calculate stationary, quasistationary, and nonstationary precipitation frequency estimates, where the choice of spatial and parameter covariates can be easily added, replaced or removed. Furthermore, the covariates can be used to model temporal nonstationarity within the data and beyond the observational period. It models both AMS and PDS data.

The proposed method extracts AMS or PDS data at the gauge locations and calculates the CMIP5 covariate at those locations. Additional covariates that may help with the spatial distribution, such as the PRISM MAP), can also be included. For a prediction point (e.g., grid cell) where precipitation frequency estimates are needed, the distance between the grid cell and the gauge locations is calculated. The weights are then assigned based on the distance, calculated using the triweight kernel function, where the weights decrease to a certain maximum radius (e.g., station distance of >50 miles = 0 weight). Other factors can be included in the regional weighting, such as the elevation difference, MAP, and distance from the coast.

The parameters of the distribution (eg. GEV location, scale, and shape) for each prediction grid cell are then calculated using the equations above (Eq. 2–4), where the values (a_0, a_1, \dots, a_n), (b_0, b_1, \dots, b_n), and (c_0, c_1, \dots, c_n) are optimized by minimizing the negative log-likelihood function of the probability distribution using the weighted observed station data (and, if wanted, the penalized shape parameter). Figure 2 shows an example of how the negative log-likelihood and parameters of the GEV distribution are optimized at one grid cell using the weighted pooled

data (AMS and two covariates) from gauges located within the 50-mile radius. In this example, the weights are assigned based on the distance calculated using the triweight kernel function, while the RCP delta radiative forcing and the square root of the PRISM mean annual precipitation (MAP) are used as covariates. In this example, RCP delta radiative forcing covariate accounts for the temporal nonstationary and can model the changes throughout observations and beyond the observations.

The image in the top right panel in Figure 2 shows the extent of the region for the prediction cell where the gauges that contribute the data are marked in red. The color scale of the weighting distribution (dark to light blue to pink) shows how data from gauges closer to the prediction point get significantly higher weights than gauges close to a 50-mile radius. The gauge data that are closer to a 50-mile radius barely contribute any information to the prediction cell given that assigned weights are almost zero. In this example, the weighting distribution looks symmetrical, but that may not be true if MAP or distance to the coast is used as spatial covariates along with the distance from the prediction location. The four graphs in the left panel of Figure 2 show the optimization of the negative log-likelihood and parameter coefficient values (Eq. 2–4). The table excerpt in the bottom right panel shows the input data for the final iteration and final parameter coefficient values. The initial parameter coefficient values are set to the stationary values, and parameter coefficients, a_0 , a_1 , a_2 , b_0 , b_1 , b_2 , and c_0 were optimized until the negative log-likelihood function was minimized. Note that each extreme event is weighted separately in the optimization procedure, allowing the gauge data with the shorter records to be used in the analysis without merging it with the nearby gauge. This process is advantageous compared to the current methodology because it simplifies the process of data quality control. Appendix A-1 details the model recommendation and additional work performed during this project phase.

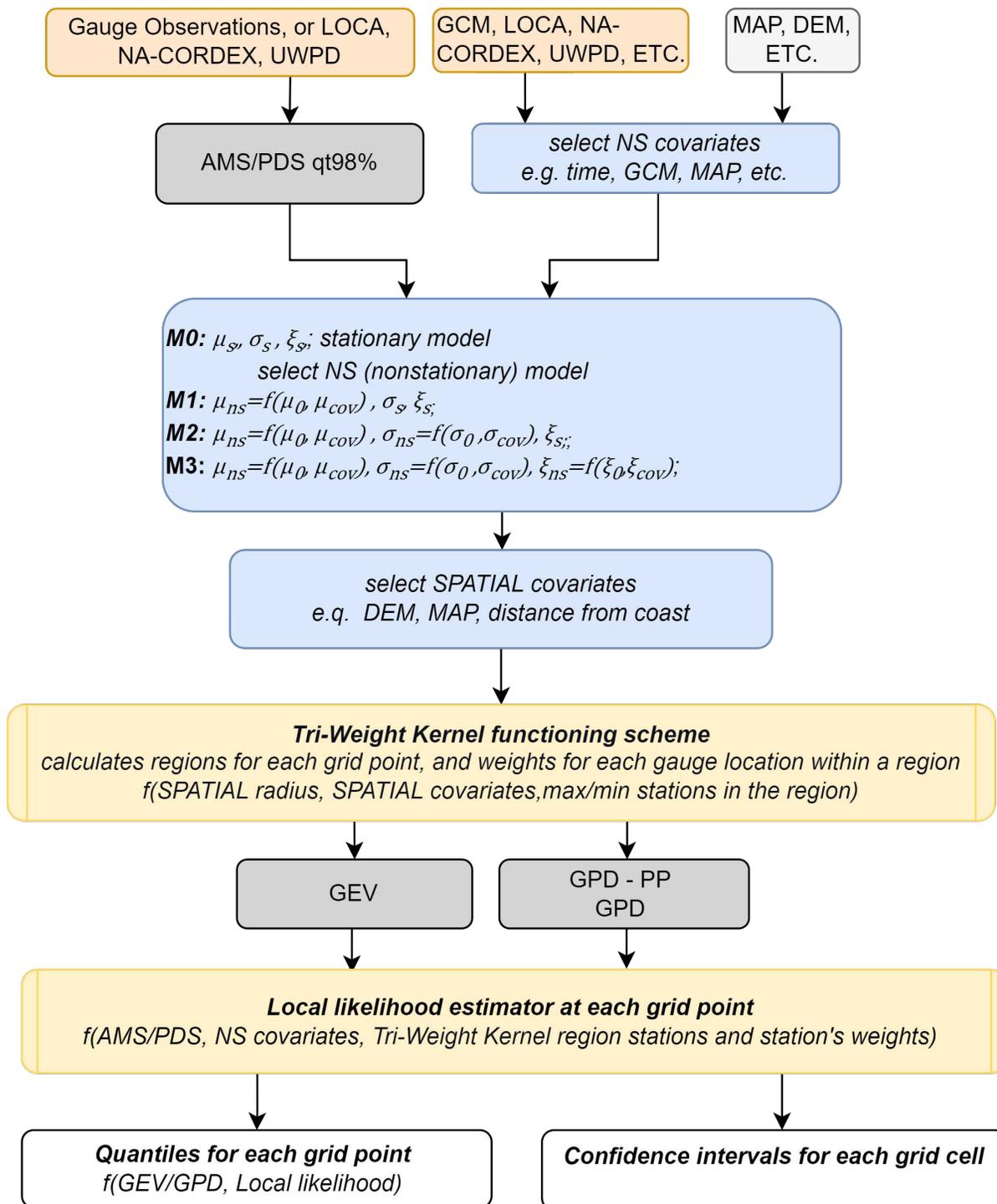


FIGURE 1: Summary schematic of the proposed Penn State nonstationary methodology [AMS, annual maximum series; DEM, digital elevation model; GCM, global climate models; GEV, generalized extreme value distribution; GPD, generalized Pareto distribution; GPD- PP, generalized Pareto distribution Poisson Point Process; MAP, mean annual precipitation; NS, nonstationary; PDS, partial duration series; QS, quasistationary].

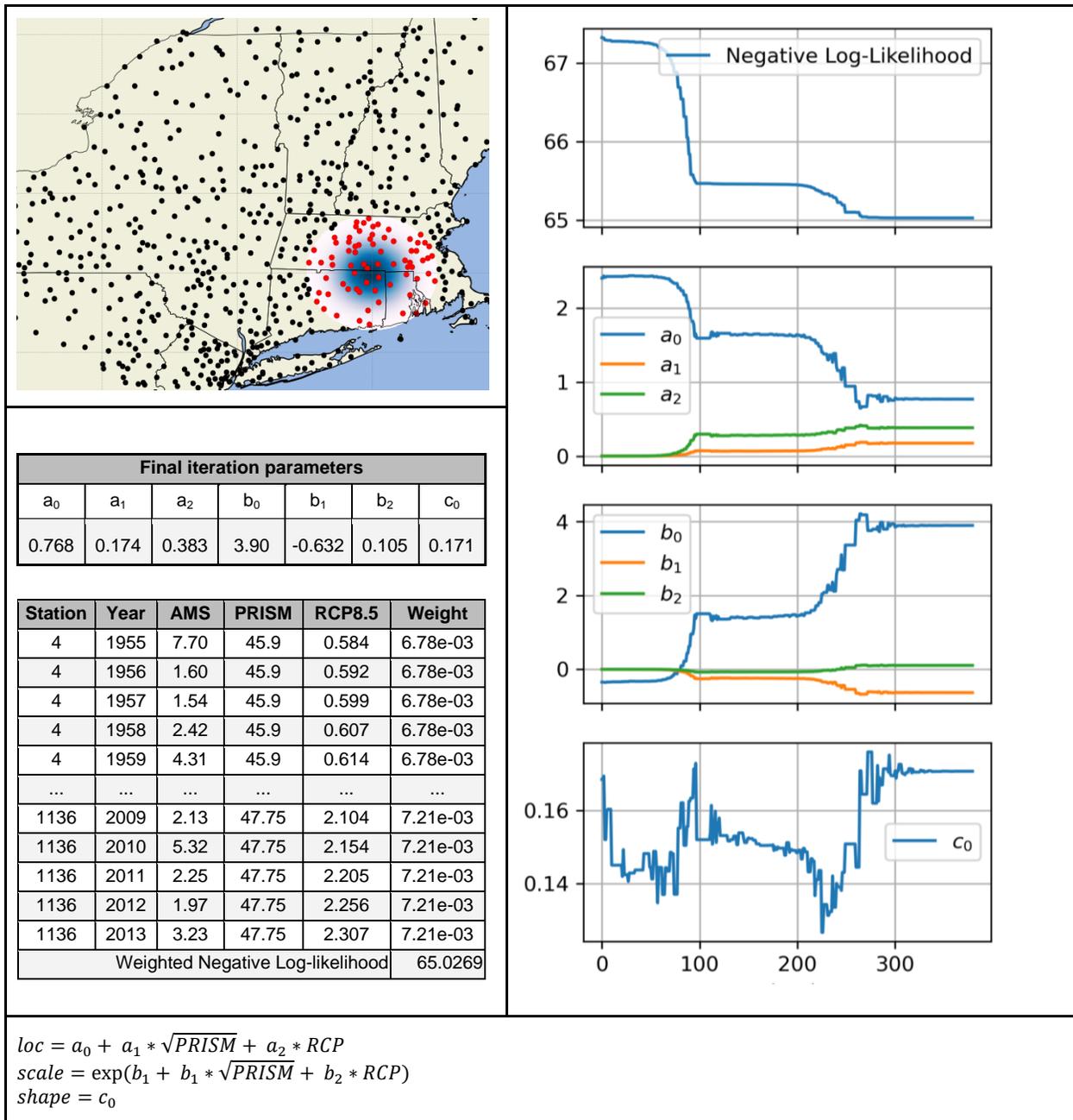


FIGURE 2: Example of how the proposed methodology couples the regionalization and optimization of the nonstationary parameters using covariates through Eq. 2–4. Top left panel, extent of the region (gauges marked in red) and the color scale of the weighting distribution (dark to light blue to pink) for one grid cell in the pilot project area; right panel, optimization iteration and convergence of the negative log-likelihood and parameter coefficient values (Eq. 2–4) for the prediction point shown in the top left panel; bottom left panel, excerpt of input data for the final iteration. In the bottom panel, the equations for GEV location, scale and shape are shown, with their respective covariates.

3. Phase II: Climate model datasets for the calculation of future precipitation frequency estimates

As part of the second phase of the project, a team of scientists from the University of Illinois Urbana-Champaign and the University of Wisconsin-Madison, led by Dr. Momcilo Markus and Dr. Shu Wu, looked at the usability of different climate model datasets for the calculation of future precipitation frequency estimates. The objective of this phase of the study was to assess what climate model datasets and covariates should be utilized in the model developed by the Penn State University in the first phase of the project.

During the period 2018–2020, this research team evaluated four downscaled climate model datasets, including one that was specifically developed for this project, through a comparison with the observed data and developed future precipitation frequency estimates utilizing these downscaled climate projection data. As part of this research, this team also introduced the quasistationary (QS) model to evaluate the generalization accuracy of the nonstationary (NS) model developed by Penn State University. As a result, the final deliverable from this research phase shows an intercomparison of the future AMS- and PDS-based precipitation frequency estimates based on three climate models and two methodologies. The future precipitation frequency estimate intercomparison was performed for the Northeastern United States, for 1-hour, 1-day, and 10-day durations.

Additional work has been performed as part of this research and is presented in the final report included in Appendix A-2. In this section, we summarize the work performed and evaluate the recommendations proposed by this research team.

3.1. Overview of the Phase II research project

The research team from the University of Illinois Urbana-Champaign and the University of Wisconsin-Madison assessed the four datasets downscaled from CMIP5 climate models. The three statistically downscaled datasets considered were as follows: University of California (San Diego)'s Statistical Localized Constructed Analogues (LOCA, Pierce et al. 2014), United States Bureau of Reclamation's Bias-Correction and Constructed Analogs Version 2 (BCCAv2; Maraun et al., 2010), and University of Wisconsin-Madison Probabilistic Downscaling (UWPD; Lorenz 2015; Notaro et al. 2011). The dynamically downscaled dataset considered is the National Center for Atmospheric Research (NCAR) North American Coordinated Regional Downscaling Experiment (NA-CORDEX, 2019). The LOCA, BCCAv2, and NA-CORDEX datasets are downloadable from the public domains, while the UWPD dataset was specifically developed for this research project.

Ideally, for the precipitation frequency analysis needs, all downscaled datasets would have the same high spatial resolutions and would be based on the same climate model pairings, number of realizations, and scenarios. Instead, the four datasets have mixed spatial and temporal resolution, different climate model pairings, and a various number of realizations (Wu et al., 2019). Furthermore, the NA-CORDEX dataset simulation is still ongoing, and at the time this research was performed, only six CMIP5 model simulation results were available. Table 1 summarizes the differences and similarities of the downscaled models considered and the

abbreviations adopted throughout this document for each dataset. For more details on the four downscaled climate model datasets and CMIP5, see Section I of the report in Appendix A-2.

TABLE 1: General summary of the four downscaled climate model datasets considered in Phase II of this research study.

Full name	Downscaling approach	Highest temporal resolution	Spatial resolution (km)	No. of realizations for:	
				RCP4.5	RCP8.5
Statistical Localized Constructed Analogues (CMIP5 LOCA)	Statistical	Daily	~6	32	32
North American Coordinated Regional Downscaling Experiment (CMIP5 NA-CORDEX)*	Dynamical	Hourly	~25, 50	14	54
Bias-Correction and Constructed Analogs (CMIP5 BCCAv2)	Statistical	Daily	~12	20	20
University of Wisconsin-Madison Probabilistic Downscaling (Center for Climatic Research, David Lorenz, 2019) (CMIP5 UWPD)	Statistical	Hourly	~12	308	336

RCP, representative constriction pathway.

*On November 19, NA-CORDEX fixed a problem with the bias-corrected version of the NA-CORDEX data in their archive. In this study, the raw NA-CORDEX data were used, which were not impacted by this error

The research team evaluated the four downscaled datasets (NA-CORDEX, LOCA, BCCAv2, and UWPD) for usability in the precipitation frequency analysis. The assessment was done by comparing the downscaled models with the observed data based on the accuracy of simulating AMS and by contrasting the AMS-based or PDS-based model-based precipitation frequency estimates against observed data-based estimates for both the Northeast and state of Texas project regions for the historical period (hindcast accuracy).

The analysis showing the downscaled models' ability to accurately simulate the AMS data for the 1950–2005 period is described in detail in Section II of the report in Appendix A-2. In summary, the research team calculated the 1950–2005 mean observed AMS at the gauge

locations and compared it to the model-based mean AMS at the gauge locations. To account for the variation between different CMIP5 realizations and models, the research team also computed the median and mean AMS values and the lowest and highest AMS values within realizations and compared them to observations. Among the four datasets, the analysis in this section revealed that the UWPD dataset most often performed better than the other three datasets regarding hindcast accuracy. The LOCA and BCCAv2 datasets both show a dry bias, but only the LOCA datasets exhibit a quasi-linear dry bias that can be corrected. On the other hand, the BCCAv2 dry bias is nonlinear, making a bias correction hard to achieve.

Section III of the report in Appendix A-2 describes a hindcast analysis contrasting AMS and PDS model-based precipitation frequency estimates against observation-based estimates for both the Northeast and state of Texas project regions. The analysis also involves looking at the three different types of thresholds for the PDS and the downscaled datasets' ability to mimic observed PDS thresholds for the historical period (1950–2005). The goodness-of-fit (GOF) statistics revealed that the four downscaled climate models have a reasonable ability to simulate extreme time series data (AMS and PDS) and fair skill in simulating precipitation frequency estimates, and the GOF statistics diminish with the increase in return periods (Appendix A-2, Figures 3.35–3.38). For the NOAA Atlas 14 application, this research team concludes that the fair skill in simulating the higher precipitation frequency estimates should be considered inadequate.

Among the four datasets considered, the UWPD, LOCA, and NA-CORDEX data show higher skill than the BCCAv2 data. The UWPD performed better than the other three with respect to hindcast accuracy, which may also be an impact of having the highest number of ensemble runs. The UWPD and NA-CORDEX datasets have both hourly and daily time series data, while LOCA provides only daily data. The LOCA dataset, however, has the highest spatial resolution and is widely used by researchers and local, state, and federal government agencies. Because of an inconsistent dry bias and inconsistent historical patterns, the BCCAv2 dataset was excluded from further analysis. The NA-CORDEX showed promising results but was initially excluded from further analysis because of the small ensemble size. However, it was added back for the analysis of 1-day durations for the Northeastern region results. Overall, the analysis in the second phase of the project showed that the UWPD performed better than the other datasets, but this dataset did not significantly outperform the other datasets considered.

Overall, the research team concluded that the level of skill of the climate models to simulate the extreme rainfall time series and precipitation frequency estimates for the historical period differs between datasets, between regions, and between time series data analyzed. The skill level for all four datasets decreases with an increase in the ARI. The decrease in estimating higher ARIs could be attributed to the datasets' low spatial resolution and inability to account for the terrain and orographic precipitation effect, as well as the climate model data's inability to simulate the most extreme precipitation events. The analysis reveals that, as other studies have found, the climate models can be a good tool to account for the change in future climate, but that using the model-based precipitation frequency estimates directly may not be adequate for design as absolute values, particularly for the wide range of NA14 design applications. Instead, the downscaled dataset could be applied as a relative change between the present and the future

precipitation frequency estimates (quantile delta change method) that can be used to extend the estimates developed based on past observations into the future.

Section IV of the report in Appendix A-2 introduces the quasistationary (QS) model. This model was introduced to evaluate the generalization accuracy of the nonstationary (NS) model developed in Phase I of this project given that this methodology is commonly used by the research and engineering community for calculating future precipitation frequency estimates (The National Climate Assessment, Easterling et al., 2017; A.T. and Castellano, C.M., 2017; ISWS Bulletin 75, Angel, J. and M. Markus, 2020; New York State Future IDF, DeGaetano, 2017). The QS model, the piecewise stationary method, uses extreme time series data from the downscaled climate model data (projections) to calculate the future precipitation frequency estimates. The QS approach divides the data into subperiods of time and calculates the precipitation frequency within those periods using the stationary methodology. In this research, the QS approach calculates the future precipitation frequency estimates based on the climate model AMS or PDS data for rolling 50-year time periods between 1951 and 2100 every 10 years. The QS method is also applied to the historic (“reference”) period, 1951–2000 (Table 2).

The QS model approach is based on the local likelihood approach developed by Penn State University, which was reduced to the stationary method by setting the parameters for the covariates to zero. In this way, these two models are equalized in all processes (regionalization, parameter optimization, etc.), except in the way the model parameters are represented. However, it is important to mention that the NS and QS models are calculated on different periods and utilize different information from the downscaled climate model datasets. For example, future precipitation frequency estimates using the QS method are based on the AMS or PDS data from each member of the ensemble of models, and then precipitation frequency estimates are averaged for each dataset. The estimates are also calculated for rolling 50-year periods, indirectly allowing all three distribution parameters to vary in time. The confidence intervals are determined by calculating the range from all member runs. In the NS model, 1951–2005 historical gauge AMS or PDS data are used, where relationships between covariates (eg. ensemble average annual precipitation) and the 1951-2005 gauge AMS (or PDS) are calculated, and then extrapolated to the 2006-2100 period using the same covariate. In the NS model, only two out of three parameters vary in time (location and scale). The ensemble average annual temperature and radiative forcing was also tested as covariates.

The final results from the second phase of this research project show the future changes (quantile delta change technique) in the AMS-based and PDS-based future precipitation frequencies using both the QS and NS frameworks on the UWPD and the LOCA datasets for both the Northeast project areas. The spatial grids of the base and future estimates as well as the future changes in 1-hour, 1-day, and 10-day precipitation frequency for the Northeast are detailed in Section V of the report.

A summary of the results, their interpretation, and further discussion are provided below. The analysis of the 1-day NA-CORDEX future precipitation estimates for the Northeast project area has been added to this summary, even though the results were not included in the final University of Illinois Urbana-Champaign and University of Wisconsin-Madison report.

TABLE 2: Target years for comparisons between datasets using the quasistationary method.

Target year	Time period
1975	1951–2000
1985	1961–2010
1995	1971–2020
2005	1981–2030
2015	1991–2040
2025	2001–2050
2035	2011–2060
2045	2021–2070
2055	2031–2080
2065	2041–2090
2075	2051–2100

3.2. One-day and ten-day durations

The summary of the project results shows the future changes in the AMS-based and PDS-based future precipitation frequencies using both the QS and NS methods on the UWPD, LOCA, and NA-CORDEX datasets for the Northeast project areas. The future changes in the 2-year and 100-year estimates have been calculated for the entire Northeast project area for two emission scenarios (RCP4.5 and RCP8.5) and are shown in Figure 3. Overall, the regional average 2-year 1-day estimates show an increase of between 5% and 22% by the end of the century, while the 100-year 1-day estimates show an increase of between 7% and 30%. This average rate of increase is for the entire project area but the range of increase is much wider and varies significantly spatially.

For the 1-day duration, the UWPD dataset, on average, has the highest rate of increase in comparison to the LOCA and NA-CORDEX datasets, while the NA-CORDEX has the lowest rate of increase for both methods considered. However, the highest rate of increase varies between the datasets, depending on the ARIs or emission scenario considered. Furthermore, the variability between the datasets varies considerably spatially, although some of the variability may also be an impact of the difference between the simulation runs. The UWPD dataset has over 300 realizations for each emission scenario, while the other datasets have less than 55.

For the same method (QS), the 2-year AMS RCP8.5 estimates for LOCA, NA-CORDEX, and UWPD show average regional future changes of 17.9%, 17.9%, and 21.1%, respectively by the end of century. The 100-years estimates for the same datasets, emission scenarios, and time periods equal 24.7%, 18.6%, and 28.8%, respectively. The variability increases further for higher ARIs. For 1,000-years estimates (see Appendix A-2), the increase rate for these datasets is 31.9%, 20.5%, and 35.8%, respectively.

The results in Figure 4 for the Northeast 10-day duration also show major variability between datasets (LOCA and UWPD) and methods (QS and NS). For the 10-day duration, the variability of the dataset has a higher impact on the future precipitation frequency estimates than the variability introduced by the methods. The precipitation frequency increases based on LOCA are higher than those found for UWPD using both the QS and NS methods. The QS method produces overall higher estimates than the NS method, but the difference between LOCA and UWPD based on the QS method is the most pronounced. For example, the regional average 100-year 10-day AMS estimates in 2075 under RCP4.5 for the LOCA using the QS method show an approximate 19.4% increase, while UWPD using the QS method shows an approximate 13.4% increase. The same quantile for LOCA using the NS method results in an approximate 10.4% increase.

The magnitude of the percent change in the future estimates varies between the duration and ARIs. For example, using QS estimates, the greatest average regional future change for 100-year 1-day estimates under RCP8.5 is calculated using the UWPD dataset, while for 100-year 10-day estimates under RCP8.5 is calculated using the LOCA dataset. The average regional rate of increase seems to be more consistent for the UWPD datasets, where the rate of future increase is consistently higher for the 1-day duration than for the 10-day duration for all return periods and emission scenarios considered. The LOCA dataset shows more variability; the percent change for RCP4.5 shows a higher increase for the 10-day duration than for 1-day durations for both 2-year and 100-year estimates. In contrast, the RCP8.5 percent change for the 10-day duration is greater than that for the 1-day duration for the 100-year estimates and approximately the same for the 2-year estimates.

One reason for the variability of the LOCA dataset in respect to emission scenarios and return periods may be the fact that the annual maximum events extracted from the RCP4.5 in the 2020–2040 period are significantly higher than the annual maximum events extracted from the RCP8.5 realizations for the same period. Based on the trends extracted from the delta radiative forcing, the precipitation in the two emission scenarios is the same until 2002 and starts slowly diverging from 2020 to the end of the century. The NS method is less sensitive than the QS method to these types of climate-based AMS artifacts. The QS method, where estimates are based on 50-year data periods, is more sensitive to the AMS data variability between the emission scenarios.

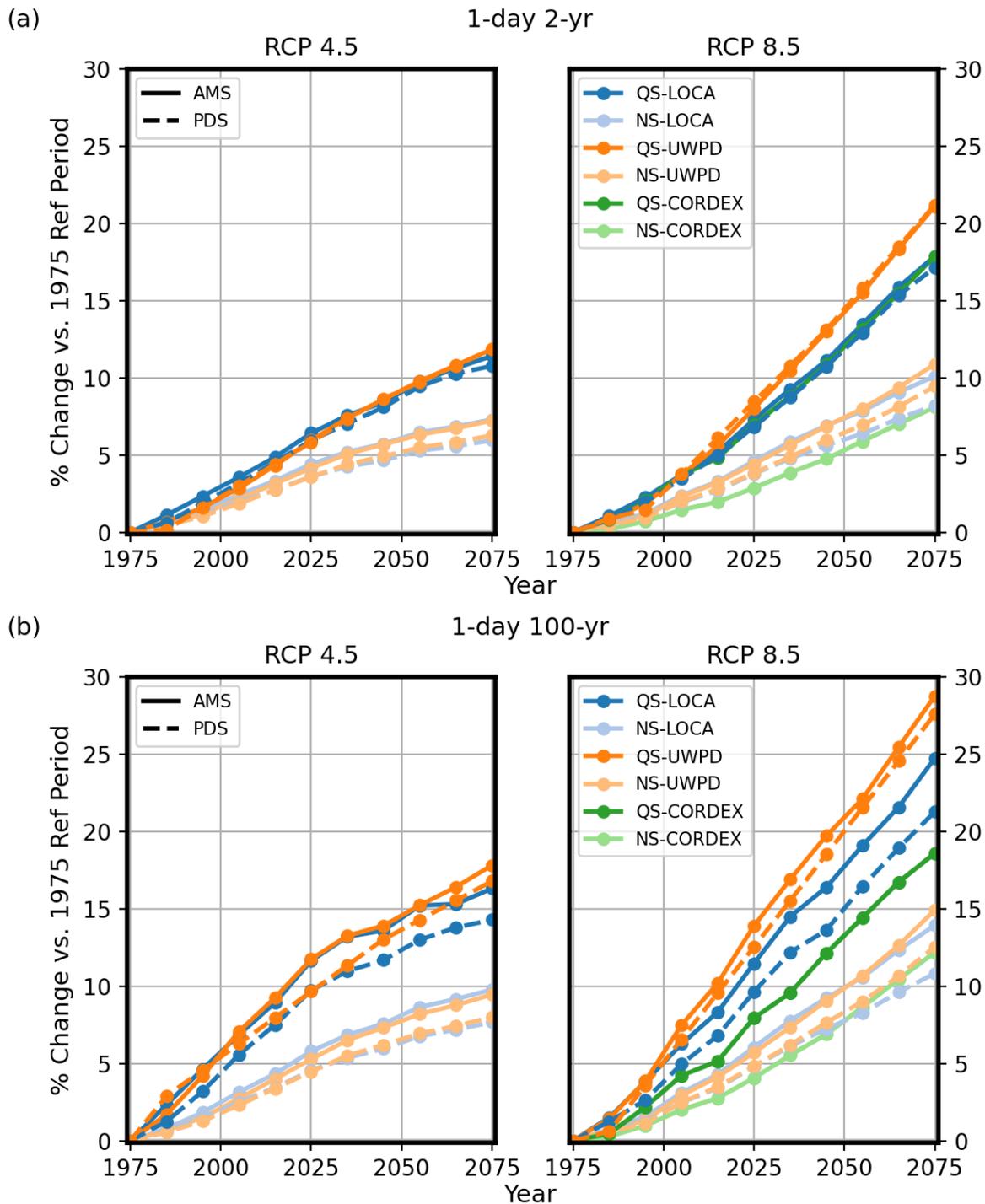


FIGURE 3: Rate of average increase in the future (a) 2-year and (b) 100-year 1-day precipitation frequency estimates for the Northeast United States for the RCP 4.5 and RCP 8.5 emission scenarios and for the base period (1951–2000), centered around 1975.

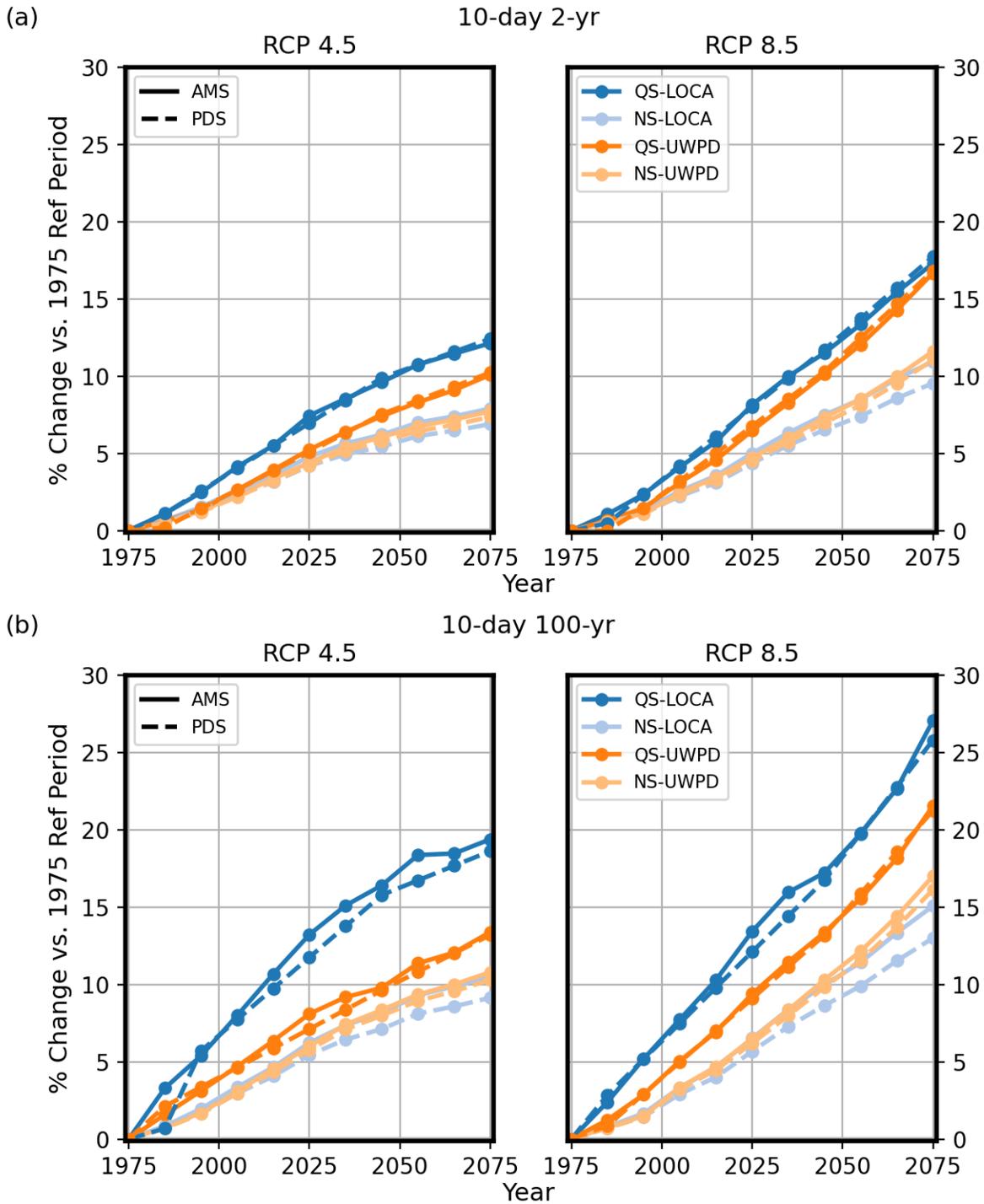


FIGURE 4: Rate of average increase in the future (a) 2-year and (b) 100-year 10-day precipitation frequency estimates for the Northeast United States under the RCP 4.5 and RCP 8.5 emission scenarios and for the base period (1951–2000), centered around 1975.

3.3. One-hour duration

The final reports from the University of Illinois Urbana-Champaign and the University of Wisconsin-Madison for the 1-hour duration in the Northeast region show that the difference between the NS and QS methods is more substantial for this duration (Figure 5). The regional average 100-year 1-hour NS estimates result in an approximate 4% increase in 2075 under RCP4.5 while using the QS method results in an approximate 12% increase.

The lower rate of increase for the NS method than in the QS model is most likely due to the mismatch in how the two models use the downscaled climate model datasets. The QS model uses model-based AMS data, while the NS model uses gauge-based AMS data where climate model-based covariates are used to forecast extreme precipitation past historical observations. The difference for the two methodologies is most pronounced at the hourly duration due to the sparse historical hourly gauge network. The HDSC conducted a complementary analysis to assess the differences between the two methods. Section 4 of this report shows the results of this analysis and insight into the difference between the two methods using the daily LOCA AMS.

The 1-hour NS estimates are calculated using the 1-day mean annual precipitation ensemble as a covariate for UWPD and LOCA NS analysis. In contrast, the UWPD QS used the hourly data directly from this climate model to calculate the future increase. It is essential to mention that the climate models available at hourly temporal resolution are more limited. The UWPD dataset has three members for hourly temporal resolution and 14 members for daily temporal resolution for each model considered. Furthermore, the dynamically downscaled NA-CORDEX dataset was not used for this analysis, either using the QS or NS method. Given that this dataset is dynamically downscaled, it would be essential to calculate the future hourly estimates and compare them to UWPD using both the QS and NS methods based on the entire downscaled climate model extreme time series data, even if it has limited the small number of members.

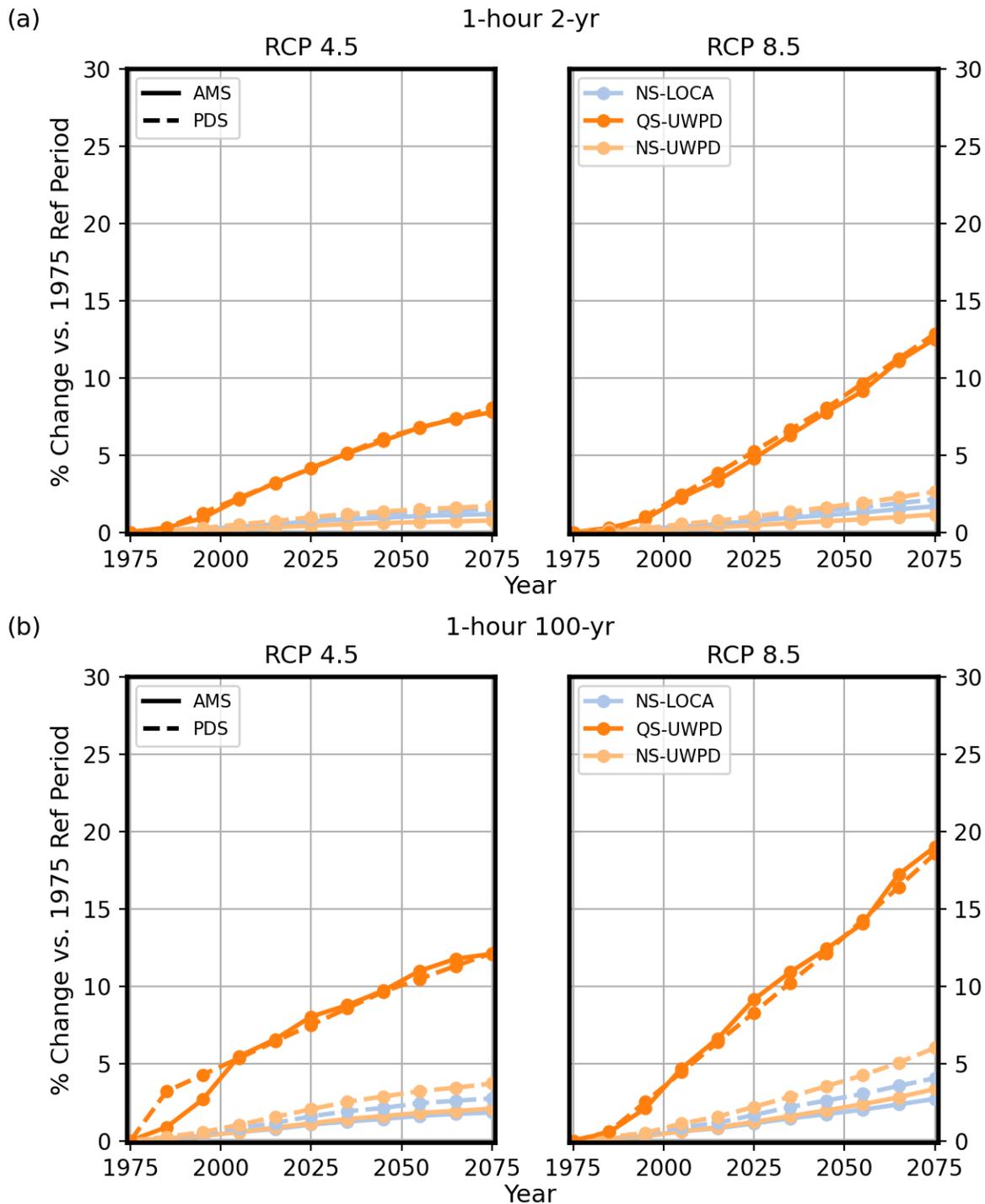


FIGURE 5: Rate of average increase in the future (a) 2-year and (b) 100-year 1-hour precipitation frequency estimates for the Northeast United States under the RCP 4.5 and RCP 8.5 emission scenarios and for the base period (1951–2000), centered around 1975.

4. Complementary analyses

The HDSC conducted several analyses to further complement the work performed in phases I and II of the project. These analyses are described here and their results are reported.

The project's Phase II results show that future estimates depend strongly on the methodology being applied. For example, using the same downscaled climate datasets, the NS method shows approximately half the rate of increase in future daily estimates compared to the QS method. The difference between the NS and QS methodologies is even more pronounced for hourly durations. The NS method shows almost no change (4%) in the future hourly estimates compared to the base period. In contrast, the QS method shows an approximately 12% increase for the Northeast using the UWPD dataset. Given these discrepancies, we performed a rigorous comparison of the two models, QS and NS.

To truly compare the two methods' performance, we applied LOCA AMS data to both NS and QS methods only in the Northeastern project area. Thus, we made modifications to the NS method. Instead of extrapolating past the historical observations, we applied the same type of downscaled climate extreme time-series data to the NS method to match the data applied to the QS method. Following three modifications to the NS method were made and then compared to the QS method:

1) Apply the NS model to LOCA AMS data for the entire 1950–2099 period.

The NS distribution parameters were calculated using the model-based data, similar to QS from 1950 to 2099, instead of using the gauge data for the base period only. This way, both methods were applied to the same time series dataset and period.

2) Allow the shape parameter to vary with covariates in the NS model (NS3).

The QS method, in effect, allows the shape parameter to vary (e.g., each stationary data chunk has a different shape parameter), while the fixed shape parameter was used for the NS. In this analysis, we calculated the NS estimates, where the location and scale vary with respect to covariates (NS2) and where the location, scale, and shape vary with respect to covariates (NS3).

3) Test choice of covariates in the NS method (time and RCP delta radiative forcing).

To test the sensitivity of the NS method to the choice of covariates, we applied two different covariates to the NS model (both NS2 and NS3). The two covariates tested were a simple "year" and RCP delta radiative forcing. Note that application of covariates over the period of climate-model time series is used to inform the temporal trend in data and is fit through the LOCA AMS data distribution parameters over the entire 1950–2099 period.

In total, we tested four NS models, shown in Table 3, and compared them to one QS model using the 32 simulations in the LOCA AMS dataset. The estimates using the five models were calculated using a regional GEV approach, where regions are constructed using a 50 km (~31 miles) radius, and weights are assigned to data within the region using a triweight kernel. The reference periods for the five methods in this analysis are slightly different from the analyses in Section 3 (Table 3). The data used starts in 1950, and ends in 2099. The 2099 ending date was chosen due to some of the 32 LOCA models not having data through 2100.

TABLE 3: Nonstationary (NS) model comparison in reference to quasistationary (QS) model.

Method	Parametrization	Emission	Covariate
QS	$\mu = \mu_0$ $\ln(\sigma) = \ln(\sigma_0)$ $\xi = \xi_0$	RCP4.5, RCP8.5	None
NS2–time	$\mu = \mu_0 + \mu_1 \times year$ $\ln(\sigma) = \ln(\sigma_0 + \sigma_1 \times year)$ $\xi = \xi_0$	RCP4.5, RCP8.5	Year
NS2–forc	$\mu = \mu_0 + \mu_1 \times forcing$ $\ln(\sigma) = \ln(\sigma_0 + \sigma_1 \times forcing)$ $\xi = \xi_0$	RCP4.5, RCP8.5	Radiative forcings
NS3–time	$\mu = \mu_0 + \mu_1 \times year$ $\ln(\sigma) = \ln(\sigma_0 + \sigma_1 \times year)$ $\xi = \xi_0 + \xi_1 \times year$	RCP4.5, RCP8.5	Year
NS3–forc	$\mu = \mu_0 + \mu_1 \times forcing$ $\ln(\sigma) = \ln(\sigma_0 + \sigma_1 \times forcing)$ $\xi = \xi_0 + \xi_1 \times forcing$	RCP4.5, RCP8.5	Radiative forcings

We compared the five models through a regional average analysis and by visualizing spatial patterns. The results of the comparison are shown below. Overall, following a more consistent approach, the future increase rate between the NS and QS models matches more closely.

All methods compare closely for both emission scenarios at 2-year and 100-year ARIs (Figure 6). The QS method shows a higher percent change at 100-year compared to the NS2 and NS3 methods. The NS3 methods show higher percent changes compared to the NS2 methods at the 100-year estimates and more closely match the QS method. The ranges of the regional averages (shaded area of Figure 6) are largest for the 100-year ARI and RCP 8.5 emission scenario. The QS method has the largest range, followed by the NS3 methods. Both the percent change means and ranges are likely the result of allowing changes in the shape parameter in the NS3 methods. The QS method also allows similar changes in the shape parameter because a stationary estimate is calculated every 10 years (for each 50 year chunk), enabling all three parameters to change.

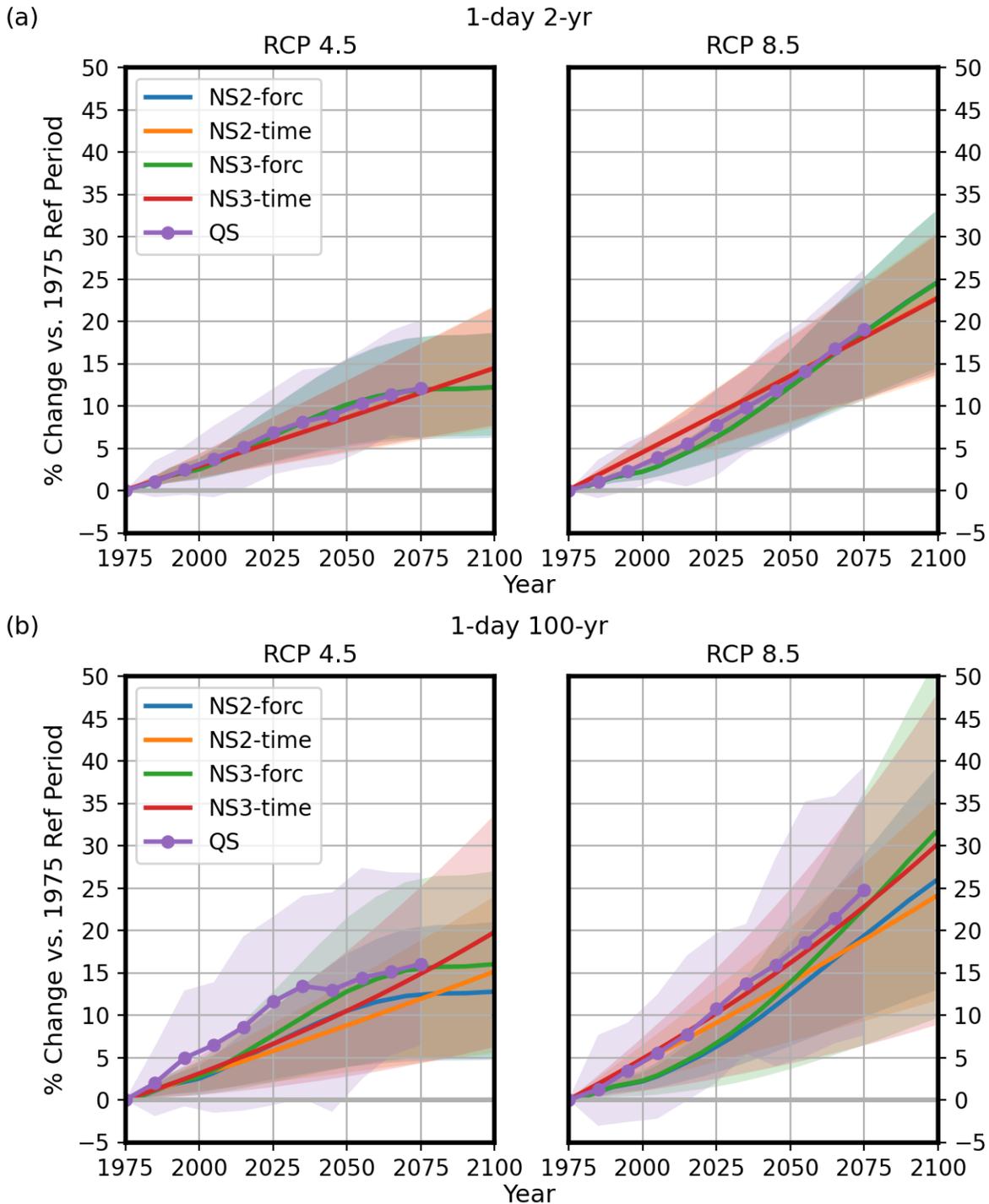


FIGURE 6: Rate of average increase in the future (a) 2-year and (b) 100-year 1-day precipitation frequency estimates for the Northeast United States under the RCP4.5 and RCP8.5 emission scenarios compared to the base period (1950–1999), centered around 1975. Values are in percent change from the base period. Each line represents the regional average of all 32 LOCA models for each method and emission scenario. The shading represents the 90% confidence interval range of the 32-model regional averages.

It is important to look at the spatial variability of the percent changes in addition to the regionally averaged values (Figure 7a to 7e). In the figures, the variability of the percent changes for the 2-year ARI are similar for all five methods, with RCP8.5 having higher overall values. At the 100-year ARI, the spatial variability is much larger for the QS, NS3-time, and NS3-forc methods, among which the QS method has the largest spatial variability. This is most pronounced in the 100-year RCP8.5 emission scenario. Once again, the difference is likely due to how the shape parameter is allowed to vary in time with the QS, NS3-time, and NS3-forc methods. A second reason may be that the 100-year QS recurrence intervals are extrapolated beyond the length of the chosen 50-year AMS period, introducing a greater degree of uncertainty in these estimates than NS methods that use the entire period's data at once. The difference in sample size between the methods affects the higher ARIs (e.g., 100-year) more, due to the uncertainty of the shape parameter. Other studies suggest that it is reasonable to use a stationary (El Adlouni et al., 2007; Nielsen-Gammon, 2020) or a preconditioned shape parameter (Martins and Stedinger, 2000), especially for nonstationary MLE.

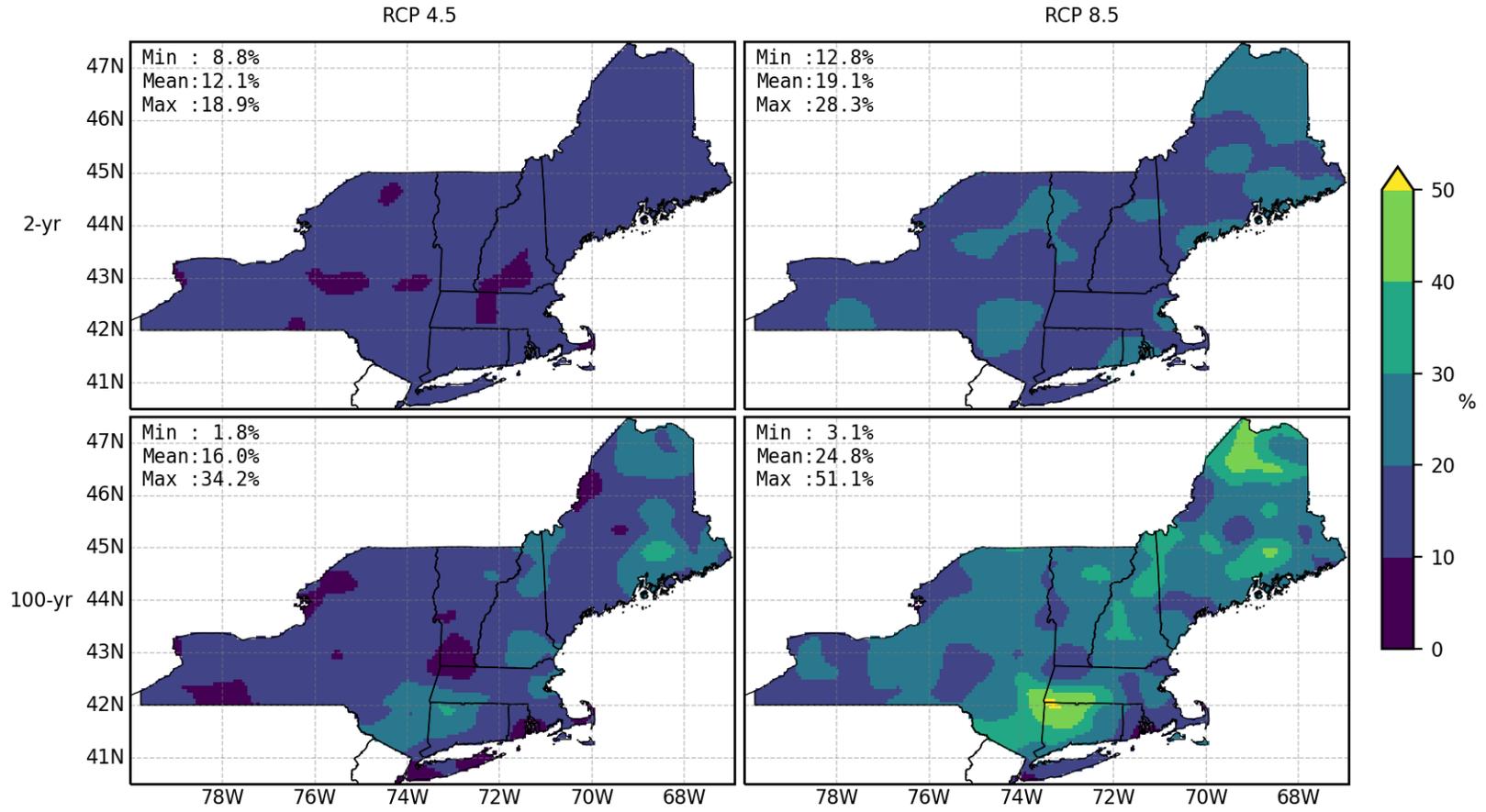
The spatial variability of the percent changes at higher ARIs can have a large impact on engineering decisions. To illustrate this, we zoom into southern New York and look at the spatial patterns in the RCP8.5 2-year and 100-year 1-day estimate percent changes (Figures 8 and 9). The percent changes for Brooklyn, NY, and a location 50 miles (~80 km) north are shown for each method (Table 4). For the 2-year ARI, the percentage changes are within 2%–3% between the two locations for all methods. Conversely, for the 100-year ARI, the two locations are within ~5%–6% for the NS2 methods, ~23% for the NS3 methods, and ~27% for the QS method. The percent changes for the 100-year estimates over a relatively short distance (50 miles) are unrealistic and problematic for engineering applications. These types of spatial artifacts can be eliminated by applying the NS method and limiting the shape parameter to its stationary form (NS2 method), implementing spatial smoothing postprocessing, extending the percent increase spatial patterns from a lower ARI to higher ARIs, or any combination thereof.

TABLE 4: 2-year and 100-year ARI average percent change comparison between four NS and one QS methods using the LOCA 32-model dataset between Brooklyn, NY, and a point 50 miles north under the RCP8.5 emission scenario.

Method	Avg % change at 2-year ARI		Avg % change at 100-year ARI	
	Brooklyn	50 miles N	Brooklyn	50 miles N
NS2-forcing	16.9	19.4	16.7	21.5
NS2-year	16.4	18.6	16.2	20.6
NS3-forcing	17.7	19.1	8.3	31.8
NS3-year	17.1	18.1	8.8	32.1
QS	18.5	19.0	8.9	36.5

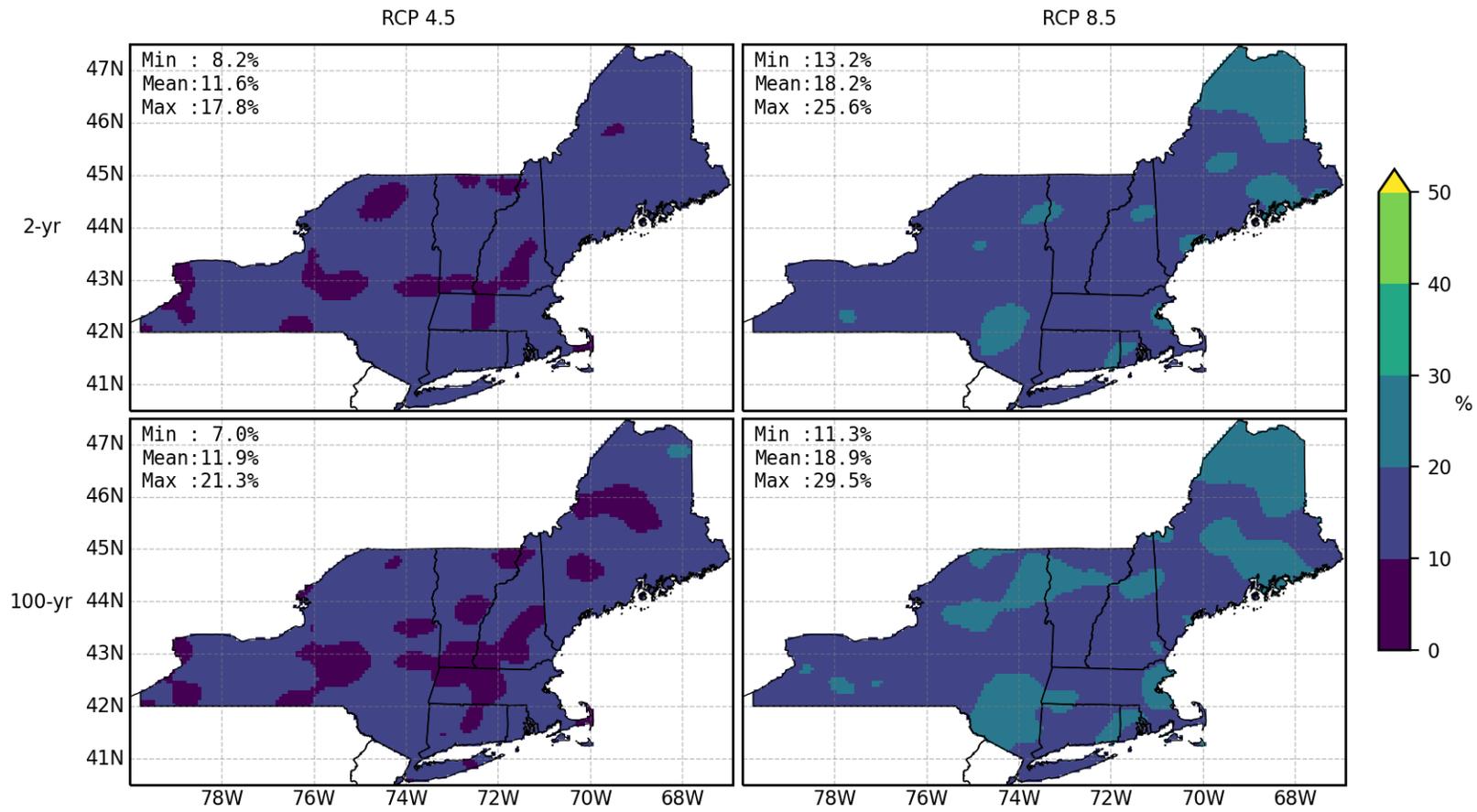
(a)

QS



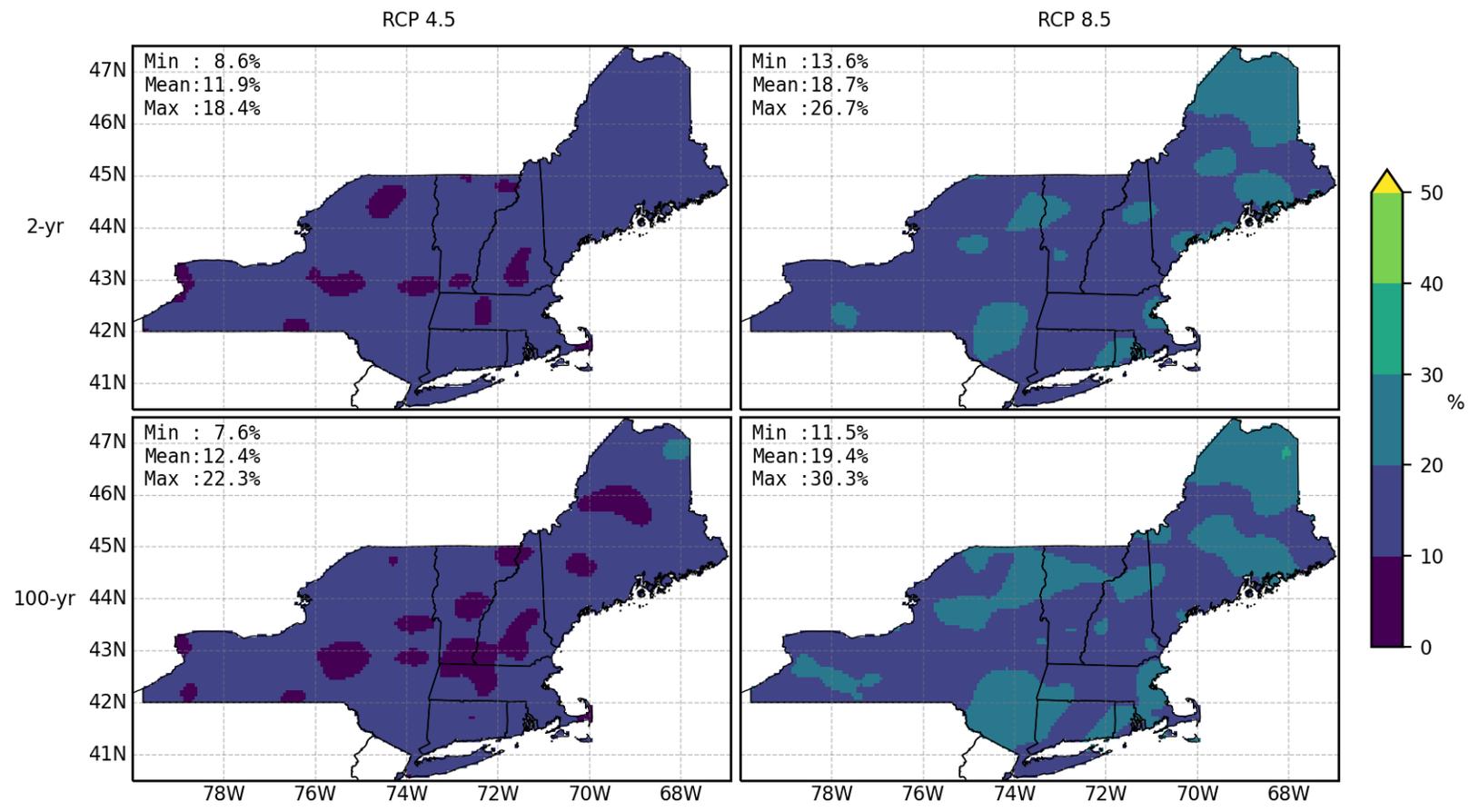
(b)

NS2-time



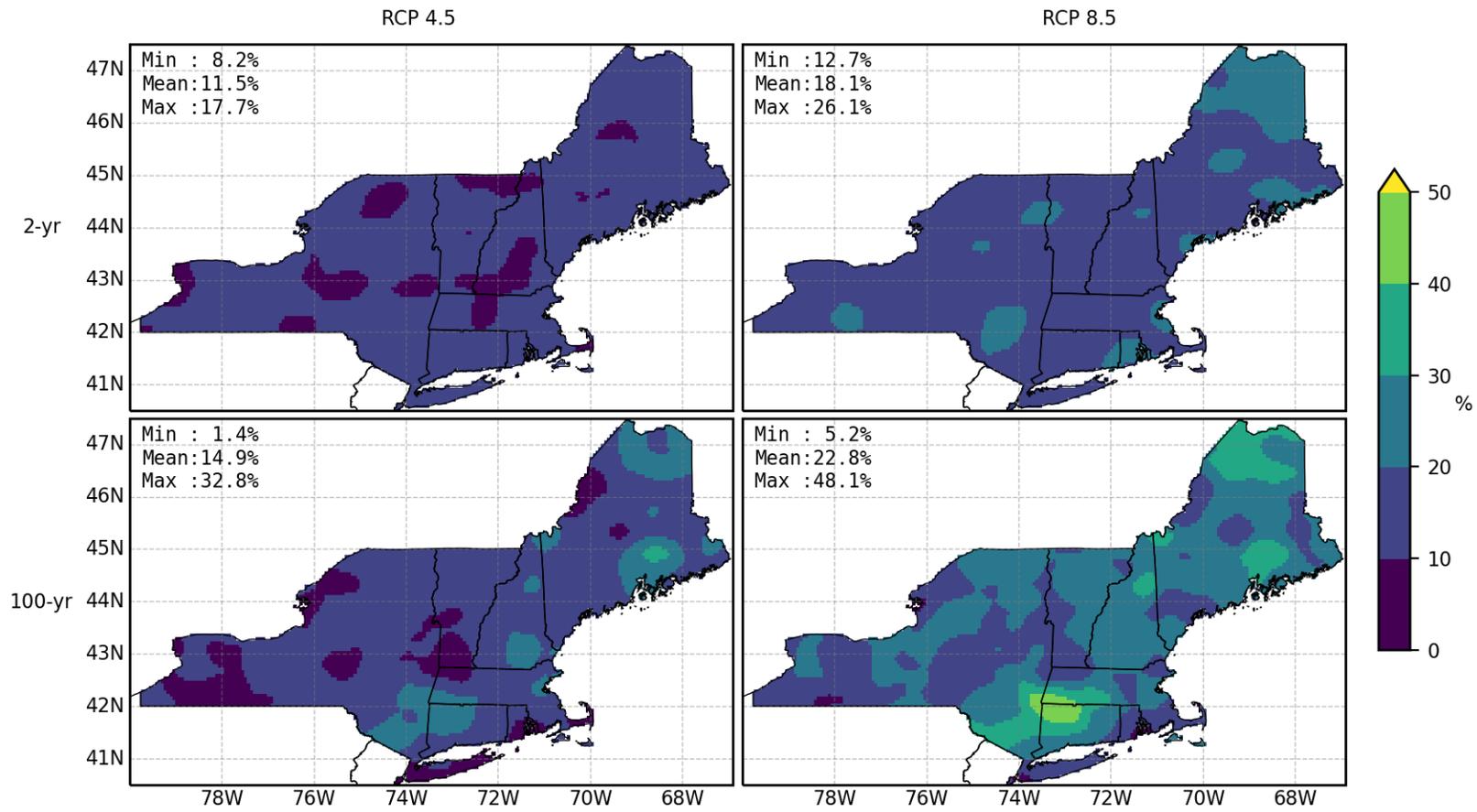
(c)

NS2-forc



(d)

NS3-time



(e)

NS3-forc

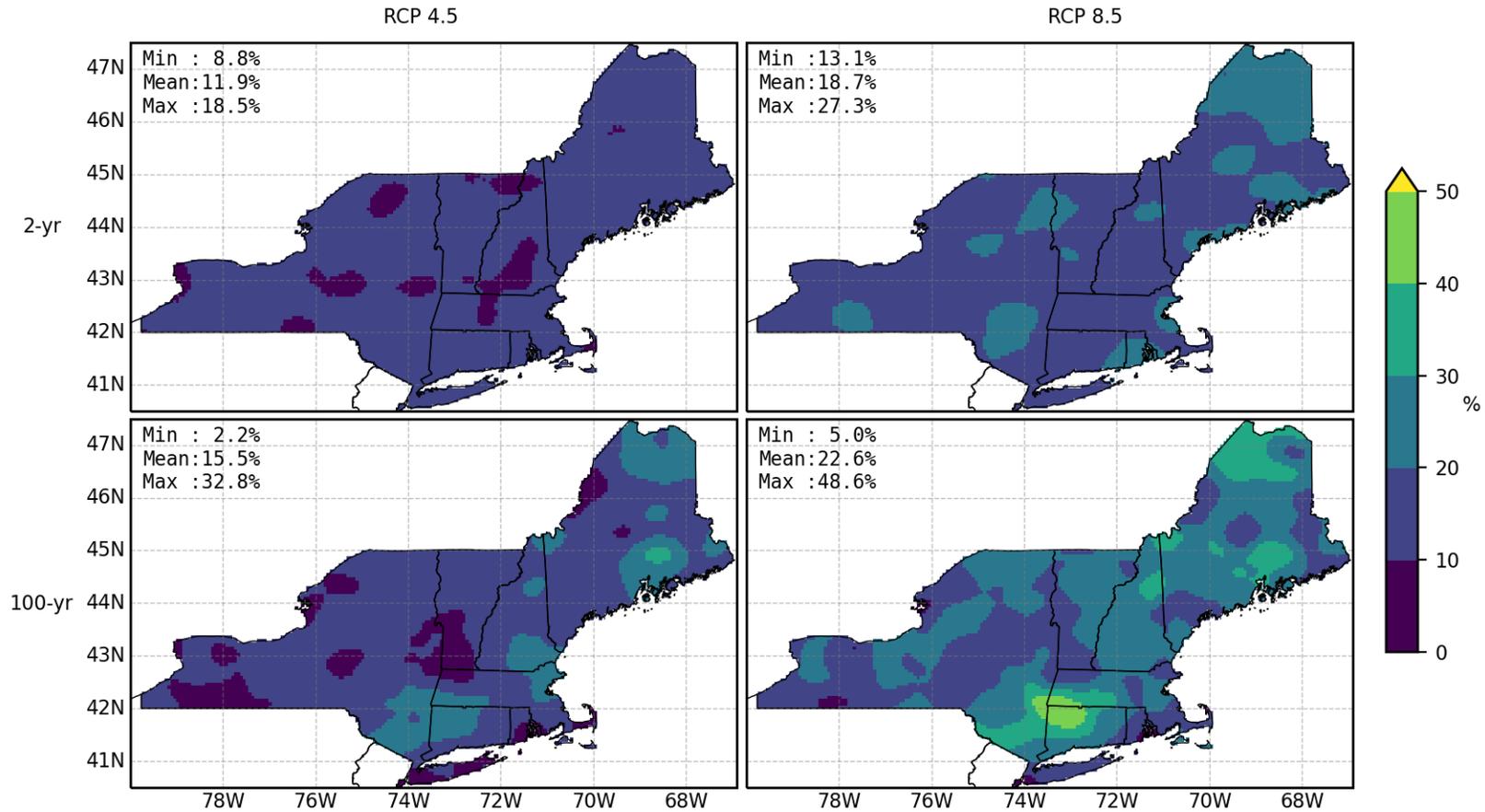


FIGURE 7: Spatial maps of the LOCA 32-model average percent change in the future (reference year 2075) 1-day precipitation frequency estimates for the Northeast United States. The different methods shown are (a) QS, (b) NS2-time, (c) NS2-forcing, (d) NS3-time, and (e) NS3-forcing. Values are in percent change from the base period (1975). The minimum, mean, and maximum values are shown in the top left corner of each map.

RCP 8.5, 2-year 1-day

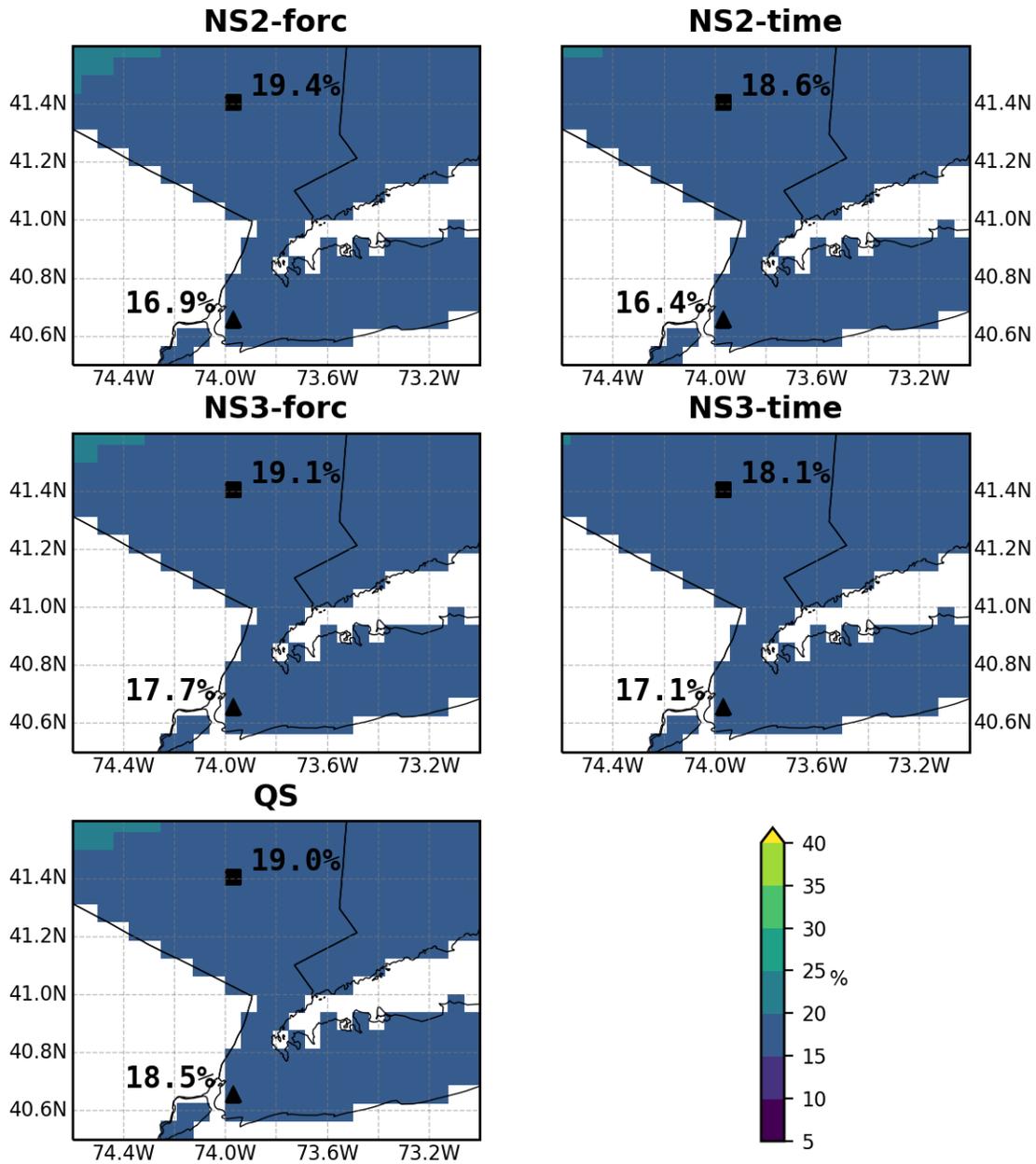


FIGURE 8: Spatial map zoomed into southern New York for LOCA 32-model average percent changes for 2075 versus the 1975 reference period for 2-year 1-day estimates under the RCP8.5 emission scenario. The triangle presents the percent change for Brooklyn, NY. The square indicates the percent change for a location 50 miles (~80 km) north of Brooklyn.

RCP 8.5, 100-year 1-day

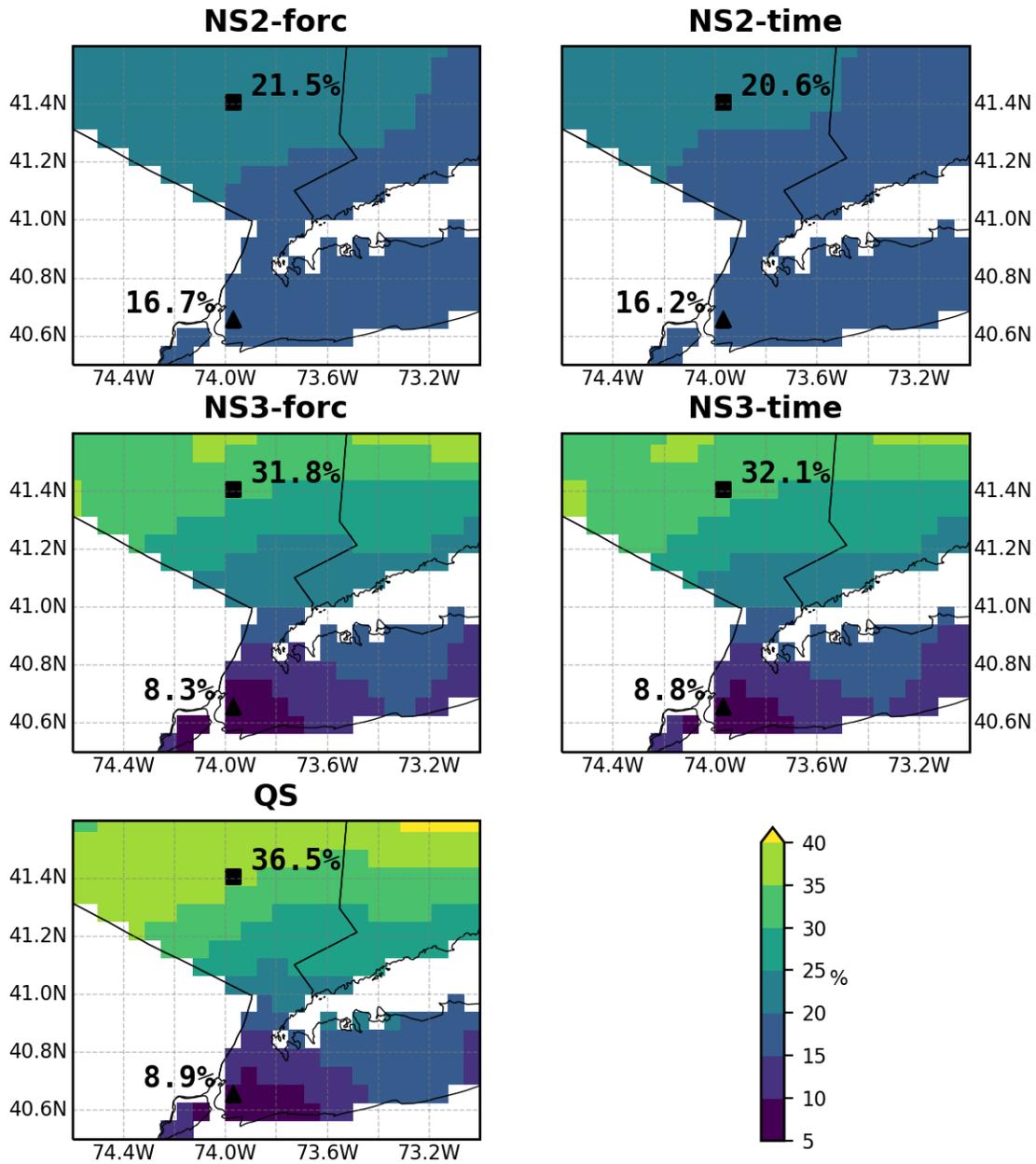


FIGURE 9: Spatial map zoomed into southern New York for LOCA 32-model average percent changes for 2075 versus the 1975 reference period for 100-year 1-day estimates under the RCP8.5 emission scenario. The triangle presents the percent change for Brooklyn, NY. The square indicates the percent change for a location 50 miles (~80 km) north of Brooklyn.

5. Conclusions and recommendations

This section summarizes the key findings from the two-phase project and assesses the added value of the newly proposed methodology with respect to traditional NOAA Atlas 14 estimates. Furthermore, it makes specific recommendations regarding the applicability of the selected methods and techniques relative to their use in future NOAA precipitation frequency analysis projects.

We have found that the local maximum likelihood approach, proposed by the researchers in Phase I of this study, is an adequate replacement for the current NOAA Atlas 14 regional L-moment statistics stationary methodology. This local (regional) maximum likelihood approach calculates the precipitation frequency estimates at grid locations by pooling information from nearby gauge locations and weighting the extreme events using a kernel function and a bandwidth. This method, when applied to the historical observations for the current period, compares well to the current NOAA Atlas 14 estimates and is recommended for national implementation. We further propose to account for the trend in historical observations, to model for the short-term nonstationary temporal changes, and integrate the terrain information.

The proposed methodology eliminates the need for trend detection in order to apply the stationary or nonstationary methodology. The proposed method can adapt to trends in data when they exist (positive or negative) or give a nearly stationary fit when a trend does not exist. If the trend does not exist, the nonstationary portion would be almost zero, providing nearly stationary parameter values. The trend in data can be informed using covariates such as radiative forcing, annual temperature or time. Section 5 of Appendix A-1 also explores using the spline function (with different levels of smoothness) to accommodate data trends without assuming a particular functional form. This flexibility is one of the reasons why we recommend this approach for national application, given that it can easily be applied to different climate regions.

This Phase II study's results show that climate models can be useful tools to account for the changes in future climate, but that using the model-based precipitation frequency estimates directly may not be adequate for design as absolute values, particularly for the wide range of NA14 design applications. The downscaled climate models still struggle to accurately capture current precipitation variability and magnitude on the gridded and regional scale, and depending on absolute values, these models can make the estimates unreliable. Instead, the downscaled datasets can be applied as a relative change between the present and future precipitation frequency estimates (quantile delta change method), which agrees with the results found in the literature. The key advantage of applying the quantile delta change method to calculate the rate of increase of future estimates is that the inherited bias of the climate model data is reduced when applied as ratios. These rate of increase of the future estimates can be applied along with the current (reference) period, which is based on more reliable spatial and temporal characteristics of the historical observations.

This study tested the proposed nonstationary method using historical extreme time-series data and extrapolating it beyond the observation period using the relationship between covariates

(e.g., ensemble average annual precipitation) and extreme historical time-series data. Unfortunately, this approach has not yielded robust results using tested covariates, particularly for hourly durations. Instead, we applied the local maximum likelihood method to calculate the future estimates needed for the calculation of the rate of future increase on the downscaled climate-model extreme time series data only. We applied this methodology in its stationary and nonstationary frameworks.

In the stationary framework, future estimates are calculated on different periods using stationary assumptions, resulting in what we have referred to throughout the report as quasistationary future estimates (QS). In its nonstationary form (NS), future estimates are also calculated on the extreme time series, where covariates capture trends in data. The QS and NS methods perform comparably when the NS method is applied to the climate-based AMS data for the entire record and when all three distribution parameters vary with the covariates. However, the NS method offers greater flexibility than the QS model because it allows for the shape parameter to be calculated over a longer record, and is capable of modeling the temporal nonstationary trend within the data. This allows more stable results, particularly on the spatial scale, which are needed for the engineering application.

This study further evaluated several covariates when the NS model is applied to the entire 1950–2099 period using the model-based extreme time series data. We found that radiative forcing or simple “year” covariates perform similarly. That said, the choice of covariate requires further investigation, mainly using the hourly data and analysis performed on several different regions of the United States. As part of our future work, we also plan to investigate further the climate models’ application to hourly durations and higher ARI (>100-year). At the time of implementation, the final estimates for the hourly durations and higher ARIs may need to rely on some type of adjustment procedure or extension of the percent change for the lower ARI to higher ARIs or any combination thereof. Similar approach is recommended by other studies in the absence of better alternatives (e.g., Kilgore et al., 2016, Mid-Atlantic Regional Integrated Sciences and Assessments, 2021; Requena et al., 2021).

The local maximum likelihood method relies on maximum likelihood parameterization. The shape parameter for this type of parameterization requires the shape parameter to be constrained due to its uncertainty. We extend the recommendation from the Penn State research team to limit the shape parameter in the GMLE approach, with the beta distribution of penalized coefficients ranging between -0.5 and 0.5 , for the NOAA Atlas 14 application. We further adopt the recommendation to use approximately two exceedances in any given year for the PDS for different regions across the continental United States and decluster both 1-day and 1-hour PDS using a 1-day separation period.

The triweight kernel function weighting scheme performs well as a regional and interpolation technique in local likelihood stationary and nonstationary frameworks. This technique calculates regions automatically based on the choice of spatial covariates and maximum allowable distance, while allowing for the easy implementation of any manual adjustments (adding pseudo-stations or specific extreme events) based on expert opinions or additional information provided in scientific meteorological documents. In this study, we used the maximum allowable distance of 50 miles without other spatial covariates. Further analysis is needed to evaluate the

choice of spatial covariates (e.g., mean annual precipitation, mean annual maximum precipitation, distance from the coast, etc.) and the maximum allowable distance value applicable for all regions across the United States. It is important to note that we have found that implementing terrain information to calculate the rate of future increases is not justifiable due to the low skill of the climate models to predict extreme rainfall at this spatial scale. Instead, terrain information should be used to calculate the current estimates based on historical observations.

Four downscaled climate datasets were evaluated in this study. The BCCAv2 dataset was not retained in this study due to low performance in comparison with observed data for the historic period. Among the three retained datasets, the University of Wisconsin-Madison, Center for Climatic Research UWPD (Dr. David Lorenz, specifically designed for this study) performed better than the University of California LOCA and NCAR NA-CORDEX in terms of hindcast accuracy, but this dataset did not significantly outperform the other datasets considered.

Based on the range of results developed through this study, the variability of future projections among the three retained datasets should be regarded as significant. The future regional projected changes compare closely at a 2-year ARI for the three retained datasets and the emission scenarios considered. The variability between the three retained datasets increases with the increase in ARIs and decrease in duration. The variability between datasets is even more pronounced spatially and often shows spatial inconsistency in the rate of increase between return periods and scenarios within one dataset and one location. The datasets' variability results in a high level of uncertainty regarding the future projected changes and, as such, needs to be considered when applied in an engineering design. Given that none of the downscaled climate models outperformed the others in this study, the use of a multi-model approach may better characterize the uncertainty associated with climate model predictions.

The use of the multi-model approach, furthermore, may provide a useful tool, given that climate models, scenarios, and downscaling techniques are continuously evolving. When this study commenced in 2016, the World Climate Research Programme's Phase 5 (CMIP5) climate model was the most up-to-date climate prediction model. At this study's completion in 2020, the next phase of climate model data (CMIP6) has become available. With the updated climate model data, the current downscaled models will also need to be updated, and new downscaled models may be developed. Thus, the choice of downscaled climate model dataset application in the NOAA Atlas 14 would depend on the availability and quality of downscaled models at the time of implementation. Given that these models, scenarios, and downscaling techniques are continually improving, developing the range of results based on several downscaled model datasets may better represent the models' prediction capabilities.

The following represent the key recommendations based on this study's results:

- 1) This proposed local maximum likelihood approach is suitable for national implementation and can streamline the production of estimates for the continental United States and affiliated territories. It is recommended to replace the current methodology based on regional L-moment statistics.
- 2) We recommend updating the current ("reference") NOAA Atlas 14 estimates using the nonstationary local likelihood approach for the base period with historical observations,

accounting short-term nonstationary temporal changes and integrated terrain information.

- 3) We recommend applying the quantile delta change method to calculate the future rate of increase and using it in conjunction with the reference estimates based on the historical observations.
- 4) The nonstationary local maximum likelihood approach methods applied to the climate-based AMS data for the entire record can be used to calculate the future rate of increase.
- 5) Among the three retained datasets, the University of Wisconsin-Madison, Center for Climatic Research UWPD (Dr. David Lorenz; specifically designed for this study) performed better than LOCA and NA-CORDEX at hindcast accuracy, but this dataset did not significantly outperform the other datasets considered.
- 6) The variability between the datasets is considered significant. Since none of the models outperform the others, a multi-model approach may better characterize climate data variability.
- 7) The climate models' application to hourly durations and higher ARIs (>100 year) requires further investigation.
- 8) Integrating the improved climate models into the process of calculating future estimates should be an integral part of the NOAA Atlas 14 development and infrastructure design requirements. The estimates and the rate of future increase should be updated at regular 10 year intervals.

References

- Angel, J. and M. Markus, 2020. Frequency Distributions of Heavy Precipitation in Illinois: Updated Bulletin 75. Illinois State Water Survey Contract Report 2020-05, Champaign, IL.
- Agilan, V., and N. V. Umamahesh, 2017. What Are the Best Covariates for Developing Non-Stationary Rainfall Intensity-Duration-Frequency Relationship? *Advances in Water Resources*, Vol. 101, pp. 11–22.
- 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>
- Coles, S., 2001. An introduction to statistical modeling of extreme values. Springer Series in Statistics. Springer-Verlag London, Ltd., London.
- Coles, S. G. and M. J. Dixon, 1999. Likelihood-based inference for extreme value models. *Extremes* 2(1), 5–23.
- Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris, 2002. A Knowledge Based Approach to the Statistical Mapping of Climate. *Climate Research* 23
- El Adlouni, S., T. B. M. J. Ouarda, X. Zhang, R. Roy, and B. Bobée, 2007. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, 43, W03410, doi: 10.1029/2005WR004545.
- DeGaetano, A.T., C.M. Castellano, 2017. Future projections of extreme precipitation intensity-duration frequency curves for climate adaptation planning in New York State. *Climate Services* 5 23–35. <http://dx.doi.org/10.1016/j.cliser.2017.03.003>.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E., 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.*, 9, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>.
- Hosking, J. R., and J. R. Wallis, 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge: Cambridge University Press.
- Kilgore, R., G. R. Herrmann, W. O. Thomas, Jr., and D. B. Thompson, 2016. Highways in the River Environment – Floodplains, Extreme Events, Risk, and Resilience. Hydraulic Engineering Circular No. 17 (HEC-17), Federal Highway Administration, FHWA-HIF-16-018.)
- Langbein, W.B., 1949. Annual floods and the partial-duration flood series. *Transactions, American Geophysical Union*, 30(6), 879–881, doi: 10.1029/TR030i006p00879.
- Lorenz, D. J., 2015: Downscaled climate projections. [Available online at <http://djlorenz.github.io/downscaling2/main.html>.]

Madsen, H., Pearson, C. P., and Rosbjerg, D., 1997. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events regional modeling. *Water Resour. Res.*, 33, 759–769.

Maraun, D., Wetterball, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Theme, M., 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.*, 48, RG3003.

Martins, E. S., and J. R. Stedinger, 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.*, 36(3), 737–744, doi: 10.1029/1999WR900330.

Mid-Atlantic Regional Integrated Sciences and Assessments, 2021. Projected Intensity-Duration-Frequency (IDF) Curve Data Tool for the Chesapeake Bay Watershed and Virginia, undated. As of September 22, 2021: <https://midatlantic-idf.rcc-acis.org>

Nielsen-Gammon, John W., 2020. Observation-Based Estimates of Present-Day and Future Climate Change Impacts on Heavy Rainfall in Harris County. Available electronically from <https://hdl.handle.net/1969.1/188617>.

Notaro, M., D. J. Lorenz, D. Vimont, S. Vavrus, C. Kucharik, and K. Franz, 2011: 21st century Wisconsin snow projections based on an operational snow model driven by statistically downscaled climate data. *Int. J. Climatol.*, 31, 1615–1633, doi:10.1002/joc.2179...

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, *Geosci. Model Dev.*, 9, 3461–3482, <https://doi.org/10.5194/gmd-9-3461-2016>, 2016

Pierce, David W., Daniel R. Cayan, and Bridget L. Thrasher, 2014. Statistical Downscaling Using Localized Constructed Analogs (LOCA). *Journal of Hydrometeorology*, Vol. 15, No. 6, pp. 2558–2585.

Perica, S., Pavlovic, S., Laurent, M. S., Trypaluk, C., Unruh, D., and Wilhite, O., 2015. Precipitation-frequency atlas of the United States. Version 2.0: Texas. National Weather Service, Silver Spring, MD, USA.

Perica, S., Pavlovic, S., Laurent, M. S., Trypaluk, C., Unruh, D., Martin, D., and Wilhite, O., 2018. Precipitation-frequency atlas of the United States. Version 3.0: Northeastern States. National Weather Service, Silver Spring, MD, USA.

Ragno, E., AghaKouchak, A., Love, C. A., Cheng, L., Vahedifard, F., & Lima, C. H. R., 2018. Quantifying changes in future Intensity-Duration-Frequency curves using multimodel ensemble simulations. *Water Resources Research*, 54, 1751– 1764. <https://doi.org/10.1002/2017WR021975>

Requena, Ana & Burn, Donald & Coulibaly, Paulin. 2021. Technical guidelines for future intensity–duration–frequency curve estimation in Canada. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*. 46. 87-104. 10.1080/07011784.2021.1909501.

Salas, J. D., J. Obeysekera, and R. M. Vogel., 2018. Techniques for assessing water infrastructure for nonstationary extreme events: A review. *Hydrological Sciences Journal* 63(3)325–352, DOI: 10.1080/02626667.2018.1426858.

The North American CORDEX Program—Regional Climate Change Scenario Data and Guidance for North America, for Use in Impacts, Decision-Making, and Climate Science. Available online: <https://na-cordex.org/> (accessed on 17 June 2019)

USGCRP, 2017. Climate Science Special Report: Fourth National Climate Assessment. Volume I [Wuebbles, D.J., D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C. Stewart, and T.K. Maycock (Eds.)]. U.S. Global Change Research Program, Washington, DC, USA, 470 pp.

USGCRP, 2018. Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II [Reidmiller, D.R., C.W. Avery, D.R. Easterling, K.E. Kunkel, K.L.M. Lewis, T.K. Maycock, and B.C. Stewart (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, 1515 pp. doi: 10.7930/NCA4.2018.

Wu, Shu, Momcilo Markus, David Lorenz, James R. Angel, and Kevin Grady, 2019. 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*, Vol. 11, No. 6, 2019, p. 1279.

Xavier, Ana Carolina Freitas et al., 2019. Selecting “the best” nonstationary Generalized Extreme Value (GEV) distribution: on the influence of different numbers of GEV-models. *Bragantia* [online]. 2019, v. 78, n. 4 [Accessed 27 August 2021], pp. 606-621. Available from: <<https://doi.org/10.1590/1678-4499.20180408>>. Epub 13 Dec 2019. ISSN 1678-4499. <https://doi.org/10.1590/1678-4499.20180408>.

Xuereb, Karin & Green, J., 2012. Defining Independence of Rainfall Events with a Partial Duration Series Approach. Proceedings of the 34th Hydrology and Water Resources Symposium, HWRS 2012.

Development and evaluation of a statistical model for
non-stationary precipitation frequency estimates with
NOAA Atlas 14

Dr. Benjamin Shaby and Gregory Bopp

Eberly College of Science, Statistics

The Pennsylvania State University

Dr. Alfonso Mejia and Sanjib Sharma

College of Engineering, Civil and Environmental Engineering

The Pennsylvania State University

June 2018

APPENDIX A-1. The Pennsylvania State University Final Report

Table of Contents

Executive Summary.....	A-1.4
1 Introduction	A-1.7
2 Data.....	A-1.8
3 At-station stationary analysis.....	A-1.9
3.1 Construction of partial duration series	A-1.9
3.1.1 Background	A-1.9
3.1.2 Data analysis	A-1.16
3.1.3 Summary and recommendations.....	A-1.22
3.2 Construction of annual maximum series	A-1.23
4 Comparison of estimation methods	A-1.25
4.1 Three commonly used estimation methods.....	A-1.26
4.1.1 Maximum likelihood	A-1.26
4.1.2 L-Moments.....	A-1.28
4.1.3 Generalized maximum likelihood	A-1.29
4.2 Estimator comparison	A-1.30
4.2.1 Simulation study	A-1.31
4.2.2 Precipitation fits.....	A-1.35
5 Modeling non-stationary extreme precipitation	A-1.38
5.1.1 AMS Model.....	A-1.39

APPENDIX A-1. The Pennsylvania State Final Report

5.1.2	PDS Model.....	A-1.39
5.1.3	Spline model.....	A-1.41
5.2	Tests for non-stationarity	A-1.42
5.3	Precipitation fits.....	A-1.44
6	Spatial Analysis.....	A-1.50
6.1	Local Likelihood Model	A-1.50
6.2	Precipitation fits.....	A-1.53
6.3	Incorporating climate model information	A-1.59
7	Summary and recommendations.....	A-1.63
8	Appendix	A-1.64
8.1	Quantile Scores	A-1.64
8.2	GMLE Penalty Function	A-1.64
8.3	Trend Detection Simulation Study	A-1.66
9	Acknowledgements.....	A-1.67
	References	A-1.68

APPENDIX A-1. The Pennsylvania State University Final Report

Executive Summary

Study Objectives

The Hydrometeorological Design Studies Center (HDSC) at NOAA's National Water Center (NWC) produces precipitation frequency estimates for the United States. These estimates are used in both private and public sectors for hydrological design and planning. Current NOAA Atlas 14 precipitation frequency estimates are based on the analysis of annual maximum series (AMS) data, use L-moments estimators for generalized extreme value (GEV) distribution parameters, and do not account for possible temporal non-stationarity. The primary objective of this study is to evaluate current NOAA Atlas 14 methods for modeling extreme precipitation and propose alternatives that account for possible non-stationarities. All data analysis for this pilot study was performed on daily and hourly precipitation data in the northeastern United States.

A commonly used alternative to AMS modeling of extreme precipitation is what is referred to as partial duration series (PDS) modeling. PDS approaches are capable of making more efficient use of data by considering more than a single observation per year. This study addresses two common subtasks involved in PDS construction: selection of a high threshold, exceedances of which approximately follow a generalized Pareto distribution (GPD), and appropriately handling weak temporal dependence, which can be broken by thinning dependent sequences. In addition, three different estimators, L-moments, maximum likelihood (MLE), and generalized maximum likelihood (GMLE) are compared for both AMS and PDS models and evaluated based on accuracy of return level estimation.

A collection of tests is used to detect temporal trends in (1) exceedance rates of a fixed, high threshold; (2) the distribution of PDS data; and (3) the distribution of AMS data. To accommodate historical trends in the distributions of extreme precipitation without assuming a particular functional form, a flexible penalized splines model is developed, wherein the distribution parameters of the GEV and GPD distributions vary smoothly in time.

Finally, a local likelihood approach for inferring the distribution of extreme precipitation at both observed and unobserved locations by pooling spatial information from nearby rain gauge locations is presented for both AMS and PDS data. The local likelihood model is capable of accommodating additional covariates that may explain differences in the distribution extreme precipitation. An example

APPENDIX A-1. The Pennsylvania State Final Report

of how climate model information can be incorporated to make forecasts of extreme precipitation in the presence of a non-stationary climate is also described.

Main Findings

A battery of threshold selection methods (mean excess plot, parameter stability plots, QQ-plots) for PDS construction were applied to daily and hourly data throughout the pilot region. A threshold fixed at the 0.98 quantile of non-zero precipitation was found to work well for most locations in this region. Run-length declustering was used to thin dependent sequences of threshold exceedances. This approach requires the selection of a run-length tuning parameter, which was chosen using out-of-sample quantile scores. For both daily and hourly data a short run-length of $r = 1$ gave the best out-of-sample quantile scores.

The popular L-moments estimator imposes an implicit constraint on the shape parameter of the GEV and GPD distributions ($\xi < 1$), which is in concordance with the typical range of inferred shape parameters for precipitation data. This constraint makes the L-moments estimator robust to small sample sizes. The MLE does not impose such a constraint, making it more sensitive to sample size. Maximum likelihood estimation, however, can easily incorporate covariates, which easily enables non-stationary distribution modeling. Generalized maximum likelihood estimation allows the modeler to constrain the inferred shape parameter to a reasonable range, while also making it possible to easily incorporate covariates. A simulation study was performed to demonstrate how these three estimators perform under varying sample sizes and shape parameters. It is shown that the GMLE is competitive with the L-moments approach for both GEV and GPD distributions, even under small sample sizes. All three estimators give relatively similar fits for stationary models applied separately to each station in the pilot region.

The tests for trends in extremes give regionally more consistent results when likelihoods are locally weighted than when considered at each station separately. Tests, both for linear changes in the location parameter of the GEV distribution and rate of exceedances of a fixed high threshold, show evidence for positive trends at the majority of stations in the pilot region. Three spline models of increasing complexity are compared for both GEV and GPD models for accommodating non-stationarities in the distribution of extremes with: smoothly varying (1) location parameter; (2) location and scale; and (3) location, scale, and shape parameters. The fitted return level estimates from models (1) and (2) capture the slight increasing trends in the data. Due to the difficulty in estimating the shape parameter under

APPENDIX A-1. The Pennsylvania State Final Report

relatively short historical records, it is advocated that the shape parameter be fixed in time rather than allow it to vary smoothly in time as in (3).

To infer the distribution of extreme precipitation at both observed and unobserved locations, a local likelihood model is employed that weights likelihood contributions of precipitation from nearby gauge locations. Variations of this approach apply equally well for modeling AMS and PDS data. A bandwidth tuning parameter that determines the range of the local neighborhoods is chosen using out-of-sample quantile scores. The fitted local likelihood models allow for the location and scale parameters of the GEV and GPD distributions to vary as a function of the covariate mean annual precipitation (MAP), which is more easily inferred. Both the fitted GEV and GPD models show similar spatial patterns. However, the GEV model exhibits greater variability in return-level estimates than the GPD model, most likely due to the smaller sample size inherent in AMS data series. Finally, an illustration for how climate model information can be incorporated by allowing GEV and GPD parameters to depend on covariates in order to account for long-term temporal non-stationarity is presented.

Recommendations and Future Work

Because of its ability to account for non-stationarity and incorporate potentially informative covariates (e.g. MAP, elevation, etc.) that may improve inference on the distribution of extreme precipitation, while also constraining the shape parameter of the GEV and GPD distributions, the GMLE is preferred over the current usage of the L-moments estimators in this context. Increasing trends in extreme precipitation were detected for most of the pilot region. AMS and PDS methods are equivalently capable of modeling temporal non-stationarity, and the two methods performed similarly on the pilot dataset. Therefore, for estimating frequency maps, the choice between AMS and PDS might reasonably be dictated by workflow considerations. For PDF methods, the 0.98 empirical quantile was a good threshold on the pilot dataset. Furthermore, the PDS analysis was insensitive to choices of declustering technique, so the simple choice of run-length declustering with a default run length of 1 was deemed adequate. By allowing the parameters of the GEV and GPD distributions to depend on climate model output it is possible to make forecasts of extreme precipitation under different climate regimes. However, the choice of climate model output variables and the functional relationship between distributional parameters and climate model variables should be investigated further.

APPENDIX A-1. The Pennsylvania State Final Report

1 Introduction

The purpose of this study is to evaluate the efficacy of existing NOAA ATLAS 14 methods for modeling extreme precipitation and propose alternatives that account for possible non-stationarities. Current NOAA ATLAS 14 methods for modeling extremes are based on the analysis of annual maxima series (AMS), which are constructed by extracting the maximum precipitation amount observed in disjoint annual blocks of time. Using AMS can be a wasteful use of data, however, when the data are available at finer granularity (daily, e.g.), as a considerable amount of data is discarded. An alternative approach to characterizing extremes that is capable of detecting fine-scale temporal features is through what is called the partial duration series (PDS), which consists of precipitation excesses of some high threshold (2 inches in per day, e.g.). We propose modeling possible time non-stationarities using a flexible splines approach fit using maximum likelihood estimation (MLE) and penalized maximum likelihood estimation, also referred to in the hydrological literature as generalized maximum likelihood estimation (GMLE). The currently used methodology is based on L-moments (LM) estimators, which are moment-based estimators, which are not easily adapted to account for non-stationarity. In the sequel, we will examine the differences between these estimation methods, which will be done in stages. First, we compare AMS and PDS methods under the assumption of stationarity and assess them based on accuracy of predicted return levels, where an N-year return level is the precipitation amount that is exceeded on average once every N years (i.e. the annual exceedance probability (AEP) of an N-year return level is $1 - \frac{1}{N}$). The expected time between consecutive exceedances of a given return level is referred to as the return period. In hydrology the return-period is sometimes referred to as the average recurrence interval (ARI), but we will use return-period throughout. We evaluate the differences among the different estimators (LM, MLE, GMLE) based on prediction error and robustness to sample size. Finally, we explore a method for modeling non-stationary long-term trends among extremes by modeling the parameters of the

APPENDIX A-1. The Pennsylvania State Final Report

distributions for AMS and PDS as smoothly varying functions of time and present a model that incorporates spatial information to reduce variability among pointwise return-level estimates.

2 Data

The data for this pilot study consist of observations from rain gauge stations located in the northeastern United States, including Maine, New Hampshire, Vermont, New York, Massachusetts, Rhode Island, Connecticut, New Jersey, and northern Pennsylvania used in NOAA Atlas 14 Volume 10. The majority of the region is heavily forested, with topography ranging from Appalachian highlands to low-lying coastal plains. Daily and hourly accumulation observations were made at 851 and 274 stations respectively. The data were obtained from the National Centers for Environmental Information (NCEI), which is maintained by NOAA. For each station, years with more than 20% missing observations were filtered out. Further, only stations with at least 30 years of data for the hourly analysis, and 50 years of data for the daily analysis, were used to ensure that record lengths were long enough to reliably estimate distributional parameters. Precipitation records range between as early as 1900 and run up to 2014. After applying these filters, 150 stations were used for the hourly analysis and 643 stations were used for the daily analysis. The locations of these stations are given in Figure 2-1. The stations are fairly uniformly dispersed throughout the region, except in Maine, where they are less dense. Because extreme hourly and daily accumulations may pose different hazards, the record length and station density differ for daily and hourly resolutions, and inference made at the daily scale does not apply to the hourly, separate analyses were conducted for each duration. The hourly data were preprocessed such that only the maximum hourly accumulation was retained for each day. After preprocessing, both the daily and hourly data consist of at most one observation per station per day.

APPENDIX A-1. The Pennsylvania State Final Report

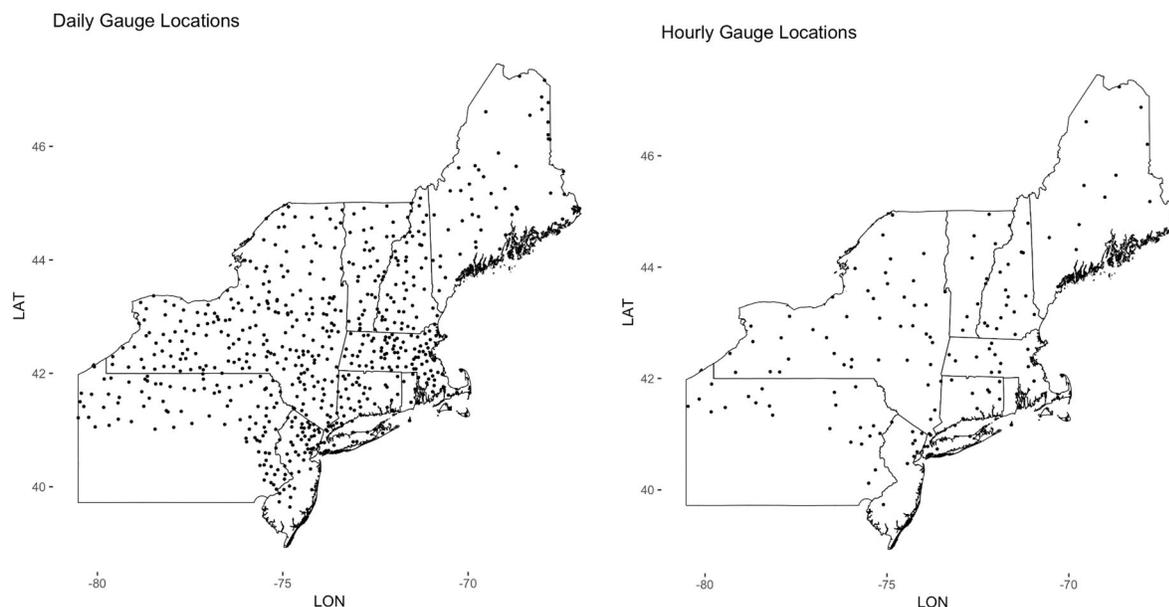


Figure 2-1: Left: 643 daily precipitation gauge locations. Right: 150 hourly accumulation precipitation gauge locations.

3 At-station stationary analysis

3.1 Construction of partial duration series

3.1.1 Background

The aim of extreme precipitation analysis is to describe the expected frequency of rare, heavy rainfall events. To characterize extreme rainfall, a natural approach is to fix a high precipitation threshold, say 2 inches per day, and evaluate the distribution of exceedances above that threshold. This method is what hydrologists refer to as a partial duration series (PDS) analysis, or what extreme value statisticians call the peaks over threshold (POT) method. Under relatively weak assumptions about the root distribution of precipitation, as the threshold increases towards the upper bound of the root distribution, the distribution of the exceedances of that threshold will become more and more similar to what is known as a generalized Pareto distribution (GPD). This asymptotic result is often used as justification for using the GPD as an approximate distribution for exceedances of finite thresholds. This distribution is different

APPENDIX A-1. The Pennsylvania State Final Report

from the limiting distribution for maxima, the generalized extreme value (GEV) distribution, which will be described in Section 3.2. For a sufficiently high threshold u and $Y = \max(X - u, 0)$, the conditional distribution of $Y | X > u$ is approximately generalized Pareto distributed. The GPD location-scale family distribution function is given by

$$F_{\xi}(z) = \begin{cases} 1 - \left(1 + \xi \left(\frac{z - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}, & \text{for } \xi \neq 0 \\ 1 - \exp\left(-\frac{z - \mu}{\sigma}\right), & \text{for } \xi = 0 \end{cases}$$

Equation 1: Generalized Pareto distribution function.

for $z \geq \mu$ if $\xi \geq 0$ and $\mu \leq z \leq -\frac{\sigma}{\xi}$ if $\xi < 0$, where $\mu \in R, \sigma > 0, \xi \in R$. The parameters μ, σ, ξ are referred to as the location, scale, and shape parameters. Considering the support of the distribution, when modeling exceedances and $Y = \max(X - u, 0)$ of a fixed threshold u , the location parameter μ is typically fixed at 0, and so we denote this family by $GPD(\sigma, \xi)$. The tail of the distribution is controlled by the shape parameter ξ . When $\xi > 0$, the distribution has a heavy upper tail. For $\xi = 0$, the upper tail decays exponentially, and when $\xi < 0$, the support of the distribution is bounded above. See Figure 3-1 for an example of a PDS of daily precipitation observed in Amherst, MA, between 2007 and 2014 using a threshold of 2.1 inches/day.

APPENDIX A-1. The Pennsylvania State Final Report

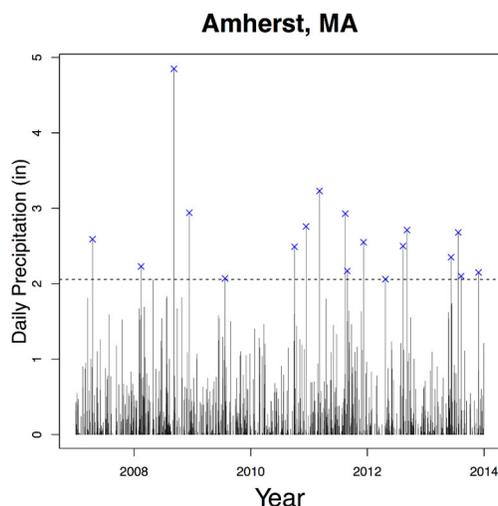


Figure 3-1: Illustration of a partial duration series (PDS) for a gauge location in Amherst, MA between the years 2007 and 2014. In this example, the threshold is set at 2.1 inches of rain per day (dashed horizontal line). Points of the PDS are designated with blue X marks.

Threshold Selection

The quality of the GPD approximation depends on the threshold that is chosen. If too low of a threshold is selected, the approximation may be poor, and estimates can be biased by the bulk of the distribution. On the other hand, if too high of a threshold is selected, it will leave little information with which to make inference about the tail distribution, and the estimator variance will be large. Finding a threshold that is sufficiently high is not always an easy task. This task is comparable to choosing a large enough block size when using a GEV approximation. Large block sizes will improve the GEV approximation, but shrink the total number of blocks, increasing variance when performing inference. There are often yearly cycles and seasonal patterns in the data that provide convenient block size options. However, such choices do not have natural analogues when considering threshold selection. In practice, one typically takes a threshold that is as low as possible but that does not invalidate the limiting GPD approximation.

Threshold selection is a classic problem in extreme value modeling that has received considerable attention. See Caeiro and Gomes (2016) and Scarrott and MacDonald (2012) for recent

APPENDIX A-1. The Pennsylvania State Final Report

reviews. Techniques for threshold selection can be broadly categorized into graphical, resampling, and model-based approaches (Scarrott and MacDonald, 2012). Model-based approaches typically require a model for both the bulk and the tail of the distribution, and since our focus here will only be on specifying a tail model, we will not review them here. The first re-sampling approach to threshold selection was proposed by Hall (1990), and is related to the Hill estimator, an estimator of the shape parameter based on differences of logs of order statistics (Hill, 1975). The Hall re-sampling method aims to select a threshold such that the corresponding mean squared error (MSE) for the Hill estimator is minimized. However, this method has been shown to be sensitive to initial conditions of the estimation procedure (Gomes and Oliveira, 2001), severely limiting its utility.

Graphical methods take advantage of the GPD's threshold stability property, which means that if exceedances of a threshold u are GPD, then exceedances of any higher threshold $u' > u$ will also be GPD. The threshold stability property is used as justification for a variety of graphical and heuristic methods. Graphical methods typically require subjective judgment on the part of the practitioner, which can become impractical if one needs to select many thresholds for different models. The three most common graphical methods, namely the mean residual life, parameter threshold stability, and graphical goodness-of-fit plots, are reviewed in Coles (2001). We briefly describe these three graphical methods next.

The mean residual life (MRL) (Yang, 1978; Hall and Wellner, 1981) measure quantifies the expected excess of a threshold u given that the random variable X is greater than u . For the GPD, the MRL is linear as a function of the threshold, which gives a natural way to choose a threshold by simply taking the first point after which the graph $E(X - u | X > u)$ versus u becomes linear. As the threshold increases from u to u' , the scale parameter for the GPD increases by $\xi(u' - u)$, and so it can be shown (Coles, 2001) that for thresholds $u' > u$,

APPENDIX A-1. The Pennsylvania State Final Report

$$E(X - u | X > u') = \frac{\sigma_{u'}}{1 - \xi} = \frac{\sigma_u + \xi u'}{1 - \xi}$$

where σ_u and $\sigma_{u'}$ are the shape parameters for GPDs using threshold u and u' respectively. Therefore, if X follows a GPD, the expectation of exceedances of u' is a linear function of u' for all $u' > u$. This method can be difficult to use in practice, as the graphs may not have a clear change-point. For high thresholds, only a small fraction of the data is used to estimate the mean excess quantity, which leads to higher variability, making it more difficult to assess that the linearity property is satisfied. The MRL plot rarely gives conclusive evidence of a good threshold when used alone but can be useful when considered in tandem with other methods. The Hill plot proposed by Drees et al. (2000) based on the classic tail index estimator of Hill (1975) is essentially a MRL plot on log-transformed observations and possesses its same drawbacks.

The threshold stability property also applies to the parameters of the GPD, a feature which can be exploited to find a reasonable threshold. The GPD is indexed by scale and shape parameters. The scale parameter grows linearly as the threshold increases according to $\sigma_{u'} = \xi(u' - u)$, whereas the shape parameter remains constant over increasing thresholds. Analogous to the MRL method, by plotting the estimates of the parameters against a sequence of thresholds, the modeler aims to find the smallest threshold after which the scale and shape parameter estimates remain roughly linear and constant, respectively. Just as was the case for the MRL plots, the parameter estimates become increasingly variable with higher thresholds, making the assessment of linearity difficult in the upper tail.

Graphical goodness-of-fit plots (e.g. qq-plots, return-level plots – see Coles, 2001), to assess how well the GPD fits for varying thresholds, are also commonly employed. Formal goodness-of-fit tests can also be used to assess the fit of a GPD to the observed exceedances conditional on a selected threshold (Bader et al., 2016). For example, one may test whether it is plausible that a sample of exceedances

APPENDIX A-1. The Pennsylvania State Final Report

follow a GPD over a range of thresholds and take the smallest such threshold that a test does not reject the plausibility of this assumption.

Declustering

Meteorological phenomena are intrinsically dynamic processes with non-trivial dependence characteristics. However, much of the classical theory for POT modeling applies to independent observations. One requirement for the limiting distribution of threshold exceedances to hold is that the process not exhibit strong long-range dependence. For example, the amount of rain that falls 1 year from now should not be strongly dependent upon how much rain fell today. This assumption seems reasonable for precipitation, however, there is some short-range (day-to-day) temporal dependence to precipitation that must be dealt with. Rather than model the dependence explicitly, for example with an explicit time series model (e.g. Bortot and Gaetan (2016); Reich et al. (2014)), one can reconcile the independence assumption by filtering dependent clusters of threshold exceedances, a process referred to as declustering. Again, there is a balance to achieve: one would like to discard as little data as possible, while sufficiently satisfying the independence assumption. Ignoring dependence can lead to an under estimation of uncertainty. In this section, we review the common methods for threshold selection and declustering and apply them to hourly and daily precipitation data. We use quantile scores (Gneiting and Raftery, 2007) to choose a run-length parameter for filtering clusters of threshold exceedances, and show that for precipitation data in the pilot study region, declustering has minimal effect on estimation. In order to apply models that assume independence, it is necessary to remove the dependence in the data by declustering. The runs declustering method (Davison and Smith, 1990) removes dependence by thinning sequences of consecutive exceedances of a fixed threshold. The declustering method works as follows: for a fixed threshold u , a cluster is defined to be the sequence of observations starting with the

APPENDIX A-1. The Pennsylvania State Final Report

first excess after a sequence of non-excesses and terminating with the last excess before a sequence of r non-excesses. For example, in a toy dataset plotted in Figure 3-2, if we take $r = 3$ and set a threshold of $u = 1.6$, the four excesses colored in red would belong to the same cluster. The next step is to thin the data by clusters. Within each cluster of excesses, only the maximum is retained, while the remaining excesses are discarded. Finally, the remaining excesses within each cluster are taken to be approximately independent and GPD. The aim is to choose r large enough such that the independence assumption is reasonable, while not so large as to discard too much data by grouping too many consecutive days together within the same cluster.

Ferro and Segers (2003) propose another method for declustering that incorporates additional information about time between threshold exceedances to account for dependence. Both of these methods, runs declustering and Ferro and Segers (2003), require that the modeler choose a tuning parameter to perform declustering. For some applications, the choice of this tuning parameter can influence the resulting inference. Because dependence typically weakens at higher values, the choice of tuning parameter may also be dependent upon the choice of threshold. In the data analysis section, we illustrate the use of the run-length declustering method for filtering precipitation data.

APPENDIX A-1. The Pennsylvania State Final Report

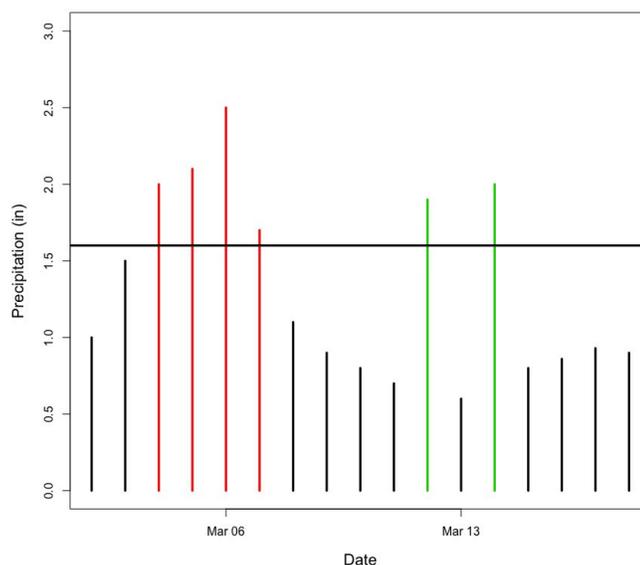


Figure 3-2: In this cluster identification example with run length $r=3$ and $u=1.6$ in, a sequence of 4 excesses are classified as belonging to a single cluster (red) and two exceedances separated by one non-exceedance are classified as belonging to another cluster (green).

3.1.2 Data analysis

In this section, we demonstrate the effects of threshold selection and declustering on model fit, and illustrate how to select a threshold for a given gauge location using the MRL, parameter stability, and goodness-of-fit plots. We illustrate the method on the PDS at two different stations, one made up of daily precipitation accumulations and one of hourly accumulations. Since the hourly accumulation data have already been preprocessed to consist of only the maximum observation from each day, declustering will have less of an effect in this case. The two examples shown here were selected to contrast when graphical methods strongly suggest a particular threshold with a more ambiguous case. The graphical methods when applied to the daily observation example do not have a clear change point from which a threshold can be selected, whereas the hourly example is more suggestive of a reasonable threshold. Although only two stations are presented here for illustration, our recommendations in Section 3.1.3 are based on an analysis of a larger collection of stations. In general, different thresholds

APPENDIX A-1. The Pennsylvania State Final Report

may be appropriate for different gauge locations, and so analysis should be applied separately for different stations. However, for the majority of stations, a threshold at the 0.98 quantile appeared adequate, and due to the large number of stations that would need to be considered, a fixed threshold at this quantile should be sufficient for obtaining reliable inference. Formal tests such as the Anderson-Darling test were considered, but have been omitted due to their poor performance as a result of sensitivity to data quantization (Deidda and Puliga, 2006), which is common in rain data. Finally, we compare various run-length declustering parameters on the basis of quantile scores (Gneiting and Raftery, 2007), which are proper scoring rules that provide a measure of discrepancy between the empirical and theoretical distributions. Lower quantile scores correspond to better distributional matches. Using this approach, we show that a declustering parameter of 1 day is sufficient in terms of out-of-sample prediction.

Daily Analysis

For this analysis, we consider observations consisting of daily precipitation accumulations. We illustrate the threshold selection methods for station 19-0120, located in Amherst, MA. The top left panel of Figure 3-3 gives an example of an MRL plot made using the extRemes R package (Gilleland and Katz, 2016), which consists of mean estimate (middle line) and 95% CI (outer bands). For the daily analysis, there is not a clear change point in the MRL plots from nonlinearity to linearity. Further into the tail, the estimate of the MRL becomes noisier due to decreasing sample size. The top right and bottom left panels illustrate parameter stability plots over a range of thresholds. The scale parameter here has been re-parameterized so that if exceedances of a given threshold are GPD, the new scale parameter will be constant over increasing thresholds. Due to the increasing variance in the parameter stability plots, it is difficult to decide at what point the parameters stabilize. One thing to note is that graphs of the scale and shape estimates will often show negative dependence (e.g. as the shape parameter decreases the scale will increase). Finally, we can assess the fit of the GPD by inspecting qq-plots (Figure 3-3: bottom

APPENDIX A-1. The Pennsylvania State Final Report

right panel) over a range of thresholds. We consider four fixed thresholds in the qq-plot: the 95th and 98th percentiles of the data when including zero precipitation observations (blue and green lines, respectively), and the 95th and 98th percentiles after excluding zero precipitation observations (red and black lines, respectively). The qq-plots show better fits for models fit using higher thresholds, especially at higher quantiles of the distribution. The red and black lines, corresponding to the 95th and 98th percentiles of non-zero observations, are both closer to the 45-degree line than the models fit using lower thresholds, which supports the choice of a threshold in those percentiles.

Next, we evaluate the effect of the run-length declustering parameter on model fit. Several run-length declustering parameters (1-5 days) are considered. For each run-length, quantile scores are computed over a range of high precipitation quantiles. Separate analyses are done over a range of different thresholds, to ensure that the declustering parameter does not depend on the selected threshold. For each threshold and run-length parameter, separate quantile scores are calculated for each station in a holdout sample. Figure 3-4 shows the average quantile scores over all stations for each run-length. The fits appear to be very similar for different declustering run-lengths, regardless of the threshold. Since there are no meaningful differences between run-lengths for hourly data, and so as to discard as little data as possible, the shortest run-length, 1 day in this case, is preferred. This conclusion may be different for other geographic regions. Even though the precipitation time series here exhibit weak temporal dependence at extreme levels, there is no particular reason to expect this to be the same elsewhere. This merely illustrates how one can use quantile scores to select a run-length declustering parameter.

Hourly Analysis

We perform the same analysis on the hourly maximum hourly precipitation accumulations. For the hourly accumulations, we will consider station 30-5801, located in New York City's Central Park. In

APPENDIX A-1. The Pennsylvania State Final Report

contrast to the daily analysis, the plots for this example are more suggestive of an adequate threshold. The estimated MRL shows nonlinearity over the set of thresholds considered, but when considering the graph above 0.7 in, the graph is more plausibly linear, however, it is difficult to select a threshold from the MRL plot alone because of the increasing uncertainty in assessing linearity of the MRL due to decreasing sample size. Estimates of both parameters appear to stabilize for thresholds above 0.7 in. Again, there is considerable variability to these plots, especially for higher thresholds, but when used in conjunction with the MRL plot, the parameter stability plot in this case suggests a consistent minimum threshold. Typically, individual graphical methods alone do not provide conclusive evidence of an optimal threshold, but when considered in tandem they can be more suggestive. For this station, a minimum threshold of at least 0.7 in appears adequate based on the MRL and parameter stability plots.

For hourly data, there is similar behavior among the quantile score curves for varying run-length declustering parameters as was seen for the daily accumulation analysis, and so the figures have been omitted.

APPENDIX A-1. The Pennsylvania State Final Report

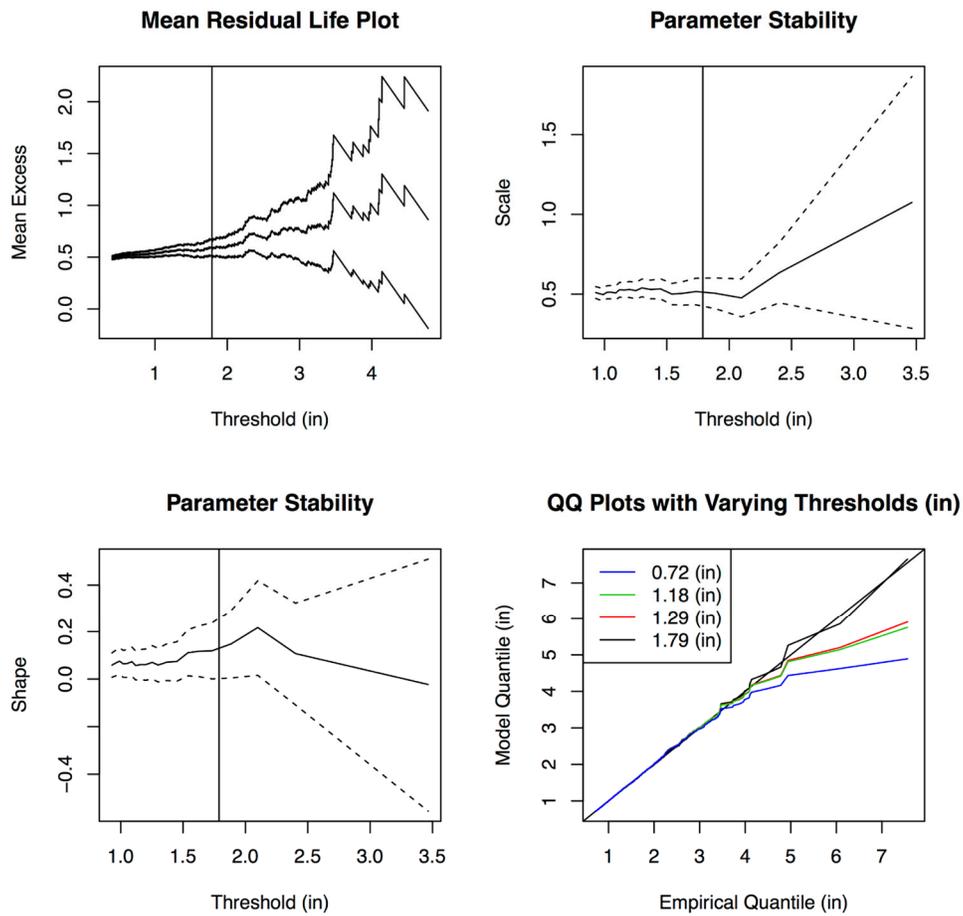


Figure 3-3: Sample daily accumulation mean residual life (MRL), parameter stability, and qq-plots (red: non-zero 95%, black: non-zero 98%, blue: with zeros 95%, green: with zeros 98%) for station 19-0120. For this sample location, the MRL and parameter stability plots give less indication of an appropriate threshold, but the qq-plots show better fit for thresholds in the range of the 95-98th percentiles of the empirical data than do the lower thresholds that were considered. Vertical lines indicate a point after which stability appears plausible by visual inspection (1.79 in).

APPENDIX A-1. The Pennsylvania State Final Report

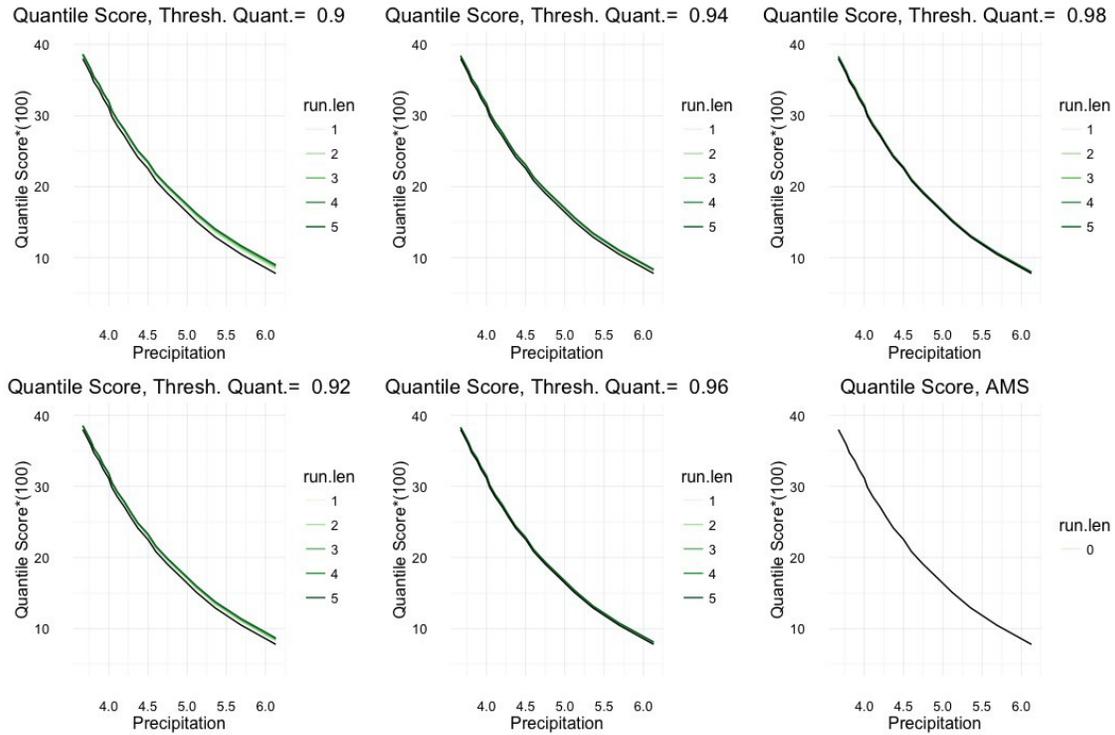


Figure 3-4: *Quantile scores for a range of run-length declustering parameters over a range of high daily accumulation precipitation quantiles. To ensure the effect of run-length does not depend on the selected threshold, the analysis is performed over a range of thresholds (panels). For reference, the quantile scores for an annual maximum series (AMS) fit are shown in the bottom right panel. Lower quantile scores are preferred. Because the quantile scores are comparable for all run-length parameters, regardless of which threshold was used, a smaller run-length is preferred to retain more of the data.*

APPENDIX A-1. The Pennsylvania State Final Report

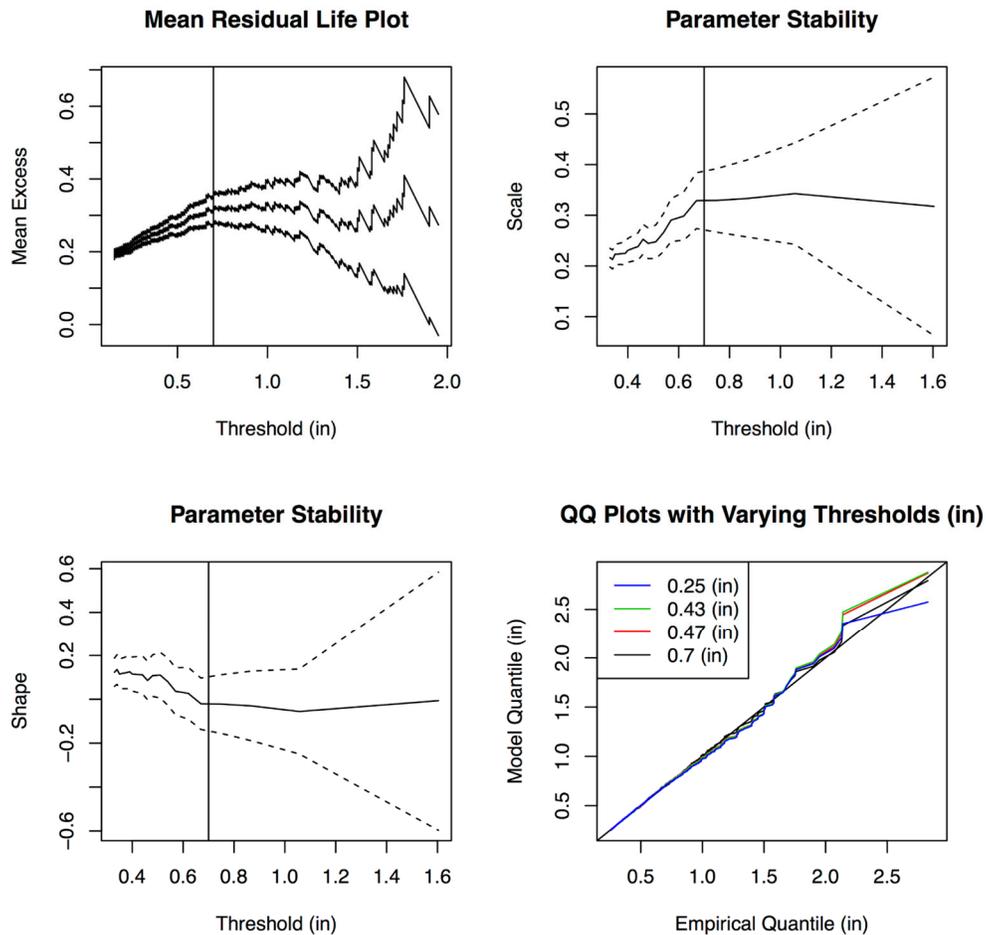


Figure 3-5: Sample hourly accumulation (in) mean residual life (MRL), parameter stability, and qq-plots (red: non-zero 95%, black: non-zero 98%, blue: with zeros 95%, green: with zeros 98%) for station 30-5801. The MRL plot becomes roughly linear and the parameter stability plots appear roughly constant after a threshold of 0.7 in. The qq-plot shows comparable model fit across all thresholds considered. Together these plots indicate that a threshold of at least 0.7 in would be adequate.

3.1.3 Summary and recommendations

A common approach to characterizing the distribution of extreme precipitation is to model exceedances of a high threshold. Asymptotic theory suggests the appropriate limiting distribution for threshold exceedances is the GPD. Approximating the distribution of threshold exceedances by this limiting distribution requires that the modeler choose an appropriate threshold. This choice is complicated by the increased variance due to sample sizes at high quantiles. Here, we illustrated how one can choose a

APPENDIX A-1. The Pennsylvania State Final Report

threshold using the mean residual life, parameter stability, and qq-plots for a single location. Based on the analysis of stations in the pilot region, thresholds in the range of the empirical 98th percentile of non-zero observations appeared most adequate for the majority of stations for both hourly and daily accumulations. In practice, because the appropriate threshold may be different for different regions, this analysis will need to be performed separately for different climatic regions.

To remove dependence in the data, declustering should be performed before model fitting. In this section, we illustrated a filtration method for removing temporal dependence called runs declustering. This method requires the modeler to select a tuning parameter that corresponds to the amount of dependent data that will be discarded. The goal is to discard as little data as possible, while sufficiently approximating the independence assumption. We illustrated how to choose this declustering parameter on the basis of quantile scores, which quantifies the correspondence between modeled and empirical quantiles. The amount of declustering applied had little effect on the resulting quantile scores in the pilot study region considered for both hourly and daily accumulation data, and so as to reduce the variance in the resulting inference, a short run-length declustering parameter ($r = 1$) is preferred for both resolutions.

3.2 Construction of annual maximum series

The AMS is constructed by partitioning a time series into disjoint annual blocks, and extracting the observed maximum precipitation from each year. A limit theorem from extreme value theory tells us that under relatively weak conditions, the distribution of the maximum $M_n = \max_{i=1, \dots, n} X_i$ of independent and identically distributed (i.i.d.) replicates X_1, X_2, \dots (daily rainfall, e.g.) from a common distribution will converge to a generalized extreme value (GEV) distribution after proper renormalization (technical detail that can be ignored for our purposes; see Resnick (1987) for details) as the number of replicates $n \rightarrow \infty$ (Fisher and Tippett (1928); Gnedenko (1943)). We can apply this limiting distribution

APPENDIX A-1. The Pennsylvania State Final Report

as an approximation for the distribution of maxima over finite block sizes ($n = 365 \text{ days}$, e.g.). As mentioned earlier, choosing a block size is analogous to choosing a threshold. The limit theorem applies as the block size goes to infinity, but we apply it as an approximation to finite sample sizes. Using annual blocks is more a matter of custom than anything else, however there are also typically annual cycles to natural phenomena like precipitation that make it a natural choice. The GEV distribution, like the GPD, is a three-parameter location-scale family with distribution function given by

$$G(z) = \begin{cases} \exp\left\{-\left(1 + \xi \left(\frac{z - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right)\right\}, & \text{for } \xi \neq 0 \\ \exp\left\{-e^{-\left(\frac{z - \mu}{\sigma}\right)}\right\}, & \text{for } \xi = 0 \end{cases}$$

Equation 2: Generalized extreme value distribution function.

where $\mu - \frac{\sigma}{\xi} \leq z < \infty$ when $\xi > 0$, $z \in R$ when $\xi = 0$, and $-\infty < z \leq \mu - \frac{\sigma}{\xi}$ when $\xi < 0$. The location and scale parameters are μ and σ , respectively, and we denote this family by $GEV(\mu, \sigma, \xi)$. Also, like the GPD, the tail of the GEV is controlled by a single parameter ξ . The three sub-families are bounded upper-tail, Weibull type ($\xi < 0$), light-tailed, Gumbel type ($\xi = 0$), and heavy-tailed, Fréchet type ($\xi > 0$). An example construction of an AMS observed in Amherst, MA between 2007 and 2014 is given in Figure 3-6.

APPENDIX A-1. The Pennsylvania State Final Report

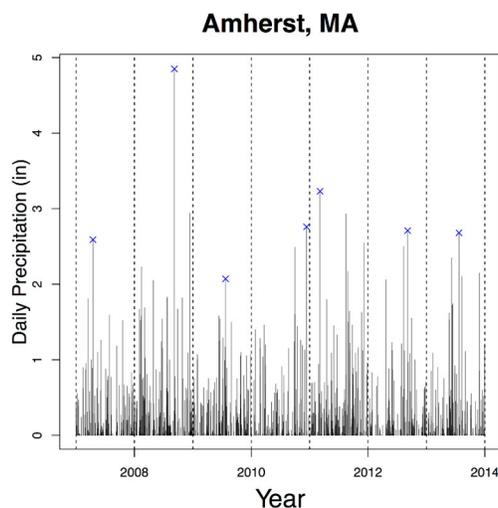


Figure 3-6: Example construction of an AMS in Amherst, MA for precipitation observed between 2007 and 2014. Annual maxima are designated by blue X marks. Vertical dashed lines delineate the years.

4 Comparison of estimation methods

If one is taking a parametric approach, as we are here, in order to make inference on quantities of practical value, such as the 100-year return level, distributional parameters (e.g. μ , σ , ξ above) must first be estimated. It is not uncommon for different estimation methods to give different results, and each has its own strengths and weaknesses. In this section, we will review and compare three common estimators for the GEV and GP distributions, the maximum likelihood estimator (MLE), linear combinations of order statistics (LM), and the generalized maximum likelihood estimator (GMLE). Broadly speaking, we would like an estimator to have the following properties: as the sample size increases the error of our estimator should decrease towards zero; if our estimator converges to the true parameter, estimators with smaller variance should be preferred; an estimator should be robust to small sample sizes (i.e. not vary wildly); finally, an estimator should have a known distribution (at least asymptotically) so that statements about uncertainty can be made. To assess these estimators, we perform a small simulation study, comparing the root mean squared error (RMSE) of predicted return

APPENDIX A-1. The Pennsylvania State Final Report

levels. We conclude the section by comparing the fits of each estimation method on the precipitation data.

4.1 Three commonly used estimation methods

4.1.1 Maximum likelihood

The method of maximum likelihood is one of the most well-known and widely used estimation methods in statistics. In order to define the MLE, we first have to introduce the likelihood function. The derivative of the distribution functions in Equations 1 and 2 are known as the GEV and GPD density functions, respectively, when viewed as a function of z . Denote one of these generically by $f(z; \theta)$, where $\theta = (\mu, \sigma, \xi)$. Assuming independence between observations, the product of densities, each evaluated at the observed data z_1, z_2, \dots, z_n , viewed now as a function of the parameters, rather than the data, is what is known as the likelihood function:

$$l(\theta; z_1, z_2, \dots, z_n) = \prod_{i=1}^n f(z_i; \theta).$$

Maximum likelihood, as the name suggests, is an estimation method whereby one estimates the parameter θ by the value which maximizes this function. We will denote the MLE by $\hat{\theta}_{MLE}$. The MLE can sometimes be found analytically, but, as is the case for the GEV and GPD, numeric optimization routines are typically required (e.g. Newton-Raphson). Note that for the GEV and GPD, boundary conditions must be accounted for when performing optimization (see the support of the GEV, e.g.). The MLE is also very easily adapted to more complex modeling situations, for example when the parameters are themselves functions of covariates, making it an attractive estimator for a broad set of problems. In addition to its broad applicability, the MLE possess desirable asymptotic properties under fairly mild regularity conditions. These properties can be summarized as follows:

APPENDIX A-1. The Pennsylvania State Final Report

Asymptotic Normality: The MLE is asymptotically normal in the sense that

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \rightarrow^L N(0, I(\theta)^{-1})$$

as the sample size n increases, where $I(\theta)$ is what is known as the Fisher information matrix (see e.g. Casella and Berger, 2002). The symbol \rightarrow^L , informally speaking, means that the distribution of the random quantity $\sqrt{n}(\hat{\theta}_{MLE} - \theta)$ becomes more and more similar to a normal distribution with increasing samples. Just as the GEV may be a useful approximation for finite block sizes, and the GPD for finite thresholds, this result is useful for obtaining approximate distributional quantities for an estimator (e.g. confidence intervals, standard errors, etc.) and for performing hypothesis tests.

Consistency: An estimator $\hat{\theta}_n$ is consistent for θ if for all $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| \geq \epsilon) = 0,$$

which means that for increasing sample size, the probability that the estimator differs from the true parameter decreases to zero. That is, even if the estimator is biased for small samples, this bias decreases to zero as the sample size grows.

Efficiency: A consistent estimator $\hat{\theta}$ is said to be efficient if it achieves the smallest possible variance among consistent estimators asymptotically (as $n \rightarrow \infty$). The smallest possible variance is what is known as the Cramér-Rao lower bound (Lehmann, 1983).

Unfortunately, these regularity conditions are not met for the entire GEV family, as the support of the distribution function depends of the distribution. These violations imply that the asymptotic properties of the MLE do not always apply. Smith (1985) summarizes the three regimes of the GEV distribution as follows:

1. $\xi > -0.5$: the MLE is asymptotically normal, consistent and efficient.

APPENDIX A-1. The Pennsylvania State Final Report

2. $-1 < \xi < -0.5$: the MLE does not have the asymptotic properties above, but it is generally obtainable.
3. $\xi < -1$: the MLE is generally unobtainable.

For precipitation data, however, the tail distribution typically falls into regime 1 (Davison et al., 2012).

When modeling extreme precipitation in the Czech Republic, Kysely and Picek (2007) note that a nonparametric test for the shape parameter supports the hypothesis that distribution is heavy tailed ($\xi > 0$). As we will see in later sections, the majority of fits at rain gauge locations in the pilot region correspond to estimates of the shape parameter to be in regime 1, $\xi > -0.5$, where the desirable asymptotic properties of the MLE hold.

Once the MLE has been found, it is easy to derive the MLE for other quantities of interest such as the N-year return level because of what is known as the invariance property of the MLE. The invariance property, briefly stated, says that if $\hat{\theta}_{MLE}$ is the MLE for θ , then for function, h , $h(\hat{\theta}_{MLE})$ is the MLE of $h(\theta)$. Since the N-year return level can be expressed as a function of the parameters, the MLE can be obtained directly once the MLE for the parameters is found.

4.1.2 L-Moments

The LM estimation method has been widely used in hydrology and climatology. These estimators are linear combinations of order statistics that can be used to summarize distributional features (e.g. scale, skewness, kurtosis). The kth order statistic of a sample of n observations is the kth smallest value in the sample. The LM estimator is analogous to the more conventional ordinary method of moments estimator (MOM), which is obtained by equating sample moments to population moments ($E(X), E(X^2), \dots, E(X^k)$). For heavy tailed distributions like the GEV (the p^{th} moment does not exist when $\xi > 1/p$) theoretical moments may not be finite, and so MOM cannot be used. Even when MOM

APPENDIX A-1. The Pennsylvania State Final Report

estimators are unavailable, LM may still be viable. The LM method is equivalent to that of using linear combinations of probability weighted moments (PWM) (Hosking, 1990), where a PWM is defined as

$$PWM_{jkl} = E(X^j F(X)^k (1 - F(X))^l)$$

where F is the distribution function for X . Letting $\alpha_k = PWM_{1k0}$, $k = 0, 1, 2, \dots$, Hosking et al. (1985) showed that, assuming $\xi < 1$, L-moment estimators for the parameters of the GEV can be found by solving linear systems of equations, constructed by equating theoretical values of α_k

$$\alpha_k = \left(\frac{1}{k+1} \right) \left(\mu - \frac{\sigma}{\xi} \left[1 - \{k+1\}^\xi \Gamma(1-\xi) \right] \right)$$

to unbiased sample estimates $\hat{\alpha}_k$ which are functions of order statistics. Unlike ordinary MOM estimators, LM estimators are robust to outliers, in part because they are based on sample order statistics, rather than raw moments. Analogous estimators exist for the GPD (Hosking and Wallis, 1987), however, the PWM estimator was recently shown to be comparatively more efficient for the GEV (Ferreira and de Haan, 2015).

4.1.3 Generalized maximum likelihood

The GMLE, introduced by Jin and Stedinger (1989), aims to balance the flexibility of the flexibility of the MLE, with the desirable performance of the LM estimator in small samples. One of the reasons the LM estimator performs well, even in small samples, is the implicit constraint on the shape parameter: $\xi < 1$. This constraint is realistic in many environmental applications. The GMLE incorporates this information through an additional weighting function on the shape parameter ξ . Instead of maximizing the joint likelihood, $l(\mu, \sigma, \xi; z)$, the GMLE maximizes the likelihood weighted by an additional weighting function $\pi(\xi)$:

$$\hat{\theta}_{GMLE} = \underset{\mu, \sigma, \xi}{argmax} l(\mu, \sigma, \xi; z) \pi(\xi)$$

APPENDIX A-1. The Pennsylvania State Final Report

Jin and Stedinger (1989) suggest taking the weighting function to be a Beta distribution with support on $[-0.5, 0.5]$:

$$\pi(\xi) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} (0.5 + \xi)^{p-1} (0.5 - \xi)^{q-1}, \quad \xi \in [-0.5, 0.5]$$

to reflect a priori knowledge of a reasonable range for the shape. The Beta distribution is a fairly flexible penalization function, where the hyperparameters p and q can be chosen to give additional weight to different regions of the support. For fixed q , increasing p gives more weight to larger values of ξ , and for fixed p , increasing q gives more weight to smaller values of ξ . See the Appendix for examples.

Confidence intervals can be obtained using bootstrapping. In the case of the GEV model, a simple non-parametric bootstrap can be constructed by resampling the annual maxima and estimating the GMLE for each bootstrap sample. For the GPD model, a non-parametric bootstrap can be done by resampling the hourly or daily precipitation accumulations. Confidence intervals can then be constructed from the quantiles of the bootstrap distribution of GMLE estimates.

4.2 Estimator comparison

Here we give some guidance on the relative strengths and weaknesses of each of the estimators described above. For small sample sizes the PWM estimators for partial duration series (PDS) parameters are more robust numerically and more efficient than the MLEs (Hosking, 1990). However, Coles and Dixon (1999) argued that the comparatively worse performance of the MLE in these cases may be due to the a priori assumption that $\xi < 1$ that the PWM asserts, and claims that when a penalized likelihood (PL) approach (similar to the GMLE, e.g.) is taken that contains similar prior information about ξ through a penalty function, the penalized approach can outperform the PWM approach in terms of bias and MSE, almost uniformly.

APPENDIX A-1. The Pennsylvania State Final Report

As mentioned earlier, the MLE, under relatively weak conditions, is asymptotically efficient, but understanding the estimator performance for finite samples is also of practical interest. Landwehr et al. (1979) compare MOM, MLE, and PWM estimators of the parameters of a Gumbel distribution based on a Monte Carlo simulation study of serially correlated Gumbel random variables. In these simulations, the MLE demonstrated the minimum variance in estimating the shape parameter in both correlated and uncorrelated cases. In these simulation scenarios, the MLE tended to achieve smaller variance, except for correlated, small sample cases. It's also worth noting that the efficiency of all three methods was diminished in serially correlated cases.

A notable shortcoming of the LM method is that estimators are not easily modified to account for non-stationarity. Because estimators are found by equating sample and theoretical LMs, they tend to be model specific, which makes the incorporation of covariates difficult. Katz et al. (2002) also comments on the apparent popularity of PWM in the hydrological literature, and speculates that this may be due to their computational simplicity for i.i.d. models and their good performance in small samples, but warns that constraining to models with finite mean may be unjustified in some practical situations. The authors note that in addition to this constraint on the shape parameter, PWM have the disadvantage of not readily allowing for the incorporation of covariates (see also Bermudez and Kotz, 2010). Uncertainty calculations tend to be fairly involved, and may not be analytically available for complex hierarchical models.

4.2.1 Simulation study

To evaluate the performance of these three estimators, we conduct a simulation study similar to that done by Martins and Stedinger (2000). In this study, we examine the root mean squared error (RMSE) of the 0.5, 0.95, and 0.99 predicted quantiles of the GP and GEV distributions from each estimator for data generated shape parameters ranging over the interval $(-1, 1)$, and sample sizes ranging 50 to 1000. The

APPENDIX A-1. The Pennsylvania State Final Report

location and scale parameters were fixed at $\mu = 0, \sigma = 1$ for all cases. Simulations were all based on 10,000 Monte Carlo replicates. For each setting, a sample size of $n = 50, 100, 500, 1000$ observations were simulated from the GEV and GP distributions with the given parameter settings, each of the MLE, LM, and GMLE estimates were found, and the corresponding return levels were calculated for each estimator. The GMLE weighting function was taken to be a uniform density on $(-1, 1)$.

Figure 4-1 shows the results of this simulation study for the GEV estimators. The median RMSE estimates are shown as solid lines for each estimator, and the 95% confidence intervals are plotted as ribbons. Each row corresponds to a different sample size, and each column corresponds to a different return level. As expected, the RMSE decreases as the sample size increases, and increases for higher return levels. All three estimators exhibit very similar performance for shape parameters near 0. For small sample sizes ($n = 50, 100$), the MLE performs worse than LM and GMLE for large shape parameters for longer return-periods. LM performs worse than the other estimators for larger sample sizes ($n = 500, 1000$) when ξ is large. Except for large shape parameters, however, all three estimators have comparable performance. Particularly, the estimators show marginal differences over the shape parameter region most commonly estimated for precipitation data, $\xi \in (0, 0.3)$.

Figure 4-2 illustrates the results of the simulation study for the GPD estimators. Here we see similar results. When ξ is near 1 the MLE performs the worst when using a small sample size, and the LM estimator is comparatively worse for larger sample sizes. For $\xi < 0.5$, all estimators perform comparably.

Consistent with regime (3) reported by Smith (1985) the MLE often fails to converge for ξ near -1. Due to the similar performance of all three estimators over $\xi \in (0, 0.3)$, the region most commonly observed for precipitation data, and the flexibility advantage that the MLE and GMLE possess over the LM estimator in terms of incorporating covariates and modeling non-stationarity, the MLE and especially the GMLE are preferred in this context.

APPENDIX A-1. The Pennsylvania State Final Report

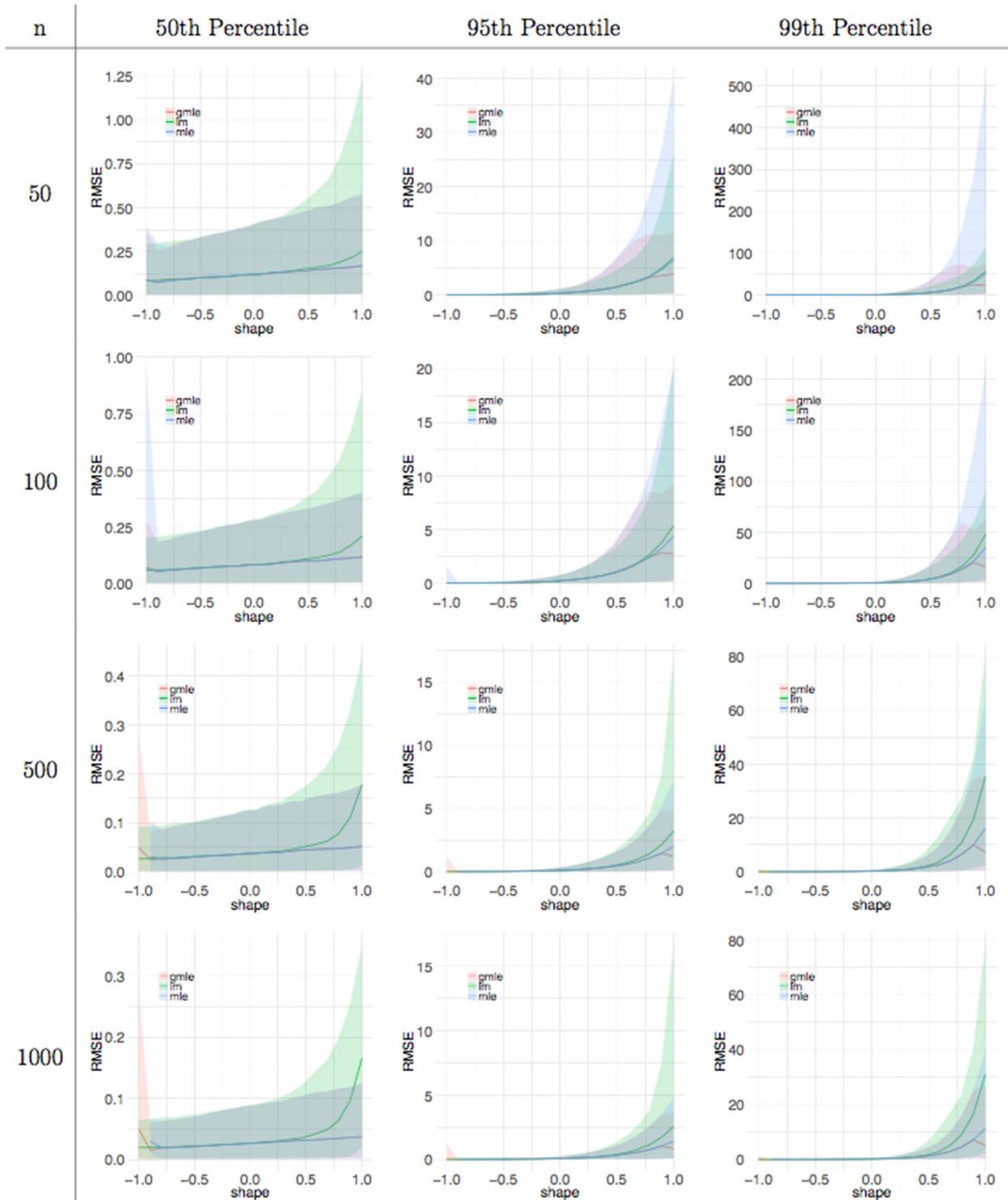


Figure 4-1: Root mean squared error (RMSE) of estimated 50th, 95th, and 99th percentiles of the GEV distribution ($\mu = 0, \sigma = 1$) for varying ξ for sample sizes of 50, 100, 500, and 1,000, each based on 10,000 Monte Carlo replicates. Lines: median RMSE, ribbons: 95% confidence intervals.

APPENDIX A-1. The Pennsylvania State Final Report

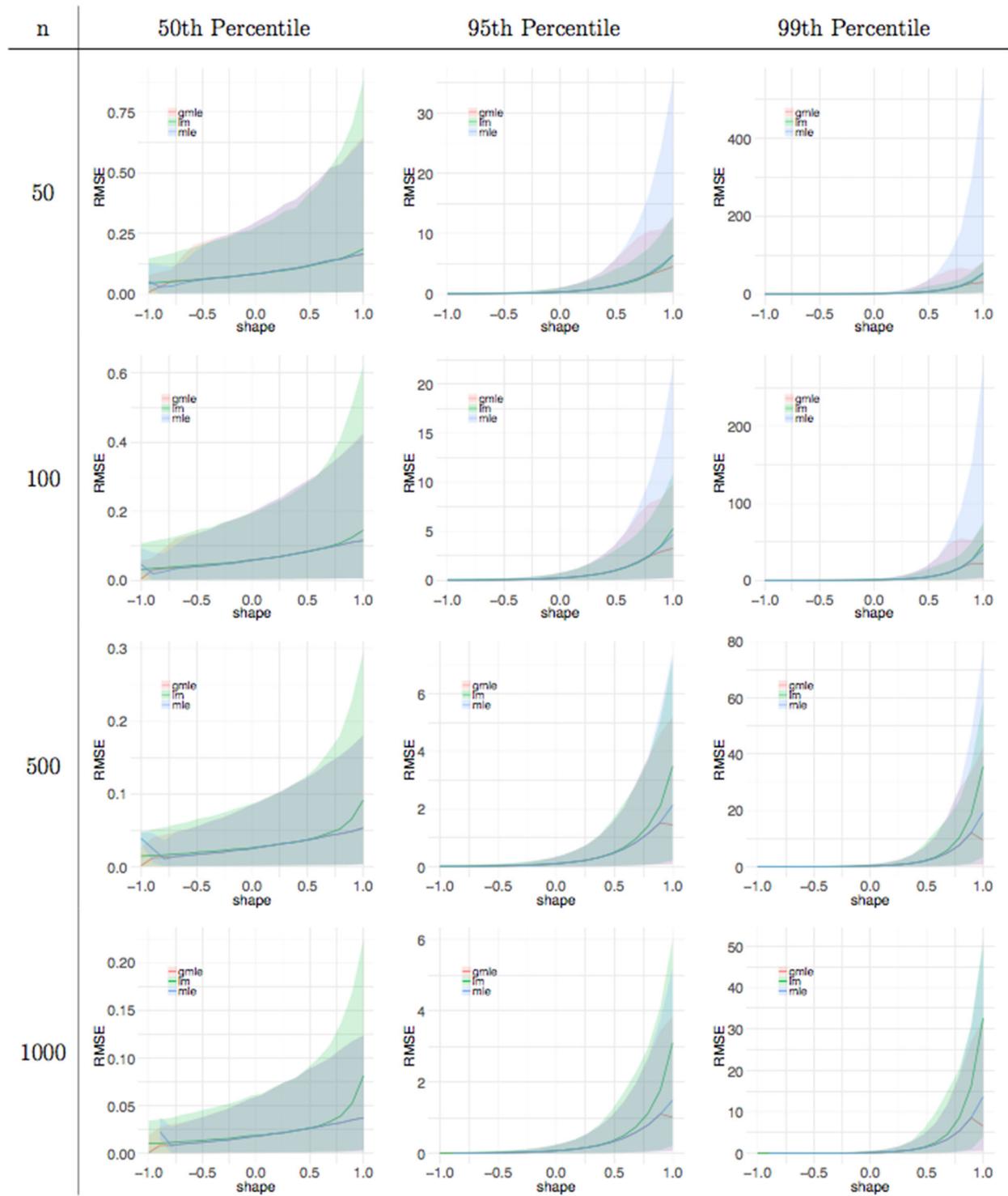


Figure 4-2: Root mean squared error (RMSE) of estimated 50th, 95th, and 99th percentiles of the GPD ($\mu = 0, \sigma = 1$) for varying ξ for sample sizes of 50, 100, 500, and 1,000, each based on 10,000 Monte Carlo replicates. Lines: median RMSE, ribbons: 95% confidence intervals.

APPENDIX A-1. The Pennsylvania State Final Report

4.2.2 Precipitation fits

Here we compare the AMS and PDS fits to the pilot data using the three estimators presented at the beginning of this section. Separate GEV and GPD models are fit for each rain gauge location in the study region. For the GMLE weighting function, we consider the Beta density given Section 4.1.3 with $p = q = 2$, to illustrate the shrinkage effect this weighting function can have on the shape parameter (See Appendix). When $p = q$, the effect of increasing these hyperparameters is to put more mass of the weighting function around zero, thereby inducing slight bias, but while decreasing the variance, resulting in an estimator that may have a lower mean squared error (MSE). Figure 4-3 shows the MLE, GMLE, and LM shape estimates for the GEV and GPD based on daily AMS and PDS at each gauge station in the pilot region plotted against station record length. For short records, the estimates are highly variable, especially for the MLE, as expected. The GMLE aims to reduce this variation by shrinking estimates towards zero, and exhibits much more robust for short record lengths. The same pattern can be seen in the hourly data, where the increased variance is more pronounced for stations with short records. We also compare the distribution of predicted 1000-year return levels for daily precipitation each station in the region obtained for each estimator (Figure 4-5). Because this return-period is so long, the differences in predictions are primarily due to differences in estimation of ξ . In particular, the 1000-year return-level MLE estimates are considerably higher at a few gauge locations than the GMLE and LM. For this reason, as discussed in the previous section, we advocate for using the GMLE. There were not significant differences between GEV and GPD based estimates of the 1000-year return-level. The maps for hourly precipitation return levels show similar consistency across estimators and have been omitted. To reiterate the conclusion made from the simulation study, because the three estimators give similar fits in most cases, and the MLE and GMLE can more easily be adapted to incorporate covariates (e.g. to capture time trends), they should be preferred to the more restrictive LM estimator in this setting. However, since the prediction of return-levels corresponding to long return-periods can be sensitive to

APPENDIX A-1. The Pennsylvania State Final Report

the shape parameter, penalizing the shape parameter gives an estimator that is less variable and more robust to small sample sizes, making the GMLE preferred over the MLE in this setting.

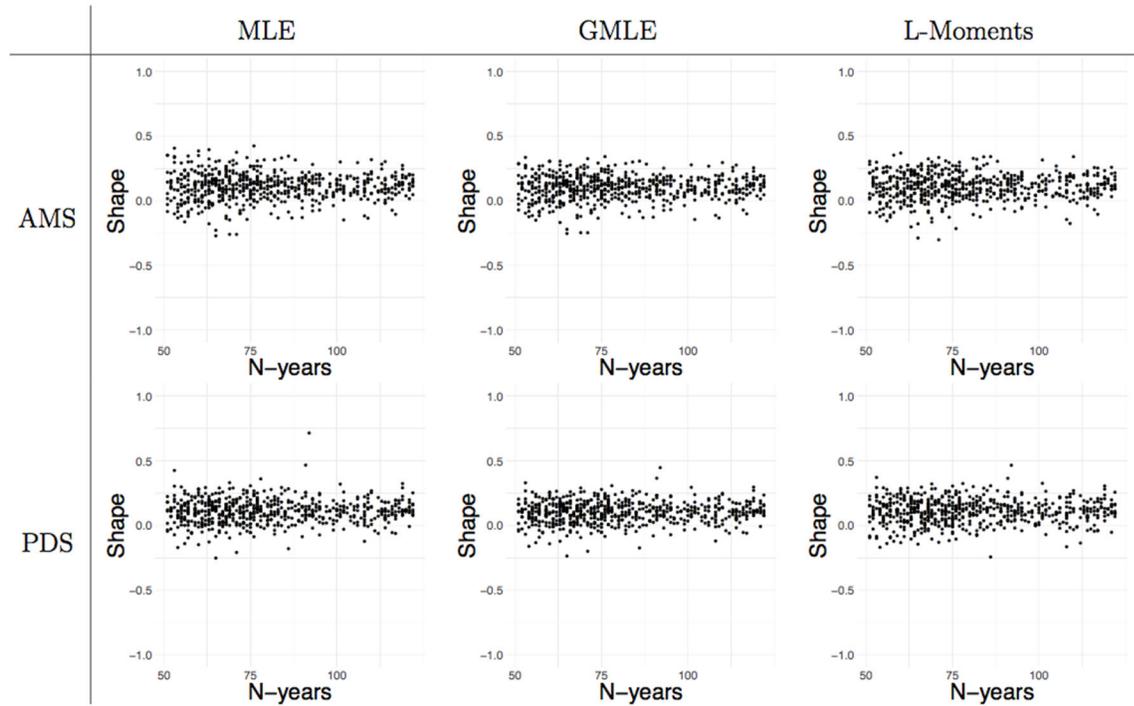


Figure 4-3: Shape estimates for the GEV (AMS) and GPD (PDS) for daily data at each station in the pilot study region based on the MLE, GMLE, and LM estimators plotted against the number of years in a stations record.

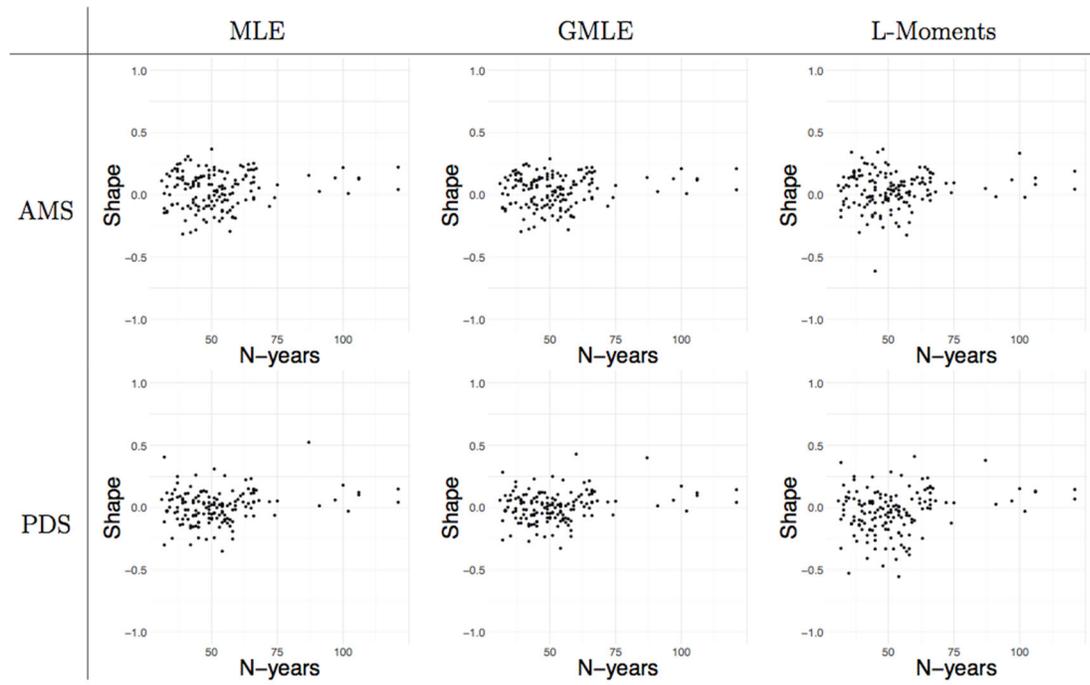


Figure 4-4: Shape estimates for the GEV (AMS series) and GPD (PDS series) for hourly data at each station in the pilot study region based on the MLE, GMLE, and LM estimators plotted against the number of years in a stations record.

APPENDIX A-1. The Pennsylvania State Final Report

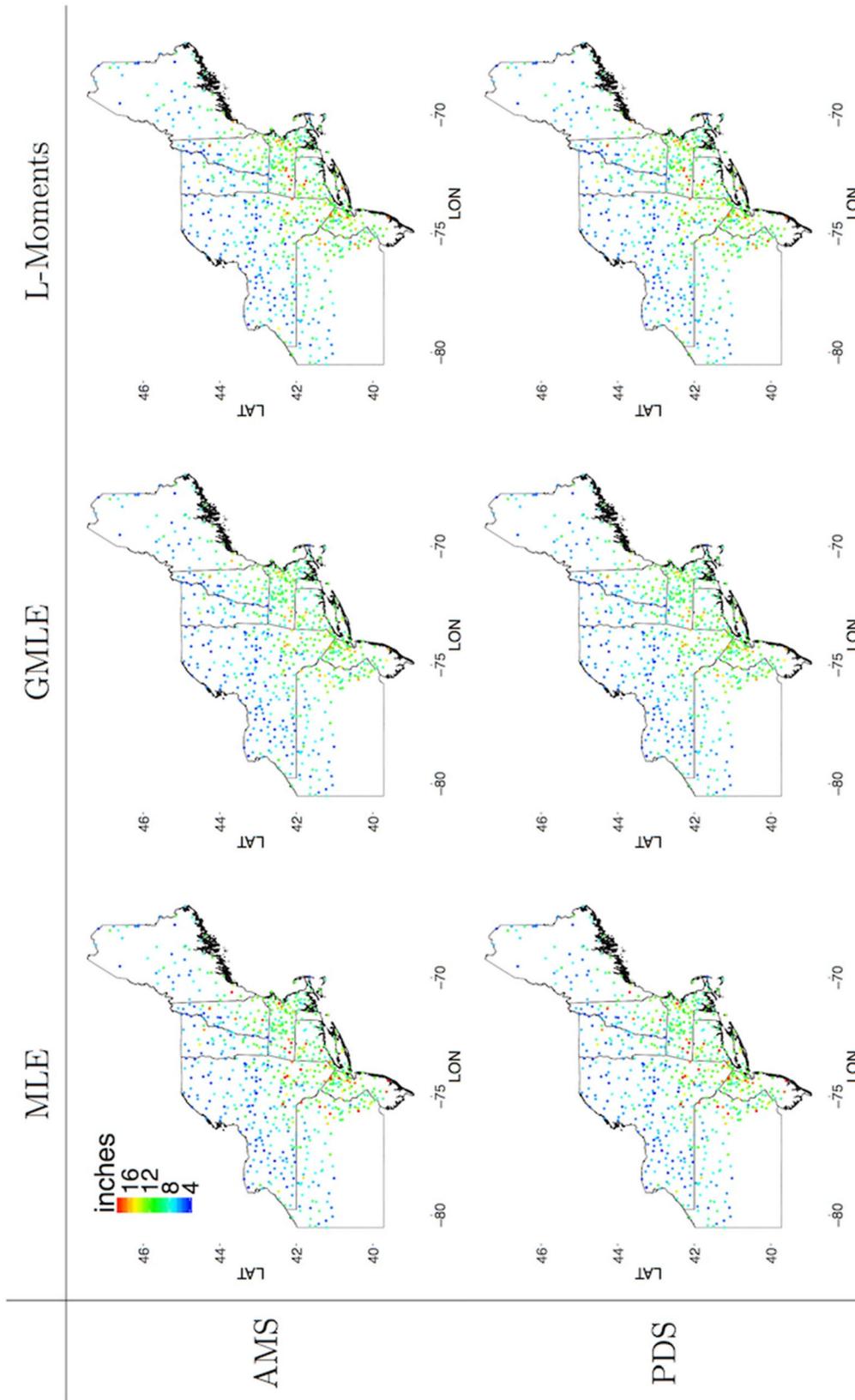


Figure 4-5: 1000-year return level estimates of daily precipitation (in) based on the MLE, GMLE, and LM estimators for the GEV (AMS) and GPD (PDS). The spatial distribution of predicted 1000-year return levels are nearly identical across estimators, suggesting that for this pilot region all three estimators give consistent fits.

APPENDIX A-1. The Pennsylvania State Final Report

5 Modeling non-stationary extreme precipitation

Most environmental phenomena are non-stationary to some degree; they may have daily, seasonal, and long-term trends. Extreme precipitation is no exception. It arises as the product of atmospheric conditions that have inherent short-term temporal dependencies and long-term trends. While accounting for short-term dependencies is beyond the scope of this study – and as we have discussed, we remove them using declustering - it is necessary to account for trends in the distribution of extreme precipitation due to a non-stationary climate. In this section, our focus will be on accounting for these long-term trends in the distribution of extreme precipitation.

The GEV and GPD models that we have described for i.i.d. data can easily be adapted to account for long-term trends using an approach analogous to the classical generalized linear model (GLM) framework (McCullagh, 1984) by allowing the parameters of distribution to relate linearly to a collection of covariates. Instead of assuming the data all come from a common distribution, we will assume that the form of the distribution is dependent upon a set of underlying covariates (e.g. time). In particular, we will assume that the parameters of the distribution (μ, σ, ξ , e.g.) depend on additional covariates. Implicit in usual concept of a return-period and return-level is the assumption of a stationary climate, which does not hold under climate change. Rootzen and Katz (2013) propose new concepts, the complementary Design Life Level and Design Life Period, for communicating risk in a non-stationary climate. These refinements aim to connect the probability of a hazardous event exceeding a specified level with some engineering time horizon of interest (e.g. the design lifetime of a levee). The Design Life Level is an upper quantile (e.g. of precipitation) that should be exceeded with a fixed probability over a specified period. In this section, although we will use the familiar language of return-levels, these return-levels should be interpreted as estimated distributional quantiles of extreme value distributions for

APPENDIX A-1. The Pennsylvania State Final Report

different points in time. For example, the 100-year return level should simply be interpreted as the $1 - 1/100 = 0.99$ quantile of annual maxima at a given point in time.

5.1.1 AMS Model

As an example, we may assume that each annual maximum follows a GEV distribution, but allow the location and scale of the GEV distribution to vary by year. Letting t denote the year, and assuming a common shape for each year, one such model would be:

$$Z(t) \sim GEV(\mu(t), \sigma(t), \xi)$$

$$\mu(t) = \alpha_0 + \alpha_1 t$$

$$\log(\sigma(t)) = \beta_0 + \beta_1 t.$$

Modeling the log of the scale as a linear function of time, instead of the scale itself, ensures that the scale parameter is always positive, as required. The shape parameter is difficult to estimate in the stationary case, and so modeling it smoothly in time can be very unstable.

5.1.2 PDS Model

We can model the GPD analogously, by letting the parameters of the distribution depend on covariates. However, in this context we will be using the Poisson point process (PPP) parameterization of the GPD. The PPP characterization unifies the GEV and GPD models and provides a natural way to model non-stationarity in threshold exceedances. A point process is a natural class of models for the random locations of point events. This approach can be applied not only to data that have spatial reference locations, but also in an abstract sense to general spaces. The latter is what is done in the case of the PPP characterization of the GPD. Here we briefly describe its construction but omit the technical details. See Coles (2001) for an overview. Suppose that x_1, x_2, \dots are the random positions of a collection of points on some set Ω . The points are said to follow a non-homogeneous Poisson point process if for all

APPENDIX A-1. The Pennsylvania State Final Report

non-overlapping subsets $A, B \subset \Omega$, the number of points falling in those subsets, $N(A)$ and $N(B)$, are independent, and $N(A) \sim \text{Poisson}(\Lambda(A))$ for $\Lambda(A) = E(N(A)) = \int_A \lambda(x) dx$, where $\lambda(x)$ is what is called the intensity function of the PPP. Suppose that Y_1, Y_2, \dots independent and well behaved in the sense that $(M_n - b_n)/a_n = (\max(X_1, \dots, X_n) - b_n)/a_n$ converges in distribution to a $GEV(\mu, \sigma, \xi)$ for some sequences a_n, b_n . We define the two-dimensional points of the points process as $N_n = \left\{ \frac{i}{n+1}, \right.$

$(Y_i - b_n)/a_n \left. \right\}$, where the first component guarantees that the time range is always mapped back to $(0,1)$ as the number of points grows. We consider sets of the form $A = (0,1) \times (u, \infty)$ where u is some high threshold. Then since by the usual GPD construction, the probability of an exceedance is

approximately $p = P\left(\frac{Y_i - b_n}{a_n} > u\right) = \frac{1}{n} \left[1 + \frac{\xi(u - \mu)}{\sigma}\right]^{-1/\xi}$, and the Y_i are assumed to be independent, the

number points falling into sets of the type above is Binomial, $N_n(A) \sim \text{Binomial}(n, p)$. As the sample size grows, this quantity converges in distribution to a Poisson limit, $N_n(A) \rightarrow^d \text{Poisson}(\Lambda(A))$ where

$\Lambda(A) = \left[1 + \frac{\xi(u - \mu)}{\sigma}\right]^{-1/\xi}$. Using this limiting model as an approximate distribution of a finite collection

of i.i.d. observations, y_1, \dots, y_n , the likelihood for the PPP model becomes $L(\mu, \sigma, \xi; y_1, \dots, y_n) =$

$\exp(-\Lambda(A)) \prod_{i=1}^n \lambda(t_i, x_i) \propto \exp\left(\left[1 + \frac{\xi(u - \mu)}{\sigma}\right]^{-1/\xi}\right) \prod_{i=1}^{N(A)} \frac{1}{\sigma} \left[1 + \frac{\xi(x_i - \mu)}{\sigma}\right]^{-\frac{1}{\xi} - 1}$. We denote this three-

parameter family by $GPD(\mu, \sigma, \xi)$. These parameters correspond directly to those of the GEV

distribution, which makes quantile estimation straightforward. Since annual blocks are typically used for

GEV estimation, to express the estimated return-levels of this model in terms of annual time-units, a

small adjustment factor of n_y , the number of years of observation, to the time scale window of the A

sets is needed. So that the parameters estimated by maximizing the likelihood $L(\mu, \sigma, \xi; y_1, \dots, y_n) \propto$

$\exp\left(-n_y \left[1 + \frac{\xi(u - \mu)}{\sigma}\right]^{-1/\xi}\right) \prod_{i=1}^{N(A)} \frac{1}{\sigma} \left[1 + \frac{\xi(x_i - \mu)}{\sigma}\right]^{-\frac{1}{\xi} - 1}$ are in the same units as those of a GEV fit to

annual maxima. Once the parameters of this PPP GPD distribution are estimated, m-year return-levels

can be found by plugging the estimated parameters into the GEV quantile function given by:

APPENDIX A-1. The Pennsylvania State Final Report

$$Q(p; \mu, \sigma, \xi) = \begin{cases} \mu + \sigma(-\log(p))^{-\xi} - 1/\xi, & \xi \neq 0 \\ \mu - \sigma \log(-\log(p)), & \xi = 0 \end{cases}$$

Note that if $\tilde{\sigma}, \tilde{\xi}$ are the scale and shape parameters for a GPD model above threshold u , the corresponding GEV (i.e. or PPP) scale σ and shape ξ are given by $\sigma = \tilde{\sigma} - \xi(u - \mu)$ and $\xi = \tilde{\xi}$, where μ is the location parameter of the GEV. Returning to our model for non-stationary GPD model, analogously to the GEV model, let $Y_i(t)$ denote the i^{th} observation in year t , and model

$$Y_i(t) | Y_i(t) > u \sim GPD(\mu(t), \sigma(t), \xi)$$

$$\mu(t) = \alpha_0 + \alpha_1 t$$

$$\log(\sigma(t)) = \beta_0 + \beta_1 t.$$

Unlike in the GEV case, where the time-scale coarseness (i.e. one observation per year) of the AMS nullified any short-term dependence features, dependence among successive PDS observations must be accounted for. For this study, we will ignore the short-term dependence by applying the declustering procedures described in Section 3.1.1 before fitting the models described here.

In the remainder of this section, we show how to account for non-linear trends, review two different methods that can be used to detect trends among the distribution of extreme precipitation and apply this framework to the pilot study precipitation data.

5.1.3 Spline model

While linear trends may often capture the major features of the data, it is possible to account for more flexible trends using an expanded set of covariates. By modeling each parameter as a linear combination of basis functions, it is possible for the parameters of the GPD or GEV to vary non-linearly in time. Linear combinations of basis functions are capable of representing a wide array of smoothly varying functions.

APPENDIX A-1. The Pennsylvania State Final Report

Taking the GEV distribution as an example again, instead of modeling the location parameter as a linear function of time, we can allow for a wider variety of time trends in the by modeling it as

$$\mu(t) = \alpha_0 + \alpha_1 t + \sum_{k=1}^K \gamma_k^{(\mu)} B_k(t)$$

where B_1, \dots, B_K are collection of basis functions (e.g. B-splines) (Wang, 2011). We can model any of the other parameters in the same way. A log link function is applied to the scale parameter to restrict it to be positive (i.e. $\log(\sigma(t)) = \beta_0 + \beta_1 t + \sum_{k=1}^K \gamma_k^{(\sigma)} B_k(t)$). To use this approach, one must decide a sufficient number of spline terms to use. In practice, one typically chooses a large number of splines, but enforces some smoothness through penalization. To ensure that the splines do not overfit the data, or so that the shape of the fitted distribution varies slowly in time, one will often penalize the roughness of the fitted spline (e.g. the square of the second derivative of $\mu(t)$). For a description of penalized B-splines, see Eilers and Marx (1996). This penalization has an analogous effect to the weighting function used by the GMLE.

5.2 Tests for non-stationarity

We look at two tests for non-stationarity in the distribution of extremes. The first is a likelihood ratio test for a linear trend in the location parameter of the GEV distribution. The second is a test for a linear change in the frequency of exceedances of a fixed high threshold.

The first test is intended to detect linear location changes in the GEV location parameter μ . For this, we use a likelihood ratio test. We assume that annual maxima $Z_t, t = 1, \dots, n$ follow a $GEV(\alpha_0 + \alpha_1 t, \sigma, \xi)$ distribution, where t denotes the year, and test for a linear trend in the location parameter $H_0: \alpha_1 = 0$ vs $H_1: \alpha_1 \neq 0$. See Casella and Berger (2002) for details on the likelihood ratio test. We emphasize that this test is intended to detect one particular type of non-stationarity in the distribution of annual

APPENDIX A-1. The Pennsylvania State Final Report

maxima – linear trends in the location of the distribution. Non-linear changes do not fall within the scope of this test, however, testing for such changes may also be of interest.

In addition to testing for changes in the magnitude of maxima, one may also be interested in changes in the frequency of extremes, and while these two things are related for most non-pathological phenomena, this next test is intended to detect changes in the frequency of extremes. That is, the aim is to test if the number of extreme events (as qualified by some high fixed threshold) is changing linearly over time. Here we assume that the number of exceedances of a fixed high threshold follow a Poisson distribution with rate $\lambda = \delta_0 + \delta_1 t$, and test for a linear change in the rate $H_0: \delta_1 = 0$ vs $H_1: \delta_1 \neq 0$ using a Wald test (Casella and Berger, 2002).

In addition to these two tests, we also consider the Mann-Kendall test for detecting trends (Mann, 1945; Kendall 1938;), and apply it to three different data time series (1) the number of exceedances per year, (2) the precipitation AMS, and (3) the precipitation PDS (see Appendix for simulation study assessing these tests).

The likelihood ratio test is a general approach that could have also been applied to detecting a linear trend in the scale and shape parameters of the GEV distribution or parameters of the GPD distribution, but changes in the shape parameter may be difficult to detect unless the precipitation record length is sizable (Cheng et al., 2014 and Lee et al., 2017). Due to the low power these tests have for detecting changes in the scale and shape of the GEV and GPD distributions, we did not consider them. Also, since the penalized spline method described above is able to adapt to trends when they exist, and give a nearly stationary fit when they don't, testing for trends in the remaining parameters does not change our estimation approach.

APPENDIX A-1. The Pennsylvania State Final Report

5.3 Precipitation fits

We apply the two tests described in the previous section to the daily precipitation data from the study region, all at a level of $\alpha = 0.05$. Tests were applied to each station separately, the GEV location test based on the AMS series and the Poisson rate test applied to counts of observations above the 0.98 quantile for the entire time series for a given station. To summarize the results of each test, stations were classified into three groups: if the null hypothesis of no linear trend was not rejected, the station was labeled as neutral. If a linear trend was detected, the station was classified according to the sign of the linear trend coefficient (α_1 or δ_1 above) – either positive or negative. The results of the test are shown in Figure 5-1. Whereas the majority of tests for the GEV location parameter failed to detect a linear trend, many of the tests for changes in the frequency of extreme precipitation events detected a linear trend. For those stations where a linear trend was detected in the frequency of extreme events, the vast majority had a positive trend – indicating an increase in the frequency of extreme precipitation in this region over time.

To reduce the variance of these tests, we also consider pooling information from nearby gauge locations (see Section 6.1 for the weighting/pooling scheme). The likelihood ratio test and Wald test can be applied in this regionalization context by weighting the likelihood contributions of each gauge location. Note that this weighted likelihood scheme does not apply to the Mann-Kendall test, which is non-parametric. The results of these tests are reported in Figure 5-2. After pooling information, the tests for the majority of locations detect a positive trend.

One way to account for trends when they exist is to fit a model like that described at the beginning of this section. To illustrate this method, we apply the penalized spline approach to AMS and PDS of daily data for the Maynard, MA gauge location used in Section 3.1.2. To enforce some consistency between the two models, we will use the Poisson point process construction of the GPD (see Coles (2001) for

APPENDIX A-1. The Pennsylvania State Final Report

details), which is a three-parameter family that is directly comparable to the GEV family. We compare the predicted 2, 100, and 1000-year return levels for the GEV and GPD models where the parameters are smoothly varying functions of time as described in Section 5.1.3. We consider three cases of increasing complexity:

M0: Stationary model

M1: Location is smoothly varying, but scale and shape are fixed across time.

M2: Location and scale each vary smoothly, but the shape is fixed across time.

M3: Location, scale, and shape each vary smoothly.

For all GPD models, we use a threshold fixed at the 0.98 quantile of non-zero observations. Other thresholds were considered and gave similar results. The penalty function used for the GMLE is applied in all instances when ξ is not varying in time, and all parameters are estimated by maximizing this penalized likelihood.

To construct confidence bands for the return-level curves, here again we use bootstrapping. For the GPD models, we resample the daily and hourly observations, blocking by year. For the GEV models, we use a parametric bootstrap. Annual maxima are simulated from the fitted GEV distribution for each year, such that the bootstrap samples have the same record length as the original sample, with one observation per year. The spline model is then fit and return levels estimated from each bootstrap sample. Note that a non-parametric bootstrap of the year and annual maxima pairs could also be used for the GEV models, but the parametric approach ensures that the bootstrap samples have one observation per year, which would not be the case for the non-parametric bootstrap. In practice, the choice of the bootstrap resampling scheme (e.g. different flavors of parametric or non-parametric bootstraps) is rarely crucial, and can be made for convenience. See Efron and Tibshirani (1994) for a review of bootstrapping.

APPENDIX A-1. The Pennsylvania State Final Report

Confidence intervals are constructed from pointwise quantiles of the bootstrap return-level estimates. Note that pointwise bootstrapped confidence intervals do not account for model misspecification, but do account for uncertainty in the return-level estimates due to sampling error.

The estimated 2, 100, and 1000-year return levels and 95% confidence bands for each model are shown in Figure 5-2. The data exhibit a very slight increasing trend, as detected by all of the tests described in the previous section. Models M2 and M3 appear to overfit the data, and especially so for the smoothly varying shape parameter model. Since estimation of the shape parameter is difficult, even in the stationary case, we recommend fixing the shape parameter across time, unless there is strong reason to believe that the shape is changing with time, and using either model M1 or M2 to account for non-stationarity. The GEV and GPD models give similar results, assuming a sufficiently high threshold and sufficiently large block size has been selected. Occasionally, the GPD model fits were slightly unstable when data quantization (rounding that occurred when the data were recorded) was present, as is the case for some of the hourly records. Further, since the GEV and GPD models give very similar results otherwise, and the GEV model does not require the modeler to make threshold selection and declustering parameter decisions for each station, we advocate for using the GEV models in this setting.

APPENDIX A-1. The Pennsylvania State Final Report

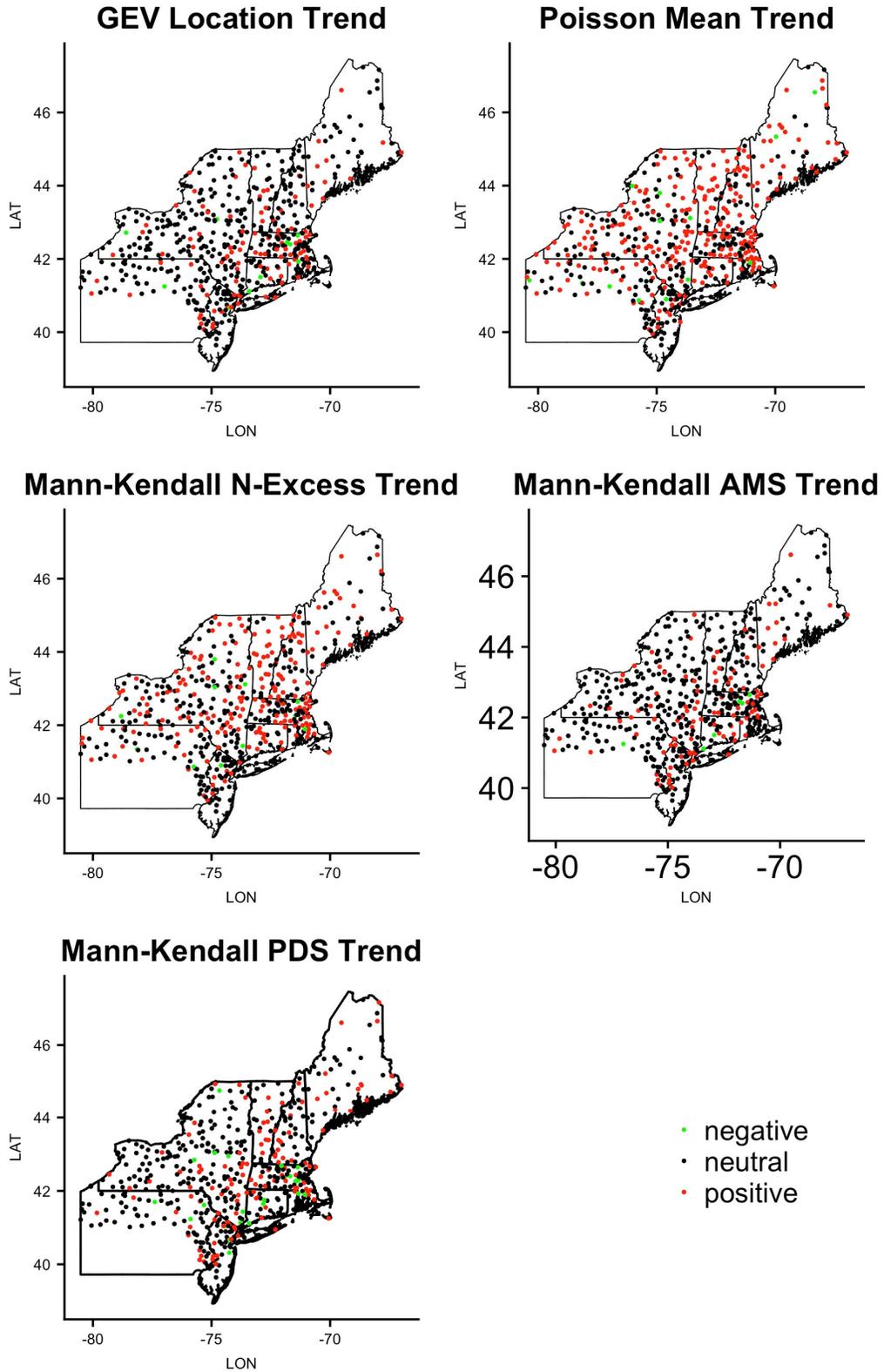


Figure 5-1: Spatial distribution of GEV location, and Poisson mean, and Mann-Kendall trend tests. If the null hypothesis of no linear trend was not rejected, the station was labeled as neutral. If a linear trend was detected, the station was classified according to the sign of the linear trend coefficient (α_1 or δ_1 above) – either positive or negative.

APPENDIX A-1. The Pennsylvania State Final Report

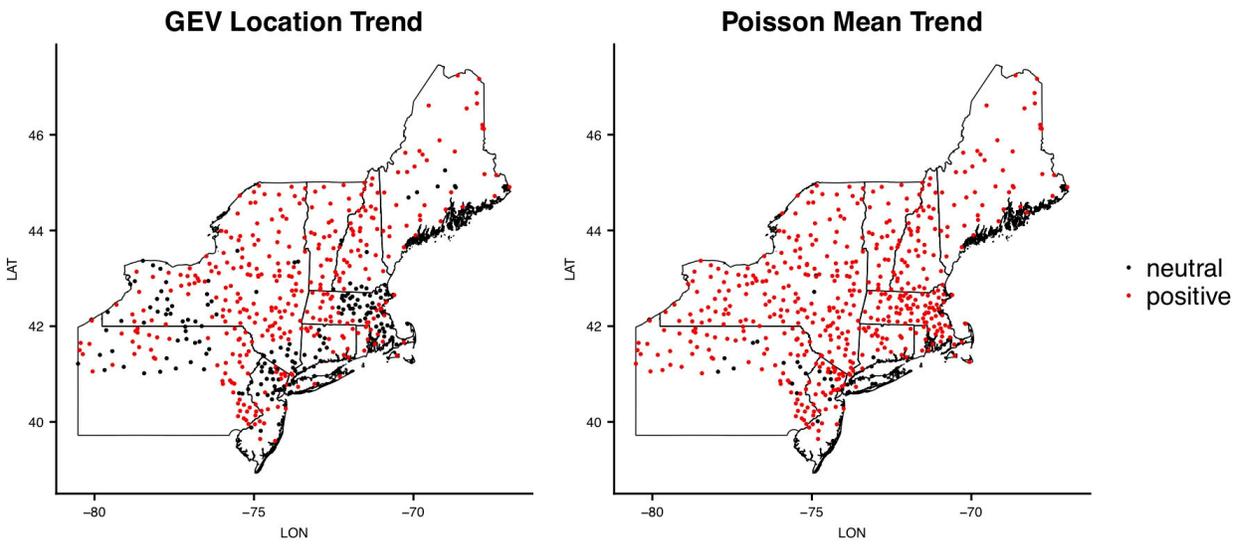


Figure 5-2 Spatial distribution of GEV location, Poisson mean, and Mann-Kendall trend tests applied to locally weighted likelihoods (regionalization). If the null hypothesis of no linear trend was not rejected, the station was labeled as neutral. If a linear trend was detected, the station was classified according to the sign of the linear trend coefficient (α_1 or δ_1 above) – either positive or negative

APPENDIX A-1. The Pennsylvania State Final Report

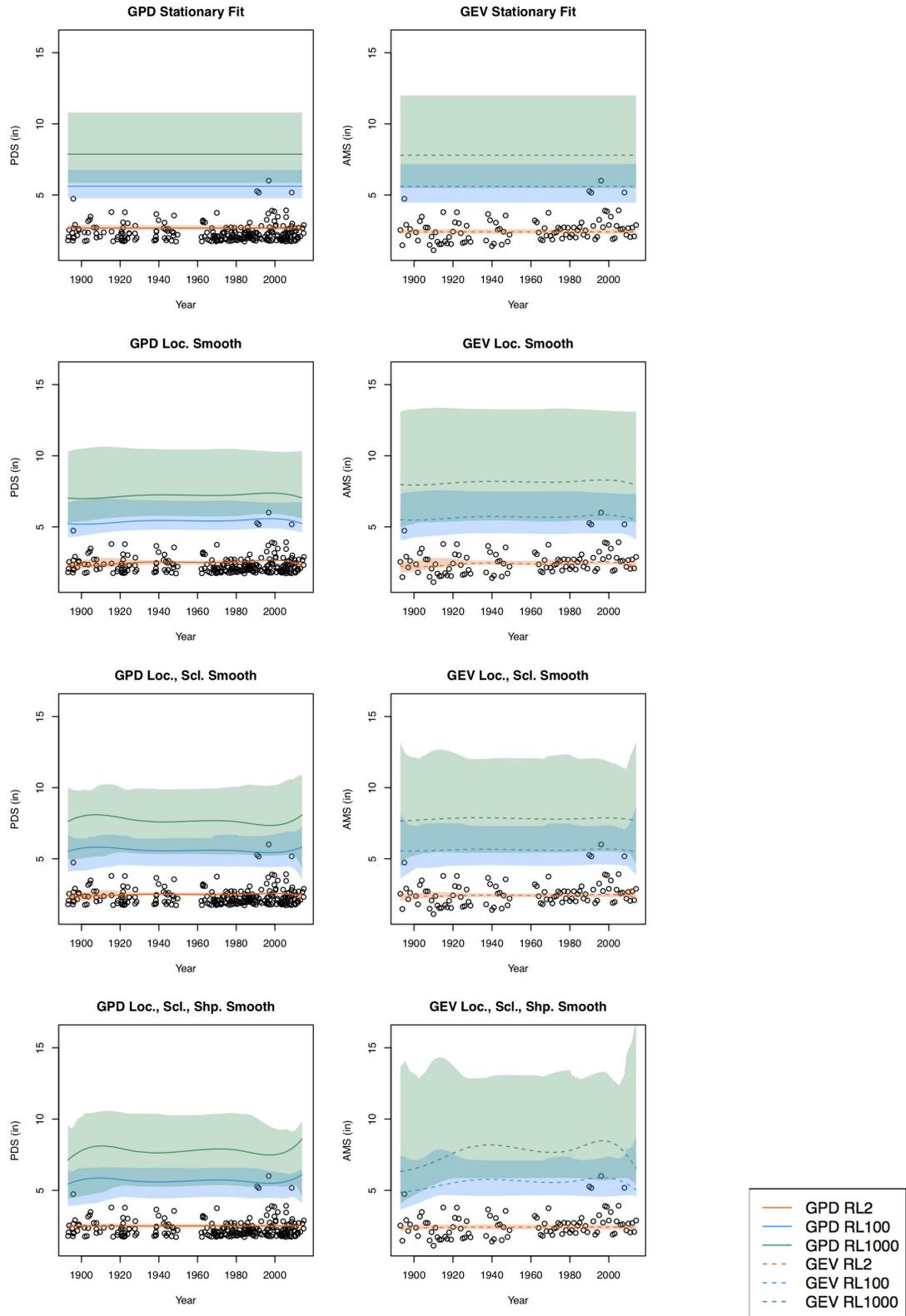


Figure 5-3: Return level (2-year: brown, 100-year: blue, 1000-year: green) estimates from GEV (right column) and GPD (left column) models M0 (top), M1 (row 2), M2 (row 3), and M3 (bottom) for PDS and AMS series from Maynard, MA are plotted over time. The 95% bootstrap confidence bands are plotted as ribbons of the same color as their corresponding point estimates. Models M2 and M3 appear to overfit the data slightly.

APPENDIX A-1. The Pennsylvania State Final Report

6 Spatial Analysis

In this section, we develop a model for spatially varying extreme precipitation. Atmospheric phenomena exhibit inherent spatial dependence, and this dependence can be used to reduce variability when doing estimation. In the last decade, considerable attention has been given to max-stable models, a class of asymptotically justified models for spatial extremes (Davison, 2012). A current limitation of these models is that they are computationally intensive, and do not scale to many spatial locations. Our focus will be on describing the return-levels of extreme precipitation at each point in space. Since the focus of this application is on adequately estimating pointwise return-levels of extreme precipitation and there are many gauge locations, we propose using a local likelihood model that incorporates information from nearby gauge locations to reduce the variation in quantile estimation.

6.1 Local Likelihood Model

The local likelihood approach applies equally well to modeling AMS as well as PDS data and can be broadly described as follows: for a given prediction location – a spatial location at which one would like to predict various return-levels – one considers only the data from nearby gauge stations, those which are within a pre-specified distance of the prediction location. As in previous sections, the MLE or GMLE can be calculated for the GEV or GPD parameters, but here using the data from all nearby stations rather than just a single station. In addition, the contribution of each observation to the likelihood is weighted by the proximity of each observation location to a prediction location such that closer observation locations are weighted more heavily than distant ones.

More concretely, denote the spatial locations of the observation gauge locations by s_1, \dots, s_J , and denote the spatial prediction locations by s_1^*, \dots, s_K^* . The local likelihood estimators are found by maximizing a weighted likelihood in a neighborhood around each prediction location. Let $N_k, k = 1, \dots, K$ denote the indices of spatial locations in a neighborhood of prediction location s_k^* defined as:

APPENDIX A-1. The Pennsylvania State Final Report

$$N_k = \{j \in (1, \dots, J): d(s_j, s_k^*) < \epsilon\}$$

where d is a distance function (e.g. Euclidean, Haversine, etc.) and ϵ is a bandwidth parameter that determines the neighborhood for which stations will receive positive weight. Those outside of this neighborhood have zero weight, and therefore no influence. Each N_k gives a set of indices for the observation stations that have non-zero weight when doing estimation at site s_k^* . The bandwidth tuning parameter dictates the trade-off between spatial smoothness and local variation under the model. For larger bandwidths, the model will pool data from more stations for estimation at a given prediction location, but potentially smooth out local variation. Smaller bandwidths will be less biased by distant, unrepresentative stations, but result in highly variable estimators, potentially identifying spurious spatial signals.

The local likelihood estimator maximizes a weighted product of individual likelihoods, where nearby observations have higher weight. For example, when modeling AMS data, denoting the i th observation at the j th spatial location by z_{ij} , the local likelihood estimator of the GEV parameters at prediction location s_k^* is given by

$$(\hat{\mu}_k, \hat{\sigma}_k, \hat{\xi}_k) = \underset{\mu, \sigma, \xi}{\operatorname{argmax}} \sum_{j \in N_k} \sum_i w_{jk} \log(f(z_{ij}; \mu, \sigma, \xi))$$

where weights w_{jk} are calculated as a function of the distance between s_j and s_k^* , and $f(z_{ij}; \mu, \sigma, \xi)$ is the GEV density at annual maximum z_{ij} (see Section 4.1.1). Analogous to the GMLE described in Section 4.1.3, $f(z_{ij}; \mu, \sigma, \xi)$ can be replaced by $f(z_{ij}; \mu, \sigma, \xi)\pi(\xi)$ to shrink ξ towards zero. For our analysis, weights are calculated by applying a triweight kernel function to the distances. The triweight kernel function is given by

$$K(r) = \left(\frac{\epsilon^2 - r^2}{\epsilon^2} \right) I(|r| < \epsilon)$$

APPENDIX A-1. The Pennsylvania State Final Report

where, as above, ϵ is the bandwidth parameter. The weights, $w_{jk} = \frac{K(d(s_j, s_k^*))}{\sum_{j \in N_k} K(d(s_j, s_k^*))}$ where $d(s, s')$ is the Haversine (great-circle) distance between two points, are normalized such that at each prediction location, the weights sum to one ($\sum_{j \in N_k} w_{jk} = 1$). Figure 6-1 illustrates this weighting scheme for a given prediction location indicated by a black point. Only those observation locations within the range indicated by the colored disc would have positive weight, where closer ones are weighted more heavily than distant ones, represented by color gradient from red (large weight) to green (small weight). Observation locations outside of this disc have zero weight.

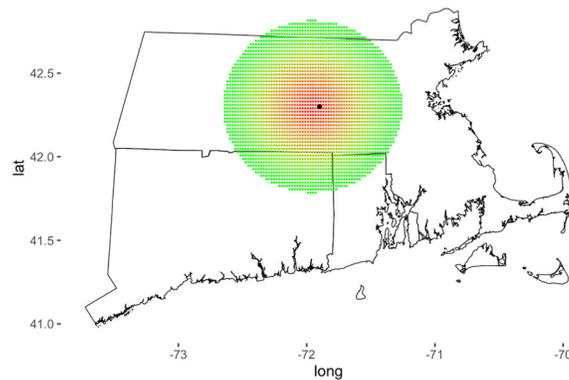


Figure 6-1: Illustration of weighting scheme used for local likelihood model. For a given prediction location, here indicated by a black point, observations from gauge stations within a fixed radius of the prediction location are given positive weight (colored region), where closer points are weighted more heavily than distant ones, as indicated by the gradient of red (large weight) to green (small weight).

The same approach can be taken for modeling PDS data by replacing the likelihood above with that of a GPD, where, as in Section 5, we use the 3-parameter Poisson point process GPD characterization. However, just as in the single station analysis, one needs to select a high threshold, above which the GPD assumption is not badly violated. Instead of fixing a single threshold across the entire region, we take a moving window approach. For each prediction location, a threshold is fixed at a high sample quantile of all of the stations in the local neighborhood of that prediction location. Due to the moving

APPENDIX A-1. The Pennsylvania State Final Report

window, the threshold is free to vary in space and reflect changes in precipitation extremes in different regions, while still borrowing spatial information to reduce estimator variance.

As in Section 5, by allowing the parameters of the GEV and GP distributions to vary as smooth functions of covariates, additional climate and temporal information can be incorporated into the model. In Section 5, we considered time covariates, but these are by no means the only useful covariates that may describe changes in the underlying distribution of extreme precipitation. More generally, in both the GEV and GPD models, set

$$\mu(X_\mu) = X_\mu \boldsymbol{\beta}_\mu, \quad \log(\sigma(X_\sigma)) = X_\sigma \boldsymbol{\beta}_\sigma, \quad \xi(X_\xi) = X_\xi \boldsymbol{\beta}_\xi,$$

where the X_μ , X_σ , and X_ξ are covariate design matrices, and $\boldsymbol{\beta}_\mu$, $\boldsymbol{\beta}_\sigma$, and $\boldsymbol{\beta}_\xi$ are vectors of parameters.

Because estimation at different prediction locations can be done independently of each other, this method is computationally very fast as it can be done in parallel. Further, since estimation at any prediction location depends only on data from nearby gauge locations, this method is scalable to many prediction locations.

6.2 Precipitation fits

Here we illustrate the fit of the local likelihood model to the AMS and PDS precipitation data over the pilot region. Daily precipitation return levels are estimated at 6,908 grid locations in the pilot region. Leave-one-out cross-validation was used to select a bandwidth that minimized the quantile score (for the $\tau = 0.98$ quantile), from which a bandwidth of 57.4 miles was chosen (see Figure 6-2). For hourly data, the optimal bandwidth was chosen to be 82.1 miles (Figure not shown). For the majority of prediction locations, this fixed bandwidth was sufficient. However, since some prediction locations are somewhat far from the nearest gauge location, the bandwidth was adjusted such that no window includes fewer than 10 gauge locations.

APPENDIX A-1. The Pennsylvania State Final Report

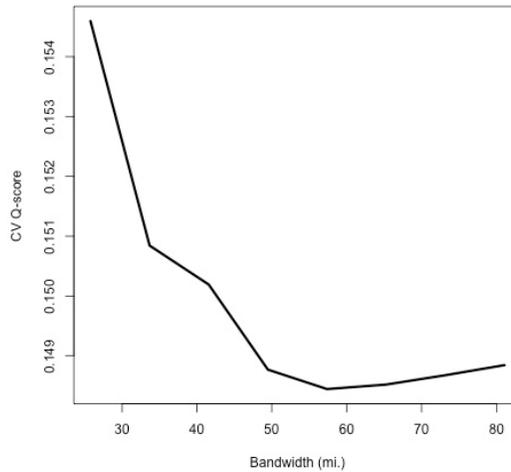


Figure 6-2: Cross-validation Q-scores for daily precipitation as a function of bandwidth for the 0.98 quantile using leave-one-out cross validation. The optimal bandwidth for hourly precipitation was selected to be 82.1 miles.

As can be seen in Figure 6-3, this bandwidth results in a balance of localization and smoothness to the estimated return-level surfaces. For a given prediction location and bandwidth, the PDS data was constructed by examining threshold exceedances above the 98th percentile of all observations from stations in the local neighborhood of that prediction location.

For additional flexibility, the location and scale parameters of the GEV and GPD local likelihood models vary as functions of mean annual precipitation (MAP) over the period 1981-2010 (PRISM, 2017), but the shape parameter does not vary as a function of covariates. Specifically, denoting these covariates generically by X , for both GEV and GPD models, we set

$$\mu(X) = \alpha_0 + \alpha_1 MAP$$

$$\log(\sigma(X)) = \beta_0 + \beta_1 MAP,$$

but not allow ξ to vary as a function of covariates. As in Section 4.1.3, we use the GMLE penalty function for ξ . In each case, the parameters are estimated by maximizing the weighted likelihood at each prediction location as described earlier.

APPENDIX A-1. The Pennsylvania State Final Report

As an aside, note that often the same storm will have caused the yearly maximum precipitation event at several gauge locations in a single area. This dependence will not bias the pointwise parameter estimates but will have an effect on the estimated standard errors, since the number of effectively independent samples is smaller under dependence than independence. Note however, that the bootstrapping scheme described in the next section gives unbiased estimates of the standard errors even in the presence of spatial dependence, and so this is not a concern.

To construct confidence intervals for the return-levels, a similar bootstrapping scheme to that of Section 5.1.3 is used. Parametric bootstraps are used for both the GEV and GPD models, now blocking by spatial location, rather than year. Bootstrap samples of annual maxima are drawn from the fitted GEV model at each gauge station, and for the GPD, model threshold exceedances equal in number to that of the original sample are drawn for each gauge stations. In either case, for each bootstrap sample, the local likelihood model is refit, and return-levels calculated. The 95% confidence intervals of the return-levels at each prediction location are then taken to 0.025 and 0.975 quantiles of the bootstrap return-level estimates. Local likelihood estimates and 95% confidence intervals of the 2, 100, and 1000-year return levels for each prediction location are shown in Figure 6-3 for the GEV fits and Figure 6-4 for the GPD fits (the general spatial pattern for hourly data is similar). Both models give similar spatial patterns, where the most extreme estimated precipitation return-levels are along the coast and in pockets in Connecticut and southern New York. There are some discernable differences between the two fits, especially at higher return-levels, where the shape parameter has more influence. The GPD model appears to have slightly less stable fits than the GEV does, especially in regions where the number of gauge locations is small. This instability was exacerbated for the GPD model when the data are highly quantized, as is the case for much of the hourly data. Because of this, we advocate for the use of the GEV local likelihood model over the GPD when for this precipitation application. Finally, note that the proposed method does not enforce any smoothness constraints on the return-level surfaces, and

APPENDIX A-1. The Pennsylvania State Final Report

because of this it is possible that this estimation method will give more variable estimates than those of NOAA Atlas 14. Smoothness can be enforced using a post-processing spatial filter, but this is outside of the scope of what is considered here.

APPENDIX A-1. The Pennsylvania State Final Report

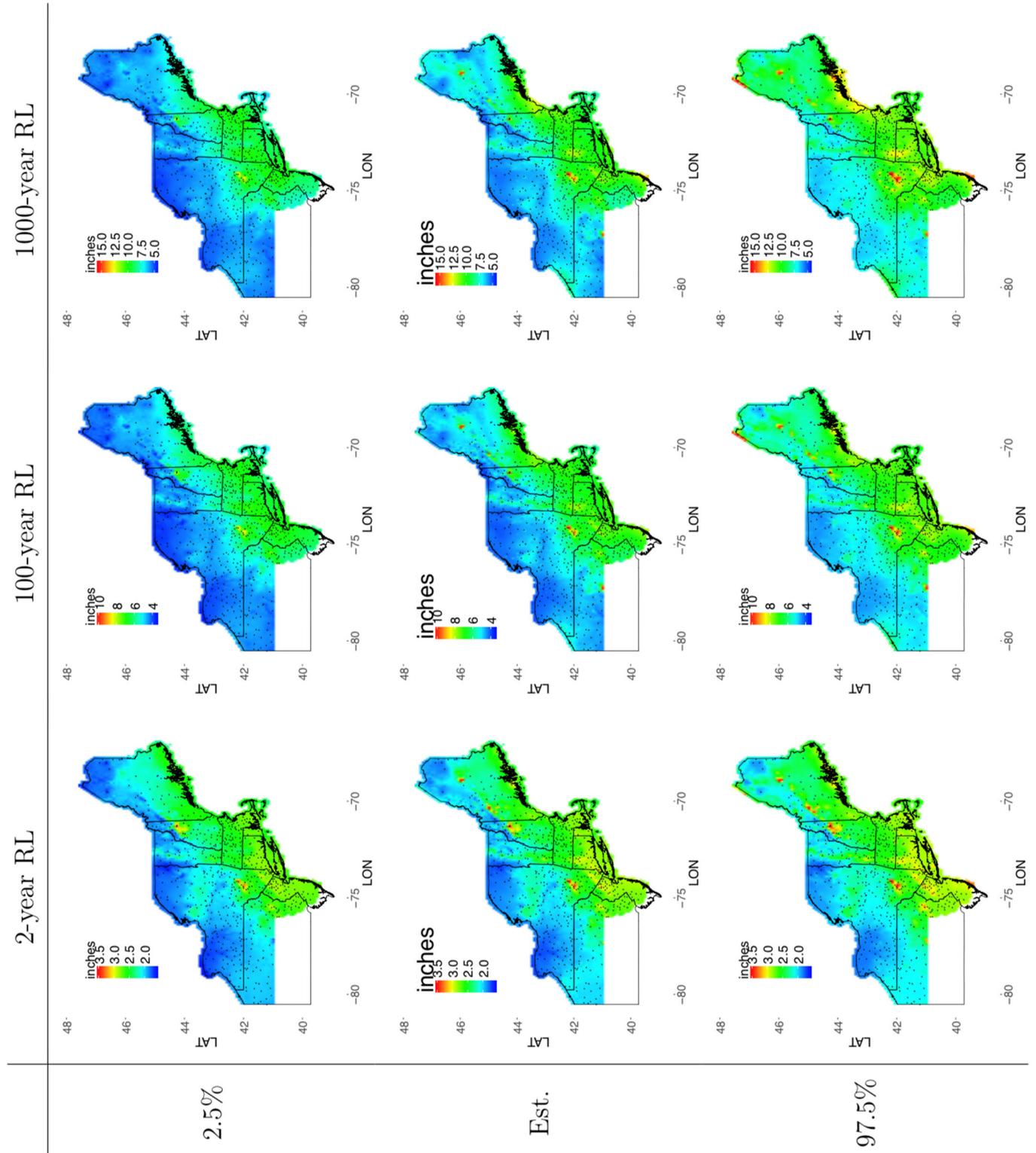


Figure 6-3 Pointwise local likelihood estimates and 95% bootstrap confidence intervals of 2, 100, and 1000-year daily accumulation return-levels over the pilot study region for GEV fit to AMS data. The overall coastal-inland spatial pattern for hourly data is similar, with pockets of extreme precipitation in the Catskill Mountains and Mount Washington. Black points correspond to gauge locations.

APPENDIX A-1. The Pennsylvania State Final Report

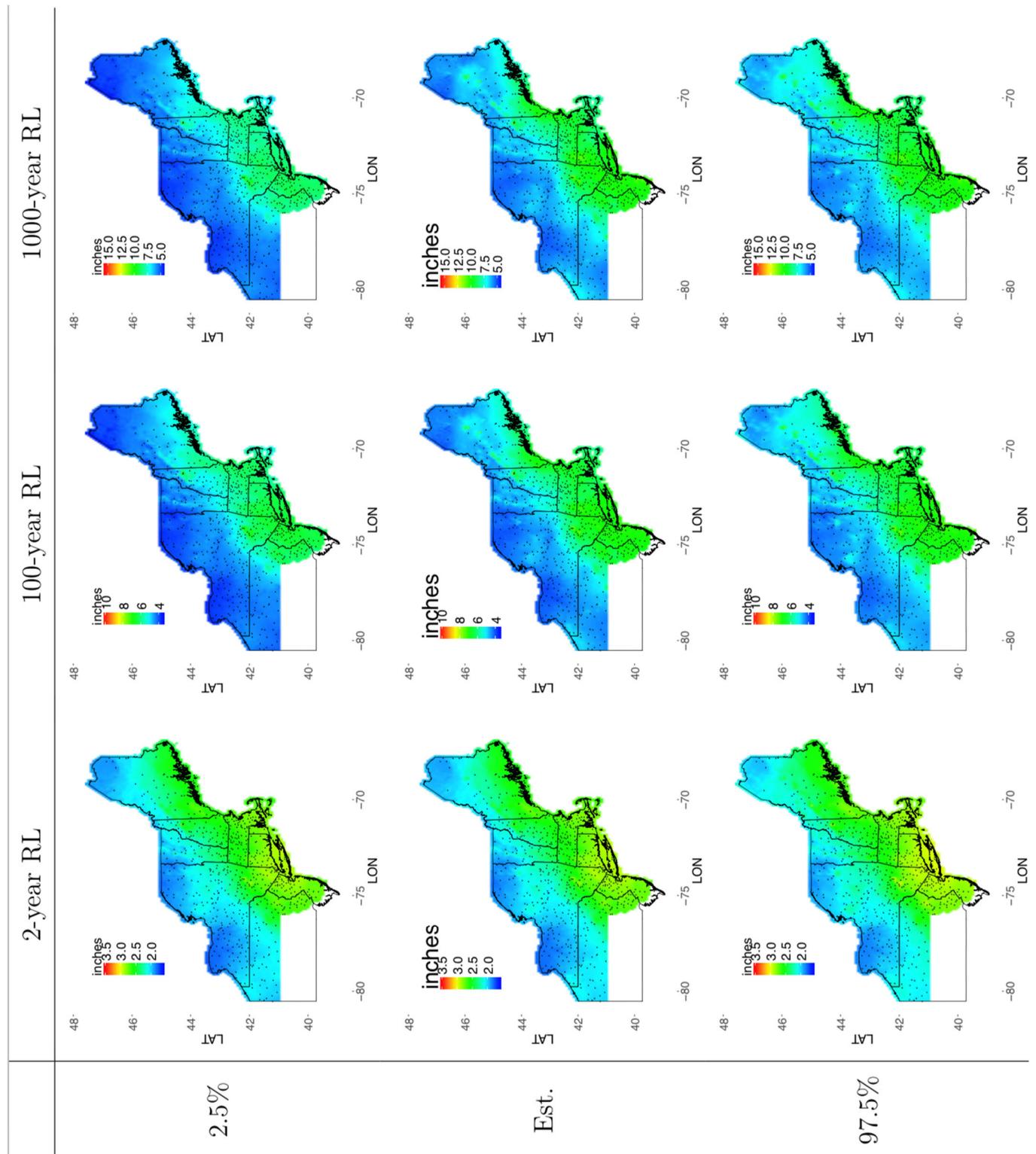


Figure 6-4 Pointwise local likelihood estimates and 95% bootstrap confidence intervals of 2, 100, and 1000-year daily accumulation return-levels over the pilot study region for GPD fit to PDS data. The overall coastal-inland spatial pattern for hourly data is similar, with pockets of extreme precipitation in the Catskill Mountains and Mount Washington. Black points correspond to gauge locations.

APPENDIX A-1. The Pennsylvania State Final Report

6.3 Incorporating climate model information

Rather than include time as a covariate in the model directly, climate model output can be used to incorporate the relationships between precipitation and other climate factors as well as make projections based on different climate scenarios. Just as in ordinary linear regression, extrapolation of linear trends beyond the range of observed data is suspect at best. Doing so is especially troubling for heavy tailed distributions, as small changes in the parameters can lead to wildly different return-level estimates. Precipitation and other atmospheric phenomena are subject to strong, often non-linear, interdependencies. In the presence of a changing climate, forecasting changes in the distribution of extreme precipitation using a linear time trend ignores these interdependencies. Climate model approximations of these dynamics provide a more robust means of projecting atmospheric conditions. By relating the distribution of extreme precipitation to the output of a climate model, the non-stationary signal captured by the climate model can be leveraged to make more reliable forecasts of extreme precipitation.

As an example, we can use non-stationary climate model forecasts of temperature to predict changes in extreme precipitation. There are strong physical justifications for why temperature is a good predictor of precipitation. Climate models run under different CO_2 emission scenarios can be used to forecast temperature in time, which itself may be changing non-linearly. By relating temperature to the distribution of extreme precipitation, forecasts of future temperatures from a given climate model run can be used as inputs to predict the distributions of extreme precipitation under a given climate regime. That is, predictions of the distribution of extreme precipitation are free to vary non-linearly in time in a way that is rooted to the atmospheric dynamics captured by the climate model.

In this section, we demonstrate how this can be done, but do not wish to suggest a particular climate model or set of climate model factors that should be used. The climate model data used for this

APPENDIX A-1. The Pennsylvania State Final Report

illustration come from the C20C+ Detection and Attribution project

(<http://portal.nersc.gov/project/m1517/data/LBNL/CAM5-1-2-025degree/All-Hist/est1/v1-0/mon/atmos/tas/run001/>) As an example, we use climate model output of mean annual temperature

(MAT) reanalysis on a 25 km grid, averaged over the period 1996-2005. The mismatch in the grids of the temperature data and precipitation grids was resolved by assigning each observation and prediction location a mean annual temperature from the nearest center on the climate model raster. This results in a single MAT covariate value for each prediction and spatial location for all time points. For this model, now we let the location and scale parameters be a function of both MAP and MAT and use an intercept only model for the shape. That is, for both the GEV and GPD models, the parameters take the form

$$\mu(X) = \alpha_0 + \alpha_1 MAP + \alpha_2 MAT$$

$$\log(\sigma(X)) = \beta_0 + \beta_1 MAP + \beta_2 MAT$$

To illustrate the effect of each covariate for the GEV model, spatial maps of the estimated parameters are shown in Figure 6-6 (GPD maps are similar, but not shown for brevity). Fitted MAP parameters have the most spatial variation for both the location and log(scale) parameters, whereas the climate model MAT covariate appears to have little effect on the distribution of extreme precipitation. As such, we would advocate for also considering other climate model variables in practice.

APPENDIX A-1. The Pennsylvania State Final Report

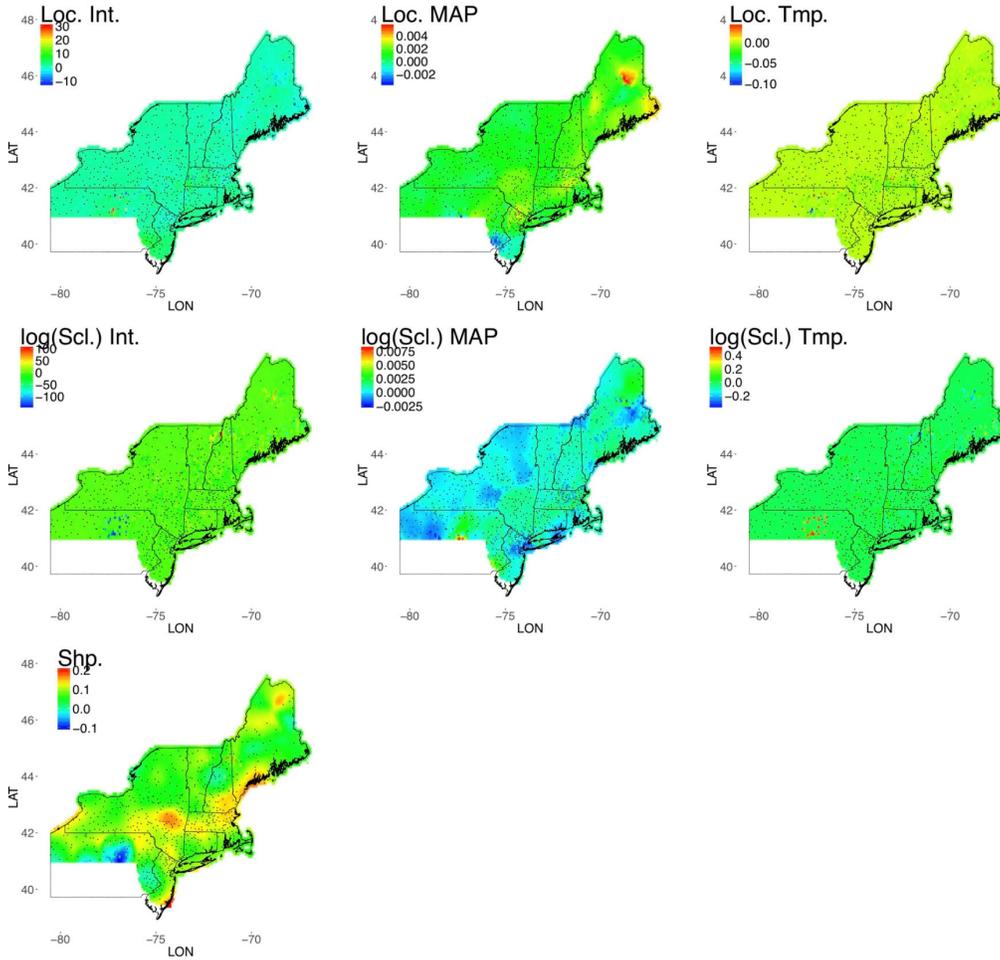


Figure 6-5 Spatial maps of local likelihood estimates of GEV model parameters.

To make projections, we can apply climate model output covariates based on future climate scenarios to our fitted model. For illustration, we use MAT based on a stationary climate run over a 10-year future run. Note that here we use historical MAP, but in practice, one would replace these with climate model projections of MAP to reflect different climate scenarios. The historical and future 100-year return-level estimates from this example model are shown in Figure 6-7. As expected, since mean annual temperature did not have a large effect in our fitted model and the future climate scenario was based on stationary conditions, there are not major differences between the historical projections and future ones. We emphasize that these return-level projections given in this section not intended to be

APPENDIX A-1. The Pennsylvania State Final Report

accurate, but merely to suggest an approach to incorporate climate model information in a spatial model for extremes. More thought should be given to which climate model variables to use, the functional form they should take in relating them to the parameters of extreme value distributions, and which future climate scenarios to use.

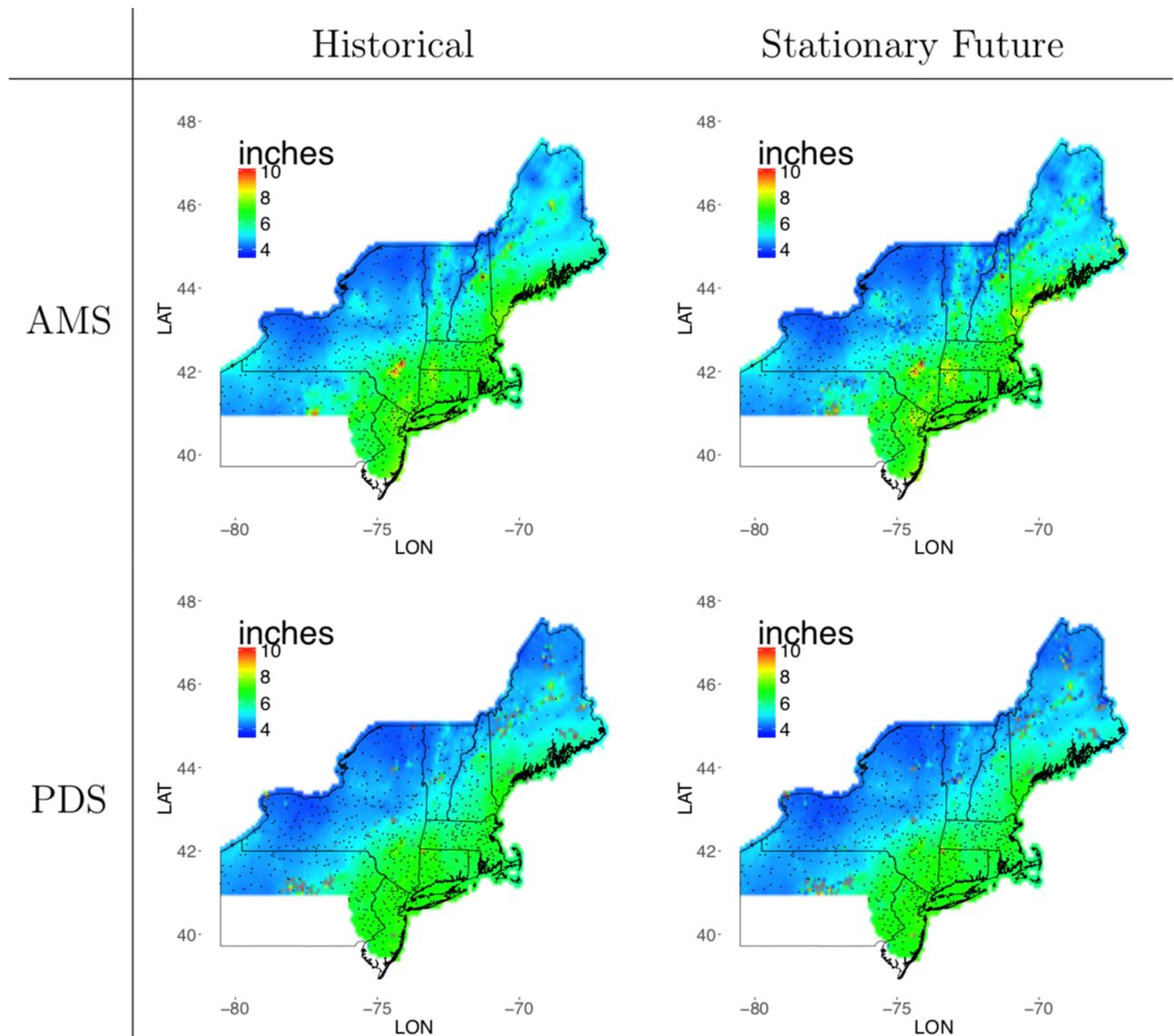


Figure 6-6 Historical and future 100-year return-level estimates based on a stationary climate for GEV (AMS) and GPD (PDS) models. Grey points correspond to locations where the optimization algorithm failed to converge.

APPENDIX A-1. The Pennsylvania State Final Report

7 Summary and recommendations

The aim of this study is threefold: (1) evaluate practical differences between AMS and PDS approaches, (2) examine the differences among the common estimators for the GEV and GPD parameters, and (3) to propose a flexible model for non-stationarity in the distribution of extremes over time.

To evaluate fixed threshold PDS approaches, a high threshold must be selected for which the GPD approximation is reasonable for the distribution of exceedances. We illustrated how a battery of threshold selection tools can be used to achieve this end and found that for the majority of stations in the pilot study region, the 98th percentile of non-zero observations appeared to be an adequate threshold. Tail inference for both the GEV and GP distributions can be highly sensitive to estimation of the shape parameter. A simulation study comparing the RMSE of predicted quantiles showed that the three most common estimators - MLE, GMLE, and LM – exhibit comparable performance over the range of shape parameters most commonly observed for precipitation data. Due to the additional flexibility of the MLE and GMLE to easily incorporate covariates, we advocate for using these approaches in this context. When applied to the pilot study data, the models for AMS and PDS data gave nearly identical predictions in most cases for all three estimation methods. This behavior is expected, as the GEV is the asymptotic distribution for renormalized maxima over an increasing block size exactly when the GPD is the limiting distribution for threshold exceedances (as the threshold goes to the upper bound of the underlying distribution) (Leadbetter and Rootzen 1983). We compare three models for non-stationarity, where the parameters of the GEV and GPD are smoothly varying functions of time using splines. Modeling the shape parameter as a smoothly varying function of time lends itself to overfitting the data, and so we advocate fixing the shape parameter in time. Finally, to predict return-levels at observed and unobserved spatial locations, we propose using a parallelizable local likelihood method that can easily be applied to many prediction locations and can be used for either GPD or GEV models. The GEV model

APPENDIX A-1. The Pennsylvania State Final Report

gave slightly more stable predictions, which aligned closely with existing NOAA Atlas 14 predictions, especially in the presence of data quantization, giving it an edge over the GPD model for this application.

8 Appendix

8.1 Quantile Scores

A quantile score is a quantity that can be used to measure the discrepancy between the observed and predicted upper tail of a distribution (Gneiting and Raftery, 2007). The quantile score for the τ th quantile is defined by

$$S^\tau(x) = 2[I\{x < \hat{q}(\tau)\} - \tau][\hat{q}(\tau) - x],$$

where $\hat{q}(\tau)$ is the predicted τ th quantile, and $I\{x < y\}$ is an indicator for the event that the observation x is less than y

$$I\{x < y\} = \begin{cases} 1, & \text{if } x < y \\ 0, & \text{if } x \geq y \end{cases}.$$

In practice, a quantile score estimate is obtained by averaging the above quantity over a collection of holdout samples x_1, \dots, x_n . Smaller quantile scores correspond to better predictive performance. Note that the expected quantile score is zero when $\hat{q}(\tau)$ is the τ th quantile of the distribution of X . Moreover, models that capture the true τ th quantile will have quantile scores which are no greater than those which do not.

8.2 GMLE Penalty Function

In this section, we illustrate the effect of the tuning parameters of the shifted Beta distribution used to penalize the shape parameter in the GMLE estimation method. Recall the Beta density is of the form

APPENDIX A-1. The Pennsylvania State Final Report

$$\pi(\xi) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} (0.5 + \xi)^{p-1} (0.5 - \xi)^{q-1}, \quad \xi \in [-0.5, 0.5].$$

The shift by 0.5 centers the density at 0, as opposed to the more typical form of the Beta distribution, which is centered at 0.5. The parameters p and q are shape parameters which give more weight to different regions of the distribution. For this application, we use a symmetric penalty around zero, and so we only consider the case when they are both equal. In this case, increasing p and q , concentrates more mass around 0, whereas $p = q = 1$, corresponds to a uniform distribution on $[-0.5, 0.5]$. The more concentrated the beta density is around zero, the stronger the penalization effect is on the shape parameter. See Figure 8-1 for an illustration of the shifted Beta density for different parameters. Penalty parameter for the GMLE models were chosen to minimize the quantile scores ($\tau = 0.98$) for the daily precipitation data. The selected penalty for both GPD and GEV models (stationary) was $p = q = 2$.

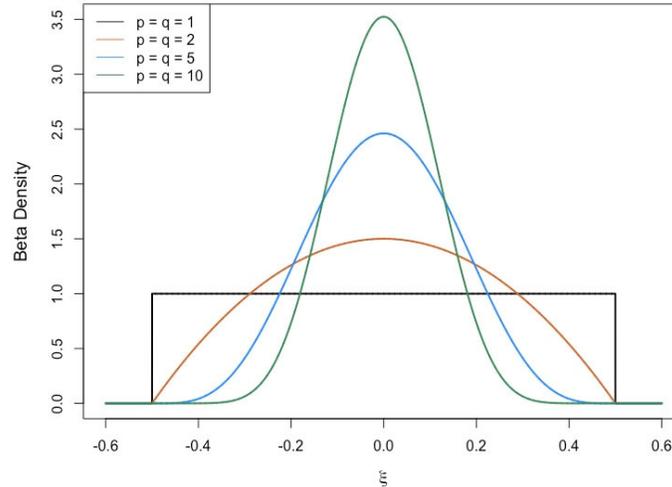


Figure 8-1 GMLE penalization function (Beta density) under different parameter settings. Functions which are more concentrated around zero impose stronger penalties on the shape parameter.

APPENDIX A-1. The Pennsylvania State Final Report

8.3 Trend Detection Simulation Study

The nonparametric Mann-Kendall test is commonly used to test non-stationarity of hydro-climatic time series (Cheng et al., 2014; Wi et al., 2016). A challenge with trend analysis is that test results can depend strongly on the chosen EVD (Wi et al., 2016). In addition to the Mann-Kendall test we consider two other tests for trends for extreme rain, a Wald test for linear trend in the mean number of threshold exceedances per year and a likelihood ratio test (LRT) for a linear trend in the GEV location parameter. Let N_t be the number of exceedances of a high threshold in year t . Modeling $N_t \sim \text{Poisson}(\lambda_t)$ where the annual mean λ_t is modeled through link function $\log(\lambda_t) = \alpha_0 + \alpha_1 t$. The usual GLM Wald test is applied for testing the hypothesis $H_a: \alpha_1 \neq 0$. To test for a trend in the location of a GEV, letting Y_t denote the precipitation annual maximum for year t , an LRT is applied to test $H_a: \beta_1 \neq 0$ for $Y_t \sim \text{GEV}(\beta_0 + \beta_1 t, \sigma, \xi)$. These three tests were compared when the data were generated under the following conditions: $n \times 365$ observations were simulated for varying record lengths of $n=20$ and 100 years from the model $Y_{it} \sim \text{GEV}(\beta t, 1, 0)$, $t = 1, \dots, n$ denotes the year, and $i=1, \dots, 365$ denotes the day of the year. AMS, PDS series, and the number of exceedances of a high threshold (0.95 empirical quantile) were constructed for each simulated dataset. Non-parametric Mann-Kendall tests were done on the AMS, and PDS series, the Poisson Wald test on the number of exceedances per year, and the LRT test on the AMS series. The linear trend term β varied from 0 to 0.1. All tests were performed at level 0.05. The estimated probability of rejection for each test based on 10,000 Monte Carlo replicates is plotted in Figure 8-2. The Mann-Kendall test applied to AMS and PDS series appear to have lower probability of detection than the other two tests in this scenario. The Mann-Kendall test applied to the number of exceedances per year is comparable to the Poisson Wald test for a change in the rate of exceedance, and aims to detect the same feature, but the Poisson Wald test is for a linear change in the mean, whereas the Mann-Kendall is for general monotonic trends.

APPENDIX A-1. The Pennsylvania State Final Report

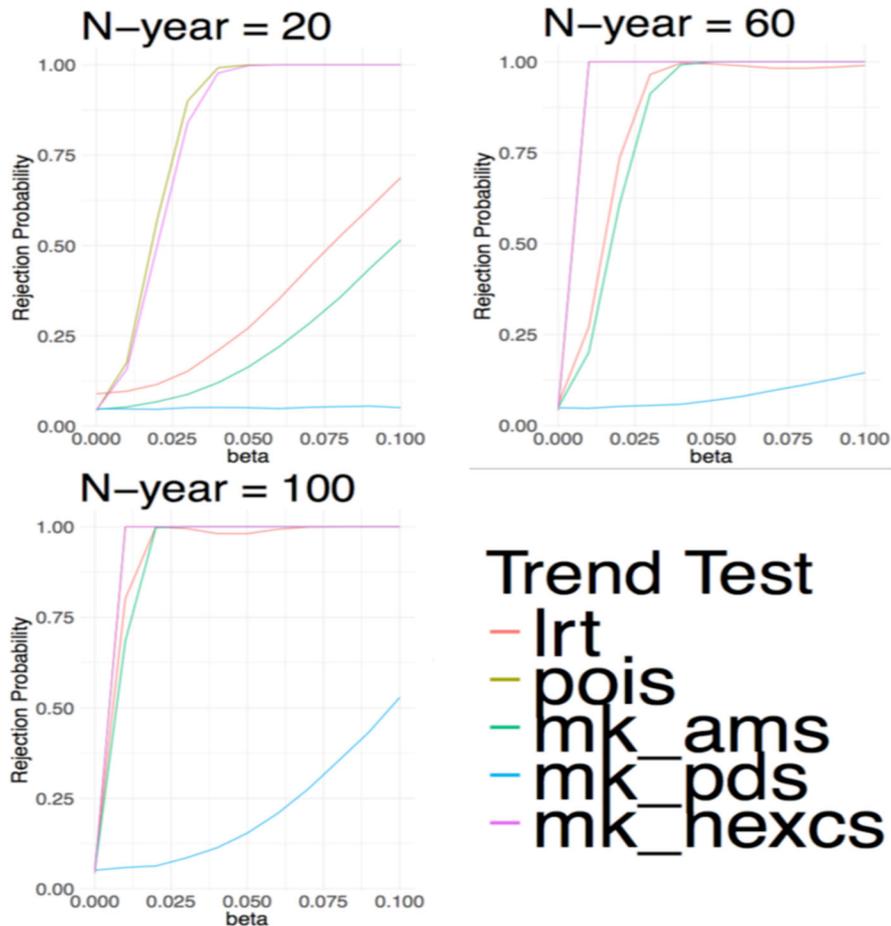


Figure 8-2 Power curves for the LRT (red), Poisson GLM Wald test (gold), Mann-Kendall on the mean number of exceedances per year (purple), Mann-Kendall AMS (Green), Mann-Kendall PDS (Blue). Note that in the N-year = 60 and 100 cases, the pink and gold lines overlap, and the gold line is obscured. All tests appear to be properly calibrated (probability of rejection for $\beta=0$ case is close to the 0.05 level).

9 Acknowledgements

This report was supported by FHWA grant Z17-20337. This research used science gateway resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Computations for this research were performed on the Pennsylvania State University's Institute for CyberScience Advanced CyberInfrastructure (ICS-ACI). This content is solely the responsibility of the authors and does not necessarily represent the views of the Institute for CyberScience.

APPENDIX A-1. The Pennsylvania State Final Report

References

- Bader, B., J. Yan, and X. Zhang (2016). Automated threshold selection for extreme value analysis via goodness-of-fit tests with application to batched return level mapping. [arXiv preprint arXiv:1604.02024](#).
- Balkema, A. A. and L. de Haan (1974). Residual life time at great age. *Ann. Probability* **2**, 792–804.
- de Zea Bermudez, P., & Kotz, S. (2010). Parameter estimation of the generalized Pareto distribution—Part I. *Journal of Statistical Planning and Inference*, **140**(6), 1353-1373.
- Bortot, P. and C. Gaetan (2014). A latent process model for temporal extremes. *Scand. J. Stat.* **41**(3), 606–621.
- Bortot, P. and C. Gaetan (2016). Latent process modelling of threshold exceedances in hourly rainfall series. *J. Agric. Biol. Environ. Stat.* **21**(3), 531–547.
- Caeiro, F. and M. I. Gomes (2016). Threshold selection in extreme value analysis. In *Extreme value modeling and risk analysis*, pp. 69–86. CRC Press, Boca Raton, FL.
- Casella, G. and R. L. Berger (2002). *Statistical inference*, Volume 2. Duxbury Pacific Grove, CA.
- Chavez-Demoulin, V. and A. Davison (2012). Modelling time series extremes. *REVSTAT-Statistical Journal* **10**(EPFL-ARTICLE-180506), 109–133.
- Cheng, L., AghaKouchak, A., Gilleland, E., & Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic change*, **127**(2), 353-369.
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London, Ltd., London.
- Coles, S. G. and M. J. Dixon (1999). Likelihood-based inference for extreme value models. *Extremes* **2**(1), 5–23.
- Davison, A. C., S. A. Padoan, and M. Ribatet (2012). Statistical modeling of spatial extremes. *Statistical science*, 161–186.
- Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B* **52**(3), 393–442. With discussion and a reply by the authors.
- Deidda, R. and M. Puliga (2006). Sensitivity of goodness-of-fit statistics to rainfall data rounding off. *Physics and Chemistry of the Earth, Parts A/B/C* **31**(18), 1240–1251.
- Drees, H., L. de Haan, and S. Resnick (2000). How to make a Hill plot. *Ann. Statist.* **28**(1), 254–274.
- Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press.
- Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, 89–102.

APPENDIX A-1. The Pennsylvania State Final Report

- Ferreira, A., & De Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. The Annals of statistics, 43(1), 276-298.
- Ferro, C. A. T. and J. Segers (2003). Inference for clusters of extreme values. J. R. Stat. Soc. Ser. B Stat. Methodol. 65(2), 545–556.
- Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In Mathematical Proceedings of the Cambridge Philosophical Society, Volume 24, pp. 180–190. Cambridge University Press.
- Gilleland, E. and R. W. Katz (2016). extRemes 2.0: An extreme value analysis package in R. Journal of Statistical Software 72(8), 1–39.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. Annals of mathematics, 423–453.
- Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. J. Amer. Statist. Assoc. 102(477), 359–378.
- Gomes, M. I. and O. Oliveira (2001). The bootstrap methodology in statistics of extremes— choice of the optimal sample fraction. Extremes 4(4), 331–358 (2002).
- Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. J. Multivariate Anal. 32(2), 177–203.
- Hall, W. J. and J. A. Wellner (1981). Estimation of mean residual life. Unpublished.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. Ann. Statist. 3(5), 1163–1174.
- Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society. Series B (Methodological) 52(1), 105–124.
- Hosking, J. R. M. and J. R. Wallis (1987). Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29(3), 339–349.
- Hosking, J. R. M., J. R. Wallis, and E. F. Wood (1985). Estimation of the generalized extreme- value distribution by the method of probability-weighted moments. Technometrics 27(3), 251–261.
- Jin, M. and J. R. Stedinger (1989). Flood frequency analysis with regional and historical information. Water Resources Research 25(5), 925–936.
- Katz, R. W. (2010). Statistics of extremes in climate change. Climatic Change 100(1), 71–76.
- Katz, R. W., G. S. Brush, and M. B. Parlange (2005). Statistics of extremes: Modeling ecological disturbances. Ecology 86(5), 1124–1134.
- Katz, R. W., P. F. Craigmile, P. Guttorp, M. Haran, B. Sanso, and M. L. Stein (2013). Uncertainty analysis in climate change assessments. Nature Clim. Change 3(9), 769–771.

APPENDIX A-1. The Pennsylvania State Final Report

- Katz, R. W., M. B. Parlange, and P. Naveau (2002). Statistics of extremes in hydrology. Advances in Water Resources 25(8-12), 1287–1304.
- Kendall, M. G. (1938). A new measure of rank correlation. Biometrika 30(1/2), 81–93.
- Kysely, J. (2008). A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models. Journal of Applied Meteorology and Climatology 47(12), 3236-3251.
- Kysely, J., & Picek, J. (2007). Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic. Climate research, 33(3), 243-255.
- Landwehr, J. M., N. Matalas, and J. Wallis (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resources Research 15(5), 1055-1064.
- Leadbetter, M. and H. Rootzen (1983). Extremal theory for stochastic processes. The Annals of Probability, 431–478.
- Lee, B. S., Haran, M., & Keller, K. (2017). Multidecadal Scale Detection Time for Potentially Increasing Atlantic Storm Surges in a Warming Climate. Geophysical Research Letters, 44(20).
- Lehmann, E. (1983). Theory of point estimation, New York: Wiley and Sons. Theory of Point Estimation.
- Mann, H. B. (1945). Nonparametric tests against trend. Econometrica: Journal of the Econometric Society, 245–259.
- Martins, E. S. and J. R. Stedinger (2000). Generalized maximum likelihood pareto-poisson estimators for partial duration series. Water Resources Research 37(10), 2551–2557.
- Martins, E. S. and J. R. Stedinger (2001). Historical information in a generalized maximum likelihood framework with partial duration and annual maximum series. Water Resources Research 37(10), 2559–2567.
- McCullagh, P. (1984). Generalized linear models. European Journal of Operational Research 16(3), 285–292.
- Perica, S., D. Martin, S. Pavlovic, I. Roy, M. St. Laurent, C. Trypaluk, D. Unruh, M. Yekta, G. Bonnin (2013b). NOAA Atlas 14 Volume 9, Precipitation-Frequency Atlas of the United States, Southeastern States. NOAA, National Weather Service, Silver Spring, MD.
- Pickands, III, J. (1975). Statistical inference using extreme order statistics. Ann. Statist. 3, 119–131.
- PRISM Climate Group, Oregon State University, <http://prism.oregonstate.edu>, created 11 Jan 2018.
- Reich, B. J., B. A. Shaby, and D. Cooley (2014). A hierarchical model for serially-dependent extremes: a study of heat waves in the western US. J. Agric. Biol. Environ. Stat. 19(1), 119–135.
- Resnick, S. I. (1987). Extreme values, regular variation and point processes. Springer.
- Rootzén, H., & Katz, R. W. (2013). Design life level: quantifying risk in a changing climate. Water Resources

APPENDIX A-1. The Pennsylvania State Final Report

Research, 49(9), 5964-5972.

Scarrott, C. and A. MacDonald (2012). A review of extreme value threshold estimation and uncertainty quantification. REVSTAT 10(1), 33–60.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. Biometrika 72(1), 67–90.

Wang, Y. (2011). Smoothing splines: methods and applications. CRC Press.

Wi, S., J. B. Valdés, S. Steinschneider, and T.-W. Kim (2016), Non-stationary frequency analysis of extreme precipitation in South Korea using peaks-over-threshold and annual maxima, Stochastic Environmental Research and Risk Assessment, 30(2), 583-606.

Yang, G. L. (1978). Estimation of a biometric function. Ann. Statist. 6(1), 112–116

Methodology for Frequency Analysis of Nonstationary Precipitation Data

Momcilo Markus, Kevin Grady, and James Angel
Illinois State Water Survey
Prairie Research Institute
University of Illinois at Urbana-Champaign

Shu Wu, Fuyao Wang, and David Lorenz
Center for Climatic Research
Nelson Institute for Environmental Studies
University of Wisconsin-Madison

Seid Koric
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

May 2021

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table of Contents

Summary	A-2.5
Acknowledgments.....	A-2.7
SECTION I. Climate models	A-2.8
1.1. Introduction.....	A-2.8
1.2. Models and downscaling methods.....	A-2.9
1.2.1. Models and CMIP Executive s	A-2.9
1.2.2. Statistical and dynamical downscaling.....	A-2.11
1.3. Downscaled data sets used in this study	A-2.13
1.3.1. Dynamical downscaled data set: NA-CORDEX	A-2.13
1.3.2. CMIP5 LOCA.....	A-2.15
1.3.3. CMIP5 BCCAv2.....	A-2.15
1.3.4. CMIP5 UWPD	A-2.16
1.4. Fourth National Climate Assessment.....	A-2.17
SECTION II. Evaluation of the downscaled climate modeling data for the historical period based on extreme time series gauge data and their basic statistics	A-2.18
2.1. Results for Northeast U.S.	A-2.18
2.2. Results for Texas.....	A-2.23
2.3. Conclusions.....	A-2.27
SECTION III. Evaluation of the downscaled climate modeling daily data for the historical period 1960-2005: Comparison between precipitation frequency using observations and climate modeling downscaled data for annual maximum series (AMS) and partial duration series (PDS) methods	A-2.28
3.1. The annual maximum series (AMS) and the partial duration series (PDS).....	A-2.28
3.2. Downscaled data's ability in simulating observed PFs.....	A-2.30
3.2.1 Results for UWPD downscaled data.....	A-2.30
3.2.2 Results for LOCA downscaled data.....	A-2.40
3.2.3 Results for NA-CORDEX.....	A-2.45
3.2.4 Results for BCCAv2	A-2.49
3.3. Conclusions.....	A-2.54

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

SECTION IV. Quasi-stationary (QS) methodology and the spatial variability of the future QS precipitation frequency estimates based on the UWPD data	A-2.57
4.1. Quasi-stationary (QS) PFs estimation methodology.....	A-2.57
4.2. Future AMS and PDS QS PFs estimated based on UWPD downscaled data for 2006-2100 A-2.58	
4.2.1. Results for RCP4.5 scenario	A-2.58
4.2.2. Results for RCP8.5 scenario	A-2.64
4.2.3. Spatial distributions of projected changes for RCP4.5 and RCP8.5.....	A-2.68
4.3. Spatial distribution of future QS PFs estimated based on UWPD downscaled data for 2006-2053 and 2054-2100 for 25-year return period and a comparison with NCA4	A-2.71
4.4. Summary	A-2.75
SECTION V. Comparisons between nonstationary and quasi-stationary precipitation frequency estimates for AMS and PDS methods, LOCA and UWPD data sets, and RCP4.5 and RCP8.5 scenarios	A-2.76
5.1. Nonstationary and quasi-stationary approaches.....	A-2.76
5.2. Summary of the nonstationary results for AMS and daily data.....	A-2.77
5.2.1. The Penn State local likelihood nonstationary method (NS model) applied to raw CMIP5 data (NS-CMIP5)	A-2.79
5.2.2. NS-LOCA model	A-2.82
5.2.3. NS-UWPD model	A-2.85
5.3. Summary of quasi-stationary runs for AMS and daily data	A-2.89
5.3.1. LOCA-based runs	A-2.89
5.3.2. UWPD-based runs	A-2.92
5.4. Summary of the nonstationary and quasi-stationary results for PDS and daily data.....	A-2.94
5.4.1. Nonstationary results for PDS	A-2.95
5.4.2. Quasi-stationary results for PDS.....	A-2.97
5.5. Comparison of all methods for daily data.....	A-2.101
5.5.1. Nonstationary and quasi-stationary comparisons for 100-year return period: AMS vs. PDS, LOCA vs. UWPD, RCP4.5 vs. RCP8.5	A-2.101
5.5.2. Quasi-stationary comparisons for AMS-based method: 2-, 100-, and 1,000-year, LOCA vs. UWPD, RCP4.5 vs. RCP8.5, including confidence limits.....	A-2.103
5.6. Conclusions for daily data	A-2.106
5.7. Ten-day results.....	A-2.107
5.7.1. Nonstationary results for AMS	A-2.107
Results of NS-LOCA run.....	A-2.107
Results of NS-UWPD run.....	A-2.108

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.7.2. Quasi-stationary results for AMS	A-2.110
Results of QS-LOCA runs	A-2.110
Results of QS-UWPD run.....	A-2.112
5.7.3. Nonstationary results for PDS	A-2.113
Results of NS-LOCA run.....	A-2.113
Results of NS-UWPD run.....	A-2.115
5.7.4. Quasi-stationary results for PDS.....	A-2.116
Results of QS-LOCA run.....	A-2.116
Results of QS-UWPD run.....	A-2.118
5.7.5. Comparison of all methods	A-2.119
5.8. Hourly results.....	A-2.121
5.8.1. Nonstationary results for AMS	A-2.121
Results of NS-LOCA run.....	A-2.121
Results of NS-UWPD run.....	A-2.122
5.8.2. Quasi-stationary results for AMS	A-2.124
Results of QS-UWPD run.....	A-2.124
5.8.3. Nonstationary results for PDS	A-2.126
Results of NS-LOCA run.....	A-2.126
Results of NS-UWPD run.....	A-2.127
5.8.4. Quasi-stationary results for PDS.....	A-2.129
Results of QS-UWPD run.....	A-2.129
5.8.5. Comparison of all methods	A-2.130
SECTION VI. Conclusions and recommendations	A-2.132
APPENDIX I. Assessment of the benefits of climate model weights for ensemble analysis in three urban precipitation frequency studies	A-2.134
I.1. Introduction	A-2.134
I.2. Data.....	A-2.138
I.3. Methodology.....	A-2.141
I.4. Results	A-2.144
I.5. Conclusions	A-2.153
References.....	A-2.154

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Executive Summary

Traditional NOAA Atlas 14 precipitation frequency analysis methods assume the heavy precipitation data used in the analysis are stationary. However, observational data indicate that the intensity and frequency of heavy precipitation have been increasing in various parts of the United States. As a result, the Office of Water Prediction, Hydrometeorological Design Studies Center (HDSC), which is responsible for producing NOAA Atlas 14 estimates, is examining the suitability of various state-of-the-art methodologies for frequency analysis of nonstationary precipitation. Considering that many climate models indicate the increasing trend in intensity and frequency of heavy precipitation will likely continue in the future, using only statistics of the past observed precipitation could underestimate precipitation frequency quantiles and their confidence limits. This report describes a framework to translate future climate scenarios efficiently and accurately into a product that will be useful for the nonstationary precipitation frequency analysis methods in future NOAA Atlas 14 projects.

This report first briefly introduces the Coupled Model Intercomparison Project Phase 5 (CMIP5) models, followed by an overview of statistical and dynamical downscaling methods. Next, the four existing downscaled data sets (NA-CORDEX, LOCA, BCCAv2, and University of Wisconsin UWPD data) that were evaluated through a comparison with the observed data are discussed. In particular, climate model-based and observed data-based precipitation frequency (PF) quantiles for the Northeast and Texas project regions were compared using the annual maximum series (AMS) and partial duration series (PDS) approaches. The UWPD and LOCA climate data sets produced PFs closest to those based on the observed data and were considered more accurate than the other sources.

Two main approaches were used in studying precipitation frequency in this report. One approach divides the time period into two or more sub-periods and treats each as stationary. This piecewise stationary method is referred to as the quasi-stationary (QS) approach. The other, which is the nonstationary (NS) approach, expresses precipitation frequency distribution statistics as a function of annual precipitation, time, or other covariates. The QS method was developed by researchers at the University of Illinois at Urbana-Champaign and University of Wisconsin-Madison. The primary NS approach used in this study was originally developed by

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Penn State researchers but was later modified by the HDSC group to use station precipitation and elevation as covariates.

The QS approach was suggested in this study and compared with the NS approach. The confidence limits for the QS method were estimated based on ensembles of model runs. A part of this research was dedicated to an assessment of the benefits of model weights in an ensemble of models. Although these results were not applied in the final stage, the initial use of model weights was shown to be very promising and is recommended for future considerations.

For easy comparisons, most of the results were presented as regional averages, given the large spread of model results and relatively large spatial variability. Although the results vary depending on the method, the selection of peaks (AMS vs. PDS), event duration, recurrence interval, and location, the following general conclusions were drawn from the results. The results demonstrated that the QS method produces estimates approximately two times larger than those from the NS method. The projected precipitation frequency increases based on the QS method are more consistent with those published in the recent Fourth National Climate Assessment (NCA4).

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Acknowledgments

The authors would like to thank Sanja Perica, Mark Glaudemans, Sandra Pavlovic and Michael St. Laurent, from the Office of Water Prediction, Hydrometeorological Design Studies Center (HDSC), for their continuous support throughout the project, and for sharing their expertise, reviewing the results, and providing insightful suggestions. Lisa Sheppard (ISWS) provided editorial assistance. In addition, the authors would like to thank the National Center for Supercomputing Applications (NCSA) Industry Program at the University of Illinois and its Blue Waters project, funded by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois for hardware and application support.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

SECTION I. Climate models

1.1. Introduction

Traditional precipitation frequency estimation methods assume the heavy precipitation data are stationary. However, the observed data indicate that the intensity and frequency of heavy precipitation have been increasing in various parts of the United States. Climate models, which are the primary tools available to investigate the response of the climate system to various forcings for making climate projections over the coming century and beyond, suggest that the increasing trend in intensity and frequency of heavy precipitation will likely continue in the future (Easterling et al., 2017), leading to the examination of possible nonstationary precipitation frequency analysis. As there are no observations for the future, the climate model data sets produced by the global community of climate modelers organized by the World Climate Research Program's (WCRP's) Coupled Model Intercomparison Project (CMIP) have become valuable for studying future rainfall. It should also be noted that climate model outputs may not be directly applicable for precipitation frequency analysis given their coarse resolutions. To produce a finer resolution, another layer of downscaling is typically added before attempting any comparison with observations or interpretation of future precipitation frequency. Therefore, using the climate model output for nonstationary precipitation analysis should incorporate the knowledge gleaned from climate models, downscaling schemes, and model performances or evaluations.

This section presents a brief introduction to the Coupled Model Intercomparison Project Phase 5 (CMIP5) models and existing downscaling methods including statistical and dynamical downscaling approaches. We further evaluate the four existing downscaled data sets selected for this study. Those data sets include NA-CORDEX, LOCA, BCCAv2, and a new data set created for this study, the University of Wisconsin probabilistic downscaling (UWPD) data set.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

1.2. Models and downscaling methods

1.2.1. Models and CMIPs

Climate models or Earth system models are based on well-established physical principles and have been demonstrated to reproduce observed features of recent and past climate changes. There is considerable confidence that atmosphere-ocean general circulation models (AOGCMs) provide credible quantitative estimates of future climate change, particularly at the continental and larger scales. Fig. 1.1 depicts the evolution of climate models and major physical processes added to them over the years. The FAR, SAR, TAR, and AR4 labels on the diagrams indicate the models in use at the times of each of the four Intergovernmental Panel on Climate Change Assessment Reports (IPCC 2007). The 1970s models were relatively simple. They used just a few key features such as incoming sunlight, rainfall, and CO₂ concentration to represent Earth's climate system. The 1980s models became more sophisticated, incorporating clouds, land surface features, ice, and other elements into their calculations. Around the 1990s, simple oceans were included in models. The First Assessment Report (FAR) by IPCC was based mainly on the output of these models. Later models included more complex representations of oceans as well as more complex clouds, a broader range of atmospheric constituents, such as sulfates and aerosols, and atmospheric chemistry, vegetation that exchanges gases with the atmosphere, and other components. The latest version AR5 climate models has included the biogeochemical processes. Therefore, it may be more accurate to consider them Earth system models. Along with the complexity, the model resolutions also keep increasing (Fig.1.1), making it possible and practical to study the regional scale.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

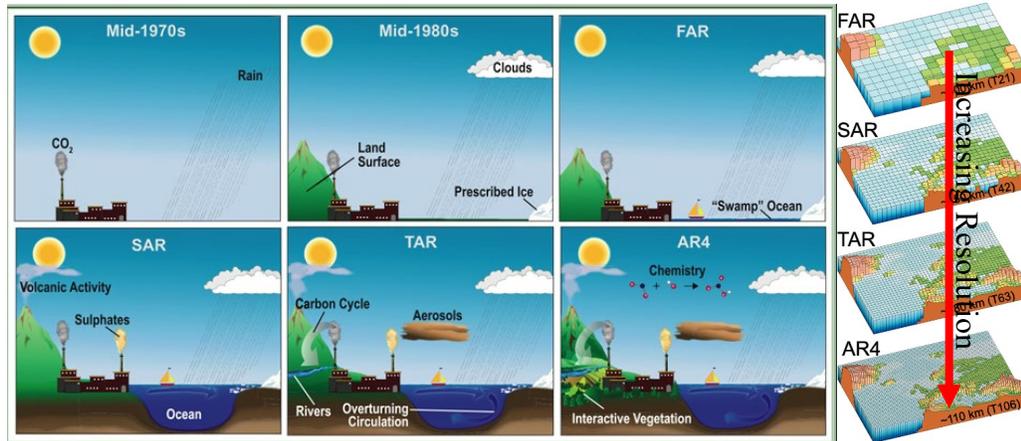


Figure 1.1. Evolution of IPCC models. Figures are adopted from IPCC (2007)

For this project, we mainly used CMIP3 and CMIP5 data. Basic information on CMIP3 and CMIP5 is presented in Table 1.1. Although the CMIP5 model output has become the standard for climate projections (Lukas et al. 2014), CMIP5 does not invalidate CMIP3. In fact, these two projections are similar (Kumar et al. 2014), and while CMIP5 is newer and superior in some ways, CMIP3 has been validated for nearly a decade, unlike the newer CMIP5. Lukas et al. (2014) writes, “It is also important to note that it took the climate science community several years to comprehensively examine and diagnose the results of the CMIP3 models, and that process is still ongoing for the CMIP5 models. Thus, while we have reason to believe the CMIP5 output is better than CMIP3 in some respects, at this stage the CMIP3 output has been more fully vetted.” However, some authors prefer using the CMIP5 data. Flato et al. (2013) writes, “There is medium evidence (single multi-model study) and medium agreement (as the inter-model difference is large) that CMIP5 models tend to simulate more intense and thus more realistic precipitation extremes than CMIP3, which could be partly due to generally higher horizontal resolution. There is medium evidence and high agreement that CMIP3 models tend to underestimate the sensitivity of extreme precipitation intensity to temperature.” Some authors (Kunkel et al. 2016) suggest using both CMIP3 and CMIP5 as a compromise.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 1.1. Key characteristics of CMIP3 and CMIP5 model projections (Markus et al. 2018)

Characteristic	CMIP3	CMIP5
Emissions scenarios	Special report on emissions scenarios (SRES) B1, A1B, A2 (IPCC 2007)	Representative concentration pathways (RCP) 2.6, 4.5, 6.0, 8.5 (IPCC 2013)
Historical climate	1880–2000	1850–2005
Projection period	2001–2100	2006–2100+
Number of modeling centers	16	30
Number of models	22	55
Number of model simulations/projections	120	250
Spatial resolutions (average grid cell size)	96–483 km (median: 260 km)	64–257 km (median: 145 km)
Timescale of archived data	Monthly	Daily and monthly

To facilitate the communication of climatic responses to certain driving forces, IPCC has grouped CMIP3 scenarios into four “storylines,” with each representing “different demographic, social, economic, technological, and environmental developments” (IPCC 2007). The A1 scenario assumes rapid economic growth followed by a decline after 2050 due to the introduction of new and more efficient technologies. This scenario has three subcategories: fossil intensive (A1FI), nonfossil energy sources (A1T), and a balance across all energy sources (A1B). The A2 assumes regionally oriented economic development and slower economic growth. The B1 low-emission scenario shows rapid changes in economic structures toward a service and information economy, with clean and resource-efficient technologies. The newer source, CMIP5, however, is not based on development scenarios. Instead, it is based on the final effects of development expressed by the representative concentration pathways (RCPs). RCPs are based on four greenhouse gas concentration trajectories. The greenhouse effect causes global warming, which is quantified through a radiative forcing and expressed as watts per square meter (W/m^2). The four RCPs are RCP2.6, RCP4.5, RCP6, and RCP8.5. They are named after a possible range of radiative forcing values in the year 2100 relative to preindustrial values (+2.6, +4.5, +6.0, and +8.5 W/m^2 , respectively). Although global warming is caused by several greenhouse gas emissions, CO₂ is the primary gas that has contributed to recent climate change.

1.2.2. Statistical and dynamical downscaling

Large-scale climate data generated by general circulation models (GCMs) are typically downscaled to spatial and temporal scales relevant to small-scale impact studies. The primary

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

downscaling approaches can be divided into statistical and dynamical downscaling (Fig. 1.2), each with its own advantages and disadvantages.

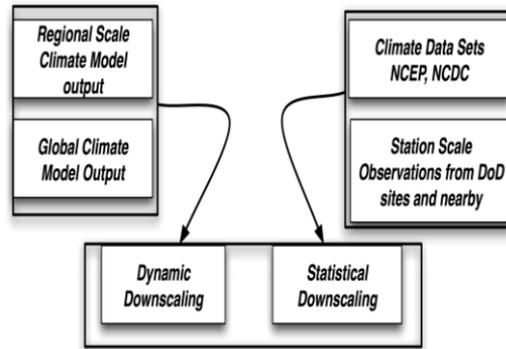


Figure 1.2. Dynamical vs. statistical downscaling (adopted from Kotamarthi et al. 2016)

Statistical downscaling consists of developing statistical relationships between large-scale predictor fields and local climate variables and then applying these relationships to coarse GCM output to approximate expected local climate features in the future. Given the low computational cost of statistical downscaling, the method can easily be applied to a diverse set of GCMs, thereby representing the envelope of future climate states. However, the approach assumes stationary statistical relationships, with the current relationships between large-scale predictors and local predictands remaining constant under climate change. Statistical downscaling is sensitive to the selection of predictors and the capability of GCMs to accurately simulate these predictors (Hayhoe 2010). Dynamical downscaling consists of the application of a limited-area, high-resolution regional climate model (RCM), forced by initial and lateral boundary conditions (LBCs) that originate from the large-scale fields in the GCM output. The fine resolution and advanced physics parameterizations of an RCM, along with the inclusion of critical dynamical processes that are absent in the parent GCM, typically provide added value to the GCM by more accurately capturing mesoscale and synoptic features (e.g., of convection) while accounting for topographic influences from mountains, lakes, and coastlines (Jones et al. 1995; Ballentine et al. 1998; Leung and Wigmosta 1999; Mearns 2003; Mearns et al. 2003; Hay et al. 2006; Anderson et al. 2007; Markus et al. 2012; Feser et al. 2011; Kanamitsu and DeHann 2011; Notaro et al.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

2015). Dynamical downscaling often leads to a better representation of extreme events, including more accurate precipitation variability and intensity (Roads et al. 2003; Mo et al. 2005; Christensen and Christensen 2007; Gutowski et al. 2010; Kanamitsu and DeHaan 2011; Kawazoe and Gutowski, 2013). The advantages of dynamical downscaling include the use of physical principles to simulate critical aspects of the local climate, including feedback among climate system components (Hayhoe 2010; Notaro et al. 2015); lack of restriction to the stationarity assumption, which may not be well justified (Fowler et al. 2007); and the vast set of potential outputted variables for analysis. However, the disadvantages of dynamical downscaling include high computational costs and thus its application to a small set of available GCMs; spurious effects from the LBCs; reliance on parameterization schemes to represent sub-grid-scale processes; and the need for debiasing of the RCM output (Brown et al. 2008; Hayhoe 2010). Numerous studies compared statistical and dynamical downscaling approaches and found similar levels of skill (Spak et al. 2007; Schoof et al. 2009; Manzanas et al. 2018), leading many researchers to recommend combined dynamical and statistical downscaling to generate high-resolution climate change scenarios for impact modeling (Yoon et al. 2012; Guyennon et al. 2013). More materials on models, CMIP runs, and downscaling schemes can be found from IPCC publications and other reports such as Kotamarthi et al. (2016).

1.3. Downscaled data sets used in this study

We used four different data sets in our comparison with observations including one dynamical downscaled data set and three statistical downscaled data sets. The following are brief introductions to these data sets and references information.

1.3.1. Dynamical downscaled data set: NA-CORDEX

The NA-CORDEX data archive contains output from regional climate models (RCMs) run over a domain covering most of North America using boundary conditions from global climate model (GCM) simulations in the CMIP5 archive. These simulations were run from 1950 to 2100 with a spatial resolution of 0.22°/25 km or 0.44°/50 km. Table 1.2 shows the global and regional models included in this project. Note that this project is still ongoing, and not all the

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

simulations have been finished. For our study, we used the downscaled results from two global models, namely the CanESM2 and EC-EARTH. We analyzed six realizations in total. As more model outputs become available in the future, it will be possible to add new outputs into our analysis. For further information about this data set, we refer to the NA-CORDEX website (<https://na-cordex.org/>).

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 1.2. Global and regional models used by NA-CORDEX

	CRCM5 (UQAM)	CRCM5 (OURANOS)	RCA4	RegCM4	WRF	CanRCM4	HIRHAM5	
ERA-Int	0.44° 0.22° 0.11°	0.44°	0.44°	50km 25km	50km 25km	0.44° 0.22°	0.44°	RCP
HadGEM2-ES				50km 25km	50km* 25km*			4.5 8.5
CanESM2	0.44° 0.44° 0.22°	0.22°†	0.44°			0.44° 0.22°		4.5 8.5
MPI-ESM-LR	0.44° 0.22° 0.44°	0.22°†		50km* 25km*	50km 25km			4.5 8.5
MPI-ESM-MR	0.44° 0.22°							4.5 8.5
EC-EARTH‡			0.44° 0.44°				0.44° 0.44°	2.6 4.5 8.5
GFDL-ESM2M		0.22°†		50km 25km	50km* 25km*			4.5 8.5

(For details: <https://na-cordex.org/simulation-matrix>)

1.3.2. CMIP5 LOCA

The Localized Constructed Analogs (LOCA) data set is based on a statistical scheme that produces downscaled estimates suitable for hydrological simulations using a multi-scale spatial matching scheme to pick appropriate analog days from observations. We used the downscaled results from 32 models with a 1/16-degree (~ 6 km by 6 km) resolution. For further information about this data set, we refer to the LOCA website (<http://loca.ucsd.edu/>).

1.3.3. CMIP5 BCCA_{v2}

Bias-Correction and Constructed Analogs (BCCA) is another widely used downscaled data set. We used the downscaled results from 20 CMIP5 models with a 1/8-degree resolution. Compared with BCCA version 1, BCCA_{v2} data have corrected a dry bias over much of the contiguous United States, and more especially over the central and eastern US. For further

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

information about this data set, we refer to the following website https://gdo-dcp.ucllnl.org/downscaled_cmip_projections/techmemo/Downscaled_Climate_Projections_Addendum_Sept2016.pdf (Maurer et al. 2010; Gutmann et al. 2014).

1.3.4. CMIP5 UWPD

This downscaled data set is generated by Dr. David Lorenz at the Center for Climatic Research, University of Wisconsin-Madison using a probabilistic approach in which the large scale in a climate model does not determine the precise values of the downscaled variables. Instead, the large scale determines the likelihood of potential values of the downscaled variables, or in other words, the large scale determines the probability density function (PDF) of the downscaled variables. This approach has several strong advantages compared with a more traditional deterministic approach in terms of realistic extremes without inflation, flexibility to control covariability in time, space, and between variables, and preservation of extremes when interpolating to a grid, etc. For further information about this data set, we refer to Wu et al. (2019)

Although these data sets have been evaluated for different applications, given our special target to use these data sets for nonstationary precipitation frequency analysis, we re-examined their abilities to reproduce the observed characteristics by station observations provided by the Office of Water Prediction's (OWP) Hydrometeorological Design Studies Center (HDSC). The following section will examine the consistency in terms of basic statistics (Section II) and precipitation frequency (PF) based on an annual maximum series (AMS) and a partial duration series (PDS) (Section III).

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

1.4. Fourth National Climate Assessment

The National Climate Assessment (Easterling et al., 2017) has shown that in many regions of the United States, including the Northeast, the risk of increased heavy precipitation has become higher. Moreover, climate modeling-based scientific studies indicate that this trend will continue. Fig. 1.3 shows the projected percent change in the 20-year return period for daily precipitation for the mid- (left maps) and late 21st century (right maps). Results are shown for a lower scenario (top maps; RCP4.5) and for a higher scenario (bottom maps; RCP8.5). These results are calculated from the LOCA downscaled data. (Fig.7.7 of Easterling et al. 2017)

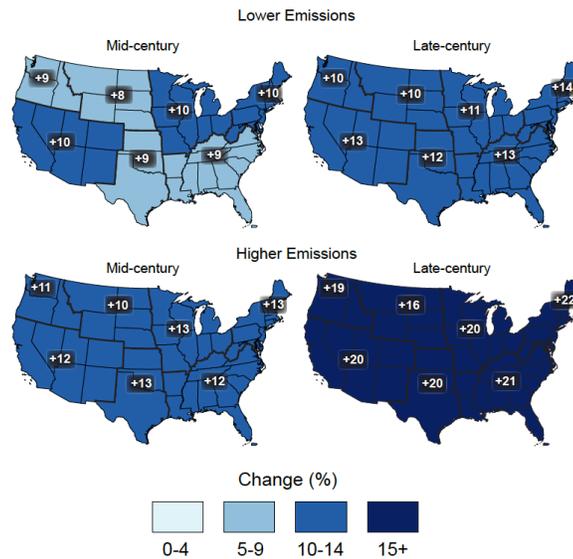


Figure 1.3. Projected change in the 20-year return period amount for daily precipitation for mid- (left maps) and late 21st century (right maps)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

SECTION II. Evaluation of the downscaled climate modeling data for the historical period based on extreme time series gauge data and their basic statistics

In collaboration with the HDSC, we identified the Northeast US and Texas where extreme precipitation events are projected to increase in the future as two target regions for our comparison study.

2.1. Results for Northeast U.S.

Observed change in the magnitude of the top 1-percentile daily rainfall per year has increased by 70% for the period of 1958-2012 in the Northeast (Karl et al. 2009). We selected 498 stations in which to examine the AMS time series, precipitation frequency, and their comparisons with the downscaled model output.

Fig. 2.1 shows the locations of these stations together with the stations used for the UWPD downscaled data set. We calculated the 1960-2005 mean AMS for each station based on the observational data first. Since the downscaled data have many realizations, we calculated the mean, median, lowest, highest, and closest to the observation value for the downscaled data. Figs. 2.2-2.3 show the scatter plots for these statistics. The scatter plots for the mean AMS (Fig. 2.2) show that all the downscaled data sets have a very high spatial correlation ($R > 0.8$) with the observational data, reflected by the linear pattern of the dots. However, LOCA and BCCAv2 data have systematic dry biases, where almost all the dots are above the line $y=x$. In contrast, the NA-CORDEX and UWPD data are closer to $y=x$, particularly for the UWPD. This suggests that these two data sets have smaller biases in terms of reflecting the observed mean AMS.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

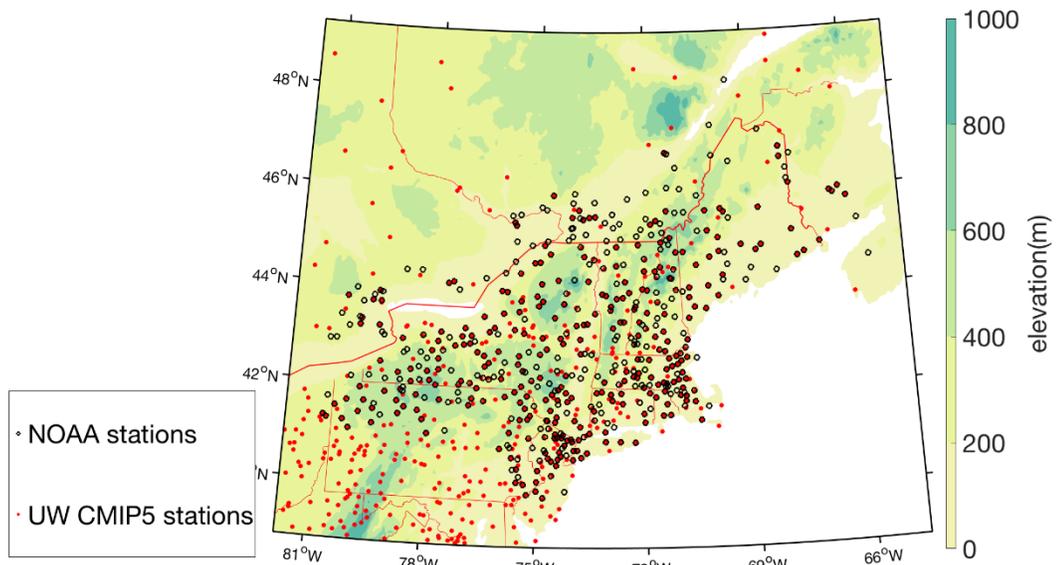


Figure 2.1. NOAA and UWPD stations for Northeast US

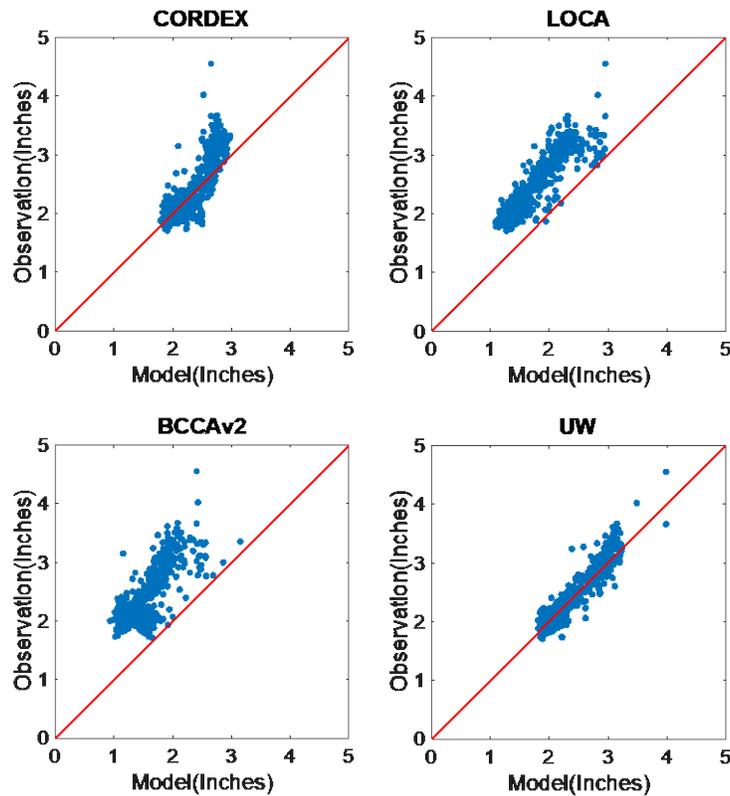


Figure 2.2. Scatter plot for 1960-2005 mean AMS calculated based on station observed data and downscaled data. Each point represents one station in the Northeast US.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

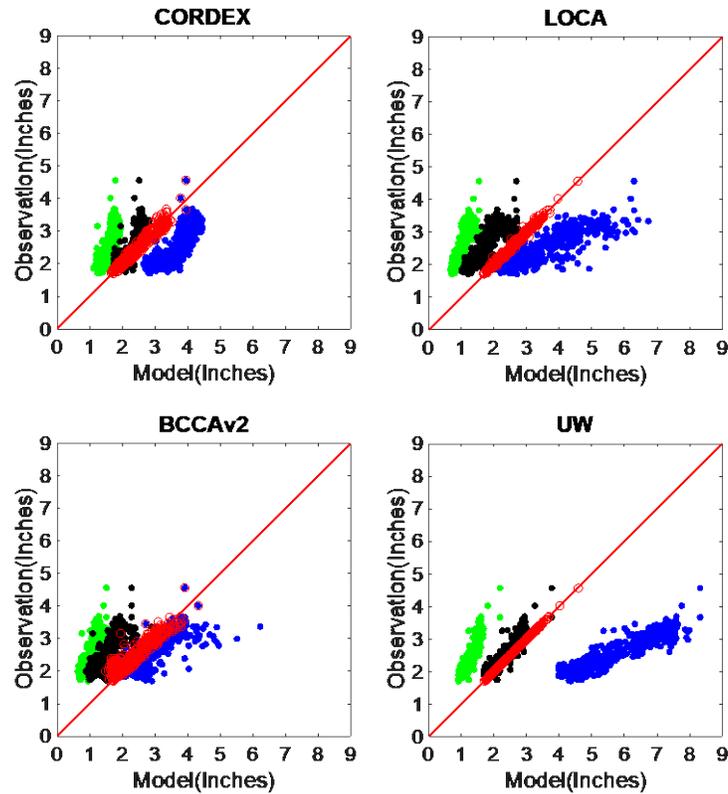


Figure 2.3. Same as Figure 2.2 but for different quantiles of the 1960-2005 mean AMS for downscaled data sets (green dots: lowest values; black dots: median values; red dots: closest values to observations; blue dots: highest values) for the Northeast US

The CMIP5-based model simulations produce results with a large variability among different models (e.g., Chapter 9 IPCC, 2007). Even for the same model, different realizations may simulate the observations in different ways and with different accuracies. To examine the variability, we compared different statistics across the ensemble of downscaled data with the observations (Fig. 2.3). The scatter plots for the median value (black dots in Fig. 2.3) are similar with that of the mean (shown in Fig. 2.2), and similarly, the NA-CORDEX and UWPD data still outperform the LOCA and BCCAv2 data. It is interesting to examine the scatter plots for the lowest and highest values, where the UWPD and LOCA have ranges higher than those of the BCCAv2 and NA-CORDEX. Another interesting option is to use the closest points (red dots) to the observations for each ensemble of the data sets.

The downscaled data sets also use the observations (overlap stations) to fit their models; therefore, the consistency should not be surprising. To understand how well the downscaled data simulate the stations that are not used when carrying out downscaling (non-overlap stations), we

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

compared the AMS scatter plots for overlapped and non-overlapped stations (Fig. 2.4) based on UWPD data. Our selection of UWPD data is based on Figs. 2.2-2.3, where UWPD data show the closest features to the observations. Among the 498 NOAA stations, 299 stations' data have been used to generate the downscaled data sets. However, the features of the scatter plots for overlapped and non-overlapped stations are quite similar, implying that the inherent features of the physical system are indeed captured by the downscaled data.

Besides the 1960-2005 mean AMS, we also compare model-generated and observed data by using the Q-Q plot for all values in the 1960-2005 period (Figs. 2.5-2.6). Again, similar features are evident: the LOCA and BCCAv2 data show systematic dry biases, while the NA-CORDEX and UWPD data show no significant biases. One interesting feature is that the LOCA and BCCAv2 data sets are more linearly related to observations, indicating that they can be easily transformed to the similar distribution with observations. This can be useful when analyzing their projection runs by applying the same transforms.

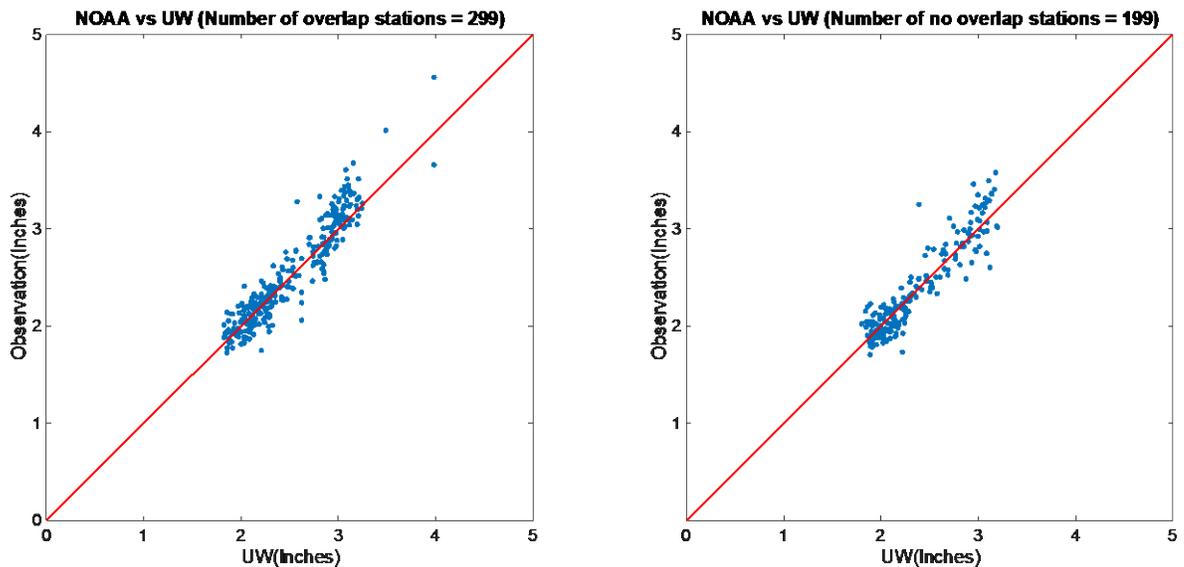


Figure 2.4. Scatter plot for overlapped and non-overlapped NOAA stations in the Northeast US

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

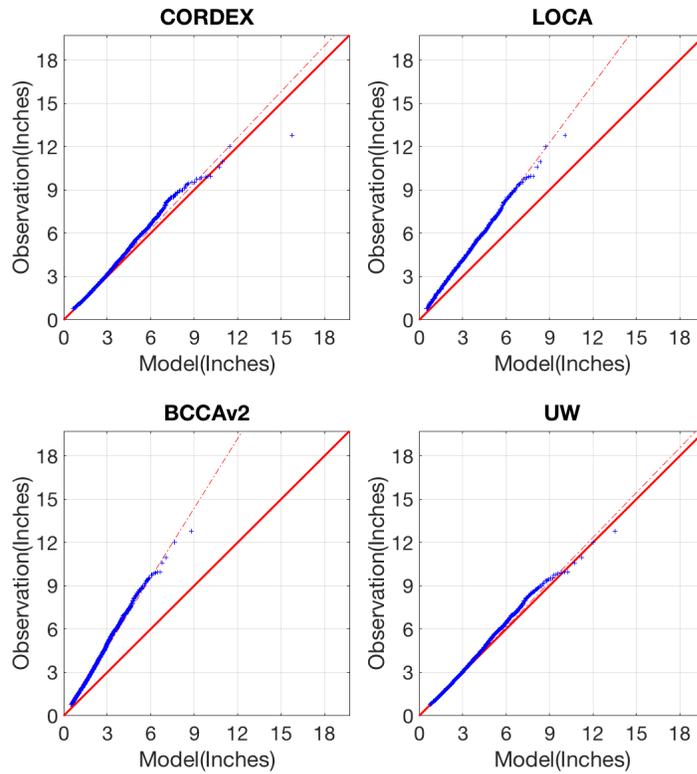


Figure 2.5. Q-Q plot for different downscaled data sets in the Northeast US

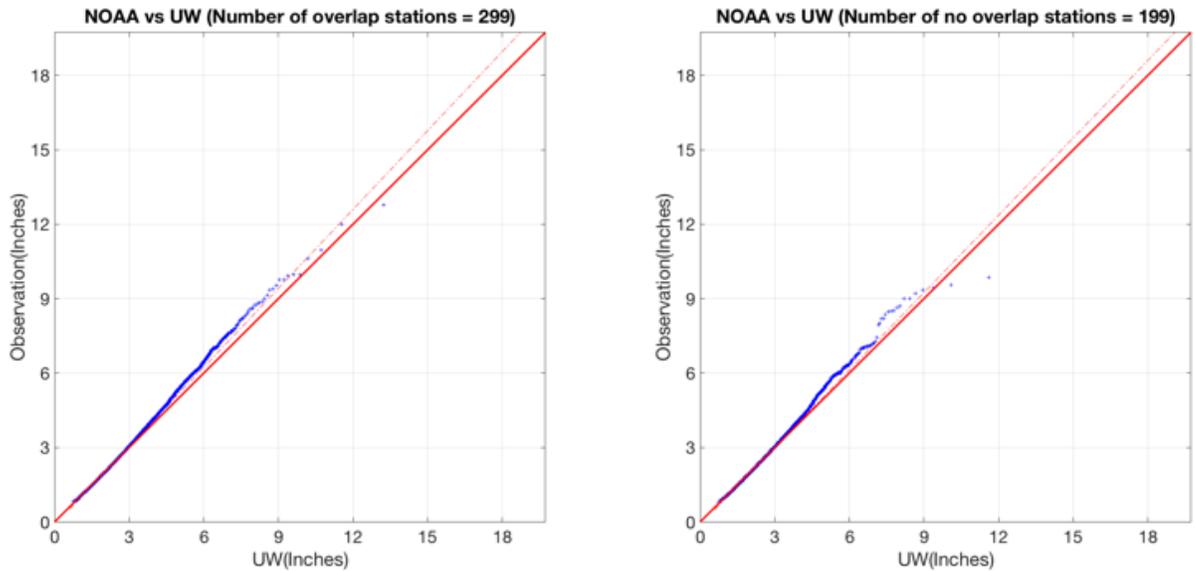


Figure 2.6. Q-Q plot for overlapped and non-overlapped NOAA stations in the Northeast US

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

2.2. Results for Texas

A parallel analysis (Figs. 2.7-2.12) was also performed for the Texas region, where we selected 616 stations (Fig. 2.7). Similar to the Northeast, the UWPD data show the best simulating results in terms of features of the AMS. The other three downscaled data sets all show systematic dry biases for the AMS, but for the PF, the NA-CORDEX performs better than the LOCA and BCCAv2.

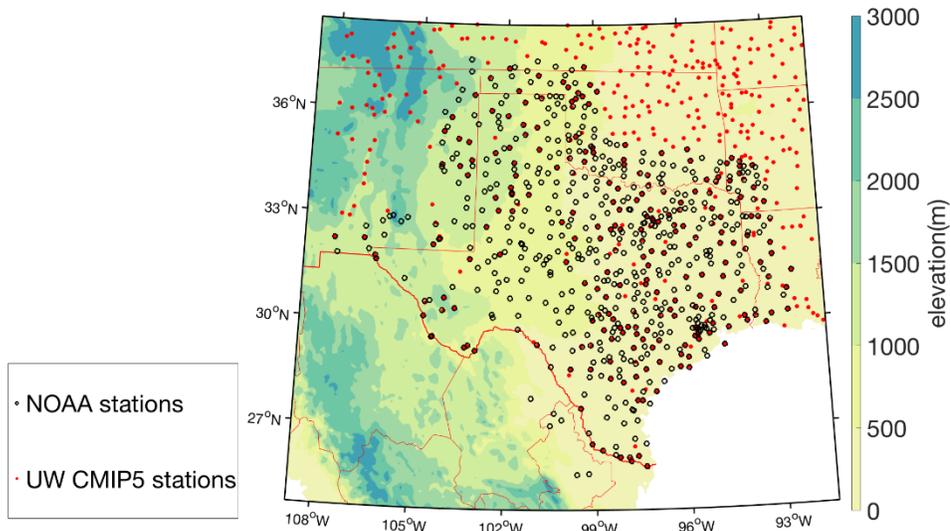


Figure 2.7. NOAA and UWPD stations for Texas

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

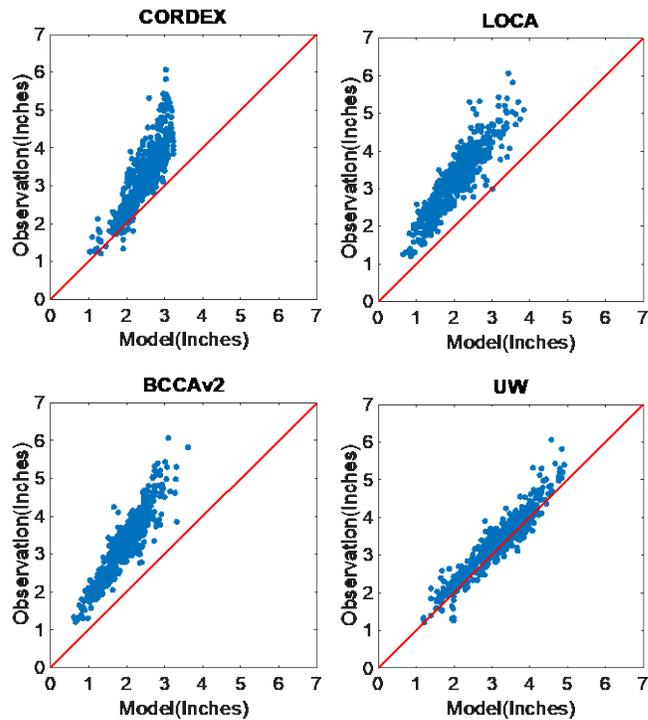


Figure 2.8. Scatter plot for the 1960-2005 mean AMS calculated based on station observed data and downscaled data. Each point represents one station in Texas

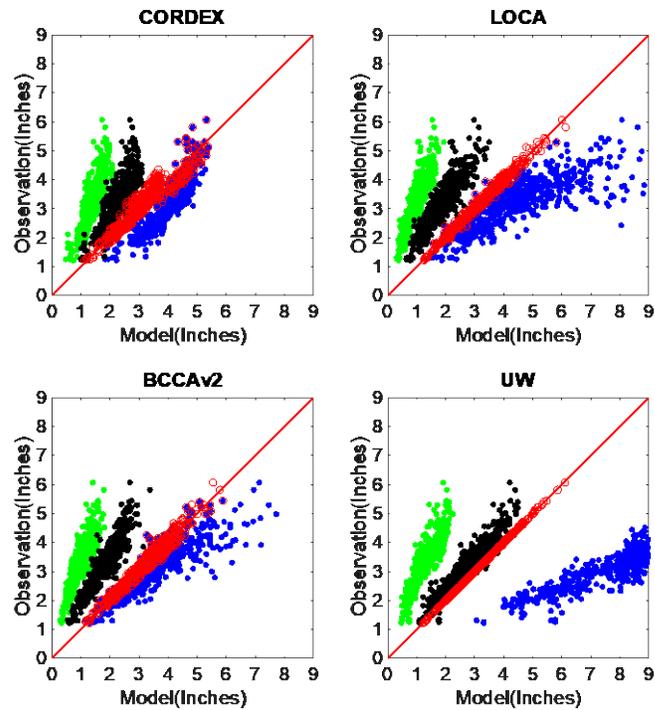


Figure 2.9. Same as Figure 2.8 but for different quantiles of the 1960-2005 mean AMS for downscaled data sets (green dots: lowest values; black dots: median values; red dots: closest values to observations; blue dots: highest values) for Texas

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

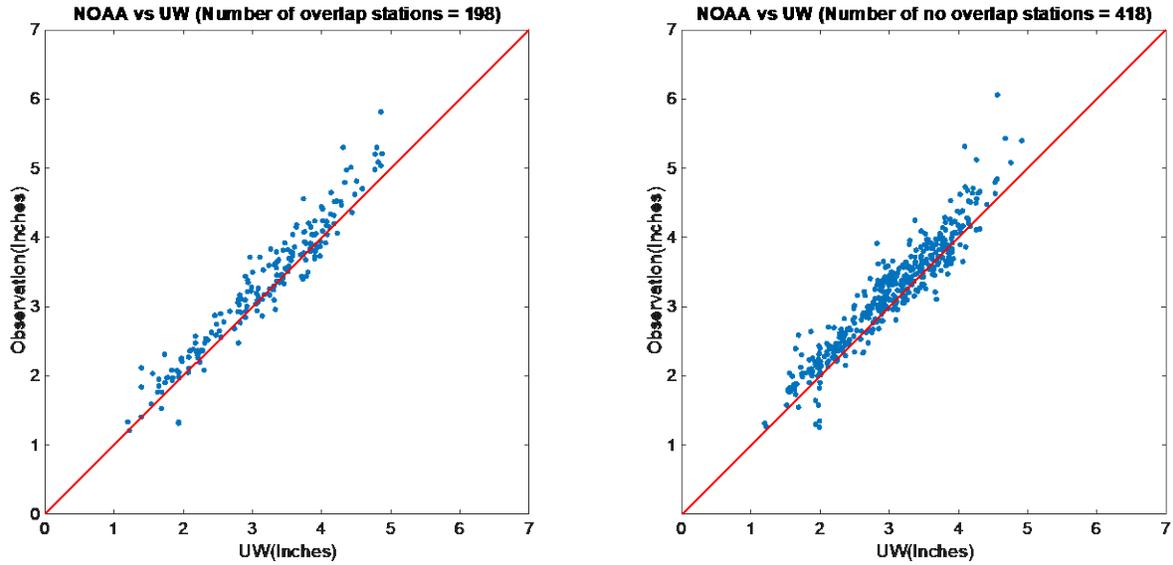


Figure 2.10. Scatter plot for overlapped and non-overlapped NOAA stations in Texas

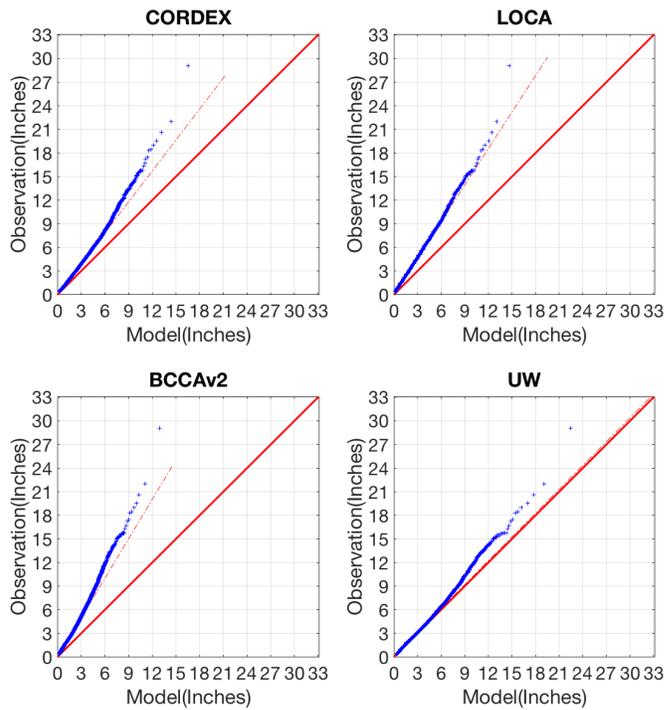


Figure 2.11. Q-Q plot for different data sets in Texas

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

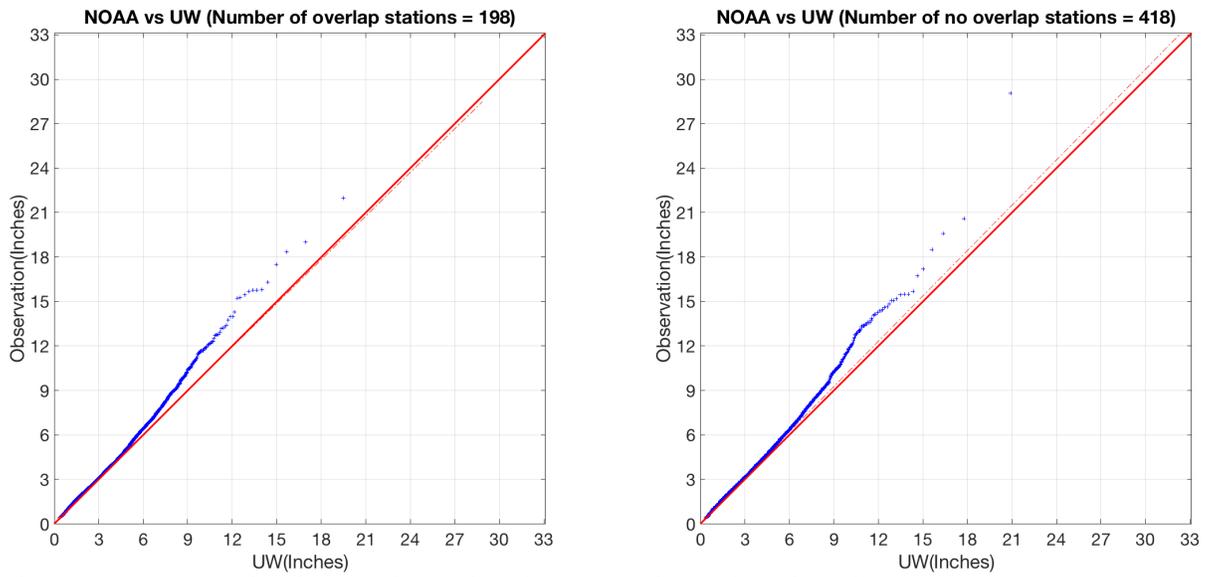


Figure 2.12. Q-Q plot for overlapped and non-overlapped stations in Texas

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

2.3. Conclusions

In this section, we presented an introduction to the CMIP5 models and an overview of existing downscale methods, including statistical and dynamical downscaling approaches. Using the annual maximum series (AMS), we compared statistics calculated based on rain gauge observed data with those based on four different downscaled data sets, namely the dynamical downscaled NA-CORDEX data, the LOCA data, the BCCAv2 data, and the UWPD CMIP5 (UWPD) data. Of the four data sets, three (NA-CORDEX, LOCA, and BCCAv2) were developed by different agencies and are publicly available, while the fourth one (UWPD CMIP5) has been developed in this project by University of Wisconsin researchers using a novel statistical downscaling approach. The Northeast and Texas project areas defined by NOAA Atlas 14 volumes (Perica et al. 2018, 2019) have been selected as test data sets for our methodology. Our initial analysis for both regions indicated that UWPD data demonstrated a better performance over the three other downscaled data sets.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

SECTION III. Evaluation of the downscaled climate modeling daily data for the historical period 1960-2005: Comparison between precipitation frequency using observations and climate modeling downscaled data for annual maximum series (AMS) and partial duration series (PDS) methods

We present further analysis using the annual maximum series (AMS) and partial duration series (PDS) to calculate PFs. The PDS is often called the peaks over threshold (POT) method. The NOAA Atlas 14 reports (Perica et al. 2018, 2019) also have both AMS and PDS PFs. Both methodologies were first applied to the observed data, but we also applied it to the downscaled data. For PDS analysis, which involved a user-defined selection of threshold, we explored three different thresholds and carried out a systematic comparison for both observational and downscaled data. Using both AMS- and PDS-based PFs, the goal of this section is to determine the skill of downscaled data in reproducing the statistical frequency estimates of the observed data. AMS-based precipitation frequency estimates used the generalized extreme value (GEV) distribution, and the PDS-based estimates used the generalized Pareto distribution (GPD). The historical period selected for this analysis is 1960-2005 to match the range of the available climate model downscaled data. We used the point precipitation estimation method, where only the target stations that observed extreme time series are used to estimate the distribution parameters using the maximum likelihood estimation (MLE) method. Because each point used ensembles of models and multiple model runs, the ensemble means were used as robust estimates of precipitation frequencies.

3.1. The annual maximum series (AMS) and the partial duration series (PDS)

The annual maximum series (AMS) and the partial duration series (PDS) are both recommended for frequency analysis in hydrology (e.g., Madsen et al. 1997). The AMS is generated based on the maximum value of each year. A typical length of an AMS equals the number of years. Therefore, all PFs based on AMS calculations use just one datum each year with all the other data eliminated, potentially wasting the valuable observed data. In contrast,

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

PDSs are generated by retaining the values over a pre-selected threshold. Therefore, the length of the PDS is determined by the threshold, which may vary dramatically across different stations. In general, the threshold is selected such that, on average, several events (2-4) occur per year. Therefore, the length of a PDS is usually several times larger than that of the AMS. More observational data are used to estimate PFs based on the PDS.

The PDS uses more data, thus it may produce more significant estimates of PFs. However, it appears that retaining more non-extreme values could potentially reduce the effects of extremes in the final analysis. Instead of developing the estimates based on one time-series data set, it is important to retain both AMS- and PDS-based estimates to provide insight into the uncertainties of PFs. For this reason, we provide comparisons between the downscaled data and both AMS- and PDS-based estimates that we used to test the downscaled data's ability to simulate historical observations.

In practice, calculating AMS- and PDS-based estimates is quite different because different distribution functions are associated with AMS and PDS series, namely the Generalized Extreme Value (GEV) distribution and the Generalized Pareto Distribution (GPD). To get PFs for the same return periods, the AMS return periods often need to be modified (e.g., Langbein 1949) by the formula:

$$T(AMS) = (1 - e^{-\frac{1}{T(PDS)}})^{-1}$$

For example, a return period of 2 years for the PDS is related to a return period of 2.54 years for the AMS. Therefore, to make a direct comparison of AMS- and PDS-based PF analyses, we need to use consistent return periods. For our case we used “T = [2 5 10 25 50 100 200 500 1000]” years for PDS analysis and “T = [2.54 5.52 10.51 25.50 50.50 100 200 500 1000]” years for AMS analysis so that our PDS and AMS results can be compared directly. Note that this conversion is developed empirically, and it only applies to stationary conditions.

With the help of the NOAA Hydrometeorological Design Studies Center, we obtained the new precipitation data for the Northeast US and Texas. There are 1218 stations in the Northeast US and 1231 stations in Texas. Stations with daily observations and with over 80% temporal coverage during the period of 1960-2005 were selected for analysis (Figs. 2.1-2.12). We analyzed 753 stations in the Northeast and 547 stations for Texas.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

3.2. Downscaled data's ability in simulating observed PFs

In this section, we evaluated the downscaled data's ability to reproduce the observed PFs. We show a PF comparison based on the AMS and PDS series. AMS results are based on the MLE method. We explored all four different statistically and dynamically downscaled data sets, namely the UWPD data, LOCA data, BCCAv2 data, and NA-CORDEX data. The results are described below.

3.2.1 Results for UWPD downscaled data

As we introduced in Section 1, the UWPD downscaled data were generated especially for this project. The data have a spatial resolution of 0.1 degree, and the downscaling scheme is based on the probabilistic approach in which the distribution parameters are estimated by logistic multi-variable regression equations and noises are generated by advection-diffusion equations. In this data set, 24 CMIP5 model outputs were downscaled, and three realizations were generated for each model, resulting in a 72-member ensemble. Fig. 3.1 shows the PDS thresholds comparison between the UWPD downscaled data and the observational data for the Northeast US. The station order number (x-axis) is an order number from the list of stations and does not have any statistical meaning. The spread of UWPD data well covers the observed values, suggesting the downscaled data have no significant biases, and the PDF is a possible representative of the real-world PDF. The average values across all the stations are also very close.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

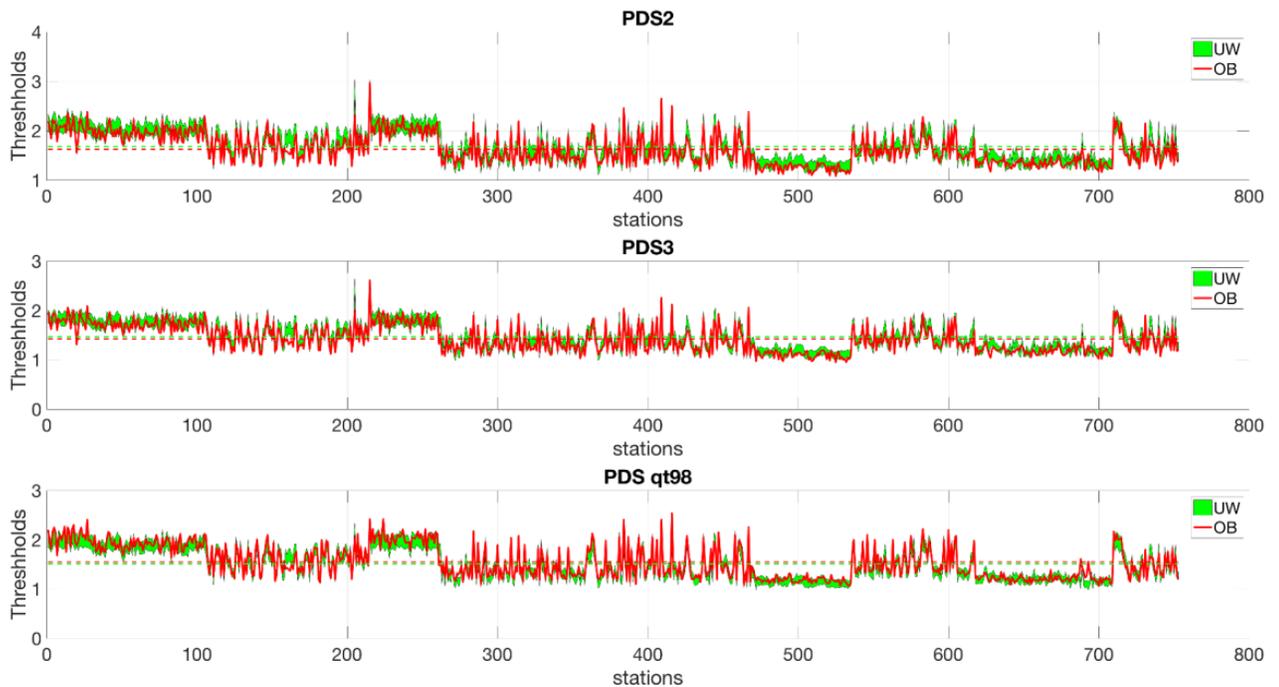


Figure 3.1. PDS thresholds for Northeast US based on UWPD data

The thresholds match and so do the PFs. Figs. 3.2-3.4 show the PF comparison for three different return periods for the Northeast US. PFs calculated based on four different methods (AMS, PDS2, PDS3, PDS qt98) were compared systematically. Since UWPD data have 72 ensemble members, providing a chance for us to calculate a range of possible values, the ensemble spread is also shown by the blue bars in the figures, while the ensemble values are shown by the red dots. To assist in the comparison, lines of “ $y=x$ ” are also plotted in red in each figure. More points above the line “ $y=x$ ” means the observed values are larger than the downscaled values. Therefore, the downscaled data have dry biases and vice versa. Figs. 3.2-3.4 show that the most outstanding feature is that most of the points are close to the line “ $y=x$ ”, suggesting the modeled PFs are consistent with the observed PFs. Quantifications using correlation coefficients and root mean square error (RMSE) further confirm this conclusion, even though the skill indeed decreases as return periods increase. The uncertainties of downscaled data reflected by the length of the blue bars also increase as the return periods increase. Note that

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

for a small group of stations with the largest PFs, the downscaled data still show dry biases as we have seen from Section II.

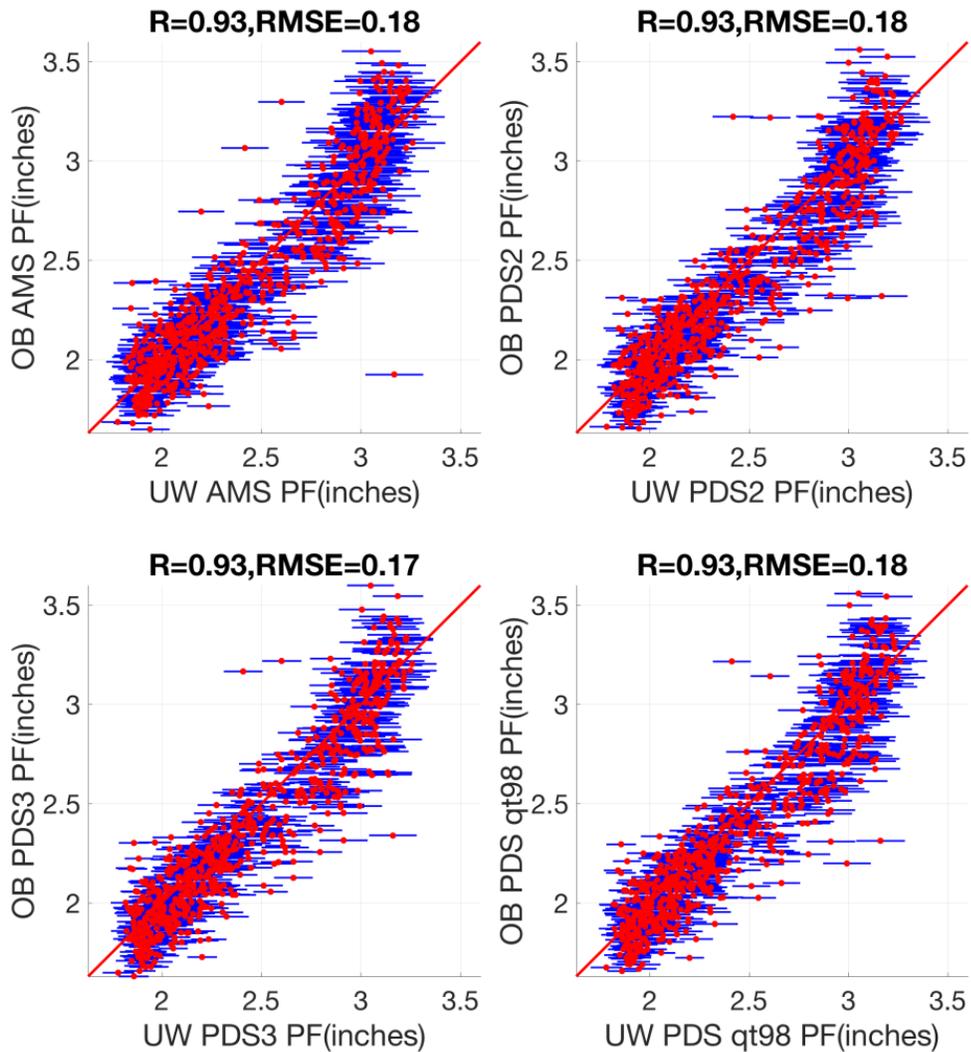


Figure 3.2. UWPD CMIP5 data: PFs scatter plot for Northeast US. The red line represents $y=x$. Results for return period = 2 years.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

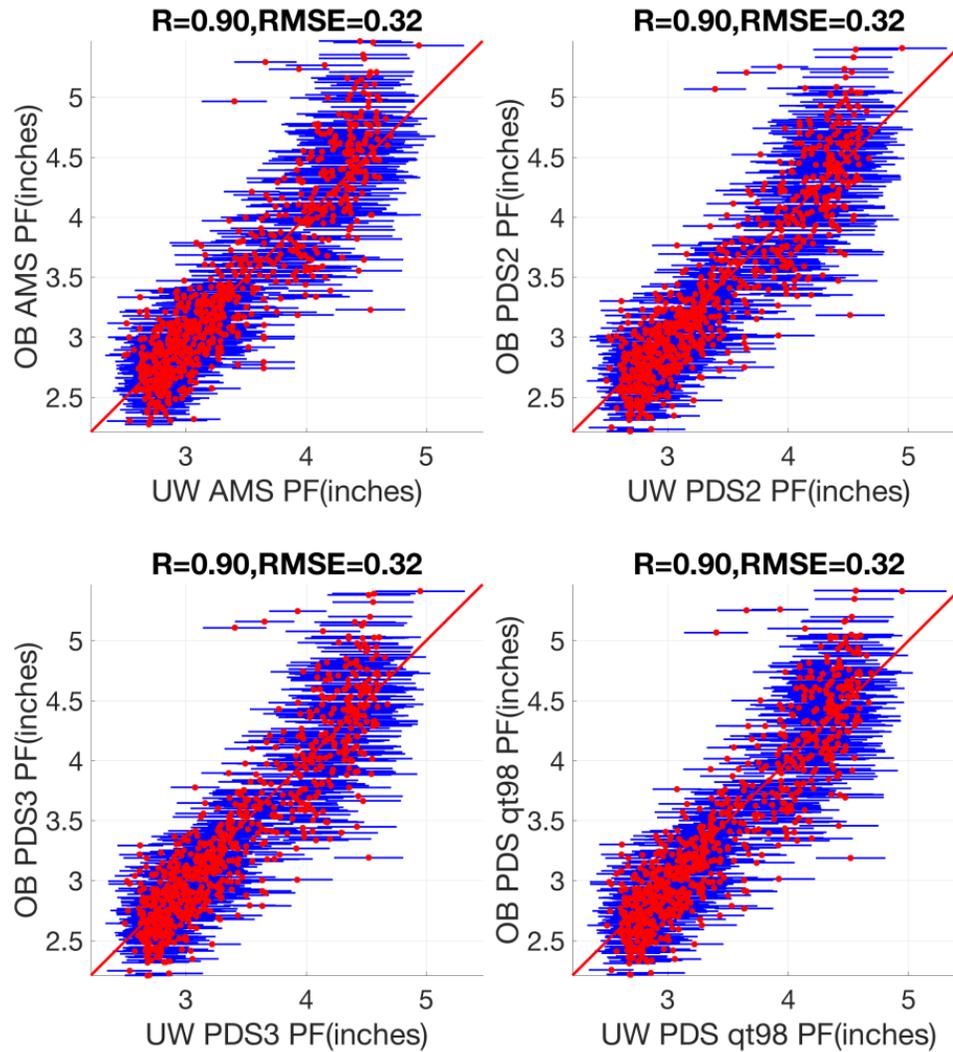


Figure 3.3. UWPD CMIP5 data: PFs scatter plot for Northeast US. The red line represents $y=x$. Results for return period = 10 years.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

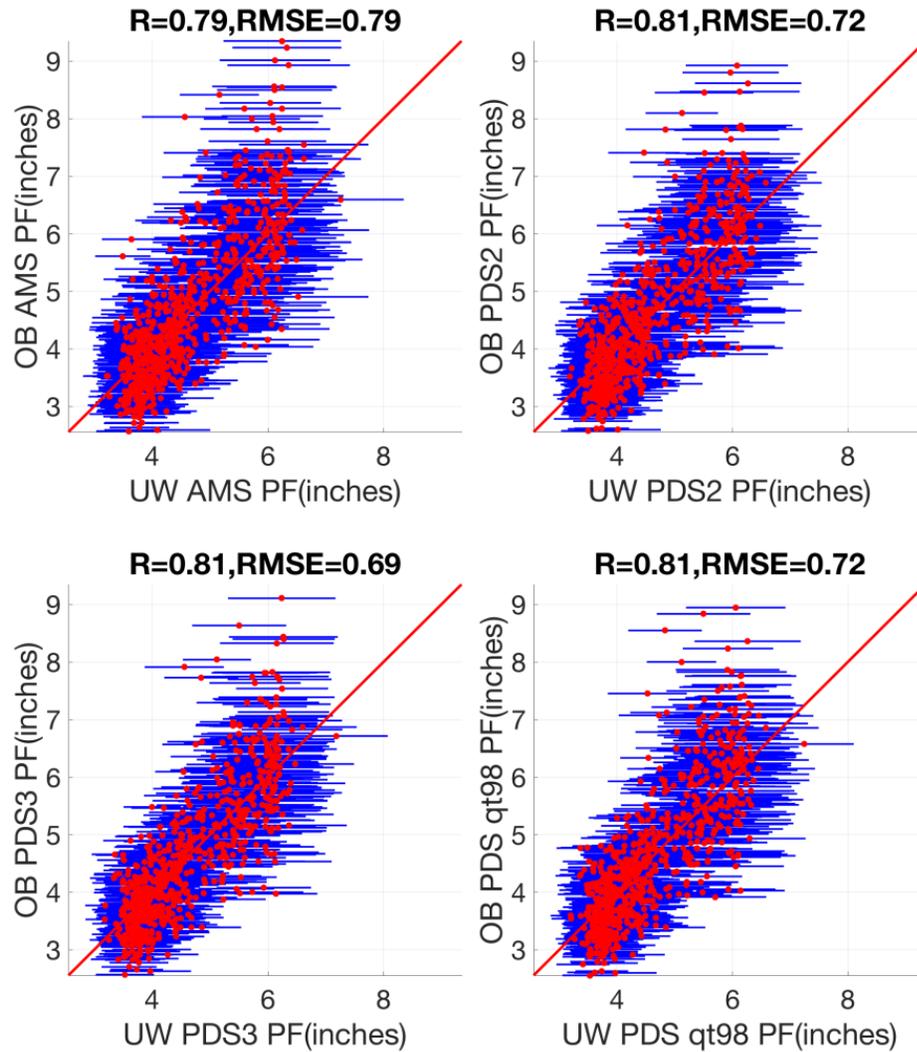


Figure 3.4. UWPD CMIP5 data: PFs scatter plot for Northeast US. The red line represents $y=x$. Results for return period = 50 years.

Results for Texas (Figs. 3.5-3.8) show a similar feature overall. The spread of the downscaled data is larger than that for the Northeast US. The correlations are somewhat higher than those for the Northeast US for 2-year and 10-year return periods, but smaller for the 50-year return period.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report



Figure 3.5. PDS thresholds for Texas. UWPD data

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

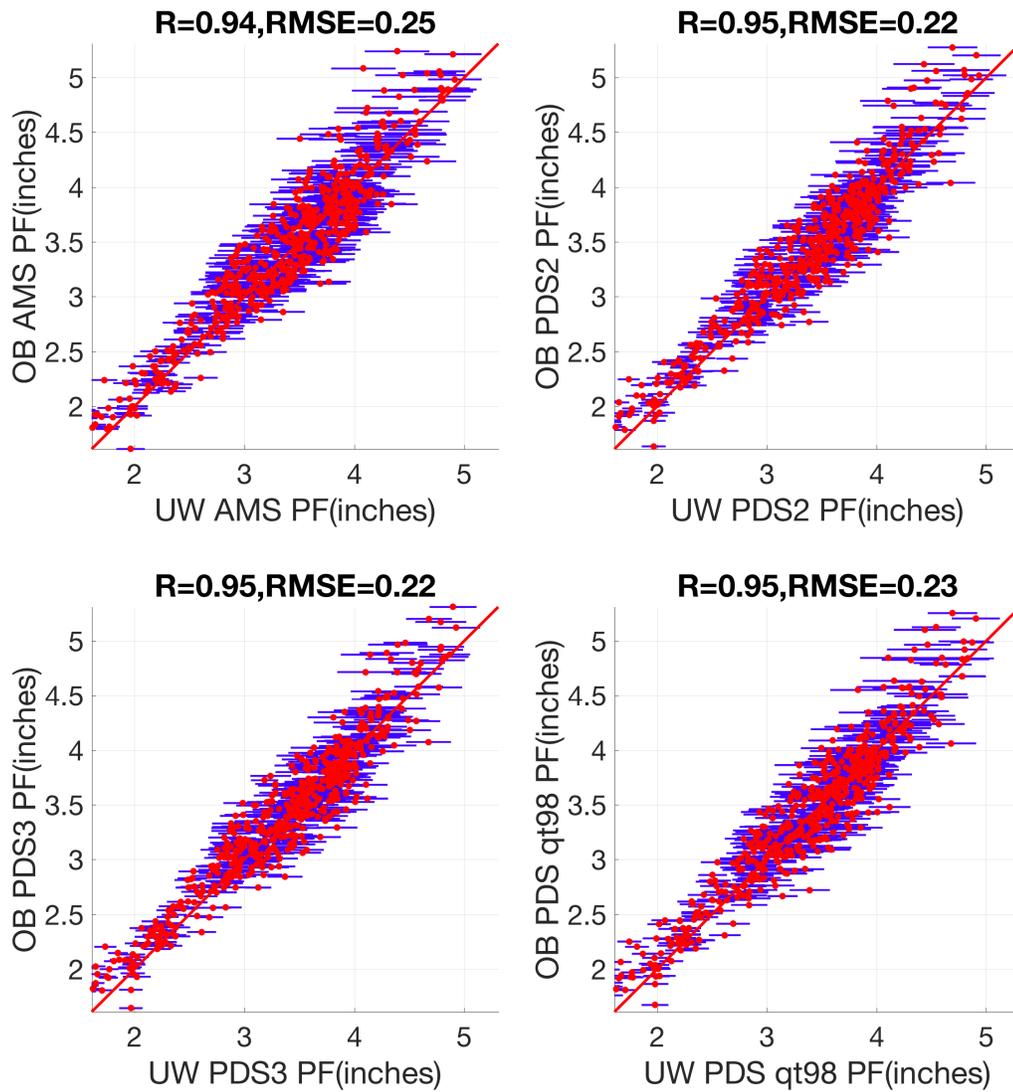


Figure 3.6. UWPD CMP5: PFs scatter plot for Texas. The red line represents $y=x$.

Results for return period = 2 years

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

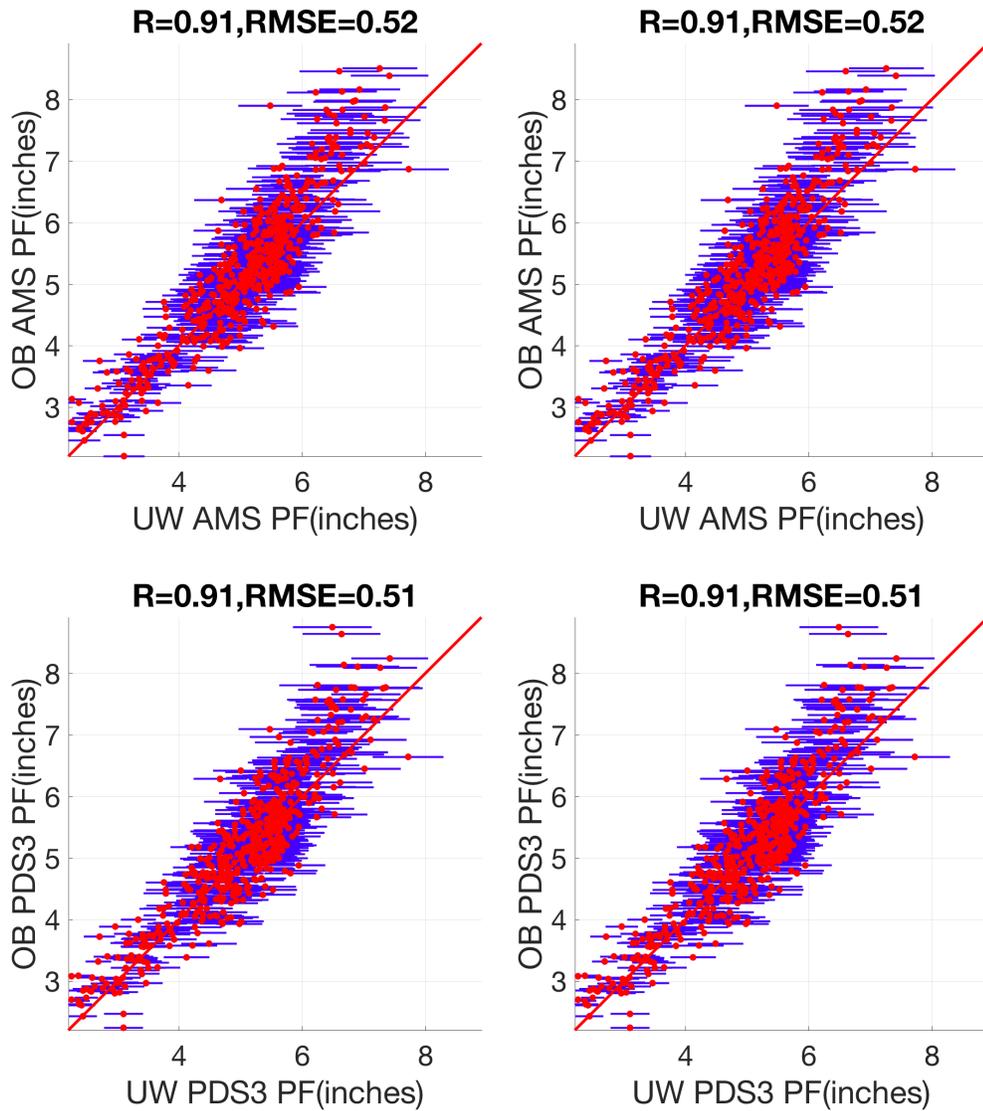


Figure 3.7. UWPD: PFs scatter plot for Texas. The red line represents $y=x$. Results for return period = 10 years

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

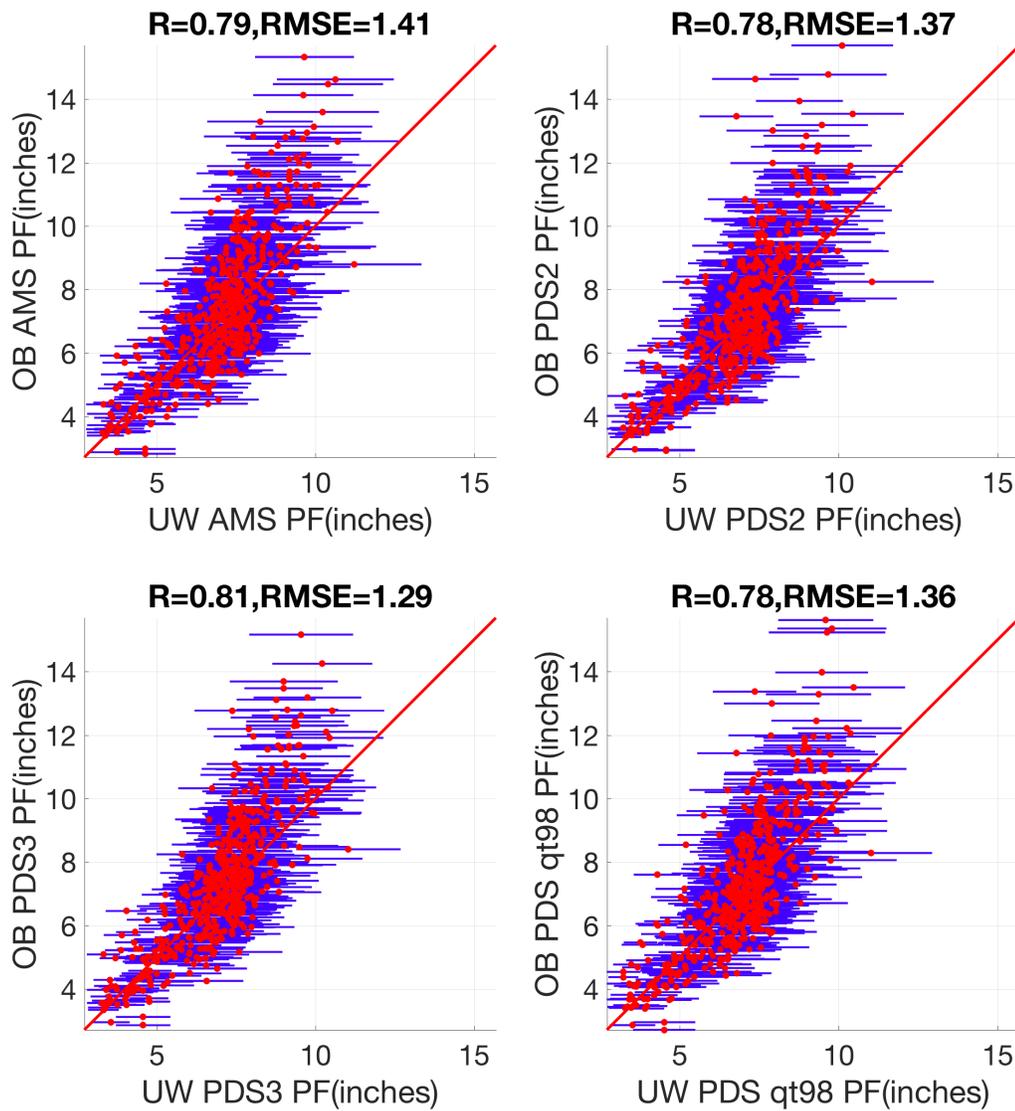


Figure 3.8. UWPD: PFs scatter plot for Texas. The red line represents $y=x$. Results for return period = 50 years

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

In summary, a comparison was made between the UWPD downscaled data and the observed data in reproducing the PF features for the historical period of 1960-2005. Four different methods were tested to calculate the PFs, namely the AMS, PDS2, PDS3, and PDS qt98. All the results generally showed that the downscaled data can capture this major feature of the observations. The confidence, however, does decrease as the return periods increase (Figs. 3.9-3.10). This might be acceptable since the uncertainties of the observed PFs for longer return periods are also large. Nevertheless, our results indeed demonstrate that downscaled data do have values to complement the observational data for PF analysis.

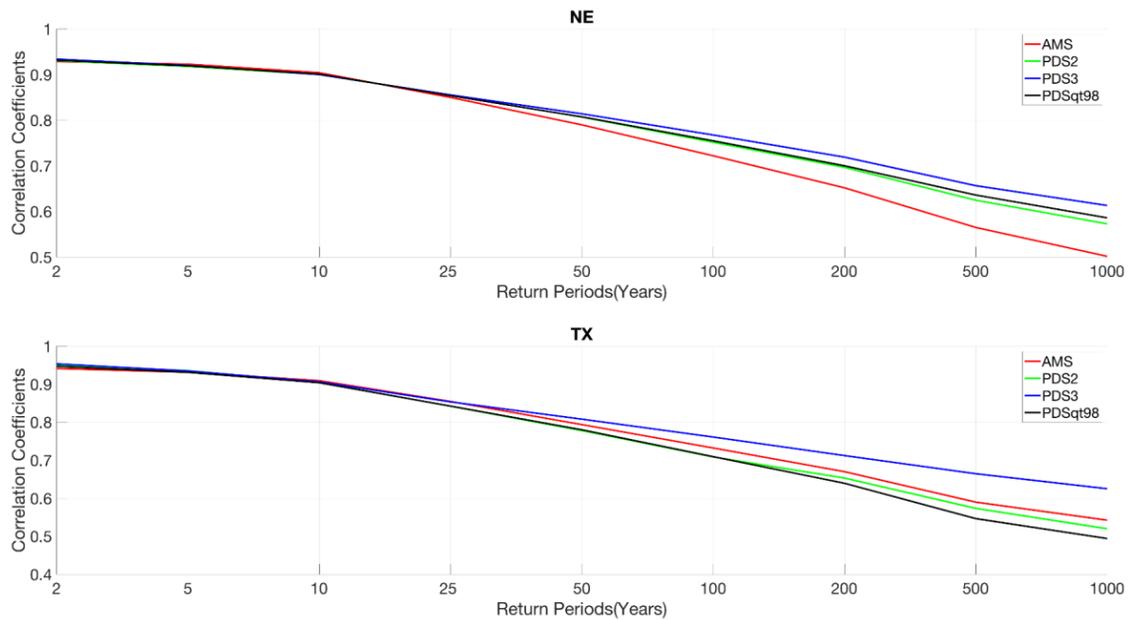


Figure 3.9. Correlation skill for all the return periods. UWPD data

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

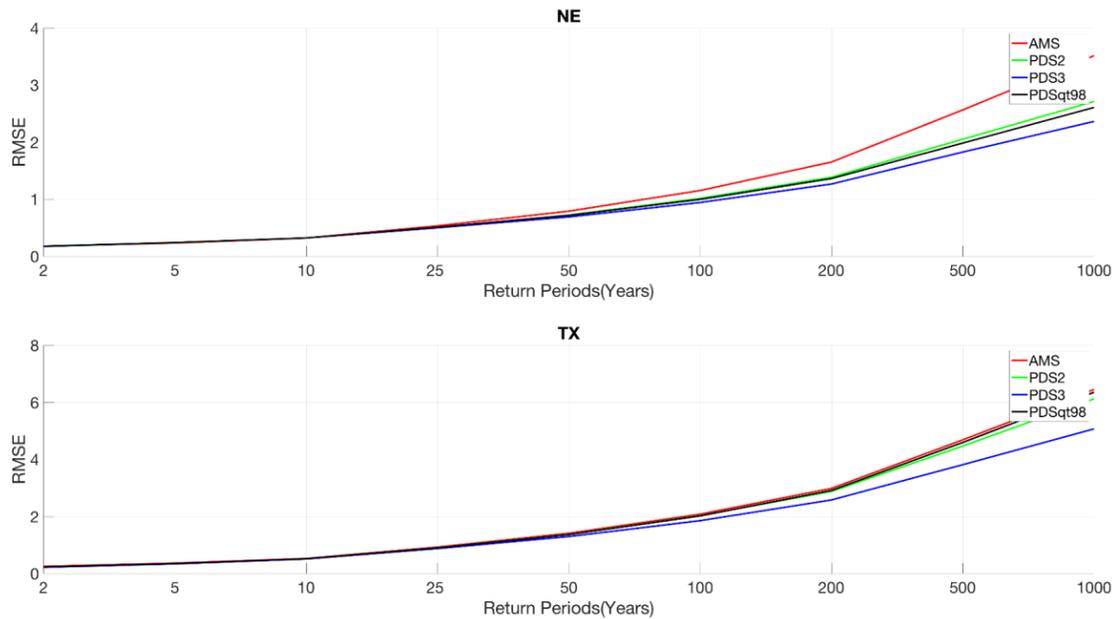


Figure 3.10. RMSE skill for all the return periods. UWPD data

3.2.2 Results for LOCA downscaled data

Besides the downscaled data we generated for this project, there are also some other widely used downscaled data sets. They all have their unique features and applications. To meet our goal, we explored these data sets as well so that we could obtain a complete picture of the performances of downscaled data. In this section, we show the results based on the LOCA downscaled data.

Figs. 3.11-3.14 show the results for the Northeast US and Figs. 3.15-3.18 show the results for Texas. Overall, the LOCA data have significant dry biases, but the correlation skill is very high. This data set could provide some useful information after some bias adjustment.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

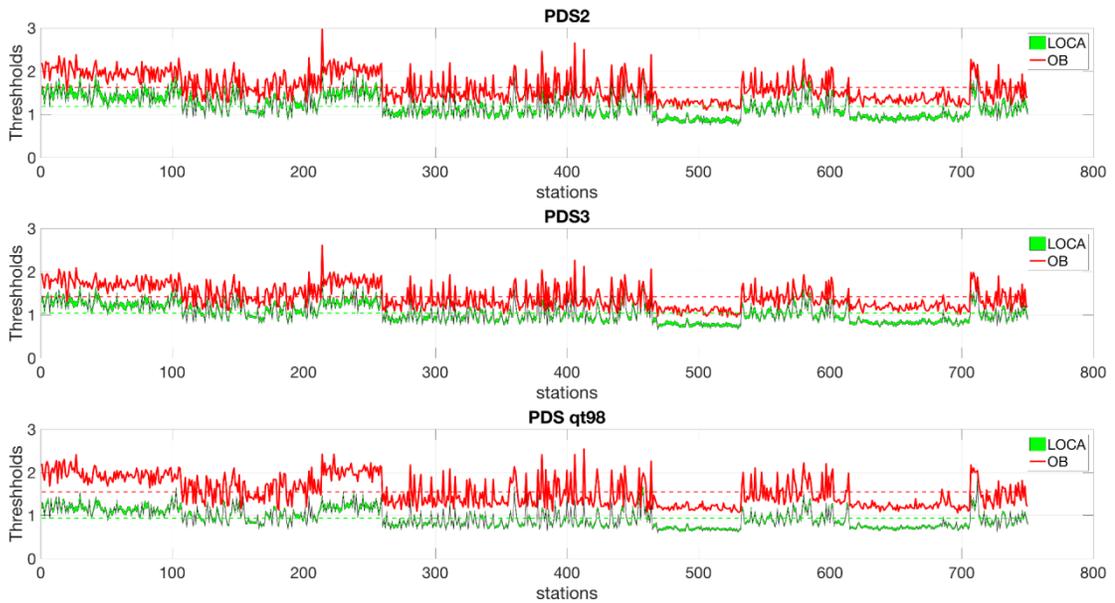


Figure 3.11. PDS thresholds for Northeast US. LOCA data

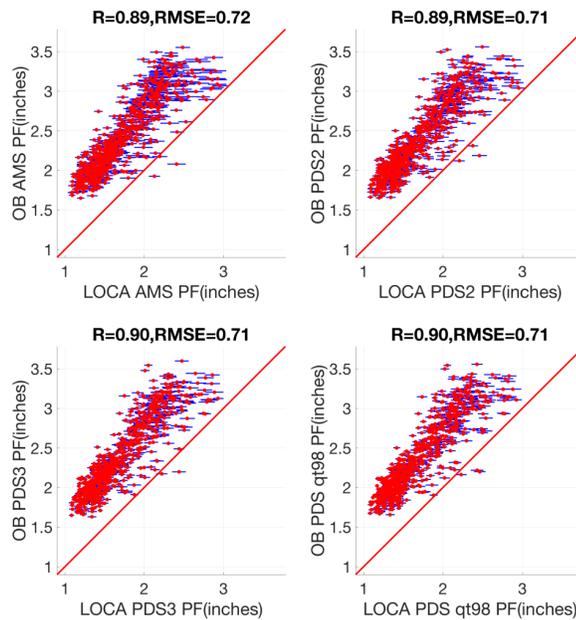


Figure 3.12. PFs scatter plot for Northeast US. The red line represents $y=x$. Results for return period = 2 years. LOCA data.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

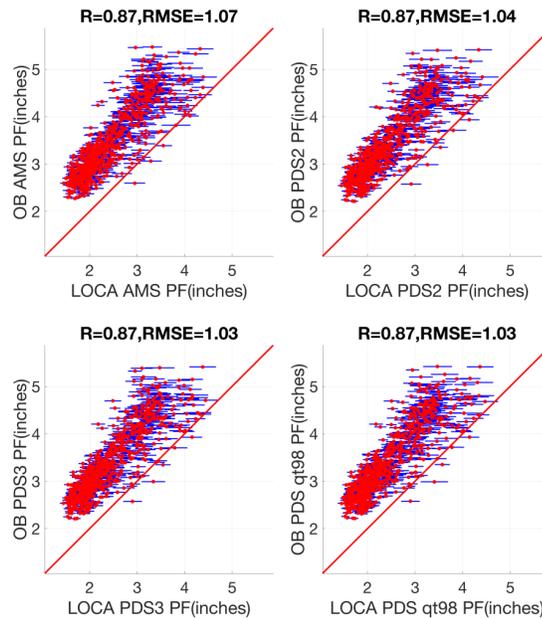


Figure 3.13. PFs scatter plot for Northeast US. The red line represents $y=x$. Results for return period = 10 years. LOCA data.

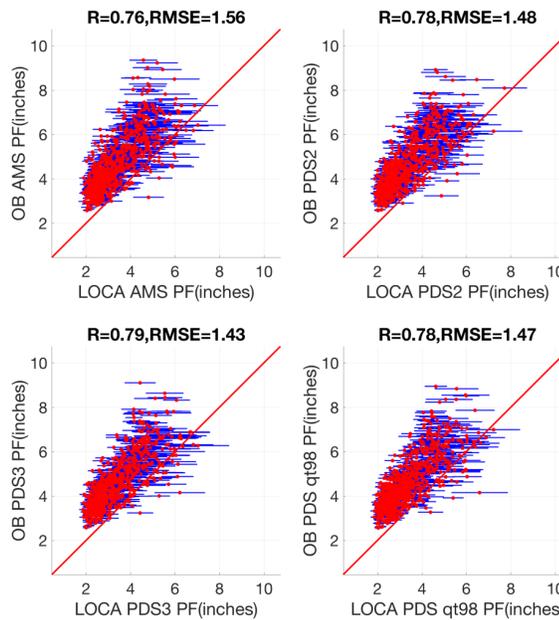


Figure 3.14. PFs scatter plot for Northeast US. The red line represents $y=x$. Results for return period = 50 years. LOCA data.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

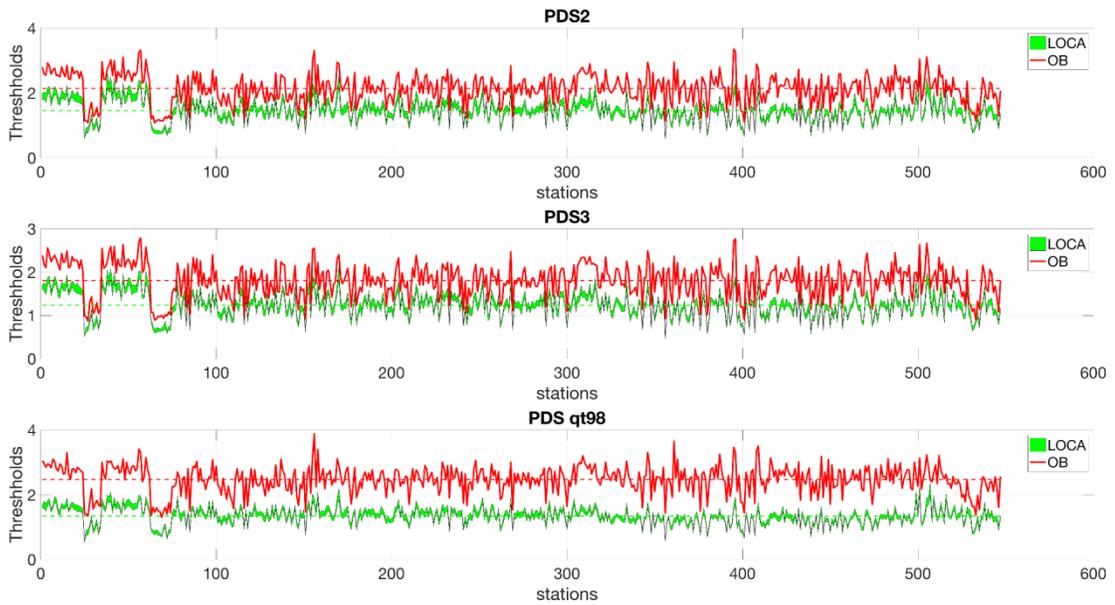


Figure 3.15. PDS thresholds for Texas using the LOCA data set.

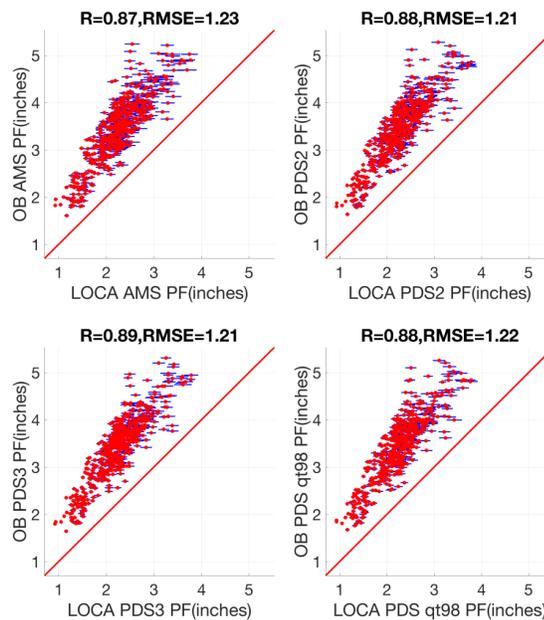


Figure 3.16. PFs scatter plot for Texas. The red line represents $y=x$. Results are shown for return period = 2 years using the LOCA data set.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

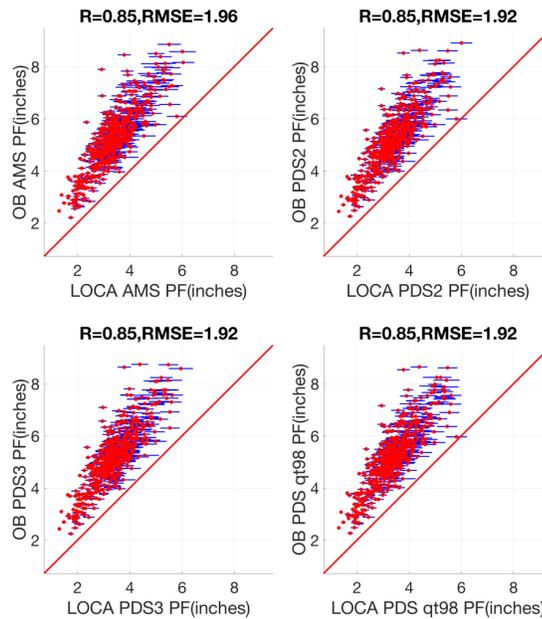


Figure 3.17. PFs scatter plot for Texas. The red line represents $y=x$. Results are shown for return period = 10 years using the LOCA data set.

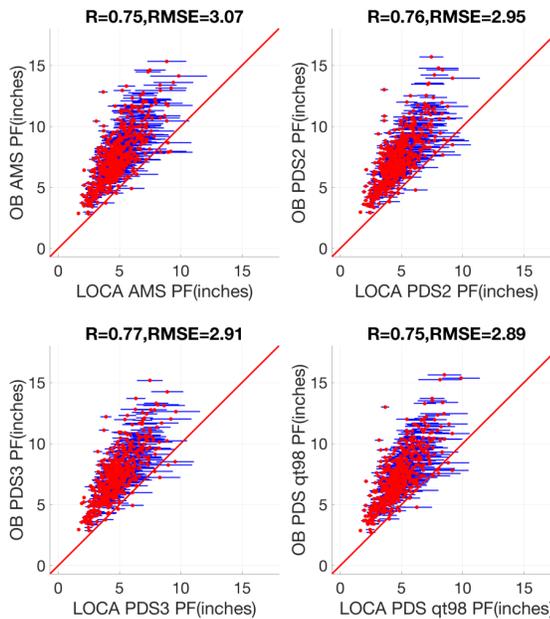


Figure 3.18. PFs scatter plot for Texas. The red line represents $y=x$. Results are shown for return period = 50 years using the LOCA data set.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

3.2.3 Results for NA-CORDEX

The NA-CORDEX data set is the only dynamically downscaled data set with a mixed resolution of 0.44 and 0.22 degrees. Ideally, the whole data set should consist of six climate models paired with seven regional models with three different resolutions (0.11, 0.22, and 0.44). However, this is an ongoing project. We were able to acquire six simulation results from the NA-CORDEX. The following shows the comparison with the observed data. Overall, the dynamically downscaled data sets have a larger spread than the statistically downscaled data. Results (Figs. 3.19-3.26) show that biases for the Northeast US are not very significant, while dry biases are somewhat higher for Texas.

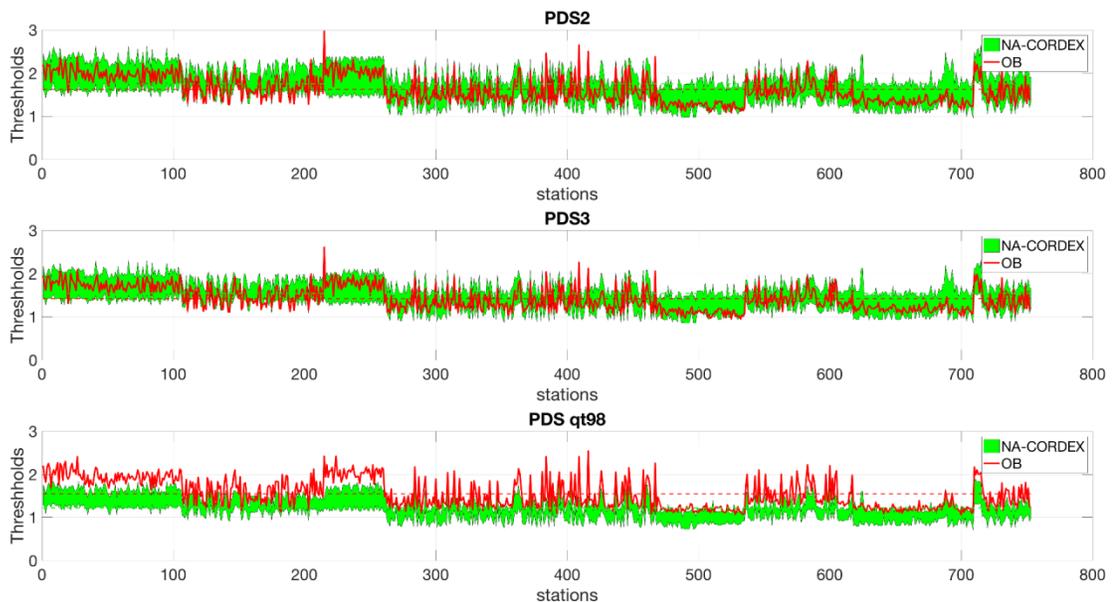


Figure 3.19. PDS thresholds for Northeast using the NA-CORDEX data set.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

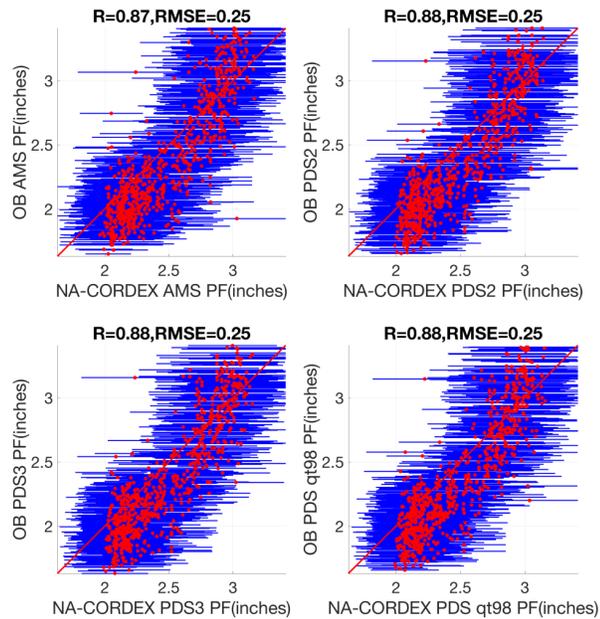


Figure 3.20. PFs scatter plot for Northeast US. The red line represents $y=x$. Results for return period = 2 years using the NA-CORDEX data set.

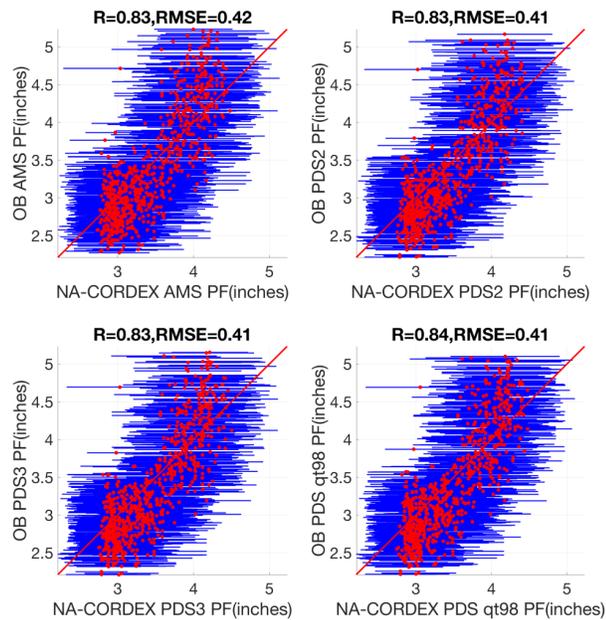


Figure 3.21. PFs scatter plot for Northeast US. The red line represents $y=x$. Results for return period = 10 years using the NA-CORDEX data set.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

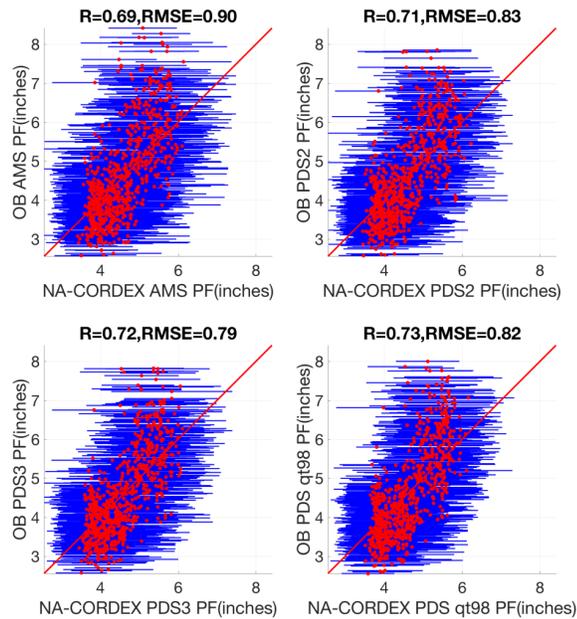


Figure 3.22. PFs scatter plot for Northeast US. The red line represents $y=x$. Results are shown for return period = 50 years using the NA-CORDEX data set.

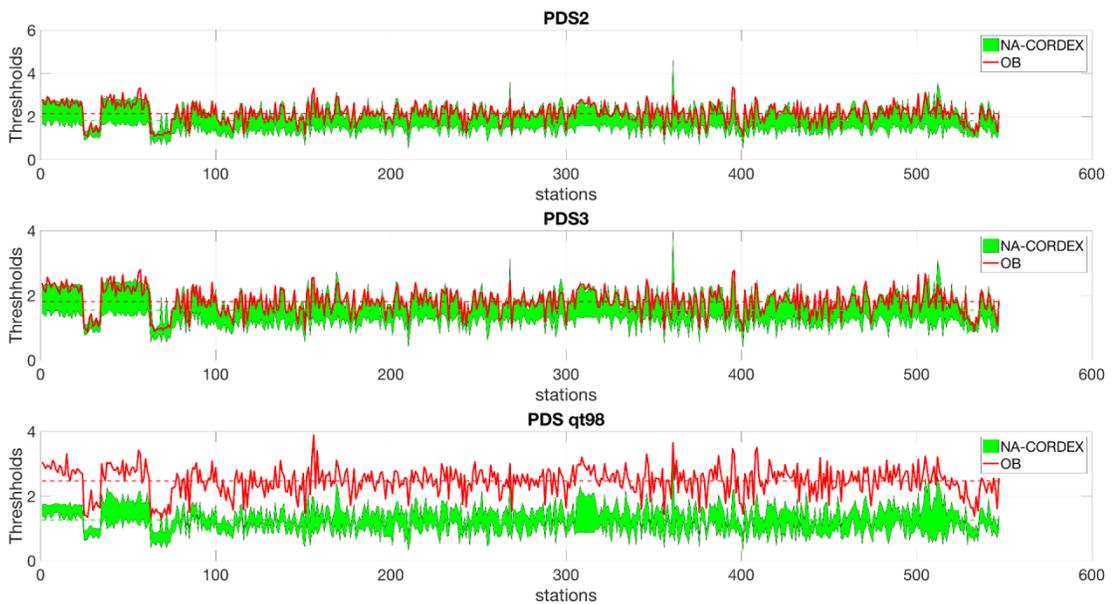


Figure 3.23. PDS thresholds for TX using the NA-CORDEX data set.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

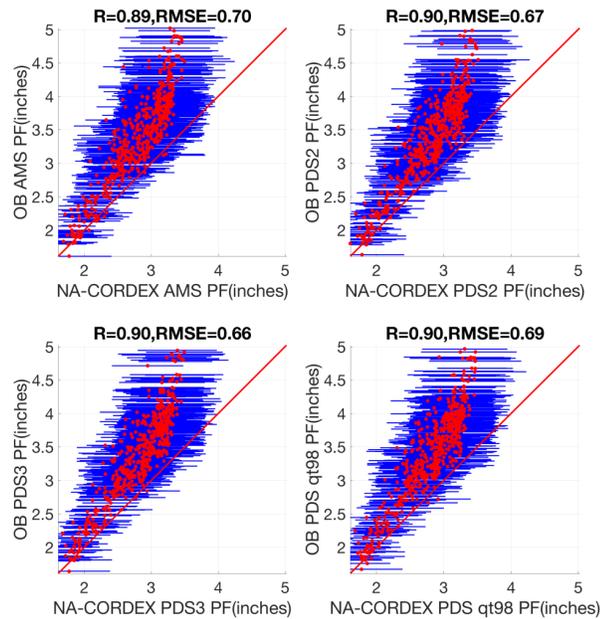


Figure 3.24. PFs scatter plot for Texas. The red line represents $y=x$. Results are shown for return period = 2 years using the NA-CORDEX data set.

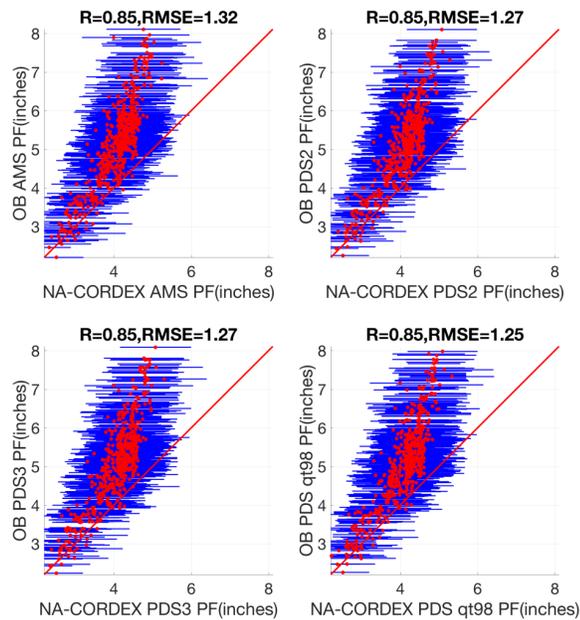


Figure 3.25. PFs scatter plot for Texas. The red line represents $y=x$. Results are shown for return period = 10 years using the NA-CORDEX data set.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

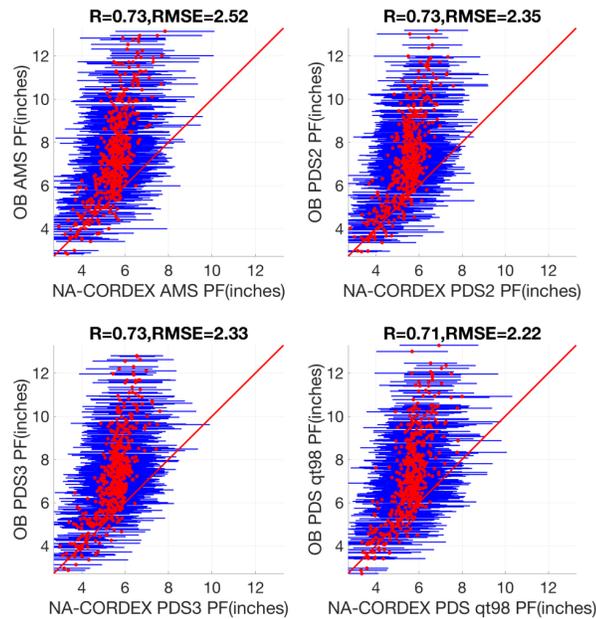


Figure 3.26. PFs scatter plot for Texas. The red line represents $y=x$. Results are shown for return period = 50 years using the NA-CORDEX data set.

3.2.4 Results for BCCAv2

BCCAv2 is another widely used downscaled data set. Like the LOCA, our results (Figs. 3.27-3.34) showed that the BCCAv2 data also have systematic biases in simulating the observed PFs, and again the correlation skill is acceptable.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

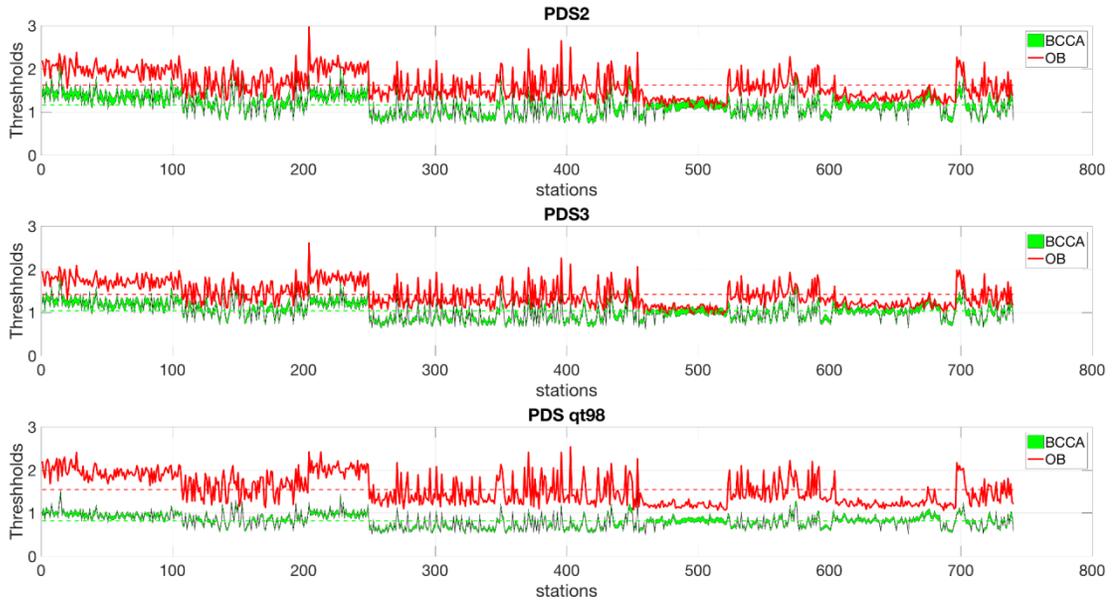


Figure 3.27. PDS thresholds for Northeast US using the BCCAv2 data set.

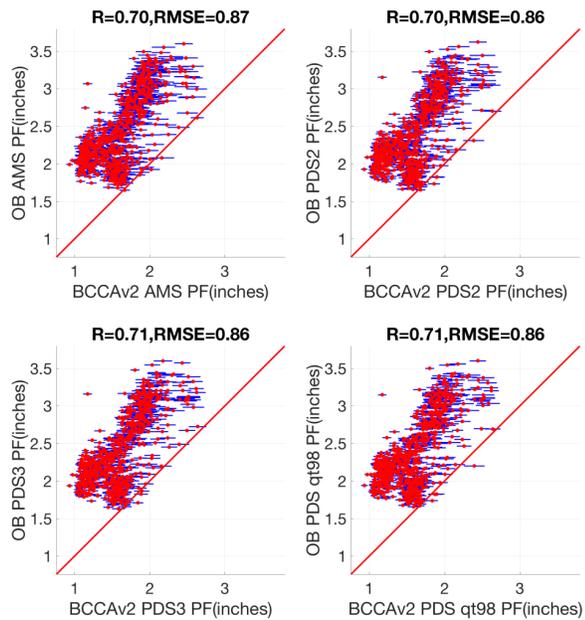


Figure 3.28. PFs scatter plot for Northeast US. The red line represents $y=x$. Results are shown for return period = 2 years using BCCAv2 data set.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

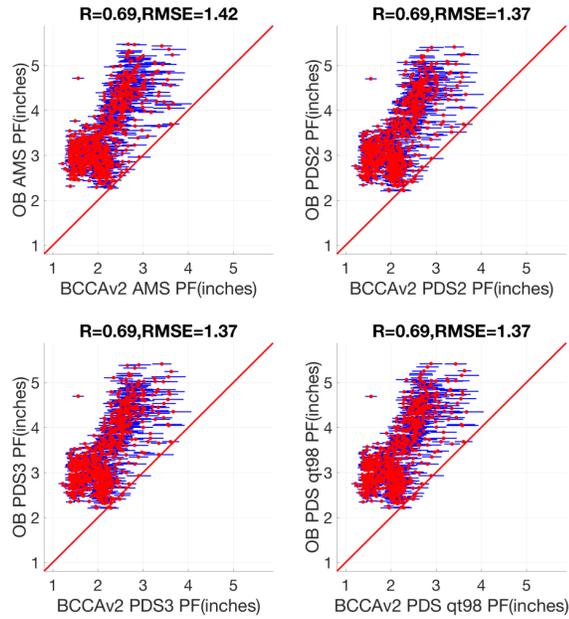


Figure 3.29. PFs scatter plot for Northeast US. The red line represents $y=x$. Results are shown for return period = 10 years using the BCCAv2 data set.

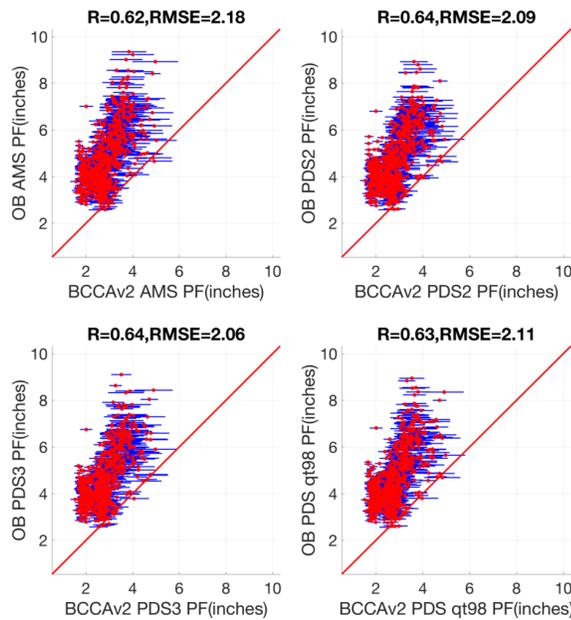


Figure 3.30. PFs scatter plot for Northeast US. The red line represents $y=x$. Results are shown for return period = 50 years using the BCCAv2 data set.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

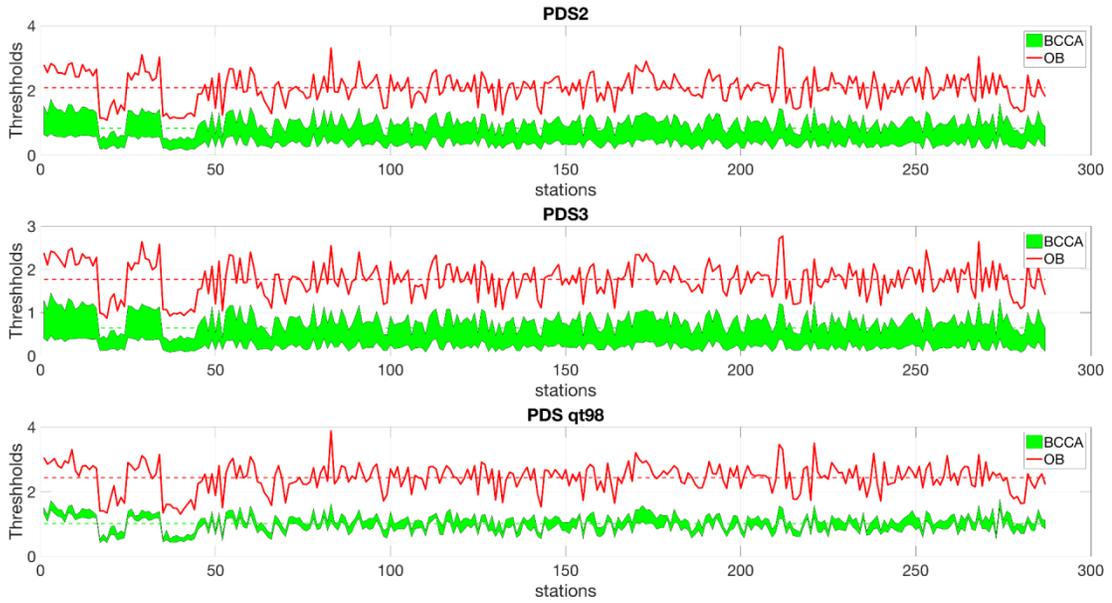


Figure 3.31. PDS thresholds for Texas using the BCCAv2 data set.

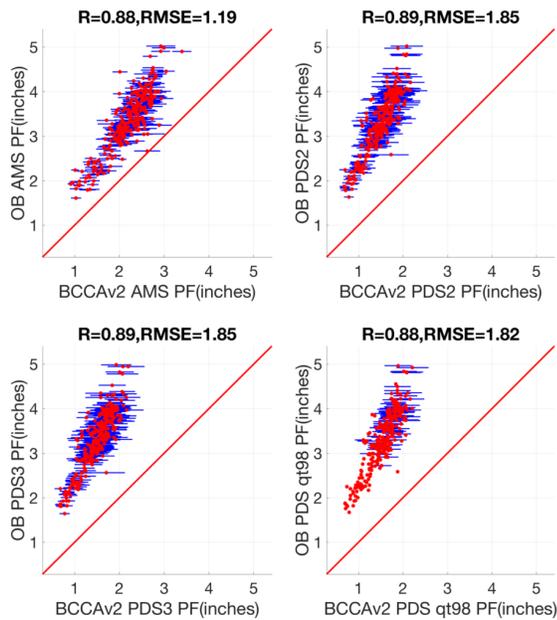


Figure 3.32. PFs scatter plot for Texas. The red line represents $y=x$. Results are shown for return period = 2 years using the BCCAv2 data set.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

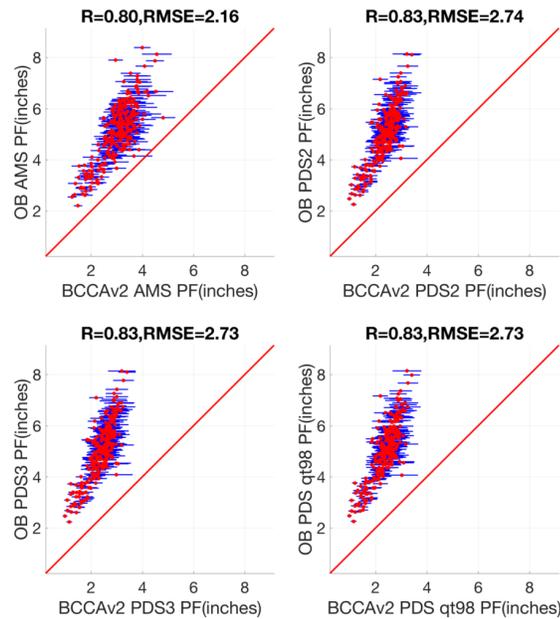


Figure 3.33. PFs scatter plot for Texas, BCCAv2, showing $y=x$. Results are shown for return period = 5 years using the BCCAv2 data set.

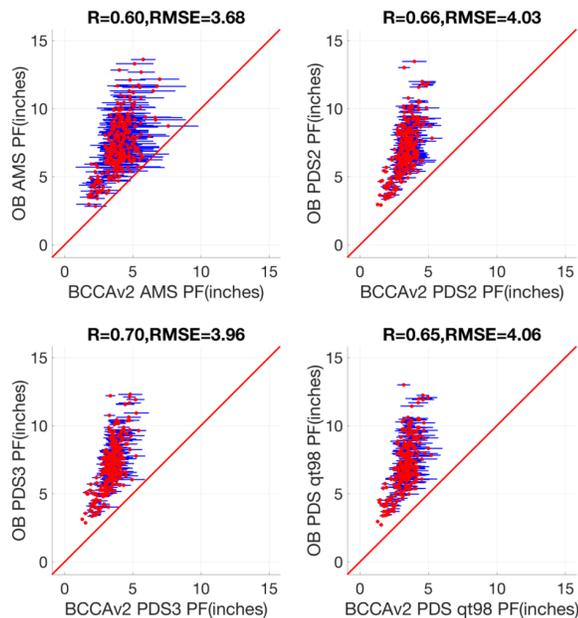


Figure 3.34. PFs scatter plot for Texas. The red line represents $y=x$. Results are shown for return period = 50 years using the BCCAv2 data set.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

3.3. Conclusions

In summary of this section, we compared PFs calculated based on four different downscaled data sets (described in Section I) with those based on observational data. We explored four different methods to calculate PFs, one AMS based and three PDS based. In general, our results showed that all the downscaled data capture the observed PFs fairly well in terms of correlation coefficients and RMSE with UWPD data outperforming the other three data sets (Figs. 3.35-3.38) in the Northeast and in Texas. The UWPD, LOCA, and NA-CORDEX data show a higher skill related to correlation coefficients as compared to the BCCAv2 data. Both the LOCA and BCCAv2 data show significant dry biases.

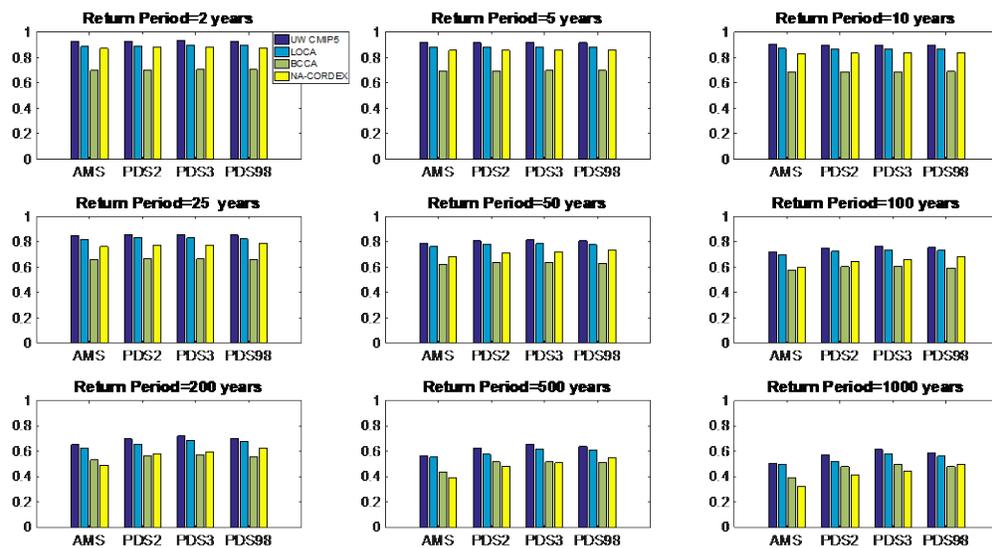


Figure 3.35. Comparison of correlation coefficients between modeled and observed PFs for Northeast US

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

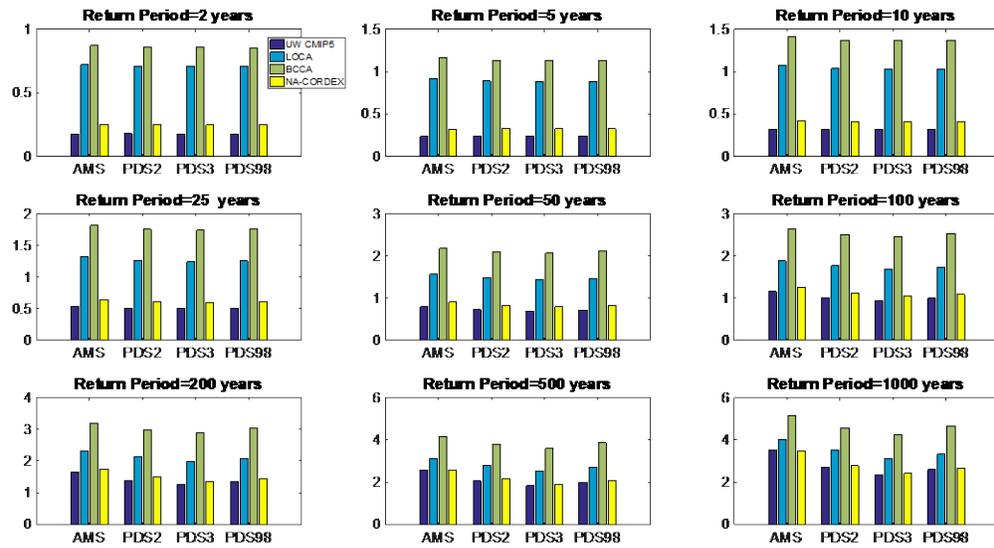


Figure 3.36. RMSE for modeled PFs for Northeast US

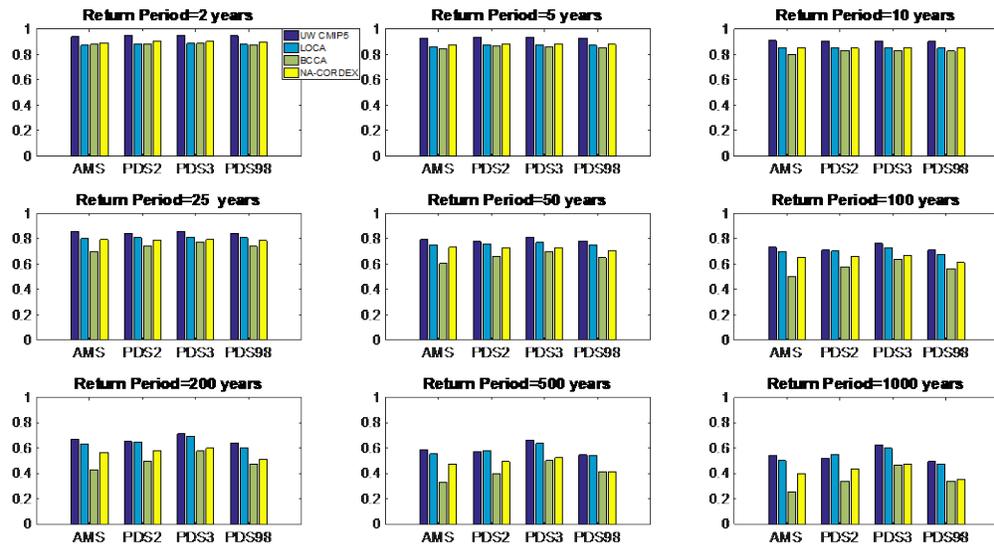


Figure 3.37. Comparison of correlation coefficients between modeled and observed PFs for Texas

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

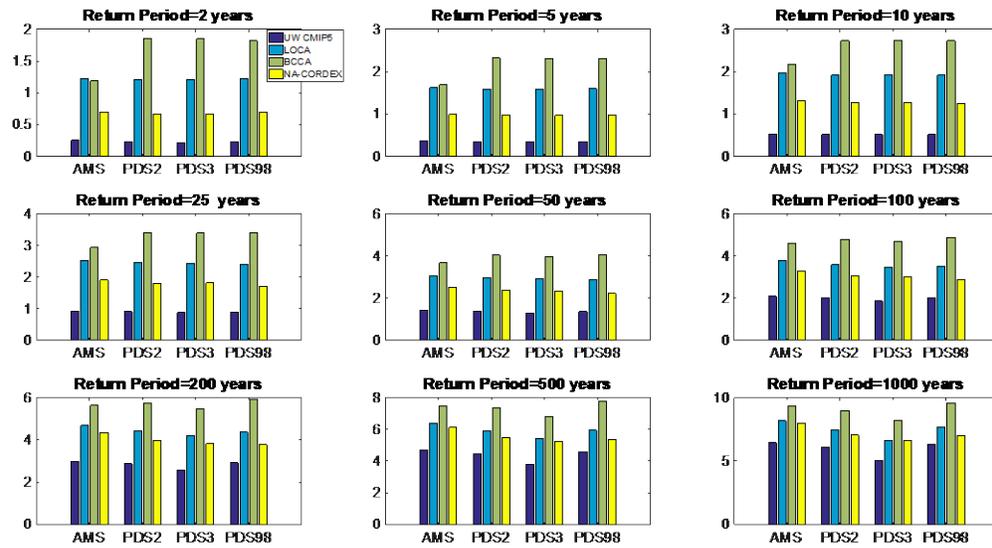


Figure 3.38. RMSE for modeled PFs for Texas

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

SECTION IV. Quasi-stationary (QS) methodology and the spatial variability of the future QS precipitation frequency estimates based on the UWPD data

4.1. Quasi-stationary (QS) PFs estimation methodology

One of the two main approaches used in studying nonstationary precipitation frequency divides the time period into two or more sub-periods and treats each as stationary. In this report, this piecewise stationary method is referred to as the quasi-stationary (QS) approach. The other approach, denoted as the nonstationary (NS) method, presented in Section V, expresses precipitation frequency distribution statistics as a function of selected explanatory variables or covariates. The QS method was developed by researchers at the University of Illinois at Urbana-Champaign and the University of Wisconsin-Madison. The primary NS approach used in this study was originally developed by Penn State researchers but later modified by the HDSC group to use station precipitation and elevation as covariates.

The QS method in this section, similar to Section III, uses the point precipitation estimation method, in which only extreme time series observed at a single station are used to estimate the distribution parameters using the maximum likelihood estimation (MLE) method. Ensemble means at each point were used as frequency estimates. The means based on large numbers of ensemble members and their multiple runs minimize the effects of variability among climate models and outliers.

In this section, an estimate of the future changes in precipitation frequencies was first obtained using the quasi-stationary approach and UWPD data for the entire future period provided by climate models (2006-2100), followed by dividing the same period into two approximately equal parts, 2006-2053 and 2054-2100, and applying the same approach for each part. Future PFs were based on both PDS and AMS approaches.

This section provides an initial estimate of the projected precipitation frequency changes in the Northeast and in Texas based on the QS approach, future projected UWPD data, and both AMS and PDS methods. This section also illustrates how the projected increases vary spatially and provides a comparison to the NCA4. More detailed comparisons between NS and QS using UWPD and LOCA data sets are presented in Section V.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

4.2. Future AMS and PDS QS PFs estimated based on UWPD downscaled data for 2006-2100

Our comparisons between the UWPD and observed data indicated that the UWPD can capture the major features of the observed extreme events. This motivates us to further explore UWPD's downscaled results for the CMIP5 projection runs. As the historical period, the UWPD generates three realizations for each CMIP5 model, resulting in 22 models*3 realization = 66 members for the RCP4.5 scenario and 24 models*3 realization = 72 members for the RCP8.5 scenario. Each member covers the temporal period of 2006-2100. Although there were trends for each time series, in the first iteration we analyzed the projected PFs by treating the whole period as a stationary data set without further dividing it into subsets. These results could be interpreted as an average PF for the period 2006-2100. In the next step, we performed a similar analysis for 2006-2053 and 2054-2100. Later in this report, a more detailed analysis is presented. PFs were calculated for each decade to determine a more accurate evolution of PFs. The results were then compared with the nonstationary method selected to calculate PFs in this study.

4.2.1. Results for RCP4.5 scenario

For the Northeast US, the RCP4.5 PDS thresholds are larger than in the historical period (Fig. 4.1). All the points in the PFs scatter plot (Figs. 4.2-4.4) are below the line of “ $y=x$ ”, meaning the projected frequency estimates are higher than those of the historical period. The red dots in the figures represent the ensemble mean values, while the blue bars represent the uncertainty range estimated as the standard deviation of all ensemble values. Also shown in the figures are the best least squares fit lines in green. This line can give rough estimates of future PF values based on their historical values. Note that since our x-coordinate represents the future values, the least squares fit is optimized according to minimization along the x-direction. Overall, the green lines are roughly consistent with the red lines, meaning the absolute changes are similar across all the stations. It should be noted that each chart shows a couple of outliers.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

This result could be due to the unstable parameters calculated by the Maximum Likelihood method. These results can be adjusted by spatial smoothing.

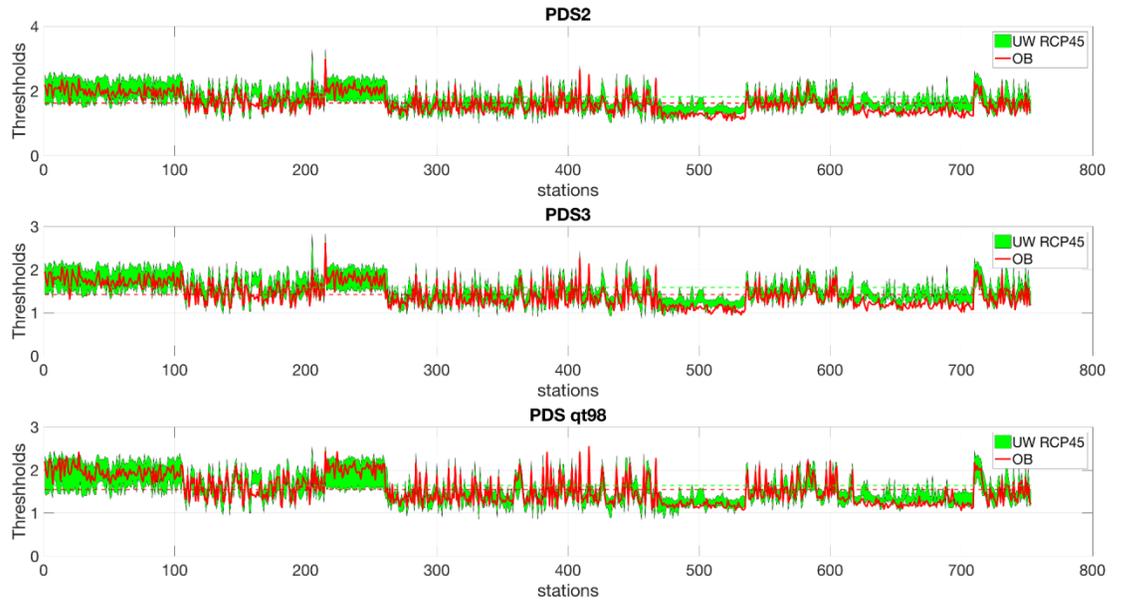


Figure 4.1. PDS thresholds for Northeast US using the projected data. UWPD RCP4.5.

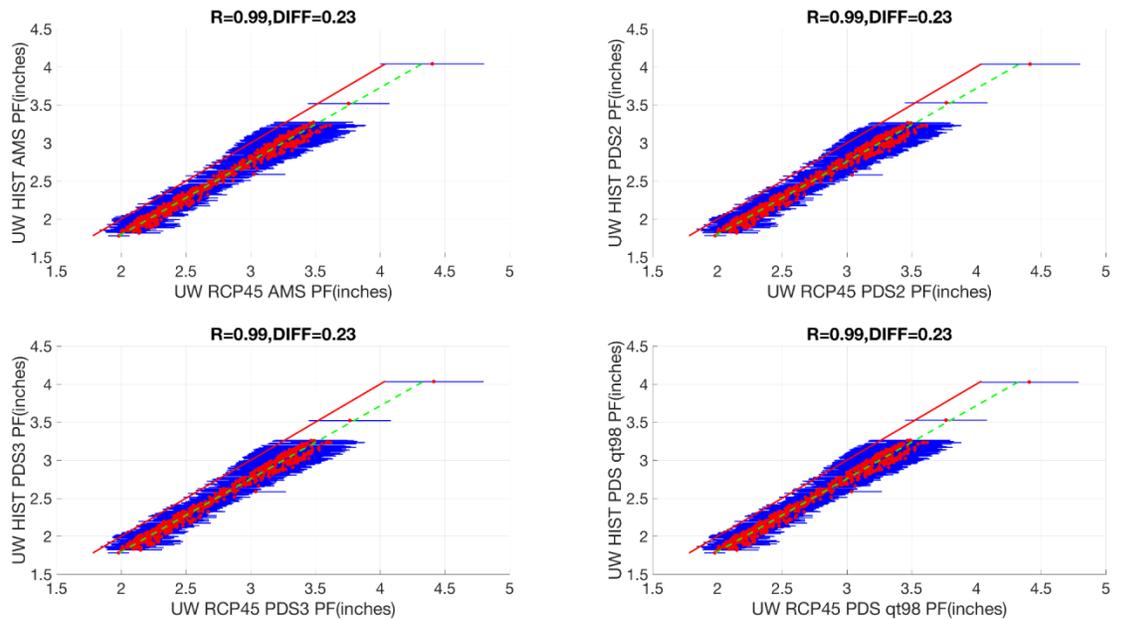


Figure 4.2. PFs scatter plot for Northeast US. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 2 years. UWPD RCP4.5.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

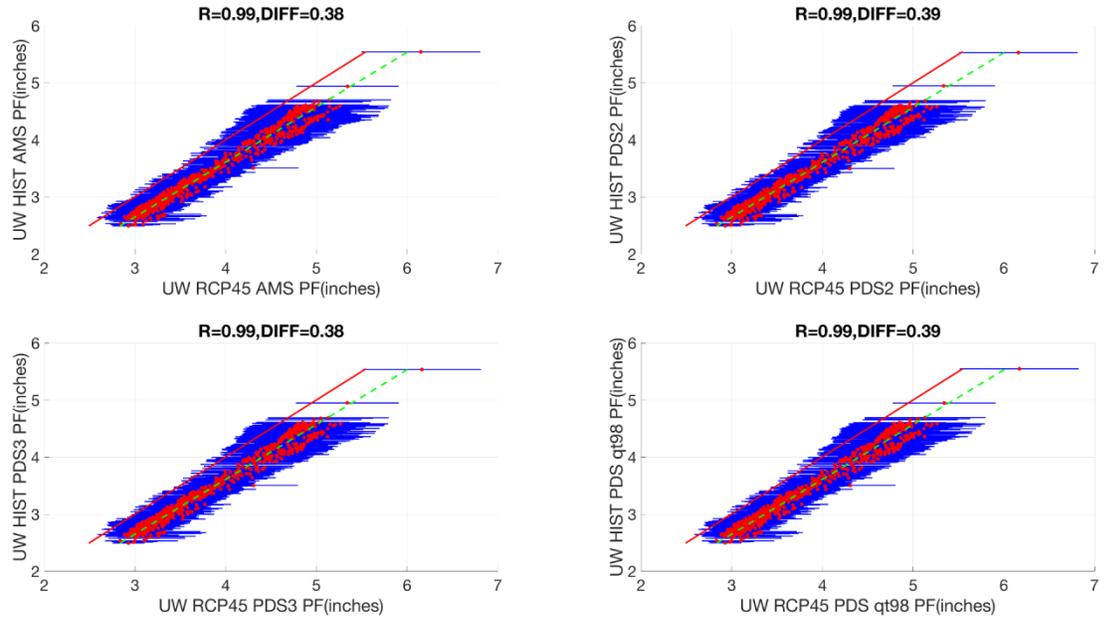


Figure 4.3. PFs scatter plot for Northeast US. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 10 years. UWPD RCP4.5.

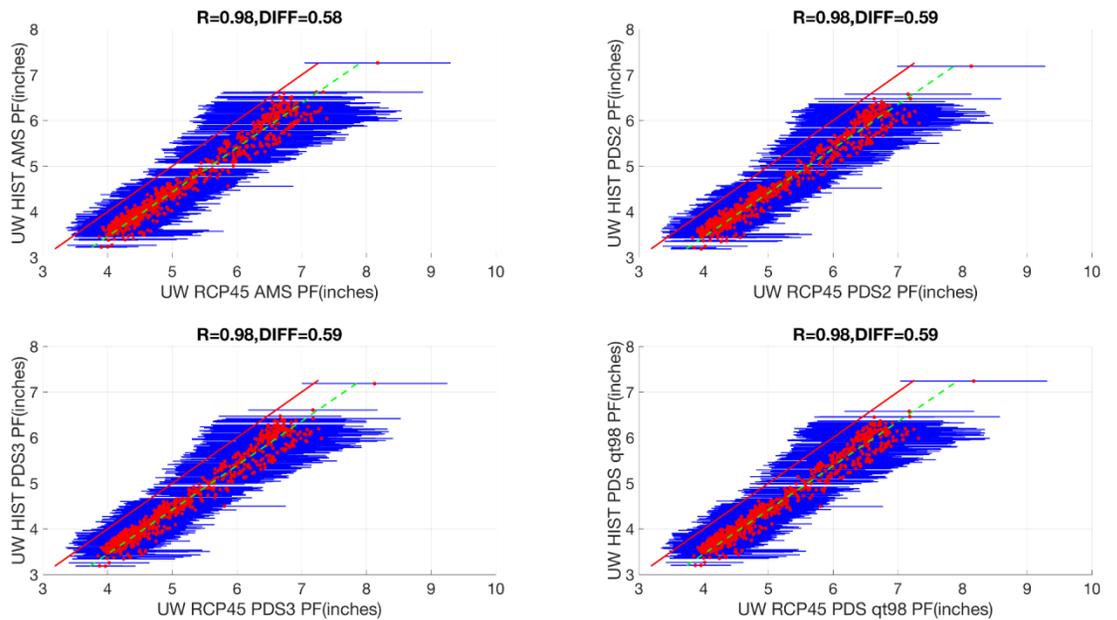


Figure 4.4. PFs scatter plot for Northeast US. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 50 years. UWPD RCP4.5.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

The changes of Texas under the RCP4.5 scenario appear to be smaller than those in the Northeast US. The RCP4.5 PDS thresholds (Fig. 4.5) are very similar to that of the historical results. There is a slight increase for PFs (Figs. 4.6-4.8). The green lines are not parallel with “ $y=x$ ”, rather they tilt down toward the x-direction indicating that stations that have larger PFs historically tend to increase more.

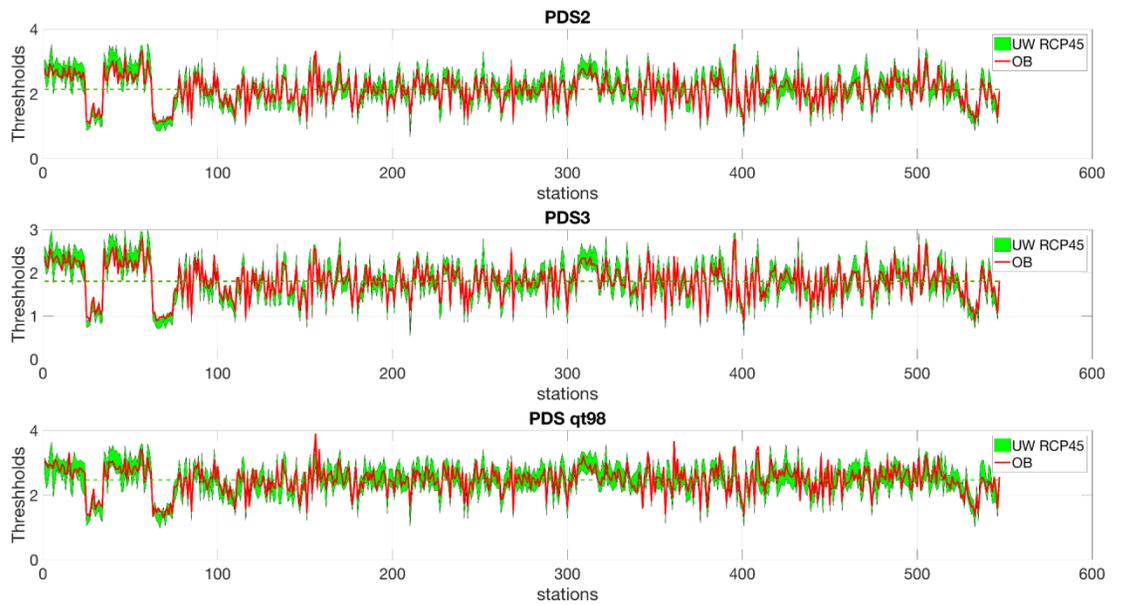


Figure 4.5. PDS thresholds for Texas using projected data. UWPD RCP4.5.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

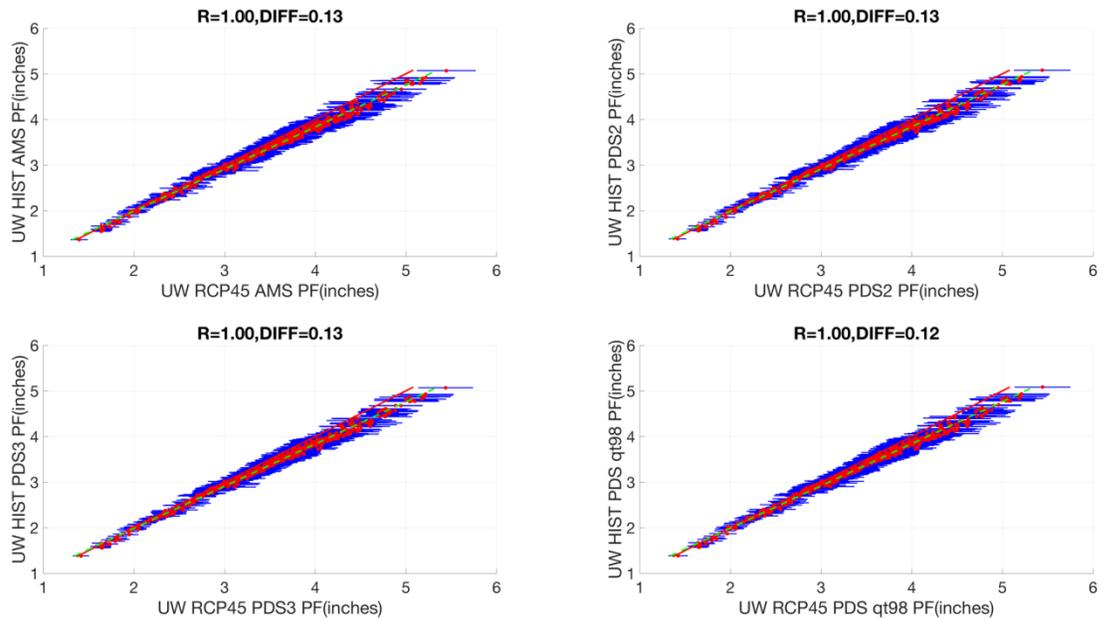


Figure 4.6. PFs scatter plot for Texas. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 2 years. UWPD RCP4.5.

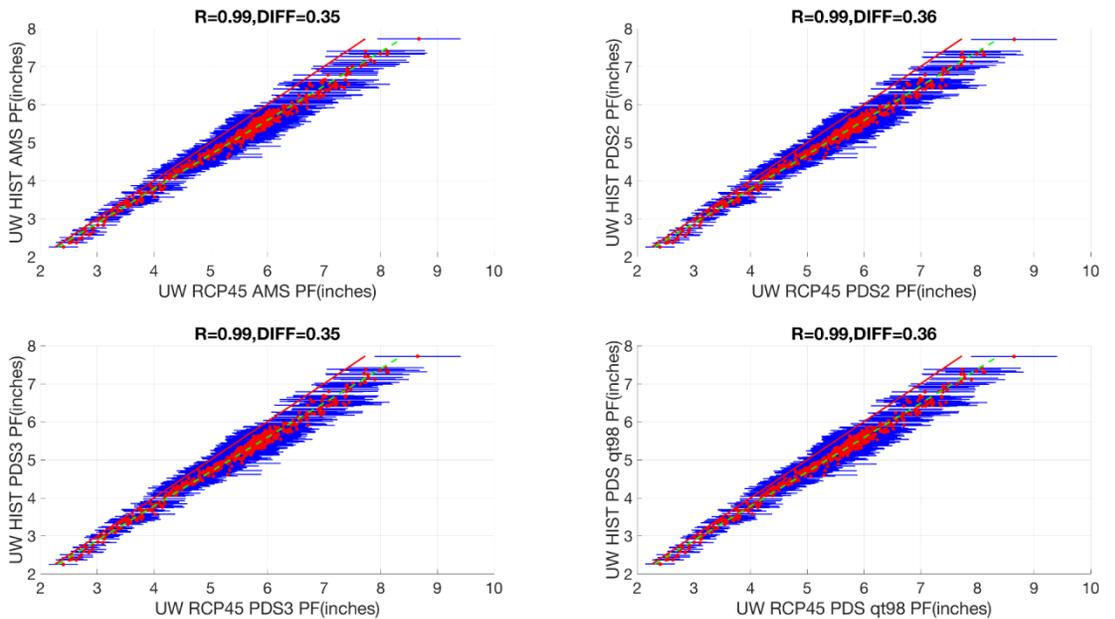


Figure 4.7. PFs scatter plot for Texas. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 10 years. UWPD RCP4.5.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

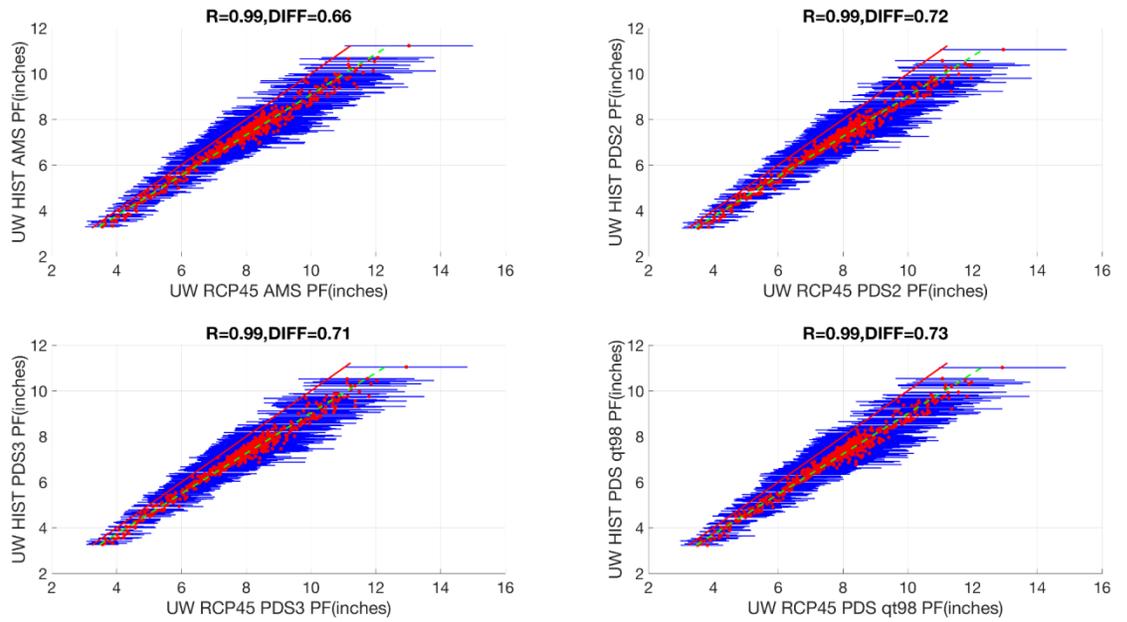


Figure 4.8. PFs scatter plot for Texas. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 50 years. UWPD RCP4.5.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

4.2.2. Results for RCP8.5 scenario

The RCP8.5 results (Figs. 4.9-4.16) have similar features with RCP4.5, but the relative increases with respect to the results based on the observed data are larger. The red dots in Figs. 4.10-4.12 and Figs. 4.14-4.16 represent the ensemble mean values, while the blue bars represent the uncertainty range estimated as the standard deviation of all ensemble values.

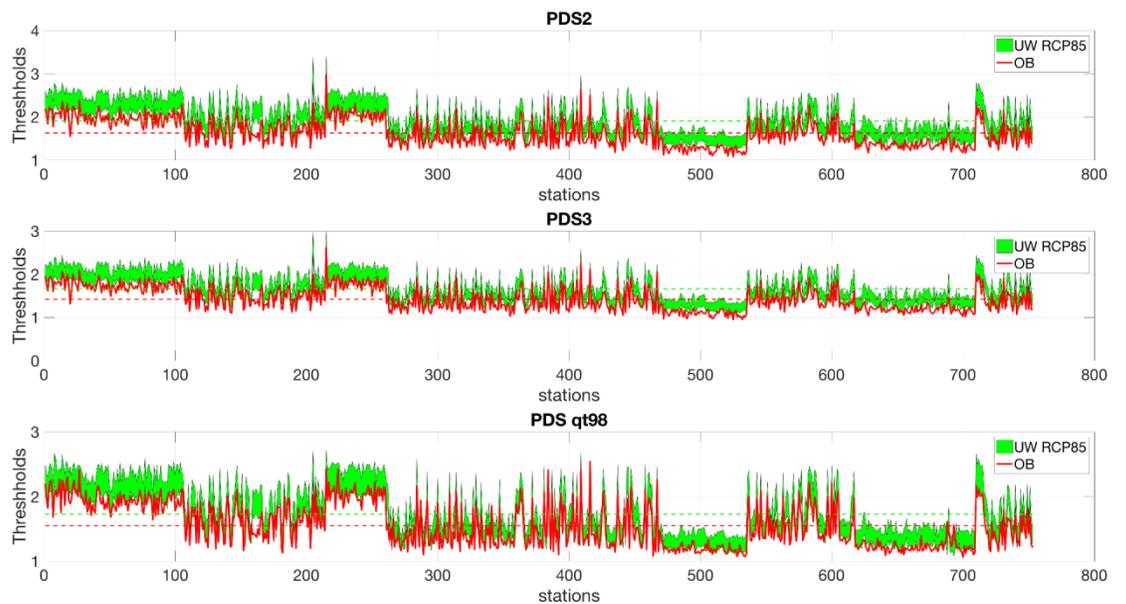


Figure 4.9. PDS thresholds for Northeast US using projected data based on UWPD CMIP5 RCP8.5.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

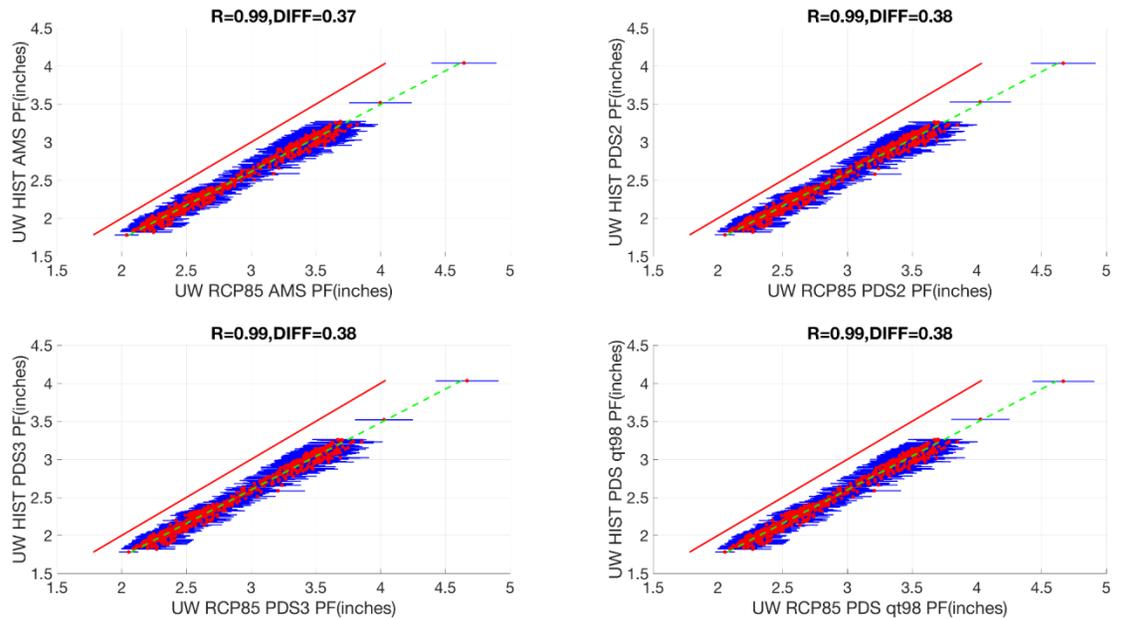


Figure 4.10. PFs scatter plot for Northeast US. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 2 years. UWPD RCP8.5.

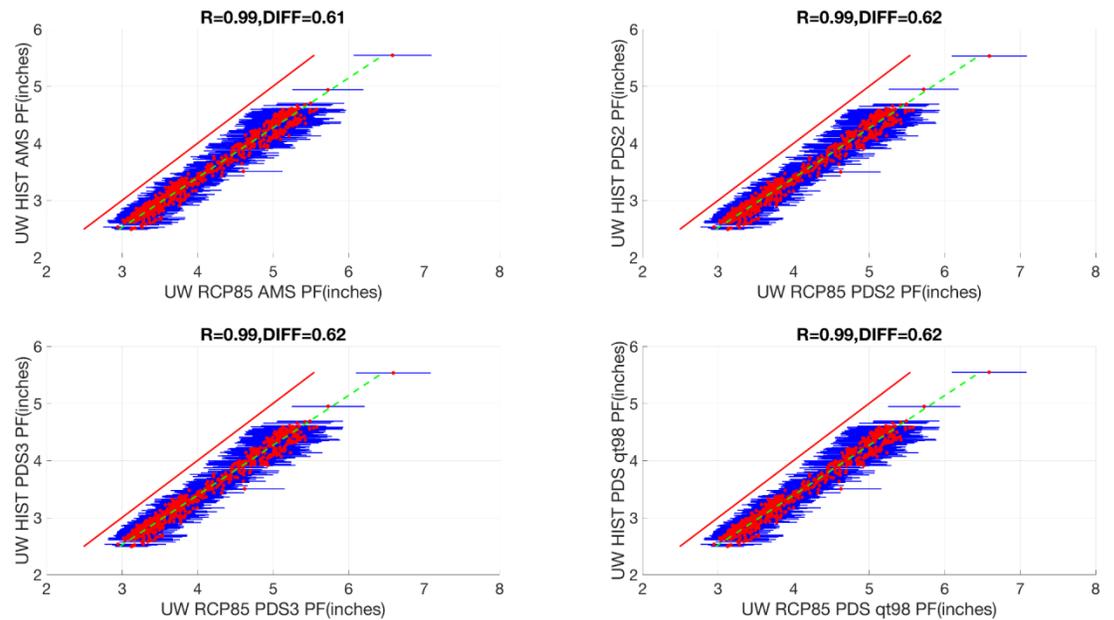


Figure 4.11. PFs scatter plot for Northeast US. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 10 years. UWPD RCP8.5.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

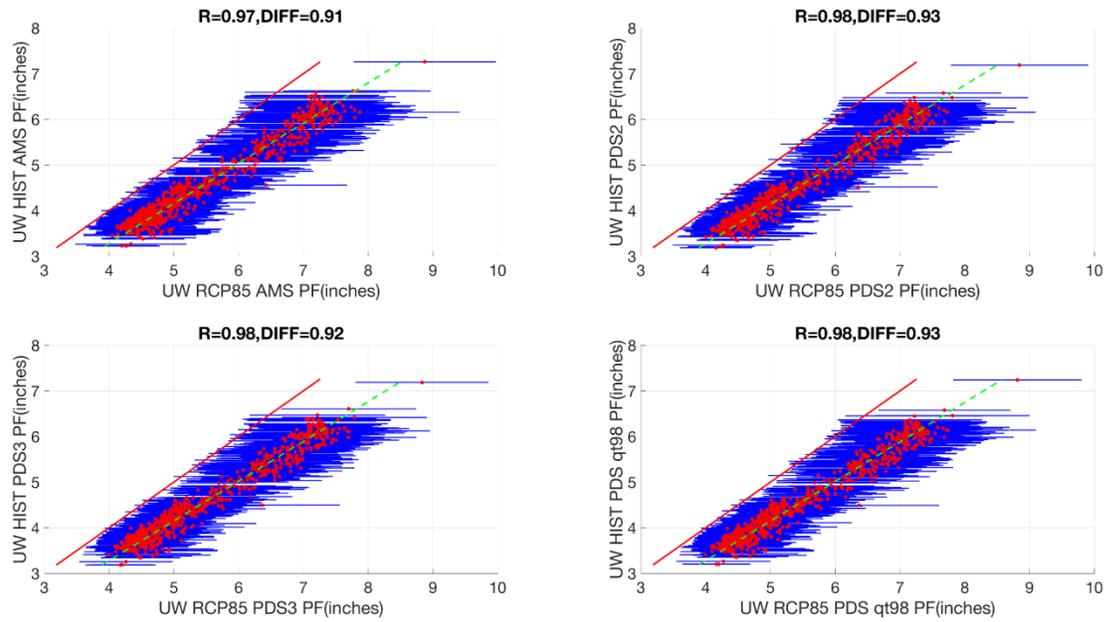


Figure 4.12. PFs scatter plot for Northeast US. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 50 years. UWPD RCP8.5.

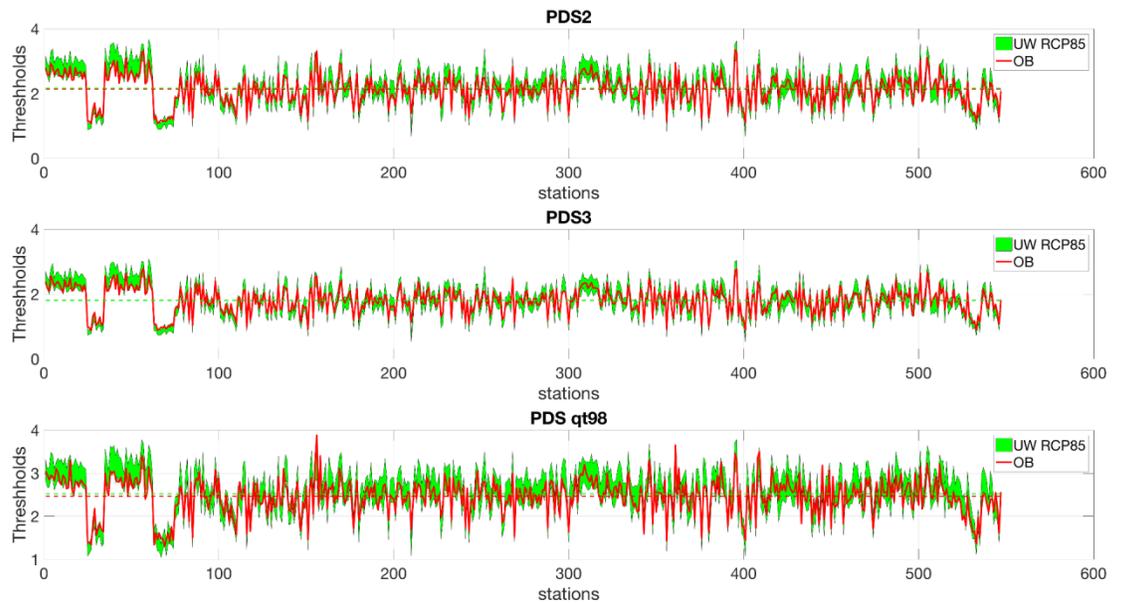


Figure 4.13. PDS thresholds for Texas using the projected UWPD RCP8.5 data.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

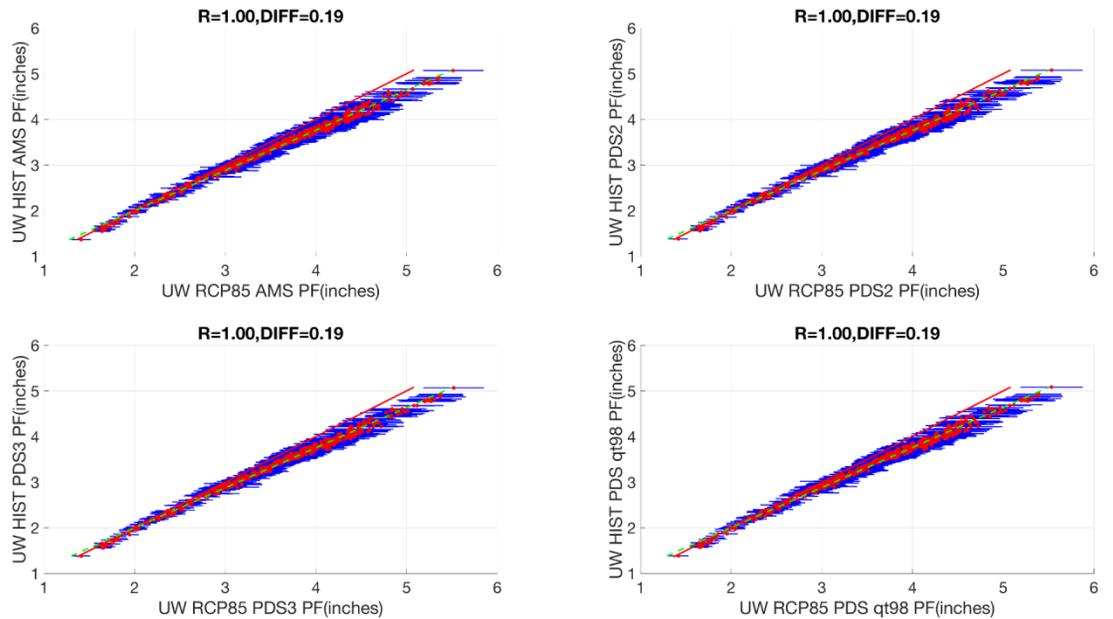


Figure 4.14. PFs scatter plot for Texas. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 2 years. UWPD RCP8.5.

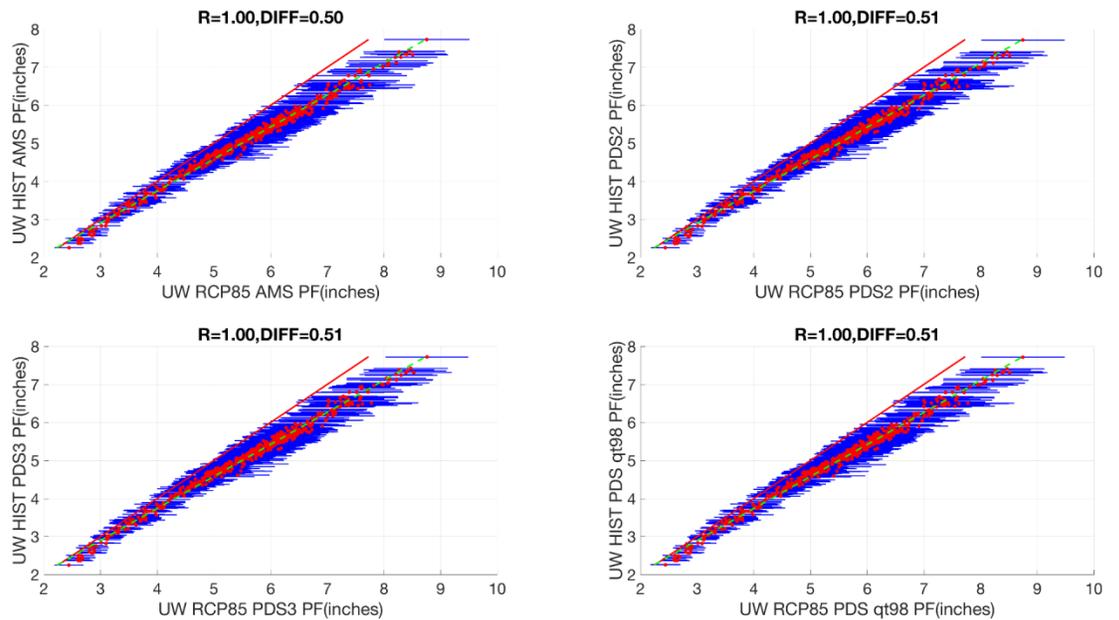


Figure 4.15. PFs scatter plot for Texas. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 10 years. UWPD RCP8.5.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

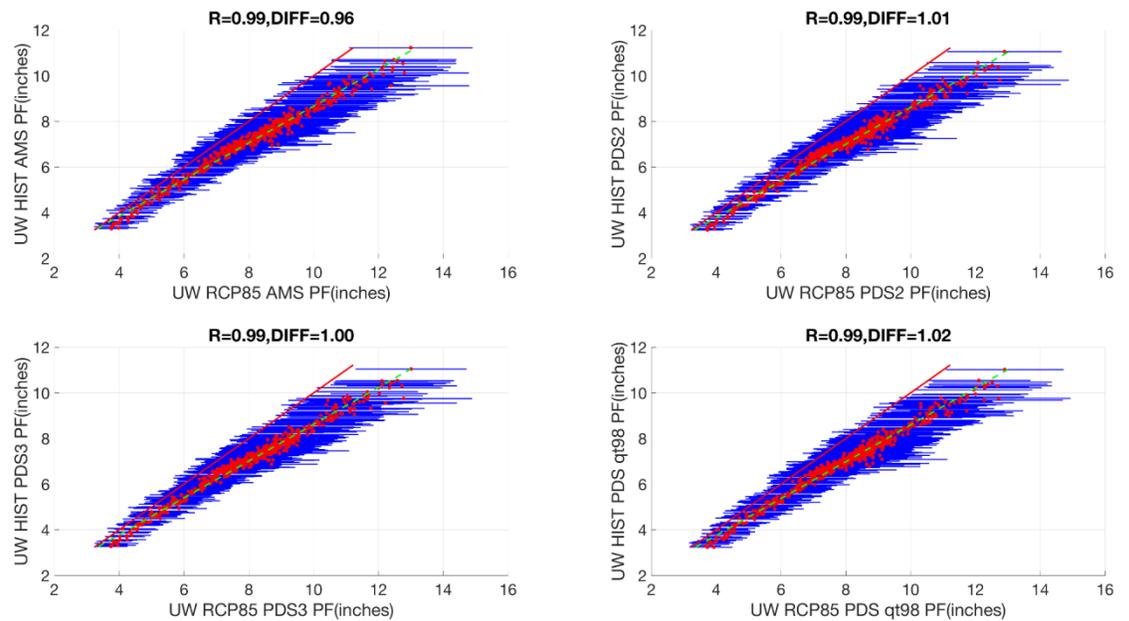


Figure 4.16. PFs scatter plot for Texas. The red line represents $y=x$. Green dashed line represents the best least squares fit. Results for return period = 50 years. UWPD RCP8.5.

It can be concluded that AMS and three PDS methods produced very comparable results when tested using the projected UWPD data. This conclusion is valid for both the Northeast (Figs. 4.10-4.12) and Texas (Figs. 4.14-4.16). Not only are the ensemble mean values (red) based on AMS and the three PDS methods similar, but also the standard deviations of the four approaches are very comparable.

4.2.3. Spatial distributions of projected changes for RCP4.5 and RCP8.5

In this section, we explore the spatial distributions of the projected PF change. From the above context, we have seen that the three PDS thresholds produce approximately similar results. Therefore, we select PDS3 as a representative of PDS analysis to explore the spatial distributions of PFs. Figs. 4.17-4.19 show the AMS and PDS PF distribution for the Northeast US for return periods of 2, 10, and 50 years. PFs based on observed data display larger values along the coastal area in a northeast and southwest direction. The PF values gradually decrease toward the inland area. The differences between the UWPD historical PFs and observational PFs are less than 10%

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

for most of the stations. As no clear spatial patterns associated with the error distribution were detected, it could be concluded that these differences are likely due to random errors. Compared with the historical PFs, the RCP4.5 results show a systematic increase across almost all the stations with the amplitudes ranging from 10-15% in the northeast to 5-10% in the southwest. The RCP8.5 shows a more significant increase, which adds approximately 5% over the RCP4.5 results with the spatial pattern remaining similar.

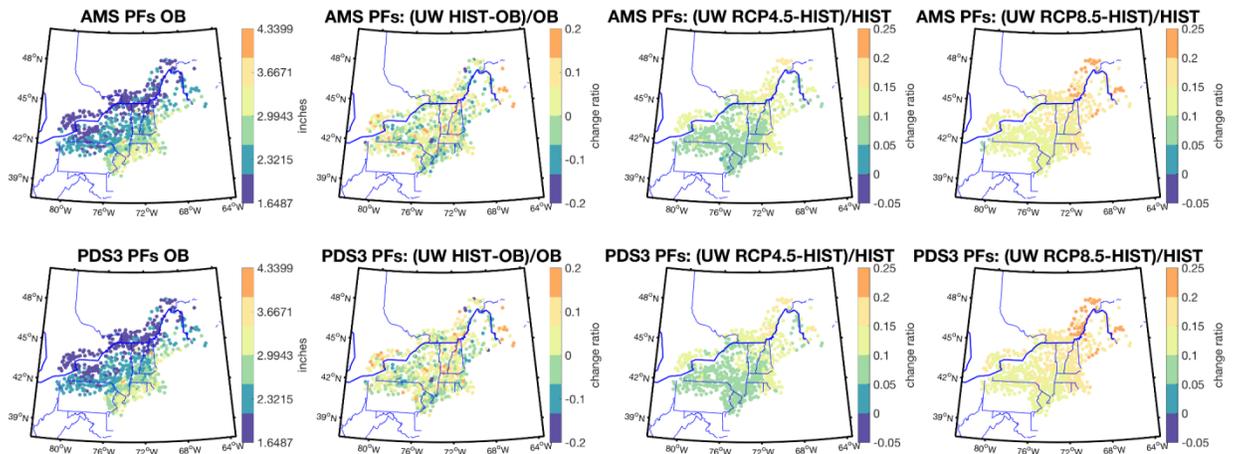


Figure 4.17. Spatial distribution of PFs for return period of 2 years. Results for Northeast US.

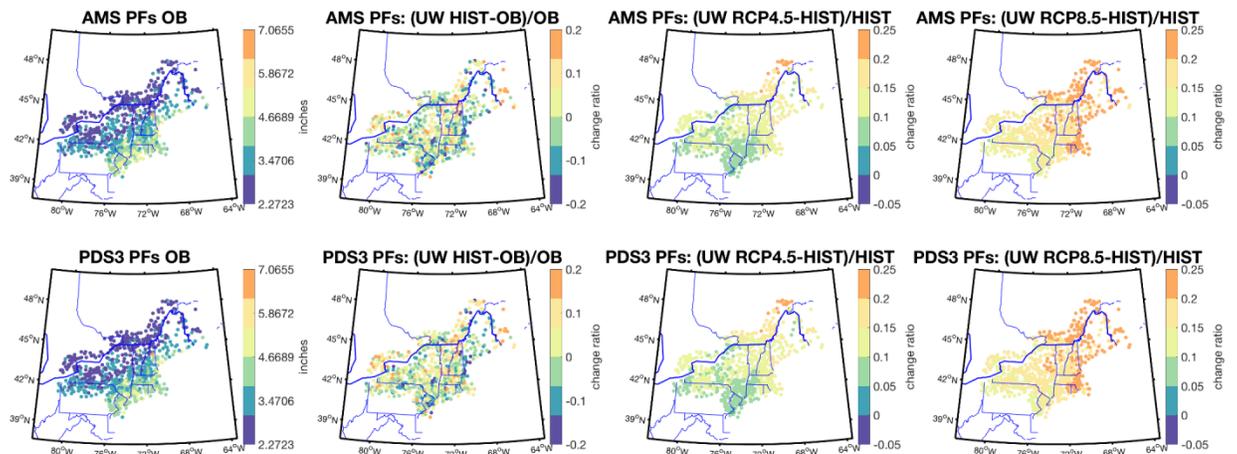


Figure 4.18. Spatial distribution of PFs for return period of 10 years. Results for Northeast US.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

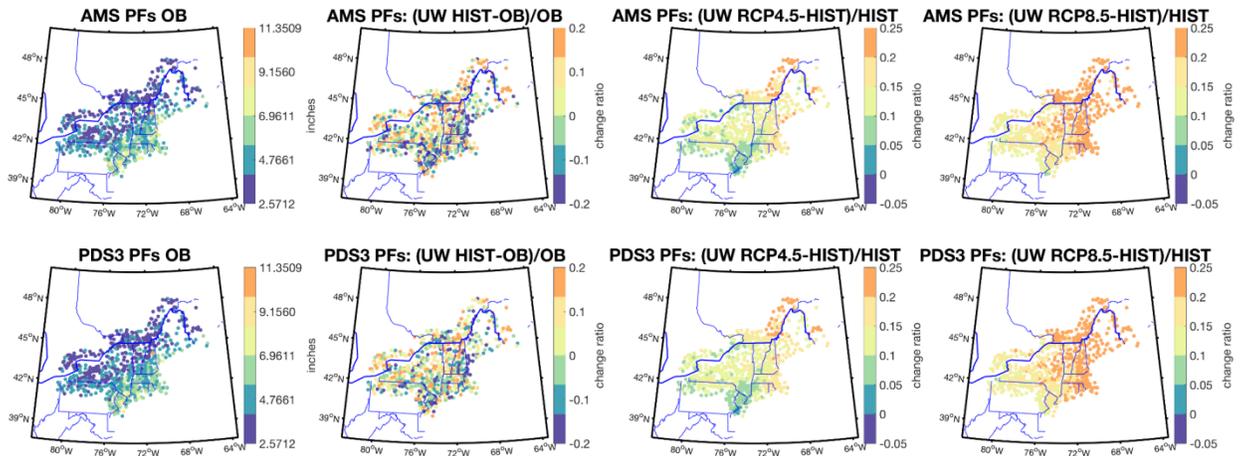


Figure 4.19. Spatial distribution of PFs for return period of 50 years. Results for Northeast US.

PF estimates in Texas are shown in Figs. 4.20-4.22. The largest values are also along the coastal area and gradually decrease along the southeast and northwest direction. For a 2-year return period, the difference between the historical downscaled data and observations are small (<5%). The RCP results also show a marginal increase with a small group of stations located in the Northeast even showing a decrease. For longer return periods (10 and 50 years), the increases of RCP results are getting more systematic, and again the RCP8.5 has a larger increase than the RCP4.5.

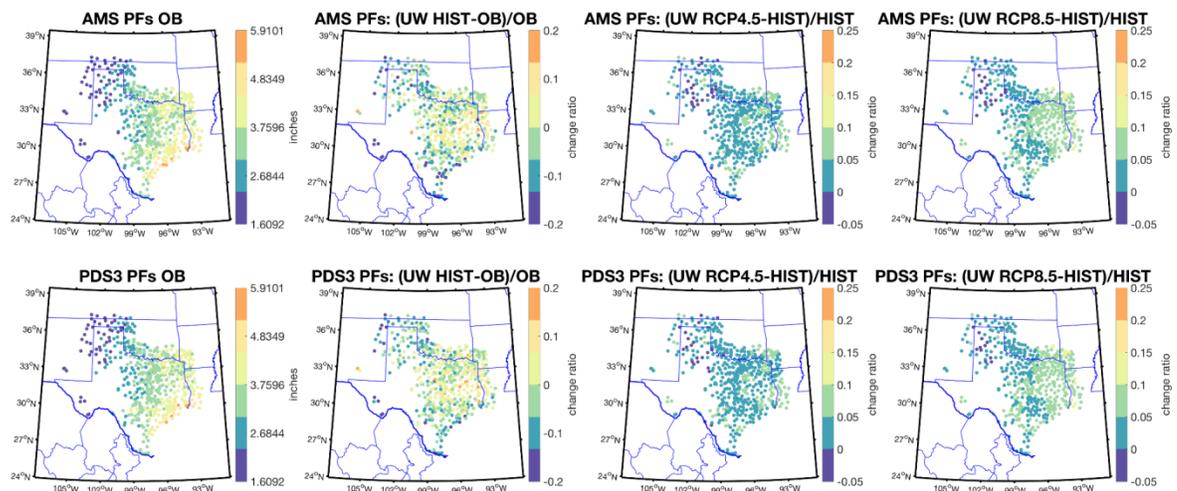


Figure 4.20. Spatial distribution of PFs for return period of 2 years. Results for Texas.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

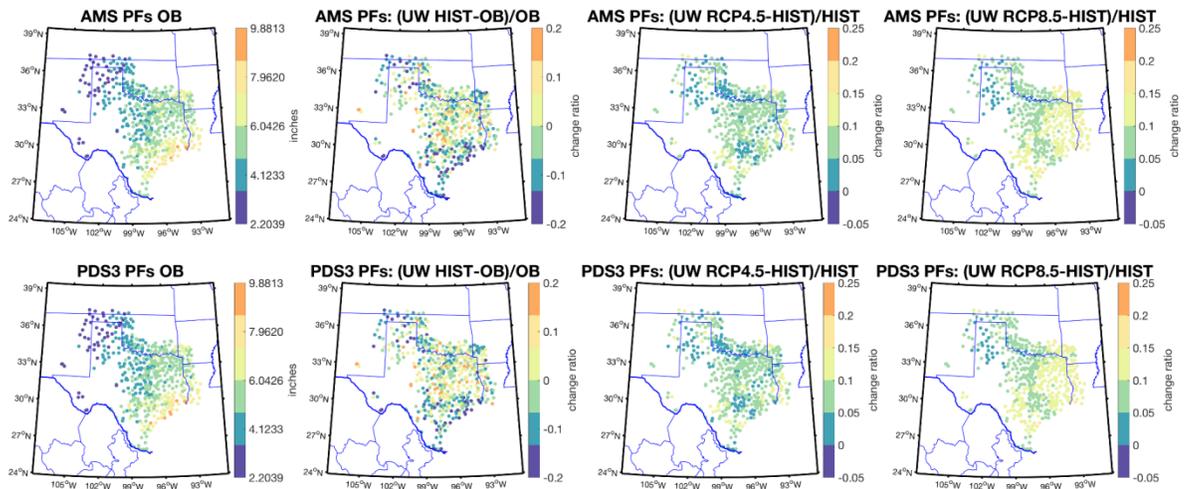


Figure 4.21. Spatial distribution of PFs for return period of 10 years. Results for Texas.

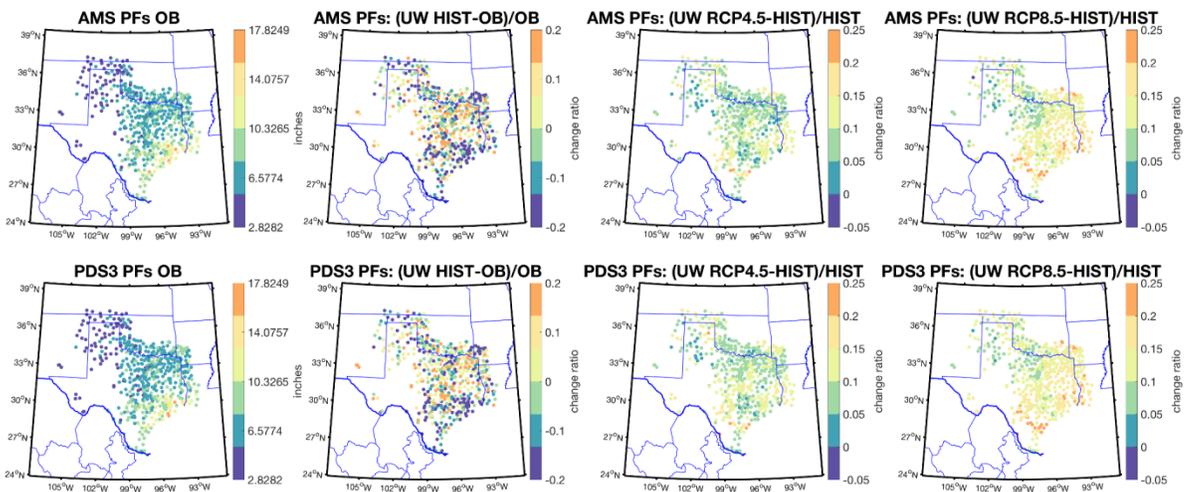


Figure 4.22. Spatial distribution of PFs for return period of 50 years. Results for Texas.

4.3. Spatial distribution of future QS PFs estimated based on UWPD downscaled data for 2006-2053 and 2054-2100 for 25-year return period and a comparison with NCA4

As opposed to the above framework, where the 2006-2100 period was treated as a whole to calculate the QS PFs, in this segment, we further divided the whole time period into two, 2006-2053 and 2054-2100, to study if and how much the PFs changed between the two periods. We also compared the results of these two periods with a historical period (1960-2005) to

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

evaluate the possible changes of PFs. For the low emissions scenario (RCP4.5) in the Northeast US (Fig. 4.23), 25-yr PFs increase by about 5%-15% and 10%-25% for the first and second half of the 21st century, respectively. For the high emissions scenario (RCP8.5), PFs increase by about 10%-20% for the first half of the 21st century and 20%-25% for the second half. While for Texas (Fig. 4.24) the overall increases of PFs are smaller, under both RCP4.5 and RCP8.5 scenarios, PF changes are about 0-15% (5%-15%) for the first half of the 21st century and 5%-15% (15%-25%) for the second half. Regional averages for the Northeast (Wu et al. 2019) and those published in NCA4 (Easterling et al., 2017) (Fig. 3.1) are shown in Table 4.1. Our results are similar to those published in the NCA4, although there are some differences in terms of different return periods (25-year vs. 20-year), the areal extent represented by averages, the methodology, and the time period used in the analysis. Results based on the PDS3 series (Fig. 4.25-4.26) are similar to those based on the AMS.

Table 4.1. Summary comparison between Easterling et al. (2017) and present study. The numbers denote future projected percent increases.

	RCP4.5		RCP8.5	
	Mid-century	Late century	Mid-century	Late century
NCA4 (Easterling et al. 2017)	10	14	13	22
Present study (Wu et al. 2019)	9.14	14.80	12.07	24.48

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

2006-2053

2054-2100

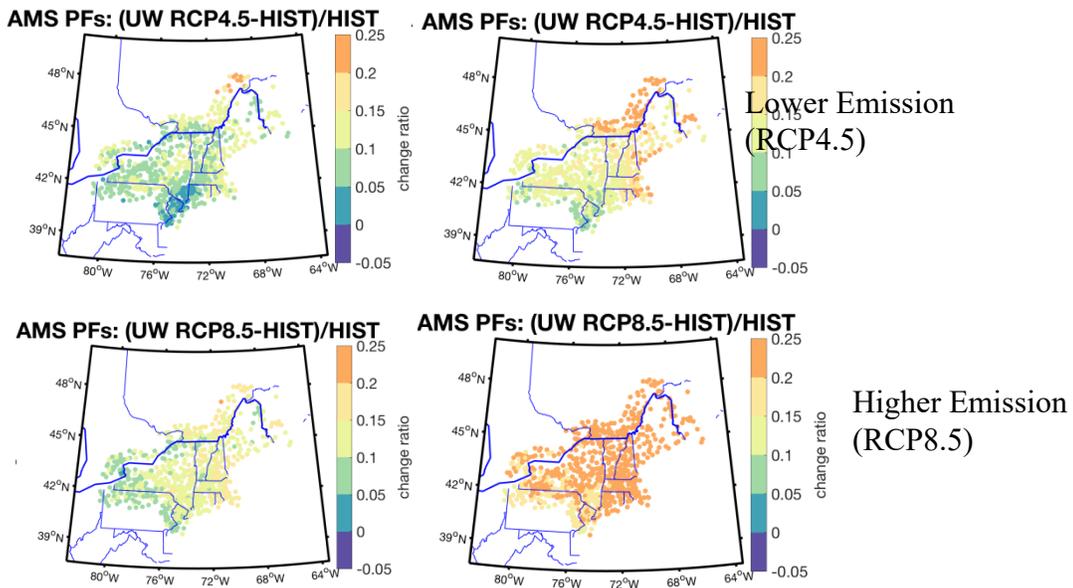


Figure 4.23. Future AMS PF changes for return period = 25 years. Results for Northeast US.

2006-2053

2054-2100

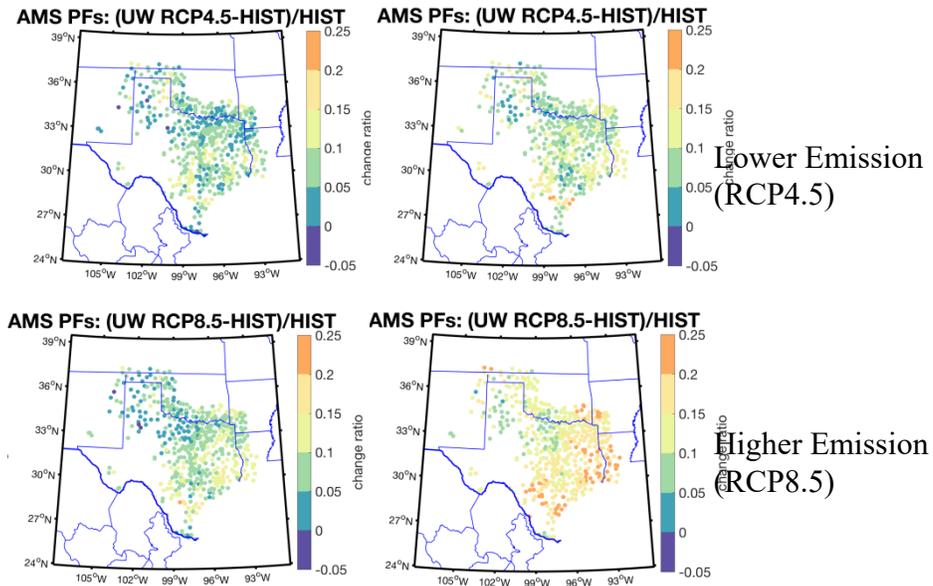


Figure 4.24. Future AMS PF changes for return period = 25 years. Results for Texas.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

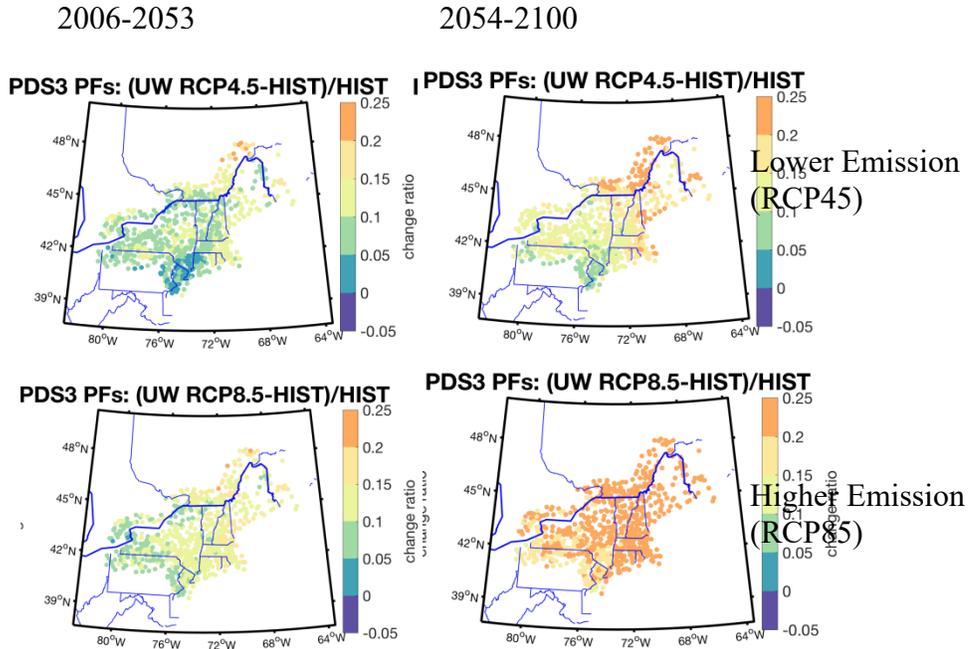


Figure 4.25. Future PDS PF changes for return period = 25 years. Results for Northeast US.

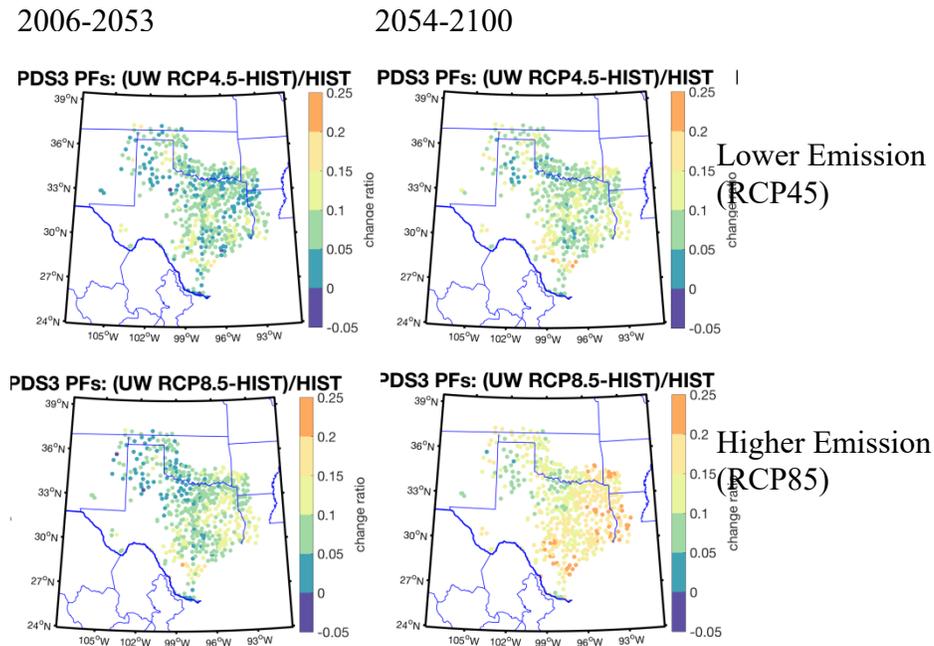


Figure 4.26. Future PDS PF changes for return period = 25 years. Results for Texas.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

4.4. Summary

The results based on AMS and three PDS methods using the UWPD data were very comparable to each other in both the Northeast and in Texas. Spatial distributions indicate that the northeastern parts of the Northeast will have the largest increases, similar to the southwestern section of Texas, while on average the projected increases in the Northeast were larger than those in Texas. Spatial averages indicate that for both regions, the point QS method produced results similar to the NCA4. This similarity is further supported for the Northeast by the results of Wu et al. (2019), shown in Table 4.1.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

SECTION V. Comparisons between nonstationary and quasi-stationary precipitation frequency estimates for AMS and PDS methods, LOCA and UWPD data sets, and RCP4.5 and RCP8.5 scenarios

In this section we first tested the NS method using the coarse CMIP5 precipitation ensemble average as a covariate (NS-CMIP5), then the NS method using each climate downscaled data set as a covariate (NS-UWPD, NS-LOCA), and quasi-stationary (QS-LOCA, QS-UWPD), which uses a direct statistical frequency analysis of each member of the downscaled data from the ensemble of models. These comparisons included projected daily 2-year, 100-year, and 1,000-year frequency estimates using AMS and PDS methods for LOCA and UWPD datasets and for climate scenarios RCP4.5 and RCP8.5. Instead of the NS-CMIP5, which uses a covariate obtained as an average of all raw (large grid-cell) CMIP5 data for each year, a modified nonstationary model (NS) that runs ensembles of statistically downscaled data instead of the climate models was developed. This model produced somewhat consistent ranges in relative increases for the runs, but the quasi-stationary approach produced an approximately two times larger percent increase in the future than those based on the NS approach. The percent increase for QS results is comparable with the results presented in the 2017 Fourth National Climate Assessment (NCA4) (Fig. 1.3), while the projected percent increases based on NS are much smaller.

5.1. Nonstationary and quasi-stationary approaches

We tested two approaches to study the effects of increasing heavy precipitation on precipitation frequency. One approach, referred to as the quasi-stationary (QS) approach, divides the period of record into two or more time periods and treats each of them as stationary. The other, nonstationary (NS) approach expresses statistics (e.g., frequency distribution parameters) as a function of covariates. This section presents the results of several comparisons between the QS and NS approaches. In addition to the observed data, two climate model-generated data sets

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

(LOCA and UWPD) were used. The results were based on both the annual maximum series (AMS) and the partial duration series (PDS).

A nonstationary method developed by Penn State researchers Ben Shaby, Alfonso Mejia, and Gregory Bopp was used in this project. The method is described in the report titled, “Development and Evaluation of a Statistical Model for Non-stationary Precipitation Frequency Estimates with NOAA Atlas 14.” HDSC shared the report and the code in R to the authors for possible modifications and comparative analyses with other approaches. Herein, this method is referred to as the Penn State method. For consistent comparisons in this section, the Penn State model was used for both QS and NS cases. The original nonstationary mode of the method uses different covariates and expresses frequency distribution parameters as a function of these covariates. The resulting frequency estimates change with time. The QS method adopted in this section is a variant of the NS Penn State method, in which the parameters for the covariates are set to zero during the optimization process, reducing the method to its stationary option. It assumes stationarity for a predefined length of time. In this study, this length was assumed to vary and included 10-, 30-, and 50-year periods. For example, for 50-year windows, the windows were shifted by 10 years, and possible consecutive “stationary” time periods would be 2011-2060, 2021-2070, and 2031-2080, representing their middle year, 2035, 2045, and 2055, respectively.

5.2. Summary of the nonstationary results for AMS and daily data

Basic information on the three nonstationary runs compared in this study, NS-CMIP5 (raw large grid-cell data), NS-LOCA, and NS-UWPD is presented in Table 5.1. All runs in this section used the AMS approach. Figs. 5.1-5.6 show the results of the NS approach for raw, non-downscaled CMIP5 data for 2-, 100-, and 1,000-year recurrence intervals. It used 1900-2005 data to calibrate nonstationary model parameters. The PF estimates are smaller than those of the other approaches. However, the same model with 1950-2005 data for calibration produced

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

questionable results, such as a decrease of the regional average of 100-year quantile of 1.5% in 2075 compared to 1975. Consequently, the NS-CMIP5 model was not recommended at this time. Future evaluation is recommended to explore its usability and sensitivity to calibration record length. Instead of raw large grid-cell data, we used downscaled CMIP5 data as covariates in the subsequent NS runs.

Figs. 5.7-5.12 show the NS AMS-based runs for LOCA, and Figs. 5.13-5.18 show the NS AMS-based runs for the UWPD. Regional averages are presented in Tables 5.2-5.3, showing that although LOCA produced somewhat larger frequency estimates for RCP4.5 and slightly smaller estimates for RCP8.5, the results show that LOCA and UWPD produced relatively similar results. In both cases, the precipitation shows a large degree of spatial variability, reflecting the variability in climate in the region, with coastal areas having larger frequency estimates compared with the inland areas.

Table 5.1. Nonstationary AMS runs

Test Name	Training Time Period	Covariate (MAP)
NS-CMIP5 (RAW)	1900-2015 (AMS is from observations)	ensemble mean of raw CMIP5
NS-LOCA	1950-2005 (AMS is from observations)	ensemble mean of LOCA CMIP5
NS-UWPD	1950-2005 (AMS is from observations)	ensemble mean of UWPD

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.2.1. The Penn State local likelihood nonstationary method (NS model) applied to raw CMIP5 data (NS-CMIP5)

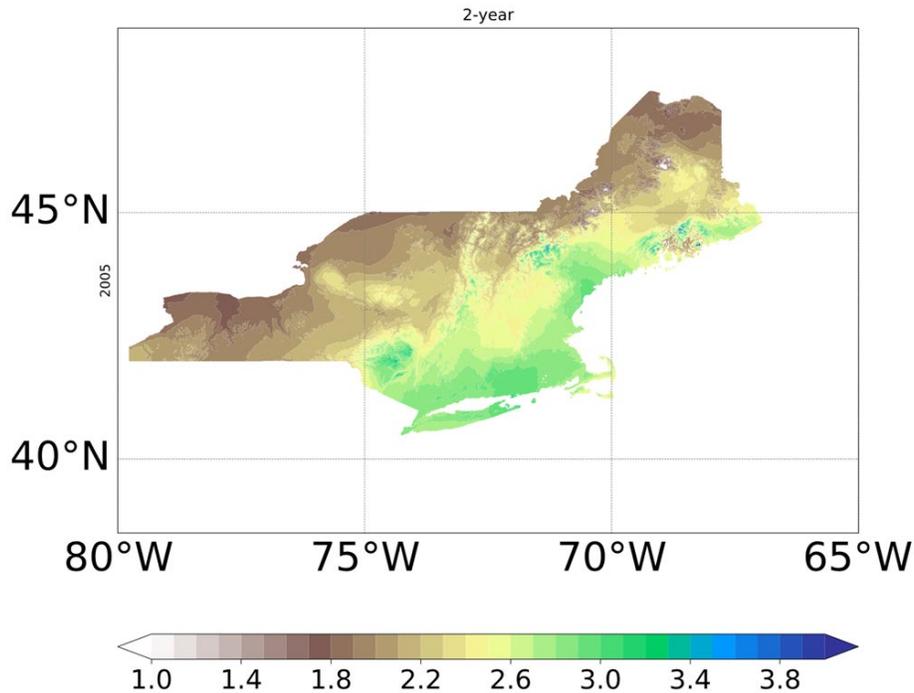


Figure 5.1. Isohyetal map showing 1-day 2-year precipitation in inches. The calculations were based on the 1900-2015 raw CMIP5 PF runs, representing the year 2005.

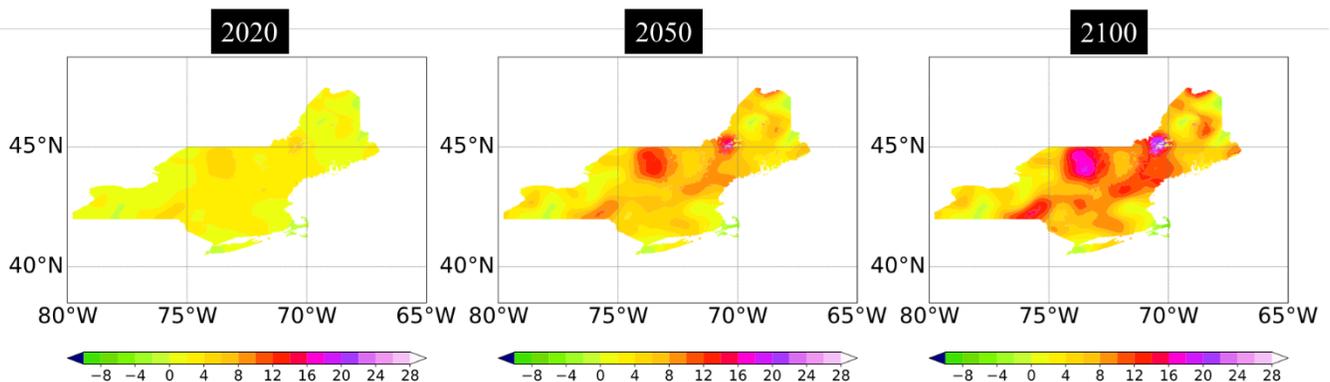


Figure 5.2. Map showing relative percent increases of 1-day 2-year precipitation for 2020, 2050, and 2100, compared to the baseline run, representing 2005, shown in Figure 5.1

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

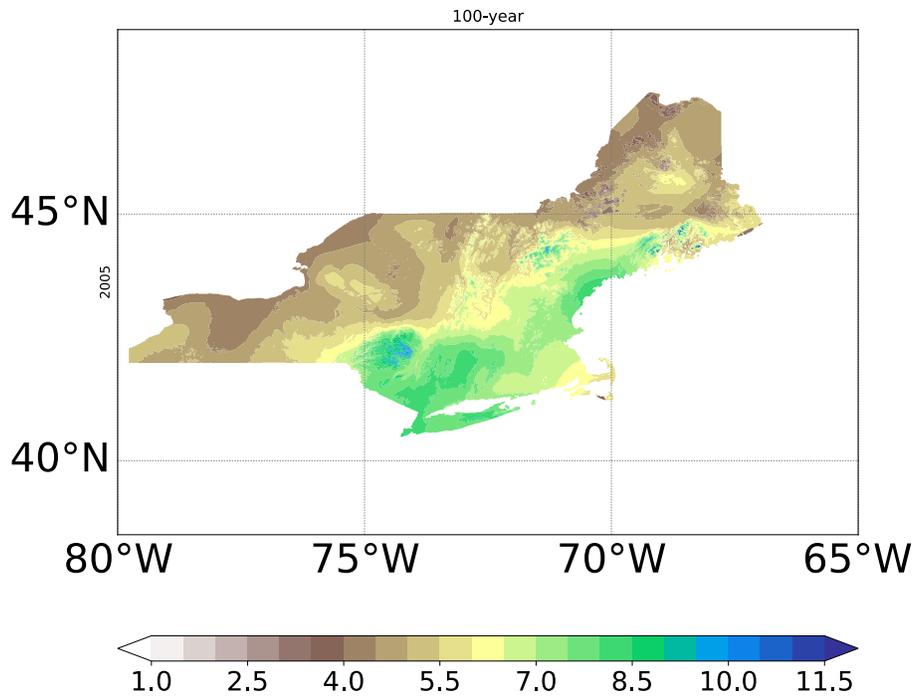


Figure 5.3. Isohyetal map showing 1-day 100-year precipitation in inches based on the 1900-2015 NS-CMIP5 PF runs, representing the year 2005

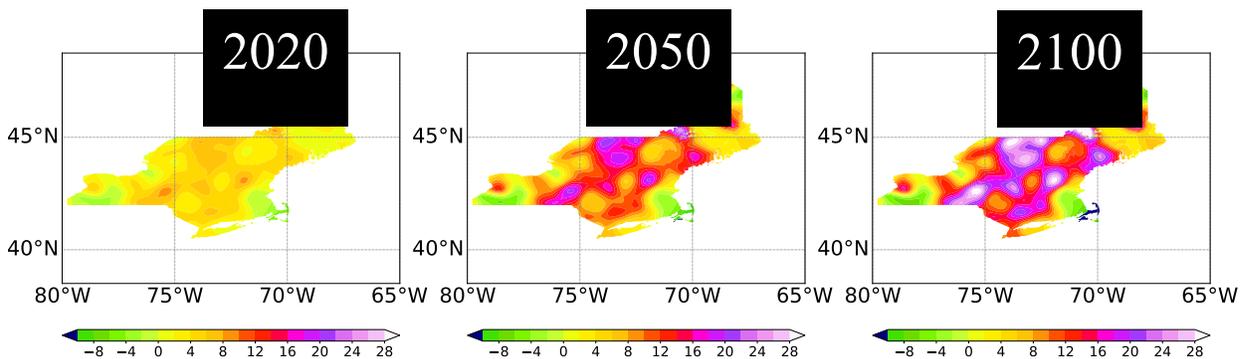


Figure 5.4. Map showing relative percent increases of 1-day 100-year precipitation for 2020, 2050, and 2100, compared to the baseline run, representing 2005, shown in Figure 5.3

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

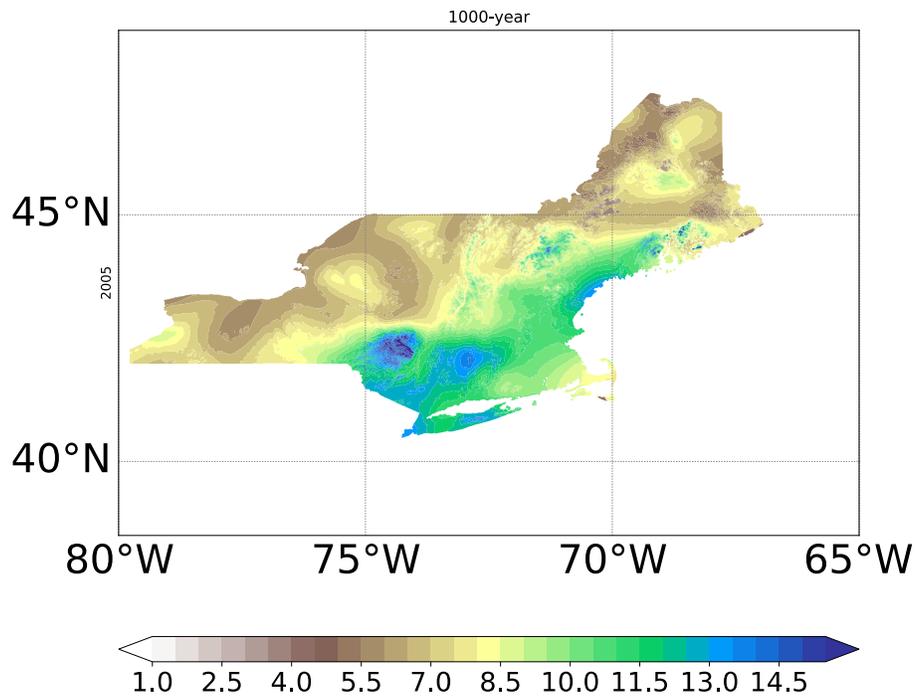


Figure 5.5. Isohyetal map showing 1-day 1000-year precipitation in inches based on the 1900-2015 NS-CMIP5 PF runs, representing the year 2005

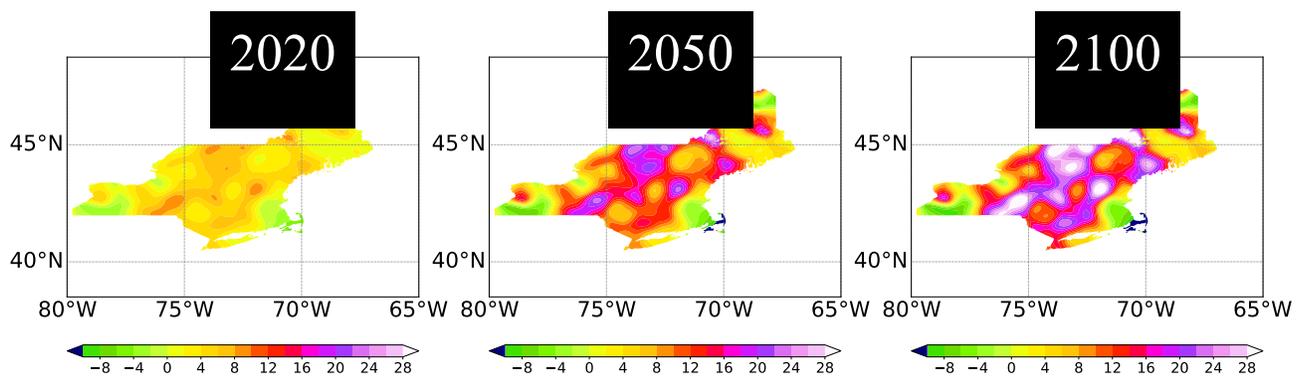


Figure 5.6. Map showing relative percent increases of 1-day 1000-year precipitation for 2020, 2050, and 2100 compared to the baseline run, representing 2005, shown in Figure 5.5

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.2.2. NS-LOCA model

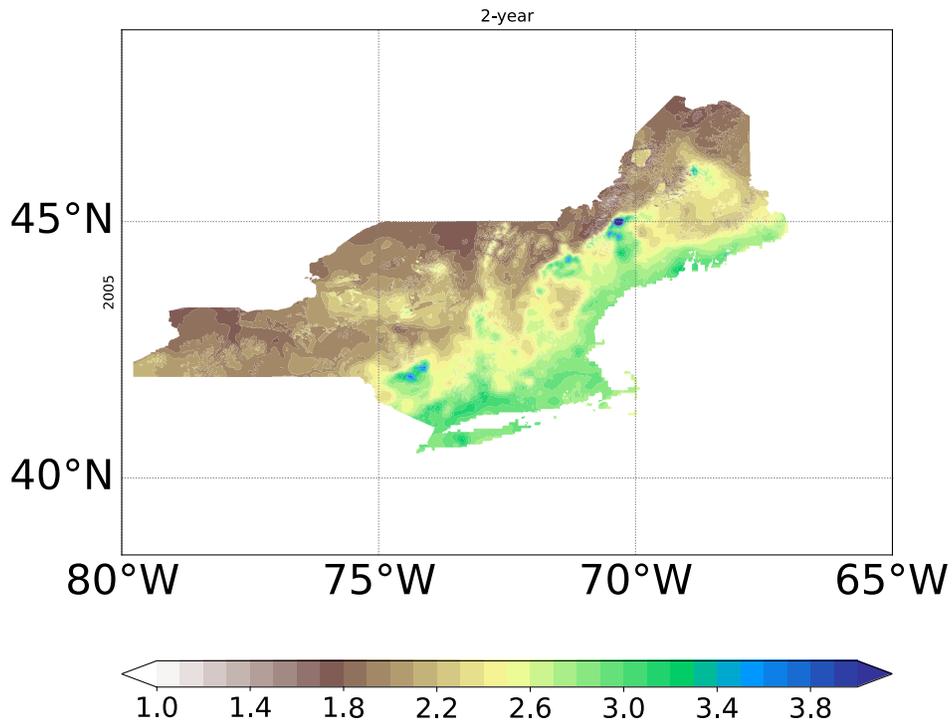


Figure 5.7. Isohyetal map showing 1-day 2-year precipitation in inches based on the 1950-2005 NS-LOCA PF runs, representing the year 2005

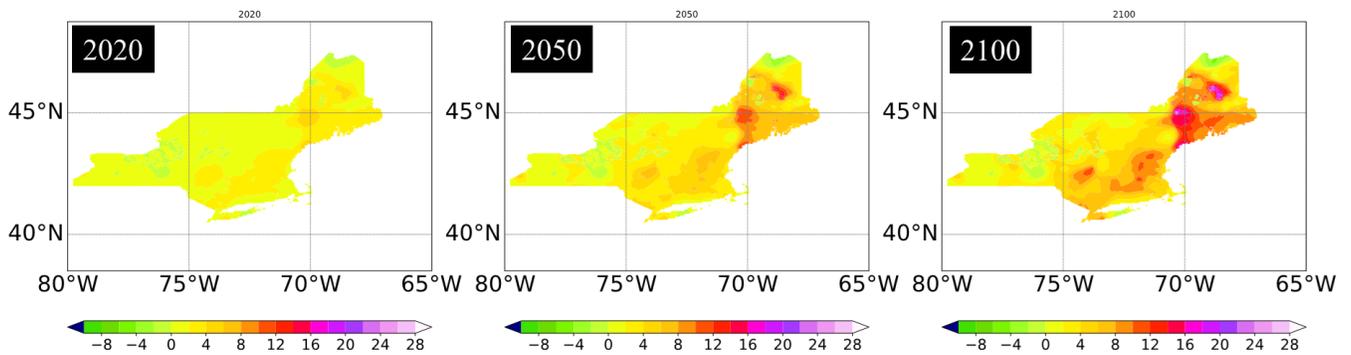


Figure 5.8. Map showing relative percent increases of 1-day 2-year precipitation for 2020, 2050, and 2100 compared to the baseline run, representing 2005, shown in Figure 5.7

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

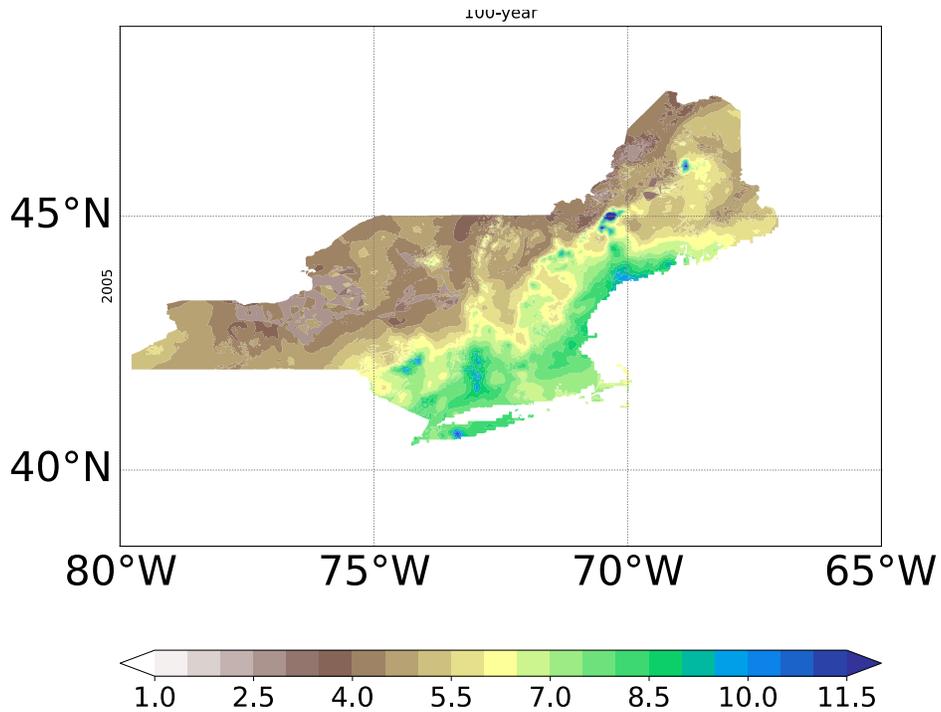


Figure 5.9. Isohyetal map showing 1-day 100-year precipitation in inches based on the 1950-2005 NS-LOCA PF runs, representing the year 2005

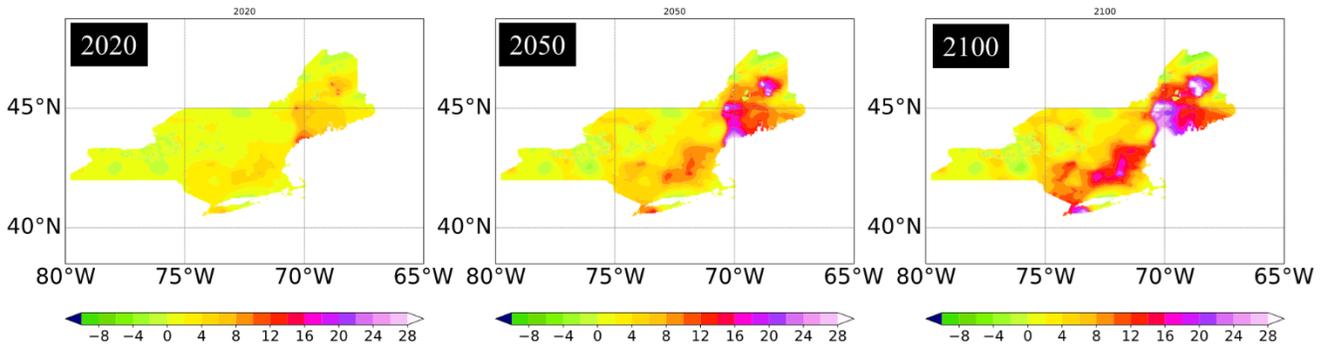


Figure 5.10. Map showing relative percent increases of 1-day 100-year precipitation for 2020, 2050, and 2100 compared to the baseline run, representing 2005, shown in Figure 5.9

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

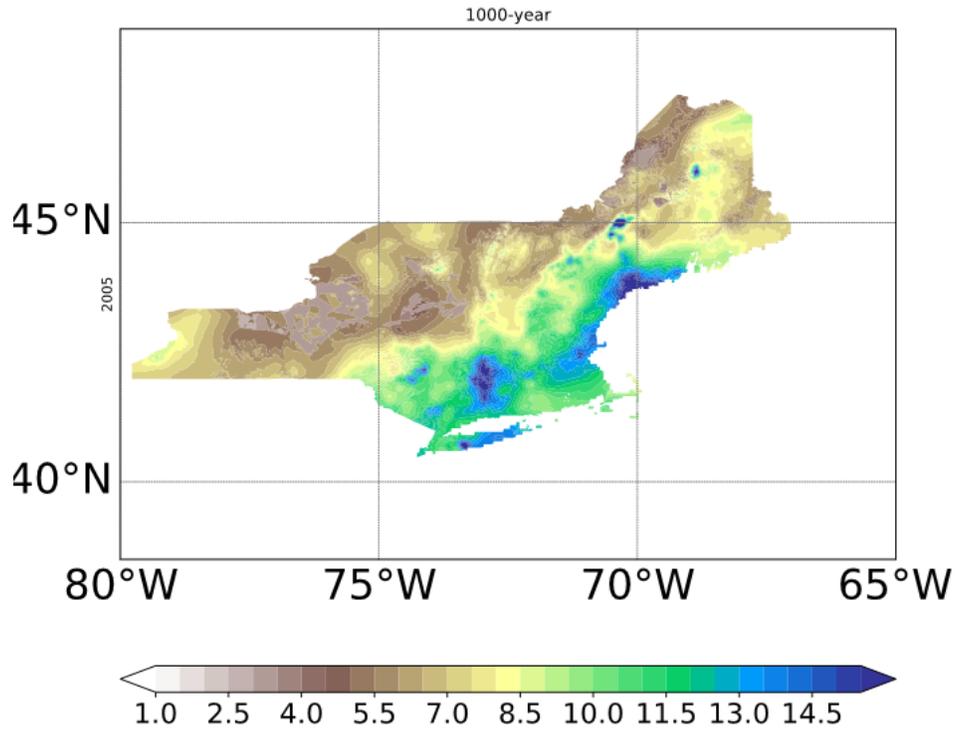


Figure 5.11. Isohyetal map showing 1-day 1000-year precipitation in inches based on the 1950-2005 NS-LOCA CMIP5 PF runs, representing 2005

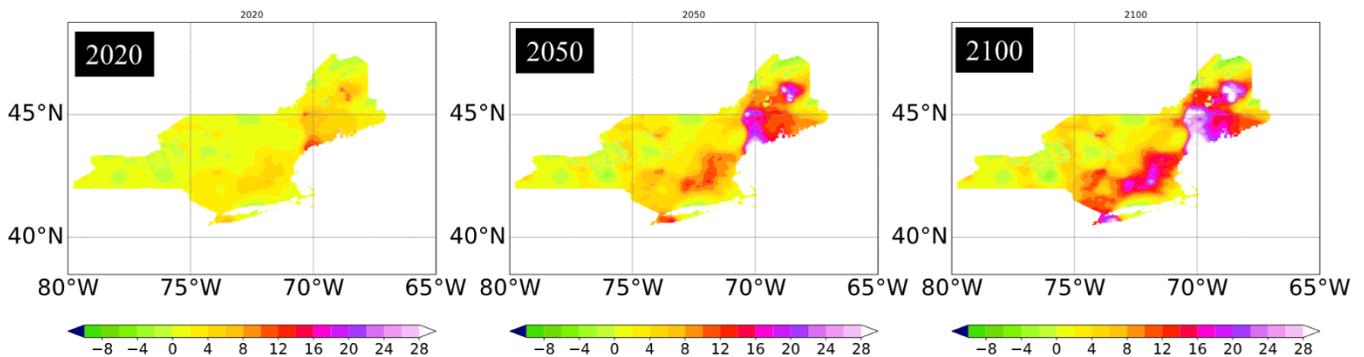


Figure 5.12. Map showing relative percent increases of 1-day 1000-year precipitation for 2020, 2050, and 2100, compared to the baseline run, representing 2005, shown in Figure 5.11

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.2. Projected changes for NS-LOCA averaged over the entire region

AMSRC45	record	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
	1900-2015	1961-2010	1971-2020	1981-2030	1991-2040	2001-2050	2011-2060	2021-2070	2031-2080	2041-2090	2051-2100
NS-LOCA	2-year	0.61	1.42	2.42	3.34	4.42	5.21	5.73	6.47	6.85	7.30
NS-LOCA	100-year	0.79	1.84	3.14	4.34	5.79	6.84	7.58	8.65	9.16	9.78
NS-LOCA	1000-year	0.83	1.92	3.29	4.55	6.07	7.18	7.97	9.10	9.64	10.30
AMSRC85	record	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
	1900-2015	1961-2010	1971-2020	1981-2030	1991-2040	2001-2050	2011-2060	2021-2070	2031-2080	2041-2090	2051-2100
NS-LOCA	2-year	0.59	1.20	2.39	3.29	4.62	5.85	6.91	7.84	9.06	10.10
NS-LOCA	100-year	0.75	1.56	3.11	4.30	6.05	7.77	9.23	10.58	12.33	13.94
NS-LOCA	1000-year	0.79	1.63	3.26	4.50	6.35	8.17	9.72	11.15	13.01	14.73

5.2.3. NS-UWPD model

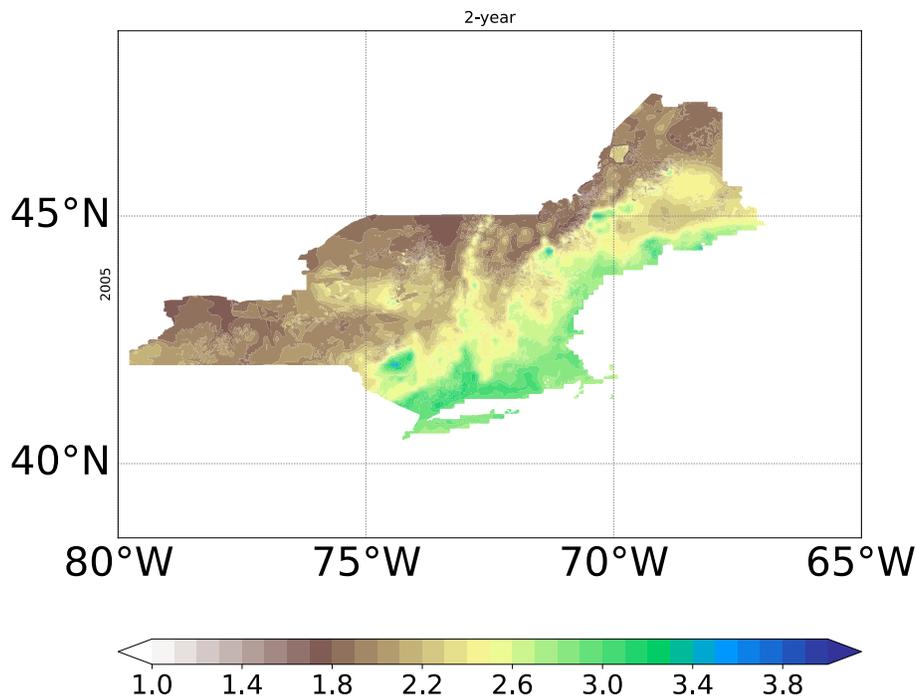


Figure 5.13. Isohyetal map showing 1-day 2-year precipitation in inches based on the 1950-2005 NS-UWPD PF runs, representing 1975

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

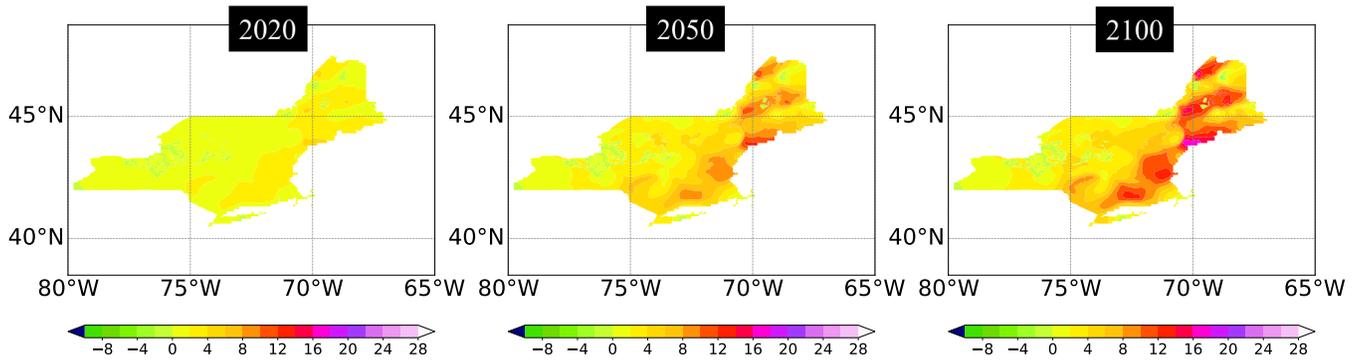


Figure 5.14. Map showing relative percent increases of 1-day 2-year precipitation for 2020, 2050, and 2100 compared to the baseline run, representing 1975, shown in Figure 5.13

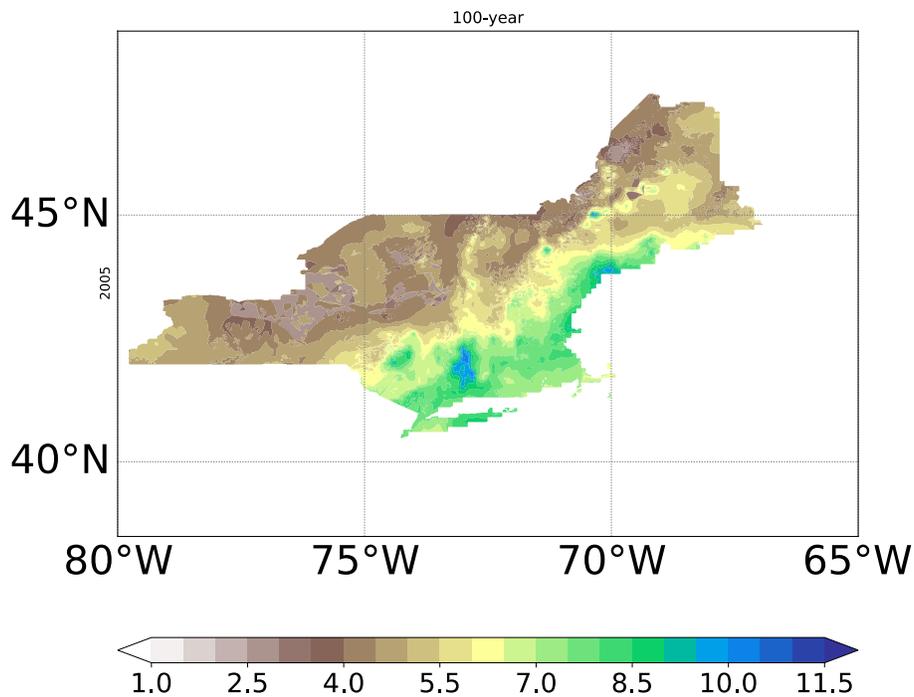


Figure 5.15. Isohyetal map showing 1-day 100-year precipitation in inches based on the 1950-2005 UWPD PF runs, representing 1975

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

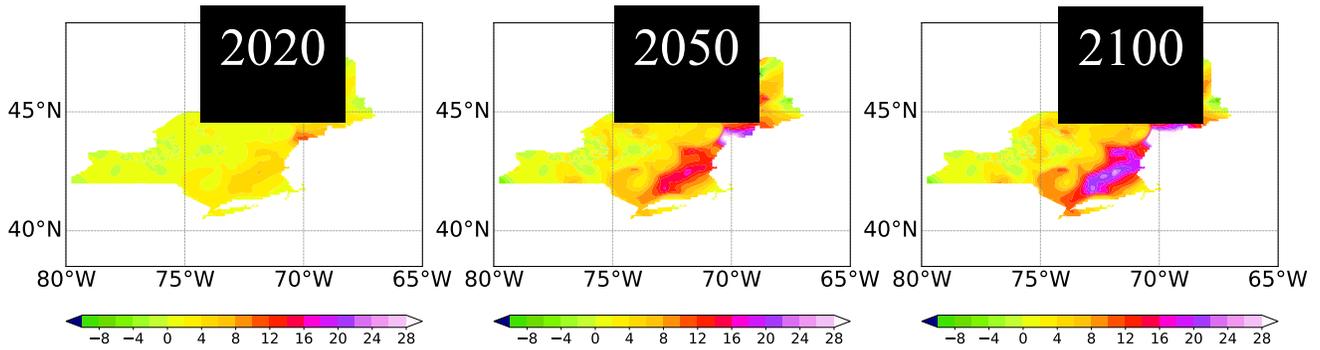


Figure 5.16. Map showing relative percent increases of 1-day 100-year precipitation for 2020, 2050, and 2100 compared to the baseline run, representing 1975, shown in Figure 5.15

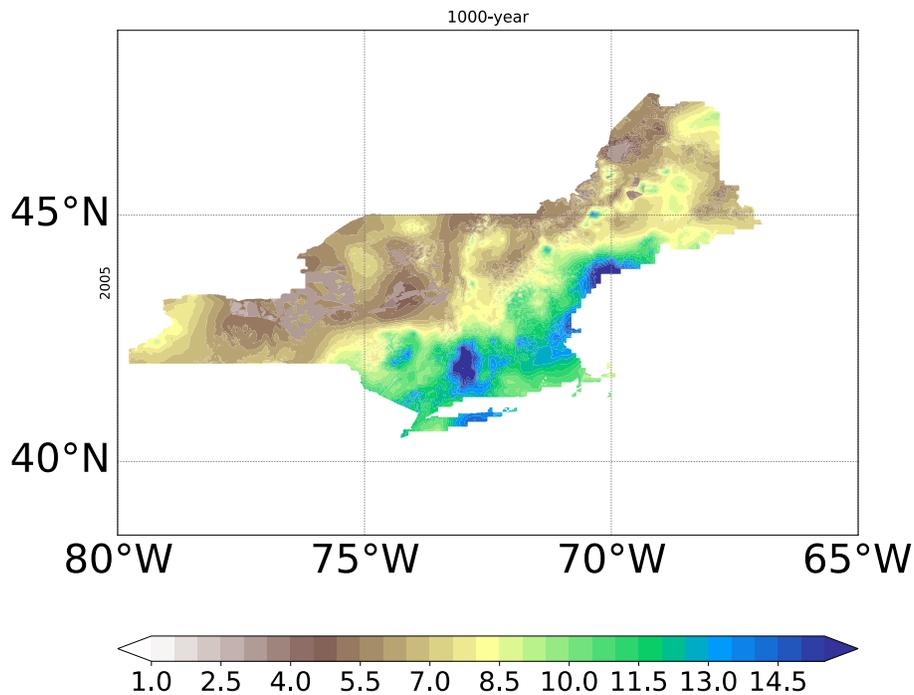


Figure 5.17. Isohyetal map showing 1-day 1000-year precipitation in inches based on the 1950-2005 NS-UWPD PF runs, representing 1975

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

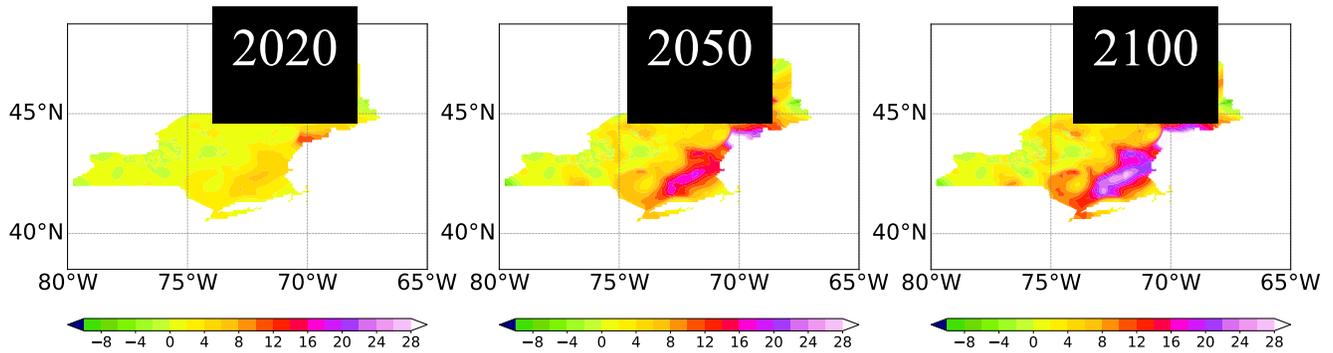


Figure 5.18. Map showing relative percent increases of 1-day 1000-year precipitation for 2020, 2050, and 2100 compared to the baseline run, representing 1975, shown in Figure 5.17

Table 5.3. Projected changes for NS-UWPD averaged over the entire region

AMS, RCP45	record	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
	1900-2015	1961-2010	1971-2020	1981-2030	1991-2040	2001-2050	2011-2060	2021-2070	2031-2080	2041-2090	2051-2100
NS-UWPD	2-year	0.51	1.18	2.15	3.15	4.15	5.08	5.68	6.32	6.73	7.23
NS-UWPD	100-year	0.65	1.50	2.72	4.00	5.28	6.51	7.32	8.20	8.77	9.46
NS-UWPD	1000-year	0.67	1.57	2.84	4.17	5.52	6.81	7.67	8.60	9.20	9.93
AMS, RCP85	record	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
	1900-2015	1961-2010	1971-2020	1981-2030	1991-2040	2001-2050	2011-2060	2021-2070	2031-2080	2041-2090	2051-2100
NS-UWPD	2-year	0.52	1.05	2.30	3.22	4.43	5.64	6.87	8.00	9.37	10.88
NS-UWPD	100-year	0.66	1.32	2.93	4.13	5.70	7.34	9.05	10.68	12.64	14.92
NS-UWPD	1000-year	0.69	1.38	3.07	4.32	5.97	7.70	9.51	11.24	13.33	15.77

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.3. Summary of quasi-stationary runs for AMS and daily data

Quasi-stationary runs based on the AMS using LOCA (QS-LOCA) and UWPD (QS-UWPD) revealed future trends in precipitation frequency estimates. For LOCA, Fig. 5.19 shows an isohyetal map for 1-day 2-, 100-, and 1000-year precipitation using the QS approach and representing the year 1975. Fig. 5.20 and Table 5.4 show the percent increase for RCP 4.5 and RCP 8.5 for 1-day 2-, 100-, and 1000-year precipitation with respect to the “base year” 1975, showing the 95% confidence limits. Similarly, for QS-UWPD, Fig. 5.21 shows an isohyetal map for 1-day 2-, 100-, and 1000-year precipitation, representing the year 1975. Fig. 5.22 and Table 5.5 show the percent increase for RCP 4.5 and RCP 8.5 for 1-day 2-, 100-, and 1000-year precipitation with respect to the “base year” 1975 and showing the 95% confidence limits. The results based on the LOCA data set were generally similar to those using the UWPD data set, but UWPD produced somewhat larger frequency estimates. All results for RCP8.5 increased with time, while RCP4.5 had some minor decreases as a result of decadal fluctuations. Scenario RCP8.5 produced similar (early in the calibration period) or larger estimates (later in the calibration period) than RCP4.5. The increasing difference between the two is particularly notable in the later decades of the 21st century.

5.3.1. LOCA-based runs

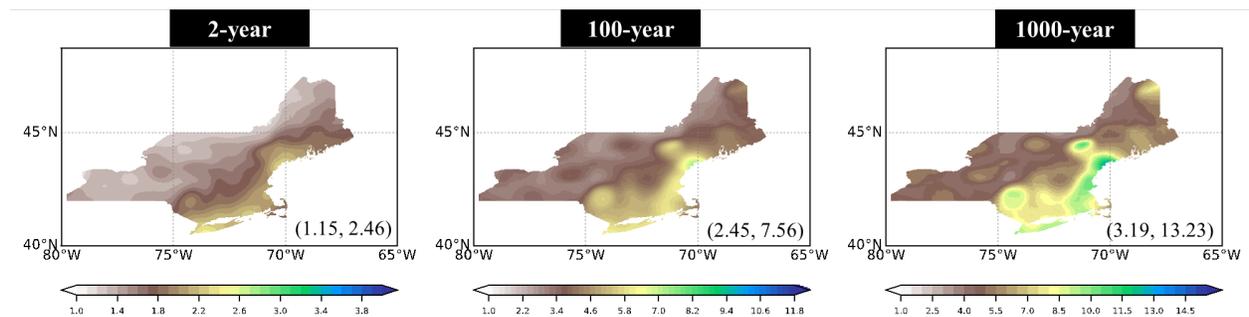


Figure 5.19. Isohyetal map showing 1-day 2-, 100-, and 1000-year precipitation in inches based on the 1950-2005 QS-LOCA runs, representing the year 1975

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

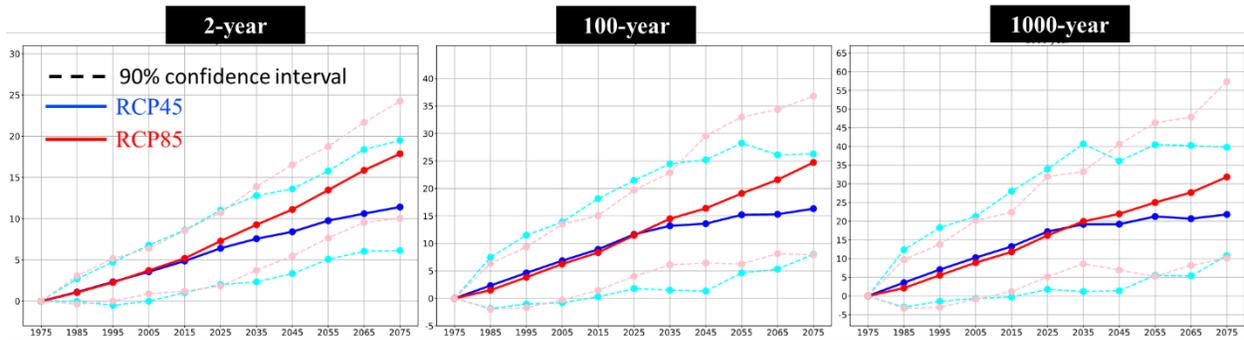


Figure 5.20. Percent increase for RCP 4.5 and RCP 8.5 for 1-day 2-, 100-, and 1000-year precipitation with respect to the 1950-2005 QS-LOCA runs, representing the year 1975. The figures also include 95% confidence limits.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.4. QS-LOCA runs: Numeric values of data shown in Figure 5.20

	1985 1961- 2010	1995 1971- 2020	2005 1981- 2030	2015 1991- 2040	2025 2001- 2050	2035 2011- 2060	2045 2021- 2070	2055 2031- 2080	2065 2041- 2090	2075 2051- 2100
RCP 4.5										
2-year	1.11 (-0.02, 2.71)	2.34 (-0.52, 4.74)	3.56 (-0.01, 6.79)	4.87 (1.04, 8.61)	6.42 (2.01, 11.03)	7.57 (2.33, 12.79)	8.41 (3.34, 13.61)	9.76 (5.08, 15.79)	10.62 (6.05, 18.39)	11.42 (6.14, 19.50)
100-year	2.34 (-1.84, 7.48)	4.64 (-1.03, 11.50)	6.85 (-0.78, 13.90)	8.93 (0.30, 18.15)	11.64 (1.79, 21.45)	13.20 (1.50, 24.48)	13.60 (1.33, 25.22)	15.21 (4.63, 28.25)	15.31 (5.31, 26.13)	16.32 (8.02, 26.27)
1000- year	3.56 (-2.97, 12.34)	7.05 (-1.42, 18.29)	10.26 (-0.66, 21.30)	13.24 (-0.30, 28.02)	17.22 (1.77, 33.95)	19.17 (1.20, 40.68)	19.21 (1.39, 36.13)	21.27 (5.51, 40.48)	20.69 (5.37, 40.27)	21.81 (1.08, 39.80)
RCP 8.5										
2-year	1.05 (-0.36, 3.06)	2.26 (-0.02, 5.17)	3.72 (0.86, 6.37)	5.19 (1.22, 8.53)	7.29 (1.84, 10.74)	9.27 (3.75, 13.89)	11.12 (5.48, 16.55)	13.48 (7.66, 18.77)	15.88 (9.55, 21.68)	17.88 (10.04, 24.26)
100-year	1.50 (-1.96, 3.21)	3.83 (-1.68, 5.28)	6.26 (-0.27, 6.64)	8.33 (1.48, 8.72)	11.47 (4.02, 10.92)	14.49 (6.12, 13.94)	16.39 (6.43, 16.70)	19.09 (6.26, 19.34)	21.58 (8.14, 21.92)	24.74 (7.92, 24.50)
1000- year	2.12 (-3.29, 9.75)	5.56 (-2.99, 13.80)	8.91 (-0.87, 20.14)	11.79 (1.19, 22.41)	16.20 (5.15, 31.97)	19.96 (8.64, 33.26)	21.97 (6.94, 40.60)	25.02 (5.25, 46.32)	21.68 (8.19, 47.84)	31.85 (10.12, 57.39)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.3.2. UWPD-based runs

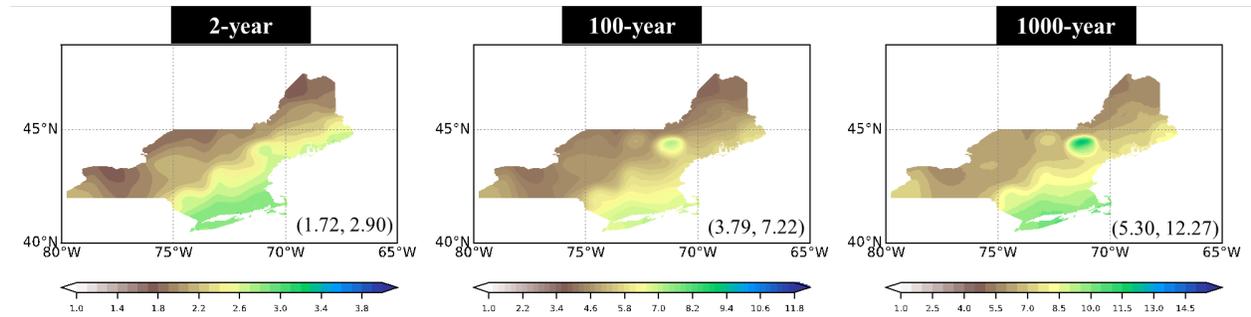


Figure 5.21. Isohyetal map showing 1-day 2-, 100-, and 1000-year precipitation in inches based on the 1950-2005 QS-UWPD runs, representing the year 1975

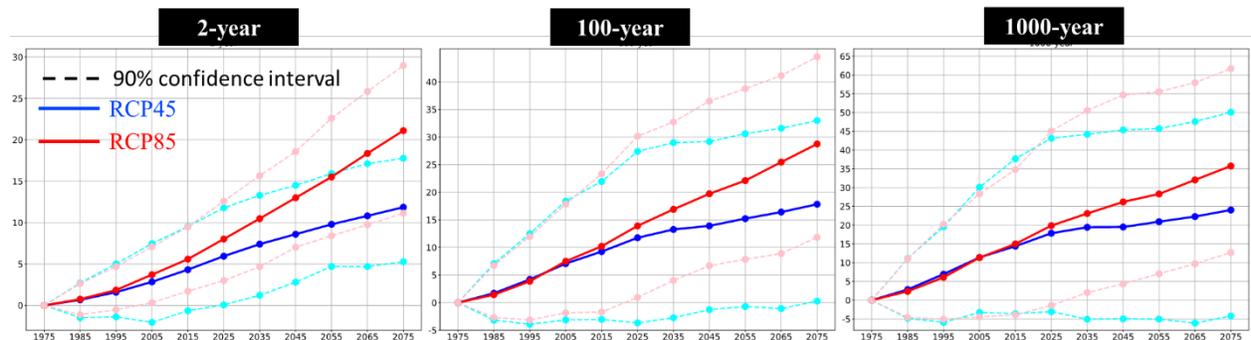


Figure 5.22. QS-UWPD runs: Percent increase for RCP4.5 and RCP8.5 for 1-day 2-, 100-, and 1000-year precipitation with respect to the 1950-2005 UWPD QS runs, representing the year 1975. The figures also include 95% confidence limits.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.5. QS-UWPD runs: Numeric values of data shown in Figure 5.22

	UWPD	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
	QS	1961- 2010	1971- 2020	1981- 2030	1991- 2040	2001- 2050	2011- 2060	2021- 2070	2031- 2080	2041- 2090	2051- 2100
RCP											
4.5		0.67	1.60	2.84	4.32	5.95	7.40	8.60	9.79	10.81	11.85
	2-year	(-1.49, 2.72)	(-1.38, 5.01)	(-2.04, 7.47)	(-0.66, 9.53)	(0.07, 11.79)	(1.22, 13.29)	(2.81, 14.50)	(4.70, 15.92)	(4.69, 17.12)	(5.27, 17.78)
	100-year	1.70 (-3.17, 7.06)	4.20 (-3.92, 12.48)	7.07 (-3.12, 18.40)	9.25 (-3.04, 21.94)	11.75 (-3.66, 27.40)	13.26 (-2.73, 28.98)	13.91 (-1.23, 29.19)	15.23 (-0.71, 30.61)	16.41 (-1.07, 31.60)	17.82 (0.27, 33.00)
	1000-year	2.85 (-4.83, 11.13)	6.94 (-5.90, 19.54)	11.39 (-3.26, 30.11)	14.39 (-3.58, 37.68)	17.85 (-3.05, 43.15)	19.43 (-5.07, 44.18)	19.49 (-4.95, 45.37)	20.92 (-5.05, 45.73)	22.28 (-6.10, 47.60)	24.04 (-4.20, 50.07)
RCP											
8.5		0.75	1.85	3.71	5.58	8.02	10.48	12.99	15.51	18.35	21.11
	2-year	(-1.13, 2.64)	(-0.51, 4.68)	(0.34, 7.08)	(1.71, 9.47)	(3.01, 12.59)	(4.68, 15.66)	(7.02, 18.57)	(8.41, 22.60)	(9.74, 25.84)	(11.14, 28.96)
	100-year	1.43 (-2.69, 6.74)	3.87 (-3.14, 11.96)	7.49 (-1.85, 17.86)	10.20 (-1.70, 23.31)	13.90 (0.94, 30.15)	16.90 (4.03, 32.73)	19.72 (6.70, 36.51)	22.11 (7.85, 38.77)	25.46 (8.89, 41.12)	28.76 (11.84, 44.54)
	1000-year	2.35 (-4.57, 11.05)	6.08 (-5.07, 20.23)	11.34 (-4.44, 28.33)	14.99 (-3.89, 34.79)	19.87 (-1.40, 45.10)	23.10 (2.04, 50.53)	26.19 (4.35, 54.67)	28.31 (7.06, 55.56)	32.05 (9.70, 57.96)	35.77 (12.70, 61.70)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.4. Summary of the nonstationary and quasi-stationary results for PDS and daily data

The test names and basic descriptions of comparisons for PDS are shown in Table 5.6. Nonstationary results are presented in this section first. For LOCA, Fig. 5.23 shows an isohyetal map for 1-day 2-, 100-, and 1000-year precipitation using the NS PDS approach, representing the year 1975. Fig. 5.24 and Table 5.7 show the percent increase for RCP4.5 and RCP8.5 for 1-day 2-, 100-, and 1000-year precipitation with respect to 1975. Similarly, for UWPD, Fig. 5.25 shows an isohyetal map for 1-day 2-, 100-, and 1000-year precipitation using the NS approach and representing the year 1975. Fig. 5.26 and Table 5.8 show the percent increase for RCP4.5 and RCP8.5 for 1-day 2-, 100-, and 1000-year precipitation with respect to the “base year” 1975.

Following the results for the NS approach, the results for the quasi-stationary method (QS) are presented for LOCA and UWPD, with the basic information shown in Table 5.9. Fig. 5.27 shows an isohyetal map for 1-day 2-, 100-, and 1000-year precipitation using LOCA data and the QS PDS approach and representing the year 1975. Fig. 5.28 and Table 5.10 show the percent increase for RCP4.5 and RCP8.5 for 1-day 2-, 100-, and 1000-year precipitation with respect to 1975. Corresponding results for UWPD were shown in Figs. 5.29-5.30 and Table 5.11.

Both the NS PDS and QS PDS methods showed more monotonic future increases compared with the NS AMS and QS AMS approaches, where decadal fluctuations were more visible. However, similar to the NS AMS vs. QS AMS cases, the NS PDS method also produced notably lower percent increases than the QS PDS. For example, the percent increase for 2075 based on the NS PDF for the 100-year recurrence interval using LOCA data and scenario RCP4.5 (Table 5.7) was 7.67%, while the corresponding result based on QS PDS (Table 5.10) was 14.29%. Other comparisons also indicate that QS estimates are approximately two times larger than those based on NS.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.4.1. Nonstationary results for PDS

Table 5.6. Nonstationary PDS runs

Time Period (used to estimate parameters)	Test Name	Covariate (MAP)
1950-2005 (PDS2 is from observation)	NS-LOCA	ensemble mean of LOCA
	NS-UWPD	ensemble mean of UWPD

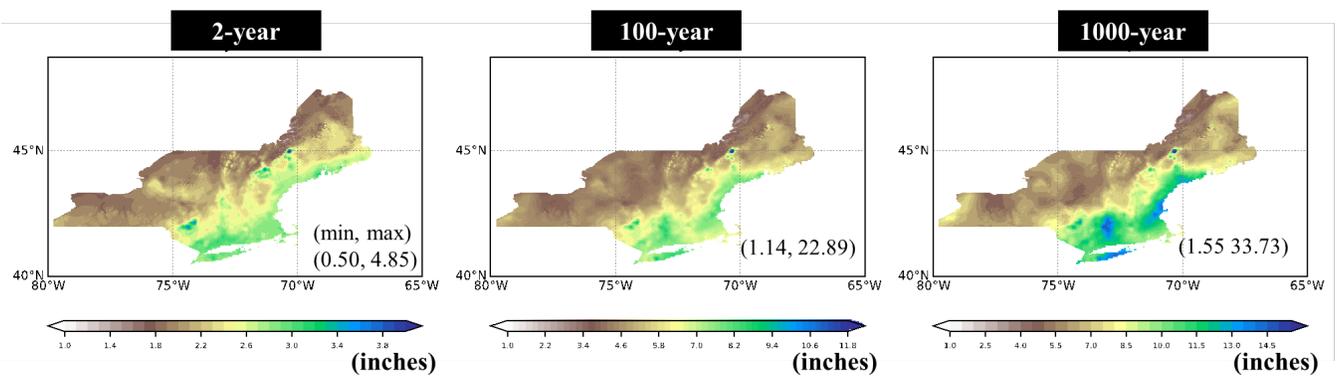


Figure 5.23. LOCA CMIP5 1-day PF for 1975 (representing average for 1951-2000)

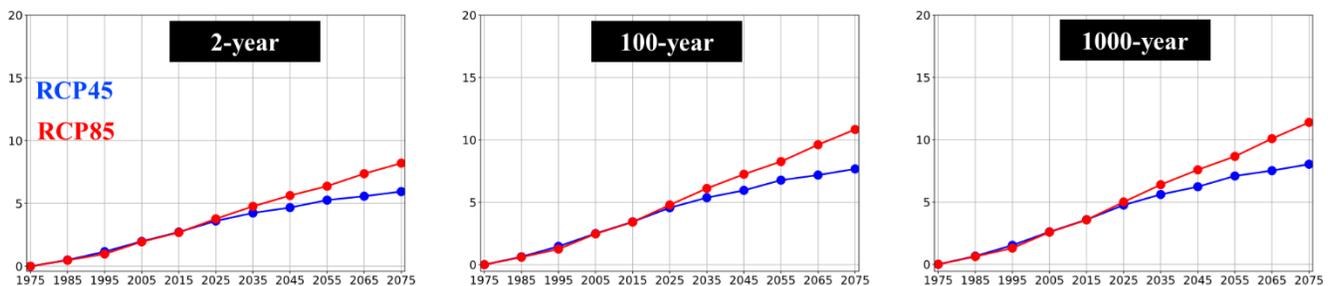


Figure 5.24. LOCA CMIP5 regional average of PF increase ratio (%) for 2-, 100-, and 1000-year events for RCP4.5 and RCP8.5

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

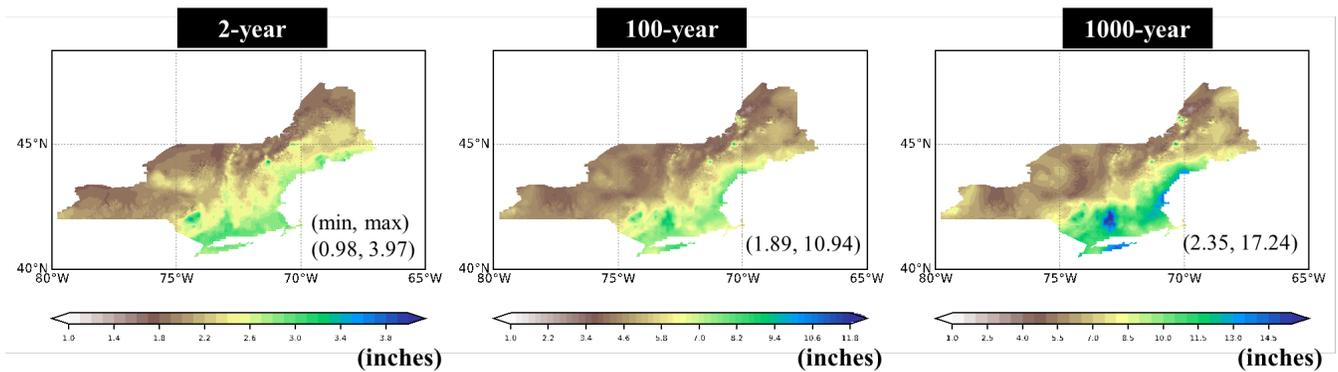


Figure 5.25. NS-UWPD 1-day PF for 1975 (representing the average for 1951-2000)

Table 5.7. NS-LOCA regional average of PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events. PDS2 1-day analysis was used for RCP4.5 and RCP8.5.

1950-2005 NS-LOCA 1-day PDS2	1985 (1961- 2010)	1995 (1971- 2020)	2005 (1981- 2030)	2015 (1991- 2040)	2025 (2001- 2050)	2035 (2011- 2060)	2045 (2021- 2070)	2055 (2031- 2080)	2065 (2041- 2090)	2075 (2051- 2100)
RCP 4.5	2-year	0.50	1.16	1.98	2.73	3.61	4.24	4.67	5.27	5.95
	100-year	0.63	1.47	2.50	3.43	4.56	5.37	5.95	6.77	7.67
	1000-year	0.66	1.53	2.6	3.58	4.76	5.61	6.23	7.09	8.04
RCP 8.5	2-year	0.48	0.98	1.95	2.69	3.78	4.77	5.63	6.38	7.37
	100-year	0.60	1.24	2.47	3.41	4.79	6.11	7.24	8.26	9.61
	1000-year	0.63	1.30	2.58	3.57	5.00	6.39	7.59	8.66	10.09

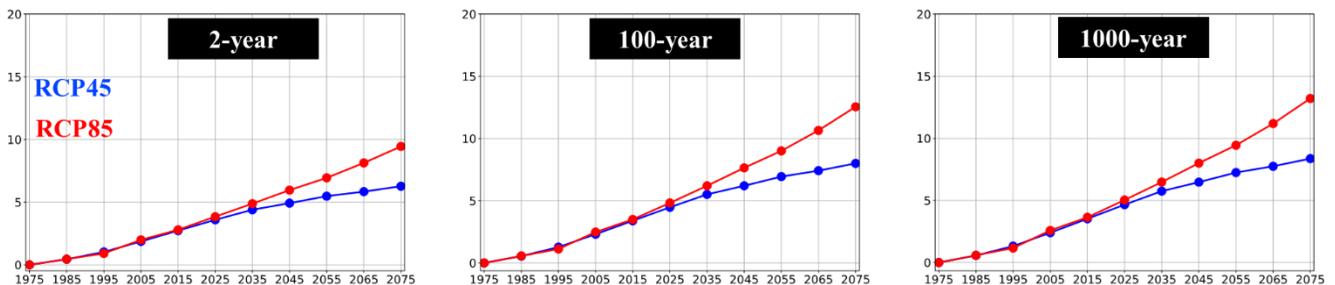


Figure 5.26. UWPD regional average of PF increase ratio (%) for 2-, 100-, and 1000-year events for RCP4.5 and RCP8.5

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.8. NS-UWPD regional average of PF increase ratio (%) relative to 1975 (average of 1950-2005) with 90% confidence limits, for 2-, 100-, and 1000-year events for RCP4.5 and RCP8.5

1950-2005 1-day PDS2		1985 (1961- 2010)	1995 (1971- 2020)	2005 (1981- 2030)	2015 (1991- 2040)	2025 (2001- 2050)	2035 (2011- 2060)	2045 (2021- 2070)	2055 (2031- 2080)	2065 (2041- 2090)	2075 (2051- 2100)
RCP 4.5	2-year	0.44	1.03	1.86	2.73	3.59	4.39	4.92	5.48	5.84	6.27
	100- year	0.55	1.27	2.31	3.38	4.47	5.51	6.20	6.94	7.42	8.00
	1000- year	0.57	1.32	2.40	3.52	4.66	5.75	6.47	7.25	7.75	8.37
RCP 8.5	2-year	0.45	0.91	1.99	2.79	3.84	4.88	5.96	6.94	8.13	9.45
	100- year	0.56	1.12	2.48	3.49	4.82	6.20	7.64	9.01	10.65	12.55
	1000- year	0.58	1.16	2.59	3.64	5.03	6.48	8.00	9.45	11.19	13.21

5.4.2. Quasi-stationary results for PDS

Table 5.9. Quasi-stationary PDS runs

Test Name	Ensemble Size	Target Years
QS-LOCA	32	1975-2075 every 10 years
QS-UWPD	308 for RCP 4.5 336 for RCP 8.5	1975, 2005, 2045, 2075

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

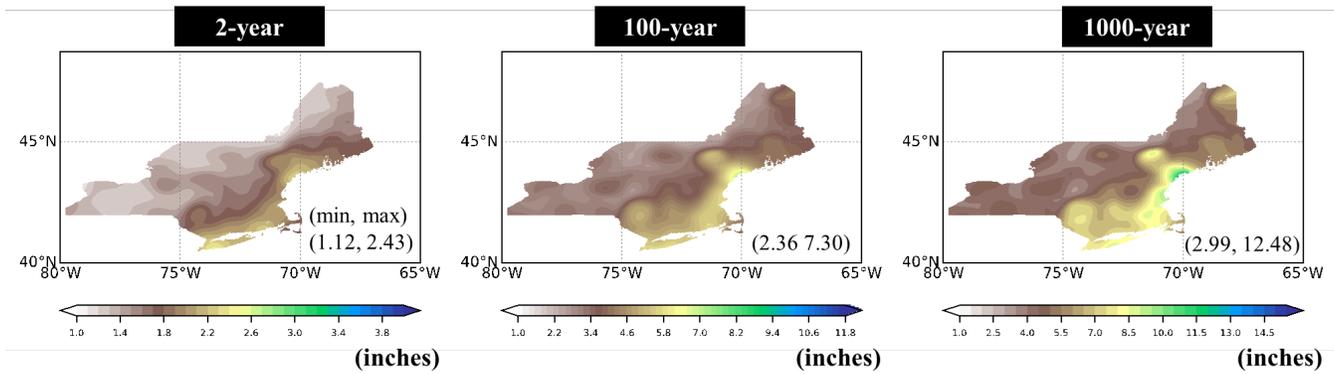


Figure 5.27. QS LOCA 1-day PF for 1975 (representing average for 1951-2000)

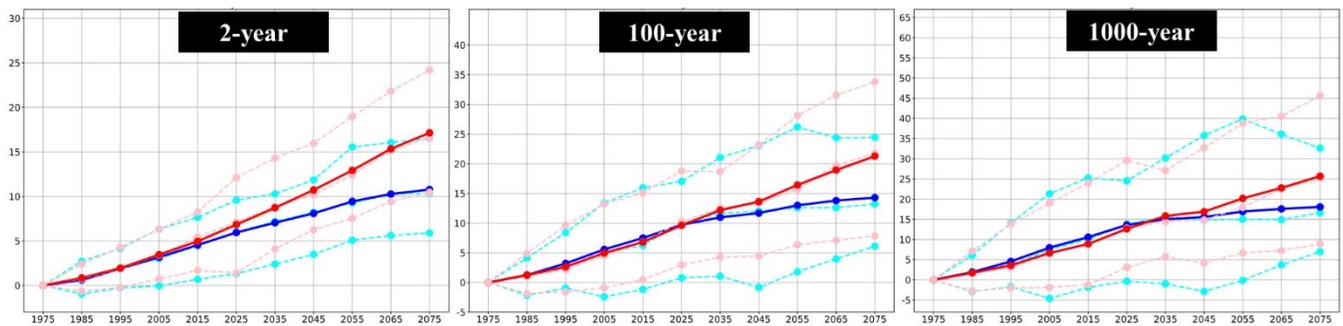


Figure 5.28. QS-LOCA regional average of PF increase ratio (%) for 2-, 100-, and 1000-year events for RCP 4.5 (blue) and RCP 8.5 (red) with median and confidence limits (dashed lines)

Table 5.10. QS-LOCA regional average of PF increase ratio (%) relative to 1975 (average of 1950-2005) with 90% confidence limits for 2-, 100-, and 1000-year events for RCP4.5 and RCP8.5 scenarios

QS-LOCA 1-day PDS2	1985 (1961-2010)	1995 (1971-2020)	2005 (1981-2030)	2015 (1991-2040)	2025 (2001-2050)	2035 (2011-2060)	2045 (2021-2070)	2055 (2031-2080)	2065 (2041-2090)	2075 (2051-2100)	
RCP 4.5	2-year	0.62	1.90	3.17	4.54	5.94	7.04	8.09	9.44	10.27	10.77
		(-0.96, 2.71)	(-0.27, 4.18)	(-0.05, 6.32)	(0.69, 7.65)	(1.32, 9.58)	(2.40, 10.29)	(3.50, 11.84)	(5.08, 15.56)	(5.61, 16.03)	(5.90, 16.54)
	100-year	1.25	3.20	5.57	7.50	9.72	10.97	11.69	13.00	13.79	14.29
	(-2.15, 4.16)	(-0.95, 8.40)	(-2.41, 13.39)	(-1.16, 15.96)	(0.78, 17.03)	(1.06, 21.05)	(-0.84, 23.06)	(1.81, 26.19)	(3.98, 24.38)	(6.11, 24.45)	
	1000-year	1.91	4.58	7.98	10.59	13.60	15.01	15.57	16.93	17.59	18.07
	(-2.89, 6.13)	(-1.76, 14.08)	(-4.61, 21.33)	(-1.86, 25.26)	(-0.33, 24.53)	(-0.96, 30.20)	(-2.86, 35.78)	(-0.11, 39.85)	(3.72, 36.08)	(6.98, 32.62)	
RCP 8.5	2-year	0.84	1.95	3.46	4.96	6.83	8.72	10.73	12.93	15.34	17.14
		(-0.64, 2.38)	(-0.23, 4.29)	(0.74, 6.31)	(1.68, 8.24)	(1.41, 12.10)	(4.11, 14.31)	(6.26, 15.97)	(7.55, 18.97)	(9.43, 21.83)	(10.47, 24.20)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

100-year	1.24	2.64	4.97	6.80	9.63	12.18	13.62	16.44	18.93	21.28
	(-1.86, 4.88)	(-1.57, 9.63)	(-0.88, 13.23)	(0.46, 15.12)	(2.97, 18.78)	(4.27, 18.67)	(4.43, 23.17)	(6.36, 28.13)	(7.10, 31.56)	(7.84, 33.82)
1000-year	1.76	3.58	6.66	8.89	12.64	15.83	16.90	20.18	22.79	25.69
	(-2.75, 7.14)	(-2.01, 13.84)	(-1.83, 19.05)	(-1.21, 23.95)	(3.12, 29.64)	(5.79, 27.12)	(4.15, 32.75)	(6.69, 38.80)	(7.28, 40.54)	(8.89, 45.62)

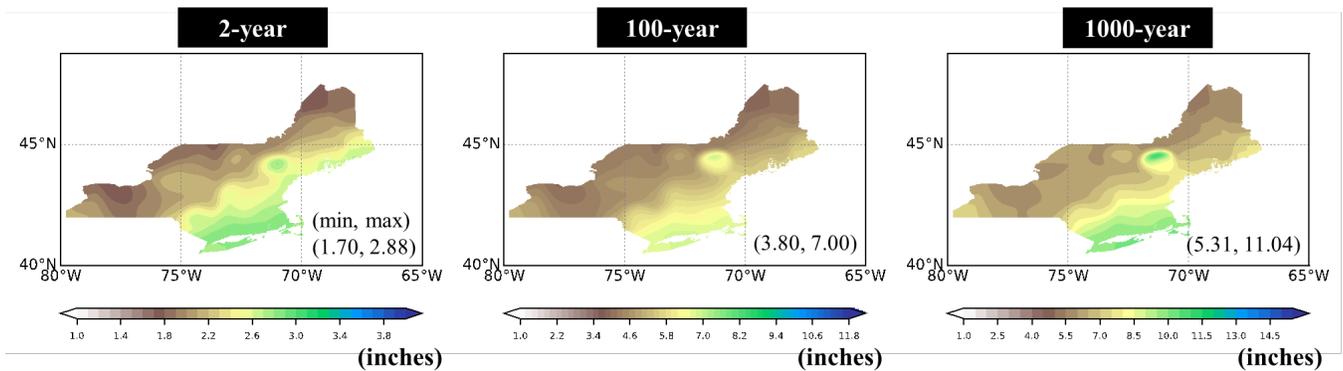


Figure 5.29. UWPD QS 1-day PF for 1975 (representing average for 1951-2000)

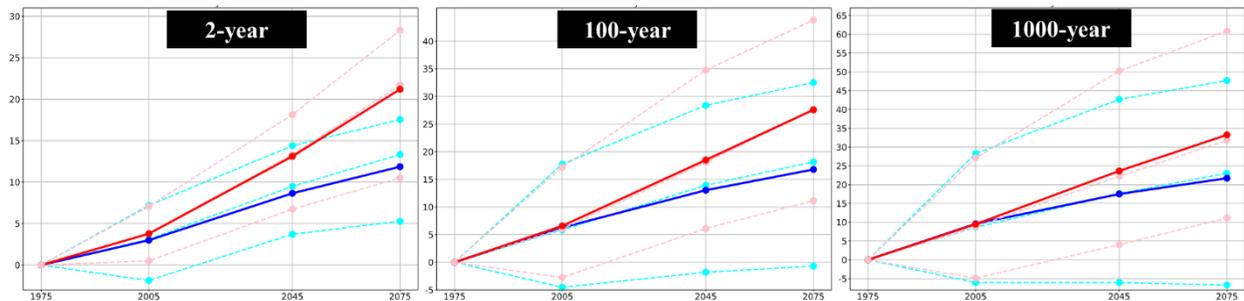


Figure 5.30. UWPD QS regional average of PF increase ratio (%) for 2-, 100-, and 1000-year events for RCP4.5 (blue) and RCP8.5 (red) with median and confidence limits (dashed lines)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.11. QS-UWPD regional average of PF increase ratio (%) relative to 1975 (average of 1950-2005) for 2-, 100-, and 1000-year events. PDS2 1-day results are shown for RCP4.5 and RCP8.5

QS-UWPD		2005	2045	2075
1-day PDS2		(1981-2030)	(2021-2070)	(2051-2100)
RCP 4.5	2-year	2.99	8.65	11.86
		(-1.86, 7.19)	(3.71, 14.38)	(5.28, 17.55)
	100-year	6.27	13.03	16.79
		(-4.52, 17.76)	(-1.80, 28.38)	(-0.69, 32.50)
	1000-year	9.58	17.51	21.72
		(-6.05, 28.29)	(-6.05, 42.68)	(-6.71, 47.69)
RCP 8.5	2-year	3.78	13.11	21.20
		(0.50, 7.05)	(6.74, 18.14)	(10.51, 28.30)
	100-year	6.57	18.51	27.59
		(-2.76, 17.19)	(6.10, 34.77)	(11.14, 43.82)
	1000-year	9.47	23.65	33.25
		(-4.88, 27.12)	(4.06, 50.25)	(11.13, 60.85)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.5. Comparison of all methods for daily data

5.5.1. Nonstationary and quasi-stationary comparisons for 100-year return period: AMS vs. PDS, LOCA vs. UWPD, RCP4.5 vs. RCP8.5

A comparative summary shows some of the most relevant runs. The nonstationary and quasi-stationary results are calculated for all target years shown in Table 5.12. Some of the most relevant runs are presented for climate scenarios RCP4.5 and RCP8.5 (Table 5.13 and Fig. 5.31). For simplicity, Table 5.13 presents results for only three selected years, 2005, 2045, and 2075. All results are illustrated for the 100-year return period.

Table 5.12. Target years for comparisons between NS and QS methods

Target Year	Time Period
1975	1951–2000
1985	1961–2010
1995	1971–2020
2005	1981–2030
2015	1991–2040
2025	2001–2050
2035	2011–2060
2045	2021–2070
2055	2031–2080
2065	2041–2090
2075	2051–2100

Fig. 5.31 shows graphs of the results for all years in Tables 5.12-5.13. The following conclusions can be drawn. The PF increasing ratio estimated by the quasi-stationary (blue, red) approach are systematically larger than the non-stationary approaches (cyan, pink). The PF increasing ratio estimated by the GEV fit (solid line/AMS approach) is slightly larger than the GPD fit (dashed line/PDS approach). F increasing ratio estimated using the UWPD data set is larger than when using LOCA, particularly for the RCP8.5 scenario.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.13. Average modeling results for 100-year return period event for scenarios RCP4.5 and RCP8.5. The results are expressed as a percent (%) increase with respect to 1975.

		2005	2045	2075
RCP4.5	PDS NS-LOCA	2.50	5.95	7.67
	PDS NS-UWPD	2.31	6.20	8.00
	PDS QS-LOCA	5.57	11.69	14.29
	PDS QS-UWPD	6.27	13.03	16.79
	AMS NS-LOCA	3.14	7.58	9.78
	AMS NS-UWPD	2.72	7.32	9.46
	AMS QS-LOCA	6.85	13.60	16.32
	AMS QS-UWPD	7.07	13.91	17.82
	RCP8.5	PDS NS-LOCA	2.47	7.25
PDS NS-UWPD		2.48	7.64	12.55
PDS QS-LOCA		4.97	13.62	21.28
PDS QS-UWPD		6.57	18.51	27.59
AMS NS-LOCA		3.11	9.23	13.94
AMS NS-UWPD		2.93	9.05	14.92
AMS QS-LOCA		6.26	16.39	24.74
AMS QS-UWPD		7.49	19.72	28.76

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

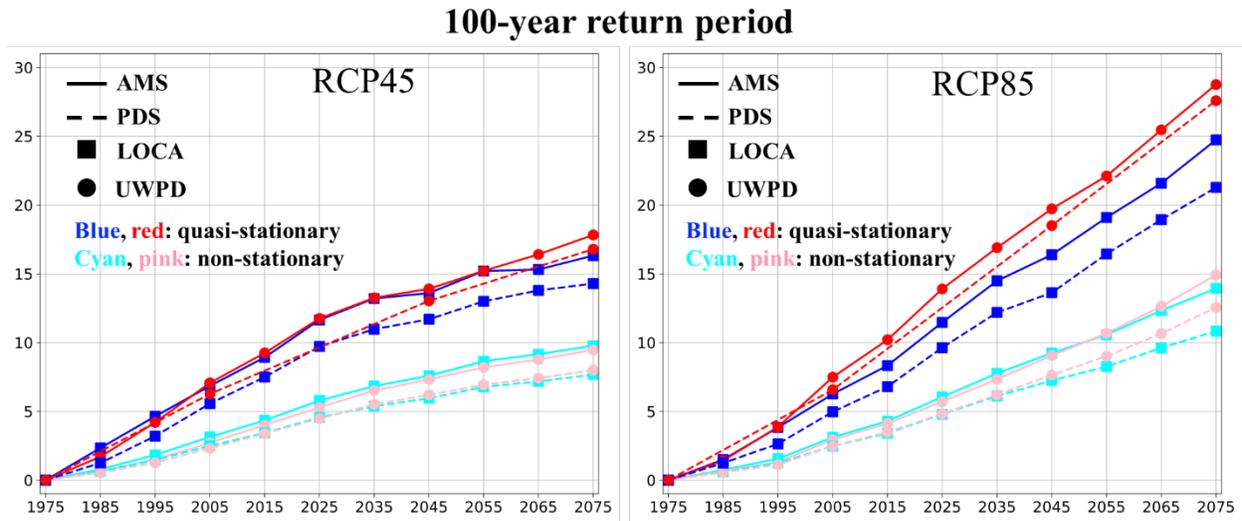


Figure 5.31. Graphs of the results for all years in Tables 5.12-5.13

5.5.2. Quasi-stationary comparisons for AMS-based method: 2-, 100-, and 1,000-year, LOCA vs. UWPD, RCP4.5 vs. RCP8.5, including confidence limits

The previous subsection compared the projected increases in a 100-year return period event for scenarios RCP4.5 and RCP8.5 and NS and QS methods. The purpose of this subsection is to test if and to what degree the increases change with the return period for the QS and AMS-based method for 2-, 100-, and 1,000-year return periods for LOCA and UWPD data and for RCP4.5 and RCP8.5 climate scenarios. Confidence limits are also shown.

Fig. 5.32 and Table 5.14 show the results for the LOCA data set. Fig. 5.33 and Table 5.15 show the same results for the UWPD data set. For both data sets, all three return periods, and both scenarios, the increase is higher with time, reaching the maximum at the end of the period, i.e., 1975. The percent increase for the 100-year period is the largest, followed by the 100-year and 2-year return periods. The LOCA and UWPD are comparable, but the UWPD produces to some degree larger projected values. For example, for the year 2075 using scenario RCP8.5 and a 1000-year event, LOCA and UWPD data show increases of 31.85% and 35.77%, respectively. Confidence intervals are wider for more distant time horizons and larger return periods.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Area averaged LOCAqs 1-day AMS-based PF ratio (%) RCP 4.5 and RCP 8.5

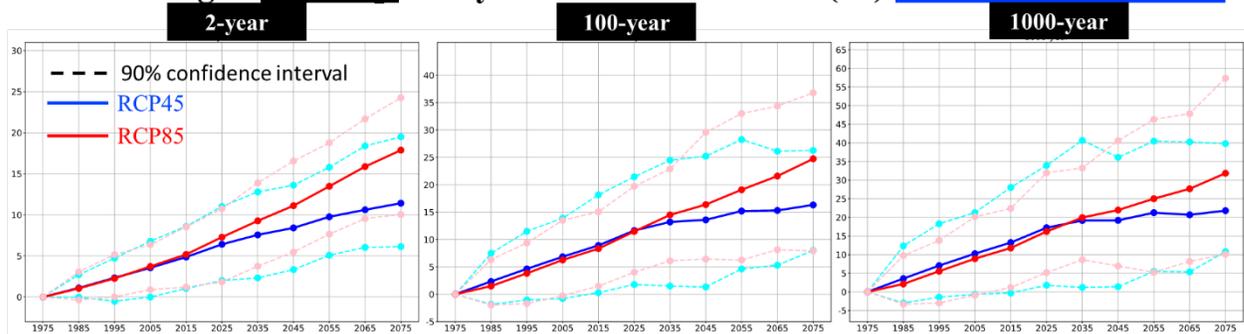


Figure 5.32. Quasi-stationary model results for AMS-based method; 2-, 100-, and 1,000-year recurrence intervals using LOCA data set, and showing RCP4.5 and RCP8.5 scenarios with confidence limits based on the distribution of ensemble results

Table 5.14. Numeric values of the results shown in Figure 5.32

QS-LOCA	1985 1961-2010	1995 1971-2020	2005 1981-2030	2015 1991-2040	2025 2001-2050	2035 2011-2060	2045 2021-2070	2055 2031-2080	2065 2041-2090	2075 2051-2100	
RCP 4.5	2-year	1.11	2.34	3.56	4.87	6.42	7.57	8.41	9.76	10.62	11.42
		(-0.02, 2.71)	(-0.52, 4.74)	(-0.01, 6.79)	(1.04, 8.61)	(2.01, 11.03)	(2.33, 12.79)	(3.34, 13.61)	(5.08, 15.79)	(6.05, 18.39)	(6.14, 19.50)
	100-year	2.34	4.64	6.85	8.93	11.64	13.20	13.60	15.21	15.31	16.32
	(-1.84, 7.48)	(-1.03, 11.50)	(-0.78, 13.90)	(0.30, 18.15)	(1.79, 21.45)	(1.50, 24.48)	(1.33, 25.22)	(4.63, 28.25)	(5.31, 26.13)	(8.02, 26.27)	
	1000-year	3.56	7.05	10.26	13.24	17.22	19.17	19.21	21.27	20.69	21.81
	(-2.97, 12.34)	(-1.42, 18.29)	(-0.66, 21.30)	(-0.30, 28.02)	(1.77, 33.95)	(1.20, 40.68)	(1.39, 36.13)	(5.51, 40.48)	(5.37, 40.27)	(1.08, 39.80)	
RCP 8.5	2-year	1.05	2.26	3.72	5.19	7.29	9.27	11.12	13.48	15.88	17.88
		(-0.36, 3.06)	(-0.02, 5.17)	(0.86, 6.37)	(1.22, 8.53)	(1.84, 10.74)	(3.75, 13.89)	(5.48, 16.55)	(7.66, 18.77)	(9.55, 21.68)	(10.04, 24.26)
	100-year	1.50	3.83	6.26	8.33	11.47	14.49	16.39	19.09	21.58	24.74
	(-1.96, 3.21)	(-1.68, 5.28)	(-0.27, 6.64)	(1.48, 8.72)	(4.02, 10.92)	(6.12, 13.94)	(6.43, 16.70)	(6.26, 19.34)	(8.14, 21.92)	(7.92, 24.50)	
	1000-year	2.12	5.56	8.91	11.79	16.20	19.96	21.97	25.02	21.68	31.85
	(-3.29, 9.75)	(-2.99, 13.80)	(-0.87, 20.14)	(1.19, 22.41)	(5.15, 31.97)	(8.64, 33.26)	(6.94, 40.60)	(5.25, 46.32)	(8.19, 47.84)	(10.12, 57.39)	

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Area averaged UWPDqs 1-day AMS-based PF ratio (%) RCP45 and RCP 8.5

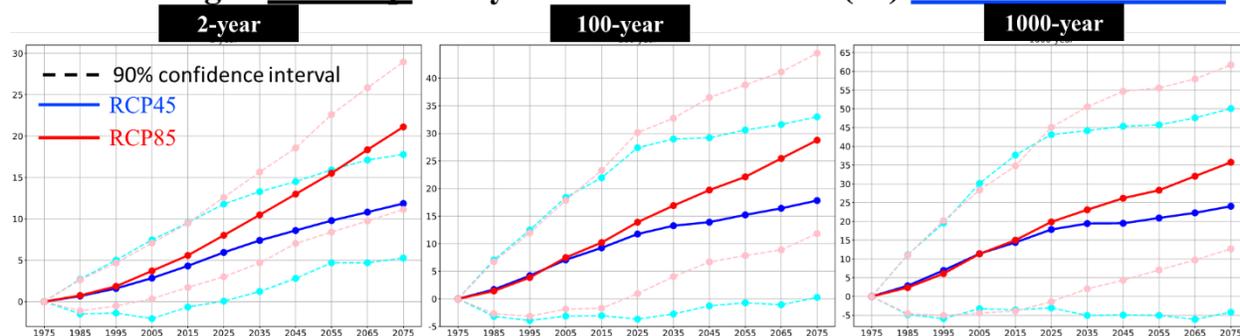


Figure 5.33. Quasi-stationary model results for AMS-based method; 2-, 100-, and 1,000-year recurrence intervals using the UWPD data set and showing RCP4.5 and RCP8.5 scenarios with confidence limits based on the distribution of ensemble results

Table 5.15. Numeric values of the results shown in Figure 5.33

	QS-UWPD	1985-1961-2010	1995-1971-2020	2005-1981-2030	2015-1991-2040	2025-2001-2050	2035-2011-2060	2045-2021-2070	2055-2031-2080	2065-2041-2090	2075-2051-2100
RCP 4.5	2-year	0.67 (-1.49, 2.72)	1.60 (-1.38, 5.01)	2.84 (-2.04, 7.47)	4.32 (-0.66, 9.53)	5.95 (0.07, 11.79)	7.40 (1.22, 13.29)	8.60 (2.81, 14.50)	9.79 (4.70, 15.92)	10.81 (4.69, 17.12)	11.85 (5.27, 17.78)
	100-year	1.70 (-3.17, 7.06)	4.20 (-3.92, 12.48)	7.07 (-3.12, 18.40)	9.25 (-3.04, 21.94)	11.75 (-3.66, 27.40)	13.26 (-2.73, 28.98)	13.91 (-1.23, 29.19)	15.23 (-0.71, 30.61)	16.41 (-1.07, 31.60)	17.82 (0.27, 33.00)
	1000-year	2.85 (-4.83, 11.13)	6.94 (-5.90, 19.54)	11.39 (-3.26, 30.11)	14.39 (-3.58, 37.68)	17.85 (-3.05, 43.15)	19.43 (-5.07, 44.18)	19.49 (-4.95, 45.37)	20.92 (-5.05, 45.73)	22.28 (-6.10, 47.60)	24.04 (-4.20, 50.07)
RCP 8.5	2-year	0.75 (-1.13, 2.64)	1.85 (-0.51, 4.68)	3.71 (0.34, 7.08)	5.58 (1.71, 9.47)	8.02 (3.01, 12.59)	10.48 (4.68, 15.66)	12.99 (7.02, 18.57)	15.51 (8.41, 22.60)	18.35 (9.74, 25.84)	21.11 (11.14, 28.96)
	100-year	1.43 (-2.69, 6.74)	3.87 (-3.14, 11.96)	7.49 (-1.85, 17.86)	10.20 (-1.70, 23.31)	13.90 (0.94, 30.15)	16.90 (4.03, 32.73)	19.72 (6.70, 36.51)	22.11 (7.85, 38.77)	25.46 (8.89, 41.12)	28.76 (11.84, 44.54)
	1000-year	2.35 (-4.57, 11.05)	6.08 (-5.07, 20.23)	11.34 (-4.44, 28.33)	14.99 (-3.89, 34.79)	19.87 (-1.40, 45.10)	23.10 (2.04, 50.53)	26.19 (4.35, 54.67)	28.31 (7.06, 55.56)	32.05 (9.70, 57.96)	35.77 (12.70, 61.70)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.6. Conclusions for daily data

A series of comparisons between NS and regional QS (modified NS by removing the nonstationary parameters) showed that both methods on an average projected precipitation frequency increase for RCP4.5 and particularly for RCP8.5. The QS model runs, however, predicted the percent increase to be nearly twice as high as the nonstationary runs. The increases based on the regional QS, used in this Section, are relatively similar to those based on the point QS (Sections III and IV), as shown in Table 5.16. The differences between the point- and regional QS methods are less than 10%, while the differences between the regional QS and NS are much larger. Regional NS results are shown in Table 5.16, last row.

Table 5.16. 100-year precipitation frequency estimates in inches for UWPD data

	RCP4.5		RCP8.5	
	~2025	~2075	~2025	~2075
Representative year				
100-year – point QS	10.79	16.90	14.18	26.75
100-year – regional QS	11.75	17.82	13.90	28.76
Relative difference	8.90%	5.44%	-1.97%	7.51%
100-year – regional NS	5.28	9.46	5.70	14.92

Confidence limits were represented by the variability of results based on different climate models in the ensemble runs. The regional average of the ratio was calculated for each ensemble run first, then the 5% and 95% quantiles to estimate the confidence limits were found. The effects of other sources of uncertainty, such as those of the statistical model structure, model parameters, sampling variability, and data observations, were not considered in this report. The confidence limits calculated in this study can be viewed as lower limits of the “true” confidence limits, but they are still significant for all cases presented in this research. Future studies should focus on multiple uncertainty sources. In addition, many studies indicated that some climate models are better than others (Knutti 2010) and thus should be given higher weights. This study used equal weights in all calculations, but our experiment suggested that model weights can be beneficial, particularly for near-future time horizons (Appendix I). Thus, future work should explore the potential of the weighted approach.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.7. Ten-day results

This subsection presents statistical frequency results for a 10-day rainfall duration using AMS and PDS methods and LOCA and UWPD data and comparing the NS and QS approaches. The results are presented for 2-, 100-, and 1000-year return periods.

5.7.1. Nonstationary results for AMS

The results of the NS method using the LOCA data and the AMS method are presented in Figs. 5.34 (isohyetal maps) and 5.35 (regional averages of PF increase ratio), followed by Tables 5.17-5.18, showing the numerical values in Fig. 5.35 for RCP4.5 and RCP8.5, respectively. The same results using the UWPD data are presented in Figs. 5.36-5.37 and Tables 5.19-5.20.

Results of NS-LOCA run

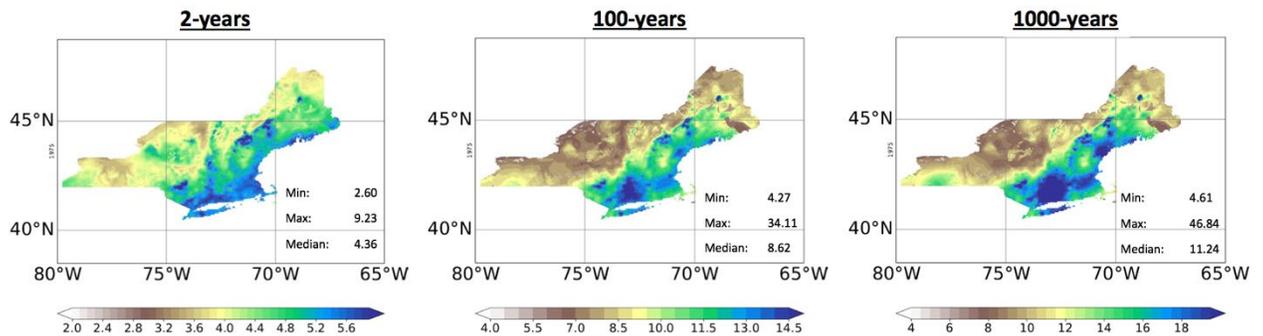


Figure 5.34. Isohyetal map showing 10-day 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the LOCA CMIP5 runs, representing 1975.

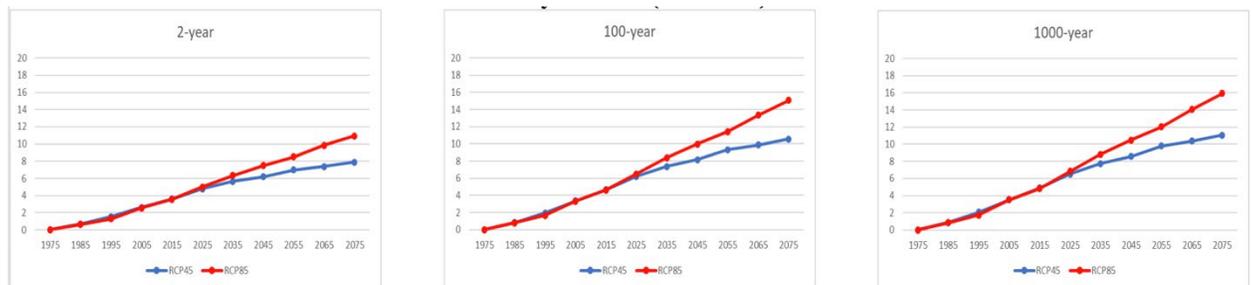


Figure 5.35. Area averaged 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on LOCA CMIP5 runs

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.17. NS-LOCA regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

LOCA cmip5 RCP4.5 10-day AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.65	1.52	2.61	3.61	4.79	5.64	6.21	6.99	7.40	7.90
100-year	0.84	1.96	3.36	4.65	6.23	7.37	8.16	9.31	9.87	10.54
1000-year	0.88	2.05	3.52	4.87	6.53	7.73	8.57	9.79	10.38	11.08

Table 5.18. NS-LOCA regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

LOCA cmip5 RCP8.5 10-day AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.62	1.28	2.56	3.55	5.00	6.33	7.49	8.50	9.83	10.94
100-year	0.80	1.66	3.33	4.61	6.52	8.38	9.97	11.43	13.34	15.08
1000-year	0.84	1.74	3.50	4.83	6.84	8.81	10.48	12.04	14.06	15.93

Results of NS-UWPD run

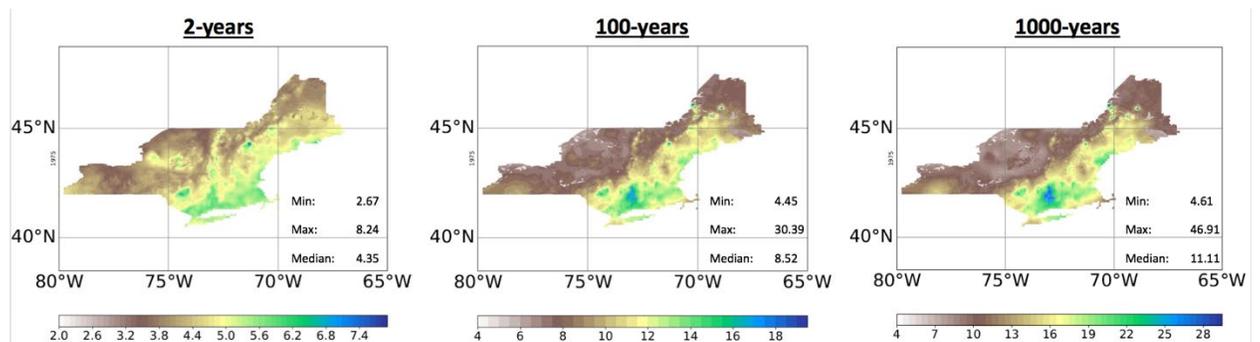


Figure 5.36. Isohyetal map showing 10-day 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the NS-UWPD runs, representing the year 1975.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

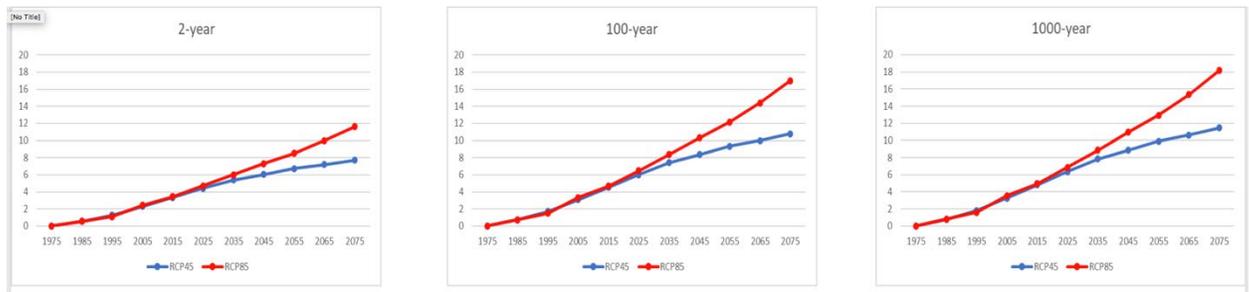


Figure 5.37. Area averaged 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on NS-UWPD runs

Table 5.19. NS-UWPD regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

NS-UWPD RCP4.5 10-day AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.54	1.26	2.28	3.35	4.41	5.40	6.06	6.74	7.19	7.72
100-year	0.73	1.70	3.09	4.54	6.01	7.41	8.35	9.35	10.00	10.79
1000-year	0.77	1.80	3.26	4.79	6.35	7.83	8.84	9.90	10.60	11.45

Table 5.20. NS-UWPD regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

NS-UWPD RCP8.5 10-day AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.55	1.11	2.43	3.43	4.72	6.00	7.32	8.51	9.98	11.60
100-year	0.75	1.50	3.33	4.69	6.48	8.35	10.30	12.15	14.40	17.01
1000-year	0.79	1.58	3.52	4.96	6.86	8.85	10.94	12.93	15.35	18.17

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.7.2. Quasi-stationary results for AMS

The results of the QS method using LOCA data and the AMS method are presented in Figs. 5.38 (isohyetal maps) and 5.39 (regional averages of PF increase ratio with confidence limits), followed by Tables 5.21-5.22 showing the numerical values with confidence limits in Fig. 5.39 for RCP4.5 and RCP8.5, respectively. The same results using UWPD data are presented in Figs. 5.40-5.41 and Tables 5.23-5.24.

Results of QS-LOCA runs

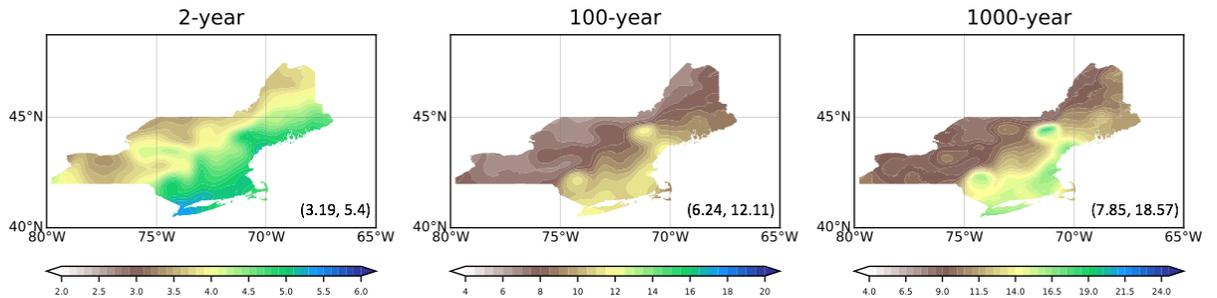


Figure 5.38. Isohyetal map showing 10-day 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the QS-LOCA runs, representing 1975.

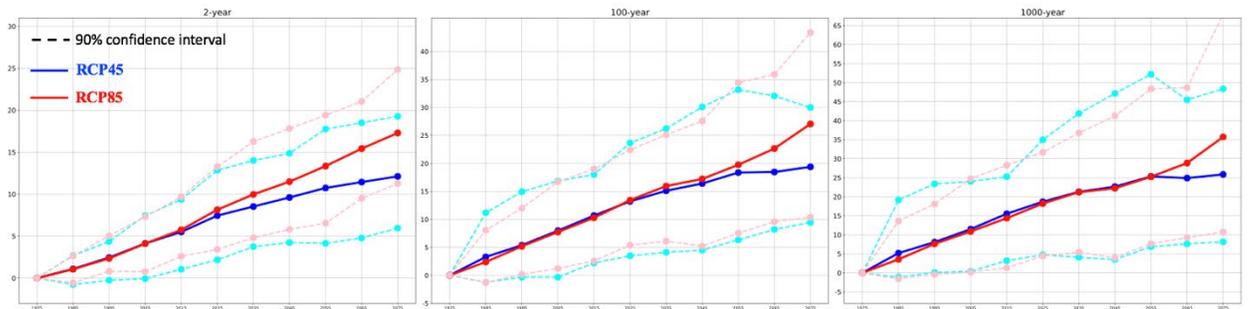


Figure 5.39. Regional averaged 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on QS-LOCA runs

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.21. QS-LOCA regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for RCP4.5 with confidence limits

QS-LOCA RCP4.5 10-day AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	1.09 (-0.75, 2.69)	2.46 (-0.25, 4.37)	4.13 (-0.06, 7.45)	5.51 (1.06, 9.36)	7.45 (2.18, 12.86)	8.54 (3.75, 14.02)	9.61 (4.24, 14.88)	10.75 (4.15, 17.76)	11.45 (4.79, 18.51)	12.13 (5.97, 19.28)
100-year	3.29 (-1.23, 11.18)	5.39 (-0.31, 14.94)	7.99 (-0.34, 16.87)	10.68 (2.22, 18.02)	13.23 (3.50, 23.65)	15.11 (4.12, 26.23)	16.41 (4.50, 30.08)	18.37 (6.33, 33.18)	18.47 (8.21, 32.07)	19.39 (9.45, 29.99)
1000-year	5.20 (-1.10, 19.14)	8.08 (0.10, 23.40)	11.49 (0.41, 24.04)	15.54 (3.22, 25.27)	18.71 (4.75, 34.98)	21.29 (4.10, 41.89)	22.67 (3.51, 47.16)	25.36 (6.89, 52.21)	24.91 (7.62, 45.49)	25.88 (8.16, 48.37)

Table 5.22. QS-LOCA regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for RCP8.5 with confidence limits

QS-LOCA RCP8.5 10-day AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	1.06 (-0.55, 2.63)	2.35 (0.82, 5.04)	4.11 (0.77, 7.31)	5.76 (2.60, 9.69)	8.15 (3.43, 13.28)	9.98 (4.81, 16.28)	11.50 (5.81, 17.82)	13.35 (6.55, 19.41)	15.43 (9.51, 21.06)	17.30 (11.27, 24.85)
100-year	2.4 (-1.26, 8.05)	5.19 (0.19, 11.98)	7.74 (1.21, 16.73)	10.29 (2.57, 19.00)	13.43 (5.39, 22.43)	15.96 (6.13, 25.12)	17.22 (5.21, 27.59)	19.75 (7.54, 34.46)	22.66 (9.57, 35.91)	27.06 (10.40, 43.39)
1000-year	3.57 (-1.56, 13.64)	7.63 (-0.41, 18.11)	10.90 (0.19, 24.76)	14.38 (1.34, 28.29)	18.23 (4.44, 31.72)	21.22 (5.43, 36.77)	22.24 (4.01, 41.25)	25.26 (7.66, 48.34)	28.87 (9.26, 48.67)	35.73 (10.77, 67.97)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Results of QS-UWPD run

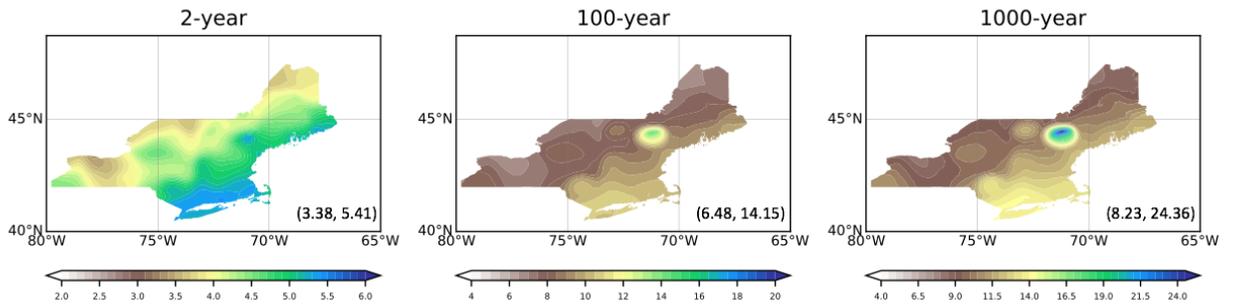


Figure 5.40. Isohyetal map showing 10-day 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the QS-UWPD runs, representing 1975.

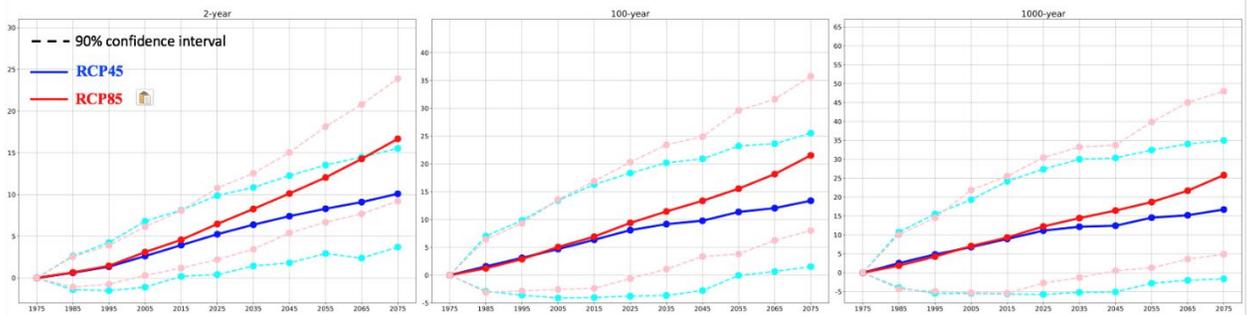


Figure 5.41. Area averaged 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on QS-UWPD runs

Table 5.23. QS-UWPD regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for RCP4.5 with confidence limits

QS-UWPD RCP4.5 10-day AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.62 (-1.40, 2.63)	1.35 (-1.52, 4.25)	2.62 (-1.10, 6.80)	3.93 (0.19, 8.13)	5.24 (0.41, 9.88)	6.39 (1.44, 10.85)	7.41 (1.81, 12.26)	8.30 (2.93, 13.53)	9.09 (2.38, 14.48)	10.09 (3.70, 15.53)
100-year	1.58 (-2.89, 7.02)	3.12 (-3.59, 9.84)	4.69 (-4.10, 13.43)	6.36 (-4.01, 16.30)	8.10 (-3.77, 18.36)	9.18 (-3.65, 20.21)	9.79 (-2.76, 20.91)	11.37 (-0.04, 23.22)	12.06 (0.65, 23.64)	13.39 (1.54, 25.54)
1000-year	2.53 (-3.92, 10.76)	4.84 (-5.44, 15.52)	6.79 (-5.47, 19.31)	8.93 (-5.59, 24.21)	11.15 (-5.74, 27.37)	12.17 (-5.18, 30.00)	12.43 (-5.07, 30.34)	14.56 (-2.78, 32.45)	15.21 (-1.99, 34.04)	16.72 (-1.61, 34.97)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.24. QS-UWPD regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for RCP8.5 with confidence limits

QS-UWPD RCP8.5 10-day AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.66 (-1.09, 2.54)	1.45 (-0.75, 3.90)	3.10 (0.30, 6.14)	4.57 (1.19, 8.07)	6.47 (2.21, 10.77)	8.27 (3.44, 5.41)	10.12 (5.41, 15.04)	12.04 (6.67, 18.14)	14.27 (7.68, 20.8)	16.68 (9.20, 23.90)
100-year	1.24 (-3.08, 6.45)	2.88 (-2.83, 9.33)	5.06 (-2.56, 13.60)	6.93 (-2.36, 16.89)	9.39 (-0.61, 20.32)	11.46 (1.09, 23.45)	13.37 (3.36, 24.90)	15.56 (3.81, 29.61)	18.17 (6.27, 31.63)	21.55 (8.03, 35.87)
1000-year	1.88 (-4.32, 10.07)	4.27 (-4.88, 14.53)	7.05 (-5.27, 21.87)	9.31 (-5.41, 25.56)	12.24 (-2.71, 30.44)	14.44 (-1.27, 33.26)	16.44 (0.55, 33.74)	18.68 (1.30, 39.85)	21.69 (3.60, 45.01)	25.81 (4.89, 47.99)

5.7.3. Nonstationary results for PDS

The results of the NS method using the LOCA data and the PDS method are presented in Figs. 5.42 (isohyetal maps) and 5.43 (regional averages of PF increase ratio), followed by Tables 5.25 and 5.26 showing the numerical values corresponding to Fig. 5.43 for RCP4.5 and RCP8.5, respectively. The same results using UWPD data are presented in Figs. 5.44-5.45 and Tables 5.27 and 5.28.

Results of NS-LOCA run

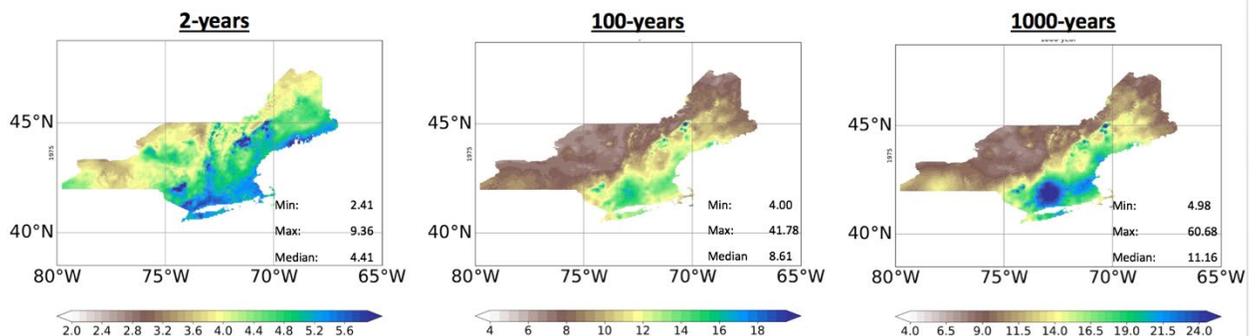


Figure 5.42. Isohyetal map showing 10-day 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the NS-LOCA runs, representing 1975.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

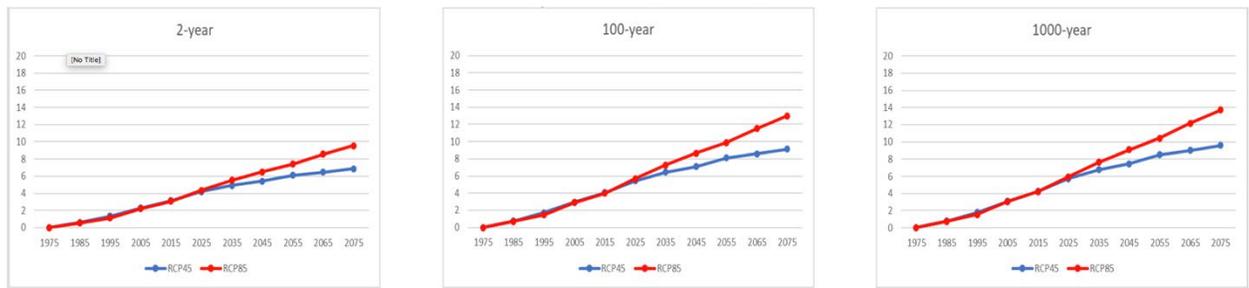


Figure 5.43. Area averaged 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on NS-LOCA runs

Table 5.25. NS-LOCA regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

NS-LOCA RCP4.5 10-day PDS2	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.57	1.33	2.28	3.15	4.19	4.94	5.43	6.12	6.47	6.89
100-year	0.74	1.71	2.95	4.07	5.45	6.44	7.12	8.11	8.58	9.14
1000-year	0.78	1.79	3.08	4.26	5.71	6.76	7.47	8.52	9.02	9.61

Table 5.26. NS-LOCA regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

NS-LOCA RCP8.5 10-day PDS2	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.55	1.11	2.23	3.08	4.36	5.52	6.52	7.40	8.57	9.54
100-year	0.71	1.45	2.91	4.01	5.68	7.28	8.64	9.90	11.53	13.00
1000-year	0.74	1.52	3.05	4.21	5.95	7.65	9.08	10.42	12.15	13.72

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Results of NS-UWPD run

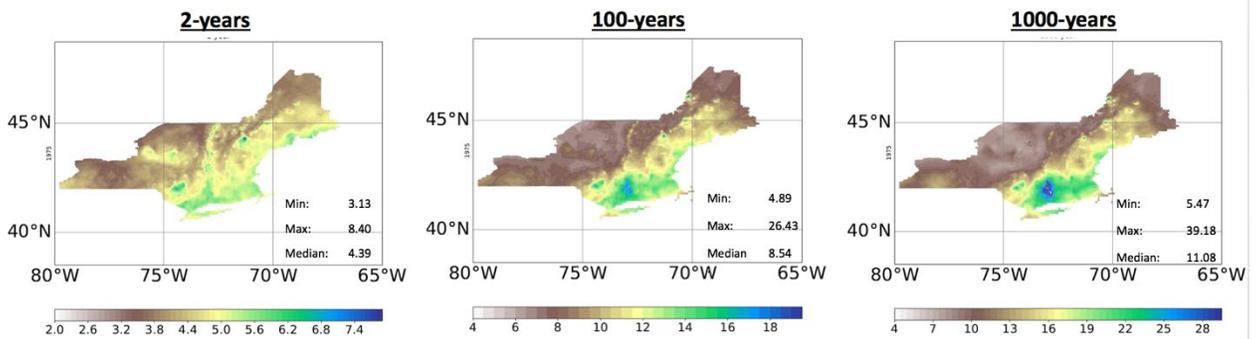


Figure 5.44. Isohyetal map showing 10-day 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the NS-UWPD runs, representing 1975.

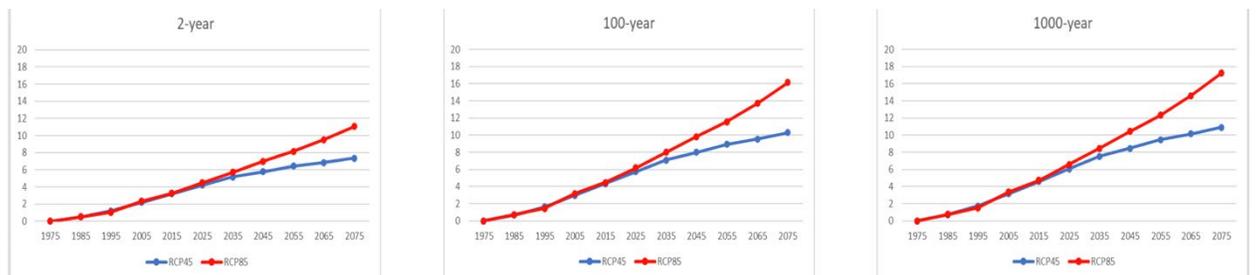


Figure 5.45. Area averaged 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on NS-UWPD runs

Table 5.27. NS-UWPD regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

NS-UWPD RCP4.5 10-day PDS2	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.51	1.20	2.18	3.20	4.21	5.16	5.78	6.43	6.85	7.36
100-year	0.70	1.63	2.96	4.35	5.76	7.10	7.99	8.94	9.56	10.30
1000-year	0.74	1.72	3.13	4.60	6.09	7.51	8.47	9.48	10.14	10.93

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.28. NS-UWPD regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

NS-UWPD RCP8.5 10-day PDS2	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.53	1.06	2.32	3.27	4.50	5.72	6.98	8.12	9.51	11.06
100-year	0.71	1.44	3.19	4.48	6.20	7.98	9.82	11.57	13.69	16.14
1000-year	0.75	1.52	3.37	4.75	6.57	8.46	10.43	12.32	14.59	17.23

5.7.4. Quasi-stationary results for PDS

The results of the QS method using LOCA data and the PDS method are presented in Figs. 5.46 (isohyetal maps) and 5.47 (regional averages of PF increase ratio with confidence limits), followed by Tables 5.29 -5.30 showing the numerical values with confidence limits corresponding to Fig. 5.47 for RCP4.5 and RCP8.5, respectively. The same results using UWPD data are presented in Figs. 5.48-5.49 and Tables 5.31-5.32.

Results of QS-LOCA run

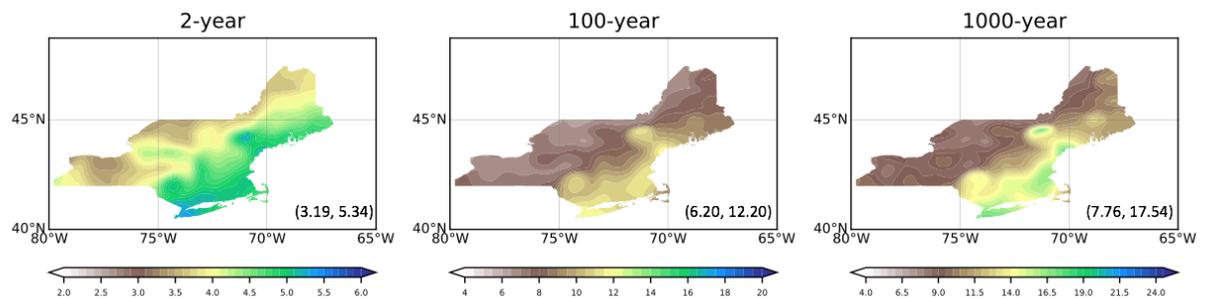


Figure 5.46. Isohyetal map showing 10-day 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the QS-LOCA runs, representing 1975.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

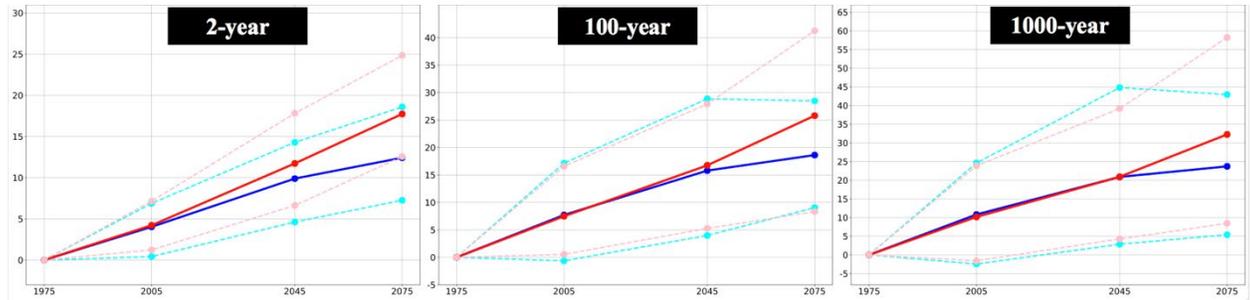


Figure 5.47. Area averaged 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on QS-LOCA runs

Table 5.29. QS-LOCA regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

QS-LOCA RCP4.5 10-day PDS2	2005 1981-2030	2045 2021-2070	2075 2051-2100
2-year	4.03 (0.42, 6.88)	9.89 (4.62, 14.29)	12.42 (7.26, 18.59)
100-year	7.72 (-0.64, 17.15)	15.79 (3.98, 28.87)	18.62 (9.04, 28.48)
1000-year	10.83 (-2.43, 24.62)	20.88 (2.86, 44.81)	23.71 (5.41, 42.94)

Table 5.30. QS-LOCA regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

QS-LOCA RCP8.5 10-day PDS2	2005 1981-2030	2045 2021-2070	2075 2051-2100
2-year	4.22 (1.23, 7.15)	11.73 (6.63, 17.82)	17.74 (12.58, 24.85)
100-year	7.48 (0.53, 16.58)	16.74 (5.27, 27.92)	25.8 (8.33, 41.26)
1000-year	10.17 (-1.58, 23.87)	20.91 (4.27, 39.19)	32.26 (8.46, 58.21)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Results of QS-UWPD run

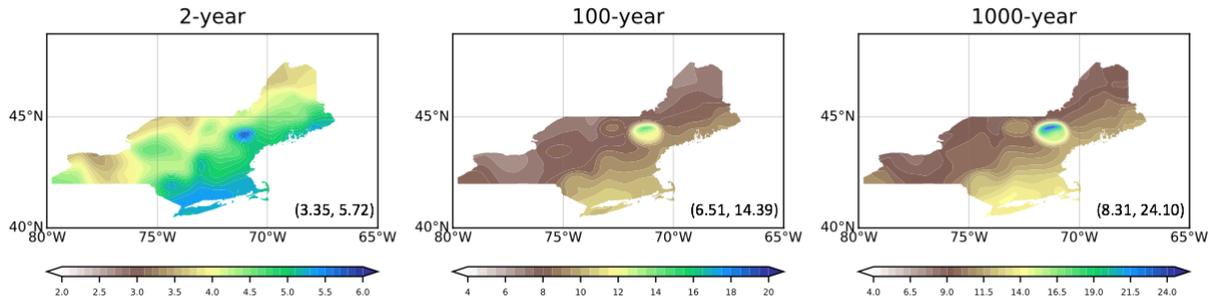


Figure 5.48. Isohyetal map showing 10-day 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the QS-UWPD runs, representing 1975.

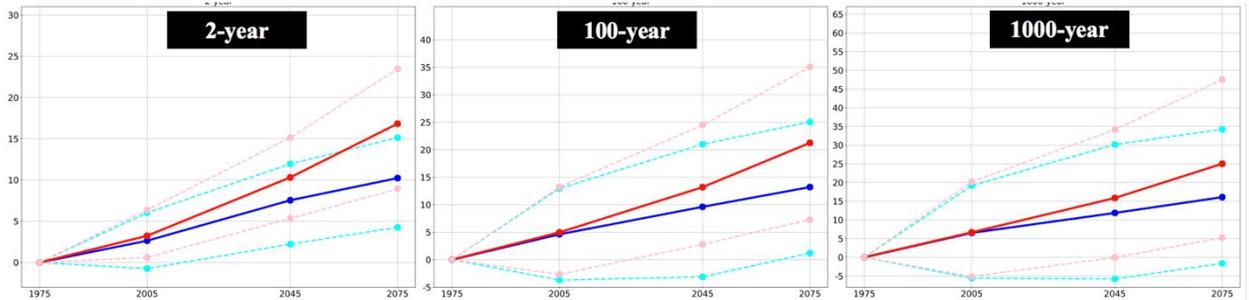


Figure 5.49. Area averaged 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on QS-UWPD runs

Table 5.31. QS-UWPD regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

QS-UWPD RCP4.5 10-day PDS2	2005 1981-2030	2045 2021-2070	2075 2051-2100
2-year	2.64 (-0.74, 6.02)	7.54 (2.22, 11.96)	10.23 (4.26, 15.13)
100-year	4.62 (-3.69, 12.97)	9.62 (-3.12, 21.00)	13.21 (1.21, 25.06)
1000-year	6.53 (-5.59, 19.18)	11.87 (-5.79, 30.18)	16.08 (-1.63, 34.24)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.32. QS-UWPD regional average of 10-day PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

QS-UWPD RCP8.5 10-day PDS2	2005 1981-2030	2045 2021-2070	2075 2051-2100
2-year	3.23 (0.61, 6.36)	10.31 (5.34, 15.15)	16.83 (8.94, 23.47)
100-year	4.97 (-2.66, 13.22)	13.19 (2.79, 24.50)	21.25 (7.29, 35.03)
1000-year	6.68 (-5.14, 20.20)	15.86 (-0.04, 34.15)	25.03 (5.24, 47.51)

5.7.5. Comparison of all methods

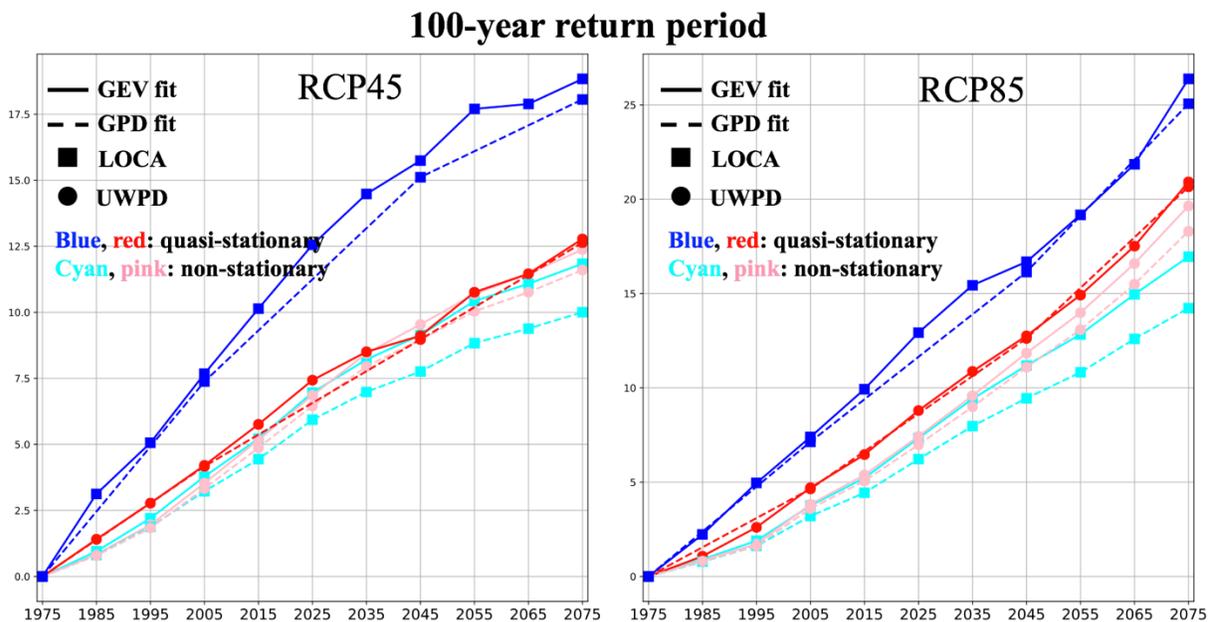


Figure 5.50. Regional average 10-day PF increase ratio (%) relative to 1975 for 100-year events based on all methods for scenario RCP4.5 and RCP8.5

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

The results in Section 5.7 show a large difference between the LOCA and UWPD data sets. Although the QS method based on LOCA produces significantly larger projected PF estimates than the NS approaches, the same results based on UWPD are smaller and closer to NS, particularly for RCP4.5 (Fig. 5.50). For the LOCA data, AMS-based 100-year increases for RCP4.5 and RCP8.5 are 19.39% and 27.06%, respectively (Tables 5.21-5.22). The corresponding 1-day values are 16.32% and 24.74% (Table 5.4), indicating that the percent increase for the 10-day period is larger than for the 1-day period. For the UWPD data, the percent increase for the 10-day period (Tables 5.23-5.24) is smaller than for the 1-day period (Table 5.5). This result could be explained by the differences in statistical downscaling methods for LOCA and UWPD, possibly producing a different spatiotemporal correlation structure, accumulation of rainy days, and the like.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.8. Hourly results

This subsection shows the projected frequency estimates for hourly data. For QS, these results include only the UWPD runs because the LOCA does not have hourly data.

5.8.1. Nonstationary results for AMS

The results of the NS method using the LOCA data and the AMS method are presented in Figs. 5.51 (isohyetal maps) and 5.52 (regional averages of PF increase ratio), followed by Tables 5.33-5.34 showing the numerical values corresponding to Fig. 5.52 for RCP4.5 and RCP8.5, respectively. The same results using UWPD data are presented in Figs. 5.53-5.54 and Tables 5.35-5.36.

Results of NS-LOCA run

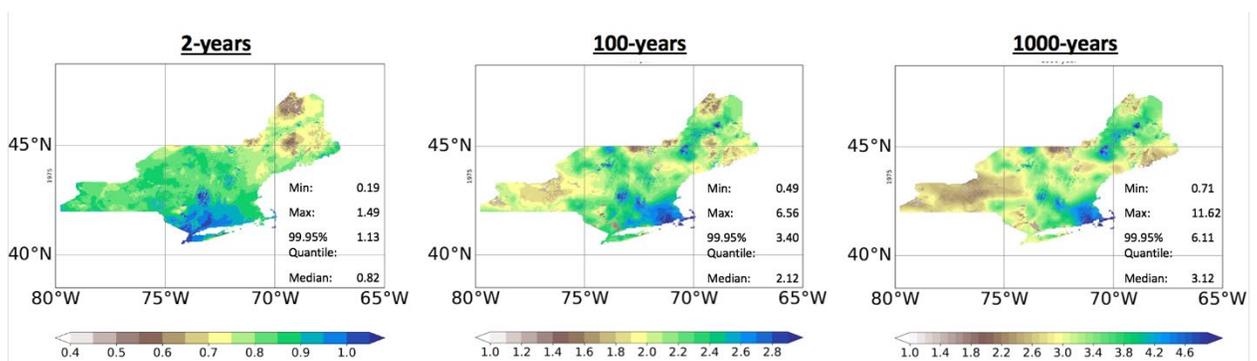


Figure 5.51. Isohyetal map showing hourly 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the LOCA CMIP5 runs, representing 1975.

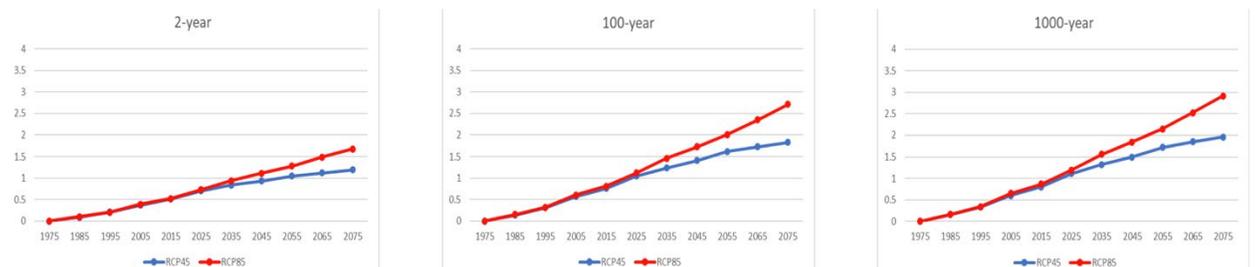


Figure 5.52. Area averaged hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on LOCA CMIP5 run

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.33. NS-LOCA regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

NS-LOCA RCP4.5 hourly AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.09	0.21	0.37	0.51	0.70	0.84	0.93	1.05	1.12	1.19
100-year	0.14	0.31	0.57	0.76	1.05	1.24	1.41	1.62	1.73	1.83
1000-year	0.16	0.33	0.60	0.80	1.11	1.32	1.50	1.72	1.85	1.96

Table 5.34. NS-LOCA regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

NS-LOCA RCP8.5 hourly AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.10	0.20	0.39	0.52	0.73	0.94	1.11	1.28	1.49	1.68
100-year	0.15	0.32	0.61	0.81	1.12	1.46	1.73	2.01	2.35	2.71
1000-year	0.16	0.34	0.65	0.87	1.19	1.56	1.84	2.15	2.52	2.91

Results of NS-UWPD run

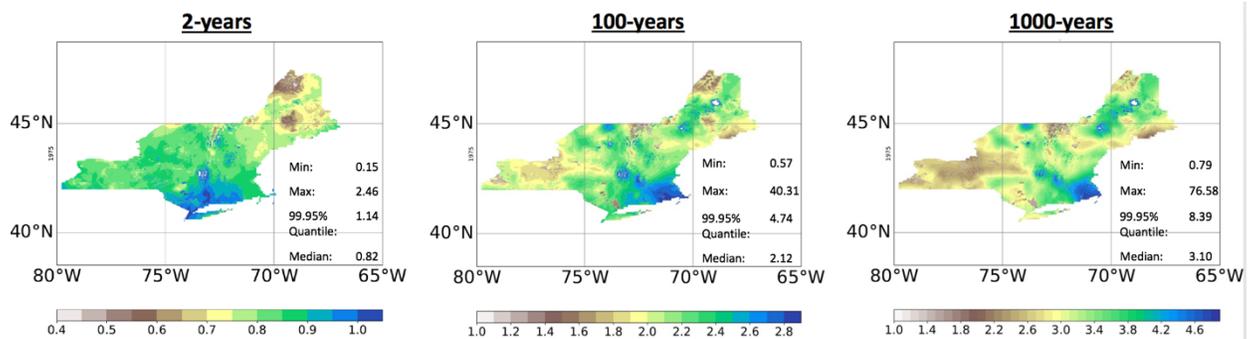


Figure 5.53. Isohyetal map showing hourly 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the NS-UWPD runs, representing 1975.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

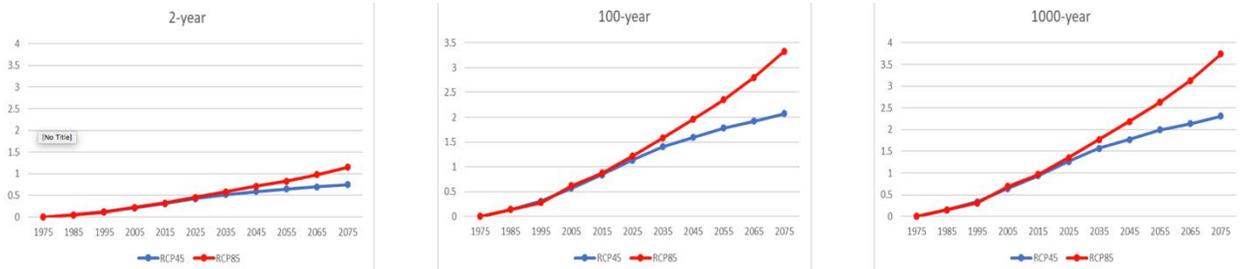


Figure 5.54. Area averaged hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on NS-UWPD run

Table 5.35. NS-UWPD regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

NS-UWPD RCP4.5 hourly AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.05	0.12	0.21	0.31	0.42	0.52	0.59	0.65	0.70	0.75
100-year	0.14	0.31	0.57	0.84	1.13	1.40	1.59	1.78	1.92	2.07
1000-year	0.15	0.34	0.64	0.94	1.26	1.57	1.77	1.99	2.14	2.31

Table 5.36. NS-UWPD regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

NS-UWPD RCP8.5 hourly AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.05	0.11	0.22	0.32	0.45	0.58	0.71	0.83	0.98	1.15
100-year	0.14	0.28	0.62	0.87	1.21	1.58	1.96	2.35	2.80	3.33
1000-year	0.15	0.31	0.69	0.97	1.35	1.77	2.19	2.63	3.13	3.74

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.8.2. Quasi-stationary results for AMS

The results of the QS method using UWPD data and the AMS method are presented in Figs. 5.55 (isohyetal maps) and 5.56 (regional averages of PF increase ratio and their confidence limits), followed by Tables 5.37- 5.38 showing the numerical values corresponding to Fig. 5.56 for RCP4.5 and RCP8.5, respectively. Similar results based on the LOCA data are not available, as this data set does not have hourly data.

Results of QS-UWPD run

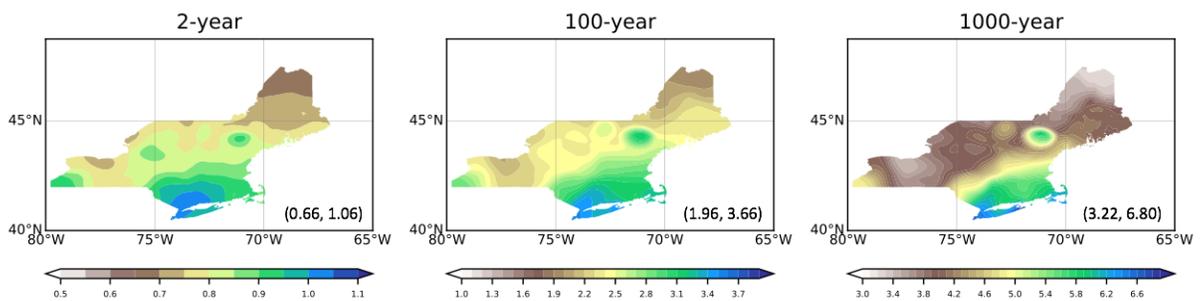


Figure 5.55. Isohyetal map showing hourly 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the QS-UWPD runs, representing the year 1975.

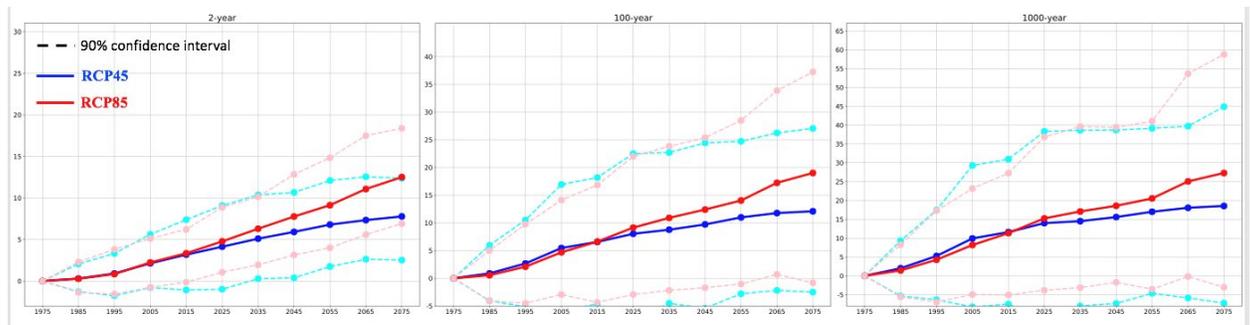


Figure 5.56. Area averaged hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on the QS-UWPD run

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.37. QS-UWPD regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

QS-UWPD RCP4.5 hourly AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.3 (-1.28, 2.06)	0.91 (-1.76, 3.32)	2.14 (-0.79, 5.61)	3.18 (-1.07, 7.39)	4.13 (-0.99, 9.08)	5.1 (0.29, 10.36)	5.92 (0.39, 10.66)	6.79 (1.75, 12.10)	7.33 (2.63, 12.55)	7.78 (2.52, 12.36)
100-year	0.88 (-4.07, 5.95)	2.67 (-5.12, 10.51)	5.46 (-5.88, 16.94)	6.56 (-4.62, 18.18)	8.03 (-7.45, 22.49)	8.77 (-4.53, 22.70)	9.72 (-5.39, 24.44)	10.99 (-2.80, 24.71)	11.77 (-2.18, 26.25)	12.09 (-2.50, 27.05)
1000-year	2.01 (-5.36, 9.30)	5.28 (-6.33, 17.45)	9.94 (-8.24, 29.30)	11.62 (-7.52, 30.96)	14.00 (-11.64, 38.32)	14.51 (-7.99, 38.65)	15.60 (-7.33, 38.71)	17.01 (-4.64, 39.17)	18.07 (-5.86, 39.73)	18.55 (-7.25, 44.89)

Table 5.38. QS-UWPD regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

QS-UWPD RCP8.5 hourly AMS	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.28 (-1.37, 2.30)	0.84 (-1.58, 3.81)	2.24 (-0.75, 5.12)	3.34 (-0.15, 6.22)	4.77 (1.06, 8.88)	6.29 (1.95, 10.13)	7.76 (3.14, 12.84)	9.13 (4.00, 14.84)	11.08 (5.60, 17.49)	12.51 (6.91, 18.38)
100-year	0.55 (-4.07, 5.01)	2.10 (-4.50, 9.77)	4.69 (-2.93, 14.11)	6.62 (-4.34, 16.82)	9.14 (-2.93, 21.99)	10.90 (-2.18, 23.82)	12.41 (-1.70, 25.38)	14.04 (-1.03, 28.49)	17.23 (0.71, 33.88)	19.01 (-0.82, 37.25)
1000-year	1.43 (-5.60, 8.25)	4.28 (-6.94, 17.33)	8.19 (-4.95, 23.18)	11.33 (-5.07, 27.26)	15.25 (-3.83, 36.85)	17.08 (-3.11, 39.63)	18.59 (-1.72, 39.43)	20.55 (-3.54, 41.04)	25.07 (-0.18, 53.66)	27.30 (-3.02, 58.77)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.8.3. Nonstationary results for PDS

The results of the NS method using the LOCA data and the PDS method are presented in Figs. 5.57 (isohyetal maps) and 5.58 (regional averages of PF increase ratio), followed by Tables 5.39- 5.40 showing the numerical values corresponding to Fig. 5.58 for RCP4.5 and RCP8.5, respectively. The same results using UWPD data are presented in Figs. 5.59-5.60 and Tables 5.41-5.42.

Results of NS-LOCA run

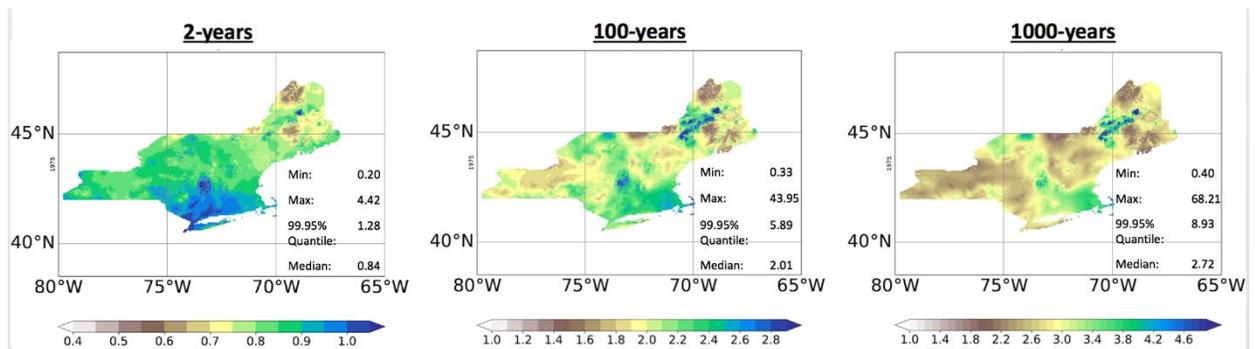


Figure 5.57. Isohyetal map showing hourly 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the NS-LOCA run, representing 1975.

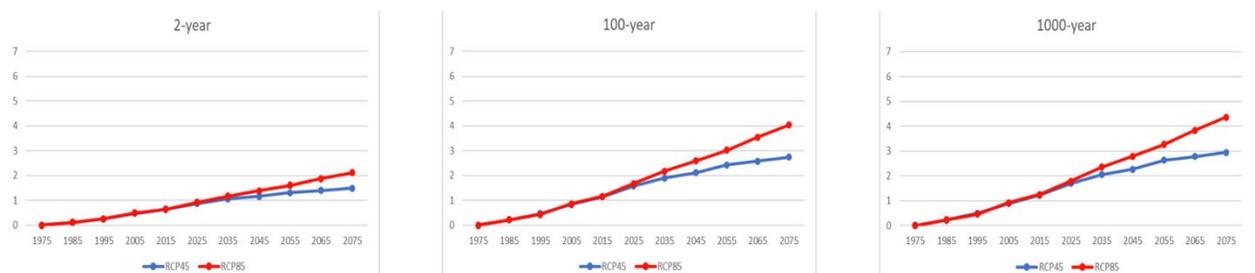


Figure 5.58. Area averaged hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on NS-LOCA run

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.39. NS-LOCA CMIP5 regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

NS-LOCA RCP4.5 hourly PDS2	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.12	0.26	0.47	0.64	0.88	1.06	1.17	1.32	1.41	1.49
100-year	0.21	0.47	0.83	1.15	1.58	1.90	2.11	2.43	2.58	2.74
1000-year	0.22	0.50	0.89	1.23	1.70	2.05	2.27	2.63	2.78	2.95

Table 5.40. NS-LOCA regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

NS-LOCA RCP8.5 hourly PDS2	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.12	0.25	0.49	0.65	0.92	1.18	1.40	1.61	1.87	2.11
100-year	0.21	0.43	0.86	1.16	1.67	2.17	2.59	3.02	3.54	4.04
1000-year	0.23	0.46	0.93	1.25	1.79	2.35	2.80	3.27	3.83	4.37

Results of NS-UWPD run

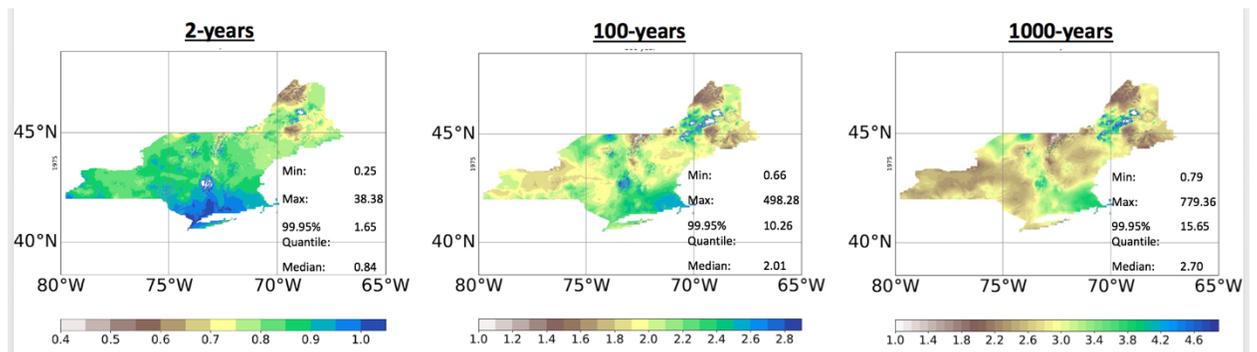


Figure 5.59. Isohyetal map showing hourly 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the NS-UWPD run, representing 1975.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

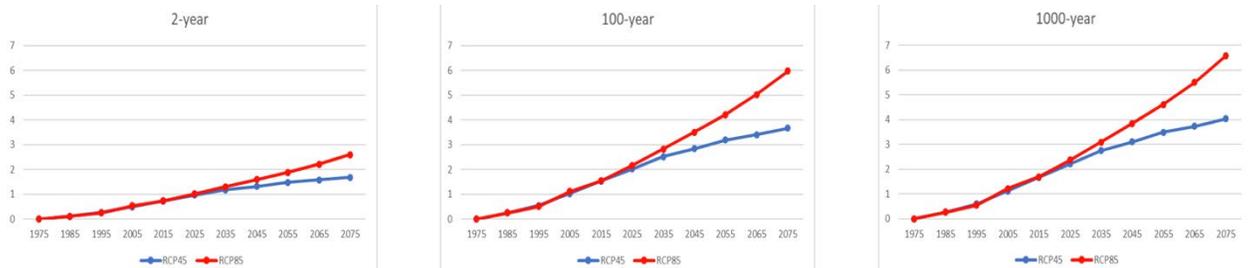


Figure 5.60. Area averaged hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on NS-UWPD run

Table 5.41. NS-UWPD regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

NS-UWPD RCP4.5 hourly PDS2	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.12	0.27	0.49	0.72	0.96	1.18	1.32	1.48	1.58	1.69
100-year	0.25	0.55	1.03	1.53	2.03	2.52	2.84	3.19	3.41	3.67
1000-year	0.27	0.60	1.12	1.67	2.21	2.75	3.10	3.49	3.73	4.03

Table 5.42. NS-UWPD regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

NS-UWPD RCP8.5 hourly PDS2	1985	1995	2005	2015	2025	2035	2045	2055	2065	2075
2-year	0.12	0.25	0.53	0.74	1.02	1.31	1.60	1.89	2.22	2.60
100-year	0.25	0.51	1.11	1.55	2.17	2.82	3.51	4.21	5.02	5.97
1000-year	0.27	0.55	1.21	1.69	2.37	3.09	3.84	4.61	5.50	6.56

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

5.8.4. Quasi-stationary results for PDS

The results of the QS method using the UWPD data and the PDS method are presented in Figs. 5.61 (isohyetal maps) and 5.62 (regional averages of PF increase ratio and their confidence limits), followed by Tables 5.43-5.44 showing the numerical values corresponding to Fig. 5.62 for RCP4.5 and RCP8.5, respectively. Similar results based on the LOCA data are not available, as this data set does not have hourly data.

Results of QS-UWPD run

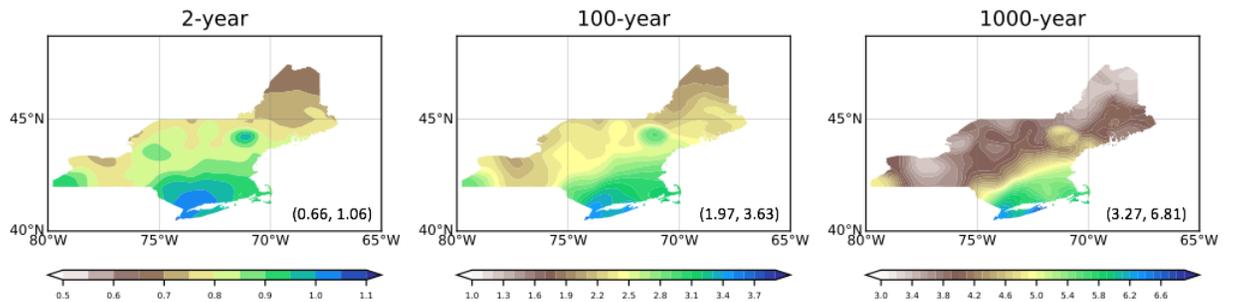


Figure 5.61. Isohyetal map showing hourly 2-, 100-, and 1000-year precipitation in inches. The calculations are based on the QS-UWPD runs, representing 1975.

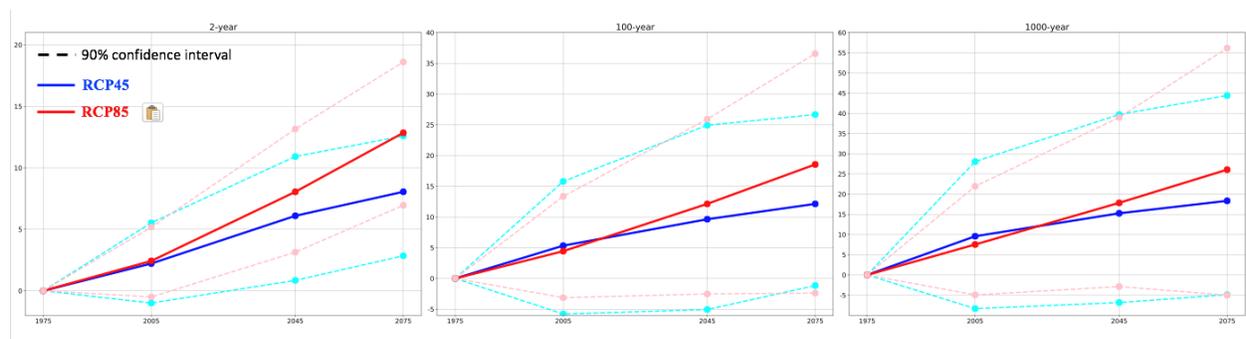


Figure 5.62. Area averaged hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events based on QS-UWPD run

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table 5.43. QS-UWPD regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP4.5

QS-UWPD RCP4.5 hourly PDS2	2005 1981-2030	2045 2021-2070	2075 2051-2100
2-year	2.22 (-1.00, 5.53)	6.10 (0.85, 10.92)	8.05 (2.85, 12.59)
100-year	5.34 (-5.79, 15.79)	9.64 (-5.05, 24.94)	12.14 (-1.16, 26.67)
1000-year	9.59 (-8.33, 28.07)	15.26 (-6.84, 39.73)	18.36 (-4.89, 44.43)

Table 5.44. QS-UWPD regional average of hourly PF increase ratio (%) relative to 1975 (representing 1950-2005) for 2-, 100-, and 1000-year events for scenario RCP8.5

QS-UWPD RCP8.5 hourly PDS2	2005 1981-2030	2045 2021-2070	2075 2051-2100
2-year	2.43 (-0.50, 5.18)	8.05 (3.14, 13.16)	12.85 (6.95, 18.61)
100-year	4.44 (-3.12, 13.35)	12.14 (-2.52, 25.92)	18.56 (-2.38, 36.58)
1000-year	7.54 (-4.94, 21.95)	17.85 (-2.86, 39.00)	26.07 (-5.01, 56.16)

5.8.5. Comparison of all methods

The hourly PF results presented in this section (Figure 5.63) show a large difference between the NS and QS methods. The results also vary for LOCA and UWPD data sets, and for AMS and PDS approaches. The QS method requires projected hourly data. It used only the UWPD hourly projected data, as the LOCA data set has only daily projected data. The NS method, however, used both the UWPD and LOCA data, as it does not require projected hourly data. It only uses projected annual data as a covariate.

The main finding in this section is that the QS method based on UWPD produced significantly larger projected PF estimates than the NS method for both data sets. For example, for the AMS-based approach, the RCP8.5 scenario, and the 100-year event, the NS-UWPD

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

projected the maximum increase of 3.33% (Table 5.36), which is much smaller than the 19.01% obtained based on the QS-UWPD model (Table 5.38). The resulting increases based on the QS were more comparable with those of the daily data. Using annual precipitation as covariates for hourly extremes, as in the NS approach, may need to be reevaluated. The NS approach also exhibited large sensitivity to location, i.e., large spatial variability. For example, 100-year precipitation in inches based on the NS-UWPD run varies between 0.66 inches and 498.28 inches (Figure 5.59), compared to a more reasonable range between 1.97 inches and 3.63 inches for QS-UWPD (Figure 5.61).

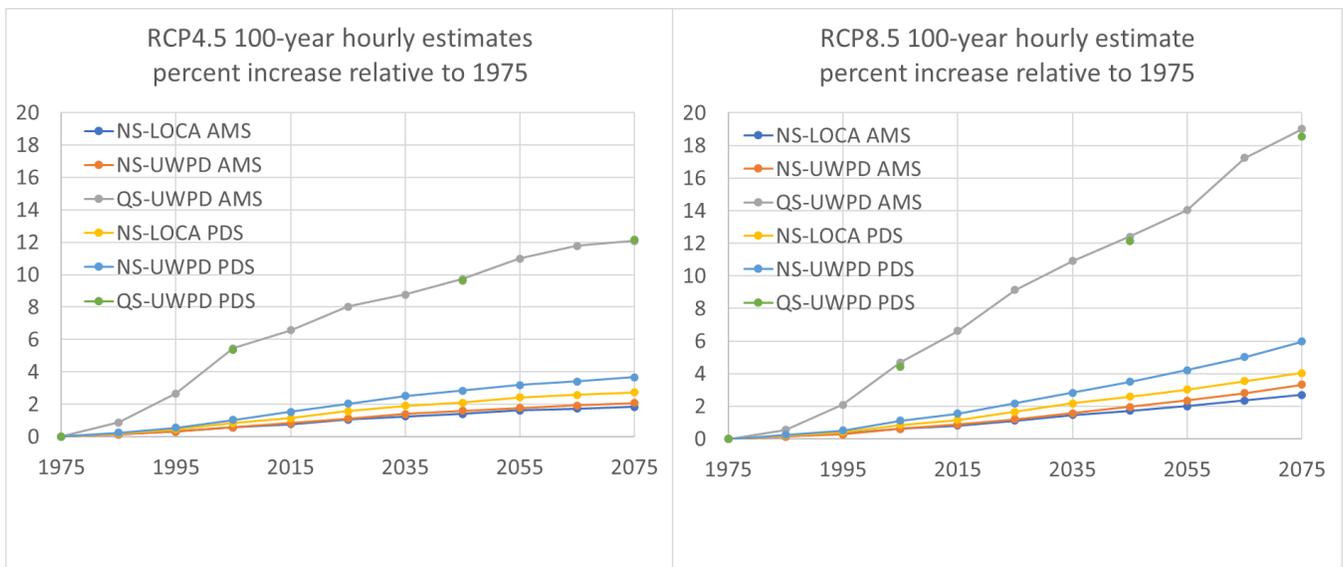


Figure 5.63. Regional average 1-hour PF increase ratio (%) relative to 1975 for 100-year events based on all methods for scenario RCP4.5 and RCP8.5

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

SECTION VI. Conclusions and recommendations

Among the three frequently used and publicly available sources (dynamically downscaled NA-CORDEX data set, statistically downscaled BCCAv2, and the LOCA data sets) and the one source of data generated for this project (the statistically downscaled UWPD data set), two of them, UWPD and LOCA, were adopted for this study. The selection was based on a) their hindcast accuracy, e.g., accuracy of simulating precipitation frequency statistics of historical data, b) availability of ensembles of models for ensemble analysis, and c) popularity and familiarity of the data set in the literature. The UWPD data set provided a higher hindcast accuracy and more ensemble runs than the LOCA data set and included hourly data unlike the LOCA data set. The LOCA had a finer spatial resolution than the UWPD (1/16 deg. vs. 1/10 deg.), but it was selected primarily because of its wide use in published studies and government reports. Both data sets are recommended for future use. In future studies, additional consideration should also be given to the benefits of dynamically downscaled data once the large-size ensembles become available.

Two main approaches, the quasi-stationary (QS) and nonstationary (NS) methods, were compared in this project. For two assumed scenarios, RCP4.5 and RCP8.5, most comparisons are done by first averaging all model results at each site, for each frequency and duration, followed by averaging all sites in the region (Northeast or Texas). In the early analyses (Sections III and IV) the QS method was applied for each point without considering the nearby stations. The final estimates at each point were averaged based on the ensemble runs consisting of multiple models and realizations. The QS method adopted in Section V, however, is a simplified version of the nonstationary (NS) Penn State method, in which the parameters for the covariates are set to zero during the optimization process, reducing the method to its stationary option. The intention of such approach was to isolate the effect of the model parameters representing nonstationarity.

The 1-day duration results demonstrated that the differences between the point-based (Sections III and IV) and the Penn State model-based (Section V) QS methods were small compared with the differences between QS and NS (Table 5.16). QS results are also closer to the published source NCA4 (Table 4.1). The NS method also exhibited sensitivity in some cases, such as the future decreases in precipitation frequency estimates (Section 5.2).

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

The Penn State method-based QS approach was also applied to precipitation frequency with 10-day and 1-hour durations. For a 10-day duration, the PF increases based on QS-LOCA are higher than those of the QS-UWPD, which are similar to NS for RCP4.5. While the QS method on average produced larger projected increases than the NS, the difference between LOCA and UWPD was significant. For hourly data, however, the difference between NS and QS was dominant. The NS method resulted in an approximate increase of 2%-4% between 1975 and 2075 under RCP4.5, which appears low compared with some other published work in which the recently observed (Winters, 2015) and projected (Easterling, 2017) hourly PF have a more significant increase. The NS method applied to hourly data also exhibited a higher spatial variability (Section 5.8) than QS.

The simplicity of directly using the downscaled and debiased climate data (QS) makes this approach favorable, particularly with the apparent stability of the results and closer alignment with the published work. Further development of this approach should include weighted ensemble analysis (Appendix I), and more accurate simulations of spatial and temporal structure of future projected data by combining dynamical and statistical downscaling.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

APPENDIX I. Assessment of the benefits of climate model weights for ensemble analysis in three urban precipitation frequency studies

In hydrology, projected climate change impact assessment studies typically rely on ensembles of downscaled climate model outputs. Due to large modeling uncertainties, the ensembles are often averaged to provide a basis for studying the effects of climate change. A key issue when analyzing averages of a climate model ensemble is whether to weigh all models in the ensemble equally or to use a weighted approach. Many studies have advocated for the latter, based on the assumption that models that are better at simulating the past, i.e., the models with higher hindcast accuracy, will give more accurate forecasts for the future and thus should receive higher weights. To examine this issue, observed and modeled daily precipitation frequency (PF) estimates for three urban areas in the United States: Boston, Massachusetts, Houston, Texas, and Chicago, Illinois were analyzed. For the comparison we used the University of Wisconsin Probabilistic Downscaling (UWPD) data set, consisting of 24 Coupled Model Intercomparison Project Phase 5 (CMIP5) models. The PFs from these models were compared with the observed PFs for a specific historical training period to determine model weights for each area. The unweighted and weighted averaged model PFs from a more recent testing period were then compared with their corresponding observed PFs to determine if weights improved the estimates. These comparisons indeed showed that the weighted averages are closer to the observed values than the unweighted averages in nearly all cases. The study also demonstrated how weights can help reduce model spread in future climate projections by comparing the unweighted and weighted ensemble standard deviations in these projections. In all studied scenarios, the weights actually reduced the standard deviations compared to the equal-weights approach. An analysis of these results' sensitivity to the areal reduction factor used to allow comparisons between point station measurements and grid-box averages is provided.

I.1. Introduction

In many cities in the United States, heavy storms have become more frequent and stronger than those used to design the existing urban drainage infrastructure, causing more frequent floods. Moreover, climate studies suggest that the heavy storms may become even more

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

frequent and intense in the future (Douglas and Fairbank 2011, Markus et al. 2012; Wuebbles et al. 2017; Um et al. 2017, 2018; Li et al. 2019). Large metropolitan areas are particularly vulnerable to climate change because of the complex interaction between climate change and urbanization (Zhang et al. 2018). To address the problem effectively, ensembles of model-generated data are often used to simulate the variability of modeling outputs for future scenarios and time horizons. When analyzing climate model ensembles, the averaged model performance is typically considered a “most likely” or “expected” outcome. A key issue, however, is whether to weigh all models in an ensemble equally, an approach often referred to as model democracy (Knutti, 2010), or to use a weighted approach when finding these averages. It can be assumed that since models all exhibit differences, in terms of dynamic cores, the parameters they emphasize, model resolutions, etc., some perform better in certain applications than others. Thus, a scheme in which these better performing models are given higher weights than the underperforming ones in theory would be better at predicting future climate than an unweighted average. The recent Fourth National Climate Assessment (NCA; Wuebbles et al. 2017) addresses model weighting and recommends a method to determine weights based on model performance (Sanderson and Wehner 2017). Although in the NCA report it is assumed that models that are better at predicting past climate will also perform better with future climates, it also outlines issues and uncertainties of weighted approaches that need to be considered. Since future performances are not known, hindcast accuracies are often used instead as a proxy for model fitness when determining weights. Thus, the models with better hindcast accuracies will receive higher weights in ensembles used for future projections.

Many studies have advocated for using a weighted model ensemble approach. Sanchez et al. (2009) and Räisänen and Ylhäisi (2012) both found that using weights improved their results. Masson and Knutti (2011) argued that often not all models in an ensemble are independent. Some models share a common underlying structure but with slightly different parameters or resolutions, especially if the models come from the same institution. Equal weights do not account for these dependencies and risk giving too much emphasis to a particular underlying model structure. Knutti et al. (2017) also argued that equal weights can allow poorer models to introduce biases and found that a weighting scheme based on both model performance and independence improved their results. Likewise, the Fourth National Climate Assessment (NCA4)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

also used model skill and independence to determine their weights (Sanderson and Wehner 2017; USGCRP 2017). They found that weighting did not strongly influence the mean projections, but they still recommended its use to guard against highly replicated but poorly performing models.

At the same time, there are some opposing viewpoints. Weigel et al. (2010) pointed out that in order to use weights, accurate knowledge of each individual model's skill is required, and unpredictable model noise must also be considered. If these uncertainties are not fully accounted for, then weights could actually do more harm than good by making the ensemble performance worse. They are concerned that this is a real possibility since there is no universal objective consensus on how to find weights, so they suggest it may be safer to use equal weights. Christensen et al. (2010) experimented with how to determine model weights but in the end suggested that it was not beneficial to use them. They argued that the subjective nature of determining weights and the associated uncertainties led to even more uncertainties during the weighting process itself. Finally, it has never been demonstrated that models with higher hindcast accuracies will indeed produce more accurate forecasts. Therefore, the question of whether to weight models remains a topic of debate.

The purpose of this study is to contribute to this debate by determining the benefits of weights in designing an experiment in which the observed data were divided into training (1961-2000 or 2005) and testing (2006-2018) data sets (Fig. I.1). This experiment allowed a cross-validation methodology to assess the model performance by comparing the known observed training-testing data. In the first step, weights were determined by comparing precipitation frequency (PF) estimates based on different climate models with those based on the training set of the observed data (Fig. I.1). Models with PF estimates closer to the observed ones received higher weights. These weights were then applied to the testing data set to determine if they provide a more accurate approximation of the observed PF. If the weighted ensemble means were closer to the observed PF than the equal-weights ensemble mean, the weighted approach would be considered beneficial. Additionally, the approach with a smaller standard deviation of the model results was considered advantageous because of smaller variability in the model results. It was hypothesized that adding weights will provide a more accurate average PF and smaller model variability.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

The climate modeling data we used in predicting future PF quantiles are based on the fine-resolution Climate Model Intercomparison Project Phase 5 (CMIP5) raw model output. To compare the input representing larger areas with estimates based on observed data representing single points, similar to Markus et al. (2018), the climate data were multiplied by the inverse areal reduction factor defined by Hershfield (1961).

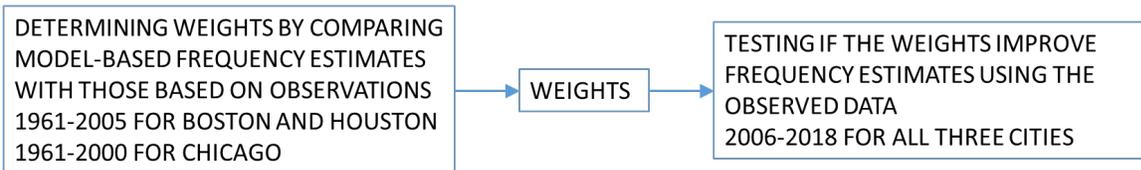


Figure I.1. Schematic of the method to determine the benefits of using weighted averages

This experiment was performed using the University of Wisconsin Probabilistic Downscaling (UWPD) data set (Notaro et al. 2014; Lorenz et al. 2016; Wu et al. 2019) consisting of 24 CMIP5 models to study three small, relatively climatically homogeneous urban areas. Using a specific period in the past as the training period, we compared the modeled PF hindcasts with the observed PFs to determine weights for each model. We then used these weights to average the modeled PFs for a more recent testing period. These results could then be compared with the unweighted averages to see which came closer to the observed PFs in the testing period to determine if and to what degree the weights improved the predictions.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

I.2. Data

For this study, we focused on the areas around three US large urban centers: Boston, MA, Houston, TX, and Chicago, IL (Fig. I.2). All three are adjacent to relatively large bodies of water. Rectangular areas surrounding each city of a few square degrees were chosen as the targets for this study. These areas were selected to be small enough so that the climate would be relatively homogeneous across the entire area, yet big enough to include at least 10 stations, to minimize the effects of potential outliers. The areas selected for Boston and Houston are contained within one climate region each as defined by NOAA Atlas 14. The Chicago area is the same as in Markus et al. (2018). Table I.1 lists the coordinates and other relevant data for each of these targeted rectangular areas.

The training period for Boston and Houston was selected to be 1960-2005 (46 years), as that was the historical period for the available UWPD data. Only stations within the study target areas with at least 80% observed coverage during this period were selected (Wu et al. 2019., The data were obtained from NOAA and were used in the development of Atlas 14. For Chicago, the selected stations are a subset of those used in Markus et al. (2017) from the Global Historical Climatology Network Daily (GHCND). Markus et al. (2017) looked at observations from 1961-2000 (40 years). That same period was selected here as the training period to take advantage of prepared observational data sets from that study. Station data for Chicago were manually checked, and only those with few missing years were retained. The criteria used were at least 80% completeness for 1961-1980 and at least 70% completeness for 1981-2000. As a result, 15 of the 30 stations were selected (Fig. I.2).

The testing period, for which the weighted and unweighted averaged modeled PFs are compared with observed values to determine which better estimates the results, was 2006-2018 for all three cities. The first year of future projections for the UWPD data is 2006, so the selected period provides as many complete years of observations as possible to compare with these projections. Following Markus et al. (2017) and the NCA4 (USGCRP 2017), we employed UWPD data for this period from two Representative Concentration Pathways (RCPs): RCP4.5 and RCP8.5. RCP4.5 represents a low-end emission scenario, while RCP8.5 represents a high-

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

end one, as the NCA4 recommended their use to cover the general range of possible emission scenarios. Observed data for this period were obtained using the Midwestern Regional Climate Center’s Application Tools Environment (cli-MATE) (<https://mrcc.illinois.edu/CLIMATE/>). However, not all stations used during the training periods had data readily available for the testing period on cli-MATE, because sometimes the observations at a station ended before or during the testing period. Thus, only a subset of training period stations for each city were used in the testing period.

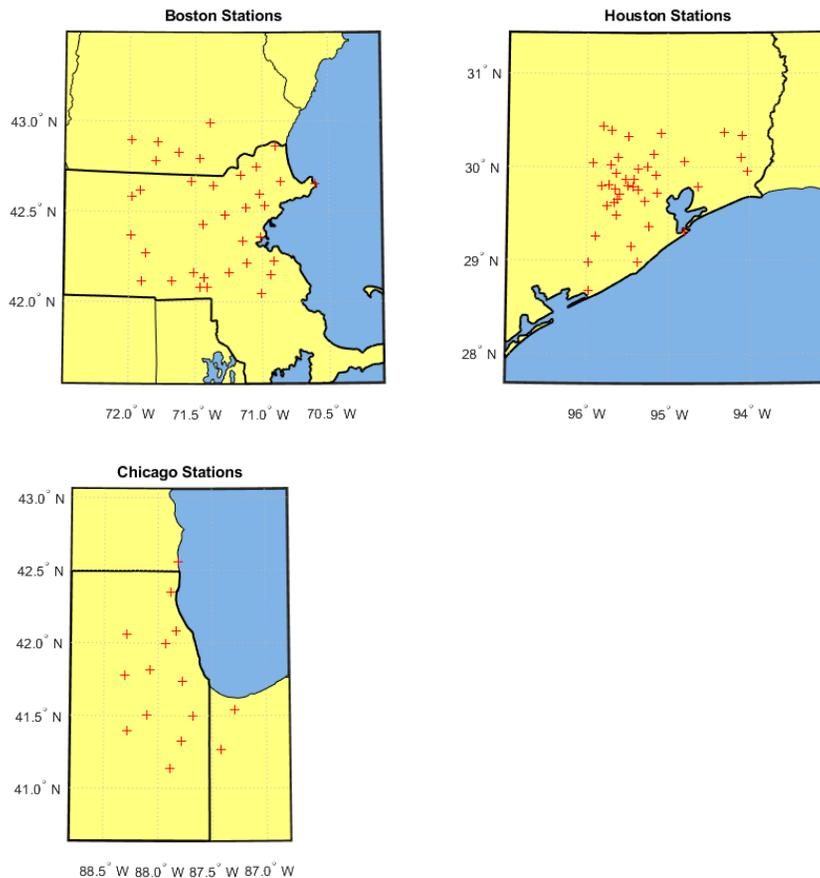
Finally, note that the testing period analysis for Houston was performed twice, once with 2017 data and once without it in both the models and observations. This is because of Hurricane Harvey, which impacted the Houston area in late August 2017, and which produced extremely high daily rainfall amounts well over 10 inches for many stations. Including such an extreme event in such a short time interval strongly influences the observed PFs, especially for longer return periods. This influence is not seen in the climate model data, leading to large disparities between the observed and modeled PFs. The results will show if and how much removing the 2017 data from consideration improves the model estimates.

Table I.1. The regions and time periods selected for this study

City	Boston, MA	Houston, TX	Chicago, IL
Coordinates of Observed Region	42-43°N × 70.5-72°W	28.5-30.5°N × 94-96°W	41-43°N × 87-88.5°W
Coordinates of Modeled Region	41-44°N × 70-73.2°W	27.8-31.6°N × 92.8-97.6°W	40-44°N × 86-90°W
# of Stations in Training Period	35	40	15
# of Stations in Testing Period	22	23	11
Training Period	1960-2005 (46 years)	1960-2005 (46 years)	1961-2000 (40 years)
Testing Period	2006-2018 (13 years)	2006-2018 (13 years); 2006-2016, 2018 (12 years)	2006-2018 (13 years)

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

The modeled data used were the raw UWPD data, consisting of 24 models (Table I.2). These models have different resolutions, with some having only a couple of grid points in each region. The closest model grid point was selected for comparison with each station for each model; however, these grid points were not always in the observed regions in Table I.1. Thus, slightly larger regions (the second row in Table I.1) were used for the models to ensure that the closest model grid point to each station in the observed region was included in the analysis. Finally, two models (CMCC-CESM and MRI-ESM1) did not have RCP4.5 data available, so two sets of weights for each city needed to be found: one without those models for RCP4.5 and one with them for RCP8.5.



APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Figure I.2. Location of selected training period stations for Boston (top left), Houston (top right), and Chicago (bottom left)

For all grid cells, due to their large size (greater than 400 square miles), the areal reduction factor (ARF) of 0.90, producing the inverse ARF equal to 1.11, was assumed based on the ARF curves in Hirschfield (1961). However, several later studies (Sivapalan and Blöschl 1998; Allen and DeGaetano 2005) indicated that the areal reduction factors in Hirschfield (1961) were too high. To accommodate the recommendations from these studies, two other ARF values (0.80 and 0.67) were also calculated to test the sensitivity of results based on the areal reduction factors.

Table I.2. UWPD models used in this study. No RCP4.5 data were available for models with an (*).

Data Set	Models Used
UWPD (24)	ACCESS1-0; ACCESS1-3; CanESM2; CMCC-CESM*; CMCC-CM; CMCC-CMS; CNRM-CM5; CSIRO-Mk3-6-0; GFDL-CM3; GFDL-ESM2G; GFDL-ESM2M; HadGEM2-CC; inmcm4; IPSL-CM5A-LR; IPSL-CM5A-MR; IPSL-CM5B-LR; MIROC5; MIROC-ESM; MIROC-ESM-CHEM; MPI-ESM-LR; MPI-ESM-MR; MRI-CGCM3; MRI- ESM1*; NorESM1-M

I.3. Methodology

The methodology of this study is depicted in Fig. I.3. We start with the training period for each city. For each station, the observed data are used to find the annual maximum series (AMS). Likewise, the AMS are found at each of the model grid points in the modeled region. From these AMSs, the PFs are found at each location using L-moments to fit the data to generalized extreme value (GEV) distributions. The return periods used are 2, 5, 10, 25, 50, and 100 years, which, after applying Langbein's (1949) formula for AMS, become 2.54, 5.52, 10.51, 25.50, 50.50, and 100 years, respectively. The observed PFs at each station are then compared with the modeled

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

PFs at the closest model grid point to each station for each model. Similar to Markus et al. (2018), a percent difference is found at each station for each model/return period combination:

$$\% \text{ Difference} = \frac{(\text{Model PF} - \text{Obs PF})}{\text{Obs PF}} \quad (1)$$

Once these differences are found, they are averaged for each model across all stations and return periods, resulting in a single value, d_i , for each of the 24 models. The model weights can be derived from this series of d_i , but here we deviate a bit from Markus et al. (2018) by using a variation of Tukey's (1977) formula:

$$w_i = \begin{cases} [1 - (\frac{d_i}{h})^3]^3, & \text{if } |d_i| \leq h \\ 0, & \text{if } |d_i| > h \end{cases} \quad (2)$$

where h is commonly the standard deviation of the set of d_i . Here we instead take it to be their standard deviation from 0, not their mean:

$$h = \sqrt{\frac{\sum_i d_i^2}{n}} \quad (3)$$

where n is the number of models used (here 24 for RCP8.5, 22 for RCP4.5). This was done because most of the series of d_i for each of the cities were large and negative, meaning that many of their absolute values were greater than their standard deviation and so, many, if not all, the weights would be 0. This is different from the more typical case for Tukey's formula where the values are more evenly distributed around 0. After finding the weights for each model using Eq. 2, they are then normalized so that their mean is 1.

After the weights are found, we then turn to the testing period. Observed and modeled PFs are found for this period in the same way as in the training period. Then, for each station and return period combination, the modeled PFs are averaged together across all 24 models (or 22 for RCP4.5), once using the weights found in the training period and once using equal weights. These averages can then be compared with their corresponding observed PFs to see whether using weights made a significant improvement overall from using equal weights.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

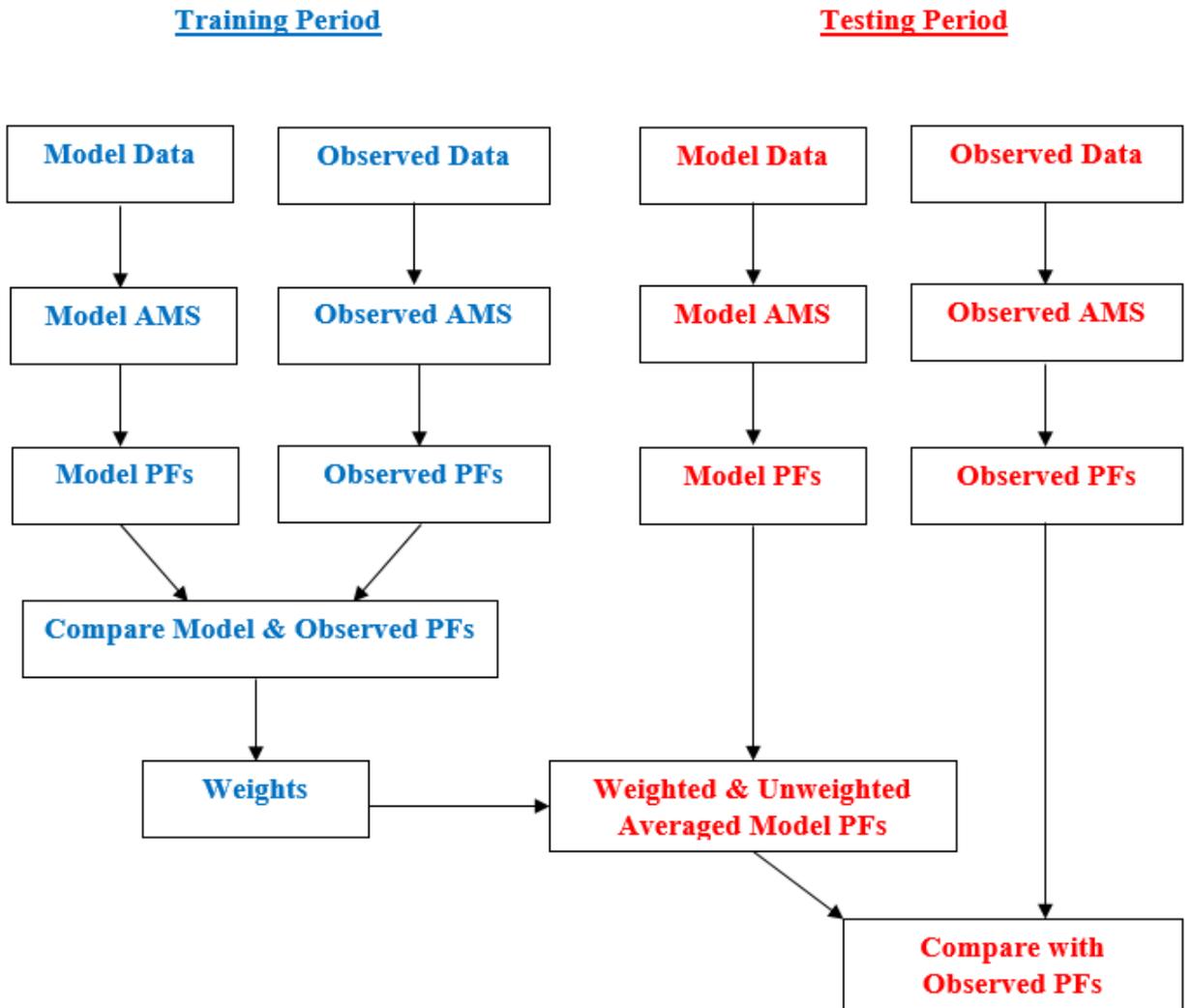


Figure I.3. Diagram of employed methodology

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

I.4. Results

Tables I.3-I.5 list the weights found for each model for each city for ARF=0.90, ARF=0.80, and ARF=0.67, respectively. All models generally showed a high degree of consistency across the three regions and for all three ARF values. Models showing high weights, regardless of ARF and location, included ACCESS1-0, ACCESS1-3, HadGEM2-CC, IPSL-CM5A-MR, IPSL-CM5B-LR, and MIROC5. Many models ended up with zero weights for each city and all ARF values (e.g., CMCC-CESM, inmcm4, and NorESM1-M), partly due to the large spread of the d_i and the large dry biases of some models. More years of data would have helped to a degree, but it is unlikely there would ever be a situation in which every model has a nonzero weight, due to the structure of Tukey's formula. One also must consider that the weights are from the training period, while the later testing period is expected to be wetter in each case.

Table I.6 shows the weighted and unweighted percent differences (Eq. 1) averaged across all return periods and stations for ARF=0.90 for Boston, Houston with the complete record, Houston without the hurricane year (2017), and Chicago. All results in Table I.6 are presented for two assumed climate scenarios (RCP4.5 and RCP8.5) and are applied to the testing period (2006-2018). For each scenario there is one value denoted as "average" for both the unweighted and weighted averages for each city. These values are averages for all 2-yr, 5-yr, 10-yr, 25-yr, 50-yr, and 100-yr frequencies. The negative values in Table I.6 indicate that the models on average underestimate the observed frequency estimates (dry bias). Small absolute values of the differences indicate that the models are accurate, and vice versa. In all cases, the percent differences for the weighted approach were smaller than those of the equal-weights approach, indicating that the weighted approach produced more accurate results than the unweighted one. The results show that all cases for most frequencies have a dry bias, except for the weighted RCP4.5 case for Boston. More importantly, the results indicate that in all cases the weighted ensemble produces a significant reduction in errors. The results for Boston show both the largest percent error reduction and highest accuracy, followed by Chicago and Houston without 2017, and Houston with 2017.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

The same statistical parameters for ARF=0.80 (Table I.7) and ARF=0.67 (Table I.8) showed similar results with a couple of exceptions, where the unweighted approach produced more accurate results. Importantly, when both approaches are inaccurate (the percent difference greater than 10-20%), the weighted approach is more accurate for all cases. When the models are accurate, but there are cases in which the unweighted approach is more accurate (shown in italics in Tables I.7-I.8), the weighted approach still produces a higher accuracy in most cases.

Table I.3. Weights for each city for both RCP4.5 and RCP8.5 using the entire historical period (1960-2005 for Boston and Houston, 1961-2000 for Chicago). Data for CMCC-CESM and MRI-ESM1 were not available for RCP4.5. ARF=0.90.

	Boston		Houston		Chicago	
	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5
ACCESS1-0	2.96	2.99	5.16	5.30	2.66	2.80
ACCESS1-3	2.00	2.06	6.62	6.81	2.00	2.11
CanESM2	2.24	2.29	0	0	1.30	1.38
CMCC-CESM	--	0	--	0	--	0
CMCC-CM	2.92	2.94	2.33	2.37	3.86	4.05
CMCC-CMS	0.69	0.76	0	0	0.81	0.87
CNRM-CM5	2.25	2.30	0.28	0.28	0.01	0.02
CSIRO-Mk3-6-0	0	0	0	0	0	0
GFDL-CM3	0	0	0	0	0	0
GFDL-ESM2G	0	0	0	0	0	0
GFDL-ESM2M	0	0	0	0	0	0
HadGEM2-CC	1.70	1.76	3.23	3.31	0.41	0.45
inmcm4	0	0	0	0	0	0
IPSL-CM5A-LR	0	0	0	0	0	0
IPSL-CM5A-MR	1.87	1.93	0.09	0.08	3.50	3.67
IPSL-CM5B-LR	2.95	2.97	3.81	3.91	5.55	5.78
MIROC5	1.53	1.60	0.05	0.04	1.68	1.79
MIROC-ESM	0	0	0	0	0	0
MIROC-ESM-CHEM	0	0	0	0	0	0
MPI-ESM-LR	0.01	0.03	0	0	0.02	0.03
MPI-ESM-MR	0	0	0	0	0.14	0.16
MRI-CGCM3	0.87	0.94	0.44	0.44	0.07	0.08
MRI-ESM1	--	1.44	--	1.46	--	0.81
NorESM1-M	0	0	0	0	0	0

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table I.4. Weights for each city for both RCP4.5 and RCP8.5 using the entire historical period (1960-2005 for Boston and Houston, 1961-2000 for Chicago). Data for CMCC-CESM and MRI-ESM1 were not available for RCP4.5. ARF=0.80.

	Boston		Houston		Chicago	
	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5
ACCESS1-0	1.57	1.59	4.96	5.05	2.82	2.91
ACCESS1-3	2.21	2.19	5.93	6.04	2.34	2.42
CanESM2	2.23	2.21	0	0	1.72	1.79
CMCC-CESM	--	0	--	0	--	0
CMCC-CM	2.06	2.05	2.59	2.61	3.45	3.56
CMCC-CMS	1.51	1.53	0	0	1.20	1.26
CNRM-CM5	2.23	2.21	0.39	0.38	0.06	0.07
CSIRO-Mk3-6-0	0	0	0	0	0	0
GFDL-CM3	0	0	0	0	0	0
GFDL-ESM2G	0	0	0	0	0	0
GFDL-ESM2M	0	0	0	0	0	0
HadGEM2-CC	2.14	2.12	3.42	3.47	0.70	0.74
inmcm4	0	0	0	0	0	0
IPSL-CM5A-LR	0	0	0	0	0.01	0.01
IPSL-CM5A-MR	2.18	2.17	0.13	0.13	3.30	3.40
IPSL-CM5B-LR	1.88	1.89	3.92	3.98	3.77	3.88
MIROC5	2.08	2.07	0.07	0.07	2.08	2.16
MIROC-ESM	0	0	0	0	0	0
MIROC-ESM-CHEM	0	0	0	0	0	0
MPI-ESM-LR	0.24	0.28	0	0	0.08	0.09
MPI-ESM-MR	0	0	0	0	0.30	0.33
MRI-CGCM3	1.67	1.68	0.58	0.57	0.18	0.19
MRI-ESM1	--	2.00	--	1.70	--	1.19
NorESM1-M	0	0	0	0	0	0

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table I.5. Weights for each city for both RCP4.5 and RCP8.5 using the entire historical period (1960-2005 for Boston and Houston, 1961-2000 for Chicago). Data for CMCC-CESM and MRI-ESM1 were not available for RCP4.5. ARF=0.67.

	Boston		Houston		Chicago	
	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5
ACCESS1-0	0	0	4.34	4.30	2.34	2.33
ACCESS1-3	0.43	0.44	4.50	4.46	2.34	2.33
CanESM2	0.07	0.08	0	0	2.30	2.28
CMCC-CESM	--	0	--	0	--	0
CMCC-CM	0	0	3.14	3.09	2.10	2.09
CMCC-CMS	3.29	3.27	0	0	2.15	2.14
CNRM-CM5	0.06	0.06	0.77	0.74	0.71	0.71
CSIRO-Mk3-6-0	0.59	0.60	0	0	0	0
GFDL-CM3	3.03	3.01	0	0	0	0
GFDL-ESM2G	0.86	0.87	0	0	0	0
GFDL-ESM2M	0.07	0.08	0	0	0	0
HadGEM2-CC	1.15	1.16	3.69	3.65	1.83	1.82
inmcm4	0	0	0	0	0	0
IPSL-CM5A-LR	2.77	2.76	0	0	0.41	0.41
IPSL-CM5A-MR	0.72	0.74	0.33	0.31	2.23	2.22
IPSL-CM5B-LR	0	0	3.96	3.91	0.07	0.07
MIROC5	1.59	1.60	0.21	0.19	2.33	2.32
MIROC-ESM	0	0	0	0	0	0
MIROC-ESM-CHEM	0	0	0	0	0	0
MPI-ESM-LR	3.65	3.63	0	0	0.78	0.78
MPI-ESM-MR	0.67	0.68	0	0	1.33	1.33
MRI-CGCM3	3.04	3.03	1.06	1.02	1.08	1.07
MRI-ESM1	--	1.99	--	2.32	--	2.11
NorESM1-M	0	0	0	0	0	0

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table I.6. Unweighted and weighted percent differences averaged across all return periods and stations for ARF=0.90

Boston	RCP4.5		RCP8.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	5.86%	-17.60%	-8.94%	-24.84%
2 yr	4.07%	-13.47%	-0.41%	-15.76%
5 yr	3.97%	-15.85%	-4.34%	-20.06%
10 yr	4.43%	-17.37%	-7.33%	-23.25%
25 yr	5.78%	-18.94%	-11.16%	-27.25%
50 yr	7.39%	-19.76%	-13.91%	-30.06%
100 yr	9.55%	-20.23%	-16.49%	-32.64%
Houston with 2017	RCP4.5		RCP4.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	-43.85%	-62.28%	-39.97%	-61.58%
2 yr	-26.31%	-51.17%	-21.32%	-48.45%
5 yr	-35.08%	-56.47%	-30.08%	-54.74%
10 yr	-41.24%	-60.40%	-36.77%	-59.42%
25 yr	-48.64%	-65.27%	-45.07%	-65.17%
50 yr	-53.66%	-68.65%	-50.75%	-69.09%
100 yr	-58.16%	-71.71%	-55.85%	-72.59%
Houston without 2017	RCP4.5		RCP4.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	-33.21%	-55.20%	-29.28%	-54.98%
2 yr	-22.60%	-49.10%	-18.02%	-46.44%
5 yr	-28.16%	-52.16%	-23.34%	-50.53%
10 yr	-31.84%	-54.30%	-27.38%	-53.55%
25 yr	-36.12%	-56.86%	-32.37%	-57.28%
50 yr	-38.99%	-58.61%	-35.78%	-59.87%
100 yr	-41.56%	-60.19%	-38.81%	-62.22%
Chicago	RCP4.5		RCP4.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	-20.11%	-38.46%	-19.60%	-35.41%
2 yr	-9.22%	-29.50%	-9.13%	-26.32%
5 yr	-13.69%	-33.55%	-13.27%	-30.46%
10 yr	-17.51%	-36.66%	-16.96%	-33.63%
25 yr	-22.77%	-40.73%	-22.15%	-37.74%
50 yr	-26.77%	-43.72%	-26.10%	-40.74%
100 yr	-30.68%	-46.57%	-29.97%	-43.60%

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table I.7. Unweighted and weighted percent differences averaged across all return periods and stations for ARF=0.80

Boston	RCP4.5		RCP8.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	13.66%	-7.21%	-0.27%	-15.36%
2 yr	14.13%	-2.56%	9.39%	-5.14%
5 yr	13.17%	-5.23%	5.00%	-9.97%
10 yr	12.87%	-6.94%	1.60%	-13.57%
25 yr	13.08%	-8.71%	-2.78%	-18.07%
50 yr	13.77%	-9.63%	-5.94%	-21.24%
100 yr	14.93%	-10.16%	-8.91%	-24.15%
Houston with 2017	RCP4.5		RCP4.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	-37.28%	-57.52%	-33.23%	-56.73%
2 yr	-18.03%	-45.01%	-12.73%	-41.94%
5 yr	-27.62%	-50.98%	-22.37%	-49.03%
10 yr	-34.38%	-55.41%	-29.71%	-54.31%
25 yr	-42.53%	-60.89%	-38.82%	-60.78%
50 yr	-48.06%	-64.69%	-45.07%	-65.19%
100 yr	-53.03%	-68.14%	-50.68%	-69.13%
Houston without 2017	RCP4.5		RCP4.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	-25.38%	-49.55%	-21.32%	-49.30%
2 yr	-13.92%	-42.68%	-9.06%	-39.69%
5 yr	-19.93%	-46.13%	-14.87%	-44.29%
10 yr	-23.89%	-48.53%	-19.27%	-47.69%
25 yr	-28.53%	-51.42%	-24.69%	-51.89%
50 yr	-31.63%	-53.39%	-28.37%	-54.80%
100 yr	-34.39%	-55.17%	-31.65%	-57.46%
Chicago	RCP4.5		RCP4.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	-12.71%	-30.69%	-12.16%	-27.27%
2 yr	-1.30%	-20.61%	-0.74%	-17.02%
5 yr	-6.00%	-25.17%	-5.42%	-21.69%
10 yr	-10.00%	-28.68%	-9.44%	-25.26%
25 yr	-15.51%	-33.26%	-14.97%	-29.89%
50 yr	-19.69%	-36.62%	-19.15%	-33.27%
100 yr	-23.79%	-39.83%	-23.21%	-36.49%

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table I.8. Unweighted and weighted percent differences averaged across all return periods and stations for ARF=0.67

Boston	RCP4.5		RCP8.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	5.96%	11.35%	0.11%	1.57%
2 yr	15.05%	16.93%	11.98%	13.83%
5 yr	10.52%	13.72%	6.48%	8.03%
10 yr	7.26%	11.67%	2.31%	3.72%
25 yr	3.38%	9.55%	-2.99%	-1.69%
50 yr	0.85%	8.44%	-6.79%	-5.48%
100 yr	-1.27%	7.80%	-10.33%	-8.98%
Houston with 2017	RCP4.5		RCP4.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	-26.42%	-49.02%	-22.50%	-48.08%
2 yr	-4.72%	-34.02%	0.81%	-30.33%
5 yr	-15.42%	-41.18%	-10.15%	-38.84%
10 yr	-23.05%	-46.49%	-18.49%	-45.17%
25 yr	-32.32%	-53.07%	-28.84%	-52.93%
50 yr	-38.65%	-57.63%	-35.96%	-58.23%
100 yr	-44.35%	-61.77%	-42.35%	-62.96%
Houston without 2017	RCP4.5		RCP4.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	-12.46%	-39.46%	-8.73%	-39.16%
2 yr	-0.06%	-31.22%	5.03%	-27.63%
5 yr	-6.52%	-35.36%	-1.55%	-33.15%
10 yr	-10.82%	-38.24%	-6.47%	-37.23%
25 yr	-15.86%	-41.70%	-12.51%	-42.27%
50 yr	-19.23%	-44.06%	-16.61%	-45.76%
100 yr	-22.24%	-46.20%	-20.25%	-48.95%
Chicago	RCP4.5		RCP4.5	
	Weighted	Unweighted	Weighted	Unweighted
Average	-4.15%	-16.83%	-2.30%	-12.72%
2 yr	7.58%	-4.74%	10.29%	-0.43%
5 yr	2.58%	-10.20%	4.58%	-6.02%
10 yr	-1.50%	-14.41%	0.19%	-10.31%
25 yr	-7.04%	-19.91%	-5.51%	-15.87%
50 yr	-11.22%	-23.94%	-9.68%	-19.92%
100 yr	-15.30%	-27.80%	-13.66%	-23.79%

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Another test has been performed to assess the effects of weights on the results. The standard deviations of the results based on all ensemble members of weighted vs. unweighted model results were calculated to determine if the weights can reduce model spread in future climate projections. The statistical quantity selected for this test was the percent reduction (PR), expressed as

$$PR (\%) = \left(\frac{SD_w}{SD_u} - 1 \right) \times 100 \quad (4)$$

where SD_w is the weighted standard deviation and SD_u is the unweighted standard deviation. A negative PR indicates that the weighted approach resulted in a smaller standard deviation, zero means that the standard deviations are equal, and a positive PR means that the weights did not reduce the standard deviation. The results for ARF=0.90, ARF=0.80, and ARF=0.67 are presented in Table I.9. The results show that the PR was negative, ranging between negative 61.6% and negative 6.7%, for all cases with the average of -37.0%, meaning that the weighted standard deviations are smaller than the unweighted ones.

Modeling uncertainties (structural, data, and scenario) were not explored herein. There are potential issues in ensemble analyses when the models are correlated (Knutti et al. 2010), but due to the weak correlations among extreme precipitation based on different models, it was assumed that the model results are independent. Also, the results can potentially be different for different weight functions. Additionally, the selection of the weight function in this study was arbitrary. Despite the simplifications, limitations, and uncertainties of the proposed approach, the hypothesis that the addition of weights will provide a more accurate average PF and smaller model variability was largely confirmed by the experimental results.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Table I.9. PR values (%) for all cases in the experiment

		Boston				Houston				Chicago			
		2006-2053		2054-2100		2006-2053		2054-2100		2006-2053		2054-2100	
Climate Scenario		RCP4.5	RCP8.5										
ARF=0.90	Return Period (years)												
	2	-41.0	-29.4	-37.7	-39.8	-54.1	-28.2	-41.8	-47.9	-21.2	-16.2	-25.6	-8.2
	5	-40.9	-32.6	-38.8	-33.6	-59.8	-35.2	-49.5	-55.8	-24.6	-20.2	-28.3	-15.2
	10	-39.5	-31.9	-38.1	-27.6	-61.6	-41.3	-53.5	-59.1	-27.0	-23.9	-29.5	-20.2
	25	-36.5	-28.4	-35.7	-19.1	-60.4	-48.6	-56.8	-57.9	-30.0	-29.1	-30.0	-26.8
	100	-33.6	-24.8	-32.9	-12.7	-57.2	-52.2	-57.9	-53.4	-32.1	-33.1	-29.6	-31.7
ARF=0.80	Return Period (years)												
	2	-21.7	-43.2	-22.1	-46.0	-42.0	-24.7	-32.6	-36.8	-43.3	-40.1	-36.7	-22.6
	5	-22.9	-44.6	-25.4	-47.5	-47.4	-30.9	-39.2	-44.0	-42.6	-41.1	-36.8	-25.9
	10	-23.9	-45.3	-27.8	-48.6	-49.4	-36.4	-43.0	-47.8	-41.6	-41.7	-35.6	-28.9
	25	-24.8	-45.0	-30.6	-48.9	-49.1	-42.6	-46.6	-48.7	-40.1	-42.6	-32.9	-33.2
	100	-25.1	-43.8	-32.4	-47.8	-47.0	-45.6	-48.3	-46.3	-38.9	-43.1	-30.3	-36.3
ARF=0.67	Return Period (years)												
	2	-41.6	-33.6	-37.8	-40.6	-51.3	-27.6	-39.9	-45.3	-24.1	-19.5	-26.2	-9.8
	5	-40.2	-34.8	-37.3	-35.0	-56.8	-34.3	-47.3	-52.9	-26.5	-22.7	-28.1	-15.7
	10	-38.3	-32.9	-36.1	-30.0	-58.5	-40.2	-51.1	-56.2	-28.1	-25.8	-28.5	-20.3
	25	-34.9	-28.3	-33.6	-23.1	-57.4	-47.0	-54.4	-55.6	-30.1	-30.3	-28.0	-26.5
	100	-31.9	-24.0	-31.1	-18.1	-54.5	-50.4	-55.6	-51.5	-31.4	-33.8	-26.9	-31.2
	100	-28.9	-19.5	-28.4	-13.5	-50.3	-51.7	-55.8	-45.6	-32.6	-37.1	-25.3	-35.6

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

I.5. Conclusions

This study indicates that the use of weights in climate ensembles can be beneficial. Weights helped bring the modeled averages closer to the observed values in the historical testing period. However, the study is limited in that it considers a small number of years in both the training and testing historical periods and a small number of stations and model grid points. More of all of these would help improve and strengthen the conclusions, especially more model grid points so that each grid point maps to only one station. Having multiple sets of raw data would also allow more models to have positive weights by finding the weights for each set and then averaging them together. Finally, the results are specific to one selected method for determining weights. Different methods would certainly result in different weights. Despite all the limitations and uncertainties, this study makes the case for continuing to investigate the use of weights and demonstrates that they can have value and could prove to be useful tools when studying the future climate.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

References

Allen, R. J. and DeGaetano, A. T., (2005). “Areal Reduction Factors for Two Eastern United States Regions with High Rain-Gauge Density.” *Journal of Hydraulic Engineering* 10 (4): 327-335.

Anderson, C. J., Arritt, R. W., and Kain, J. S. (2007). “An alternative mass flux profile in the Kain-Fritansch convective parameterization and its effects in seasonal precipitation.” *J. Hydrometeorol.*, 8, 1128–1140.

Ballentine, R. J., Stamm, A. J., Chermack, E. E., Byrd, G. P., and Schleede, D. (1998). “Mesoscale model simulation of the 4-5 January 1995 lake-effect snowstorm.” *Weather Forecasting*, 13, 893–920.

Brown, C., Greene, A. M., Block, P., and Giannini, A. (2008). “Review of downscaling methodologies for Africa climate applications.” IRI Technical Rep. 08-05, Columbia Univ., New York.

Christensen, J. H., and Christensen, O. B. (2007). “A summary of the PRUDENCE model projections of changes in European climate by the end of this century.” *Climatic Change*, 81, 7–30, <https://doi.org/10.1007/s10584-006-9210-7>.

Christensen, J. H., Kjellström, E., Giorgi, F., Lenderink, G., and Rummukainen, M. (2010). “Weight assignment in regional climate models.” *Clim. Res.*, 44(2-3), 179–194.

Douglas, E. M. and Fairbank, C. A., (2011). *Is Precipitation in Northern New England Becoming More Extreme? Statistical Analysis of Extreme Rainfall in Massachusetts, New Hampshire, and Maine and Updated Estimates of the 100-Year Storm.* *Journal of Hydrologic Engineering* 16 (3): 203-217

Easterling, D. R., Kunkel, K. E., Arnold, J. R., Knutson, T., LeGrande, A. N., Leung, L. R., Vose, R. S., Waliser, D. E., and Wehner, M. F. (2017). “Precipitation change in the United States.” In: *Climate Science Special Report: Fourth National Climate Assessment, Volume I* [Wuebbles, D. J., Fahey, D. W., Hibbard, K. A., Dokken, D. J., Stewart, B. C., and Maycock, T. K. (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, pp. 207–230, doi:10.7930/J0H993CC.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Feser, F., Rockel, B., von Storch, H., Winterfeldt, J., and Zahn, M. (2011). "Regional climate models add value to global model data: A review and selected examples." *Bull. Am. Meteorol. Soc.*, 92(9), 1181–1192.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M. (2013). "Evaluation of climate models." In: *Climate Change 2013: The Physical Science Basis. Contribution of working group I to the fifth assessment report of the Intergovernmental Panel on Climate Change* [Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C. (2007). "Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling." *Int. J. Climatol.*, 27(12), 1547–1578.

Gutmann, E., Pruitt, T., Clark, M.P., Brekke, L., Arnold, J.R., Raff, D.A., and Rasmussen, R.M., (2014). "An intercomparison of statistical downscaling methods used for water resource assessments in the United States," *Water Resources Research* 50, 50(9), 7167–7186.

Gutowski, W. J., et al. (2010). "Regional extreme monthly precipitation simulated by NARCCAP RCMs." *J. Hydrometeorol.*, 11(6), 1373–1379.

Guyennon, N. et al. (2013). "Benefits from using combined dynamical statistical downscaling approaches: Lessons from a case study in the Mediterranean region." *Hydrol. Earth Syst. Sci.*, 17(2), 705–720.

Hay, L. E., Clark, M. P., Pagowski, M., Leavesley, G. H., and Gutowski, W. J., Jr. (2006). "One-way coupling of an atmospheric and a hydrologic model in Colorado." *J. Hydrometeorol.*, 7(4), 569–589.

Hayhoe, K. A. (2010). "A standardized framework for evaluating the skill of regional climate downscaling techniques." Ph.D. dissertation, Univ. of Illinois at Urbana-Champaign, Champaign, IL, 153.

Hershfield, D. M. (1961). "Rainfall frequency atlas of the United States." Technical paper, 40.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

IPCC (Intergovernmental Panel on Climate Change). (2007). “Climate change 2007: The physical science basis.” Contribution of working group I to the fourth assessment report of the IPCC [Solomon, S. et al. (eds.)]. Cambridge University Press, Cambridge, U.K., 996.

Jones, R. G., Murphy, J. M., and Noguer, M. (1995). “Simulation of climate change over Europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries.” *Q. J. Royal Meteorol. Soc.*, 121(526), 1413–1449.

Kanamitsu, M., and DeHaan, L. (2011). “The added value index: A new metric to quantify the added value of regional models.” *J. Geophys. Res.*, 116(D11), 106.

Karl, T. R., Melillo, J. T., and Peterson, T. C. (eds.). (2009). “Global climate change impacts in the United States.” Cambridge University Press, 189 pp.

Kawazoe, S., and Gutowski, W. (2013). “Regional, very heavy daily precipitation in NARCCAP simulations.” *J. Hydrometeorol.*, 14(4), 1212–1227.

Knutti, R. (2010). “The end of model democracy?” *Clim. Change*, 102, 395–404.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A. (2010). “Challenges in combining projections from multiple climate models.” *J. Clim.*, 23(10), 2739–2758

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., and Eyring, V. (2017). “A climate model projection weighting scheme accounting for performance and interdependence.” *Geophys. Res. Lett.*, 44, 1909–1918.

Kotamarthi, R., Mearns, L., Hayhoe, K., Castro, C. L., and Wuebbles, D. (2016). “Use of climate information for decision-making and impacts research: State of our understanding.” Prepared for the Department of Defense, Strategic Environmental Research and Development Program. 55pp.

Kumar, D., Kodra, E., and Ganguly, A. R. (2014). “Regional and seasonal intercomparison of CMIP3 and CMIP5 climate model ensembles for temperature and precipitation.” *Clim. Dyn.*, 43(9–10), 2491–2518.

Kunkel, K. E., Moss, R., and Parris, A. (2016). “Innovations in science and scenarios for assessment.” *Clim. Change*, 135(1), 55–68.

Langbein, W.B. (1949). “Annual floods and the partial-duration flood series.” *Transactions, American Geophysical Union*, 30(6), 879–881, doi:10.1029/TR030i006p00879.

Leung, L. R., and Wigmosta, M. S. (1999). “Potential climate change impacts on mountain watersheds in the Pacific Northwest.” *J. Am. Water Resour. Assoc.*, 35(6), 1463–1471.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Li, Z., X. Li, Y. Wang, and S. M. Quiring. 2019. "Impact of Climate Change on Precipitation Patterns in Houston, Texas, USA." *Anthropocene* 25, ISSN 2213-3054. doi:10.1016/j.ancene.2019.100193.

Lorenz, D. J., Nieto-Lugilde, D., Blois, J. L., Fitzpatrick, M. C., and Williams, J. W. (2016). "Downscaled and debiased climate simulations for North America from 21,000 years ago to 2100 AD." *Scientific Data*, 3, 160048.

Lukas, J., Barsugli, J., Doesken, N., Rangwala, I., and Wolter, K. (2014). "Climate change in Colorado. A synthesis to support water resources management and adaptation." Technical Rep., Univ. of Colorado Boulder, Boulder, CO.

Madsen, H., Pearson, C. P., and Rosbjerg, D. (1997). "Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events regional modeling." *Water Resour. Res.*, 33, 759–769.

Manzanas, R., Gutiérrez, J. M., Fernández, J., van Meijgaard, E., Calmanti, S., Magariño, M. E., Cofiño, A. S., and Herrera, S. (2018). "Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: Added value for user applications." *Clim. Serv.*, 9, 44–56, <https://doi.org/10.1016/j.cliser.2017.06.004>.

Markus, M., Wuebbles, D.J., Liang, X.-Z., Hayhoe, K., Kristovich, D.A.R, (2012). Diagnostic analysis of future climate scenarios applied to urban flooding in the Chicago metropolitan area." *Climatic Change*, 111(3), 879–902.

Markus, M., Angel, J. R., Wang, K., Byard, G., McConkey, S., and Zaloudek, Z. (2017). Impacts of potential future climate change on the expected frequency of extreme rainfall events in Cook, DuPage, Lake, and Will Counties in Northeastern Illinois. Illinois State Water Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign, USA.

Markus, M., Angel, J. R., Byard, G., McConkey, S., Zhang, C., Cai, X., Notaro, M., and Ashfaq, M. (2018). "Communicating the impacts of projected climate change on heavy rainfall using a weighted ensemble approach." *J. Hydrol. Eng.*, 23(4): 04018804.

Masson, D., and Knutti, R. (2011). "Climate model genealogy." *Geophys. Res. Lett.*, 38(8), L08703.

Maurer, E. P., Hidalgo, H.G., Das, T., Dettinger, M.D., and Cayan, D.R., (2010). The utility of daily large-scale climate data in the assessment of climate change impacts on daily stream flow in California, *Hydrology and Earth System Sciences*, 14(6), 1125-1138.

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

Mearns, L. O. (2003). “Issues in the impacts of climate variability and change on agriculture: Applications to the southeastern United States.” *Clim. Change*, 60(1–2), 1–6.

Mearns, L. O., Giorgi, F., McDaniel, L., and Shields, C. (2003). “Climate scenarios for the southeastern U.S. based on GCM and regional model simulations.” *Clim. Change*, 60(1–2), 7–35.

Mo, K. C., Schemm, J.-K., Juang, H., and Higgins, R. H. (2005). “Impact of model resolution on the prediction of summer precipitation over the United States and Mexico.” *J. Clim.*, 18(18), 3910–3927.

Notaro, M., Bennington, V., and Vavrus, S. (2015). “Dynamically downscaled projections of lake-effect snow in the Great Lakes Basin.” *J. Clim.*, 28(4), 1660–1684.

Notaro, M., Lorenz, D. J., Hoving, C., and Schummer, M. (2014). “Twenty-first-century projections of snowfall and winter severity across central-eastern North America.” *J. Clim.*, 27(17), 6526–6550.

Perica, S., Pavlovic, S., Laurent, M. S., Trypaluk, C., Unruh, D., and Wilhite, O. (2018). “Precipitation-frequency atlas of the United States. Version 2.0: Texas.” National Weather Service, Silver Spring, MD, USA.

Perica, S., Pavlovic, S., Laurent, M. S., Trypaluk, C., Unruh, D., Martin, D., and Wilhite, O. (2019). “Precipitation-frequency atlas of the United States. Version 3.0: Northeastern States.” National Weather Service, Silver Spring, MD, USA.

Sivapalan, M. and Blöschl, G., (1998). “Transformation of Point Rainfall to Areal Rainfall: Intensity-Duration-Frequency Curves.” *Journal of Hydrology* 204: 150-167.

Räisänen, J., and Ylhäisi, J. S. (2012). “Can model weighting improve probabilistic projections of climate change?” *Clim. Dyn.*, 39(7–8), 1981–1998.

Roads, J. O., Chen, S.-C., and Kanamitsu, M. (2003). “U.S. regional climate simulations and seasonal forecasts.” *J. Geophys. Res.*, 108(D16), 8606.

Sánchez, E., Romera, R., Gaertner, M. A., Gallardo, C., and Castro, M. (2009). “A weighting proposal for an ensemble of regional climate models over Europe driven by 1961-2000 ERA40 based on monthly precipitation probability density functions.” *Atmos. Sci. Lett.*, 10(4), 241–248.

Sanderson, B. M., and Wehner, M. F. (2017). “Model weighting strategy.” In: *Climate Science Special Report: Fourth National Climate Assessment, Volume I* [Wuebbles, D. J., Fahey,

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

D. W., Hibbard, K. A., Dokken, D. J., Stewart, B. C., and Maycock, T. K. (eds.]. U.S. Global Change Research Program, Washington, D.C., USA, 436–442.

Schoof, J. T., Shin, D. W., Cocke, S., LaRow, T. E., Lim, Y.-K., and O'Brien, J. J. (2009). “Dynamically and statistically downscaled seasonal temperature and precipitation hindcast ensembles for the southeastern USA.” *Int. J. Climatol.*, 29(2), 243–257.

Spak, S., Holloway, T., Lynn, B., and Goldberg, R. (2007). “A comparison of statistical and dynamical downscaling for surface temperature in North America.” *J. Geophys. Res.*, 112(D8), 1–37.

Tukey, J. W. (1977). “Exploratory data analysis.” Addison Wesley, Reading, MA.

Um, M.-J., Kim, Y., Markus, M., and Wuebbles, D. J. (2017). “Modeling nonstationary extreme value distributions with nonlinear functions: An application using multiple CMIP5 precipitation projections for U.S. cities.” *Journal of Hydrology*, 552, 396–406.

Um, M.-J., Heo, J.-H., Markus, M., and Wuebbles, D. J. (2018). “Performance evaluation of four statistical tests for trend and non-stationarity and assessment of observed and projected annual maximum precipitation series in major United States cities.” *Water Resources Management*, 32(3), 913–933.

USGCRP. (2017). “Climate science special report: Fourth national climate assessment, volume I” [Wuebbles, D. J., Fahey, D. W., Hibbard, K. A., Dokken, D. J., Stewart, B. C., and Maycock, T. K. (eds.]. U.S. Global Change Research Program, Washington, D.C., USA, 470pp.

Weigel, A. P., Knutti, R., Liniger, M. A., and Appenzeller, C., (2010). “Risks of model weighting in multimodel climate projections.” *J. Clim.* 23 (15): 4175-4191

Winters, B. A., J. R. Angel, C. Ballerine, J. Byard, A. Flegel, D. Gambill, E. Jenkins, S. McConkey, M. Markus, B. A. Bender, M. J. O’Toole. 2015. Report for the Urban Flooding Awareness Act. Springfield, IL: Illinois Department of Natural Resources.

<http://hdl.handle.net/2142/78150>

Wu, S., Markus, M., Lorenz, D. J., Angel, J. R., and Grady, K. (2019). “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, 1279.

Wuebbles, D. J., Fahey, D. W., Hibbard, K. A., DeAngelo, B., Doherty, S., Hayhoe, K., Horton, R., Kossin, J. P., Taylor, P. C., Waple, A. M., and Weaver, C. P., (2017). Executive Summary. In: *Climate Science Special Report: Fourth National Climate Assessment, Volume I*,

APPENDIX A-2. The University of Illinois Urbana-Champaign and the University of Wisconsin-Madison Final Report

edited by D. J. Wuebbles, D. W. Fahey, K. A. Hibbard, D. J. Dokken, B. C. Stewart, and T. K. Maycock. 12-34. Washington, DC: U.S. Global Change Research Program. doi: 10.7930/J0DJ5CTG

Yoon, J. H., Leung, L. R., and Correia, J. (2012). “Comparison of dynamically and statistically downscaled seasonal climate forecasts for the cold season over the United States.” *J. Geophys. Res.*, 117(D21), D21109.

Zhang, W., Villarini, G., Vecchi, G. A., and Smith, J. A., (2018). “Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston. *Nature* 563: 384-388.