U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

NUREG/CR-7247

Research to Develop Guidance on Extreme Precipitation Frequency Estimates in Orographic Regions

Office of Nuclear Regulatory Research

AVAILABILITY OF REFERENCE MATERIALS IN NRC PUBLICATIONS

NRC Reference Material

As of November 1999, you may electronically access NUREG-series publications and other NRC records at the NRC's Library at www.nrc.gov/reading-rm.html. Publicly released records include, to name a few, NUREG-series publications; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigative reports; licensee event reports; and Commission papers and their attachments.

NRC publications in the NUREG series, NRC regulations, and Title 10, "Energy," in the *Code of Federal Regulations* may also be purchased from one of these two sources:

1. **The Superintendent of Documents**
   U.S. Government Publishing Office
   Washington, DC 20402-0001
   Internet: https://bookstore.gpo.gov/
   Telephone: (202) 512-1800
   Fax: (202) 512-2104

2. **The National Technical Information Service**
   5301 Shawnee Road
   Alexandria, VA 22312-0002
   Internet: https://www.ntis.gov/
   1-800-553-6847 or, locally, (703) 605-6000

A single copy of each NRC draft report for comment is available free, to the extent of supply, upon written request as follows:

Address: **U.S. Nuclear Regulatory Commission**
Office of Administration
Program Management and Design Service Branch
Washington, DC 20555-0001
E-mail: Reproduction.Resource@nrc.gov
Facsimile: (301) 415-2289

Some publications in the NUREG series that are posted at the NRC's Web site address www.nrc.gov/reading-rm/doc-collections/nuregs are updated periodically and may differ from the last printed version. Although references to material found on a Web site bear the date the material was accessed, the material available on the date cited may subsequently be removed from the site.

Non-NRC Reference Material

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, transactions, *Federal Register* notices, Federal and State legislation, and congressional reports. Such documents as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings may be purchased from their sponsoring organization.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at—

The NRC Technical Library
Two White Flint North
11545 Rockville Pike
Rockville, MD 20852-2738

These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—

American National Standards Institute
11 West 42nd Street
New York, NY 10036-8002
Internet: https://www.ansi.org/
(212) 642-4900

Legally binding regulatory requirements are stated only in laws; NRC regulations; licenses, including technical specifications; or orders, not in NUREG-series publications. The views expressed in contractor prepared publications in this series are not necessarily those of the NRC.

The NUREG series comprises (1) technical and administrative reports and books prepared by the staff (NUREG–XXXX) or agency contractors (NUREG/CR–XXXX), (2) proceedings of conferences (NUREG/CP–XXXX), (3) reports resulting from international agreements (NUREG/IA–XXXX),(4) brochures (NUREG/BR–XXXX), and (5) compilations of legal decisions and orders of the Commission and the Atomic and Safety Licensing Boards and of Directors' decisions under Section 2.206 of the NRC's regulations (NUREG-0750), (6) Knowledge Management prepared by NRC staff or agency contractors (NUREG/KM-XXXX).

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

NUREG/CR-7247

Research to Develop Guidance on Extreme Precipitation Frequency Estimates in Orographic Regions

Manuscript Completed: September 2025
Date Published: March 2026

Prepared by:
K. D. Holman
A. P. Verdin
D. P. Keeney

Bureau of Reclamation
Technical Service Center
Denver Federal Center
PO Box 25007
Denver, CO 80225-0007

Joseph Kanney, NRC Project Manager

Office of Nuclear Regulatory Research

ABSTRACT

Currently, many engineering design projects use Probable Maximum Precipitation (PMP) to develop Probable Maximum Floods (PMFs), which are then used to evaluate the safety of a proposed or existing facility. Orographic methods employed in hydrometeorological reports (HMRs) used to estimate PMP vary widely through time and by region, if they were employed at all. Deterministic metrics, such as PMP, do not provide decision makers with information on precipitation events less intense than PMP, yet still extreme, nor do they provide information on the expected frequency of such events. The Nuclear Regulatory Commission (NRC) requested assistance from the Department of Interior's Bureau of Reclamation (Reclamation) in developing guidance on regional extreme precipitation analyses in orographic regions. The methods presented in this report are illustrated across the Tennessee River Valley watershed. The Tennessee River Valley is a region with pronounced orographics, and thus serves as a suitable testbed for the methodologies outlined in this report.

Probabilistic precipitation information often takes the form of a precipitation-frequency relationship, which describes the depth of precipitation and the associated probability of occurrence. NOAA's Hydrometeorological Design Studies Center produces precipitation-frequency estimates across most of the United States. These estimates are published as individual regional volumes of NOAA Atlas 14 (e.g., Bonnin et al. 2006), and include precipitation-frequency estimates as rare as the 1,000-year return period, which corresponds to an annual exceedance probability (AEP) of 1/1,000 or 0.001. However, high-hazard dam safety or nuclear facility applications frequently require precipitation-frequency estimates beyond the 1,000 year return period. This research provides information useful to license applicants on acceptable methods and data sources for estimating and using precipitation-frequency analyses in orographic regions for calculating floods at critical AEPs that the applicants need to consider in evaluating siting factors, conducting Probabilistic Risk Assessments (PRAs), and in designing facilities.

The two main goals of the research were (i) to critically review orographic extreme storm methodologies used in PMP and magnitudes less than PMP in orographic regions; and (ii) to evaluate methodologies for developing precipitation-frequency estimates at AEPs significantly less than those offered in NOAA Atlas 14, in the range of 10^{-4} to 10^{-6} AEP. To meet the objectives and goals, Reclamation performed the following five major tasks:

1. Review extreme storm precipitation techniques, precipitation-frequency methods, and databases in orographic regions;
2. Develop a methodology to estimate precipitation-frequency analyses in regions of complex topography;
3. Demonstrate the precipitation-frequency methodology and provide uncertainties and confidence intervals at the regional and reactor-site scale for a pilot region in the Tennessee River Valley;
4. Transfer technology to the NRC staff via a training session, including data, software, and scripts; and
5. Complete a final report that conveys research findings.

The five tasks outlined above were accomplished through the use of existing and new technologies available within the scientific community. Task 1 was accomplished by reviewing the variety of techniques used in orographic storm analyses, including methods from the HMRs, private consultants, and additional methods from the scientific community. Along with the

HMRs, we reviewed previous precipitation-frequency data and methods as part of Task 1. We reviewed federal approaches to precipitation-frequency analyses (e.g., NOAA Atlas 14), as well as approaches used by private consultants and the scientific community. We focused primarily on a discussion of regional L-moments and Bayesian inference.

Tasks 2 and 3 included developing and demonstrating a methodology to estimate precipitation-frequency analyses in a region of complex topography. We accomplished these tasks by first combining a known objective clustering algorithm, the Self-Organizing Map (SOM), with two different regional frequency methods, L-moments and Bayesian inference. These two regional frequency methods vary widely in terms of the level of complexity (and consequently effort) and in the way in which epistemic uncertainty is estimated. The SOM algorithm used a combination of geophysical information (i.e., latitude, longitude, elevation) and observed precipitation data (i.e., mean annual precipitation, mean one-day annual maxima) to identify 14 groups (i.e., homogeneous regions) across the Tennessee River Valley watershed. The two regional frequency methods were applied to historical precipitation observations from these 14 groups located across the Tennessee River Valley watershed. Both analyses developed precipitation-frequency estimates using the generalized extreme value (GEV) distribution, a commonly used distribution. Epistemic uncertainty due to probability distribution choice was not the focus of the current research. Results suggest that the SOM algorithm is a useful tool for identifying and grouping similarly-behaved point precipitation observations. Furthermore, the frequency results from these analyses indicate that uncertainty estimates from the L-moments analysis are consistently less than the uncertainty estimates from Bayesian inference. These differences are the result of estimating uncertainty differently between the two methods. L-moments uncertainty estimates are obtained using an ad hoc bootstrap resampling routine; Bayesian inference more properly accounts for uncertainty via Monte Carlo methods, which result in posterior distributions of all model parameters.

In addition to demonstrating these methods on historical point precipitation observations, we also demonstrated the application of the two regional frequency methods on a gridded precipitation dataset, the 100-member ensemble data from Newman et al. (2015). The application of these two regional frequency methods to a gridded precipitation dataset demonstrates the utility of these methods to an alternative data format. With some modifications, the techniques developed here can be applied to any gridded or point-based precipitation dataset of interest to the users.

Task 4, transfer technology to NRC staff, was accomplished through a meeting between Reclamation personnel and NRC staff that was held at the NRC headquarters in Rockville, MD, on August 1-2, 2017. The event included two days of lectures, with allocated time for questions and discussion. An external hard drive, containing project data, software, and scripts, will be sent to NRC following project closeout. Finally, Task 5 was completed through this report, which conveys all the research findings from the Reclamation team.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

PRIMARY AUDIENCE: NRC hydrologists and engineers involved in developing regulatory guidance on application of probabilistic flood hazard assessment (PFHA) methods to the design of nuclear facilities, such as nuclear power plants (NPPs).

SECONDARY AUDIENCE: Hydrologists, engineers, reviewers, utility managers, and other stakeholders who conduct, review, or manage flood analyses and need to understand the underlying technical basis for developing probabilistic precipitation estimates in orographic regions.

KEY RESEARCH QUESTION

The purpose of this research was to investigate methods for estimating extreme precipitation in orographic regions to provide a technical basis for developing guidance on acceptable data sources and methods.

The main research questions were to investigate the following aspects of extreme precipitation:

1. Critically review orographic extreme storm methodologies used in Probable Maximum Precipitation (PMP) studies and magnitudes less than PMP in orographic regions.
2. Evaluate methods for developing precipitation-frequency estimates at annual exceedance probabilities (AEPs) in the range of 1E-4 to 1E-6 per year, and the application of methods to regions with significant orographic influence.

RESEARCH OVERVIEW

Current NRC guidance for estimating extreme precipitation is based on the deterministic concept of probable maximum precipitation (PMP). While PMP is a metric that has and continues to be used in engineering design projects across the U.S. and beyond, deterministic metrics such as PMP do not provide decision makers with information on the expected frequency of PMP-level events, nor do they provide information on the likelihood of precipitation events less intense than PMP. In addition, significant advancements in probabilistic methods have occurred during the past few decades. Consequently, some agencies, such as the Bureau of Reclamation and the U.S. Army Corps of Engineers, have started using probabilistic precipitation information for risk-based and risk-informed decision making in dam safety applications.

Probabilistic precipitation information most often takes the form of a precipitation-frequency relationship, which describes a precipitation amount associated with an “average” return period. The National Oceanic and Atmospheric Administration’s Hydrologic Design Study Center (NOAA’s HDSC) produces precipitation-frequency estimates across most of the United States. These estimates are published as individual regional volumes of NOAA Atlas 14 and include frequency estimates at annual exceedance probabilities (AEPs) as low as 1E-3. However, applications such as nuclear facility design or high-hazard dam safety often require estimates at lower AEPs (1E-4 to 1E-6).

This report describes research performed to: (1) Critically review orographic extreme storm methodologies used in Probable Maximum Precipitation (PMP) studies and magnitudes less than PMP in orographic regions; and (2) evaluate methods for developing precipitation-frequency estimates at AEPs in the range of 1E-4 to 1E-6. The Tennessee River Valley

watershed was selected as an example of a watershed with significant orographic influences on precipitation.

KEY FINDINGS: Key findings from this research include:

- The review of PMP methods reconfirmed that such deterministic methods are not well-suited to risk-informed hazard assessment.

- The review of orographic storm analysis techniques used in PMP studies (especially the historical HMR reports, but also more recent site-specific PMP studies), reveals that PMP techniques have not kept pace with major orographic storm analysis advancements within the scientific community, particularly advances in numerical simulation approaches.

- A review of data sources for precipitation frequency analysis found that there currently exists a plethora of precipitation datasets from precipitation gauge networks, in-situ observing platforms, gridded techniques applied to point observations, and reanalysis products.

- A review of available methods for precipitation frequency analysis found that the regional L-moments approach has been used in conjunction with a variety of variables, including precipitation, snow water equivalent, and streamflow. The L-moments approach has also been used outside of point observations, namely several gridded datasets. Bayesian inference methods have been used for a wide variety of purposes, including biostatistics and medicine, real estate, paleoflood reconstructions, and spatial precipitation extremes, to name a few. These two frequency methods vary widely in terms of the level of complexity (and consequently effort) and in the estimation of uncertainty. The Bayesian approach uses an iterative Monte Carlo sampling scheme to estimate uncertainty in distribution parameters, while the approach used to estimate uncertainty with L-moments is based on systematically removing a fixed number of observations, estimating the distribution parameters with the remaining observations, and repeating the process.

- The Self-Organizing Map (SOM) algorithm for identifying homogenous precipitation regions (a key concept in regional frequency analysis) was investigated using historical point precipitation observations (gauge data) from across the Tennessee River Valley watershed. Results suggest that the SOM algorithm is a useful tool for identifying and grouping similarly behaved point precipitation observations.

- The SOM algorithm was used with two different regional frequency methods (L-moments and Bayesian inference) using the same historical precipitation observations from across the Tennessee River Valley watershed. In this work, both analyses assumed the data are best described by a GEV distribution (uncertainty due to distribution choice was not a focus of this work). The uncertainty estimates from the L-moments analysis were consistently less than the uncertainty estimates from Bayesian inference.

- In addition to demonstrating the regional frequency methods on historical point precipitation observations, the application of the two regional frequency methods on a gridded precipitation dataset was demonstrated using the 100-member precipitation ensemble from Newman et al. (2015). The application of these two regional frequency methods to a

gridded precipitation dataset demonstrates the usefulness of these methods to a variety of data formats and structures. In addition, use of the ensemble allowed some extension of the effective record and for inclusion of modeled uncertainty in the precipitation observations.

WHY THIS MATTERS

Probabilistic flood hazard assessment (PFHA) methods are being investigated to risk-inform flood hazard analyses for nuclear facilities, including nuclear power plants (NPPs). In contrast to deterministic approaches, PFHA can provide best estimates of magnitudes and uncertainty for floods at critical annual exceedance probabilities (AEPs) that NPPs need to consider in evaluating siting factors, designing their facilities, and conducting probabilistic risk assessments (PRAs). Precipitation frequency analysis is one approach to developing the precipitation component of a probabilistic flood hazard assessment.

HOW TO APPLY RESULTS

This work sought to demonstrate use of selected regional frequency analysis methods, along with the SOM algorithm, in a region with significant orographic influence on precipitation. With some modifications, the techniques developed here can be applied to any gridded or point-based precipitation dataset of interest to the users.

For simplicity, this work assumed that the precipitation data are best described by a GEV distribution. In actual applications, uncertainty in the probability distribution choice, as well as other sources of epistemic uncertainty, should be considered.

LEARNING AND ENGAGEMENT OPPORTUNITIES

Users of this report may be interested in the proceedings of the NRC-RES Annual Probabilistic Flood Hazard Assessment Workshops, which are published as NRC Research Information Letters.

ABBREVIATIONS

AEP	Annual Exceedance Probability
AR	Atmospheric River
ASOS	Automated Surface Observing System
AWA	Applied Weather Associates
BOF	Broadscale Orographic Factor
CCPA	Climatology-Calibrated Precipitation Analysis
CFSR	Climate Forecast System Reanalysis
CHIRP	Climate Hazards group InfraRed Precipitation
CI	Credible Interval
CMORPH	CPC MORPHing
COOP	Cooperative Observer Program
CPC	Climate Prediction Center
CRAB	Cascade, Residual Add-Back
DEM	Digital Elevation Model
ECMWF	European Centre for Medium-Range Weather Forecasts
EOF	Empirical Orthogonal Function
FAFP	Free Atmospheric Forced Precipitation
GEFS	Global Ensemble Forecast System
GEV	Generalized Extreme Value
GHCN	Global Historical Climatology Network
GLO	Generalized Logistic
GPCP	Global Precipitation Climatology Project
HDSC	Hydrologic Design Study Center
HMR	Hydrometeorological Report
HR	Homogeneous Region
IDF	Inflow Design Flood
IDW	Inverse Distance Weighting
IMW	Inter-Mountain West
K-S	Kolmogorov-Smirnov
L-CV	Coefficient of L-variation
LP	Local polynomials
MCC	Mesoscale Convective Complex
MERRA	Modern-Era Retrospective analysis for Research and Applications
MLE	Maximum Likelihood Estimate
MODIS	Moderate Resolution Imaging Spectroradiometer
MPE	Multi-sensor Precipitation Estimates
NA14	NOAA Atlas 14
NAM	North American Mesoscale model
NARR	North American Regional Reanalysis
NASA	National Aeronautics and Space Administration
NCEP	National Centers for Environmental Prediction
NGM	Nested Grid Model
LAM	Limited Area Model
NASA	National Aeronautics and Space Administration
NLDAS	National Land Data Assimilation System
NOAA	National Oceanic and Atmospheric Administration
NPP	Nuclear Power Plant
NRL	Naval Research Laboratory

NWS	National Weather Service
OI	Optimal Interpolation
OTF	Orographic Transposition Factor
PAM	Partitioning around medoids
P-F	Curve Precipitation-Frequency Curve
PFA	Precipitation-Frequency Analysis
PMF	Probable Maximum Flood
PMP	Probable Maximum Precipitation
PQPE	Probabilistic Quantitative Precipitation Estimate
PRISM	Parameter-elevation Regressions on Independent Slopes Model
PWM	Probability Weighted Moments
QA	Quality Assurance
RAMOS	Remote Automated Meteorological Observing System
RCM	Regional Climate Model
RFA	Regional Frequency Analysis
SE	Southeast, U.S.
SOM	Self Organizing Map
SPAS	Storm Precipitation Analysis System
SSM	Storm Separation Method
SSM/I	Special Sensor Microwave/Imager
SST	Sea Surface Temperature
SYMAP	Synagraphic Mapping System
TAF	Total Adjustment Factor
TMPA	TRMM Multisatellite Precipitation Analysis
TP40	Technical Paper 40
TP 42	Technical Paper 42
TRMM	Tropical Rainfall Measuring Mission
TRVW	Tennessee River Valley Watershed
TSF	Terrain Stimulation Factor
U.S.	United States
USWB	United States Weather Bureau
WMO	World Meteorological Organization
WRF	Weather Research and Forecasting Model
WWII	World War II

1 INTRODUCTION

1.1 Background

The Nuclear Regulatory Commission (NRC) has requested research assistance in developing regional extreme precipitation-frequency analyses with associated uncertainty in orographic regions, using a region of the Tennessee River Valley including parts of Tennessee, Kentucky, West Virginia, Virginia, North Carolina, South Carolina, Georgia, and Alabama to evaluate and compare approaches. NRC General Design Criterion 2, "Design Bases for Protection Against Natural Phenomena," of Appendix A, "General Design Criteria for Nuclear Power Plants," to 10 CFR Part 50, "Domestic Licensing of Production and Utilization Facilities," requires, in part, that structures, systems, and components important to safety be designed to withstand the effects of natural phenomena such as floods, tsunami, and seiches, without loss of capability to perform their safety functions. Criterion 2 also requires that design bases for these structures, systems, and components reflect: (1) appropriate consideration of the most severe of the natural phenomena that have been historically reported for the site and surrounding region, with sufficient margin for the limited accuracy and quantity of the historical data and the period of time in which the data have been accumulated; (2) appropriate combinations of the effects of normal and accident conditions with the effects of the natural phenomena; and (3) the importance of the safety functions to be performed. Paragraphs 100.10 (c) and 100.20 (c) of 10 CFR Part 100, "Factors to be considered when evaluating sites," require that physical characteristics of the site, including seismology, meteorology, geology, and hydrology, be taken into account in determining the acceptability of a site for a nuclear power reactor. Paragraph 100.20 (c) (2) states that meteorological characteristics of the site that are necessary for safety analysis or that may have an impact upon plant design (such as maximum probable wind speed and precipitation) must be identified and characterized. Paragraph 100.20 (c) (3) states that "The maximum probable flood along with the potential for seismically induced floods discussed in 100.23 (d) (3) must be estimated using historical data." Paragraph 100.23 (d) which focuses on "Geologic and seismic siting criteria" requires investigations and detailed study of seismically induced floods and water waves. Applicants for nuclear reactor combined operating licenses (COL) must demonstrate the ability of their proposed facilities to withstand the Probable Maximum Flood (PMF), among other hazards. The demonstration is scrutinized by the NRC internally and through a public review process. Some further details on NRC design criteria are described in NRC (1977).

Currently, Probable Maximum Precipitation (PMP) is one key factor in developing PMFs. The National Oceanic and Atmospheric Administration's (NOAA's) National Weather Service (NWS) has published a series of Hydrometeorological Reports (HMRs), which are used by many Federal agencies, including the NRC, for estimating PMP in the United States (U.S.). For those regions of the U.S. where most operating reactors are located and where most new reactors are expected to be located (east of 105 degrees west longitude), the data and information that form the technical basis for these reports (HMR 51) have not been updated for over thirty years. In addition, there is a large region along the spine of the Appalachian Mountains where orographic effects may lead to deficiencies in PMP estimates.

While PMP is a metric that has and continues to be used in engineering design projects across the U.S. and beyond, advancements in probabilistic methods have occurred during the past few decades. Deterministic metrics, such as PMP, do not provide decision makers with information on precipitation events less intense than PMP, nor do they provide information on the expected frequency of PMP-level events. Consequently, some agencies, such as the Bureau of

Reclamation and the U.S. Army Corps of Engineers, have started using probabilistic precipitation information for risk-based and risk-informed decision making. Probabilistic precipitation information most often takes the form of a precipitation-frequency relationship, which describes a precipitation amount associated with an "average" return period. NOAA's Hydrologic Design Study Center (HDSC) produces precipitation-frequency estimates across most of the U.S. These estimates are published as individual regional volumes of NOAA Atlas 14 (e.g., Bonnin et al. 2006), and include frequency estimates up to 1E-3 (i.e., 1/1,000) annual exceedance probabilities (AEPs). However, currently published precipitation-frequency estimates from NOAA Atlas 14 are not yet suitable for high-hazard dam safety or nuclear facility applications, because very low AEPs (< 1/1,000) with uncertainty are needed. Although NOAA Atlas 14 includes estimates of aleatory variability via probability distributions, there has been no attempt to comprehensively quantify uncertainty (epistemic uncertainties are not characterized or quantified).

1.2 Objectives

This research provides a technical basis for developing guidance on acceptable data sources and methods for estimating precipitation-frequency in orographic regions. Precipitation-frequency analyses can be used to estimate floods at critical AEPs that nuclear power plants (NPPs) need to consider in evaluating siting factors, conducting Probabilistic Risk Assessments (PRAs), and in designing their facilities. An additional objective is to transfer findings and provide training to NRC staff, so that NRC can build internal capacity to perform and/or review extreme storm analyses.

The main goals of the research were to investigate the following aspects of extreme precipitation:

1. Critically review orographic extreme storm methodologies used in PMP and magnitudes less than PMP in orographic regions;
2. Evaluate methods for developing precipitation-frequency estimates at AEPs significantly less than those offered in NOAA Atlas 14, in the range of 10^{-4} to 10^{-6} AEP.

To meet the objectives and goals, Reclamation performed the following five major tasks:

1. Review extreme storm precipitation techniques, precipitation-frequency methods and databases in orographic regions;
2. Develop a methodology to estimate precipitation-frequency analyses in regions of complex topography;
3. Demonstrate the precipitation-frequency methodology and provide uncertainties and confidence intervals at the regional and reactor-site scale for a pilot region in the Tennessee River Valley;
4. Transfer technology to the NRC staff via a training session, including data and software scripts; and
5. Complete a final report that conveys research findings.

The tasks outlined above are accomplished using existing and new technologies available within the scientific community. We review historical approaches to storm analyses in section 2, including a review of HMRs and other PMP-related studies. In section 3, we discuss recent advancements to storm (orographic) analyses, such as data improvements and advances to numerical and statistical methods. We present and apply two methods for developing precipitation-frequency analyses using historical precipitation observations across the

Tennessee River Valley in section 4. In section 5, we adapt the methods presented in section 4 to a multi-member gridded precipitation dataset. Finally, in section 6, we provide a summary of the work presented.

2 HISTORICAL OROGRAPHIC STORM ANALYSES

Orographic precipitation is a crucial mechanism by which the atmosphere and land surface interact. Orographic precipitation is vital for natural ecosystems, as well as for the management of human water resources (Roe 2005). The physical mechanisms responsible for orographic precipitation include complex interactions among thermodynamics, fluid dynamics, micro-scale cloud processes, and large-scale patterns of the atmospheric general circulation (Roe 2005). While orographics play an important role in precipitation, topography itself does not cause precipitation. Rather, precipitation occurs when storm types that can occur anywhere (e.g. convection, frontal systems, and tropical cyclones) form near or move over complex terrain. The interaction of the atmosphere with complex terrain can aid or hinder the formation and continuation of precipitation, but topography on its own does not cause precipitation (Houze 2012). Over time, investigations of orographic precipitation have followed along three primary tracks: observations, theory, and modeling. In this chapter, we review different categories of orographic storm analyses (e.g. National Weather Service approaches and peer-reviewed manuscripts), highlight some of the various advancements made through time, and discuss limitations among those methods.

2.1 Hydrometeorological Reports and Related U.S. Federal Studies

Beginning in the 1930s, the U.S. Weather Bureau (USWB) began publishing reports that could be used to estimate extreme precipitation across the U.S. These reports were initially site-specific and included very limited spatial domains. For example, HMR 1 was specifically written to estimate the maximum possible precipitation over the Ompompanoosuc Basin above Union Village, Vermont in 1943. Over time, the spatial domain in each HMR increased, such that by 1961 an entire state (California) was covered (HMR 36). By the late 1970s, HMRs were used to estimate PMP over large regions of the U.S. HMR 55A developed the Storm Separation Method to estimate the convergence and orographic components of precipitation still in use today. The most recently published HMR was finalized in 1999 for the state of California. Rather than discussing all HMRs, we focus on those that introduced new ideas/approaches to orographic precipitation. HMRs not explicitly discussed in this report (e.g., HMR 57), do not introduce new methods or ideas related to the computation of orographic precipitation. Figure 2.1 shows the regions covered by the most up-to-date HMRs across the U.S. and Puerto Rico, while Table 2.1 includes a summary of information regarding each HMR discussed in this section.

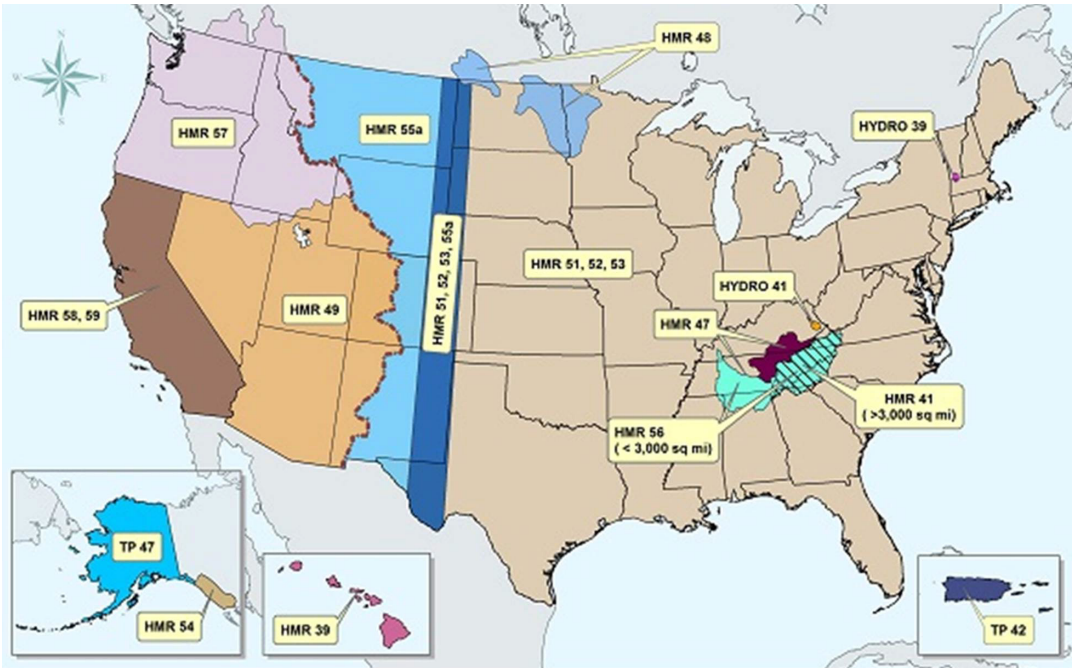


Figure 2-1 Locations of Current HMRs for the United States and Puerto Rico (HSDC2016)

Table 2-1 Summary of Information for Each Study Reviewed in Section 2.1

Document	Date Published	Region	Area Size	Durations	Orographics
TP 42	1961	Puerto Rico/ Virgin Islands	≤400 mi^2	≤ 24 hrs	1km vertical lift, all condensation considered precipitation
HMR 36	October 1961	California	≤1,000 mi^2	≤ 72 hrs	Inflow vs. outflow moisture; conservation of mass
HMR 39	May 1963	Hawaiian Islands	≤200 mi^2	≤ 24 hrs	Relates ground-slope to orographically-enhanced precipitation
HMR 41	June 1965	TN River Valley	7,980 and 21,400 mi^2	≤ 72 hrs	None
HMR 45	May 1969	TN River Basin	≤3,000 mi^2	≤ 72 hrs	First and secondary barriers contribute up to 10% per 1,000ft and 5% per 1,000ft respectively
HMR 49	1977	CO River and Great Basin	≤5,000 mi^2	≤ 72 hrs	Compares precipitation in non-orographic to highly-orographic areas; difference between

Document	Date Published	Region	Area Size	Durations	Orographics
					the two is orographic precipitation
HMR 55A	June 1988	Continental Divide to 103rd Meridian	≤5,000 mi²	≤ 72 hrs	Storm Separation Method
HMR 56	October 1986	TN River Basin	<3,000 mi²	≤ 72 hrs	Same as HMR 45 with addition of TSF, BOF, and TAF

2.1.1 Technical Paper 42

Technical Paper Number 42 (TP 42; USWB 1961a) was one of the first documents to provide instructions on how to estimate PMP over heterogeneous topography with widely varying average annual precipitation amounts. The paper focused on a region encompassing Puerto Rico and the U.S. Virgin Islands (Figure 2.2 and Figure 2.3). TP 42 stated that a hurricane would cause the largest six- to 24-hour precipitation amounts in and around Puerto Rico and the U.S. Virgin Islands. TP 42 divided the total PMP into two sub-categories: convergence and orographic based precipitation. This idea came from Fletcher (1951), who recommended dividing precipitation into orographic and non-orographic components. Fletcher (1951) however, cautioned against trying this methodology on single storms without considering the wind and humidity conditions associated with the storm.

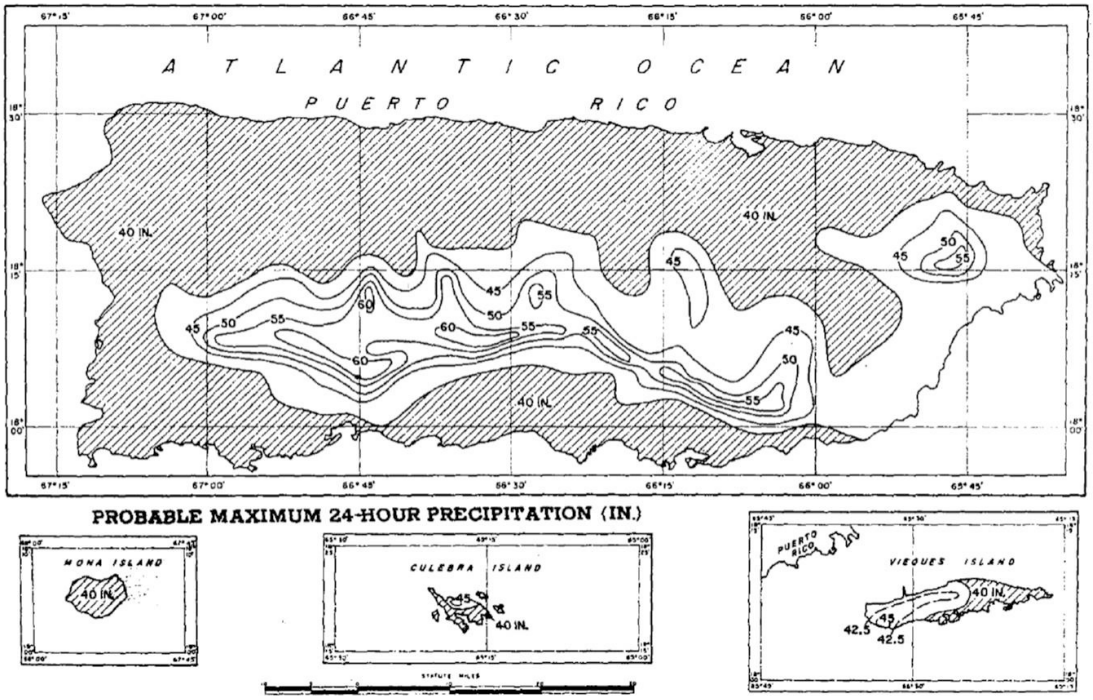


Figure 2-2 PMP Values for Puerto Rico Based on TP 42 (USWB 1961a)

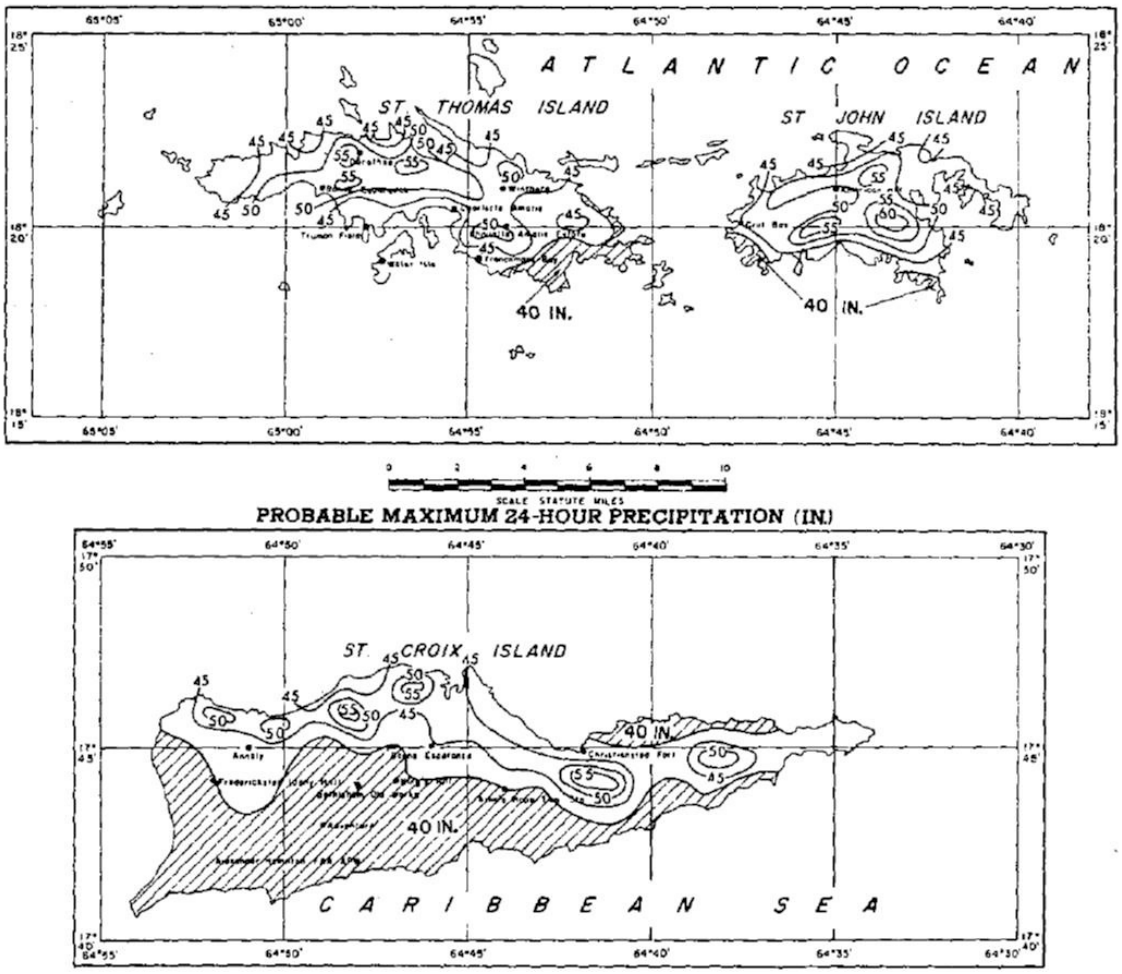


Figure 2-3 PMP Values for the Virgin Islands Used in TP 42 (USWB 1961a)

In a hurricane, winds radially converge towards the center of circulation in the lowest one km of the atmosphere. This convergence forces air to rise near the center of the hurricane. Condensation occurs in the rising air due to cooling from vertical ascent, which leads to precipitation. The estimated amount of precipitation occurring in a hurricane due to convergence was labelled convergence rainfall in TP 42 and became the theoretical basis for convergence rainfall used in the storm separation method in subsequent HMRs. The orographic component of the storm separation method was also first introduced in TP 42. The orographic component is based on lifting a saturated parcel one km (the approximate height of what is considered to be the top of the inflow layer for the elevation ranges found in Puerto Rico) and then assuming that all water vapor condenses and turns into precipitation. The convergence and orographic rainfall components are then added together to estimate the total rainfall.

There are several assumptions that are made in the storm separation method of calculating PMP. First, the convergence portion of precipitation in TP 42 is based on a tropical system. If any other type of system other than a tropical system is used, this assumption becomes invalid. Second, the air is assumed to be saturated. Saturated air at the surface of the Earth is uncommon in most precipitation events, particularly when the event occurs in a location far away from a coast or source of moist air. Third, all the condensation produced in the orographic

portion of the storm separation method is considered to be turned into precipitation that reaches the surface. This rarely happens, as dry air is usually entrained into a storm which results in some precipitation evaporating before reaching the surface. Fourth, the separation method assumes all condensation becomes precipitation and no moisture is lost at the top of the tropical system. In reality, a tropical system is not a perfect heat engine (Tsonis 2007; Holton and Hakim 2013), and some moisture is lost due to inefficiencies of the precipitation process. This loss of energy in a tropical system in is often referred to as entropy (Tsonis 2007).

It is important to note that the HMRs (in this case, TP 42) were ultimately written to provide an engineer with a conservative precipitation amount. This conservative precipitation amount was then used for design purposes for different public works (e.g., dams, culverts, bridges, etc.).

2.1.2 HMR 36 PMP in California

HMR 36 (USWB 1961b) was written at the same time TP 42 was being produced. While HMR 36 only covers one state, California, it nonetheless involves estimating PMP over a large spatial domain of complex topography with large ranges in annual average precipitation (Figure 2.4).

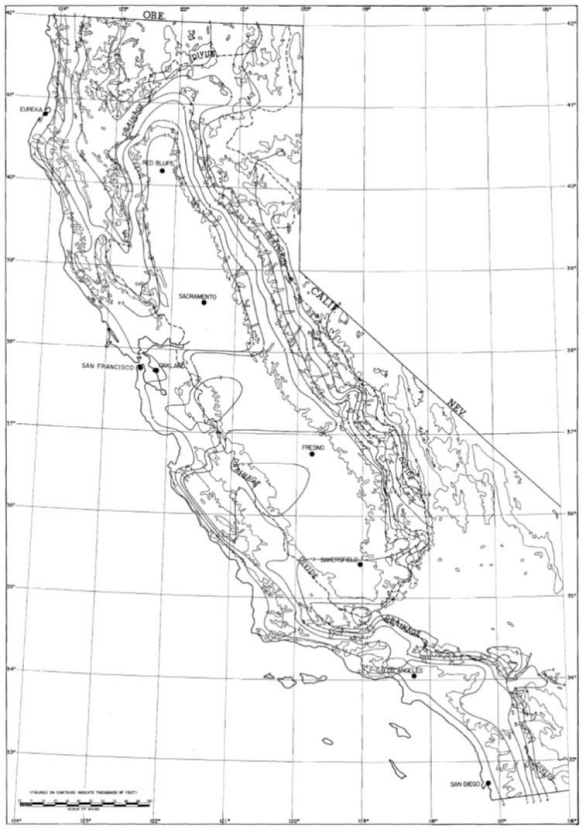


Figure 2-4 Spatial Extent Covered in HMR36 (USWB 1961b)

HMR 36 uses the storm separation method to determine convergence and orographic PMP separately and then combines them to produce overall PMP values. However, the methodology for storm separation used in HMR 36 is different than TP 42. While TP 42 uses a hurricane model for the convergence PMP, HMR 36 uses horizontal convergence, frontal systems, and instability. However, if a storm had instability as an important factor, then it was excluded from consideration when calculating convergence PMP. Convergence PMP is summarized in

Equation 1 while the mass continuity equation used in the orographic model is summarized in Equation 2.

$$\left(\frac{Precipitation_{max}}{Moisture_{max}}\right) * Moisture_{max} = convergence\ PMP \quad (1)$$

where *Precipitation$_{max}$* represents maximized precipitation based upon maximum persisting dew points, and *Moisture$_{max}$* represents maximized moisture based upon maximum persisting dew points.

$$\frac{R}{t} = \frac{0.4V_1\bar{q}_1\Delta P_1 - 0.4V_1\Delta P_1\bar{q}_2}{Y} = \frac{0.4V_1\Delta P_1(\bar{q}_1 - \bar{q}_2)}{Y} \quad (2)$$

In Equation 2, *R* represents precipitation in inches, *t* represents time in hours, V_1 represents speed of air at inflow, $\bar{q}_1$ represents average specific humidity at inflow, ΔP_1 represents pressure difference between inflow and the outflow, $\bar{q}_2$ represents average specific humidity at outflow, and Y represents distance between the inflow and outflow in miles. Note that Equation 2 is reproduced here using the exact form presented in HMR 36 (their page 41). However, as is, the units of Equation 2 do not balance out. This imbalance is the result of two missing constants, the density of water and the gravitational constant.

Orographic PMP calculations, such as Equation 2, assume that the total precipitation caused by orography is the difference between the inflow and outflow moisture as saturated air flows over a barrier. This is a conservation of mass approach to estimate precipitation and is calculated using a 3-dimensional orographic model (USWB 1945; USWB 1947). HMR 36, like TP 42 before it, assumes saturated air is present in the storm in question even though TP 42 was developed for a tropical location, while HMR 36 was developed for a non-tropical location. In reality, even when precipitation is occurring the atmosphere is rarely completely saturated. If the air is not fully saturated, the assumption that all moisture becomes precipitation as a parcel of air is lifted orographically is unrealistic. This simplifying assumption does not include many other factors (e.g., dry air entrainment, hail contamination, inefficiencies in the precipitation process) that could result in not all the moisture being converted into precipitation that reaches the ground. HMR 36 excluded storms in which instability played an important role in the generation of convergence precipitation. This exclusion has the result of decreasing the sample size of storms used in the study due to methodology preference rather than scientific reason.

2.1.3 HMR 39 PMP in the Hawaiian Islands

HMR 39 (USWB 1963) was written after TP 42 and shares much of its methodology with TP 42 since both areas are tropical (Figure 2.5). PMP estimates in HMR 39 cover durations up to and including 24 hours. PMP estimates followed the basic premise of storm separation used in TP 42 and Fletcher (1951). This included estimating a non-orographic PMP and then adjusting the base value for topography. This method of storm separation uses an empirical, simplified approach. Both Fletcher (1951) and HMR 39 do not include the theoretical knowledge of effects of complex topography on precipitation in favor of a simplified presentation that includes a single geographical distribution of rainfall. While TP 42 used a tropical system as the prototypical PMP storm for Puerto Rico, HMR 39 used an upper-level cold-core storm with embedded thunderstorms as the PMP prototype for the Hawaiian Islands. Convergence in HMR 39 is assumed to be caused by a slow-moving frontal system, instead of radial convergence caused by a hurricane in TP 42. Another marked difference was the seasonality of the PMP in HMR 39.

A cool season PMP storm was considered more appropriate for Hawaii than a warm season PMP storm.

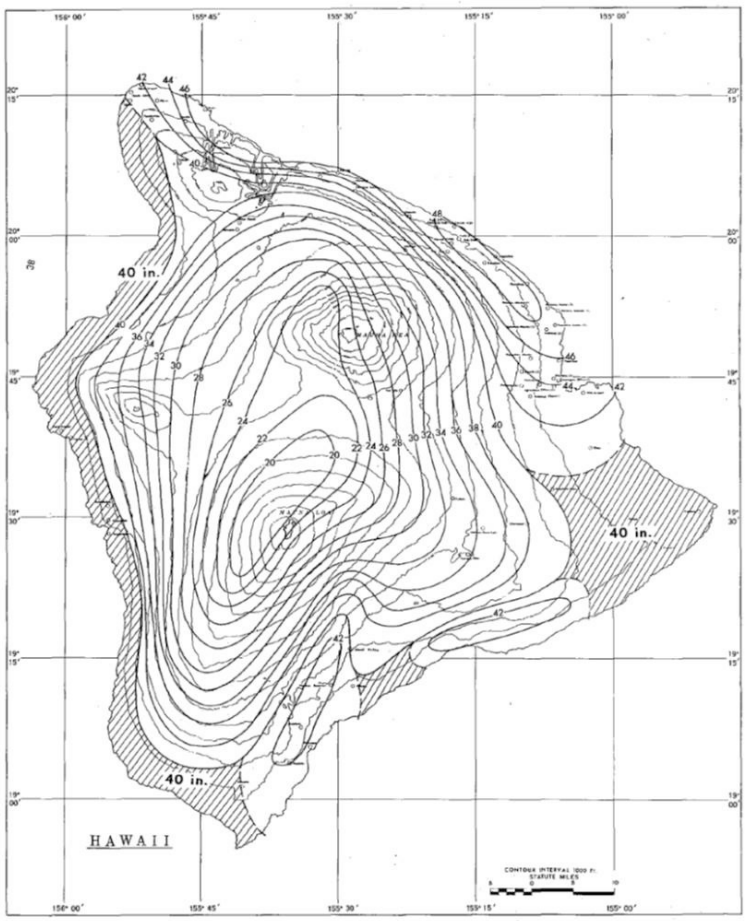


Figure 2-5 24-hour PMP Values for the Island of Hawaii from HMR 39 (USWB 1963)

Much like TP 42, HMR 39 has many assumptions in the estimation of PMP. The simplified method used, while efficient, does not take into consideration the effects of complex topography, and thus does not give as accurate an estimation of PMP as possible. The simplified method assumes saturated air is lifted over a barrier resulting in precipitation. However, since Hawaii features islands created by volcanoes, distinct peaks are present instead of extended mountain ranges, as compared to, for example, the Sierra Nevada in California. The air can move around these distinct peaks instead of being lifted over them which would result in less precipitation.

2.1.4 HMR 41 PMP in the Tennessee River Basin

HMR 41 (USWB 1965) concluded that three ingredients were necessary for a PMP-type storm across the Tennessee River basin: 1) a large upper-level storm to the west of the study area, 2) a frontal zone at the surface for convergence over the study area, and 3) a rich and continuing moisture supply. HMR 41 divided orographic precipitation into two categories based upon wind orientation: downstream (surface winds from the southwest) and upstream (surface winds from the southeast). Winds from the southeast typically produced stronger orographic precipitation, on southeast facing slopes, than winds from the southwest. Five meteorologists independently

determined that the non-orographic portion of the mean annual precipitation of the basin above Chattanooga, TN was anywhere from 45 to 52 inches, with a mean of 47 inches. This estimate attributes 8-9% of the mean annual precipitation to orographic effects. However, during extreme daily precipitation events, no change in precipitation over the Tennessee River Basin was observed due to orographics. Thus, HMR 41 did not take orographics into consideration when estimating PMP, even though the study area occurred in a mountainous region (Figure 2.6). HMR 41 assumes that orography does not play a large enough role in the study area to be considered. This is in light of HMR 41 stating that 8-9% of annual precipitation is likely due to orographic effects and the orientation of the wind in relation to orography can affect precipitation totals.

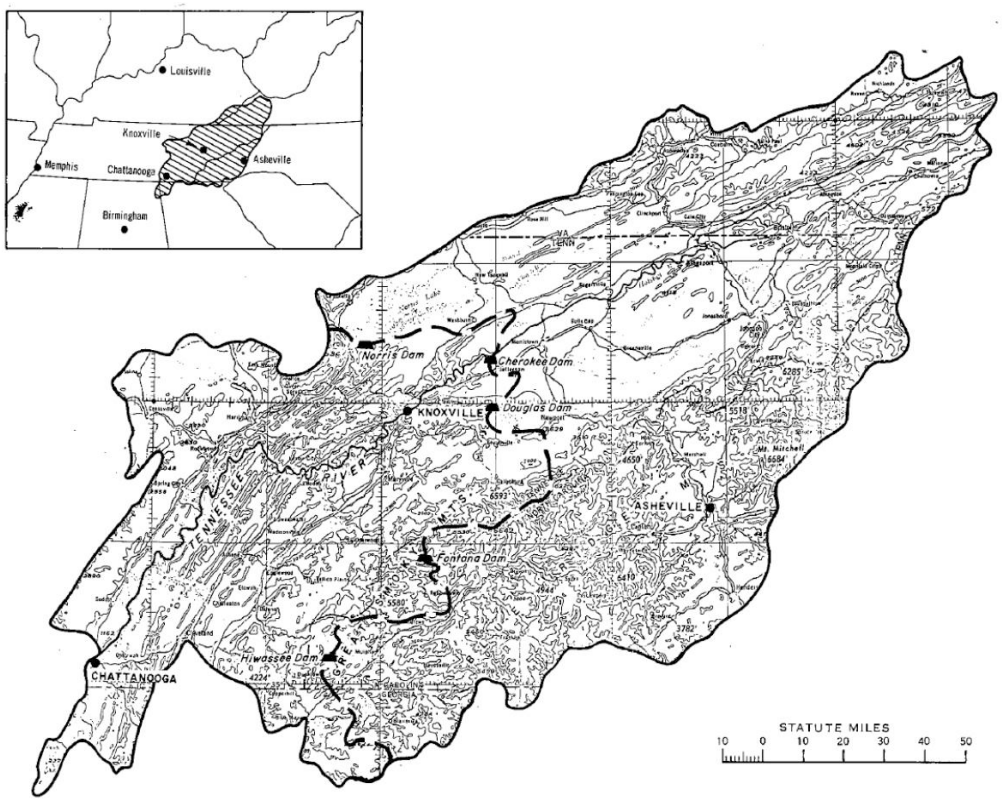


Figure 2-6 Location and Spatial Extent Covered in HMR41 (USWB 1965)

2.1.5 HMR 45 PMP in the Tennessee River Basin

HMR 45 (USWB 1969), an update to HMR 41, used three categories to describe topography in the Tennessee River watershed (Figure 2.7): 1) smooth, 2) intermediate, and 3) rough. The rough category is further delineated into first upslope, secondary upslope, and sheltered areas. The first upslope is defined as a mountain slope with no topographical barriers between it and the moisture sources of the Gulf of Mexico or Atlantic Ocean. A precipitation increase of 10% per 1,000 feet rise from sea level up to 2,500 feet (with no further increase above 2,500 feet) is added to the total precipitation to account for orographic effects on precipitation. The secondary upslope is defined as a mountain slope with barriers between it and the moisture source. A precipitation increase of 5% per 1,000 feet rise from sea level is added to secondary slopes to account for orographic effects on precipitation. A sheltered area is defined as a valley having upwind barriers 2,000 feet or higher between the valley floor and the moisture source. A

precipitation decrease of 5% per 1,000 feet of differing elevation between the crest of the barrier and the valley floor is added to account for orographic effects on precipitation.

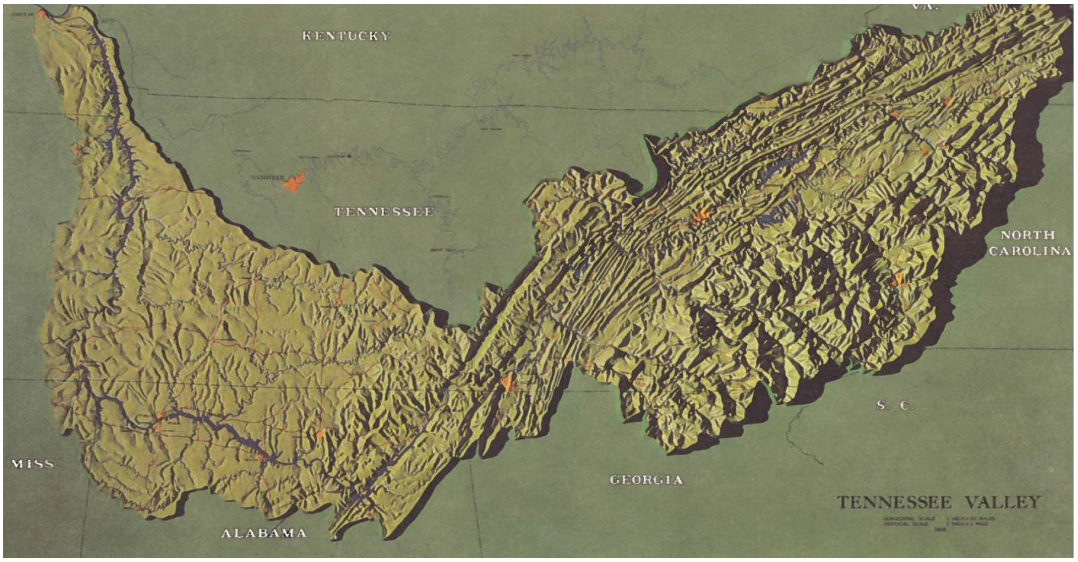


Figure 2-7 Location and Spatial Extent Covered in HMR 45 (USWB 1973)

In an addendum to HMR 45 (USWB 1973), precipitation produced by small-area thunderstorms is increased by 20% due to triggering and fixing effects caused by orography. Triggering effects include the initiation of a thunderstorm due to differential surface heating effects caused by topography while fixing effects include the slow or non-movement of a thunderstorm due to topography.

HMR 45 includes orographic effects to precipitation (as opposed to HMR 41 which did not) but does not provide a physical justification for suggesting a 10% increase per 1,000 feet of elevation (or other increase and decrease used in the report). The only explanation given is that experimentation led the authors to the results given. However, some of the results for PMP do not adequately encompass the data used in the study. For example, in Figure 2.8 the Holt, MO observed precipitation total falls outside of the adopted PMP curve.

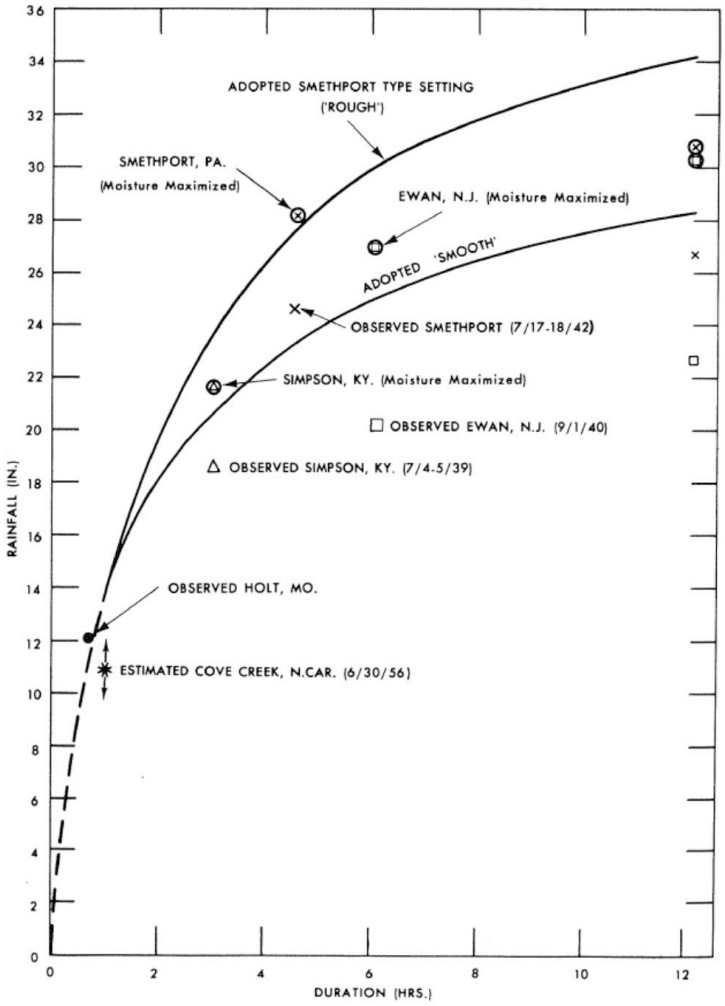


Figure 2-8 Five Square Mile Local Storm PMP Totals from HMR 45 (USWB 1973)

2.1.6 HMR 49 PMP in the Colorado River and Great Basin Drainages

HMR 49 (Hansen et al. 1977) developed a method for generalized PMP studies to estimate a first approximation of orographic PMP using a single orographic index map. This single orographic map is created by comparing an area with little orographic effect (assumed to be convergence only precipitation) to an area with a large orographic effect. The difference in precipitation between the two maps is considered to be the amount of precipitation caused by orographics. HMR 49 states that many caveats exist for the single orographic index method. For example, the method does not take into account differences in topographic effects due to monthly variation in moisture, wind, and the height of the freezing level.

2.1.7 HMR 55A PMP Between the Continental Divide and 103rd Meridian

HMR 55A (USWB 1988) developed the current methodology used for the Storm Separation Method (SSM). SSM is an empirically based idea that consists of two parts: 1) Free Atmospheric Forced Precipitation (FAFP), which was called convergence precipitation in earlier HMRs, and 2) orographic precipitation. The SSM method uses five modules (Figure 2.9), which can be thought of as a series of steps, that when combined, equal the total precipitation that

occurred in a sampled storm or for estimating PMP. Modules one through three deal with the recorded amount of precipitation and how much of it can be attributed to FAFP or orographic influences. Module four averages the results from the first three modules in an effort to achieve a better answer than any of the first three modules could arrive at independently. Module five is the most subjective of the modules and attempts to estimate the confidence in the precipitation amounts found in modules one through three. The amounts of precipitation from modules one through three that have the most confidence are given the most weight in the final calculation of total precipitation.

The SSM methodology developed in HMR 55A has many assumptions, some of which are very clearly stated in HMR 55A. The report notes that FAFP is an idealized property since no research has been used to identify which amounts of precipitation were formed by orographic forcing and which by atmospheric forcing. Module five requires a substantial amount of subjectivity which inherently means a “correct” answer cannot be known and, hence, there is no way to know which of the various techniques used produces a “correct” result most frequently. In HMR 55A, maximum persisting dew points above 74°F are used in five separate figures (Figures 4.4, 4.10, 4.11, 4.12, and 4.13) as a method to maximize precipitation. However, according to Lenderink et al. (2011), dew point values above 74°F do not indicate increased precipitation intensity and may even signify a decrease in precipitation intensity.

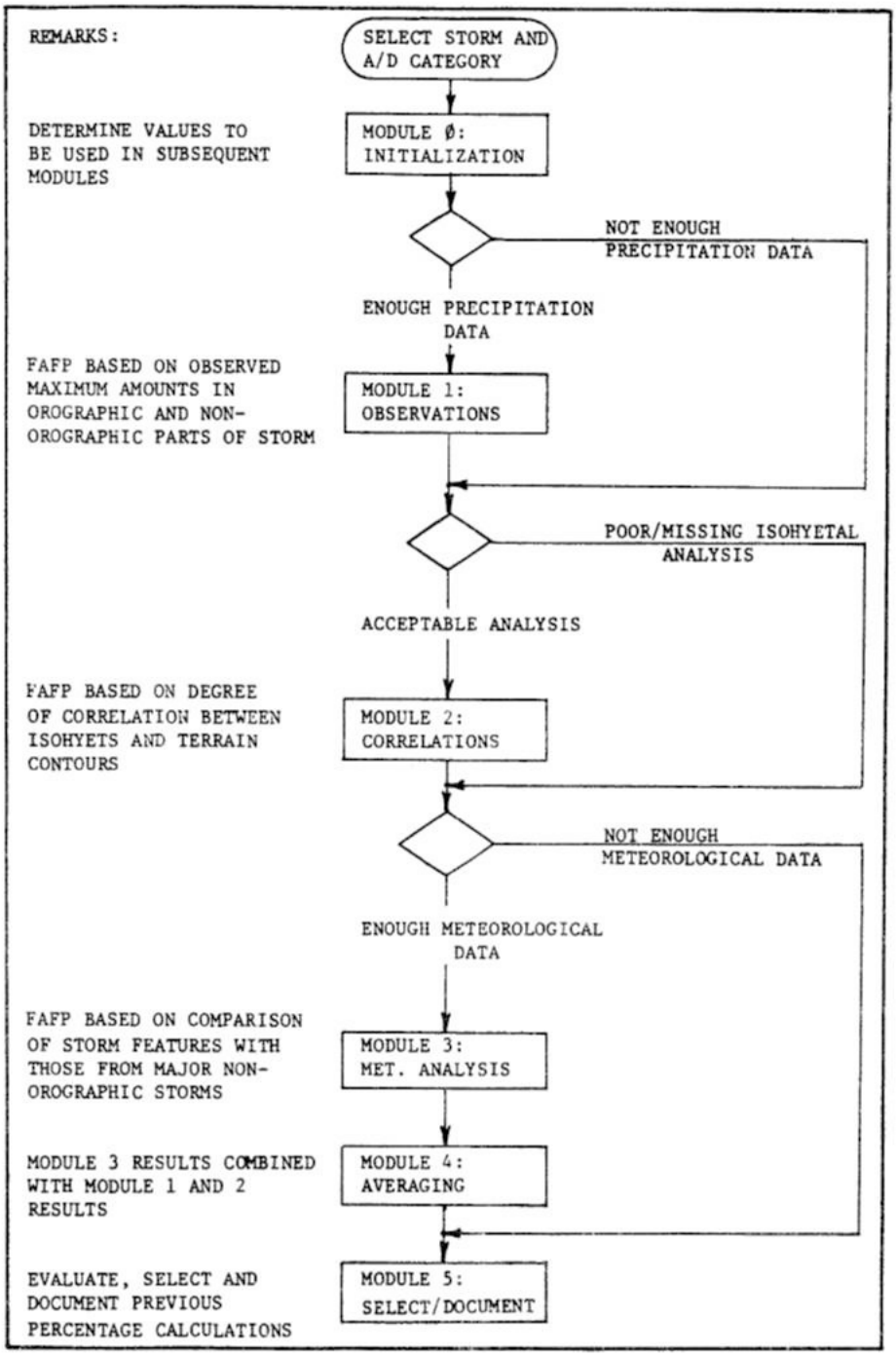


Figure 2-9 Steps in the Storm Separation Method of HMR 55A (USWB 1988)

2.1.8 HMR 56 Probable Maximum and TVA Precipitation Estimates with Areal Distribution for Tennessee River Drainages Less than 3,000 mi^2 in Area

HMR 56 (Zurndorfer et al. 1986) updates and supersedes HMR 45 but still uses most of the same methodology employed in HMR 45 to account for orographics. For example, the three same categories used in HMR 45 to describe topography (smooth, intermediate, rough) are

identical in HMR 56 in spatial extent and effect on PMP in local storms (see Table 2.1). There are some differences in PMP development between HMR 45 and 56, namely, the addition of the Terrain Stimulation Factor (TSF), Broadscale Orographic Factor (BOF), and Total Adjustment Factor (TAF). The basic concepts behind the TSF, BOR, and TAF were originally used in the addendum to HMR 45 but were not named and refined as such until HMR 56. The TSF calculates the percentage of the watershed that is classified as topographically intermediate and rough and then uses those percentages to modify the PMP accordingly. The BOF is the percentage of the watershed comprised of primary upslopes, secondary upslopes, and sheltered area. The TAF is the combination of the TSF and the BOF and is applied to the nonorographic component of the PMP to calculate the total PMP.

2.2 International Manual for Estimation of PMP

The World Meteorological Organization (WMO) published a study in 2009 (WMO 2009) reporting methodologies for estimating PMP around the world. The two main methods proposed in the WMO report were: 1) a comprehensive meteorological analysis to estimate PMP and 2) a statistical method to estimate PMP. The orographic separation method was used under the comprehensive meteorological analysis approach. The orographic separation method used in this study is very similar to the one used in HMR 36. A laminar flow model is used to calculate the amount of precipitation that would result from orographically induced precipitation. A secondary method to determine orographic precipitation (the single orographic index method introduced in HMR 49) is also mentioned as an alternative. Convergence precipitation (similar to the method used in HMR 36) is then combined with orographic precipitation to estimate the total precipitation. The statistical approach relied on a method devised by Hershfield in 1961 (Koutsoyiannis 1999; WMO 2009). The Hershfield statistical method is based upon Equation 3, where h_m is the maximum observed rainfall at the site of interest, h_n and s_n are the mean and standard deviation, respectively, of a series of n annual rainfall maxima at the site, and k_m is a frequency factor.

$$h_m = h_n + k_m s_n \tag{3}$$

Along with the assumptions listed in the HMRs above, the WMO Manual for Estimation of PMP has some cautionary remarks when estimating PMP in orographic regions. Precipitation data in mountainous locations tends to be sparse and located in valleys rather than on steep slopes. Thus, the historical precipitation record likely does not capture the full effect orographics has on precipitation. In reality, atmospheric instability likely disrupts the laminar flow assumed in the orographic model, which will result in erroneous precipitation amounts. Results from Browning et al. (1974) suggest convection is often embedded in orographic precipitation, which means precipitation totals using the orographic model in WMO 2009 may be underestimated. An overestimation of either the convergence or orographic component (likely due to cross contamination of data) will lead to precipitation amounts above what may have happened. The WMO also cautioned against using this statistical method on areas greater than 386 mi^2 (1,000 km^2). The original Hershfield method was developed in 1961 for point precipitation. WMO (2009) suggest that using it on basin sizes above 386 mi^2 will provide inaccurate results.

2.3 PMP-Related Studies by Consultants

Many site-specific PMP studies have been performed over the past 20 years. In general, these studies have been conducted by consultants (as opposed to government agencies) and have

often resulted in decreased PMP values. The Bureau of Reclamation (Reclamation), however, also conducted site specific PMP studies between the 1950s and 1970s, before adopting the generalized methods found in the HMRs in the 1980s (see section 2.3.1). We attempted to include Reclamation's site-specific studies in this report; however, they could not be located (physically or electronically).

2.3.1 PMPs Never Happen – or Do They?

Harrison (2001) gave a presentation at the 2001 Association of State Dam Safety Officials annual conference concerning the topic of whether precipitation totals ever approach or exceed PMP. Unfortunately, Harrison gave his presentation on the afternoon of September 11, 2001, which limited the dissemination of information. In his presentation (and paper from the proceedings), Harrison showed that PMP estimations from HMR 51 have been exceeded on at least two occasions (Haynes Camp, CA and Smethport, PA) with more than half a dozen examples of storms approaching PMP precipitation values (Figure 2.10). If a storm exceeds the PMP value for the area and type of storm, the PMP value is considered incorrect (Caldwell et al. 2011). When one of these storms occurred in the Mid-Atlantic region, the Virginia State Climatologist's Office stated that topographic relief was one of the factors that contributed to the extreme rainfall (Harrison 2001). The combination of instability released via convection (thunderstorms) and orographically enhanced precipitation was deemed the main causes of extreme precipitation for small areas less than 200 mi^2. For larger spatial domains, tropical systems, or their remnants, interacting with topography were deemed the main causes of extreme precipitation in the Mid-Atlantic, an argument supported by Schwarz (1970). Harrison noted that regions in the mid-latitudes (i.e., not tropical) typically experience the most extreme precipitation totals for durations less than ten hours, while tropical locations typically experience the most extreme precipitation totals for durations greater than 10 hours. In longer duration precipitation events, especially those in or near the tropics, sustained flow over topography for longer temporal durations can cause longer duration precipitation.

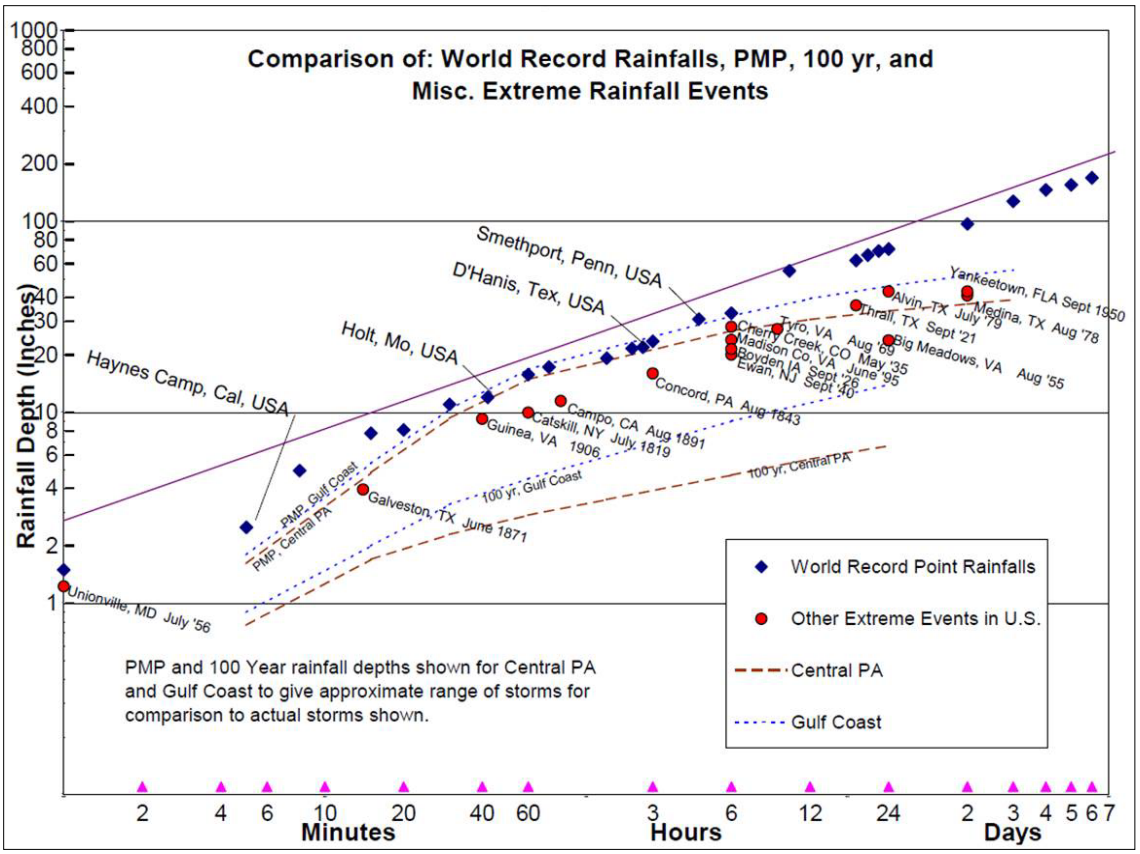


Figure 2-10 Comparison of HMR 51 PMP and Extreme Rainfall Events. Figure from Harrison (2001).

2.3.2 AWA PMP Study for Cherry Creek Reservoir

Applied Weather Associates (AWA; Tomlinson et al. 2003) conducted a PMP study for the Cherry Creek watershed near Denver, CO in 2003. The AWA study showed a 25% decrease in 24-hour PMP values for the Cherry Creek watershed when compared to values estimated in a 1995 NWS study of PMP in the Cherry Creek watershed. The differences in PMP estimates are mainly due to certain orographic effects not being utilized in the AWA report that were present in the NWS report, specifically the K-factor, which estimates the net effect orography has on total PMP rainfall. Other differences include a smaller sample size of storms used in the AWA report compared to the NWS report, areal reduction techniques, and the use of different significant figures in dewpoint temperatures. As part of the review process for the AWA report, Mr. Lou Schreiner, former Meteorologist with Reclamation, stated that Reclamation had used a procedure like AWA's to estimate site specific PMPs in the 1950s through the 1970s. Reclamation abandoned the site-specific approach in the 1980s because the generalized approach used in the HMRs was considered to be superior and gave more conservative values (Tomlinson et al. 2003). Note that during the review process for the AWA report, the smaller PMP estimates AWA calculated were considered more indicative of the uncertainty inherent with PMP rather than a flawed methodology used by AWA.

2.3.3 Gomez and Sullivan Site-Specific PMPs for New England

Gomez and Sullivan (2009) showed that site-specific PMP studies for areas in New England can produce PMP values reduced by 13-60% when compared to the generalized PMP approach of HMR 51. Moisture deficiencies caused by topographic barriers in New England (which include primary and secondary ridges) were one of the reasons PMP decreased. However, Gomez and Sullivan (2009) did not include Mesoscale Convective Complexes (MCCs) in their analysis even though MCCs can and do occur in New England (Bentley 1997). MCCs can result in extreme amounts of precipitation. The Smethport, PA storm of 17-18 July 1942, which produced 30.8 inches of rain in a four-and-a-half-hour timeframe, is believed to have been caused by an MCC (Gomez and Sullivan 2009).

2.3.4 AWA PMP Study for Arizona

In 2013, AWA (Tomlinson et al. 2013) conducted a PMP study for the Arizona Department of Water Resources to estimate PMP for the entire state of Arizona. HMR 49 is the report used to estimate PMP in Arizona (Hansen et al. 1977). The methodology used in the AWA study was similar to the 2003 PMP Study for Cherry Creek Reservoir. In both studies, the Storm Precipitation Analysis system (SPAS) was used to create high resolution spatial precipitation maps from hourly and daily precipitation reports of individual storms (Tomlinson et al. 2013). An Orographic Transposition Factor (OTF) was used to incorporate an orographic component of precipitation to the total PMP value (Equation 4, Equation 5),

$$OTF = \frac{P_o}{P_i} \tag{4}$$

where P_o is the orographically adjusted precipitation (target) and P_i is the SPAS-analyzed in-place precipitation. It should be noted that without access to the SPAS program this method is not reproducible, which is also noted in Hall (2015).

$$P_o = mP_i + b \tag{5}$$

In Equation 5, P_o is the orographically adjusted precipitation (target), m is the proportionality coefficient (slope), P_i is the SPAS-analyzed in-place precipitation, and b is the transpositional offset (y-intercept).

Instead of estimating orographic precipitation differences explicitly, OTF relies on NOAA Atlas 14 precipitation values at the 10-, 25-, 50-, 100-, 200-, 500-, and 1,000-year average recurrence intervals. A relationship is developed between a source location (non-orographic) and the target location (orographic) at these recurrence intervals for the 3-hour (local storm) and 24-hour (general and tropical storm) durations. AWA assumes that the 3-hour and 24-hour durations can be converted (via a ratio) to any other needed duration. A best-fit linear trend line is calculated based on a scatterplot of precipitation totals (the previously mentioned 10-, 25-, 50-, 100-, 200-, 500-, and 1,000-year average recurrence intervals), where the linear trend line estimates orographic precipitation based on the non-orographic totals. The primary assumption is that the derived relationship is a function of elevation (Hall 2015). HMR 49 used a method similar as HMR 36 to estimate PMP (see section 2.1.2).

2.3.5 AWA PMP Study for Wyoming

In 2014, AWA (Kappel et al. 2014) conducted a state-wide PMP study for Wyoming at the request of the Wyoming Water Development Office. This study used a methodology similar to previous AWA studies (e.g., Arizona, 2013). A question concerning the OTF arose during the

peer review for the Wyoming study and in an audit performed by the Nuclear Regulatory Commission (Hall 2015). Specifically, the OTF may double count moisture adjustment since the OTF uses NOAA Atlas 14 precipitation-frequency estimates. If the NOAA Atlas 14 precipitation-frequency estimates already include topographical influences on precipitation, then the OTF does not need to introduce any additional topographic influences. The NRC audit (Hall 2015) also noted that the transposition limits in the OTF do not follow the homogeneous regions used to develop the precipitation-frequency estimates in NOAA Atlas 14.

2.3.6 AWA PMP Study for Virginia

In 2015, AWA (Kappel et al. 2015) conducted a PMP study for the Virginia Department of Conservation and Recreation to estimate PMP for the entire state of Virginia. Similar to the AWA PMP studies for Cheery Creek Reservoir (Tomlinson et al. 2003) and Arizona (Tomlinson et al. 2013), the Virginia PMP study used an in-place OTF to estimate orographic precipitation instead of the SSM developed in HMR 55A (see section 2.1.7). The result of this different methodology is a net decrease in PMP values over most topographical areas in Virginia (Figure 2.11). The decrease in PMP values over topographical areas in the Virginia study is typical when using the OTF method, as compared to the SSM method (see section 2.3.1).

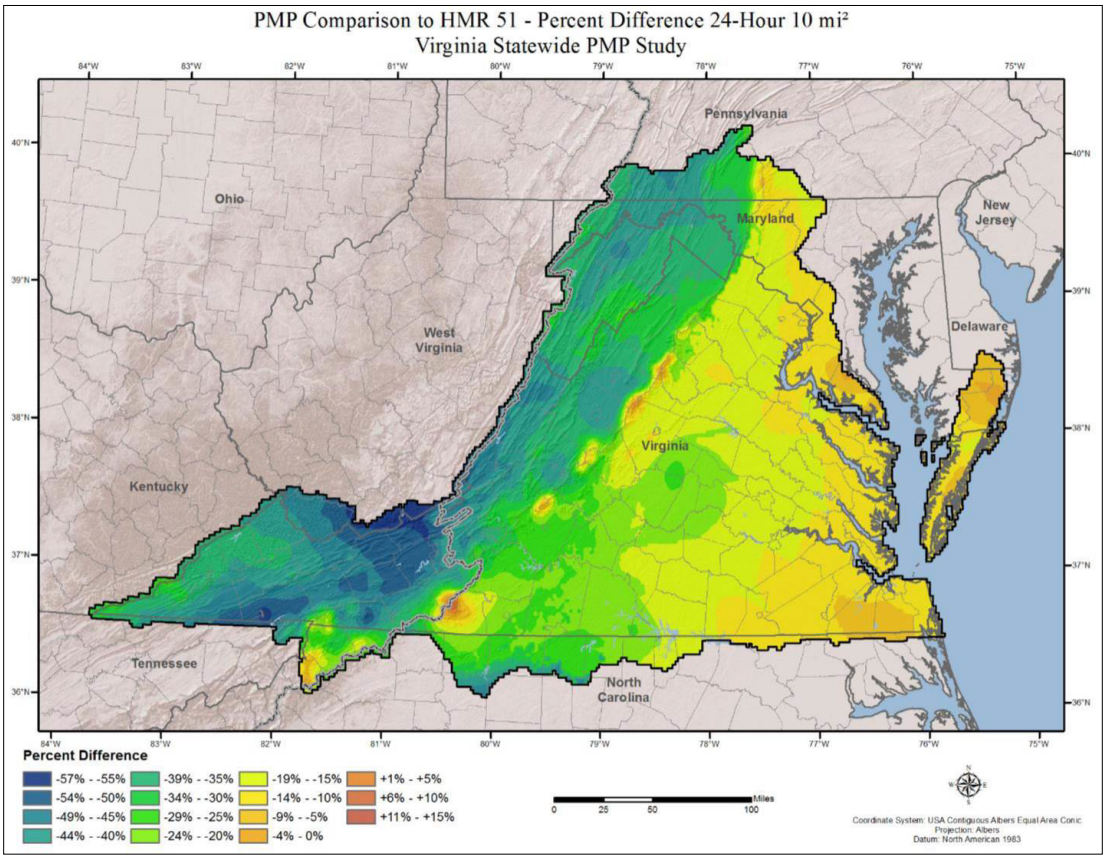


Figure 2-11 Difference in HMR 51 and AWA 24-hour 10 miles^2 PMP for Virginia. Values Relative to Totals in HMR 51. Figure from Kappel et al. (2015).

2.4 Academic Studies

Previous sections discuss historical approaches to orographic storm analyses that have been put forth by various organizations. Orographic storm analyses presented in the HMRs, for example, represent instructions on *how* to calculate PMP in different parts of the U.S. Conversely, peer-reviewed approaches to orographic precipitation analyses, which include analytical investigation of physical processes, have also been published. Rather than prescribing instructions on how to compute PMP, these studies focus on expanding working knowledge of processes that impact extreme precipitation and PMP. In this chapter, we review published orographic precipitation analyses as they have developed through time.

2.4.1 Early Studies

In addition to coordinated approaches to orographic precipitation analyses, scientists and engineers have been publishing investigations of orographic precipitation for over 100 years. During the late 1800s, some individuals started exploring the relationship between elevation and mean annual precipitation using precipitation gauges. For example, Stow (1875) collected and analyzed data from approximately 1,800 gauges located across the British Isles. The author categorized orographic precipitation and its relationship with prevailing synoptic conditions (Stow 1875). In 1911, Thiessen published a paper noting that precipitation totals on the western slopes of the Wasatch Mountains in Utah were much larger than in the remainder of the state (Thiessen 1911).

Later, Mr. J. B. Lippincott reported a relationship between elevation and mean annual precipitation in the western U.S. (Donley and Mitchell 1939). Mead (1919) extended Lippincott's work to other parts of the U.S. and Europe. However, the relationship between elevation and mean annual precipitation was not documented in the Eastern U.S. until the late 1930s. Donley and Mitchell (1939) analyzed 73 rainfall stations located above 2,500 feet in the Southern Appalachian Region, a region defined as the principal Tennessee River watershed east of the Continental Divide. The authors concluded that a definite linear relationship existed between rainfall and elevation in the Southern Appalachian Region.

2.4.2 Mid-Century Studies

By the early 1960s, scientists and engineers were aware of the need for accurate, quantitative precipitation forecasts in areas of complex topography in order to make vital decisions for flood control and water management (Elliott and Shaffer 1962). Dynamical models of orographic precipitation began surfacing during the mid- to late-1960s (Sarker 1966). In some cases, models of orographic precipitation followed the development of models capturing the impact of topography on airflow (Sarker 1965).

Sarker (1966) published results from a dynamical model of orographic precipitation based on linearized equations that assumed a saturated atmosphere with a pseudo-adiabatic lapse rate. The model was used to simulate the distribution of rainfall rates across idealized topography representing the Western Ghats in India. Sarker (1966) compared results from the dynamical model to results from the empirical model developed in HMR 36 (USWB 1963). Peak rainfall rates from the empirical model were less than those produced by the dynamical model and shifted relative to the location of the topography crest. The peak rainfall rates in the dynamic model and observations were located 10-12 km to the west of the crest (i.e. upstream), while the peak rainfall rate in the empirical model was shifted to the crest. Sarker (1966) concluded

that the dynamical model explained the rainfall distribution along the Wester Ghats orography better than the empirical model developed in HMR 36.

Schwarz (1970) explored the extreme precipitation associated with the remnants of Hurricane Camille that fell across Virginia during August 1969. The author used isohyetal techniques, a depth-duration-area analysis, satellite imagery, and radar echoes to describe the temporal and spatial evolution of precipitation during the event. In the absence of sufficient observations, the author manually drew 36-hour precipitation total isohyets (see Figure 2.12) and acknowledged that his isohyets likely overestimated and underestimated precipitation totals in different areas. While the author stated that terrain was important to the intense precipitation, he also stated that a better observation network was needed to verify this. The author also compared precipitation totals around the storm center to the generalized PMP totals calculated using HMR 33 (USWB 1956). His results showed that observed 12-hour precipitation totals around the storm center were within 80-85% of the PMP for areas up to 1,000 mi^2 (Schwarz 1970).

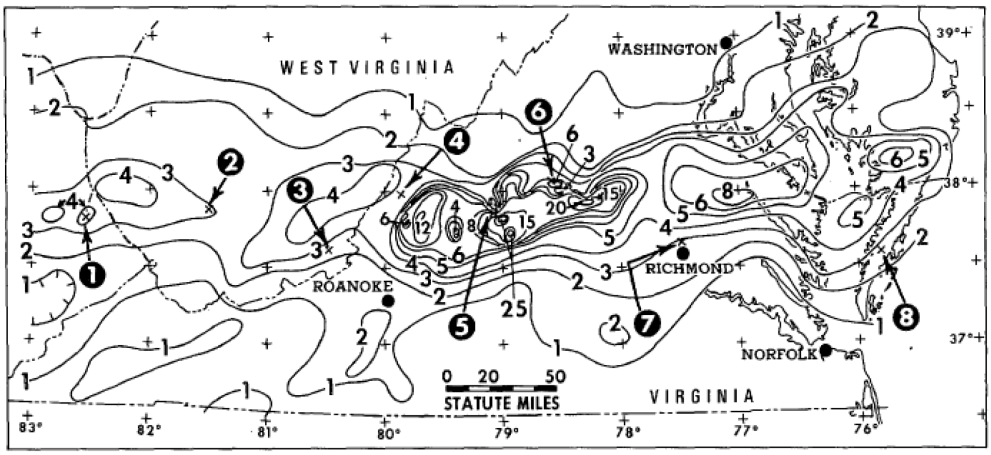


Figure 2-12 36-hour Precipitation Totals (in inches) from Hurricane Camille. See Figure 1 in Schwarz (1970)

In 1975, Hobbs et al. explored the effects of orography, namely the Cascade Range in Washington, on frontal system dynamics by examining changes in cloud configuration and microstructure. The authors used a combination of datasets, including rawinsonde data, aircraft penetrations, rain gauge measurements, and conventional observations. The authors showed that the mountains were responsible for cutting off the low-level moisture source to the frontal system, something that reduced cloudiness ahead of the frontal system and impacted the timing of precipitation.

To increase the effectiveness of its operational river forecasting program and to improve the determination of the spatial distribution of precipitation in mountainous areas, the Sacramento River Forecast Center developed a fine-mesh numerical model to simulate the mesoscale circulation and associated orographic precipitation rates over a limited area (Colton 1975). While the full model was based on the basic equations of momentum, thermal energy, and moisture, a quasi-two-dimensional version was used to perform many simulations of orographic precipitation events in California (Colton 1976). Figure 2.13 shows a schematic of the model grid system in *x-z* coordinates.

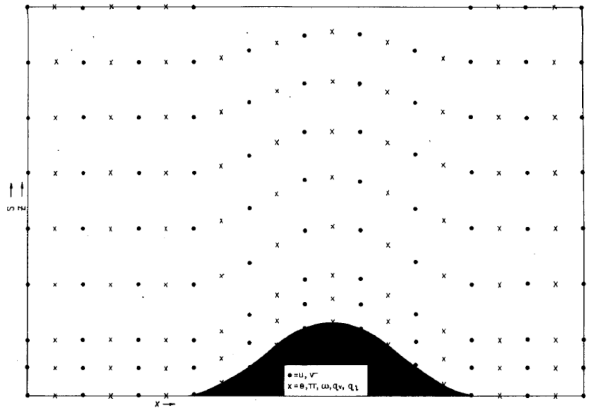


Figure 2-13 Schematic of the Grid System Used by Colton (1976)

Colton (1976) simulated five different cases of strongly orographic precipitation events in the Sierra Nevada Mountains and compared simulated precipitation rates to observed rates from official Weather Service climatological data. Results showed that in most cases, the model-derived precipitation rates agreed favorably with observed rates. The author further stated that the results justified the use of numerical models in determining the spatial distribution of precipitation in mountains regions where orographic influences dominate.

Rhea (1978) developed an operationally based orographic precipitation model designed to diagnose the effect of topography on winter precipitation in western Colorado. The two-dimensional, multi-layer, steady-state model[1] was tested for western Colorado using 13 winter seasons and forced by twice-daily upper air measurements. Results showed large discrepancies in daily precipitation between the model and observations. However, the model was able to realistically produce the frequency distribution of daily precipitation totals (Rhea 1978).

The models used by Colton (1975, 1976) and Rhea (1978) were diagnostic, meaning they neglected time-dependent processes and assumed steady-state solutions (Ezer and Mellor 1994). The atmospheric flow was calculated based on large-scale atmospheric conditions, and under steady-state conditions, the amplitude and patterns of the precipitation rates were diagnosed from that flow. Simplifications in these models mean they were unable to capture the complexity of observed transient processes that could impact extreme precipitation events, such as convection, valley circulations, and blocking (Roe 2005).

The development of radar and, more specifically, weather radar introduced new technologies to the scientific community after World War II. Bent (1946) is credited with the first U.S. publication regarding meteorological phenomena and radar echoes. Browning et al. (1974) used a combination of weather radars, autographic rain gauges, and radiosonde observations to diagnose the sub-synoptic scale structure of orographic precipitation over hills in south Wales. Autographic rain gauges represent one of two types of gauges, a tilting-siphon type or a tipping-bucket type. The authors were interested in the three-dimensional structure and evolution of precipitation upwind, over, and downwind of the hills during a wintertime warm-sector event. Their results suggested that orographic effects increased the intensity of precipitation within regions of mesoscale precipitation areas (MPAs) and increased the extent of precipitation over

[1] The term steady-state implies no change in time, i.e., the time derivative is zero.

hills and up to 100 km downwind (Browning et al. 1974). Their results further suggested that convective cells are often embedded in orographic precipitation (Browning et al. 1974).

Hill et al. (1981) expanded upon work completed by Browning et al. (1974) by exploring the structure of orographic precipitation using radar data combined with autographic rain gauge data from eight case studies in south Wales. Their results demonstrated that orographic enhancement of precipitation is strongly influenced by low-level wind speed. The largest enhancement of rainfall happened in association with strong winds and high relative humidity below 2 km. Alpert (1989) later showed that the orographic enhancement to precipitation by high mountains could be estimated under the assumption that moisture convergence from encountering a mountain slope is proportional to the precipitation enhancement.

2.4.3 Later Studies

Most methods of estimating gridded precipitation estimates from point observations fall into three categories: graphical, topographical, and numerical. Graphical methods involve mapping precipitation data, such as isohyetal mapping (Schwarz 1970) and Thiessen polygons (Thiessen 1911). Topographic methods involve calculating correlation coefficients between precipitation observations and several topographic parameters such as slope, elevation, and location of barriers (Daly et al. 1994). Numerical methods of estimating gridded precipitation are often based on developing some type of statistical model between point observations and grid-cell estimates.

Optimal interpolation (OI) is an example of a numerical interpolation scheme, where the user can estimate the value of a variable at a specific location based on nearby observations (Tabios and Salas 1985). Optimal interpolation schemes were introduced to the field of meteorology during the late 1950s and 1960s (Eliassen 1954; Gandin 1963). However, the technique was not used to interpolate precipitation observations in regions of complex terrain until the early 1990s. Bhargava and Danard (1994) used optimal interpolation to estimate one-day and three-day precipitation amounts on a 20 km grid in the upper Columbia River watershed in southeastern British Columbia. The authors stated that the discontinuous nature of precipitation, along with spatially heterogeneous observation networks, affect the application of optimal interpolation schemes to precipitation.

Alternative numerical interpolation techniques have also been used to fill gaps in observational networks of precipitation in regions of complex terrain. The Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994), for example, is an analytical tool that distributes point observations of precipitation to a regularly spaced grid using a digital elevation model (DEM). As explained more specifically in Daly et al. (1994), "… PRISM 1) estimates the 'orographic' elevation of each precipitation station using a DEM at 5-min latitude-longitude grid spacing, and 2) assigns each DEM grid cell to a topographic facet by assessing slope orientation. PRISM then estimates precipitation at each DEM cell by 3) using a windowing technique to develop a precipitation-DEM elevation regression function from nearby rainfall stations on the cell's facet; and 4) predicting precipitation at the cell's DEM elevation with this regression function." PRISM can continuously accommodate local and regional changes in the orographic precipitation regime because it uses many precipitation-elevation relationships.

Daly et al. (1994) compared orographic precipitation estimates from PRISM to estimates based on alternative interpolate techniques, including kriging, detrended kriging, and cokriging at a site in Oregon. Their results showed that PRISM estimates had the lowest cross-validation bias and absolute error compared to the other three methods (Daly et al. 1994). Furthermore, the

authors tested the application of the model across the Western U.S. (see Figure 2.14). The model was able to reproduce precipitation patterns across wide areas, including wet coastal areas and interior deserts. The model appeared to capture realistic precipitation totals on isolated mountain peaks, where no data exists (Daly et al. 1994).

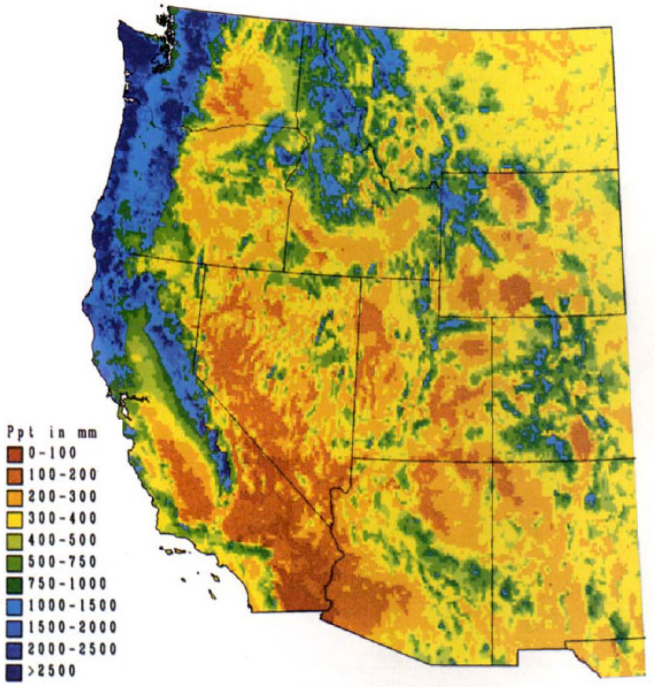


Figure 2-14 Gridded Average Annual Precipitation Totals (in mm) from PRISM. See Figure 9 in Daly et al. (1994).

In 1994, Horel and Gibson analyzed a winter storm in a region of complex topography using a hydrostatic mesoscale model referred to as the Utah Local Area Model (LAM; Paegle and McLawhorn 1983; Nicolini et al. 1993; Waldron 1994). The authors used surface and upper-air observations, satellite imagery, and a vertical profiler in conjunction with numerical simulations to better understand the dynamical processes present during the winter event. The mesoscale model had increased horizontal resolution over terrain (see Figure 2.15) and enhanced vertical resolution compared to the Nested Grid Model (NGM), which is operated by the National Meteorological Center (Horel and Gibson 1994).

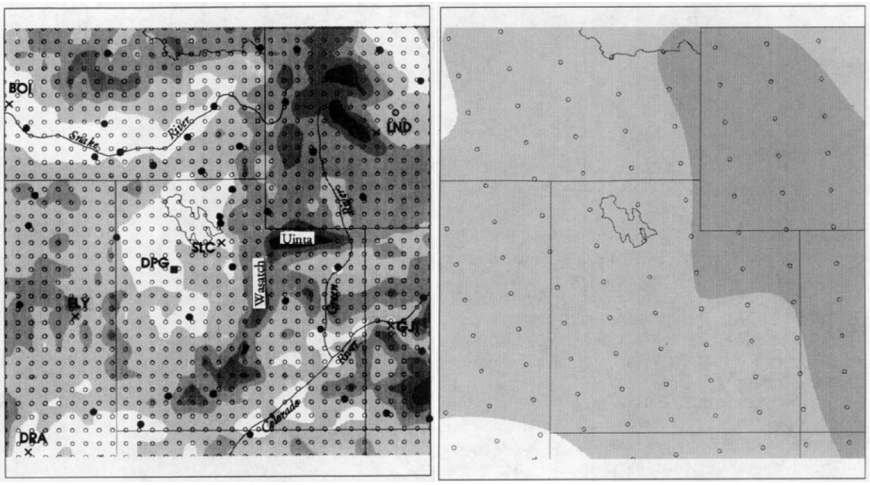


Figure 2-15 Model Grid Points from the High-Resolution Utah Local Area Model (left) and the National Meteorological Center Nested Grid Model (right). See Figure 2 of Horel and Gibson (1994).

Results from Horel and Gibson (1994) indicated that the Utah LAM produced large 24-hour precipitation amounts over mountains due to favorable upslope conditions. NGM precipitation totals along the southern slopes of the Uinta Mountains in Utah (shown in Figure 2.15) were quite small compared to estimates from the Utah LAM. Furthermore, the Utah LAM was capable of developing rain shadows downwind of major mountain barriers.

Katzfey (1995) explored the impact of horizontal resolution and different orographies on the simulation of three extreme precipitation events (greater than 700 mm) in New Zealand using a hydrostatic mesoscale model nested within the European Center for Meteorology and Weather Forecasting (ECMWF) analyses. The author used a 30-km horizontal resolution and progressively increased terrain heights, including increasing terrain heights to an artificial height of the Southern Alps. Results showed that peak precipitation was related to the orographic slope, while area-averaged precipitation was related to maximum elevation (Katzfey 1995). The author increased the horizontal resolution from 30 km to 15 km and found that peak precipitation amounts increased, though area-average precipitation totals were not significantly different.

3 RECENT ADVANCEMENTS IN OROGRAPHIC STORM ANALYSES

Since the late 1990s and early 2000s, individuals and organizations around the globe have been working to improve the datasets and methods used for analyses of precipitation, including those in regions of complex terrain. Improvements to data include changes in the observation network and instrumentation, access to radar and remotely sensed datasets, and high-resolution gridded products. Methodology improvements include major strides in numerical modeling and statistical analyses, among others. In this chapter, we discuss recent advancements to precipitation analyses, including advancements related to orographic precipitation. However, note that recent data and methodological improvements include advancements that are directly *and* indirectly related to orographic storm analyses. As a result, we discuss general advancements that may be relevant to orographic storm analyses. Table 3.1 includes a list of the major datasets discussed in this section, along with a note indicating whether the advancement may have an impact on orographic storm analyses. These advancements include various data types, including observations, radar and remotely sensed products, reanalysis datasets, and numerical models. Point observations from various platforms are considered observations. Gridded datasets are based on point observations combined with some type of interpolation scheme. Radar retrievals are considered modeled, as precipitation totals are estimated using statistical relationships. Remote-sensing datasets are also based on statistical models, and as a result, are considered modeled output. Reanalysis data is considered modeled output because the data are computed using a combination of Earth-system models (e.g., sea-ice model, land-surface model, atmospheric model, etc.) that are constrained by observations. Numerical weather model output is also considered modeled output because of the numerical techniques required to solve the equations of motion.

Table 3-1 List of Major Datasets Discussed in Section 3

Name	Type	Spatial resolution	Temporal coverage	Direct or indirect impact to orographic storm analyses
Global Historical Climatology Network	Observations	Variable	Variable	x
Dual-polarization radar	In-situ product	Variable	Variable	x
NCEP Stage IV	Gridded product	Variable	Variable	x
CMORPH	Method	0.0727° latitude x longitude	December 3, 2002 - present	x
TRMM	Remote sensing product	Variable	Variable	x
Daymet	Gridded product	1 km x 1 km	January 1, 1980 - December 31, 2013	x

Maurer	Gridded product	⅛° latitude x longitude	January 1, 1949 - December 31, 2010	x
Name	**Type**	**Spatial resolution**	**Temporal coverage**	**Direct or indirect impact to orographic storm analyses**
Livneh et al. (2013)	Gridded product	1/16° latitude x longitude	January 1, 1915 - December 31, 2013	x
Newman et al. (2015)	Gridded ensemble product	⅛° latitude x longitude	January 1, 1980 – December 31, 2012	x
NARR	Reanalysis product	32 km x 32 km	October 1, 1978 - present	x
MERRA	Reanalysis product	⅔° latitude x ½° longitude	January 1, 1979 - present	x
CFSR	Reanalysis product	½° latitude x longitude	January 1, 1979 - present	x
NLDAS-2	Reanalysis product	⅛° latitude x longitude	January 1, 1979 - present	x
WRF	Numerical weather model	Variable	Variable	x
HRRR	Numerical weather model	Variable	Variable	x

3.1 Data Improvements

3.1.1 Observations

The Global Historical Climatology Network (GHCN)-Daily is a dataset of daily climate observations recorded over global land areas (Menne et al 2012). The dataset contains climatic data from over 30,000 stations in 180 countries and territories. The five most commonly recorded variables are maximum temperature, minimum temperature, snow fall, snow depth, and total daily precipitation, although the most common single variable is total daily precipitation. Figure 3.1 shows the number of GHCN-Daily precipitation stations located across the globe through time. While the number of stations decreased across the globe between the late 1960s and 1990s, the number of precipitation stations has been increasing since then. Precipitation gauge density has also increased through time (Figure 3.2), both in the U.S. and across the globe.

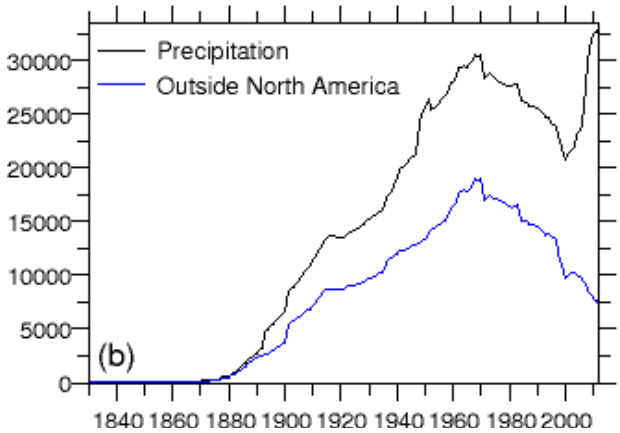


Figure 3-1 Number of GHCN-Daily Precipitation Stations Through Time. Figure Reproduced from https://www.ncdc.noaa.gov/oa/climate/ghcn-daily.

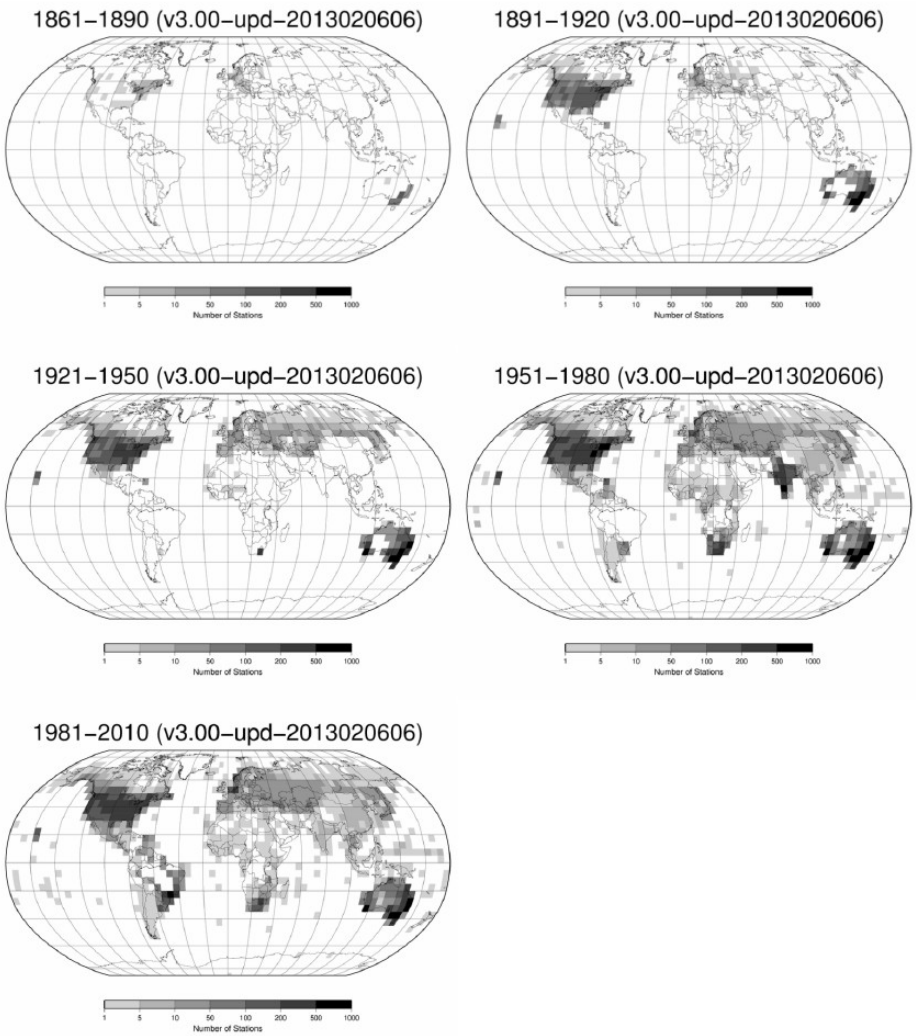


Figure 3-2 GHCN-Daily Precipitation Gauge Density Through Time. Figures Reproduced from https://www.ncdc.noaa.gov/oa/climate/ghcn-daily/index.php?name=coverage.

In addition to a general increase in the number and density of GHCN-Daily precipitation gauges, there have been improvements to the way data are collected and checked for quality assurance (QA; Durre et al. 2010). In the 1960s and 1970s, the Automated Meteorological Observing System (AMOS) and the Remote Automated Meteorological Observing System (RAMOS) were deployed across the U.S. AMOS and RAMOS only recorded ambient and dewpoint temperature, wind speed and direction, and atmospheric pressure. Through the 1970s and 1980s the NWS, Federal Aviation Administration, and the Department of Defense worked together to develop an automated sensor platform that could record more meteorological variables than AMOS and RAMOS (NOAA 1998). The Automatic Surface Observation Systems (ASOS) is an automated weather observation platform that records ambient and dewpoint temperature, wind speed and direction, precipitation type, accumulated precipitation, visibility, cloud height, atmospheric pressure, and other variables (NOAA 1998). ASOS platforms were designed to record observations using improved methods. These improvements include a windshield to minimize wind effects on the precipitation gauge, a heated tipping bucket to accurately measure liquid-equivalent mass, a redesigned tipping bucket to more accurately measure extreme precipitation, an aspirated housing unit for the ambient and dewpoint temperature measurement, an electro-optical method to determine wind speed, and redundant digital pressure transducers to record atmospheric pressure.

Data for the GHCN-Daily dataset are collected by various methods, ranging from handwritten paper forms to automated weather stations. According to Durre et al. (2010) and Menne et al. (2012), a robust QA process is needed due to the diversity in data collection methodologies. A vigorous and automated QA also allows for traceable and reproducible results, which is a necessary part of tracking the source of the data (Menne et al 2012). Durre et al. (2010) describe the QA tests applied to GHCN-Daily observations (19 in total) and group the tests into five general categories: 1) basic integrity checks, 2) outlier tests, 3) internal and temporal consistency checks, 4) spatial consistency checks, and 5) final consistency checks. Final consistency checks ensure that certain relationships still hold for each station record after all other QA tests are completed. For example, one final consistency check examines all monthly minimum temperatures to ensure none of them exceed the highest monthly maximum temperature values within the same month. All the QA tests are fully automated. Approximately 36 million (0.24%) of the more than 1.5 billion data points analyzed in the GHCN-Daily are identified as errors using these tests and removed from the dataset. The QA testing procedure has a false-positive rate of 1-2%. Thus, the GHCN-Daily dataset provides a vast amount of point observations that have been QA tested and covers a large portion of the land area across the globe, including the U.S.

There are other sources of climate data than GHCN-Daily. For example, the Food and Agriculture Organization of the United Nations (FAO) provides global climatic and real-time meteorological data (FAO 2016). FAO data are primarily used for agrometeorological data analysis and crop yield forecasting. Another source of climate data is the World Data Center for Meteorology (WDCMET; WDCMET 2016). The WDCMET acquires, catalogues and archives meteorological data from around the world to provide open access to meteorological data. WDCMET has offices in Asheville, North Carolina, Obninsk, Russia, and Beijing, China.

3.1.2 Radar, Remote Sensing, and Multi-Sensor Products

Accurate measurements of precipitation with fine spatial and temporal resolution are important for weather forecasters, scientists, decision makers, and others (Ebert et. al 2007; Duncan and Biggs 2012). As such, advances in in-situ data collection and methodologies for blending remotely sensed datasets have been and continue to be developed. These advances have led

to newly developed precipitation datasets that can be used for a range of applications. Quantitative Precipitation Estimation (QPE) is a method of approximating the amount of precipitation that occurred over a spatial area in a given amount of time. QPE can use direct measurements like observations from rain gauges or can be based upon data from a remote sensing platform, such as satellite retrieval information. In this section, we discuss advances in radar and remote sensing products and methods.

Radar was first used for a variety of reasons during World War II (WWII). As WWII continued into the mid-1940s, radar started to be used in weather applications. For example, the structure of Typhoon Cobra on 18 August 1944 was captured on radar, as the U.S. Navy Third Fleet accidently sailed directly through the storm (Figure 3.3).

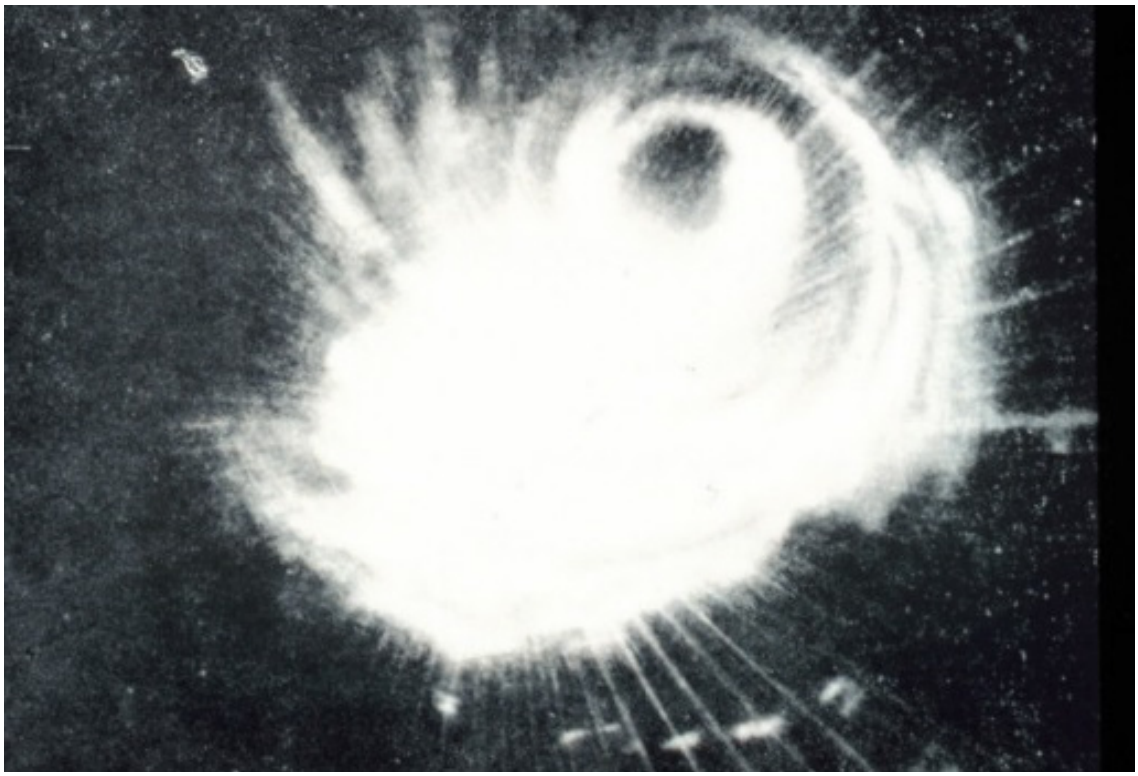


Figure 3-3 Structure of Typhoon Cobra on 18 August 1944. Radar Image from NOAA.

After WWII, weather radar usage continued to expand. From the 1950s into the early 1960s the USWB deployed weather radars across most of the U.S. east of the Rockies. The radars were considered second generation radars, meaning they could not detect the direction or velocity of wind within a storm (referred to as Doppler radar). With the advent of Doppler radar in the 1990s, meteorologists and atmospheric scientists were able to obtain high spatial and temporal resolution reflectivity data across the U.S. (Vazquez 2013). Improved methods to estimate precipitation totals were developed to go along with the higher resolution data. To estimate how much precipitation may be occurring, a relationship between the amount of energy reflected back to the radar and the intensity of precipitation was needed. This radar reflectivity-intensity relationship is called a Z-R relationship. A Z-R relationship tries to calculate how much precipitation is falling based upon the energy reflected back to the radar from the hydrometeors present in the air. Specific Z-R relationships exist based upon time of year, location across the U.S., type of weather system causing the precipitation, and other factors (List 1988; Fujiyoshi et al 1990; Vasquez 2013; Fabry 2015).

While Doppler radar and Z-R relationships helped improve precipitation estimation, it was not until the arrival of Dual-Polarization (Dual-Pol, Vasquez 2013) radar in the 2000s that radar precipitation estimation reached its current state. Dual-Pol radar sends out horizontal and vertical polarized electromagnetic waves, while Doppler radar only sends out horizontal waves. Dual-Pol radar can discriminate between solid (frozen) and liquid hydrometeors and is able to infer precipitation rates by comparing the energy reflected back to the radar with the energy from the original horizontal and vertical electromagnetic waves. This allows the individual interpreting the radar information to better estimate the character of the precipitation (e.g., snowfall, hail, graupel, rainfall) and the intensity at which the precipitation is occurring. Together, these qualities allow Dual-Pol radar to provide better instantaneous, and storm total spatial precipitation estimates when compared to Doppler radar (Vasquez 2013).

The Bollene-2002 Experiment (Delrieu et al. 2009) was developed to test how well non-Doppler radar could be used to estimate precipitation over the mountainous Cevennes-Vivarais region in France. The Regionalized and Adaptive Radar Data Processing for Hydrological Applications processing system was developed for this experiment. This system is coupled with several algorithms to determine precipitation type (convective vs. stratiform), clutter identification (non-precipitation targets), and a method to calculate reflectivity at the lowest scan level of the radar from reflectivities measured at radar scans farther aloft in the atmosphere. Five intense and long-lasting rain events from the Mediterranean Sea were used to test the validity of this project. Radar-estimated point precipitation totals were compared to precipitation observations measured from ground stations using the Nash coefficient, which assesses performance of a given estimator with respect to reference values. A Nash coefficient of one means there is perfect agreement between the estimator and the reference value. A Nash coefficient less than zero indicates the value of the estimator is as poor as a simple average of the reference values. In deep convective rain systems, the Nash coefficient was approximately 0.9. In shallow convective and frontal rain systems, the Nash coefficient was between 0.6 and 0.8. The Nash coefficient was higher in deep convective rain systems because the radar was able to sample the vertical structure of the storm better. The shallow convective and frontal rain systems did not have the same vertical extent as the deep convection, so the radar could not sample the storm vertically as well. The authors of the study attribute the positive results from an improved processing of spatial variations in the vertical profile of reflectivity.

Ground-based weather radar (e.g., Doppler and Dual-Pol radar) are often used to estimate QPE. Though QPE is most often deterministic, rather than probabilistic, there have been advancements involving probabilistic methods. For example, Kirsetter et al. (2015) used Doppler radar and point precipitation observations to create a Probabilistic Quantitative Precipitation Estimate at a five-minute temporal resolution and one-km spatial resolution. Rather than using a static Z-R relationship, the authors implemented a variable Z-R relationship, where the surface reflectivity factor Z is described by a distribution of possible R values, rather than a single static R value. By including a distribution of R values for different precipitation types instead of a fixed Z-R relationship, the authors were able to estimate uncertainty in precipitation totals from different precipitation types and to reduce systematic biases (Kirksetter et al. 2015).

NCEP produces a national precipitation product referred to as Stage IV. The national Stage IV QPE product is computed using hourly and six-hourly multi-sensor precipitation analyses (MPEs) produced by the 12 River Forecast Centers located around the continental U.S. (Figure 3.4). MPEs from the RFCs are based on radar and gauge data. Stage IV computations are run at NCEP 33 minutes past the hour, of each hour. There are a variety of temporal resolutions

available, including hourly, six-hourly, and 24-hourly accumulated precipitation totals. Data are available from January 1, 2002 to present.

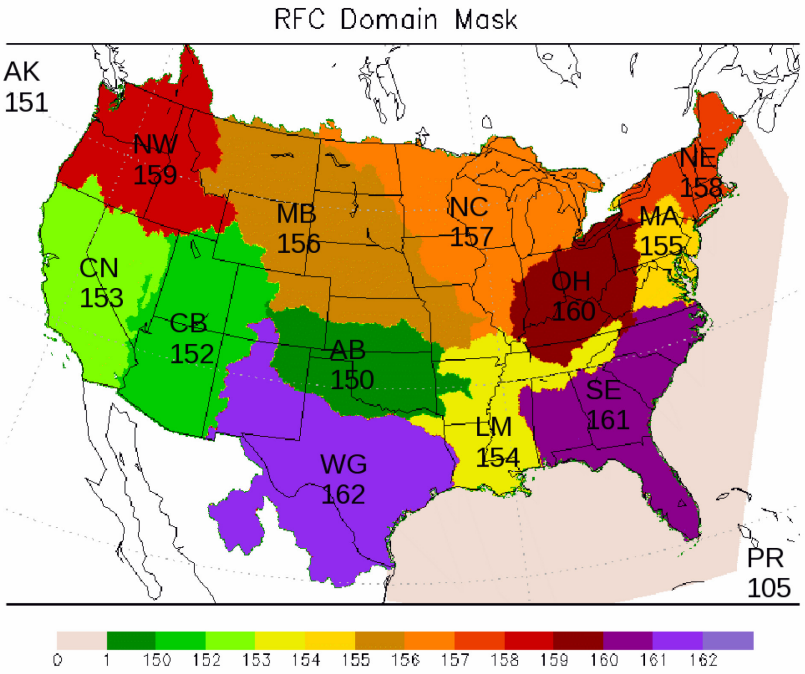


Figure 3-4 River Forecast Centers Multi-Sensor Precipitation Totals Domain Mask. Figure from http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/.

In developing and remote areas of the globe, as well as across oceans, satellite-based tools can provide estimates of precipitation that would otherwise not exist (Joyce et al. 2004; Huffman et al. 2007). In 2004, the Climate Prediction Center proposed a new technique to estimate global precipitation every half-hour using satellite data, the Climate Prediction Center MORPHing technique (CMORPH; Joyce et. al. 2004). This method uses passive microwave scans to estimate global precipitation on a 0.0727° latitude-longitude resolution. Motion vectors estimated from geostationary satellite infrared imagery are then used to propagate the estimated precipitation in space. A time-weighted linear interpolation scheme is used to estimate the shape, intensity, and path of the precipitation between half-hour measurements. Joyce et al. (2004) analyzed CMORPH precipitation estimates over Australia from December 2002 to May 2003 and found that, relative to passive and infrared estimates, CMPORPH estimates were substantially improved.

The Naval Research Laboratory developed a blended-satellite precipitation technique (NRL blended technique) that uses many of the different wavelengths measured using the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument, located aboard the Earth Observing System satellite (Turk and Miller 2005). In this technique, passive microwave and other sensors are used to filter out high cirrus clouds which are often mistakenly identified as light precipitation in satellite estimated precipitation. The NRL blended technique can be used to estimate precipitation during the day and at night.

The Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA; Huffman et al. 2007) is an example of another satellite precipitation estimation tool recently developed. This technique provides a method to combine 3-hourly multiple satellite

precipitation estimates with precipitation gauge analyses at a resolution up to 0.25º latitude-longitude. TMPA provides precipitation estimates for areas of the globe between 50ºN latitude to 50ºS latitude and is available from 1989 to present. Additional TRMM precipitation products are available through the National Aeronautical Space Agency (NASA) Goddard Earth Sciences Data and Information Services Center[2]. Ebert (2005) found that TMPA performed best in relatively heavy, convective, warm-season precipitation systems, while it tended to perform poorly in relatively light, cool-season precipitation systems that were mid-latitude in nature. Huffman et al. (2007) showed that TMPA was successful at recording when and where precipitation occurred at a sub-daily and daily time step but tended to over- and under-estimate the amount of precipitation that occurred, respectively.

The Program to Evaluate High Resolution Precipitation Products (PEHRPP; Arkin and Turk 2006) analyzed CMORPH, TMPA, NRL blended technique, and other satellite-based precipitation estimates and compared them to existing precipitation gauge data over the U.S. and Pacific Ocean. Results show the satellite precipitation estimates tend to overestimate warm season precipitation over the U.S. and underestimate precipitation over the tropical Pacific Ocean (Sapiano and Arkin 2009). CMORPH was found to have the highest correlation with gauge data over Australia and the United Kingdom with correlations as high as 0.7 (Tian et al. 2009). Tian et al. (2009) also showed that all the products missed some winter precipitation events across the U.S., likely due to low-level cloudiness which hampered the satellite's ability to detect precipitation.

3.1.3 Gridded Observations

Gauge observations remain critical for documenting the character of precipitation across global land areas (Chen et al. 2008). However, high resolution networks are still missing in some regions, particularly regions of complex terrain (Newman et al. 2015). Since the late 1990s, a number of gridded precipitation datasets have been developed based on point observations for use in both short-term and long-term analyses. Methods used for gridding historical observations include inverse distance weighting (IDW; Davis 1986) and Optimal Interpolation (Gandin 1963), among others. This section provides a review of some common gridded precipitation datasets used within the scientific community.

In 1997, Thornton et al. (1997) created an algorithm for generating daily precipitation fields (among many other variables) over large regions of complex terrain. The algorithm required digital elevation information and meteorological observations from stations. The authors demonstrated the application of this algorithm in the northwestern U.S. for a single year. Since then, however, the dataset has been extended[3] to all of the U.S. and parts of Canada and Mexico. This dataset is referred to as Daymet.

Maurer et al. (2002) released a long-term, hydrologically based gridded dataset of surface fluxes and states (including precipitation) across the U.S. and parts of Canada and Mexico. Data are available at a daily time step from January 1, 1949 to July 31, 2000. Precipitation totals in the dataset are derived from NOAA Cooperative Observer (COOP) stations with at least 20 years of data. The average station density at the time of development was one station per 700 km^2. Daily precipitation observations were gridded to the ⅛° resolution using the synagraphic mapping system (SYMAP) algorithm of Shepard (1984). Gridded precipitation

[2]https://disc.gsfc.nasa.gov/datasets?page=1&measurement=Precipitation%20Amount,Precipitation%20Rate&keywords=TRMM

[3] https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1219

totals were then scaled to match the long-term average of the PRISM climatology (Daly et al. 1994).

Members of NOAA's Climate Prediction Center (CPC) used three different objective analysis techniques to produce daily gridded precipitation datasets based on ~16,000 stations over the global land areas (Chen et al. 2008). Gridded daily precipitation data are available on a 0.25° latitude-longitude grid between January 1, 1948 and December 31, 2006. The authors employed two applications of IDW (Cressman (1959) and Shepard (1968) algorithms) and one method of OI (Gandin 1963). OI is different from other methods because the ratio of daily precipitation to climatology of precipitation is interpolated in space, rather than daily precipitation observations. A comparison analysis from Chen et al. (2008) showed that the OI method consistently performed the best among the three techniques, relative to independent gauge observations. Consequently, the CPC decided to use the OI technique to create the unified analyses of daily precipitation over global land areas.

Wuest et al. (2010) produced an hourly gridded precipitation dataset across Switzerland based on a simple methodology that combined daily point precipitation observations with hourly radar scans. More specifically, the authors temporally disaggregated daily point precipitation observations to the temporal distribution from radar scans, retaining the observed precipitation totals. Validation of the dataset indicated that errors in gridded estimates, relative to independent gauge observations, were larger in valleys than over plateaus due to the shielding effects of radar.

In 2013, Livneh et al. (2013) released a long-term hydrometeorological dataset based on applying the methods from Maurer et al. (2002) to approximately 20,000 NOAA COOP stations. Additionally, Livneh et al. (2013) improved on the Maurer (2002) dataset by extending the period of analysis (1915-2011) and increasing the spatial resolution (1/16°). As with Maurer (2002), monthly precipitation totals from the Livneh et al. (2013) dataset were scaled to match the long-term mean (1961-1990) from PRISM. Applications of the Maurer (2002) and Livneh et al. (2013) datasets generally include 1) studies that use the data directly to characterize a specific variable, 2) studies that use the data as an observational baseline, and 3) water and energy balance studies. In 2015, the Livneh et al. (2013) dataset was updated to include parts of Canada and all of Mexico. Methods applied to precipitation observations across the conterminous U.S. remained the same (Livneh et al. 2015).

Figure 3-5 shows the average one-day and three-day precipitation totals (at each grid cell) during the top 0.05% of precipitation events (N=58) between 1980 and 2011 from the Livneh dataset. The largest one-day and three-day precipitation totals exist along the Gulf of Mexico coast and a small region in southwestern North Carolina, along the Blue Ridge Mountains. The largest one-day (three-day) precipitation totals across the domain reach 120 mm (220 mm), or 4.7 inches (8.66 inches).

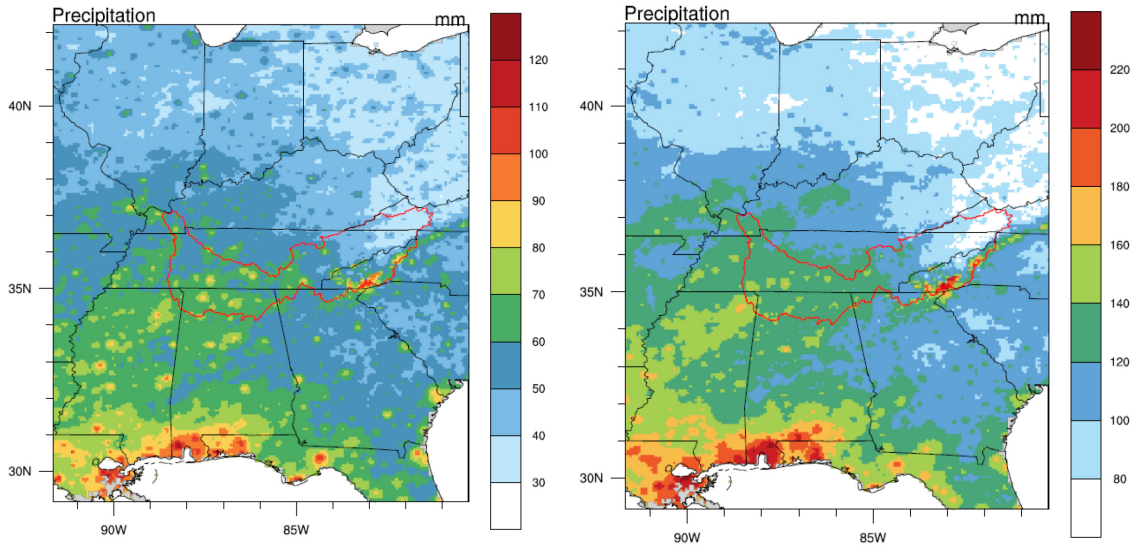


Figure 3-5 Average One-day (left) and Three-day (right) Precipitation Totals (mm) During Top 0.05% of Precipitation Events Across the Southeastern U.S. Between 1980 and 2011, Based on the Livneh et al. (2013) Dataset. The Tennessee River Valley Watershed is Outlined in Red.

Hou et al. (2014) developed a fine-scale gridded precipitation product, the Climatology-Calibrated Precipitation Analysis (CCPA), by combining daily precipitation observations from the CPC and NCEP's Stage IV precipitation data in near real-time. The six-hourly accumulated precipitation estimates in CCPA were developed by first aggregating and upscaling hourly gridded Stage IV estimates (4 km × 4 km) to daily totals on the same grid as the CPC product (0.125° latitude-longitude). Then, a linear regression relationship was developed between the upscaled Stage IV data and CPC data (at each grid separately). The linear regression relationship was used to develop adjustment factors for the Stage IV data to make its climatology more closely match the climatology of the CPC product. Finally, the adjusted estimates of Stage IV from the linear relationship were downscaled back to the original resolution of the Stage IV data (4 km × 4 km). These data are available over the continental U.S. from June 2002 through July 2009. The primary limitation of this methodology is related to heavy and extreme precipitation events in the Stage IV dataset. The Stage IV procedure tends to underestimate these heavy precipitation events, an artifact that is not corrected during the CCPA analysis procedure. As a result, the CCPA dataset underestimates some of the heavy to extreme precipitation events as well (Hou et al. 2014).

Although there are many benefits to using gridded precipitation products, data homogeneity remains a concern. Inhomogeneity in station data, and consequently gridded products, can result from changes in station location, instrumentation, observation practices, and data processing methods (Bukovsky and Karoly 2007). Guentchev et al. (2010) suggest performing stringent homogeneity testing of gridded products and that users of any gridded dataset understand the reason for its generation and limitations, in order to use the dataset in a manner consistent with its construction.

3.1.4 Gridded Ensemble Datasets

Deterministic approaches to quantitative precipitation estimation have many limitations. To address this, Clark and Slater (2006) developed a stochastic modeling framework to generate

ensemble gridded fields of precipitation in complex terrain on a 2 km × 2 km grid. The framework utilizes locally weighted regression to predict spatial variability in precipitation occurrences and totals using topographic attributes from station locations as explanatory variables. At each time step, regression models are used to estimate the conditional cumulative distribution function (CDF) of precipitation at each grid cell based on daily precipitation totals from gauges within a station network. Correlated random spatial fields (e.g., a spatial map of random, correlated percentiles) are used to sample values from the precipitation CDF at each grid cell, resulting in a spatial map of precipitation totals. The random sampling is used to generate precipitation ensembles. This framework was applied to point precipitation observations across western Colorado to produce a gridded ensemble precipitation product from 1980 to 2003. The authors found that the framework adequately reproduced the climatological gradients and spatial correlation structure in precipitation in this region (Clark and Slater 2006).

More recently, Newman et al. (2015) extended the work of Clark and Slater (2006) and developed a daily, gridded, observation-based ensemble of precipitation and temperature. The dataset, which includes 100 ensemble members, covers the contiguous U.S. at ⅛° latitude-longitude resolution for the years 1980 to 2012. Figure 3-6 shows the elevation used to develop the dataset. Like Clark and Slater (2006), Newman et al. (2015) used multivariate, locally weighted logistic regression to estimate precipitation and the probability of precipitation, at each grid cell, using the 30 closest observations.

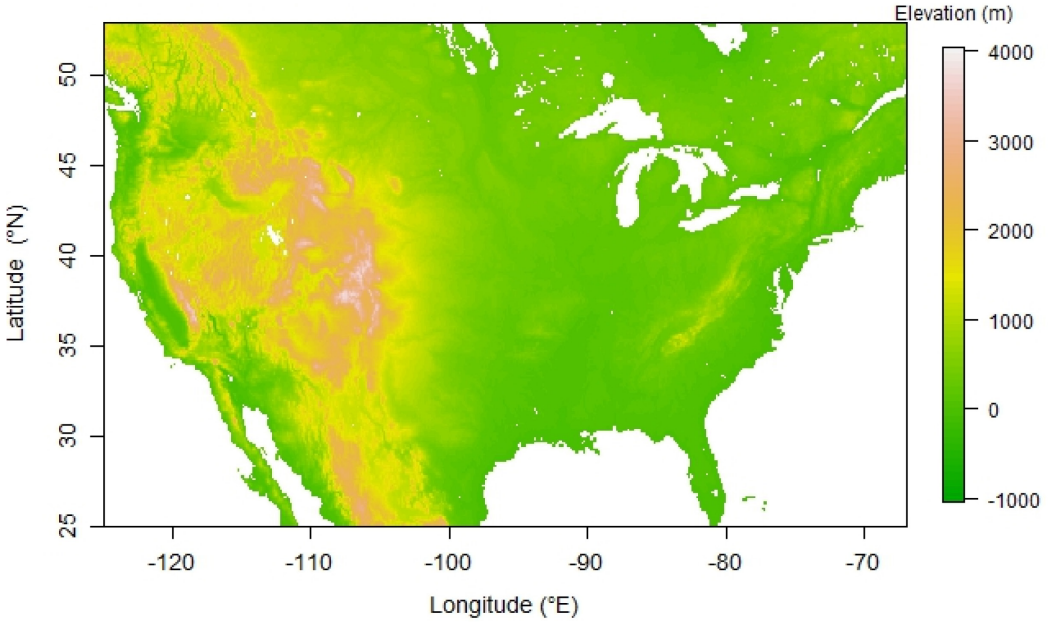


Figure 3-6 Elevations Used to Develop the Newman et al. (2015) Dataset

Figure 3-7 shows an example of daily precipitation estimates on April 8, 1983 from four of the 100 ensemble members. According to the analysis presented by Newman et al. (2015), the ensemble estimates a high probability of event occurrence given an observed event greater than zero mm. The authors also indicate that the ensemble can distinguish between events and nonevents. Using a leave-one-out cross validation approach, the authors show that the ensemble slightly underestimates precipitation in the southeastern U.S. and overestimates precipitation in the western U.S. Treatment of aleatory variability is provided by this ensemble and enables users to more properly account for uncertainty in precipitation fields. For more

details on the generation of the gridded ensemble dataset, see Newman et al. (2015) and Appendix A.

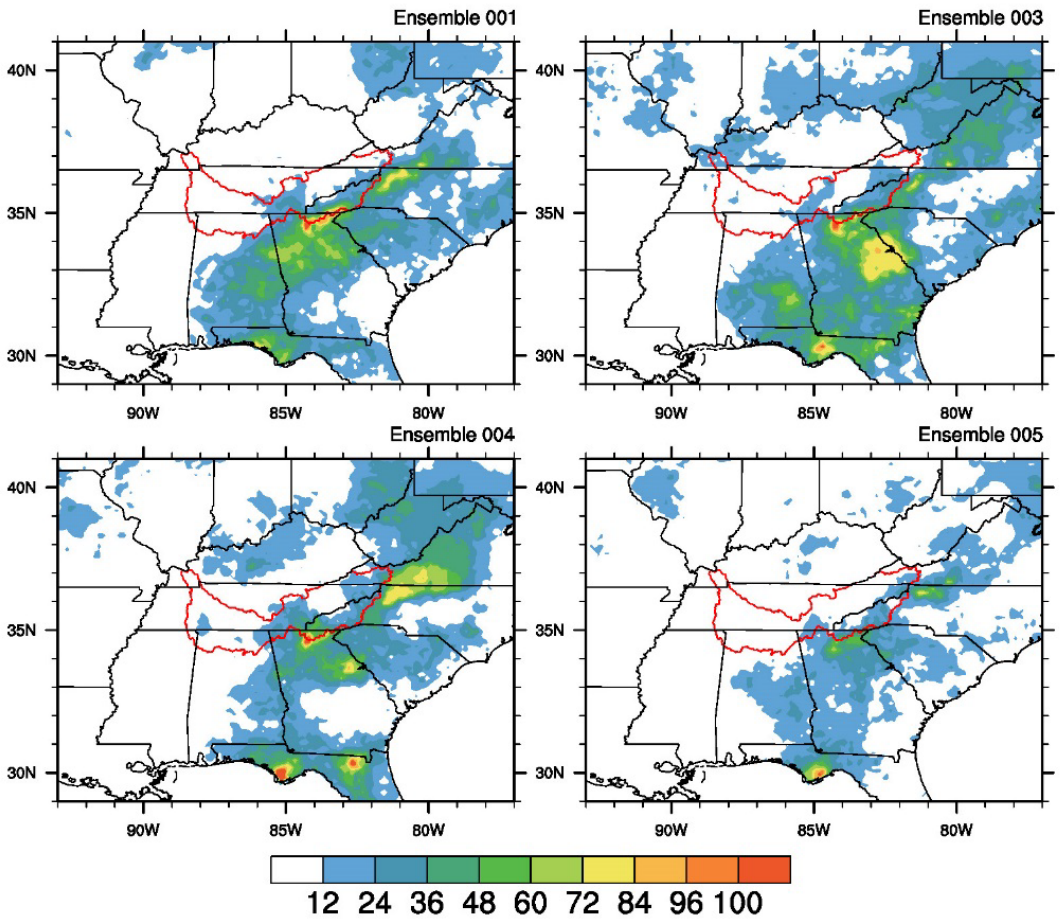


Figure 3-7 Daily Precipitation Totals (in mm) for April 8, 1983, based on Four Ensemble Members of the Newman et al. (2015) Dataset. The Tennessee River Valley Watershed is Outlined in Red.

3.1.5 Reanalysis Datasets

Atmospheric retrospective analyses, or reanalyses, integrate a number of observing systems with numerical model simulations to create temporally and spatially consistent, long-term, systematic portrayals of the climate system (Kennedy et al. 2011; Bosilovich 2013). Reanalyses provide an additional source of data for use in exploring the earth-system, including extreme precipitation events (Holman and Vavrus 2012). These datasets can be used to fill gaps in observation records and to diagnose variables that are not measured directly (Kennedy et al. 2011). In this section, we discuss some of the widely available reanalysis datasets that include daily or sub-daily precipitation estimates.

The National Centers for Environmental Prediction's (NCEP) North American Regional Reanalysis dataset (NARR; Mesinger et al. 2006) represents a relatively new, higher-resolution reanalysis dataset that is based on the Eta Model, a numerical weather prediction model, and the Eta data assimilation system (Mesinger et al. 2006). The reanalysis fields are available from October 1978 to present every 3 hours at a 32-km horizontal resolution (Bukovsky and Karoly

2007; NCEP 2016). The NARR system assimilates precipitation observations, which means that the precipitation fields are not entirely determined by large-scale variables and physical parameterizations (Bukovsky and Karoly 2007). The assimilated precipitation observations come from various data sources, depending on geographical region.

Bukovsky and Karoly (2007) evaluated the performance of precipitation from three different reanalysis datasets relative to CPC, including NARR. The authors found that average annual precipitation estimates from NARR agree well with estimates from CPC over the continental U.S. NARR accurately captures the summer maximum over the eastern half of the U.S. and the winter maximum over the western half of the U.S. The authors demonstrated the performance of NARR during an extreme precipitation event in Las Vegas, Nevada during July 1999. NARR captured the three-hour average precipitation rate during the event, as well as the spatial pattern of precipitation (as compared to radar imagery).

Holman and Vavrus (2012) explored the timing and magnitude of extreme precipitation events in the Midwest using NARR. Relative to historical observations at a single point, NARR accurately captured the seasonal cycle of daily precipitation and extreme precipitation events. NARR was also able to accurately capture the timing of the wettest 10% and wettest 5% of daily precipitation totals in Madison, WI within the period of 1981-2000 (Holman and Vavrus 2012).

NASA released the Modern-Era Retrospective analysis for Research and Applications (MERRA) in 2010. MERRA is based on the Goddard Earth Observing System Data Analysis System and covers the same period as NARR: 1979 through the present. The native resolution of MERRA is ⅔° latitude × ½° longitude in the horizontal. Three-dimensional output is available every six hours, while two-dimensional output is available hourly. Kennedy et al. (2011) compared precipitation fields from MERRA with observations recorded at the Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Southern Great Plains site during the period 1999-2001. Results indicate that MERRA has a negative monthly bias of -22.2 mm over this period. However, MERRA can capture the monthly variability of precipitation averaged over the Southern Great Plains ARM domain. Daily precipitation totals from MERRA exhibit greater disagreement with ARM observations than monthly totals over the same time period (Kennedy et al. 2010).

Bosilovich (2013) analyzed MERRA summertime precipitation totals relative to CPC data across the U.S. for the period 1979-2010. The author found that during this time MERRA displayed a high summer bias in the Southeast and Great Plains regions, along with a dry bias in the Midwest. In addition, MERRA precipitation totals contained trends that were not supported by CPC observations. However, seasonal mean precipitation totals from MERRA were positively correlated with CPC across much of the U.S.

NCEP developed the Climate Forecast System Reanalysis (CFSR) dataset in August 2004. CFSR represents an improvement over earlier products (e.g., NCEP-NCAR 40-Year Reanalysis (Kalnay et al. 1996) and NCEP-DOE reanalysis (Kanamitsu et al. 2002)) because it is the product of a coupled ocean-atmosphere-land system (Higgins et al. 2010). CFSR covers the period 1979 to present and is available at a higher resolution than previous products (approximately 0.5° latitude-longitude). Higgins et al. (2010) showed that seasonal biases in mean daily precipitation totals in CFSR are generally weaker in magnitude than earlier products. Furthermore, the authors showed that CFSR better simulates the annual cycle of the probability of precipitation (greater than 0.25 mm) across the U.S., as compared with previous reanalysis products. According to Higgins et al. (2010), the daily precipitation statistics in CFSR are, in general, improvements over the earlier versions.

The North American Land Data Assimilation System (NLDAS), a multi-institutional system, provides retrospective and real time analyses to support land surface modeling (Cosgrove et al. 2003). Hourly precipitation totals are available through NLDAS, which are derived from the NCEP CPC daily gauge-based precipitation data. First, daily CPC totals are adjusted to match PRISM and interpolated to the 1/8° NLDAS grid using a least squares distance weighting scheme. Then, the daily ⅛° precipitation data are temporally disaggregated into hourly totals based primarily on hourly temporal weights obtained from the NEXRAD Stage II precipitation (Cosgrove et al. 2003). Luo et al. (2003) explored differences in hourly precipitation totals between NLDAS and station observations from the Oklahoma Mesonet and Atmospheric Radiation Measurement/Cloud and Radiation Testbed in the southern Great Plains region between September 1996 and September 1999. The authors found that precipitation totals between the two datasets do not agree at hourly timescales. Station observations generally produced higher hourly precipitation rates than those simulated by NLDAS. However, NLDAS precipitation totals did show improved agreement with station observations on longer time scales, such as five days and monthly (Luo et al. 2003). The authors attribute this to the interpolation scheme. Nan et al. (2010) explored spatiotemporal similarities between NLDAS precipitation totals and NEXRAD Multisensor Precipitation Estimator (MPE) estimates between January 2002 and December 2007 over the Ohio River Basin. The authors used three spatial metrics to characterize the statistical behavior of spatial hourly and daily precipitation data. The authors found large dissimilarities in magnitude and spatial patterns between NLDAS and NEXRAD MPE. The authors suggested that the differences in spatial patterns arise because NLDAS spatial patterns are primarily determined by gauge observations, while the NEXRAD MPE patterns are determined by radar observations.

While reanalysis datasets have many benefits over traditional datasets, such as spatial and temporal consistency and the availability of unmeasured variables, the data represent estimates of the real atmosphere. Using reanalysis data may add inaccuracies and uncertainty to a study if the data are not used with caution, particularly when using precipitation data (Bukovksy and Karoly 2007; Bosilovich 2013).

3.2 Methodology Improvements

Beyond improvements to datasets (in-situ, gridded, and reanalyses), scientists and engineers have made major improvements to methods used to understand orographic precipitation. The application of numerical weather modeling to research and operational analyses has widely expanded, as computing capabilities have increased. Statistical methods to estimate precipitation have also improved over time. In this section, we discuss recent advancements in methods used to explore and better understand orographic precipitation.

3.2.1 Numerical Weather Models

Numerical weather prediction and modeling efforts have greatly expanded during the past 30 years. In some cases, modeling efforts have been combined with intense observational programs in order to validate numerical simulations (e.g., the NOAA Hazardous Weather Testbed Spring Experiment; Coniglio et al. 2010). As computational speed and power have improved, numerical modeling efforts have followed. For example, scientists at NCEP have reduced the horizontal resolution of the North American Mesoscale model (NAM) from 80 km in 1993 to 12 km in 2001 (Kain et al. 2008). In this section, we discuss recent applications of numerical weather models to orographic storm analyses, with specific emphasis on the Weather Research and Forecasting (WRF) model.

The Weather Research and Forecasting (WRF) project is a multi-institutional effort to develop an accurate and efficient mesoscale forecast and data assimilation system that works across a range of scales and computer platforms (Michalakes et al. 2001). The model, which incorporates advanced numerics and data assimilation techniques, is intended to advance the understanding and prediction of mesoscale precipitation systems and improve relations between the research and operational forecasting communities. In 2001, the model was a candidate to replace existing forecast models, such as the PSU/NCAR Mesoscale Model (MM5), the Eta model at NCEP, and the RUC system at the Forecast Systems Laboratory (Michalakes et al. 2001). The first release of WRF 1.0 happened on November 30, 2000. Since then, the modeling system and development and user communities have widely expanded; the user community now includes over 30,000 users in over 150 countries[4]. The latest version of the WRF modeling system is Version 3.8.1[5]. Developers of WRF 1.0 were primarily focused on simulating spatial scales between 1 and 10 kilometers (Michalakes et al. 2001). However, some more recent research applications involved horizontal scales of 10-100 meters (Moeng et al. 2007).

While the WRF model can be used to simulate idealized experiments such as flow around a barrier and baroclinic waves, most applications involve simulating previous meteorological events of interest, such as convective precipitation events (Weisman et al. 2008; Trentmann et al. 2009), frontal systems (Medina et al. 2005; Viale et al. 2013), and landfalling atmospheric rivers (ARs; Leung and Qian 2009; Smith et al. 2010; Hughes et al. 2014). These simulations are often designed to gain a better understanding of the physical processes responsible for heavy and/or extreme precipitation. For example, Smith et al. (2010) used the WRF model to explore the three-dimensional structure of water vapor fluxes during a December 2005 AR event as it traversed the Coastal Ranges and the Sierra Nevada Range in California. This storm was selected because it caused major flooding along the Russian, Napa, and Truckee Rivers. The authors demonstrated that the WRF model could capture the structure and strength of the AR over the ocean, relative to Special Sensor Microwave Imager (SSM/I) data, and over land, relative to GPS integrated precipitable water (IPW) receivers.

Smith et al. (2010) also used the WRF model to test the sensitivity of water vapor fluxes to different mountain topography configurations during the 2005 AR event. The authors explored the impact of topography on the drying ratio, which is the ratio of water vapor removed from topography to the initial water vapor content. Their results showed that during this event, the Coastal Range and Sierras had drying ratios of 25% and 28%, respectively (Smith et al. 2010). The authors also demonstrated the impact of removing all terrain over the West Coast. Results from those simulations showed that an AR still developed in the absence of topography, however the magnitude was weaker than in the control simulation.

As with Smith et al. (2010), Flesch and Reuter (2012) used the WRF model to explore two heavy precipitation events that impacted Alberta, Canada, during June 2005, and the role of terrain during these events. The authors used an “off-the-shelf” version of WRF (i.e., default parameters) to demonstrate how a student or non-meteorologist might use the model. The authors demonstrated the impact of topography on precipitation totals during the two events by comparing simulations with actual terrain of the northwestern U.S. to simulations where the coastal mountains were eliminated. Their results suggested that a reduction in mountain elevation leads to reductions in the maximum precipitation totals over the mountains and

[4] http://www.wrf-model.org/index.php

[5] http://www2.mmm.ucar.edu/wrf/users/downloads.html

foothills, along with altered synoptic characteristics. The simulations also revealed that precipitation totals outside the mountains were not impacted.

Lackmann (2013) used the WRF model to explore how enhanced water vapor and condensation under future climate change may influence mesoscale dynamics during heavy precipitation events. The author focused on the May 2010 severe flooding event that struck parts of the south-central U.S., where the central and western portions of Tennessee were severely impacted. The synoptic pattern during this event involved a slowly evolving upper-level ridge (trough) over eastern (western) North America, with persistent lower-level southerly flow from the Yucatan Peninsula and Gulf of Mexico. The pre-cold-frontal low-level jet accompanied the warm, moist airstream on the western boundary of the cyclonic warm sector (Lackmann 2013).

Using three nested domains, Lackmann (2013) performed a control simulation of the observed event and used a pseudo-global warming approach (Kawase et al. 2009) to estimate a future version of the event. The author also performed an additional simulation to isolate the role of terrain in generating the precipitation during the historical event. The author concluded that terrain, particularly the Mexican Plateau, played a prominent role in this historical precipitation and flooding event. Precipitation totals in the no-terrain simulation were dramatically reduced. Maximum precipitation values in the no-terrain simulation were less than one-third of the precipitation totals from the control simulation.

Using the WRF model, Hughes et al. (2014) simulated a January 2010 AR event that resulted in extreme precipitation across Arizona. The authors explored the sensitivity of modeled precipitation across Arizona to terrain height across Baja California. The WRF simulations showed that larger amounts of precipitation occurred across the Baja, with lower amounts downstream over Arizona, in the presence of raised topography. Increasing the terrain of the Baja California resulted in less water vapor and water vapor transport downstream (over Arizona). Similarly, Mahoney et al. (2016) used the WRF model to investigate the role of elevation in the 2013 Colorado floods. The authors explored the influence of model terrain across Colorado on precipitation by running simulations where 1) all terrain was removed, 2) all terrain was set to a constant height of 1610 m, 3) all terrain was set to half the actual height, 4) all terrain was capped at 1610 m, and 5) all terrain height was increased by 25%, and comparing results with a control simulation. Simulations with terrain height increased by 25% resulted in decreased maximum precipitation of 10-20%, although the location of the maximum precipitation remained the same.

Beyond observational and theoretical applications, the WRF model has also been used in the context of operational forecasting (White et al. 2012) and in support of decision making (Sankovich et al. 2012; Mueller and Mahoney 2016). During January 2009, record flood storage from a winter storm seriously affected Howard A. Hanson Dam, located near Seattle, Washington. After the event, USACE staff discovered two depressions had formed on the upstream face of the right abutment of the dam, resulting in a major increase in the risk of catastrophic flooding downstream of the dam. Soon after, NOAA implemented a rapid response effort to help communities at risk. These efforts consisted of gathering information on forecast needs, implementing new technologies, synthesizing specialized inputs to the forecast process, and collecting feedback (White et al. 2012). New technologies were implemented in the form of numerical modeling tailored to the West Coast, among others. Scientists implemented Version 3.0 of the WRF model to produce mesoscale numerical forecasts to aid in early AR detection.

Reclamation's Denver Office has also started to explore the benefits and applications of WRF simulation to hydrologic analyses for dam safety. For example, Sankovich et al. (2012) utilized the WRF model in a study of Green Mountain Dam. The authors explored the ability of WRF to reproduce two large precipitation events near Green Mountain Dam watershed, in Colorado, and to perturb model parameters for potential climate change studies. The authors concluded that the WRF model was able to simulate extreme events near Green Mountain Dam, a region of complex topography, and is beneficial to extreme storm analysis required for dam safety hydrologic hazard analyses.

Mueller and Mahoney (2016) used WRF to generate a series of precipitation ensembles for several historical precipitation events that impacted the Taylor Park Dam watershed in Colorado for use in a Reclamation hydrologic hazard analysis. Each precipitation ensemble includes 18 members, which were created using variations in microphysics schemes, lateral boundary conditions, and stochastic perturbations. Figure 3.8 displays the total accumulated precipitation surrounding the Taylor Park Dam watershed over the period June 16-20, 1995, based on six different simulations with varying initial and lateral boundary conditions from the Global Ensemble Forecast System[6] (GEFS). In the future, precipitation from the ensemble members will be used by Reclamation staff to estimate uncertainty in historical flooding events in the watershed.

[6] https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-ensemble-forecast-system-gefs

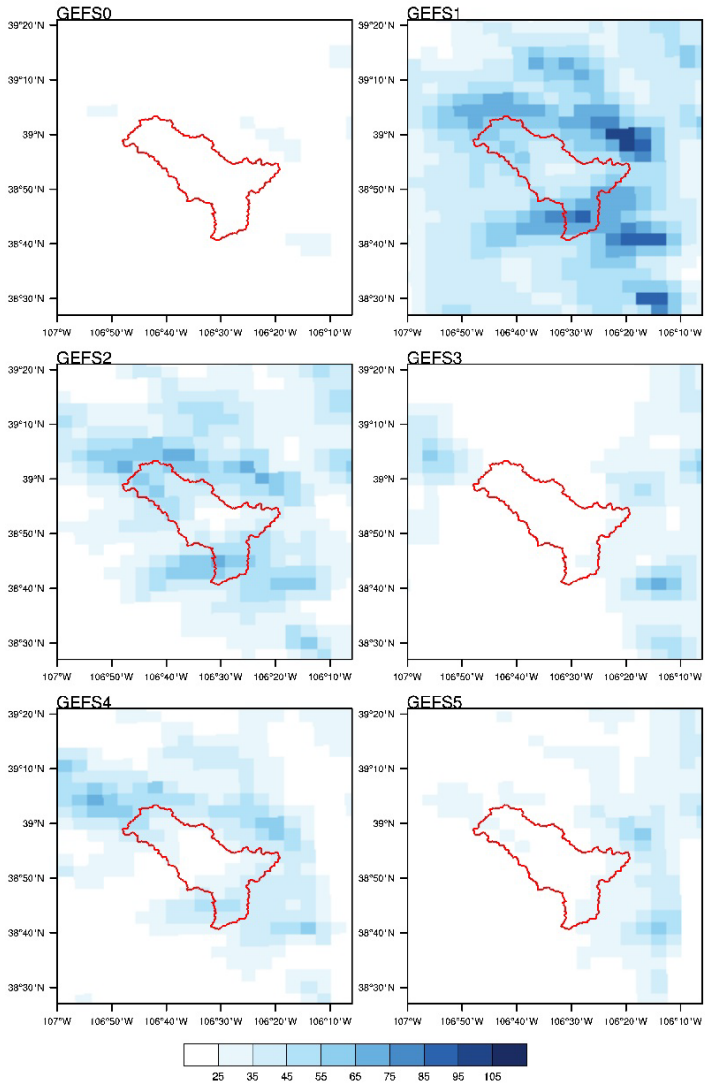


Figure 3-8 Total Accumulated Precipitation (in mm) from 16-20 June 1995 from Six WRF Simulations with Varying Inititial and Boundary Conditions from the GEFS. The Taylor Park Dam Watershed is Outlineed in Red.

3.2.2 Statistical Methods

There have been many kinds of statistical models used to estimate orographic precipitation over the years, a brief summary of which is provided in this section. Most of these statistical methods share some common assumptions, the most frequent assumption of which is that the precipitation totals at a given location are proportional to the ground surface elevation and other physiographic variables at that location, such as slope, aspect, and distance to the coast, to name a few. These assumptions have been applied to a wide array of models for the common goal of estimating precipitation in orographic regions.

The most widely used geostatistical model is kriging, which has numerous associated extensions including ordinary kriging, co-kriging, kriging with external drift, indicator kriging, to name a few (see Chiles and Delfiner 1999). In kriging theory, the prediction at an unobserved

location is calculated as a weighted linear combination of the available data; the weights are calculated in such a way to minimize the expected squared error (Cressie 1990).

Goovaerts (2000) presented three kriging models (simple kriging with varying local means, kriging with external drift, and collocated co-kriging) to illustrate the use of elevation in spatial rainfall estimation. Diodato (2005) showed that including terrain elevation data and a topographic index in an ordinary co-kriging model of average annual precipitation vastly improved results from an ordinary kriging model, because co-kriging can account for the properties of the landscape. Haberlandt (2007) used kriging with external drift and indicator kriging with external drift to interpolate hourly rainfall using rain gauges, radar estimates, and elevation. Masson and Frei (2014) compared linear regression to kriging with external drift, using predictors of topographic height and slope at several spatial scales, a stratification by types of circulation classification, and a predictor for wind-aligned topographic gradients. They found that a single predictor (i.e., elevation) may be sufficient, because kriging accounts for spatial autocorrelation, and therefore can be more effective than using many predictors.

Manz et al. (2016) explored several different methodologies – linear regression, residual inverse distance weighting, ordinary kriging, residual ordinary kriging, and kriging with external drift – to blend the mean monthly climatologies of the Tropical Rainfall Measuring Mission (TRMM; Kummerow et al. 1998) with rain gauge measurements in a tropical region of the Andes Mountains. Cross-validation and catchment water balances (i.e., runoff ratios) were used to validate the different methodologies. However, the authors found that there was very high uncertainty and poor performance in regions of very low station density. Overall, Manz et al. (2016) determined that the quality of the gauge network is the most critical aspect of selecting a methodology to blend precipitation data in regions of complex topography.

Castro et al. (2014) developed a two-step process for spatial estimation of daily precipitation in mountainous terrain with scarce data networks. The first step used IDW on observed precipitation occurrence; the second step fitted separate linear regression models for windward and leeward grid cells. The main methodological improvement made by the authors was classifying each grid cell by its "terrain orientation," i.e., windward or leeward oriented slopes. The authors noted that separating and classifying the grid cells by terrain orientation significantly reduced the overall bias in the precipitation estimates and contributed to the skill of the model in capturing the spatial discontinuity of precipitation in mountainous terrain.

Nonparametric methods, such as local polynomials (LP; Cleveland 1979), have been shown to capture local nonlinearities in spatial precipitation data while offering a competitive alternative to parametric methods. In the LP methodology, *k*-nearest neighbors (i.e., observations) of the point of interest are identified. Weighted least squares is then used to fit a local regression of order *p*, which is used to provide an estimate of the function at the desired point. The variables, *k* and *p*, are dynamic and can be determined by Generalized Cross Validation (Craven and Wahba 1978). LP can be preferred to traditional parametric methods because LP is purely data-driven and thus does not make any assumptions about the true distribution of the data. LP has been successfully used for spatial estimation of precipitation (Rajagopalan and Lall 1998). For example, Verdin et al. (2016) found that LP and ordinary kriging were comparable in their ability to blend pentadal (5-day) and monthly total satellite-based precipitation estimates with gauge observations in a vast, mountainous Central and South American study region. The LP and ordinary kriging models used latitude, longitude, and elevation as predictor variables, and were fitted for specific pentads and months in 2009. The residuals (i.e., gauge observation minus satellite estimate) were calculated for each gauge location, then the LP and ordinary kriging models were used to estimate the residual at the spatial resolution of the satellite

estimate. The estimated residual field was added to the original satellite estimate to produce the blended product. Leave-one-out cross-validation showed that LP and ordinary kriging performed similarly. However, the authors provided specific scenarios where one model would be preferred over the other (Verdin et al. 2016).

Other statistical methods for orographic precipitation estimation can involve merging (or "blending") datasets of different spatial resolutions and with different spatial support. Radar- and satellite-based gridded products, reanalyses, numerical model output, and networks of gauge observations are the most commonly blended datasets. Satellite-derived precipitation products take advantage of the relationship between satellite-observed infrared brightness temperatures, microwave radiation, outgoing longwave radiation, and precipitation intensities to estimate gridded precipitation values (e.g., Arkin and Ardanuy 1989). Because these products report areal averages, they tend to underestimate extreme events (AghaKouchak et al. 2011). Hence, there has been much research on methods to blend data sources of different spatial and temporal scales for an improved gridded precipitation product, a brief summary of which is provided below.

The first widely available blended product was the Climate Prediction Center Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997), which incorporated satellite-derived estimates, rain gauge analyses, and numerical model output. Through the years, the Global Precipitation Climatology Project (GPCP) has offered a suite of products, and has continually improved upon them to implement increasingly fine temporal ranges (monthly, pentadal, daily, respectively) and spatial resolutions (2.5°×2.5°, 2.5°×2.5°, 1.0°×1.0°, respectively) and unbounded global scales (Adler et al. 2003; Xie et al. 2003; Huffman et al. 2001, respectively).

Xie and Xiong (2011) provided a novel approach for blending datasets, wherein they matched the probability density functions of satellite-based precipitation estimates with those of rain gauge analyses to correct for bias. Then an optimal interpolation technique was used to combine the bias-corrected satellite estimates with gauge data. The authors demonstrated this methodology on data over all of China for a five-month period from April to September 2007.

Vila et al. (2009) developed a stepwise method for merging daily TRMM satellite estimates and gauge data for the South American continent, which experiences intense orographic precipitation due to vast coastal mountain ranges and warm moist air from the intertropical convergence zone. The authors used a combination of additive and multiplicative bias correction schemes to minimize the overall bias. The resulting satellite-gauge products were tested for different seasons and densities of station networks. Rozante et al. (2010) presented a similar blending technique, known as MERGE, which used the Barnes objective analysis method (Barnes 1973) to interpolate and combine TRMM satellite estimates with gauge data over the same South American study region. The MERGE algorithm was validated over the entire continent, but performance was most notable in subsamples of the continent where gauge networks were sparse.

Sinclair and Pegram (2005) use Conditional Merging (see Ehret 2003) to blend radar-based precipitation estimates with rain gauge data. This technique uses ordinary kriging to obtain the mean field of the rain gauges, while the bias of the interpolated values is reduced by using the spatial structure as reported by radar. Bárdossy and Pegram (2013) presented two copula-based methodologies – Gaussian copulas and unsymmetrical v-copulas, both which use elevation information – for spatially interpolating daily- to yearly-scale precipitation data. Simply put, a copula can be thought of as a link function or a transformation of the data to be modeled. Copulas can ease the modeling complexities when considering heavy-tailed (e.g., precipitation)

dependent variables by transforming the data to the space of its (uniformly distributed) cumulative distribution function, which is much more well behaved than the data in its original space. Copula theory is rooted in the fact that it is possible to express any multivariate joint distribution in terms of a univariate marginal distribution (Sklar 1959).

Bayesian methods are attractive because they robustly quantify the uncertainty associated with the process being modeled. Bayesian kriging, a specific application of Bayesian methods, has long been used in a variety of fields (e.g., Omre 1987). Verdin et al. (2016a) used Bayesian kriging (see Banerjee et al. 2004) to blend the Climate Hazards group InfraRed Precipitation (CHIRP; Funk et al. 2014) with a large network of in situ observations over Central America and northern South America. The in-situ observations were modeled as a function of the CHIRP estimate for the containing grid cell. Uncertainty was quantified by invoking computationally intensive Markov chain Monte Carlo sampling routines on the model parameters, which resulted in thousands of parameter values (known as a *posterior distribution*). The posterior distributions of model parameters were used to generate a posterior distribution of precipitation at each grid cell.

Bayesian spatiotemporal modeling is also an effective method for blending datasets of distinctly different spatial scales. Jin et al. (2014) applied the Bayesian spatiotemporal model form of Le and Zidek (2006, chap. 10) to blend satellite-based estimates with rain gauge observations in data-sparse southwestern Canada. Sansó and Guenni (1999) developed a predictive Bayesian spatiotemporal model for monthly rainfall amounts at 80 stations in Venezuela. Additional examples of “Bayesian melding” are presented in Fuentes and Raftery (2005) and Berrocal et al. (2010), though their focus is on air quality data.

No one statistical model is inherently better than the other, as each research question requires in depth analysis before proceeding. Careful attention should be paid to the spatial and temporal properties of the available data, the planned use of the resulting orographic precipitation data, any assumptions that would be violated, and which statistical method would best represent the data.

3.2.3 Additional Methods

Previous sections highlight advancements in the application of numerical weather modeling and statistical methods to estimate precipitation in orographic regions. However, many scientific advancements include a combination of tools and disciplines to create a synergistic improvement, rather than a gain in just one subject of interest. These collaborative improvements can be difficult to narrow down to a single genre or field of study, so we discuss them together in this section.

Isopercental analysis is a method for spatially interpolating and distributing a parameter whose behavior is spatially non-linear in the original units of measurement, such as precipitation. Isopercental analysis relies on the availability of prior regional precipitation information, which are used as base maps. The prior regional precipitation information often takes the form of gridded datasets, such as mean annual maxima, or 10-year or 100-year precipitation totals for specific durations. Isohyets are produced by comparing (i.e., dividing) a historical precipitation event of interest to one of the base maps (producing unitless spatial maps). These scaled historical events can then be transposed over a watershed of interest. After transposition, the scaled historical event is multiplied by the base map to produce transposed estimates of precipitation over the watershed of interest. Isopercental analysis represents a way of

characterizing a previous storm relative to historical extreme values for the same area (Micovic et al. 2014).
The Mediterranean region is a complex mix of a rich moisture source (e.g., Mediterranean Sea), strong synoptic-scale storms, and local topographic barriers. This combination can lead to localized flooding due to heavy precipitation events. Rudari et al. (2004) developed composite maps of 500-hPa geopotential height, surface wind, and water vapor flux in an effort to determine under which synoptic-scale regimes heavy precipitation events occur. They discovered that local topography helped to create secondary mesoscale features that aided in the development of localized heavy precipitation.

Wright et al. (2013) explored spatial and temporal variability in extreme precipitation, and its relationship to heterogeneous land cover and urban drainage networks. This study used the Hydro-NEXRAD (Wright et al. 2013) system, a 10-year, 15-min, 1 km^2 radar estimated rainfall dataset, to research how land cover and urban drainage networks affect the variability of a flood response. This project combined stochastic storm transposition (SST) with a catalog of storm events from the Hydro-NEXRAD system to estimate the intensity-duration-frequencies (IDF) of precipitation events over different watersheds in the region surrounding Charlotte, North Carolina. The SST method used in this study is from Foufoula-Georgiou (1989). The conditional probability that a storm of given characteristics (size, shape, duration, depth of rainfall) will occur at a location within the watershed is estimated using a multivariate spatial point process framework. A comparison of return periods for rain rates between Wright et al. (2013) IDF estimates using SST and those from NOAA Atlas 14 are shown in Figure 3.9. Wright et al. (2103) concluded that the SST methodology employed in their study produced comparable results to NOAA Atlas 14 precipitation-frequency estimates.

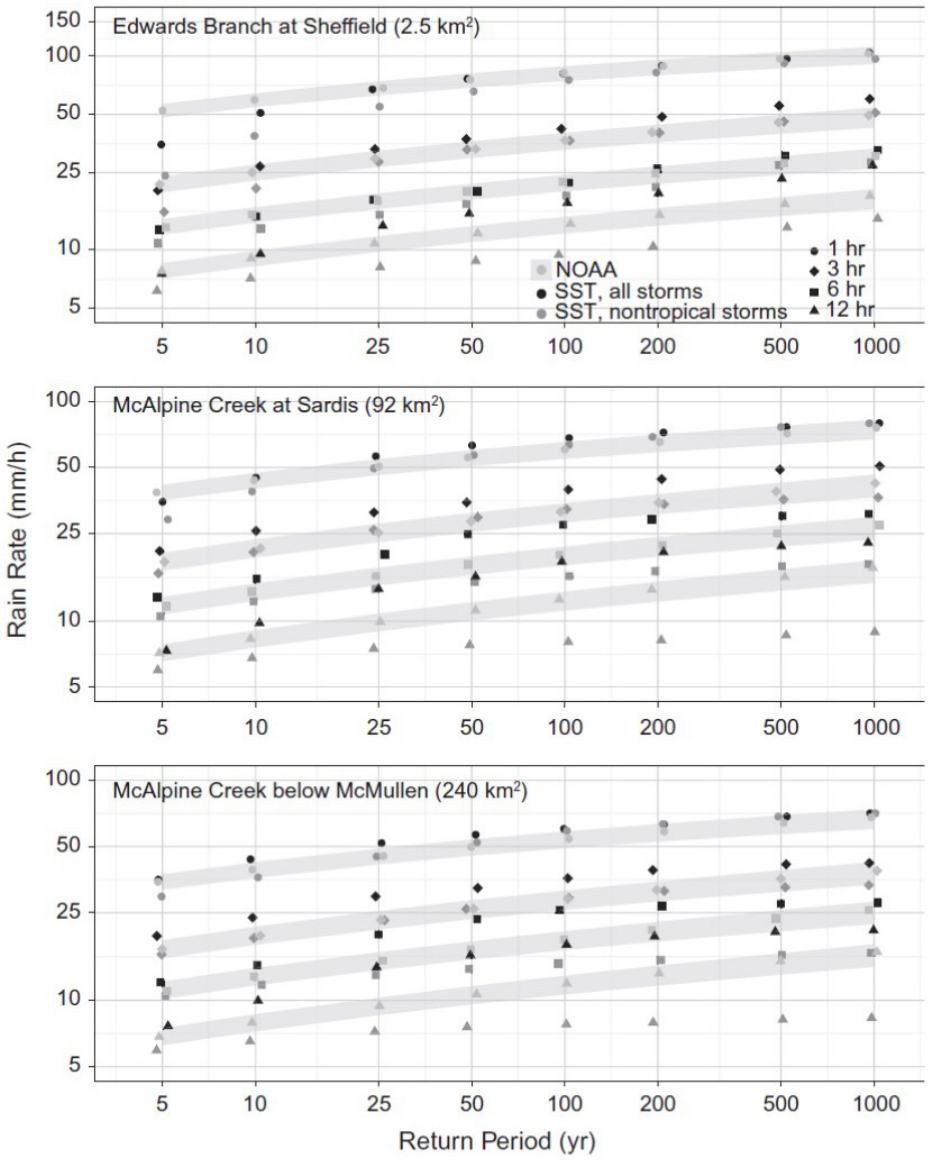


Figure 3-9 Comparison of SST-Based and NOAA Atlas 14-Based IDF Estimates for Edwards Branch at Sheffield (top), McAlpine Creek at Sardis (middle), and McAlpine Creek below McMullen (bottom). Light Gray Shading Indicates 90% Confidence Levels for the NOAA Atlas 14 Estimates.

The U.S. Intermountain West (IMW) has many large mountain ranges that often impede moisture transport into the region. Alexander et al. (2015) explored how moisture typically reaches the IMW during winter months via preferred moisture pathways. Back trajectory and Empirical Orthogonal Function (EOF) analyses were used to identify these preferred moisture pathways through the IMW. Back trajectory analyses track the horizontal and vertical path of an air parcel through the atmosphere during a specific amount of time. Back trajectory analyses allow the user to determine where moisture for a given storm originated and the pathway the moisture took through the IMW (Alexander et al. 2015; Bracken et al. 2015). The result of these analyses was a determination of the preferred moisture pathways through portions of the IMW (Figure 3.10) under certain synoptic conditions.

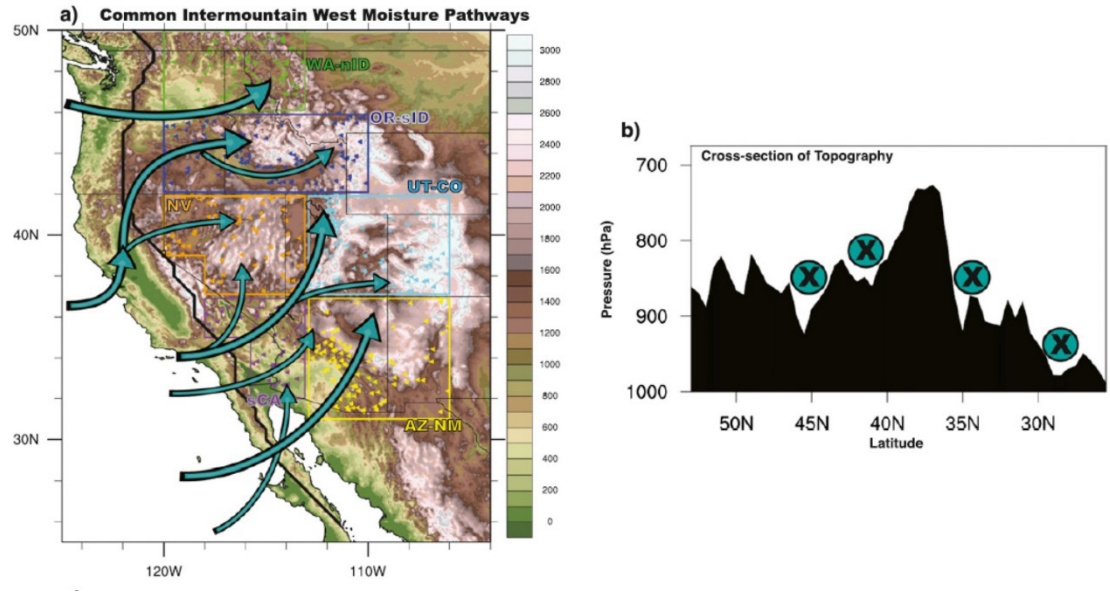


Figure 3-10 Schematic Map Showing Major Moisture Pathways from the Pacific Ocean into the IMW (a) and Topographic Cross-Section Along the Black Line in Map (b). Figure from Alexander et al. (2015).

Bracken et al. (2015) used a clustering analysis to identify homogeneous regions in the Western U.S., a back trajectory analysis to identify dominant moisture sources and pathways for each homogenous region and studied whether these pathways vary with El-Niño/Southern Oscillation regime. The authors used an alternative clustering approach to k-means clustering called partitioning around medoids (PAM; Kaufman and Rousseeuw 1990). A variogram called F-Madogram was used to assess pairwise dependence between time series of precipitation maxima (Bernard et al. 2013), instead of a more classical method such as Euclidian distance. The use of PAM and F-Madogram allowed an efficient and nonparametric way of clustering precipitation maxima and is referred to as the modified extremes clustering method for large geographic regions (Bracken et al. 2015). The results of this research show that homogeneous regions and their associated back trajectories can be identified using a combination of tools.

Self-organizing Maps (SOM) are a nonlinear neural network algorithm that employ independent learning to arrange input data into representative groups, where adjoining patterns are more alike than distant patterns. Swales et al. (2016) utilized the SOM algorithm to explain relationships between large-scale synoptic patterns and extreme precipitation events over the U.S. IMW. They discovered that ~70% of the extreme precipitation events across the IMW were associated with rarely occurring (~1.5%) synoptic conditions. The two types of synoptic conditions that lead to the greatest precipitation include a landfalling, zonally propagating trough that causes a band of moisture to traverse southward along the West Coast, and a stationary ridge over the West Coast that causes a persistent flow of moisture into the Pacific Northwest. The SOM algorithm was also used to show preferred moisture pathways using integrated vapor transport in the IMW based on synoptic conditions. This valuable information shows how moisture is transported into areas of complex topography and where preferred pathways may be located given certain synoptic conditions.

4 PRECIPITATION-FREQUENCY ANALYSES

4.1 Introduction

Deterministic precipitation analyses, such as PMP, have been used in design engineering projects for decades, without updates to the data, assumptions, and methods used to develop the procedures. For example, HMR 51, which covers most of the eastern U.S., has not been updated since 1978. Yet, there have been many advancements within the atmospheric and hydrologic communities during the past few decades that could be applied to improve such deterministic concepts. However, the ultimate limitation of using deterministic analyses is the result and application of a single value. While deterministic analyses can provide "theoretical" upper limits for precipitation, they alone do not provide probabilistic information for risk-based decision making. Federal agencies, such as Reclamation and the U.S. Army Corps of Engineers, have begun the transition to probabilistic methods for decision making.

Precipitation-frequency analyses are probabilistic methods that play important roles in many water resources planning projects, including culvert design, flood plain management, and river impoundments. These probabilistic methods estimate precipitation depths (also referred to as return levels) for specific durations associated with various return periods. A *T*-year precipitation event (*T*-year return level) is an event that is expected to occur, on average, once every *T*-years (return period). For example, a 100-year event has a 1/100, or 1% chance of occurring every year, and the depth of a 100-year event is referred to as the 100-year return level. Several different methods exist for estimating precipitation-frequency, including probability weighted moments (PWMs; Greenwood et al. 1979), L-moments (a subset of PWMs; Hosking 1990), maximum likelihood estimators (Fisher 1912), and Bayesian inference (e.g., Cooley et al. 2007; Bracken et al. 2016), among others.

In this section, we review previous precipitation-frequency studies, including those produced by government agencies, consultants, and researchers in the scientific community. The studies discussed here neglect to distinguish between rainfall and precipitation, which are technically different. Observations of total precipitation include solid and liquid precipitation as a single number. Consequently, analyses performed on precipitation gauges located in regions that receive snowfall, such as orographic regions, may include contributions from solid precipitation. Individuals interested in rainfall observations, as opposed to total precipitation, should contact the agency responsible for the operating and maintaining the observing platforms. We then discuss and demonstrate the application of two different precipitation-frequency estimation methods in an orographic region, specifically, the Tennessee River Valley (TRV). While these methods are demonstrated on data within the TRV, they have been applied to watersheds across the globe. With proper preparation and understanding of the driving mechanisms behind extreme precipitation in the study region of choice, the methods developed in this study are transferable.

4.2 Previous Studies

4.2.1 Federal Studies

Organized precipitation-frequency analyses have been performed across the U.S. since the early 1930s. For example, Yarnell (1935) published some of the first precipitation-frequency maps across the U.S. for durations of five minutes to 24 hours, at return periods of two- to 100-

years. The analysis, performed by the U.S. Department of Agriculture, was motivated by the need for economical designs of farm-terrace systems, farm-drainage systems, culverts, storm-sewer systems, and other engineering works (Yarnell 1935). The report focused on high-intensity, short duration precipitation events, as opposed to moderate-intensity and long duration. The author analyzed precipitation observations from 211 automatic Weather Bureau gauges located at cities across the U.S. between 1888 and 1933.

Later, the Weather Bureau produced a rainfall frequency atlas of the entire US in 1963, Technical Paper No. 40 (hereafter TP40; Herschfield 1963). The authors used the method of moments to fit a Fisher-Tippett Type I distribution (also referred to as a Gumbel distribution) to annual maximum data and estimate partial duration series. Precipitation depths were computed for durations of 30 minutes to 24 hours for return periods from one- to 100-years. Shortly after TP40 was produced, the Weather Bureau released Technical Paper 49 (hereafter TP49; Miller et al. 1964). TP49 was prepared for the Soil Conservation Service to support planning and design in the Watershed Protection and Flood Prevention Program. The report focused on precipitation events from two- to 10-day durations at return periods from two- to 100-years. The analysis included data from 94 stations located across the U.S. Precipitation-frequency estimates were computed using the method of moments. However, the data records were considered too short to compute any moments higher than the second moment. Consequently, the authors restricted return periods to values equal to or less than 100-years. As with TP40, the authors fit a Fisher-Tippet Type I distribution to annual maximum time series and estimated results for partial duration series using scale factors.

More than a decade later, the NWS produced precipitation-frequency analyses for 37 states located from North Dakota to Texas and eastward (Frederick et al. 1977). Maximum annual precipitation observations from between 200 and 1900 stations were used to develop frequency results at durations from five- to 60-minutes, at return periods from two- to 100-years. A different number of observations was used for different durations. The analysis included observations from first-order NWS gauges with an average period of record equal to 60 years. In an effort to remain consistent with previous studies, the authors fit a Fisher-Tippet Type I distribution to annual maximum time series. Unlike previous analyses, the authors produced gridded final precipitation estimates on a 2° latitude × 2° longitude grid using a multi-step smoothing technique.

In the early 1970s, NOAA's Hydrometeorological Design Studies Center started producing precipitation-frequency analyses for domains across the U.S. For example, NOAA Atlas 2 (Miller et al. 1973) consists of five volumes, each of which contains gridded maps of six- and 24-hour point precipitation estimates for Montana, Wyoming, Idaho, Washington, and Oregon[7] for two- and 100-year return periods. NOAA Atlas 2 was developed to address weaknesses in previous frequency analyses in the realm of orographics (Miller et al. 1973). As with TP40, the authors use the method of moments to fit a Fisher-Tippett Type I distribution to annual maxima and partial duration series.

Regardless of methodology, a lack of observations increases the difficulty of quantifying the probability and intensity of rare precipitation events. As a result, regional techniques have been developed to improve accuracy (Trefry et al. 2005). Regional precipitation-frequency analyses "trade space for time" by using data from similar gauges to derive estimates for any single gauge within a homogenous region. In general, the steps of a regional frequency analysis include identifying a homogeneous region (within which all observations are assumed to be

[7] http://www.nws.noaa.gov/oh/hdsc/currentpf.htm

described by a single frequency distribution aside from a site-specific scaling factor), screening the data within the homogeneous region for errors, selecting a frequency distribution that describes the data well and produces robust quantile estimates, and estimating the regional frequency distribution (Trefry et al. 2005). Organized precipitation-frequency analyses across the U.S. transitioned to regional methods with the publication of the numerous volumes contained within NOAA Atlas 14 (e.g., Bonnin et al. 2006; Perica et al. 2013).

In general, precipitation estimates from volumes within NOAA Atlas 14 (hereafter NA14) are characterized by multiple improvements over the analyses presented in TP40 and NOAA Atlas 2, including improved data quality (in terms of station density and period of record), methods for selecting a frequency distribution, and techniques used for spatial interpolation. More specifically, volumes in NA14 use the regional L-moments approach for selecting and parameterizing probability distributions. Multiple previous studies suggest that L-moments, which represent linear combinations of PWMs, are less sensitive to outliers and are more efficient than the method of moments used in previous studies (e.g., TP40 and NOAA Atlas 2). For example, Landwehr et al. (1979) use PWMs to estimate parameters for the Gumbel distribution and found them superior to conventional moments and maximum likelihood estimates (MLE). Hosking et al. (1985) suggest that the upper quantiles obtained from PWMs are preferable to MLE because MLE have large biases and standard deviations. Finally, Hosking (1990) shows that L-moments are less sensitive to outliers and can be more efficient than the method of moments estimators.

NA14 is divided into volumes based on geographic boundaries of the country. Consequently, precipitation-frequency estimates for the Tennessee River Valley watershed include values from two NA14 volumes, Volume 2 (Bonnin et al. 2006) and Volume 9 (Perica et al. 2013). We focus our discussion of NA14 methods on those two volumes, as they are the most relevant to our study region. NA14 Volume 2 includes precipitation-frequency estimates for the domain highlighted in Figure 4.1 (the Ohio River basin and surrounding states) for durations between five-minutes and 50-days at average recurrence intervals of one- through 1,000-years. The precipitation-frequency estimates in NA14 Volume 2 supersede those estimates from TP40, NWS HYDRO-35, and TP49 (Bonnin et al. 2006). The authors employ regional L-moments on observations from seven different precipitation datasets recorded throughout the region. Homogenous regions are first defined by climate, seasonality of extreme precipitation, type of precipitation, topography, and homogeneity. Later, the regions are refined using objective metrics from Hosking and Wallis (1997). In some cases, a single site is treated as a homogeneous region. Final homogeneous regions used in NA14 Volume 2 are shown in Figure 4-2.

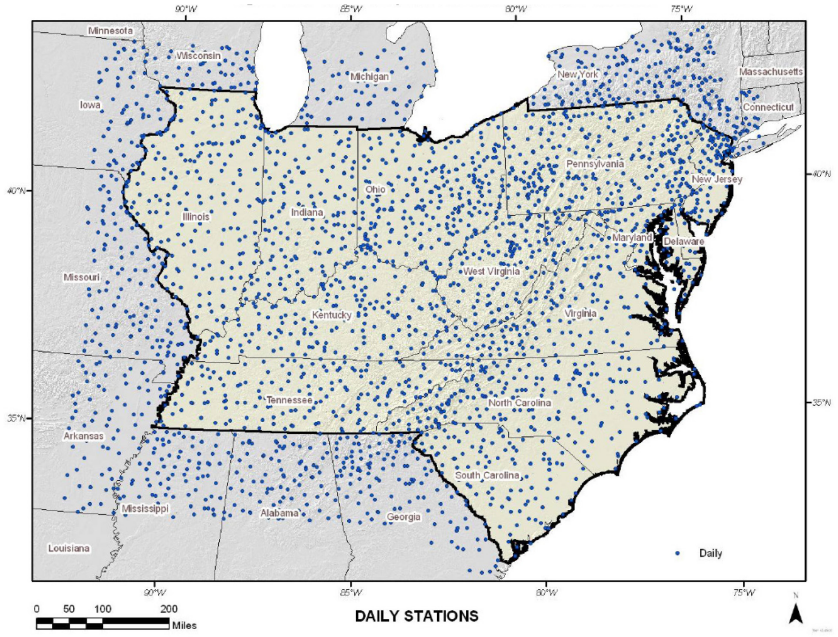


Figure 4-1 Daily Stations Used in NOAA Atlas 14 Volume 2. Volume 2 Domain is Outlined in Black. Figure from Bonnin et al. (2006).

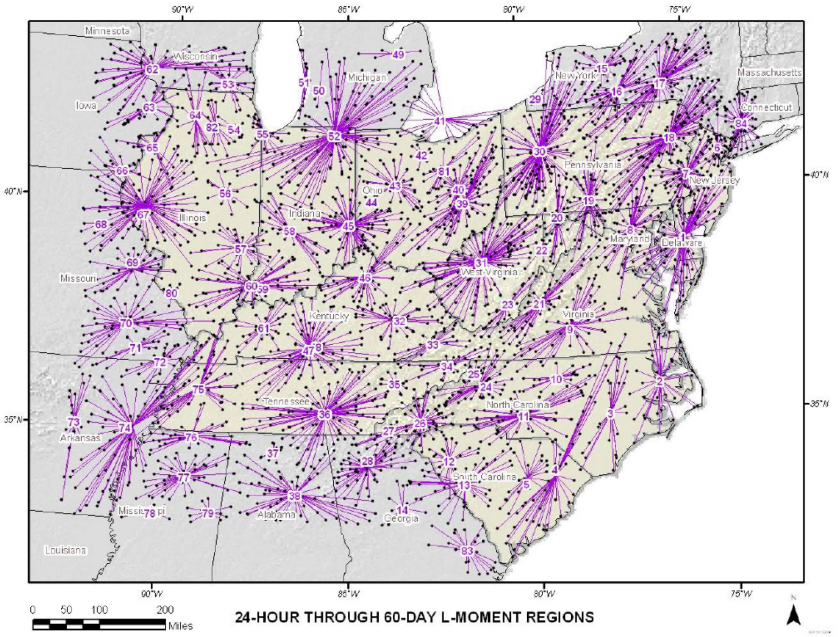


Figure 4-2 Homogeneous Regions for Daily Durations Used in NOAA Atlas 14 Volume 2. Figure from Bonnin et al. (2006).

Bonnin et al. (2006) use three different tests to identify the most appropriate statistical distribution for a set of gauges including Monte Carlo sampling, a "real-data" test, and root-mean-square error test. However, in some cases, the authors employ a different distribution than the one that was identified using the three tests. This manual distribution selection is used to reduce "bull's eyes and/or gradients in precipitation-frequency estimates between regions" (Bonnin et al. 2006).

Overall, the generalized extreme value (GEV) distribution is used for 68 regions, the generalized normal (GNO) is used for 10 regions, and the generalized logistic (GLO) is used for eight regions.

Gridded precipitation-frequency estimates from NA14 Volume 2 are developed using a procedure referred to as the cascade residual add-back (CRAB) grid derivation. The procedure involves producing a grid of mean annual maximum precipitation from PRISM and using that grid to estimate precipitation magnitudes for rarer events. Ultimately, for an *x*-duration event (e.g., 24-hour), the mean annual maximum precipitation fields are used to estimate the two-year return period, the two-year return period fields are used to estimate the five-year return periods, and so on for all return periods. The authors justify this procedure using an empirical linear relationship identified between *x*-duration 50-year precipitation totals and *x*-duration 100-year precipitation totals. However, instances arise during this procedure where, for example, the initial estimate of the 48-hour *y*-year precipitation total is greater than the 72-hour *y*-year precipitation total. In those cases, the 72-hour precipitation total is set to 1.01 times the 48-year precipitation total. A similar multiplication factor is applied to frequency errors (e.g., if the *x*-duration, 50-year total is greater than the *x*-duration, 100-year value). The implications of this heuristic scale factor are visible in the results of Figure 4.3. Precipitation depths for the two-, three-, and four-day durations at return periods greater than 25-years appear artificially similar. Finally, a smoothing factor is applied to the final precipitation totals depending on local topography. Precipitation estimates are available across the domain on a 30 second spatial resolution.

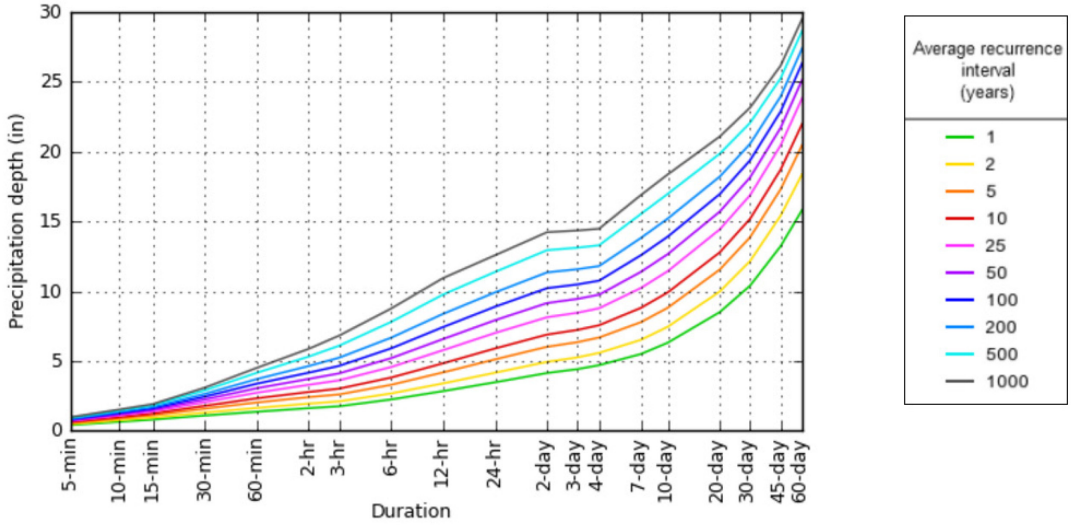


Figure 4-3 Depth Duration Frequency Curves for a South Carolina Site (34.9133°N, -82.7159°W) Based on NOAA Atlas Volume 2[8]

NA14 Volume 9 was published more recently than NA14 Volume 2 and, as a result, includes methods that differ from those presented in NA14 Volume 2. For example, NA14 Volume 9 includes observations from precipitation gauges located within one degree of the domain (Figure 4.4), while a distinct regional boundary is not explicitly defined in NA14 Volume 2. A second major difference between the two volumes involves the spatial extent over which precipitation-frequency estimates are created. NA14 Volume 9 computes regional precipitation-frequency relationships at each gauge separately, where a homogeneous region is assumed to

[8] Figure from http://hdsc.nws.noaa.gov/hdsc/pfds/pfds_map_cont.html?bkmrk=sc.

be the 10 closest precipitation gauges. However, the final number of gauges included in a homogenous region is ultimately allowed to vary. NA14 Volume 2 computes precipitation-frequency relationships over large regions, not at each gauge separately. The third difference between the two Volumes is in the choice of frequency distribution. NA14 Volume 2 allows the frequency distribution to vary spatially (from a selection of five distributions), while NA14 Volume 9 forces all data to a GEV distribution.

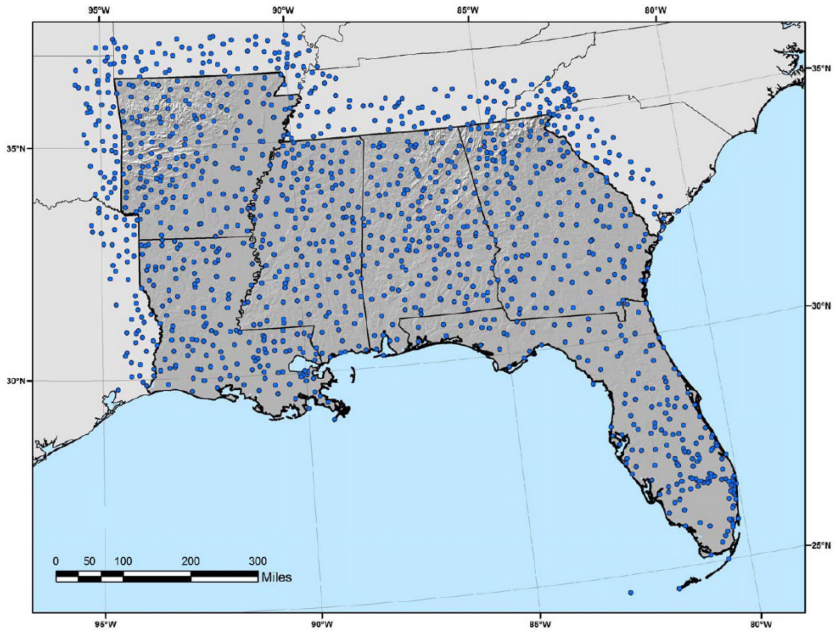


Figure 4-4 Daily Stations used in NOAA Atlas 14 Volume 9. Volume 9 Domain is Outlined in Black. Figure from Perica et al. (2013).

The final difference between NA14 Volume 2 and Volume 9 relates to the methods used to spatially interpolate precipitation totals. In Volume 9, the PRISM dataset is used to spatially interpolate station observations of mean annual maximum precipitation for each duration. Next, at-site ratios between two-year estimates and mean annual maximum precipitation are spatially interpolated to the grid using a modified version of Thiessen polygons. These gridded ratios are then multiplied by gridded mean annual maximum precipitation to produce gridded estimates of the two-year precipitation totals. This process is repeated for larger return periods (RP_{i+1}), using the previous return period (RP_i) as a predictor and spatially interpolating ratios. Similar to Volume 2, a dynamic filter is applied to precipitation-frequency grids, such that elevation gradients and the proximity to a coastline determine the degree of smoothing. The 1.01 scale factor described above is also used in Volume 9 to "correct" inconsistencies in precipitation-frequency totals. This approach differs from the CRAB approach used in Volume 2.

4.2.2 Non-Federal Studies

4.2.2.1 L-Moments

Outside of government-coordinated analyses, regional L-moments (sometimes referred to as the index flood procedure) have been implemented in several precipitation and snow studies. Schaefer (1990) uses mean annual precipitation to construct homogeneous regions while performing a regional frequency analysis (RFA) for precipitation totals in Washington State. Alila (1999) uses regional L-moments to estimate precipitation return periods for durations

ranging from five-minutes to 24-hours across Canada using observations from 375 gauges. According to heterogeneity measures, the author assumes that all gauges can be treated as a single region (i.e., the entire country is one region) and characterized by a single regional growth curve (i.e., one homogeneous region). The author suggests that there is very little spatial variability in two of the L-statistics, supporting the use of one homogeneous region. Fowler and Kilsby (2003) employ regional L-moments to estimate return periods of one- to 10-day annual maximum precipitation in nine regions across the United Kingdom. The analysis includes precipitation observations from 204 gauges with data from 1961 to 2000. Assuming a GEV distribution, the authors calculate precipitation totals for return levels using a ten-year moving window and several fixed decades (e.g., 1961-1970, 1971-1980, etc.). Their results show significant decadal changes in five- and 10-day precipitation events across the region.

Kysely and Picek (2007) use a regional L-moments approach to quantify precipitation return periods for one- to seven-day annual maxima across the Czech Republic. The authors use a clustering algorithm, along with some subjective decision making, to group long-term precipitation gauges into four distinct regions across the country. Precipitation maxima in three regions were best described by the GEV distribution, while precipitation maxima in the fourth region were best described by the generalized logistic (GLO).

4.2.2.2 Bayesian Inference

In recent years, there have been numerous advancements in the field of frequency analysis, including precipitation- and flood-frequency analyses. Most notable are those advancements that exploit iterative processes to quantify uncertainty, such as methods based on Bayesian inference. The benefits of using Bayesian inference techniques for precipitation-frequency are numerous, from the quantification of uncertainty at unmonitored locations (i.e., where no data is available), to the flexibility to include data from a variety of sources with distinct treatments for each. Bayesian inference is not a new concept. In fact, Bayes' theorem (Bayes and Price, 1763), upon which Bayesian inference is based, was published in 1763 by Richard Price, a friend of the late Reverend Thomas Bayes. It was not until recently that the significant increase in personal computing power helped to bring Bayesian inference to the foreground of the research and operations community.

From a spatial perspective, Bayesian inference can be a very powerful tool for quantifying the full distribution of uncertainty at unmonitored locations. Some of the most prominent research on the subject was presented by Cooley et al. (2007), who analyzed the spatial distribution of extreme precipitation levels for the front range of the Rocky Mountains in Colorado. Since then, there has been a renewed interest in spatial Bayesian applications. Reich and Shaby (2012) use a hierarchical max-stable model with climate model output in the east coast to examine spatially varying GEV parameters, with a max-stable process for the data dependence level. Ghosh and Mallick (2011) model gridded precipitation data over the entire U.S., for annual maxima at a 5° latitude x 5° longitude resolution (43 grid cells) and copula for data dependence, incorporating spatial dependence directly in a spatial model on the data, not parameters. A copula (the Latin word for link) is a multivariate distribution function that enables the efficient modeling of correlated, non-normal data. Sang and Gelfand (2009) and Cooley and Sain (2010) model over 1,000 grid cells of climate model output using spatial autoregressive models.

We propose using a Bayesian inference RFA framework as a replacement for the commonly used L-moments, which is quite possibly the most frequently used parameter estimation technique for regional precipitation-frequency analysis. By using a Bayesian inference

framework, we can more properly account for parametric uncertainty, while simultaneously reducing the computational expense of bootstrapping that comes with an L-moments uncertainty analysis.

4.3 Regional Frequency Analyses

In this section, we perform two types of regional precipitation-frequency analyses within the Tennessee River Valley watershed (TRVW). The first approach is based on regional L-moments (Hosking and Wallis 1997), while the second approach is based on Bayesian inference. Both analyses begin with the same set of historical precipitation observations and assume that all data follow a GEV distribution, a distribution that is commonly used to model annual maximum precipitation from observed data (Overeem et al. 2009; Hanel and Buishand 2010). Distribution selection can have a large impact on precipitation-frequency estimates, particularly at the tails. However, the purpose of this study is not to demonstrate the impact of frequency distribution selection on precipitation-frequency estimates; the epistemic uncertainty due to distribution choice was not the focus of this work. If a user is interested in exploring additional sources of uncertainty throughout the precipitation-frequency analysis (PFA) process, frequency distribution selection should be considered. Section 4.3.1 discusses the data processing steps taken to establish the precipitation dataset used in both analyses. Section 4.3.2 provides an objective clustering method to address the orographic effects on precipitation in the TRVW. Section 4.3.3 and 4.3.4 discuss the regional L-moments and Bayesian analysis, respectively. A comparison of results from the two methods is presented in 4.4.

4.3.1 Data

We analyze daily total precipitation from gauges in the GHCN-Daily network that are located within the TRVW. Currently, there are 1,503 gauges located in the watershed; however not all gauges have sufficient period of record and coverage for analysis. We restrict our analysis to gauges that have 85% data coverage (i.e., a minimum of 85% of 365 days are not missing) for a minimum of 10 years. These criteria reduce the number of available stations from 1,503 to 179. Figure 4.5 shows the spatial distribution of GHCN daily gauges across the TRVW used in the precipitation-frequency analyses. Note that one may choose to include precipitation gauges outside the TRVW boundary, as precipitation patterns do not necessarily conform to basin boundaries.

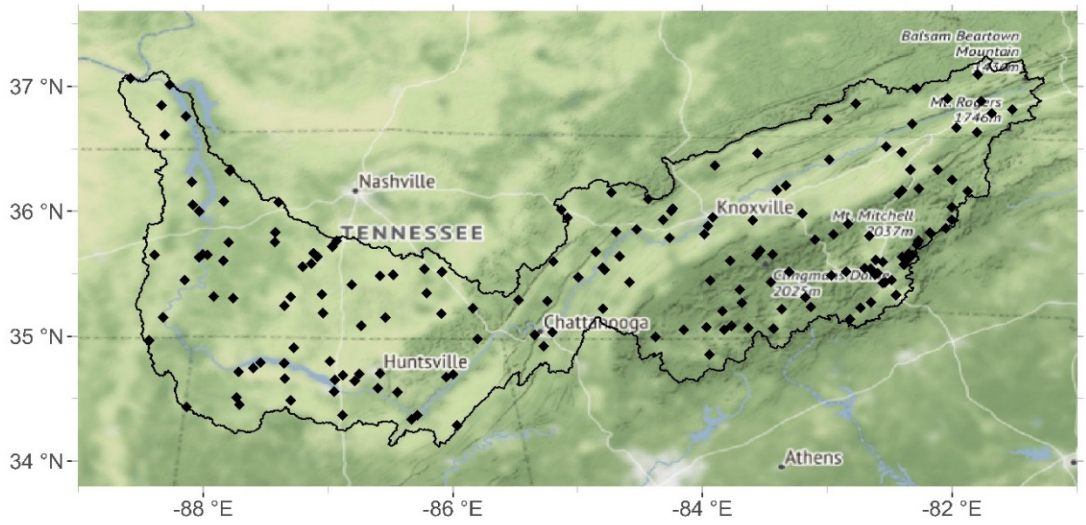


Figure 4-5 GHCN-Daily Precipitation Gauges within the TRVW that Have a Minimum of 85% Data Coverage for a Minimum of 10 years

One-day, two-day, and three-day annual maximum values from the gauges shown in Figure 4.5 were quality-controlled to ensure accuracy. Anomalously large and small annual maximum values, namely those values that are greater than the 95th percentile or less than the 5th percentile were verified against data tables maintained by the National Climate Data Center (NCDC). Additional quality metrics may be applied to observations, however the primary goal in a regional frequency analysis is to retain all large precipitation observations.

The TRVW is large and vast, with significant variability in weather and climate, as well as topography. Additionally, moisture sources and forcing mechanisms vary by location. Most notably, the eastern half of the TRVW has pronounced orographic effects, which have proven to be problematic for regional frequency analysis (RFA). To this end, we define several sub-regions in the watershed to identify groups of stations with frequency distributions that are approximately identical, aside from a site-specific scale factor (Norbiato et al. 2007).

4.3.2 Homogeneous Region

Rather than computing a single precipitation-frequency relationship for the entire TRVW, which encompasses multiple elevation and climactic zones where precipitation observations likely behave differently, we define sub-regions (also referred to as homogeneous regions or regions) in the watershed over which to compute the precipitation-frequency analyses. The purpose of this step is to identify groups of stations with frequency distributions that are identical (Norbiato et al. 2007). A homogeneous region (HR) can be a geographic area, or it can be a collection of sites having similar statistical characteristics (Wallis et al. 2007).

There are multiple approaches available to define homogeneous regions for use in RFAs, including both subjective and objective approaches. Some subjective approaches involve identifying regions of similar mean annual precipitation, seasonality of extreme events, and forcing mechanisms (Caldwell et al. 2014; Keeney et al. 2015). Previous studies in the TRVW developed homogenous regions based on climate, seasonality of extreme precipitation, and forcing mechanism of precipitation (Bonnin et al. 2006; Schaefer et al. 2016). For example, regions in NOAA Atlas 14 Volume 2 (Figure 4-2) are delineated using climate, seasonality of

extreme precipitation, type of precipitation, topography, and homogeneity (Hosking and Wallis 1997; Bonnin et al. 2006). Bonnin et al. (2006) use the heterogeneity measure, H1, a metric of standard deviation, to refine the gauges within regions that were originally developed using multiple variables. In instances where H1 exceeds a value of 2 (a lower threshold than what is specified by Hosking and Wallis (1997)), the authors manually identify and filter out individual problematic gauges in the region. In some instances, high H1 values are acceptable without modifying gauges or region delineations (Bonnin et al. 2006).

In this study, we identify homogeneous regions in the TRVW using an objective, nonlinear neural network algorithm known as the Self-Organizing Map algorithm (SOM; Kohonen 1990). We apply the SOM algorithm to physical attributes of gauges – latitude, longitude, and elevation – and climatological precipitation data – average annual total precipitation and average annual one-day maximum precipitation. For real applications, one would want to perform sensitivity analyses to arrive at the optimal collection of attributes for a specific application. We employ this objective method to define homogenous regions because the TRVW is expansive, includes regions of highly variable terrain, and extreme precipitation events can be caused by different physical mechanisms. This analysis demonstrates the value of using an objective method to define homogenous regions for regional precipitation-frequency analyses in a region of complex terrain and forcing mechanisms. While the methods used to delineate homogeneous regions in previous studies are acceptable (e.g., Bonnin et al. 2006; Schaefer et al. 2016), the SOM algorithm represents a more advanced, objective approach to delineation.

The SOM algorithm is an example of an artificial neural network that identifies and converts complex relationships among high-dimensional data into geometric (or spatial) relationships on a low-dimensional grid (Kohonen et al. 1996; Cavazos 2000). The SOM consists of one input layer (the training data), and one output layer, also referred to as the Kohonen layer (Figure 4.6). The SOM algorithm derives patterns from the training data, and as a result, cannot identify patterns that are not included in the training data (Cavazos 1999). The topological structure of the SOM output layer can be one- or two-dimensional and is typically organized as a lattice (Lin and Chen 2006).

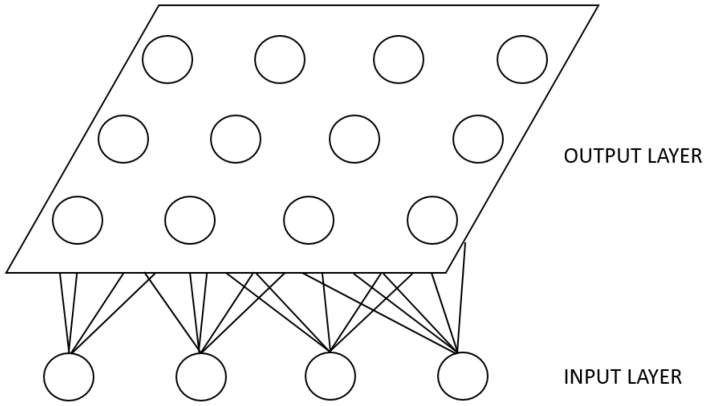


Figure 4-6 Schematic of the SOM Architecture. Adapted from Lin and Chen. (2006)

Lin and Chen (2006) provide a comprehensive review of the SOM algorithm, which we summarize here. Assume the input layer has M input vectors and is denoted by

$$X = [x_1, x_2, x_3, \dots, x_M]^T \quad (6)$$

The output layer includes N output neurons, $u_j, j = 1,2, \dots, N$. Weights are computed from each input vector to each output layer neuron (connecting lines in Figure 4.6), $w_{ij}, i = 1,2, \dots, M, j = 1,2, \dots, N$. The weight vector of each output neuron has the same dimension as the input vector and can be written as

$$W_j = [w_{1j}, w_{2j}, w_{3j}, \dots, w_{Mj}]^T, j = 1,2, \dots, N \quad (7)$$

Training begins by initializing all weights using any number of options, such as random values, empirical/orthogonal functions, or linear combinations. The SOM algorithm computes a similarity measure, most often Euclidean distance, between the input vector, X, and the weight vector, W_j, of each neuron, u_j, denoted by d_j, where

$$d_j = \|X - W_j\| = \sqrt{\sum_{i=1}^{M} (x_i - w_{ij})^2} \quad (8)$$

The output neuron u_j with the weight vector that is the smallest distance from the input vector is the winner. The weights of the winning neuron are adjusted toward the input vector. In addition, neurons surrounding the winning neuron are also adjusted. The influence of a neuron on neighboring neurons decays symmetrically from the winning neuron location. The Gaussian function is a typical choice of topological neighborhood function,

$$h_j = exp\left(-\frac{\|u_j - u_j^*\|^2}{2\sigma^2}\right) \quad (9)$$

where h_j is the topological neighborhood, σ is the 'effective width' of the topological neighborhood, and u_j^* is the winning neuron. The change to the weight vector W_j is computed as

$$\Delta W_j = \eta h_j (X - W_j) \quad (10)$$

where η is the learning rate parameter and h_j is the topological neighborhood.

Lin and Chen (2006) apply the SOM algorithm to precipitation observations across Taiwan in an effort to identify homogeneous regions across complex terrain for use in a regional L-moments analysis. The authors apply the SOM algorithm to 16 variables from each gauge, which include latitude, longitude, elevation, mean annual precipitation, and average monthly precipitation for each month separately. The authors standardize each variable (see their Table 2) such that equal weight is given to each variable. The authors utilize a 12x12 SOM grid, with random initiation and Gaussian neighborhood updates. Their results show that daily precipitation gauges across Taiwan can be categorized into eight different regions. The authors use a discordancy measure from Hosking and Wallis (1997) to verify that each sub-region is acceptably homogeneous. Note that the selection of variables, SOM grid size, initiation and neighborhood updates, and discordancy measure are all subjective decisions the authors made. Such subjective decisions may affect the outcomes to an unknown degree and provide a source of epistemic uncertainty that should be addressed.

We follow the approach presented in Lin and Chen (2006) to develop homogeneous regions across the TRVW. We identify and compute a similar set of standardized variables for the 179 GHCN gauges located in the watershed and run the SOM algorithm on those variables. The standardized variables include latitude, longitude, elevation, average annual precipitation, and

average one-day annual maximum precipitation. We include mean annual one-day maximum precipitation because the focus of the regional frequency analysis is on one-day annual maximums. If the user were to analyze multiple-day (e.g., two-day, three-day, etc.) precipitation totals, the standardized variables would need to be modified to best reflect the focus. Because there is no official guidance on the appropriate SOM grid size, we employ a SOM grid that is 5x3. We use a rectangular grid topology that is initialized using linear combinations of the input fields and updated based on a Gaussian neighborhood function. Numerous grid geometries were considered in an iterative sensitivity testing framework, which included subjective decision making regarding the distribution and population of each SOM grid. The final grid size, 5x3, was selected because it provided intuitive results: large and widespread groups in the western region of the TRVW, while in the eastern region of the TRVW the groups are smaller and more variable. Additionally, 5x3 was the first grid size wherein one node was assigned no stations, which can be considered an inflection point between too few nodes (ignoring orographics) and too many nodes (subjectively combining multiple nodes, as in Lin and Chen (2006)). The initial learning rate is 0.05 and decreases linearly. The 5x3 SOM map was constructed after a total of 2,000,000 iterations, which takes approximately 10 minutes to run on a standard personal computing laptop.

Frequency results from the SOM analysis are shown in Figure 4-7, which indicate how many precipitation gauges map to each node of the SOM grid. The majority of input vectors map to the lower left, upper left, and lower right of the SOM grid. The neighborhood function of the SOM algorithm ensures that similar nodes are located near one another, such that nodes nearby show similar characteristics while nodes on opposite sides of the SOM map show dissimilar (or even opposite) characteristics.

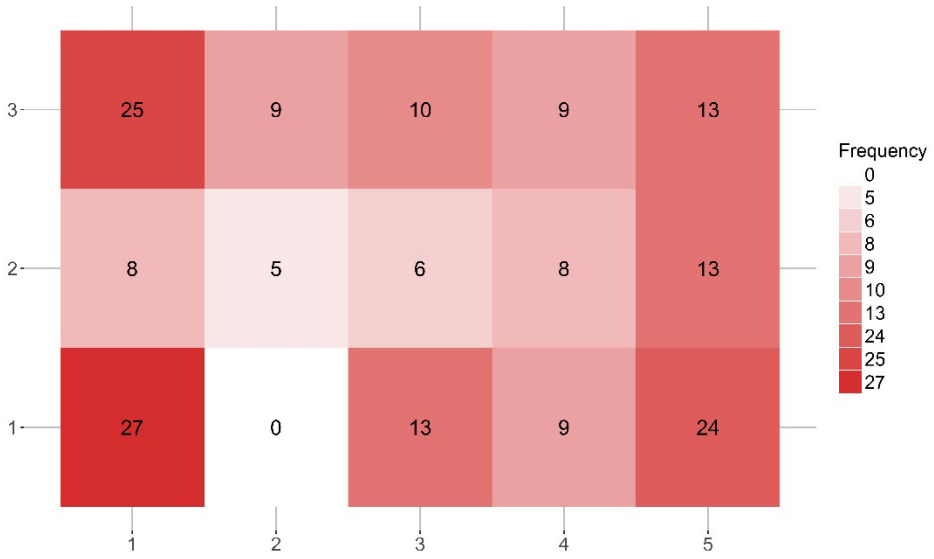


Figure 4-7 Frequency Output from 5×3 SOM Analysis. Each Node Value Represents the Number of Precipitation Gauges that Map to that SOM Node.

Our application of the SOM algorithm differs from that of Lin and Chen (2006) in that we assume each output neuron (i.e., node) of the 5×3 SOM is its own homogeneous region (HR). Thus, all stations that map to a specific node (Figure 4-8) are assumed to represent a distinct HR. The difference between Figure 4-7 and Figure 4-8 is the color scheme; both figures include frequency information. Assuming each node represents a single HR reduces the amount of

subjectivity that comes with determining which nodes to group together as a single HR, as in Lin and Chen (2006). However, the exact number of groups and the dimensions of the SOM node map are subjective quantities and can vary depending on the user's interest and prior knowledge of the study region.

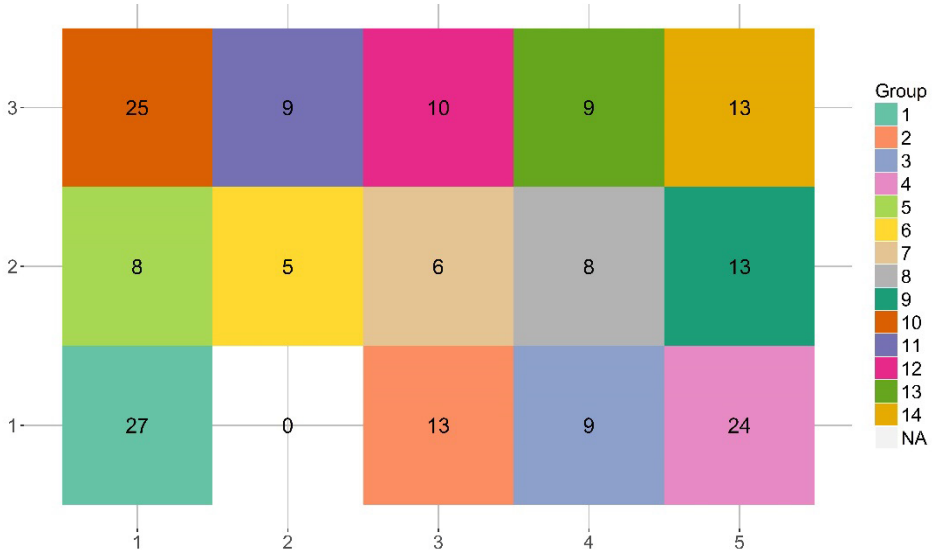


Figure 4-8 Fourteen Independent Groups Identified in the 5×3 SOM Grid. Each Node Value Represents the Number of Precipitation Gauges that Map to that SOM Node.

Each precipitation gauge shown in Figure 4.5 maps to a single SOM node shown in Figure 4-7 and Figure 4-8, while no precipitation gauges map to the node located at x=2, y=1. The number of precipitation gauges that map to each node varies from a maximum of 27 for group 1 (x=1, y=1) to a minimum of five for group 6 (x=2, y=2). We color each GHCN gauge based on its associated group from Figure 4-8, with results shown in Figure 4-9. These results suggest that applying the SOM algorithm to the five scaled variables from 179 GHCN gauges generally produces spatially coherent and consistent subgroups across the TVRW. The exception to the spatial coherence of these groups is evident in the Appalachian Mountains of eastern Tennessee, where there is a mixing of groups. Hosking and Wallis (1997) elaborate that homogeneous regions that are "chosen to be sets of contiguous sites, based on administrative areas or major physical groups of sites… seem arbitrary and subjective and the resulting regions rarely give the impression of physical integrity." Thus, homogeneous regions need not be geographical but rather based on similarities of the physical mechanisms of extreme precipitation between sites. If fewer homogeneous regions were desired, individual groups could be combined. However, we analyze all 14 groups separately in order to capture differences in precipitation-frequency relations in this region of complex topography. The RFAs in the following sections are based on the fourteen groups defined in Figure 4-8 and Figure 4-9. In the remainder of the text, we use the terms "group" and "homogeneous region" interchangeably.

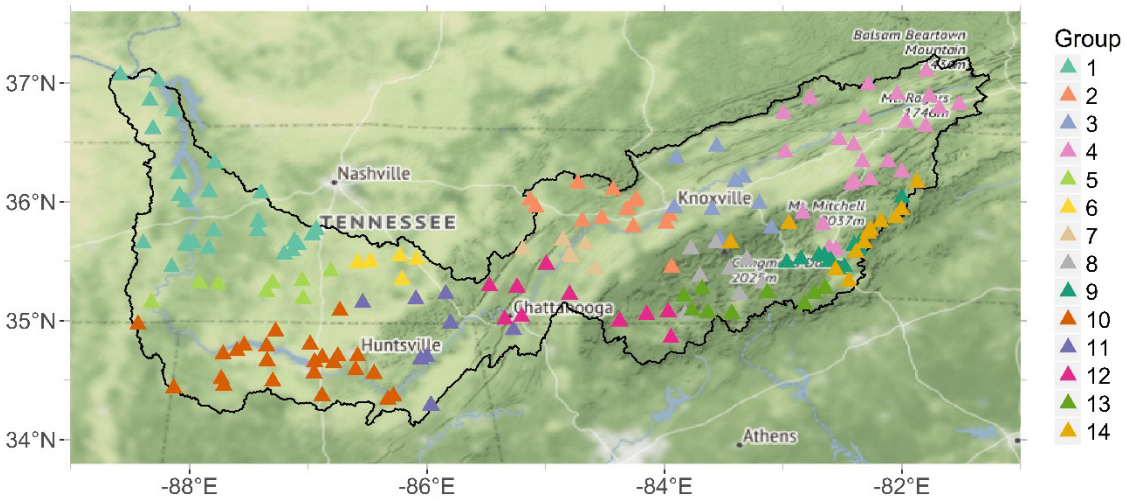


Figure 4-9 Fourteen Homogeneous Regions across the TRVW Defined by Output from the 5 × 3 SOM Analysis. Colors Correspond to Homogeneous Regions from Figure 4-8.

The HRs developed in this study and shown in Figure 4.9 vary from the HRs included in NA14 Volume 2 and Volume 9. NA14 Volume 9 treats each precipitation gauge as a separate HR such that a comparison between results is unjustified. However, NA14 Volume 2 includes equivalent homogeneous regions (e.g., Figure 4.2). NA14 Volume 2 includes a number of small HRs along the Appalachian Mountains in western North Carolina, while the analysis presented here includes more. Furthermore, NA14 Volume 2 includes a large HR located in the southeastern half of Tennessee, whereas results from the current study include more HRs across this region. The total number of HRs from this study exceeds those presented in NA14 Volume 2.

A major benefit to the SOM algorithm is its ability to "project" new data onto the trained SOM grid. For instance, a proposed site for a facility may not have available precipitation data, and therefore the proposed site would not be associated with any of the HRs defined by the SOM algorithm. However, given a vector or matrix of the site's attributes (i.e., latitude, longitude, elevation, average annual total precipitation, and average annual one-day maximum precipitation), the trained SOM model assigns the proposed location to one of the HRs using its vector of site attributes. Average annual total precipitation and average annual one-day maximum precipitation may be available from a variety of sources, such as the PRISM dataset (Daly et al. 1994). Projecting new data onto the trained SOM grid can be very useful, as it reduces the subjectivity involved in determining the risk of extreme precipitation at proposed facility sites.

4.3.3 L-Moments Analysis

Site-specific precipitation-frequency analyses often suffer from a lack of observed data for fitting a distribution to extreme values, an obstacle that can be overcome by using a regional approach (Bocchiola et al. 2008). Regional frequency analyses (RFAs) utilize data from sites other than the site of interest, while assuming that the data from all sites, when properly scaled, can be described by a single frequency distribution. Some suggest that using data from more than one site, or the "space-for-time" substitution mentioned previously, has the potential to improve the

accuracy of final quantile estimates as compared to single-site analyses (Hosking and Wallis 1997).

The regional L-moments algorithm is a common technique used to estimate site-specific precipitation-frequency relationships (Bonnin et al. 2007; Wallis et al. 2007) using annual or seasonal maximum precipitation observations. Each step in an RFA represents a unique process that involves multiple sub-steps, tools, and site-specific decisions. For example, all observations should be examined for errors and inconsistencies. The next step of an RFA involves identifying a homogeneous region. A *region* is defined as a set of sites whose frequency distributions are approximately the same (not necessarily a geographical region). The homogeneity of a region can be tested by calculating summary statistics of at-site data and comparing the between-site variability of these statistics with what would be expected from a homogenous region (Hosking and Wallis 1997). The next stage in an RFA involves specifying an appropriate regional frequency distribution. Most investigations determine the appropriate regional frequency distribution by applying some goodness-of-fit test. Goodness-of-fit tests often involve computing summary statistics from the data and testing whether these values are consistent with what would be expected if the data were a random sample from a postulated distribution (Hosking and Wallis 1997). According to Cong et al. (1993), L-moment goodness-of-fit tests are more robust than single-site goodness-of-fit tests because they use regional data to discriminate between alternative distributions. The regional L-moment algorithm computes L-statistics at each site separately and combines those estimates to produce a regional average. Regional L-statistics are used to compute distribution parameters for frequency distributions, which are assumed valid for every site in the homogeneous region.

Treyfry et al. (2005) present the computations of a regional L-moments analysis, which we summarize here. L-moments are based on a random sample of a random variable *X* of size *n* arranged in ascending order (i.e., $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$). The L-moment analysis begins by estimating the population probability weighted moment, β_r, where an unbiased estimator of β_r is

$$b_r = n^{-1} \sum_{j=r+1}^{n} \frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} x_{j:n} \tag{11}$$

The first four sample L-moments (also referred to as L-statistics) are given by

$$\lambda_1 = b_0 \tag{12}$$
$$\lambda_2 = 2b_1 - b_0 \tag{13}$$
$$\lambda_3 = 6b_2 - 6b_1 + b_0 \tag{14}$$
$$\lambda_4 = 20b_3 - 30b_3 + 12b_1 - b_0 \tag{15}$$

where λ_1 and λ_2 are measures of the sample mean and scale of the distribution, respectively. The remaining L-moments are used to determine the sample L-moment ratios, which are

$$\tau_r = \lambda_r / \lambda_2 \tag{16}$$

where τ_3 and τ_4 are sample L-skewness and L-kurtosis, respectively. The sample L-CV is given by

$$\tau = \lambda_2 / \lambda_1 \tag{17}$$

The next stage of the analysis involves screening the data. In this study, we employ the discordancy and heterogeneity metrics from Hosking and Wallis (1997). With *N* sites in a region, let

$$u_i = \left[t^{(i)}\ t_3^{(i)}\ t_4^{(i)}\right]^T \tag{18}$$

be a vector with t, t_3, and t_4 L-moment ratio estimates for site i. The matrix of sums of squares and cross-products is defined as

$$A = \sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T \tag{19}$$

The discordancy statistic for site i is given by

$$D_i = \frac{1}{3}N(u_i - \bar{u})^T A^{-1}(u_i - \bar{u}) \tag{20}$$

For a region with 15 or more sites, site i is considered discordant if $D_i \leq 3$.

Heterogeneity measures are used to assess whether the sites within a given region can be reasonably treated as homogeneous. The heterogeneity measures are designed to measure between-site variations in sample L-moments (L-CV, L-skewness, and L-kurtosis) for the group of sites with what is expected for a truly homogenous region. Hosking and Wallis (1997) argue that L-CV is most important because between-site variations in L-CV have a much larger impact than variations in L-skewness and L-kurtosis on the variance of the final estimates. Repeated simulations of a homogeneous region with sites having record lengths equal to those of the observed data allow the user to gain estimates of what is expected from a truly homogeneous region. The mean and standard deviations of L-statistics from those repeated simulations are used to estimate heterogeneity.

In order to compute the L-CV heterogeneity metric, the user must compute the regional average L-CV,

$$t^R = \sum_{i=1}^{N} n_i t^{(i)} / \sum_{i=1}^{N} n_i \tag{21}$$

Then, compute the weighted standard deviation of the at-site sample L-CVs,

$$V = \left\{\sum_{i=1}^{N} n_i \left(t^{(i)} - t^R\right)^2 / \sum_{i=1}^{N} n_i\right\}^{1/2} \tag{22}$$

After, the user fits a kappa distribution to the regional average L-moment ratios and simulates many realizations of synthetic regions with N sites, each having the same kappa distribution as its frequency distribution. The V metric is computed for each simulation, and from all simulations, a mean and standard deviation of V is computed, referred to as μ_V and σ_V, respectively. The L-CV heterogeneity measure is then computed as,

$$H = \frac{(V - \mu_V)}{\sigma_v} \tag{23}$$

Hosking and Wallis (1997) declare a region to be “acceptably homogeneous” if H is less than one, “possibly heterogeneous” if H is greater than or equal to one and less than two, and finally, “definitely heterogeneous” if H is greater than or equal to two. Note that the heterogeneity and discordancy measures (H and D) are different. The heterogeneity metrics are defined using a combination of observations and L-moments from all gauges, combined with synthetic series, whereas the discordancy metric is computed for each gauge separately. The discordancy metric is typically used to flag individual stations, which can be further inspected. The

heterogeneity metric applies to an entire region. If a region is deemed "definitely heterogeneous", the user should evaluate which gauges are retained in the analysis and which gauges are removed. The discordancy metric can help identify those gauges for removal.

The "regional" aspect of a regional L-moments analysis involves fitting a distribution using the method of L-moments with sample L-moment ratios from each site in a homogeneous region, weighted by record length and combined to yield a set of regional average L-moment ratios. Assume that we have *N* sites in a region, with site *i* having record length n_i, sample mean $\lambda_1^{(i)}$, and sample L-moment ratios $\tau^{(i)}, \tau_3^{(i)}, \tau_4^{(i)}$, and $\tau_5^{(i)}$, the regional average L-moment ratios are given by

$$t^R = \sum_{i=1}^{N} n_i t^{(i)} \Big/ \sum_{i=1}^{N} n_i \tag{24}$$

$$t_r^R = \sum_{i=1}^{N} n_i t_r^{(i)} \Big/ \sum_{i=1}^{N} n_i \quad \text{r=3, 4, 5.} \tag{25}$$

With the regional mean λ_1^R scaled to 1, the regional quantile function $\hat{q}(.)$ can be derived, and final quantile estimates can be obtained by multiplying the regional quantile estimates by the at-site mean (Trefry et al. 2005). Quantile estimates at site *i* for nonexceedance probability *F* (i.e., *F=1-1/T*, where *T* represents years) is expressed as

$$\widehat{Q}_i(F) = \lambda_1^{(i)} \hat{q}(F) \tag{26}$$

where $\widehat{Q}_i(F)$ represents the final, site-specific quantile estimates.

Regional L-moments are used to compute the three parameters of the GEV distribution (namely the location parameter μ, the scale parameter α, and the shape parameter ξ) used to describe the annual or seasonal maximum precipitation data. The regional L-moment estimate of the shape, scale, and location parameters are described as

$$\xi^R = 7.8590c^R + 2.9554(c^R)^2, c^R = \frac{2}{3+t_3^R} - \frac{\ln 2}{\ln 3} \tag{27}$$

$$\sigma^R = \frac{\lambda_2^R \xi^R}{(1-2^{-\xi^R})\Gamma(1+\xi^R)} \tag{28}$$

$$\mu^R = \lambda_1^R - \frac{\sigma^R}{\xi^R}[1 - \Gamma(1+\xi^R)] \tag{29},$$

respectively, where $\Gamma()$ denotes the gamma function. For ξ≠0, the quantile function of the fitted regional GEV distribution is

$$\hat{q}(T) = \mu^R + \frac{\sigma^R}{\xi^R}\left\{1 - \left[-\ln\left(1-\frac{1}{T}\right)\right]^{\xi^R}\right\} \tag{30}$$

For ξ=0, the quantile function, also referred to as the regional growth curve (valid for all locations within a homogeneous region), is given by

$$\hat{q}(T) = \mu^R + \sigma^R \ln\left[-\ln\left(1-\frac{1}{T}\right)\right] \tag{31}$$

Finally, the regional *T*-year estimate for site *i* is

$$\widehat{Q}_i(T) = \lambda_1^{(i)} \hat{q}(T) \quad (32)$$

In the current study, we demonstrate the application of a regional L-moments approach to precipitation observations across the TRVW to estimate return levels for one-day annual maximum precipitation events. We utilize the software, R (R Core Team 2016), along with the lmomRFA[9] package (Hosking 2015), to compute the individual and group L-moments. By default, the lmomRFA computes RGCs for six different probability distribution functions, including the generalized logistic, generalized extreme value, generalized Pareto, generalized normal, Wakeby, and log-Pearson Type III. However, since the purpose of this project is not to understand uncertainty in the probability distributions, we analyze results based on the GEV distribution. The GEV distribution was consistently selected as one of the six distributions that describes the one-day annual maximum precipitation data sufficiently well. We use quality-controlled observations from GHCN-Daily dataset (see section 4.3.1). We develop homogeneous regions using the SOM algorithm applied to five standardized variables (see section 4.3.2 and Figure 4.9). Using data from each group separately, we compute regional frequency distributions for each of the 14 homogeneous regions. We test discordancy among sites in each group separately using Equation 20 (Hosking and Wallis 1997). Observations from stations that are flagged as discordant, specifically, a discordancy value greater than three, are removed from the analysis. Table 4.1 includes a summary of the discordant stations. Two different groups included discordant stations, groups 1 and 10, which included two and one discordant station(s), respectively.

Table 4-1 Summary of Gauges Used in the L-Moments Analysis of One-Day Annual Maximum Precipitation Across the TRVW

Group	Starting number of gauges	Number of discordant gauges removed
1	27	2
2	13	0
3	9	0
4	24	0
5	8	0
6	5	0
7	6	0
8	8	0
9	13	0
10	25	1
11	9	0
12	10	0
13	9	0
14	13	0

Figure 4.10 shows the distribution of one-day at-site means for each group. This figure clearly shows that the magnitude of one-day annual maximum precipitation events varies throughout the watershed. For example, the largest one-day at-site mean in the watershed occurs in group 14, the group with stations located in the Appalachian Mountains, in far eastern Tennessee. This group also has the largest range of at-site means. Alternatively, groups 3 and 4, which

[9] https://cran.r-project.org/web/packages/lmomRFA/lmomRFA.pdf

include sites primarily located in the rain-shadow region of eastern Tennessee, north of the Appalachians, have the lowest values of one-day at-site means. One-day at-site means in the remaining groups generally fall between these groups.

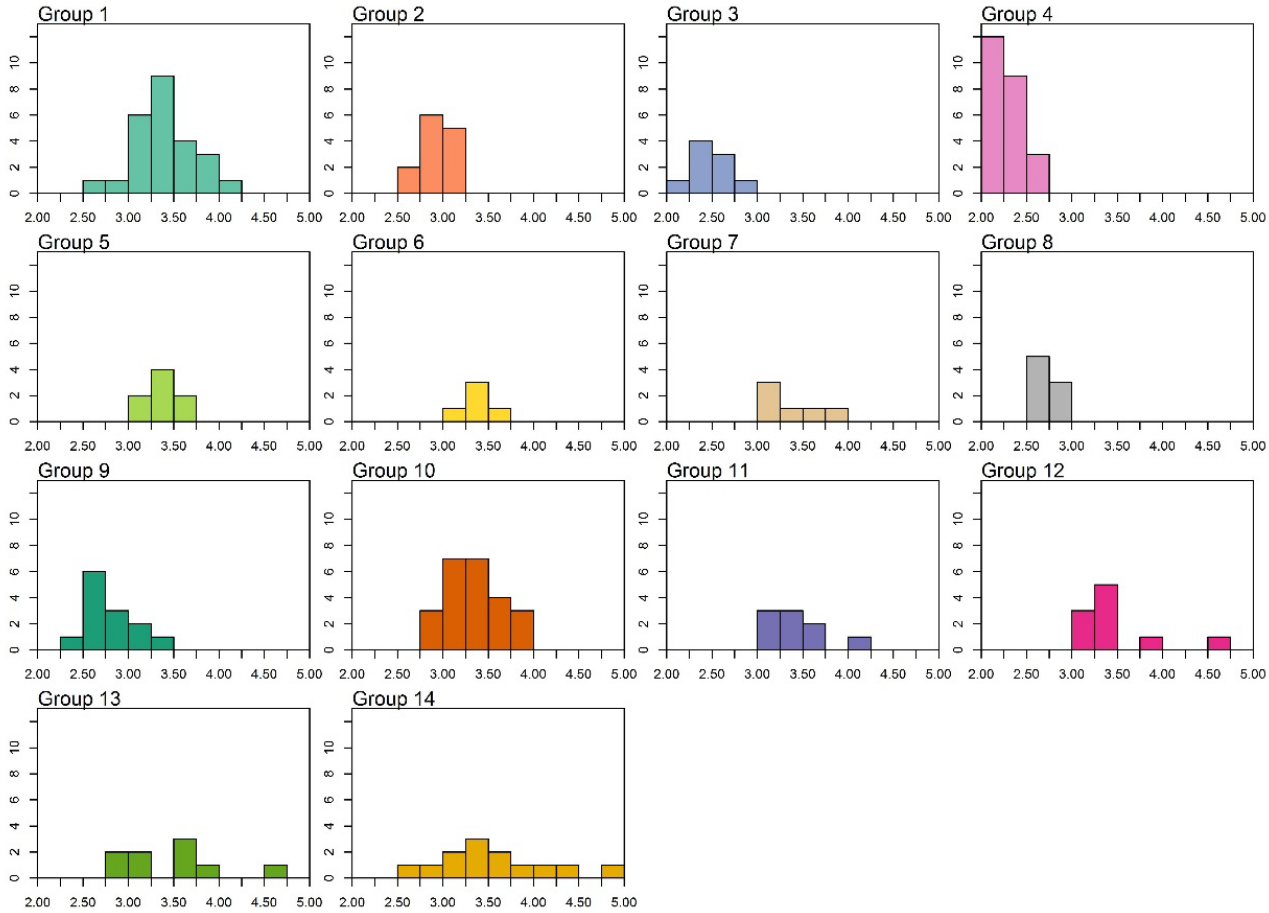


Figure 4-10 Distribution of One-Day at-Site Mean Precipitation (in inches) for Each Group. The X-axis Represents the Range of Means (inches), while the Y-axis Indicates the Number of Sites. Colors Correspond to the Groups from Figure 4-9.

One-day dimensionless regional growth curves (RGCs) for each of the 14 groups are shown in Figure 4-11. Each curve is based on the GEV distribution (defined above). Note, RGCs are scaled by at-site mean precipitation values to produce site-specific precipitation frequency analyses. Clearly, there is a great deal of variability in RGCs among the regions defined across the TRVW. For instance, the RGC for group 14 has the largest magnitude at an AEP of 1E-4, followed by group 12 and group 7. Groups 3, 2, and 4 have the lowest magnitudes at an AEP of 1E-4. Beyond magnitudes, the shapes of the RGCs also vary among groups, from relatively flat (group 3) to highly skewed (group 14). This is made possible by using the GEV distribution, which can capture other distributions through variations in the shape parameter.

Estimates of precipitation at the 10,000-year return period are rarely, if ever, presented in the literature due to the compounding effects of aleatory variability and epistemic uncertainty (e.g., model selection, data quality, data availability, etc.). Indeed, previous works (Bocchiola et al. 2008, and references therein) suggest that users should use caution when estimating

precipitation return levels at return periods greater than two times the period of record. We equate the total number of observations in each group to the length of the period of record. The period of record varies by group, with a minimum of 222 observations, and a maximum of 1,310 observations. As a result, our true confidence ranges from 444- to 2,620-years. Although we produce 10,000-year P-F estimates, they must be interpreted with caution, as the considerable amount of epistemic uncertainty at this return period must not be ignored.

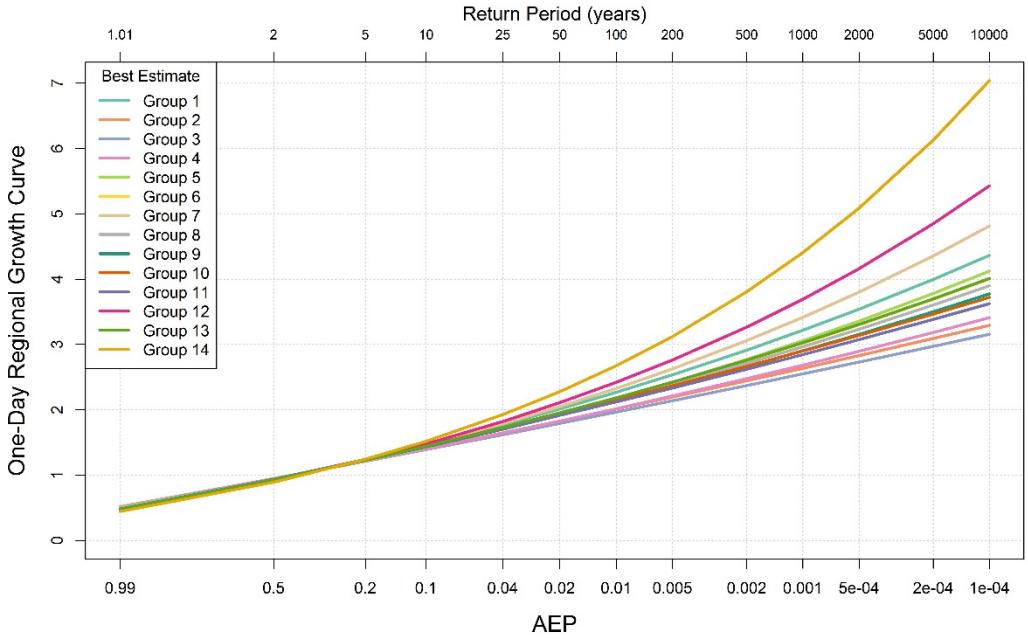


Figure 4-11 Dimensionless GEV Regional Growth Curve for Each Homogeneous Region Across the TRVW Based on One-Day Annual Maximum Precipitation. Colors Correspond to the Groups from Figure 4-9.

In order to produce an at-site precipitation-frequency analysis, an RGC must be scaled by an at-site mean. Figure 4-12 shows an at-site precipitation-frequency analysis for each site included in group 14. Each relationship is valid for a single site. The only difference among sites is the at-site means, which ranges from 2.6 inches to 4.88 inches. The range of one-day precipitation totals among sites at an AEP of 1E-4 (10,000-year return period) is 16 inches. Similar results are available for other groups, though they are not shown for the sake of brevity.

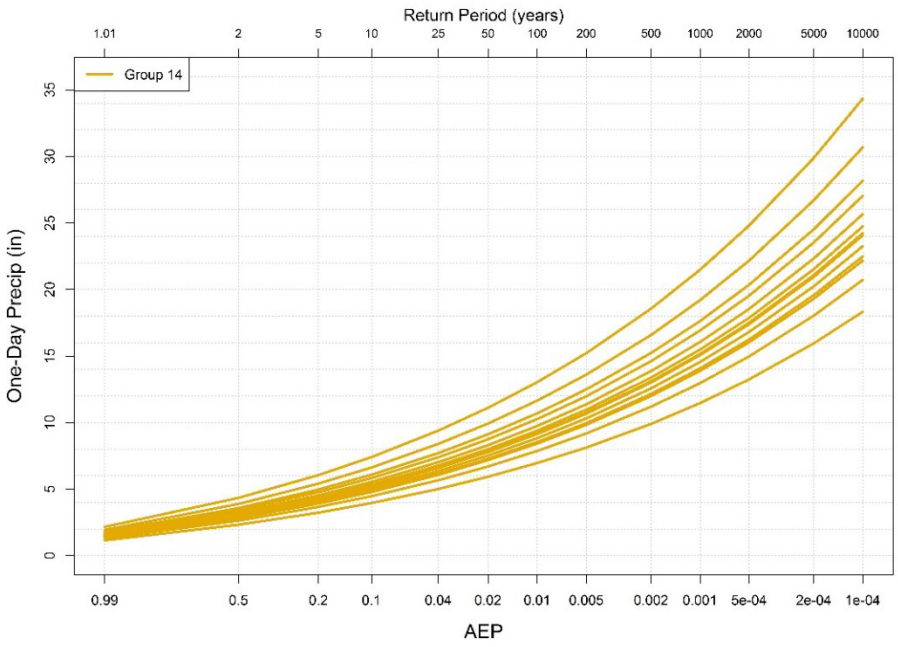


Figure 4-12 One-Day at-Site Precipitation-Frequency Relationships for those Sites Included in Group 14. Color Corresponds to the Group in Figure 4-10.

4.3.3.1 L-Moments Uncertainty

There are two major sources of uncertainty in any RFA, aleatory variability and epistemic uncertainty, which are fundamentally different from each other (Merz and Thieken 2005). Aleatory variability, which can be thought of as natural variability, stems from variability in the underlying stochastic processes and often refers to quantities that are inherently variable over time, space, or populations. Epistemic uncertainty, also referred to as ignorance, results from incomplete knowledge of the system under investigation. Merz and Thieken (2005) contend, along with others, that epistemic uncertainty can be reduced, while aleatory variability cannot. In this study, we explore only sources of epistemic uncertainty in the statistical parameters used to define the RGCs, rather than aleatory variability.

We estimate epistemic uncertainty in the RGCs using a bootstrapping resampling approach with replacement applied to historical precipitation observations. First, we randomly identify 10% of non-missing annual maximum values, remove those values, and re-compute the site-specific and regional L-moments and associated GEV RGCs. Secondly, we randomly select 10% of the reporting stations, remove all observations from those stations, and again compute the site-specific and regional L-moments and associated GEV RGCs. We use a threshold of 10% of stations because that is standard practice at Reclamation. In practice, one may choose to consider alternate thresholds. We perform these resampling computations 100 times each, resulting in 200 RGCs for each group. From this collection of 200 curves, we compute the 5th and 95th percentiles of the RGC at each return period. Results from this analysis are shown in Figure 4-13.

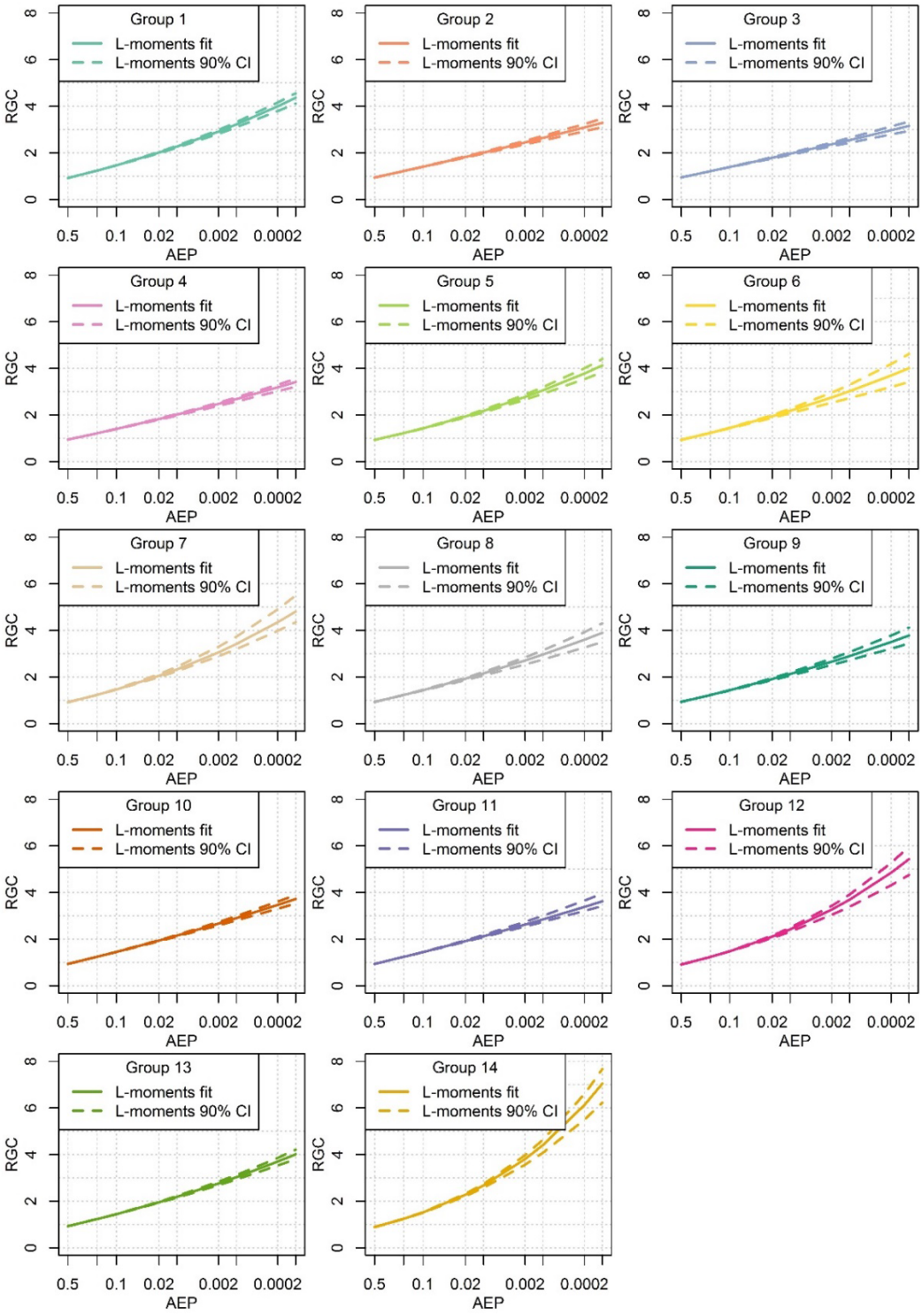


Figure 4-13 One-day RGCs and 90% Confidence Intervals for Each Group. The X-axis Represents Annual Exceedance Probability, and the Y-axis Represents the Magnitude of the Dimensionless RGC. Colors Correspond to the Groups from Figure 4-9.

Results in Figure 4-13 indicate that the range of uncertainty in the RGCs, as estimated by this bootstrapping resampling procedure, varies by region. Groups with fewer gauges, such as group 7, have wider uncertainty bounds at an AEP of 1E-4 than groups with more gauges, such as group 1. Group 6, which is composed of only five gauges, has the largest range of uncertainty at an AEP of 1E-4. Groups that contain gauges located across complex topography (e.g., Group 14) also have ranges of uncertainty that are greater than groups within non-orographic regions, such as the western region of the TRVW. This is a source of uncertainty that, although inconvenient for decision making, cannot be reduced further without additional data. If a user were interested in reducing this uncertainty, the homogeneous regions could be redefined to include fewer groups with more gauges in each group. However, the potential reduction in uncertainty from defining fewer groups may minimize the detail given to precipitation gauges located in orographic regions and ultimately result in orographic influences being minimized or ignored.

In addition to presenting the group-specific RGCs and associated uncertainty, we compute two site-specific precipitation-frequency analyses with associated uncertainty estimates for two example sites located in the TRVW. This is accomplished by scaling the RGC for each corresponding group by the at-site mean from the site of interest. Due to differences in the shape of the RGCs of groups 7 and 10, we compare site-specific precipitation-frequency curves for these groups. Figure 4-14 shows the distribution of the three parameters used to describe the GEV distribution for group 10 and group 7 based on the bootstrapping resampling approach. The scale parameters for both groups are consistently greater than zero, meaning that the distributions are lower-bounded. Note, however, that the definition of the shape parameter in the L-moments methodology is inconsistent with the literature, in that there is a factor of negative one (-1) in the L-moments definition. We multiply all shape parameters by a factor of negative one for presentation purposes and to be consistent with the sign of the parameter in the Bayesian analysis, to be discussed in section 4.3.4.

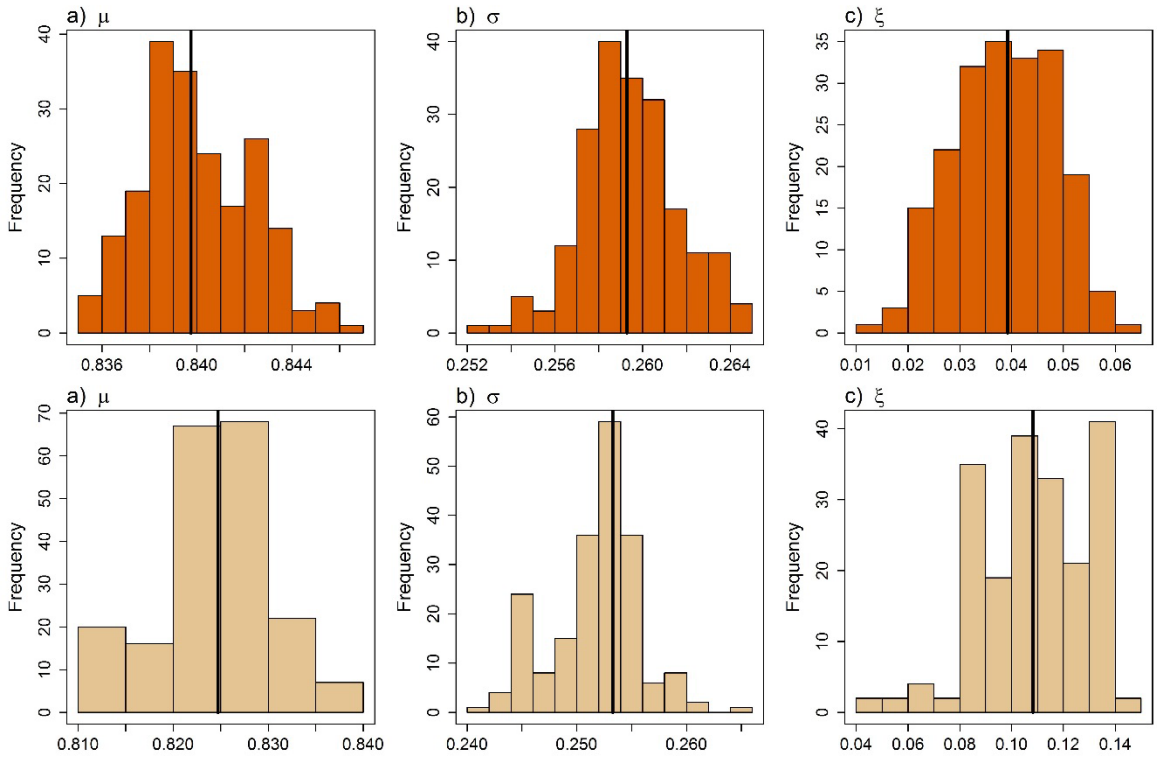


Figure 4-14 Distribution of Location (a), Shape (b), and Scale (c) Parameters Used to Describe the GEV Distribution for Group 10 (top) and Group 7 (bottom). Sample Median Indicated by the Vertical Black Line. Note Different X- and Y-axes.

Figure 4-15 shows the site-specific precipitation-frequency relationships developed for group 7 and group 10, using the parameter estimates from Figure 4-14. These results are site-specific, which means they represent precipitation totals over a 10 mi^2 area (Schreiner and Riedel 1973). These relationships were developed by scaling each respective group's RGC by an at-site mean, namely the long-term average annual maximum one-day precipitation total for the site of interest. Return levels for the sample site in group 10 range from 7.2 inches at the 100-year return period, to 12.5 inches at the 10,000-year return period. The range of the 90% confidence limits is 0.22 inches and 1.27 inches for the 100- and 10,000-year return periods, respectively. Return levels for the sample site in group 7 are greater than those for the site in group 10 and range from 7.6 inches at the 100-year return period, to 15.7 inches at the 10,000-year return period. As with the best estimates, those estimates developed using all available observations, the range of the 90% confidence limits at the group 7 site is 0.62 inches and 3.75 inches for the 100- and 10,000-year return periods, respectively. The P-F curve on the left has far less uncertainty, due to its location on the western side of the watershed, thus it is less affected by orographics. The P-F curve on the right has greater shape (as can be seen by its concave appearance and overall greater estimates at rare return periods). The greater uncertainty in the rare return period estimates can be attributed not only to orographics but also a smaller sample size.

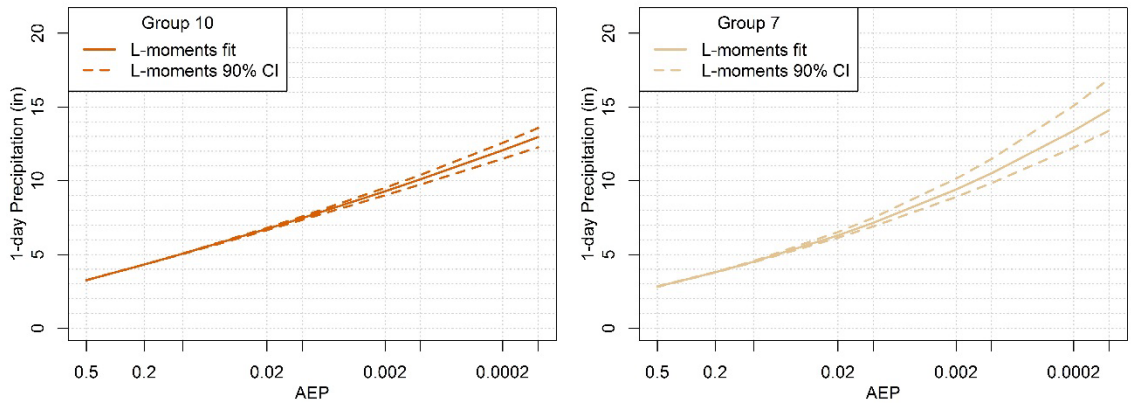


Figure 4-15 One-day Precipitation-Frequency Results for an Example Station in Group 10 (left) and an Example Station in Group 7 (right) Based on L-Moments. Colors Correspond to the Groups from Figure 4-9.

4.3.4 Bayesian Analysis

As mentioned in section 4.1, the RFA methodology may be used with any parameter estimation technique. In this subsection, we illustrate the benefits of using Bayesian inference to build distributions of RFA parameters. Bayesian inference uses a sampling scheme (typically a form of Monte Carlo), a priori information (prior distributions), and likelihood functions to build distributions of RFA parameters (posterior distributions). This concept is based on the inherent uncertainty that comes with parameter estimation techniques. Once the Monte Carlo sampling scheme begins, it must be allowed sufficient time to warm up. This is also known as the "burn-in". Upon completion of the burn-in, the sampling scheme has converged on the optimal parameter values, and subsequent samples that pass acceptance criteria (e.g., Metropolis algorithm; Metropolis et al. 1953) are included in the posterior distribution. The burn-in will vary with application and is usually determined by visual examination of trace plots (Carpenter et al. 2016), which describe the temporal evolution of the sampled (and accepted) values.

For this application of Bayesian inference, we use the rstan[10] package (Stan Development Team 2016a, 2016b) in the R statistical library (R Core Team 2016). The maximum likelihood estimation (MLE) method is used to define strictly informative prior distributions. The standard deviations of these prior distributions are set to twice the standard errors of the MLE. Setting the standard deviation to twice the standard errors of the MLE was a subjective decision made to provide significant information via the prior distributions, but also consider the possibility that the standard error of the MLE underestimates parametric uncertainty. Sensitivity testing indicates that there are no significant differences between the modes of the posterior distributions built using strictly informative (MLE) and flat uninformative (Uniform) prior distributions; however, there is a reduction in variance (i.e., parametric uncertainty) with the use of strictly informative priors (see section 4.3.4.1). The No-U Turn Sampler (Hoffman and Gelman 2014), which uses a form of Hamiltonian Monte Carlo, is used to build the posterior distributions. After visual inspection of test model runs, the burn-in period is defined as 5,000 iterations, followed by 5,000 iterations to build the posterior distributions. To be clear, there are 5,000 parameter estimates that make up each posterior distribution. This, in turn, leads to 5,000 frequency curves, which is a robust way of quantifying parametric uncertainty.

[10] ftp://cran.r-project.org/pub/R/web/packages/rstan/rstan.pdf

We begin with the same quality-controlled dataset as that used in the L-moments analysis. One very important and fundamental difference of the Bayesian method as compared to L-moments is that once quality control has been done, there are no further tests for homogeneity. That is, the Bayesian method could be including data that the L-moments method discards as discordant. Another important difference between the Bayesian method and L-moments is that we are only considering the GEV distribution in the Bayesian method, due to its flexibility and ability to mimic a number of extreme value distributions. By default, the regional L-moments approach considers six different distributions. As before, all maxima are scaled by their site-specific mean and combined into a pool of observations, to which GEV parameters are estimated using rstan and the No-U Turn Sampler.

For each homogeneous region, the scaled annual maxima (e.g., one-day annual maxima divided by the at-site mean) are combined to a single vector and used to drive the sampling scheme. Random samples over all real space are used to initialize the sampling scheme, which is an effective way to determine whether there are multiple optimal values of the parameters. That is, if the posterior distributions are multi-modal, it is likely there is no single optimal parameter set. Figure 4-16 shows the posterior distributions for the location (μ), scale (σ), and shape (ξ) parameters for the RGC of two sample groups, groups 10 and 7 (see Figure 4-9). For comparison, the MLEs are also shown as vertical lines. The range of the x-axes represents the range of parameter values for all of the fourteen groups (not shown). Note the MLEs and modes of the posterior distributions are consistent, however the full posterior distributions provide empirical representations of parametric uncertainty, which are invaluable for risk-informed decision making.

Given the posterior distributions of GEV parameters, it follows that RGCs may be calculated from the quantiles of the associated GEV distributions. As there are 5,000 distinct parameter values in each of the posterior distributions (i.e., μ , σ , and ξ), it follows that there are 5,000 unique RGCs. These 5,000 unique RGCs can collectively be thought of as the RGC posterior distribution. The median of the RGC posterior distribution can be used as the “best fit”, while the 5th and 95th percentiles represent the 90% credible interval (CI). The concept of a credible interval is akin to the traditional confidence interval, though the difference is the credible interval is empirical (calculated from the 5,000 unique RGCs in the posterior distributions), while the confidence interval is theoretical (estimated using assumptions of standard error and normality). Figure 4-17 shows the median RGCs and associated 90% CI for the fourteen groups. As can be seen, there is a wide variety of shapes and ranges of uncertainty in the RGCs between the groups, which indicates that there is significant variability in one-day annual maximum precipitation in the TRVW.

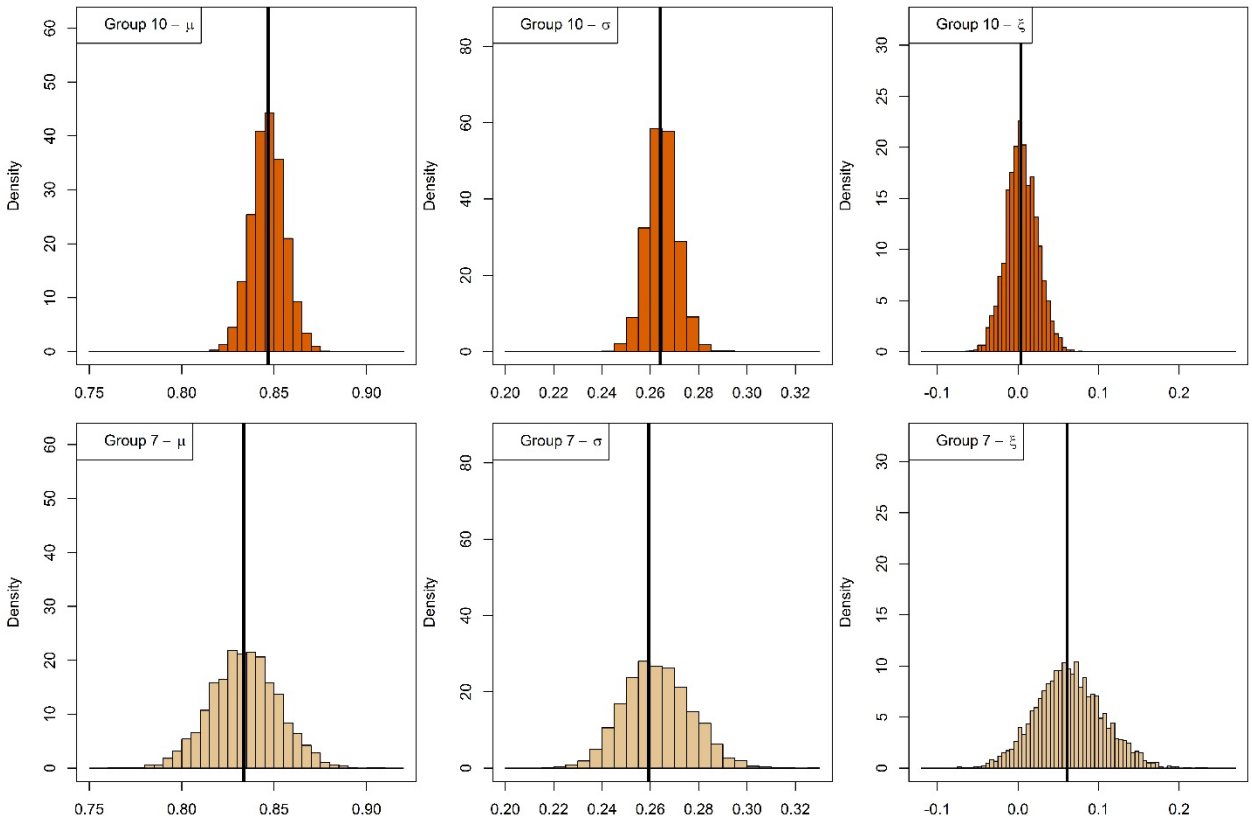


Figure 4-16 Posterior Distributions of Location (μ, left), Scale (σ, center), and Shape (ξ. right) Parameters for Group 10 (top panels) and Group 7 (bottom panels). Maximum Likelihood Estimates are Shown as Vertical Lines. Colors Correspond to the Groups from Figure 4-9.

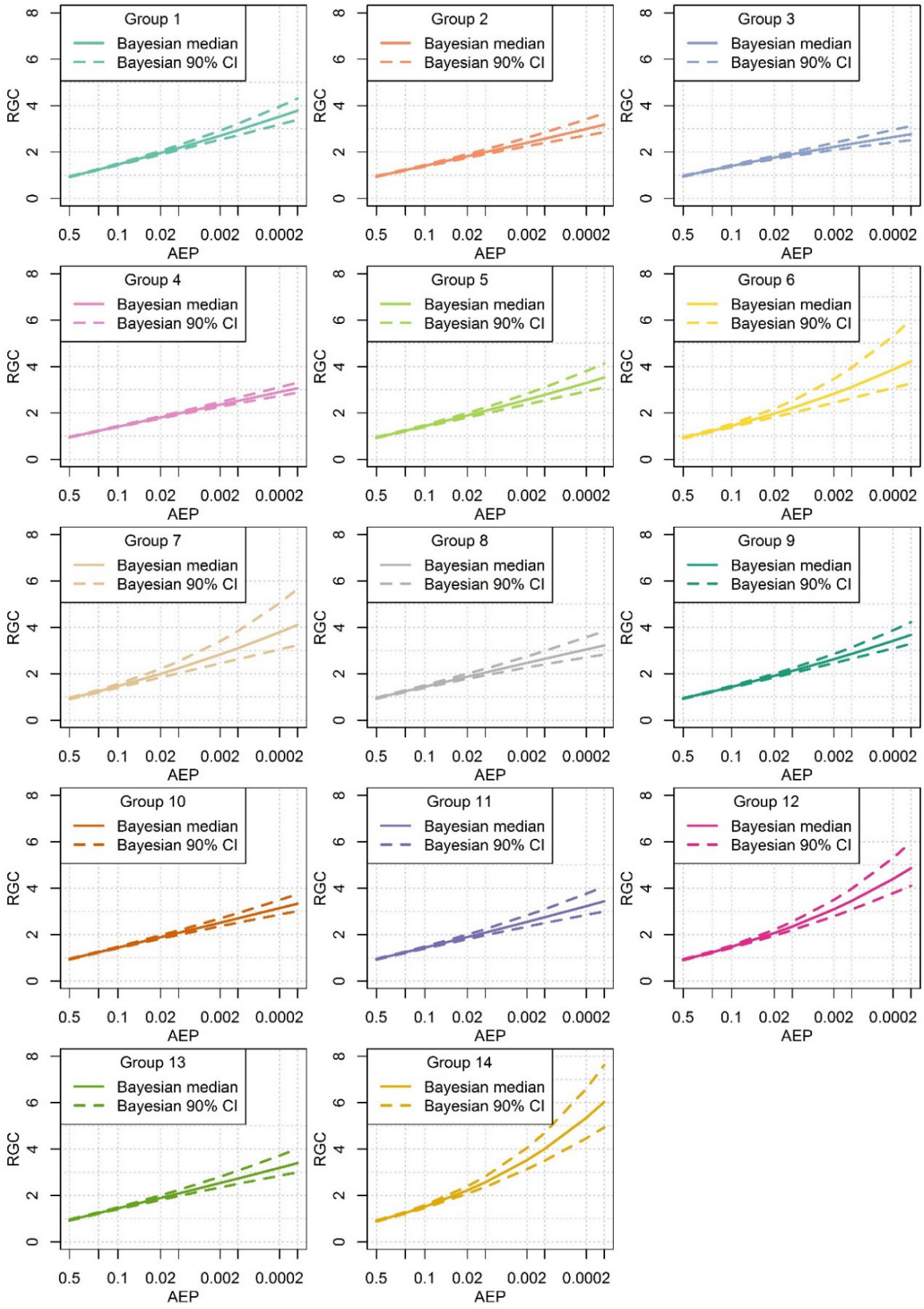


Figure 4-17 Bayesian Method Median Regional Growth Curves for all Groups. Colors Correspond to the Groups from Figure 4-9.

The information contained within an RGC is assumed to be a scaled, dimensionless precipitation-frequency (P-F) relationship that describes all locations within its defined HR. To obtain a site-specific P-F curve, the RGC is multiplied by the at-site mean that was used to scale the data prior to analysis. As an example, Figure 4-18 shows two P-F curves, one specific to a GHCN-Daily site located in group 10; the other specific to a site located in group 7. As can be seen, the shapes of the P-F curves are identical to those of the RGCs (see Figure 4-17). The P-F curve on the left has far less uncertainty, due to its location on the western side of the watershed, thus it is less affected by orographics. The P-F curve on the right has greater shape (as can be seen by its concave appearance and overall greater estimates at rare return periods). The greater uncertainty in the rare return period estimates can be attributed not only to orographics but also a smaller sample size. As can be seen in Figure 4-9, group 10 covers a far larger region, and therefore contains more stations, than does group 7. Figure 4-8 quantifies this, noting that group 10 contains 25 stations, while group 7 contains only six stations.

As mentioned before, we produce estimates at the 10,000-year return period, however, our true confidence ranges from 444- to 2,620-years, depending on the group (see section 4.3.3). In summary, we equate the total number of observations in each group to the length of the period of record. The period of record varies by group, with a minimum of 222 observations, and a maximum of 1,310 observations. As a result, our true confidence ranges from 444- to 2,620-years. Although we produce 10,000-year P-F estimates, they must be interpreted with caution, while the considerable amount of epistemic uncertainty at this return period must not be ignored.

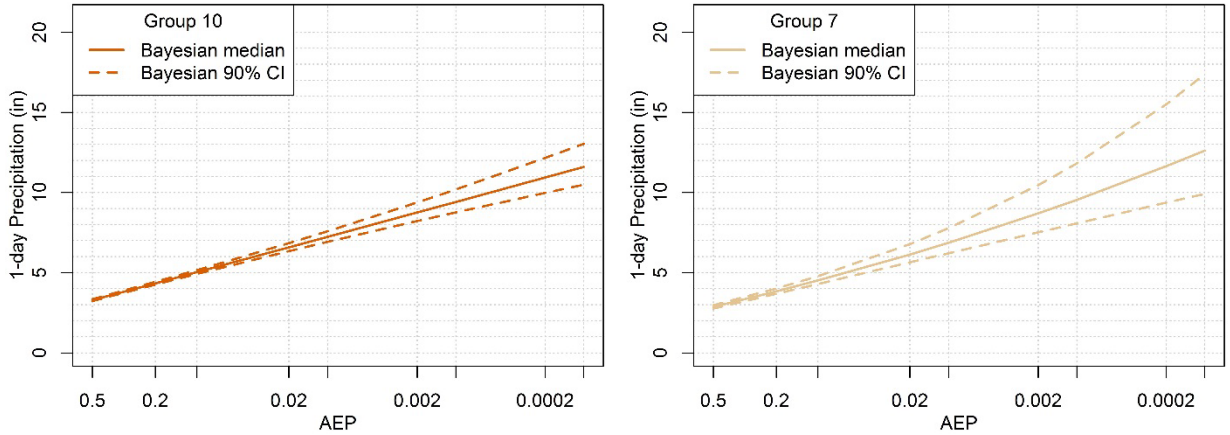


Figure 4-18 Bayesian Method Precipitation-Frequency Curves for an Example Station in Group 10 (left) and an Example Station in Group 7 (right). Colors Correspond to the Groups from Figure 4-9.

4.3.4.1 Sensitivity Testing

It is of interest to ensure that the Bayesian model is defined such that uncertainty is properly modeled and quantified. Because a posterior distribution is developed as the product of a prior distribution and a likelihood function (e.g., GEV), it is prudent to determine the sensitivity of the posterior to the definition of the prior. By comparing the mode and spread of the posteriors, it can be determined, both qualitatively and quantitatively, whether an informative prior will have a significant impact on the outcome of the posteriors. A qualitative analysis could include a visual inspection of all posteriors on the same plot and a comparison of median and credible interval RGCs calculated from the posteriors. Additionally, a quantitative analysis should be included to substantiate the findings of a qualitative analysis. Thus, a Kolmogorov-Smirnov test may be

performed, wherein the null hypothesis is that two separate groups of data come from the same distribution. A large p-value (> 0.05 at the 95% confidence level) would fail to reject the null hypothesis that posteriors produced by Models 1 and 2 (defined below) were drawn from the same continuous distribution and, therefore, the posteriors are insensitive to the definition of priors. A p-value less than 0.05 would indicate that the posteriors were not drawn from the same distribution, which could indicate either a difference in mean, standard deviation, or both.

To this end, two distinct models were defined, which differ by the definition of priors. Model 1 is given uninformative Uniform priors, which span all real bounded space:

$$\mu_i \in [-2, +2],$$

$$\sigma_i \in [0,2], \text{ and}$$

$$\xi_i \in [-0.5, +0.5],$$

for $i = 1, \ldots, N_g$, where N_g represents the number of groups, or homogeneous regions, in the analysis. For this application, $N_g = 14$. Model 2 is given informative truncated-Normal priors; the Normal distributions are truncated at the bounds defined above. Consistent with the previous definition of informative priors from section 4.3.4, the means of the priors are centered on the MLEs; the standard deviations of the priors are set to twice the standard error of the MLEs.

Figure 4-19 shows the priors and posteriors for group 10 and group 7 using Models 1 and 2. As can be seen, the posteriors have very similar modes (and medians), while the ranges are notably tightened with an informative prior (Model 2). Based on this analysis, the use of a well-defined prior distribution is one way to reduce the range of uncertainty in the posterior distributions.

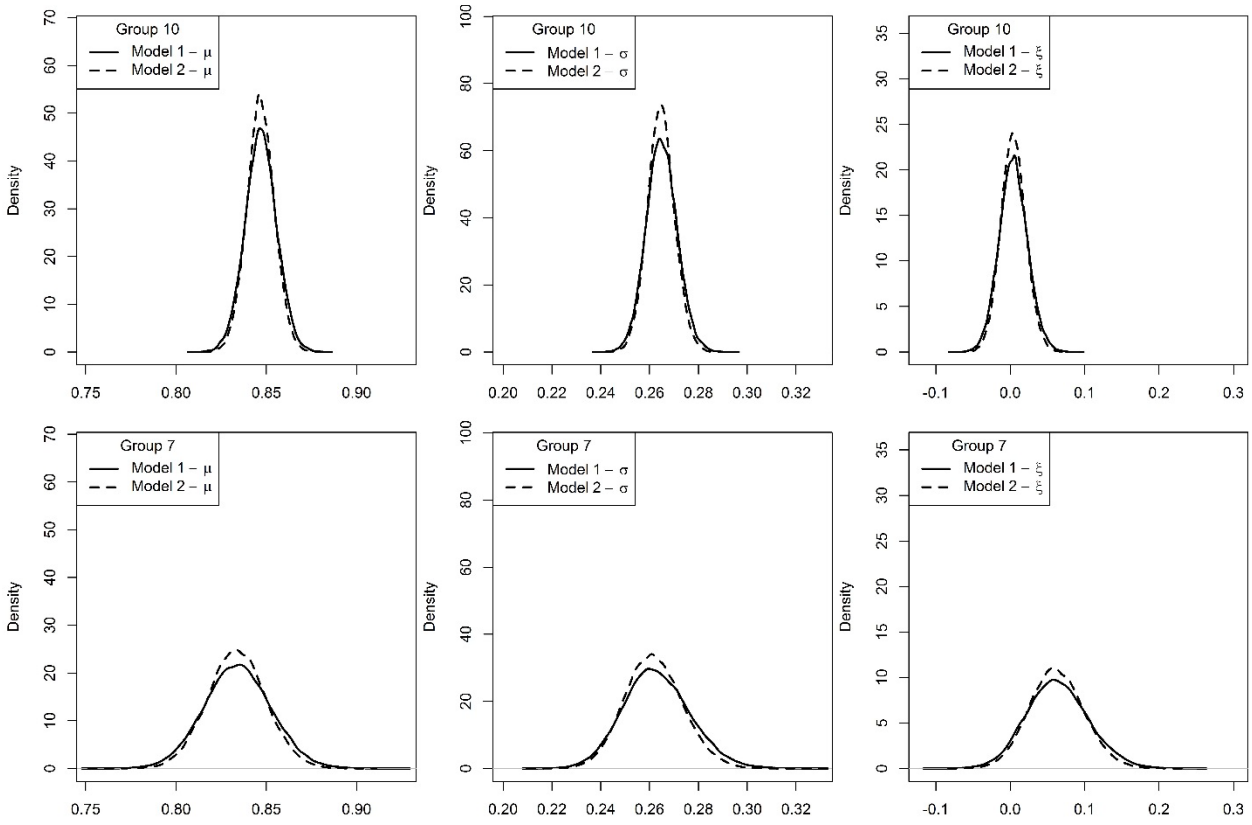


Figure 4-19 Posterior Distributions of GEV Parameters (from left to right: μ, σ, ξ) for Group 10 (top row) and Group 7 (bottom row), from Models 1 and 2

Figure 4-20 shows the RGCs for Groups 10 and 7, as computed from the posteriors from Models 1 and 2. It can be seen that Model 2, which has informative priors, has consistently less uncertainty in the rare events than does Model 1, albeit marginally. Thus, when prior information is available, it should be included in the model, even for marginal benefit.

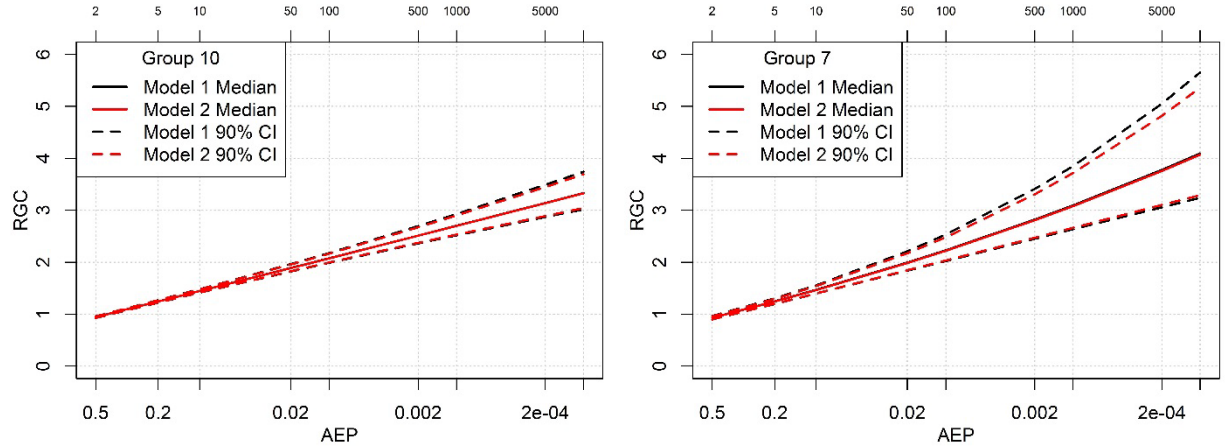


Figure 4-20 Dimensionless RGCs for Group 10 (left) and Group 7 (right), as Computed from Model 1 (black) and Model 2 (red)

Kolmogorov-Smirnov (K-S) tests were performed for the posteriors of the GEV parameters and the RGCs produced from them. Both Models 1 and 2 generated posterior distributions of 5,000 samples, thus for each K-S test we are comparing the distribution of 5,000 samples from Model 1 to 5,000 samples from Model 2. For all K-S tests, the p-values are essentially zero, indicating that the samples from Models 1 and 2 are not drawn from the same continuous distribution. As mentioned previously, the mode (and median) of the posterior distributions are similar between Models 1 and 2, while the range (and variance) of the posterior distributions shrink from Model 1 to Model 2. These K-S tests confirm that the definition of informative priors has significant effect on the variance of the posteriors.

Note that the uncertainty quantified by Bayesian inference is not necessarily the total uncertainty of the modeled process. The ranges of uncertainty provided in the RGCs and P-F curves are a quantification of parameter estimation uncertainty. Other sources of uncertainty, including model selection, data quality, data availability, to name a few, are not explicitly modeled in this application of a Bayesian framework.

4.4 Discussion

In section 4.3, we present two alternative methods to estimate regional precipitation-frequency analyses in regions of complex terrain across the TRVW: an L-moments approach and a Bayesian inference approach. Resulting RGCs from both methods are used in conjunction with at-site mean precipitation observations to produce site-specific precipitation-frequency relationships. In this section, we discuss some of the similarities and differences in results between the two methods.

The primary (and possibly only) similarity between these two methods is the pool of observations used to inform them. Each frequency analysis began with the same homogeneous region and pool of historical observations. However, in the regional L-moments analysis, we flag and remove gauges that are discordant. The Bayesian approach does not use a discordancy measure to filter observations. As a result, there are groups that have a differing number of gauges between the two methods. This difference, which may sound insignificant, can ultimately impact the RGCs developed for a region.

RGCs from the two methods are shown in Figure 4-13 and Figure 4-17. At first glance, the two sets of RGCs appear similar. For example, the RGCs between the two methods are characterized by similar curvature (as dictated by the GEV shape parameter). However, groups 4, 8, and 13 appear to have differing curvature. The L-moments RGC for group 4 appears to have an upper bound, while the same does not appear true for the group 4 Bayesian RGC. In addition to differences in curvature, there are also many differences in the magnitude of RGCs. For example, the L-moments best-estimate RGC magnitude for group 14 at the 10,000-year return is approximately 7, while that from the Bayesian model is 6. While the difference between these two methods sounds small, there are implications when these two RGCs get scaled by an at-site mean.

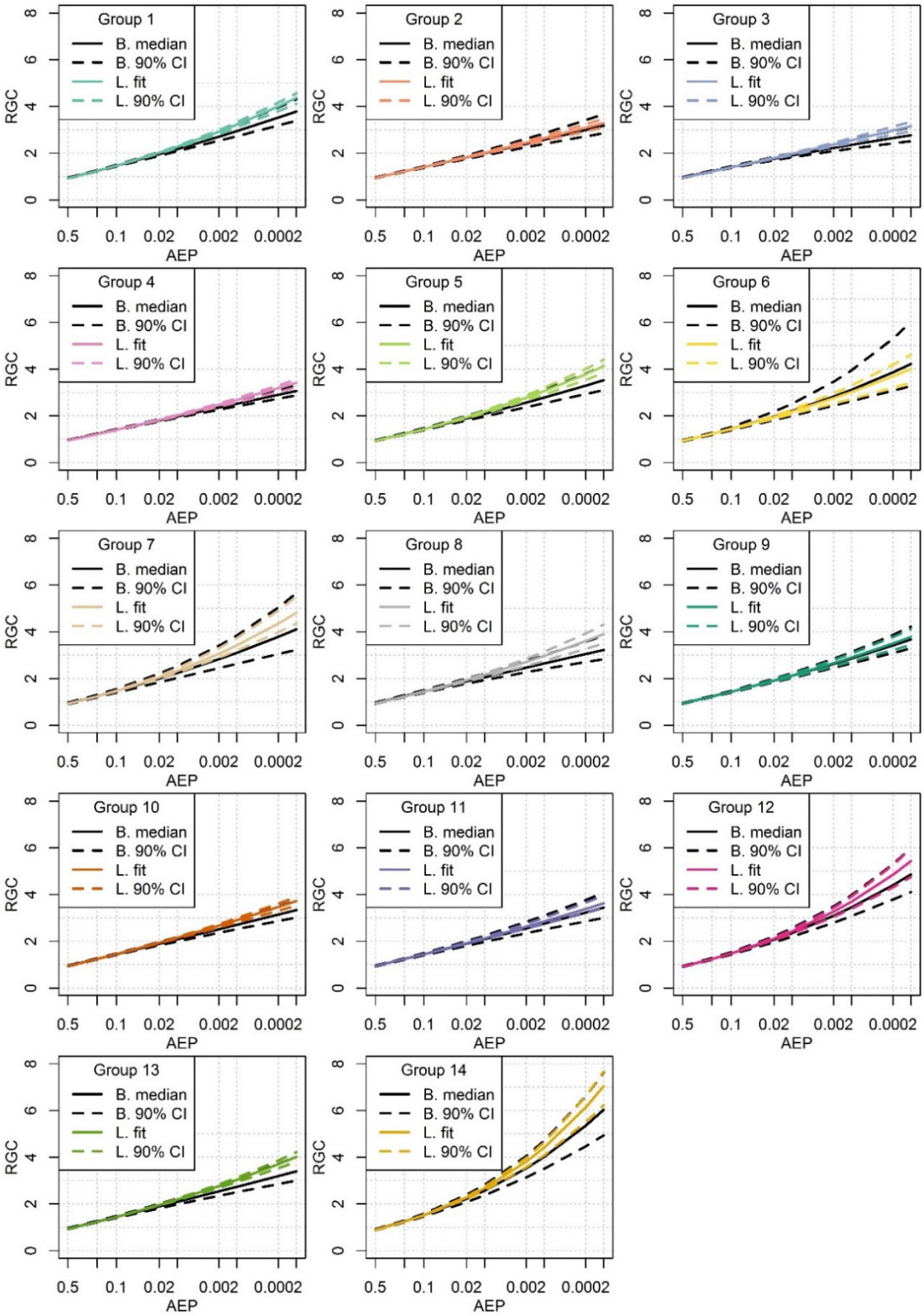


Figure 4-21 Group-Specific RGCs and Associated Uncertainty Based on the Regional L-Moments Approach (color) and Regional Bayesian Approach (black)

Results in Figure 4-21 also highlight noticeable differences in the range of uncertainty between the two methods. At the 10,000-year return period, the range of uncertainty from the Bayesian analysis exceeds the range of uncertainty from the L-moments analysis for each group. These differences are the result of estimating uncertainty differently. The L-moments uncertainty analysis is based on resampling the pool of historical observations, which primarily captures uncertainty in the RGC as a function of precipitation observations. Conversely, the Bayesian uncertainty analysis estimates are empirically based, and come as a result of the Monte Carlo resampling scheme.

Differences in uncertainty ranges are also visible in the results of Figure 4-15 and Figure 4-18, although influenced by each at-site mean. An illustration of these differences can be seen more clearly in Figure 4-21 because the RGC is a dimensionless representation of the site-specific precipitation-frequency curves in Figure 4-15 and Figure 4-18. The range of uncertainty in the RGC for group 7 is nearly double for the Bayesian analysis as compared to the L-moments analysis. The lower confidence bound as estimated by L-moments is greater than the median estimate from the Bayesian analysis. Similarly, the upper bound as estimated by L-moments is less than that from the Bayesian analysis. The range of uncertainty at the 10,000-year return period is approximately 3.70 inches from L-moments, and approximately 7.5 inches from the Bayesian analysis. The L-moments uncertainty bounds are extremely narrow, likely due to the estimation method. The bootstrap resampling procedure represents more of a sensitivity tool, rather than a formal uncertainty method. In this regard, Bayesian inference methods provide more trustworthy estimates of uncertainty. One may suggest that lower ranges of uncertainty can help improve decision making; however, this is not always true.

Another issue to discuss is stationarity, the idea that natural systems exist and vary within a constant envelope of variability (Milly et al. 2008). Stationarity assumes that a given variable has a time-invariant probability density function, the properties of which can be estimated from the observational record. Probability density function estimation errors are often acknowledged under stationarity, but are considered reducible by including additional observations, using more efficient estimators, or including regional and/or paleo records (Milly et al. 2008). The water management community has designed and operated infrastructure under the assumption of stationarity for decades. Similarly, both regional frequency analyses presented in section 4.3 follow the assumption of stationarity. However, observed trends and future projections of air temperature, precipitation, streamflow, and extreme events across the globe (IPCC AR5) now support the claim that stationarity is dead (Milly et al. 2008). As a result, future research should consider the use of tools that incorporate time-varying conditions.

There are multiple sources of uncertainty in any precipitation-frequency analysis, including assumptions regarding stationarity, the dataset selected for analysis, the identification of stations, the period of record, the method and selection of the “best” frequency distribution, the computation of an at-site mean, and the method used to estimate uncertainty. While limited estimates of uncertainty may lead decision makers to assume the system is better understood than it is in actuality, there is no absolute way to prove one set of precipitation-frequency estimates and uncertainty bounds are indisputably better than another set.

5 GRIDDED FREQUENCY ANALYSES

5.1 Introduction

Observation-based precipitation-frequency analyses are inherently limited by the spatial distribution and availability of gauge records. In some regional frequency analyses, regional growth curves (which are assumed valid over an entire homogenous region) are scaled by at-site means, to produce site-specific precipitation-frequency relationships that are often assumed valid over 10 mi^2 (Schreiner and Riedel 1973). A problem arises with this method when station data are unavailable at a location of interest. For example, there are no GHCN-Daily gauges that meet the data criteria discussed in section 4.3.1 in the region northeast of Knoxville, Tennessee (Figure 5-1). Gridded datasets, however, ameliorate the problem of missing data. Figure 5-2 shows average annual precipitation estimates between 1980 and 2012 from the combined Newman et al. (2015) ensemble. While there are weaknesses to gridded precipitation datasets (e.g., Daly et al. 1994), they represent a unique opportunity to study extreme precipitation events in regions of complex topography where point observations are often incomplete and sparse. Gridded datasets also represent a potential input for more advanced hydrologic models that require gridded input fields. Note that the "best" gridded dataset for any application depends on a number of factors, including, but not limited to, the required spatial and temporal resolution of the dataset.

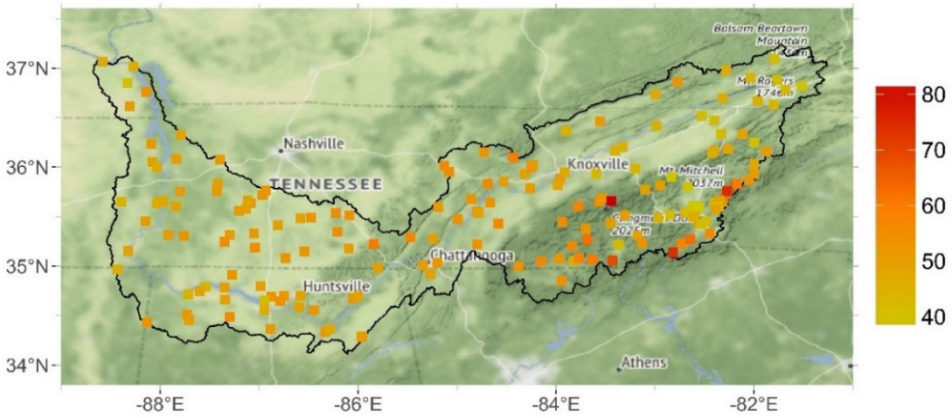


Figure 5-1 Average Annual Precipitation (inches) at GHCN-Daily Gauges with 85% Data Coverage for 10 or More Years

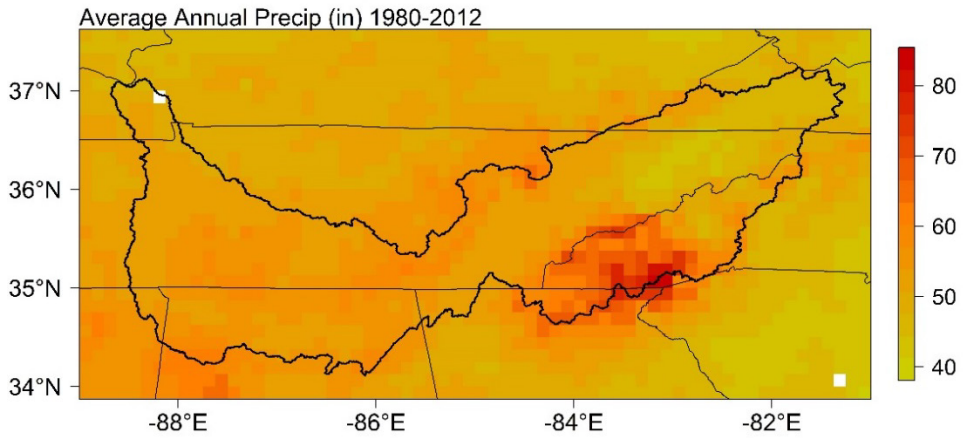


Figure 5-2 Average Annual Precipitation (inches) Between 1980 and 2012 from the Combined Newman et al. (2015) Dataset. Missing Values Represent Grid Cells that are Primarily Water.

Recently, gridded frequency analyses have been completed using a variety of precipitation datasets. Overeem et al. (2009) estimate precipitation return levels for durations of 15-minutes to 24-hours across the Netherlands using an 11-year (1998-2008) radar dataset available at 2.4 km horizontal resolution. Radar data are adjusted toward point observations by combining an hourly mean-field bias adjustment with a daily spatial adjustment. The authors assume the data follow a GEV distribution, where only the location parameter varies spatially (i.e., the scale and shape parameters are constant in space). The authors compare the 10- and 50-year quantiles from the radar dataset to those estimated using point observations. Results indicate that precipitation return level estimates from the radar dataset exceed those from point observations by 7-9% at the 24-hour duration. The opposite is true at the hourly duration; point-observation return levels exceed those from the radar estimates by 6-8%. This study represents a general advancement to PFA.

Unlike Overeem et al. (2009), Früh et al. (2010) perform seasonal precipitation-frequency analyses using two gridded precipitation datasets valid between 1971 and 2000. The gridded datasets are based on interpolating point observations recorded in southwest Germany to two different regional climate model (RCM) grids, while explicitly accounting for the underlying topography (Früh et al. 2007). Fruh et al. (2010) combine the peak-over-threshold approach with the method of L-moments to produce estimates of summer and winter precipitation magnitudes at the 10-year return period. The authors explore sensitivities in precipitation magnitudes to the choice of distribution function, specifically the generalized Pareto distribution and the general kappa distribution. Their results indicate that the kappa distribution produces precipitation magnitudes that are closer to ranked precipitation observations at two sites in Germany.

Bracken et al. (2016) produce gridded (⅛° latitude × longitude) precipitation return periods and uncertainty estimates across the western U.S. using a spatial Bayesian hierarchical model. The authors model the annual maximum precipitation data using a Gaussian elliptical copula (Schoelzel and Friederichs 2008) with GEV marginal distributions. Spatial regression parameters are allowed to vary in space, accounting for the large climatic variations across the western U.S. Results indicate that the strongest covariate for the shape parameter (ξ) is elevation, suggesting that assuming a constant shape parameter over large domains may be problematic.

A number of other studies estimate precipitation magnitudes at various return periods using output from RCMs for the purpose of exploring potential changes in extreme precipitation events under future climate projections (Semmler and Jacob 2004; Nikulin et al. 2010; Hanel and Buishand 2010; Mladjic et al. 2011; Dominquez et al. 2012). For instance, Nikulin et al. (2010) analyze changes in Scandinavian summer and winter precipitation extremes simulated by the Rossby Center Regional Climate Model forced by six different GCMs between a historical time period (1961-1999) and future time period (2071-2100). The authors assume the seasonal maxima of precipitation follow a GEV distribution and estimate the distribution parameters using L-moments and the maximum likelihood method. The authors use the L-moments approach as the main approach for estimating return periods due to its simplicity and efficiency. However, the authors verify L-moments results with three maximum likelihood models assuming: stationary inputs, a linear trend in the location parameter, and linear trends in both location and scale parameters. The authors find no significant differences among the methods (and ultimately present only results from the L-moments analysis). The analysis is performed on each grid cell separately, where the horizontal resolution of each simulation is 0.44° latitude × longitude grid (approximately 49 km). These studies represent general advancements in the field of PFA.

Mladjic et al. (2011) explore projected changes in extreme precipitation characteristics over Canada using an ensemble of five historical and future simulations performed with the Canadian Regional Climate Model. The study focuses on changes in return levels of one-, two-, three-, five-, seven-, and 10-day annual (April through September) maximum precipitation amounts between a current period (1961-1990) and a future period (2040-2071). Two complementary methods are applied to annual maximum precipitation data to estimate return levels: a regional frequency analysis based on predefined climate regions and an individual grid box analysis. In both analyses, the authors utilize the Z test developed by Hosking and Wallis (1997) to identify the most appropriate frequency distribution. Results from the analysis suggest statistically significant increases in event magnitudes for seven out of the 10 climate regions defined across the country. This study represents a general advancement in the field of PFA.

In this section, we build on previous studies and our results from section 4 by performing two precipitation-frequency analyses, L-moments and Bayesian, applied to the 100-member ensemble dataset of Newman et al. (2015). This dataset, with all 100 members, represents an opportunity to estimate precipitation magnitudes across the entire TRVW at return periods greater than those provided in NOAA Atlas 14. We estimate return periods greater than those provided by NOAA Atlas 14 (i.e., greater than 1,000-year return period) because we assume that records from each ensemble member and grid cell represent independent regional observations from sites within the grid cell (100 ensemble members, 33 years of data, N=3,300). With a time series of 3,300 records, we can estimate return periods as rare as 6,600-years (Bocchiola et al. 2008). Note that the assumption of independence among members is not entirely valid, as the temporal correlations of annual maximum time series among all ensemble members and grid cells range from -0.63 to 0.95. The median correlation among all ensemble members and grid cells is 0.08, while the mean is 0.1. Independence in the annual maximum precipitation time series is true for some combinations of ensemble members, but clearly not all. We present results for the 10,000-year return period with a note of caution, as these results are beyond the range of confidence (two times the period of record with dependence among ensemble members). However, the confidence in the 10,000-year estimates from a time series with 3,300 records is greater than what could be obtained through traditional datasets and methods. Although we focus our gridded analysis on the Newman dataset, the methods used may be applied to any valid gridded precipitation dataset of interest.

The remainder of this section is organized as follows. Section 5.2 includes a review of the Newman dataset and two methods used to estimate grid-cell specific precipitation-frequency analyses. L-moments and Bayesian results are presented in section 5.3. Finally, we include with a comparison of L-moments and Bayesian results in section 5.4.

5.2 Data and Methods

As discussed in section 3.1.4, Newman et al. (2015) produced a 100-member ensemble dataset of historical precipitation and temperatures across CONUS from 1980 to 2012. This dataset is available at a daily time step on a ⅛° latitude × longitude grid. Newman et al. (2015) compare average daily precipitation from the ensemble mean with three alternative gridded precipitation datasets, NLDAS (Xia et al. 2012), Daymet (Thornton et al. 2014), and Maurer (Maurer et al. 2013). Differences among the four precipitation products are small, with a few regions of consistent differences located in areas of complex topography. Newman et al. (2015) show that, in some cases, the ensemble produces too much lee-side precipitation, as compared with the three other products.

We perform frequency analyses on one-, two-, and three-day annual maximum values at each grid cell separately using two methods, an L-moments approach and a Bayesian inference approach (for details on these methods, see sections 4.3.3 and 4.3.4). For both cases, annual maximum precipitation data from every ensemble member (at each grid cell) are aggregated to produce a single time series (hereafter referred to as the combined ensemble). As there are 33 years of data and 100 ensemble members[11] in the Newman et al. (2015) dataset, we ultimately have 3,300 annual maximum precipitation values at each grid cell. We assume that annual maxima for each duration and grid cell follow a GEV distribution and estimate precipitation totals at the 5-, 25-, 50-, 100-, 500-, 1,000-, and 10,000-year return periods.

Estimates of precipitation at the 10,000-year return period are rarely, if ever, presented in the literature due to the compounding effects of aleatory variability and epistemic uncertainty (e.g., model selection, data quality, data availability, etc.). Indeed, previous works (Bocchiola et al. 2008, and references therein) suggest that users should use caution when estimating precipitation return levels at return periods greater than two times the period of record. We assume that annual maximum precipitation events from each ensemble member are regional observations within the grid cell (resulting in 3,300 annual maximum observations within each grid cell). As a result, our true confidence limit in P-F estimates is at the 6,600-year period. Although we produce 10,000-year P-F estimates, they must be interpreted with caution because the considerable amount of epistemic uncertainty at this return period cannot be ignored. With this in mind, we quantify the epistemic uncertainty in gridded estimates using the same approaches described in sections 4.3.3 and 4.3.4. Although results are available for each method, duration, and return period, we present L-Moments and Bayesian inference results only for the one-day duration for brevity.

We compare gridded precipitation magnitudes with existing estimates from NOAA Atlas 14 (hereafter referred to as NA14). We compare precipitation magnitudes from the Ohio River

[11] Officially, the Newman dataset includes 100 ensemble members. However, we consistently had trouble downloading three members (ensemble 13, 56, and 71). These members were ultimately left out of the analysis. Since this work represents a demonstration of frequency data and methods, this lack of members does not impact our conclusions. We refer to all ensemble members (e.g., 100) throughout the entire document.

Basin (ORB) and Southeastern States (SE) regions in NA14 against results from the combined ensemble method. Gridded point estimates from NA14 are available online through the Precipitation-Frequency Data Server[12] (Bonnin et al. 2006). The spatial resolution of NA14 output is different between the ORB and SE regions, while the spatial resolution of precipitation fields from the two regions differ from the Newman ensemble. Consequently, we regrid (or up-scale) gridded precipitation estimates from both NA14 regions (30-second or roughly 0.5 × 0.5 miles) to the same horizontal resolution of the Newman ensemble (⅛° latitude × ⅛° longitude) using linear interpolation. For example, Figure 5-3 shows the native ORB and SE NA14 (gridded) point precipitation totals for the 48-hour, 1,000-year event. The bottom panel of Figure 5-3 shows an example of those fields after the values have been re-gridded and combined across the region of interest. The regridding was applied to each NA14 duration and return period.

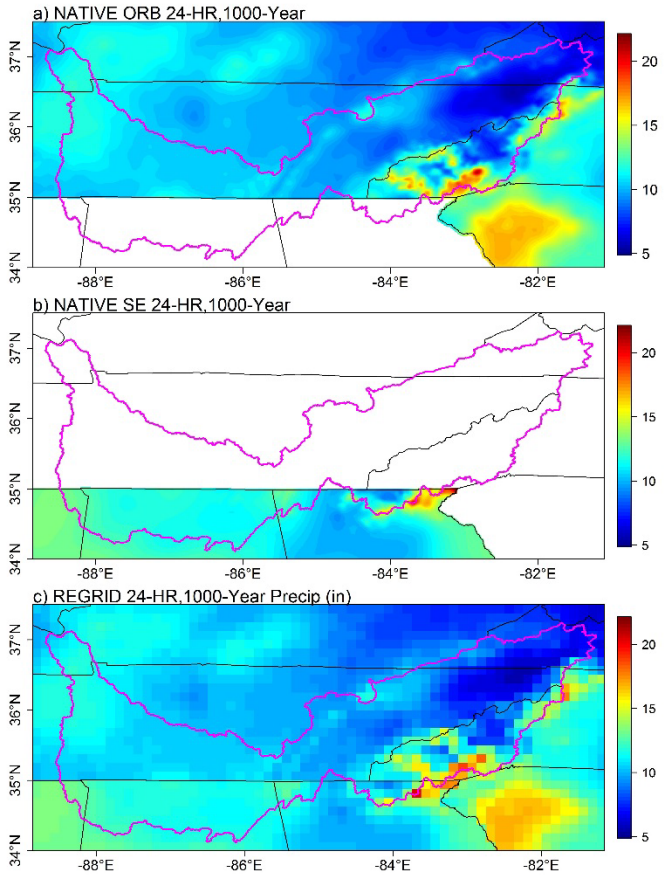


Figure 5-3 48-hour, 1,000-year Precipitation Totals (inches) across the Ohio River Basin from NOAA Atlas 14 Volume 2 (a), Across the Southeast from NOAA Atlas 14 Volume 9 (b), and Re-Gridded Estimates (inches) from Both Regions on ⅛° Latitude × Longitude Grid (c). The Magenta Line Indicates the Extent of the TRVW.

Two additional steps are required before NA14 magnitudes can be compared with computed magnitudes from the Newman ensemble. The methods employed when developing NA14 PFE involve multiplying one- and two -day observed precipitation totals by scale factors (1.13 and 1.04)

[12] http://hdsc.nws.noaa.gov/hdsc/pfds

in an effort to estimate 24- and 48-hour totals, respectively. The authors of NA14 suggest these scale factors correct daily precipitation totals, which are defined by time, for precipitation that occurs before or after a calendar day. The authors do not employ a scale factor for 72-hour totals. Thus, we divide NA14 estimates by the respective duration-specific scale factor to eliminate this correction, which is not used in the Newman ensemble. The second adjustment required to allow comparisons between NA14 and the Newman ensemble relates to spatial scale. NA14 estimates were developed to represent point estimates. Regridding these estimates to the Newman resolution does not change the fact that the NA14 estimates were computed as point estimates. As a result, we utilize three duration-specific areal-reduction factors (ARFs) to "coarsen" the magnitude of NA14 estimates to the spatial resolution of a single Newman grid cell. We utilize the empirical relationship for ARF based on work from Leclerc and Shaake (1972),

$$ARF = 1 - \exp\left(at^b\right) + \exp\left(a\,t^b - cA\right),$$

where *t* is event duration (in hours) and *A* is the area (in km^2). The coefficients *a*, *c*, and *b* were empirically fit with *a*=-1.1, *c*=2.59×10-2, and *b*=0.25 (Leclerc and Shaake 1972). Based on this expression from Leclerc and Shaake (1972), and an area of 155 km^2 for a single Newman grid cell, the 24-, 48-, and 72-hour ARFs are 0.9140, 0.9457, and 0.9601. We multiplied NA14 estimates by these scale factors to areally-reduce the precipitation magnitudes to the same resolution of the Newman dataset. Alternative expressions and/or methods exist for estimating ARFs. However, exploring the impact of ARF magnitude on differences between NA14 and the Newman dataset is beyond the scope of this report.

5.3 Results

5.3.1 Comparison with PRISM

Although Newman et al. (2015) compare precipitation results from the combined ensemble with three alternative gridded precipitation datasets, we present an additional comparison of average annual and seasonal precipitation totals from the combined ensemble with one additional dataset, PRISM. PRISM data, which are available from 1980 to 2012, were regridded to the spatial resolution of the Newman dataset in the same manner as the NA14 estimates (for details, see section 5.2).

Average annual precipitation from the combined Newman ensemble and PRISM dataset are shown in the top two panels of Figure 5-4. Differences between the two datasets are shown in the bottom panel of Figure 5-4. The combined ensemble captures the general pattern of annual average precipitation across the domain (the spatial correlation is 0.89), with high totals in topographic region of southwestern North Carolina and the rain shadow region (Konrad 1994) in northeastern Tennessee. Average annual precipitation totals from the combined ensemble are greater than the PRISM dataset in southwestern North Carolina (up to 20 inches greater) and in northeastern Tennessee (less than 10 inches greater). However, average annual precipitation totals from the combined ensemble are slightly less than PRISM across the remainder of the TRVW (Figure 5-4). The average, maximum, and minimum difference in average annual precipitation across the TRVW between the two datasets is 0.01 inches, 17.84 inches, and -12.91 inches, respectively. The large negative differences between the two datasets are primarily located in the region of greatest elevation, namely in western North Carolina. One potential explanation for the lower combined ensemble values in this region is the course resolution. The ensemble dataset may not be able to capture such large precipitation values at a resolution of ⅛°. The large positive differences, which indicate precipitation values from the

combined ensemble that exceed PRISM estimates, agree with findings from Newman et al. (2015). The authors state that the ensemble overestimates precipitation on the leeward side of some mountains.

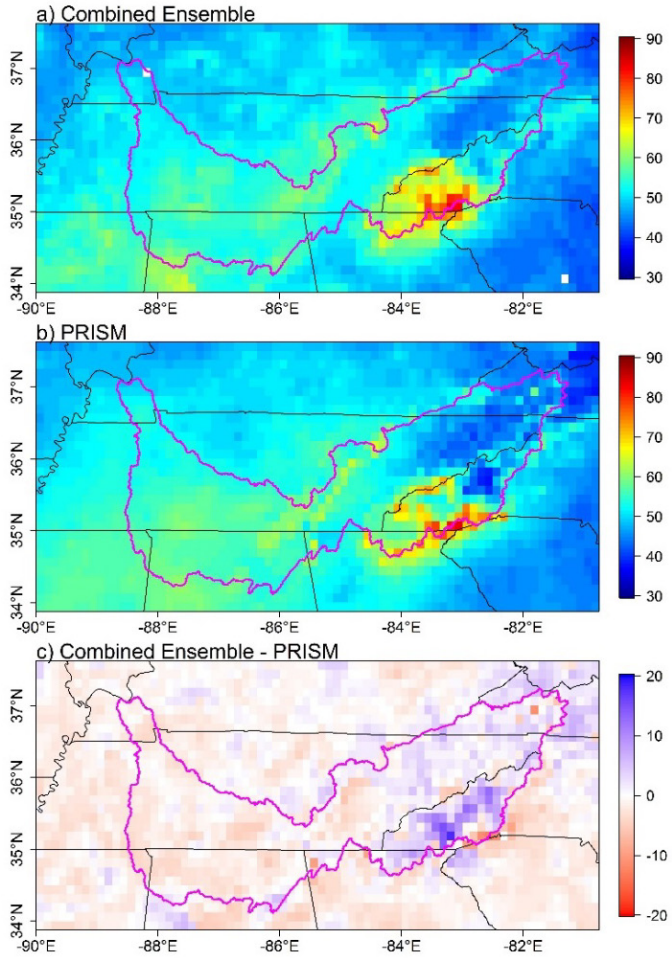


Figure 5-4 (Top) Average Annual Precipitation (inches) from the Combined Ensemble (1980-2012). (Middle) Average Annual Precipitation (inches) from PRISM Dataset (1980-2010). (Bottom) Difference in Average Annual Precipitation (inches) Between the Vombined Ensemble and PRISM Datasets. (Combined Ensemble Minus PRISM)

We present differences in seasonal precipitation, in addition to differences in annual precipitation. Figure 5-5 shows the average difference in average total precipitation during each season between the combined ensemble and PRISM dataset. Differences in seasonal precipitation between the two datasets vary in sign and magnitude throughout the year. For example, the combined ensemble tends to estimate wetter conditions in the orographic region of southwestern North Carolina during spring, summer, and fall, a signal that is present in the annual difference plot, as well (bottom panel of Figure 5-4). Seasonal precipitation totals from the combined ensemble exceed the PRISM dataset during spring and fall in the rain shadow region of northeastern Tennessee. However, the combined ensemble underestimates precipitation in that region during winter and summer.

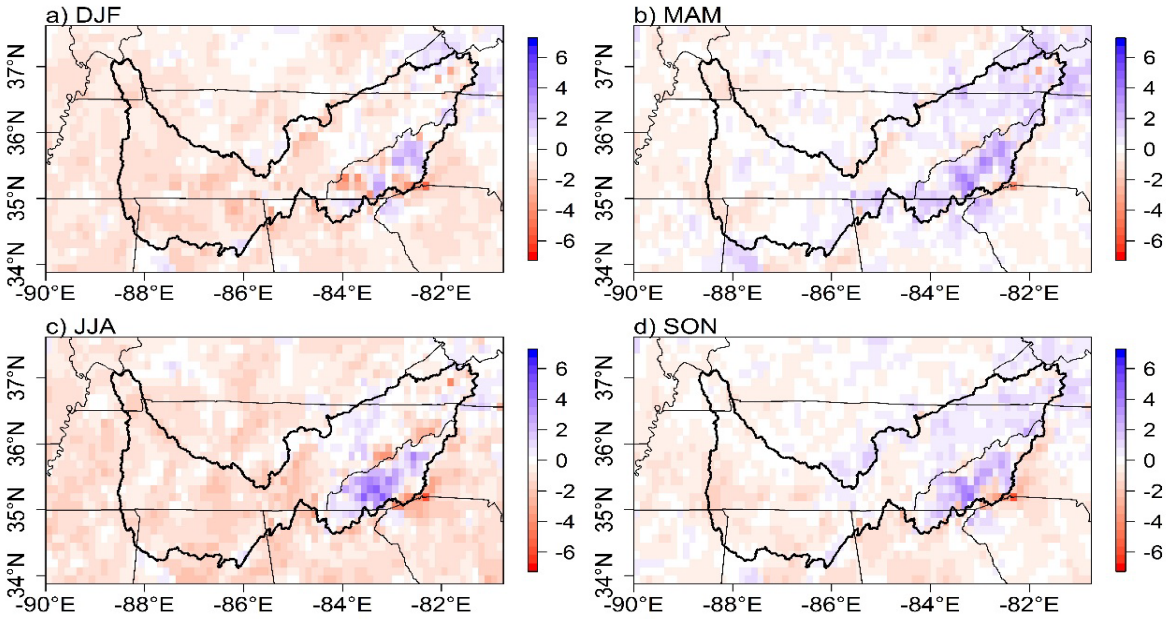


Figure 5-5 Differences in Average Total Precipitation (inches) Between the Combined Ensemble (1980-2012) and PRISM Dataset (1980-2010) During a) December Through February, b) March Through May, c) June Through July, and d) September through November. Differences are Computed as the Combined Ensemble Minus PRISM.

5.3.2 L-Moments Analysis

Ensemble average one-, two-, and three-day annual maximum precipitation totals are shown in Figure 5-6. As with average annual precipitation, the ensemble dataset captures the region of enhanced annual precipitation totals across the southern Appalachians, as well as the rain-shadow region in northeastern Tennessee. Average annual one-, two-, and three-day precipitation totals across the TRVW range from 1.6 to 4.3 inches, 2.4 to 6.4 inches, and 2.8 to 7.3 inches, respectively. The time series of annual maximum precipitation at each grid cell (for each duration) is used as input to the L-moments algorithm. Each grid cell is treated as an individual homogeneous region. Other approaches group neighboring grid cells in an effort to create homogeneous regions from grid cell data (e.g., Mladjic et al. 2011). However, we restrict our analysis to each grid cell separately, under the assumption that the 100 ensemble members represent neighboring sites and therefore will adequately represent this variability. Grouping grid cells into HRs would likely result in higher spatial correlation of results (i.e., reduced spatial variability) and sharp gradients in precipitation return periods among neighboring regions (see Figure 5 of Mladjic et al. 2011).

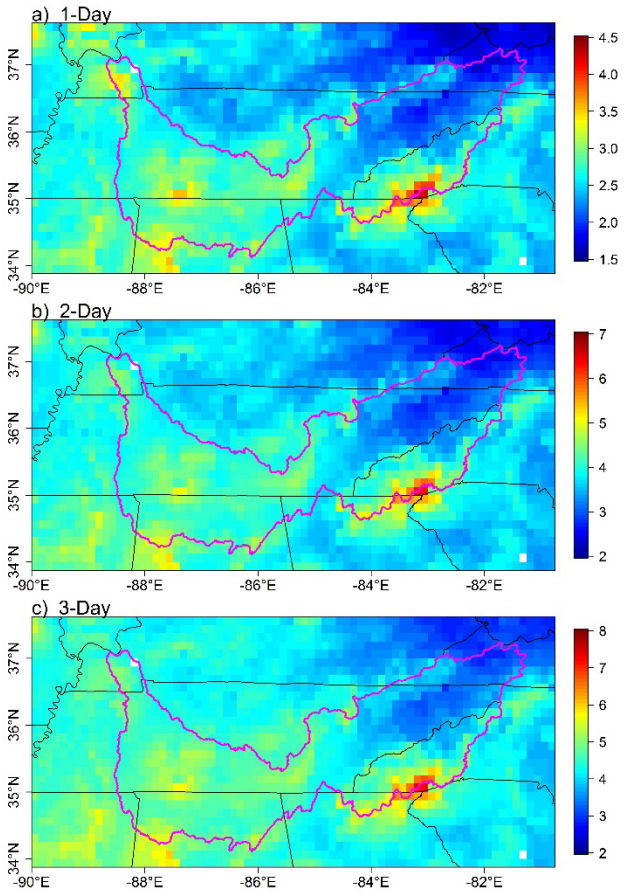


Figure 5-6 Ensemble Average Annual Maximum Precipitation (inches) for a) One-Day, b) Two-Day, and c) Three-Day Precipitation Events. Note the Different Scale for Each Duration.

We present the spatial distributions of the GEV parameters (μ, σ, and ξ) developed using one-day annual maximum precipitation data from the combined ensemble (Figure 5-7). Overeem et al. (2009) assume that the scale (σ) and shape (ξ) parameters of the GEV distribution do not vary spatially across the Netherlands, a region with relatively homogeneous topography. This assumption is relatively uncommon, and very clearly not justified across the TRVW (Figure 5-7). For example, the scale parameter (σ; middle plot in Figure 5-7) is characterized by relatively low values north of the TRVW (~0.22) and relatively high values in the northwest region of the domain (~0.28). The shape parameter (ξ; bottom plot in Figure 5-7) varies in magnitude and sign across the region of interest. Large negative values characterize the western lobe of the watershed, the far eastern part of the watershed, and large areas of North and South Carolina. Clearly, assuming that any one of the GEV parameters is constant across space in this region of complex topography is unjustified (e.g., Bracken et al. 2016). In addition to spatial variability, these parameters vary by precipitation duration (e.g., two- and three-day totals; not shown).

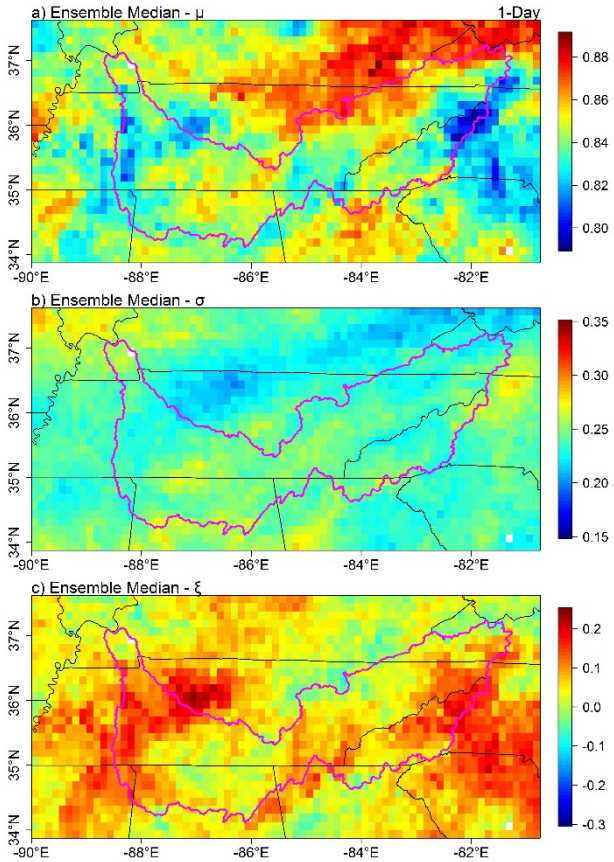


Figure 5-7 Spatial Distribution of the GEV Parameters, a) Location (μ), b) Scale (σ), and c) Shape (ξ), Valid for One-Day Precipitation Maxima Estimated Using L-Moments

Rather than presenting results for each duration and return period, we present results for the one-day duration at the 100-, 1,000-, and 10,000-year return periods. Figure 5-8 shows results of the gridded L-moments analysis for one-day precipitation totals at the 100-year and 1,000-year return periods. At both return periods, the combined ensemble captures enhanced precipitation along the Appalachians, near the border of North Carolina, South Carolina, and Georgia, with lower magnitudes present north of this region, in far eastern Tennessee. However, at these two return periods, the combined ensemble produces precipitation totals that are consistently lower than estimates from NA14 across the TRVW (bottom two plots in Figure 5-8). The largest difference between the combined ensemble and NA14 at the 100-year (1,000-year) return period is approximately -4 inches (-6 inches). One potential explanation for the large negative differences in precipitation totals at these return periods between the two datasets could be the differences in gauges included in both datasets, along with period of record. NA14 precipitation estimates are based on point observations from a longer period of time; some point observations used in the NA14 volumes extend back to the late 1800s. Conversely, the combined ensemble is based on point observations between 1980 and 2012. Another potential explanation for the large differences is resolution. The combined ensemble represents a gridded precipitation data product. Gridded data products are known to consistently underestimate precipitation extremes (e.g., Hofstra et al. 2010; Mladjic et al. 2011; King et al. 2013). Clearly, this common issue is also a concern with the combined ensemble.

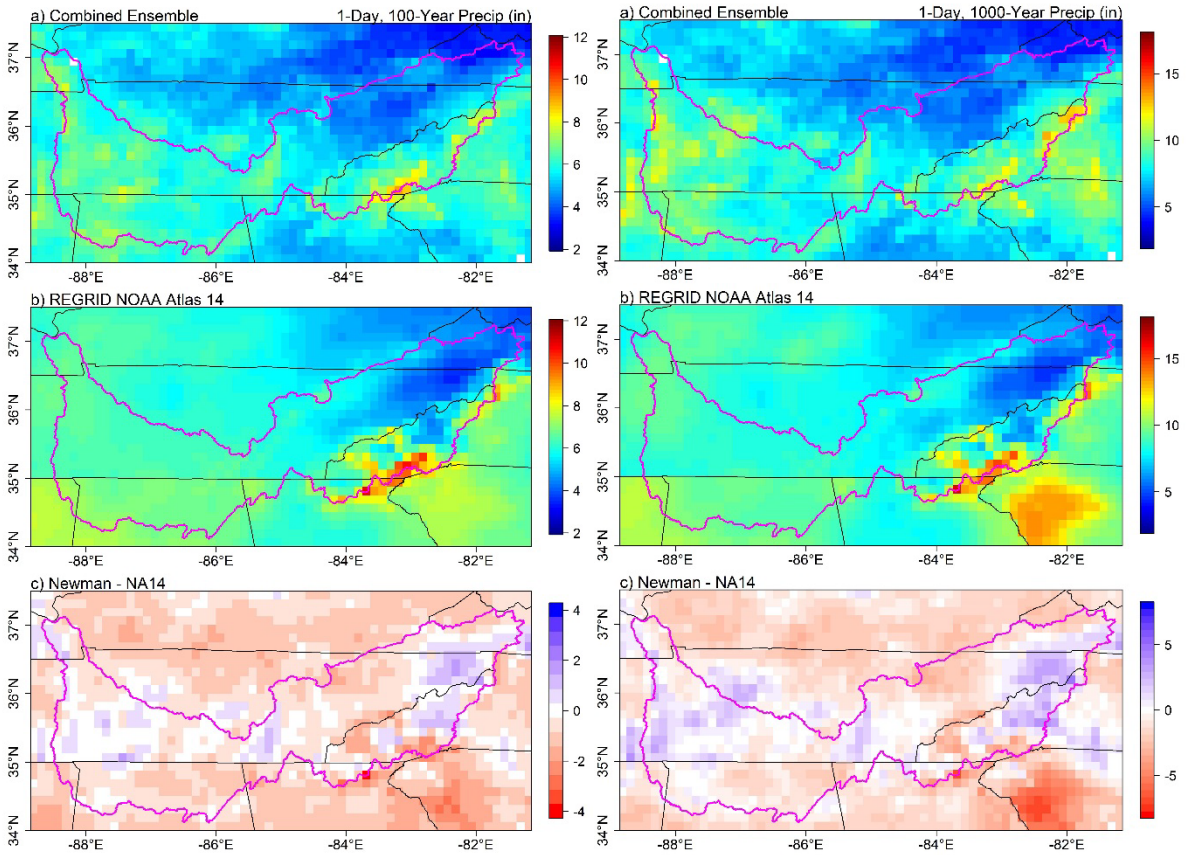


Figure 5-8 One-Day, 100-Year (left) and 1,000-Year (right) Precipitation Totals (inches) from a) the Combined Ensemble (L-Moments) and b) NOAA Atlas 14. c) The Difference Between the Two Datasets, Calculated as the Combined Ensemble Minus NOAA Atlas 14. The Magenta Line Indicates the Extent of the TRVW.

Figure 5-9 shows the spatial correlation between the NA14 field and each Newman ensemble member (box and whisker values), as well as the combined ensemble (red squares) for one-day precipitation totals at each return period available in NA14. This figure indicates that members of the Newman ensemble can reproduce the spatial structure of one-day precipitation return levels captured in NA14. The highest spatial correlations exist between the two datasets at the five-year return period and decrease as return period increases. Spatial correlations between the combined ensemble and NA14 indicate that spatial correlations from the combined ensemble exceed those produced from the individual ensemble members at each return period. Similar figures for the other durations (two- and three-day; not shown) show very similar results. Results in this figure also suggest that combining the ensemble members allows for a better spatial representation of precipitation events across the domain of interest.

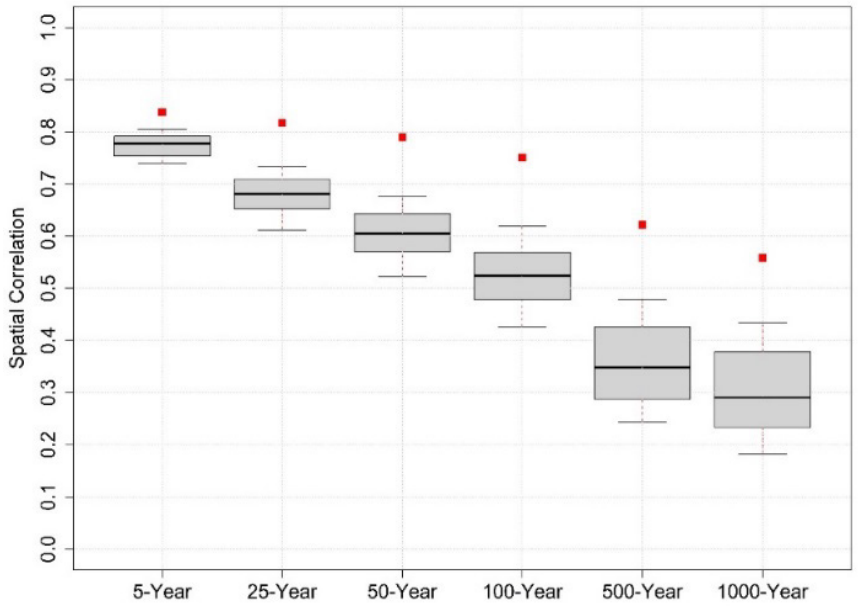


Figure 5-9 Distribution of Spatial Correlation Values Between Individual Ensemble Members (N=100) and NA14 for the One-Day Duration Across the Entire Domain. Median Represented by Thick Black Lines. Horizontal Bounding Boxes Represent the 25th and 75th Percentiles. Dashed Whiskers Extend to the 10th and 90th Percentile Values. The Red Squares Represent the Spatial Correlation Value Between the Combined Ensemble (all members) and NA14.

NA14 produce precipitation-frequency estimates only to the 1,000-year return period because the authors assume the uncertainty levels beyond this return period are insurmountable. While this return period may seem extreme for some design engineering projects (e.g., culverts), this return period can also be viewed as 'frequent' for others. Engineers and risk-management teams working on high-hazard structures, such as dams and nuclear power plants, commonly need probabilistic precipitation information beyond the 1,000-year return period. We use data from the combined ensemble to estimate one-, two-, and three-day precipitation totals at the 10,000-year return period (Figure 5-10). This return period is not included in NA14, or many scientific publications, because of the expected magnitude of uncertainty at this low of an AEP. However, we include this return period in our analysis because some risk assessments applied to various design structures, including dams and nuclear power plants, require this information. Results shown in the top panel of Figure 5-10 suggest that the spatial distribution of precipitation totals at the one-day, 10,000-year return period is very similar to that of more frequent events (lower return periods), although the precipitation magnitudes are clearly greater. As with the one-day, 100- and 1,000-year totals (Figure 5-8), there are two general regions of extreme precipitation magnitudes, one in the western part of the watershed and one in the eastern part of the watershed. At the 10,000-year return period, one-day precipitation totals in these regions exceed 20 inches. The spatial patterns of precipitation for the two- and three-day 10,000-year return periods (middle and bottom panels of Figure 5-10, respectively) very closely resemble that of the one-day event.

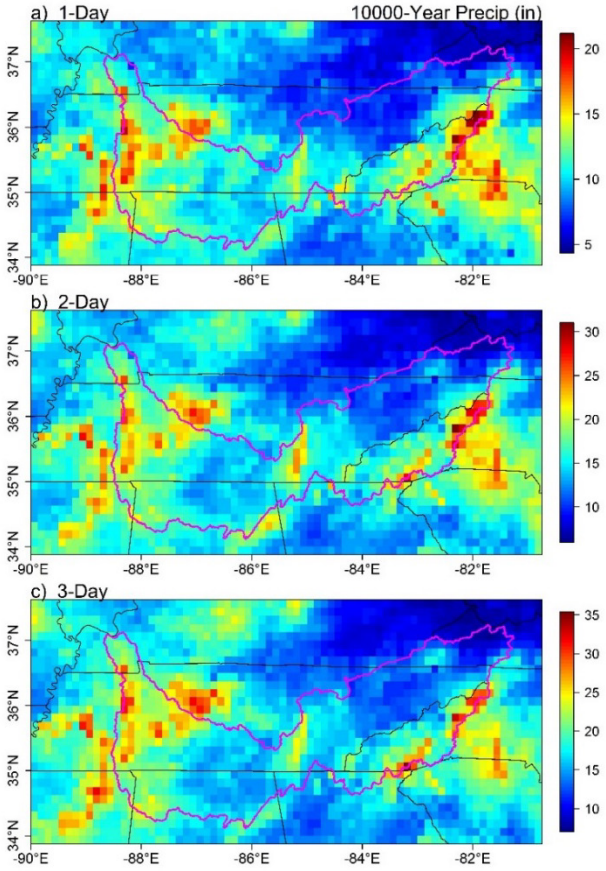


Figure 5-10 Combined Ensemble Estimates of the a) One-day, b) Two-day, and c) Three-day Precipitation Totals (inches) at the 10,000-Year Return Period Based on an L-Moments Approach. Note the Different Colorscale for Each Duration.

5.3.2.1 Gridded L-Moments Uncertainty

We estimate epistemic uncertainty from the combined ensemble using a similar approach as that presented in 4.3.3.1. However, by analyzing precipitation records from a 100-member ensemble, we are already accounting for aleatory variability in the historical precipitation record (Newman et al. 2015). The approach used for the gridded precipitation dataset involves applying a resampling routine to remove 10% of the annual maximum precipitation values (from the 3,300 records) and re-computing the L-moments and RGCs; this is done 100 times. More details on the resampling routine can be found in section 4.3.3.1. Figure 5-11 shows the distribution of the three parameters used to describe to GEV distribution, valid across the TRVW. Distributions of the location (μ) and scale (σ) parameter appear almost normal, while the normality of the shape (ξ) parameter is questionable. The shape parameter distribution includes positive and negative values, indicating that distributions in the watershed are both upper- and lower-bounded. However, this figure combines parameter values across the entire watershed. Analyzing the parameter values spatially would provide additional information on the existence of any upper limits.

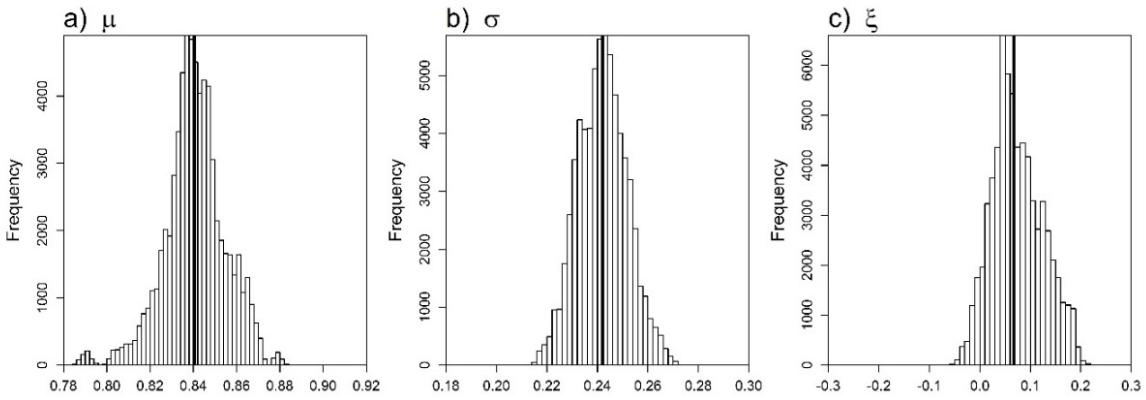


Figure 5-11 Distribution of the Three Parameters used to Describe the GEV Distribution at Grid Cells (from all Ensemble Members) Located in the TRVW. Distributions are Based on 100 Monte Carlo Simulations on Each Ensemble Member. The Vertical Black Line Indicates the Distribution Median.

Figure 5-12 shows the spatial estimates of confidence limits from the L-moments analysis. Areas of large uncertainty follow regions of high precipitation values (e.g., Figure 5-8 and Figure 5-10). The greatest range in precipitation estimates is in the eastern portion of the watershed, along the Blue Ridge Mountains. The western portion of the watershed is also characterized by a large range in precipitation estimates relative to the remainder of the domain, though small in absolute terms. Uncertainty estimates in precipitation totals at the 1,000-year return period range from near zero to one inch. At the 10,000-year return period, the uncertainty estimates in precipitation totals range from near zero to two inches. This incredibly narrow range of uncertainty in precipitation estimates at the 1,000- and 10,000-year return periods is likely the result of reduced variability in the individual time series of annual maximum precipitation among the combined ensemble members. Removing all observations from a subset of ensemble members does not dramatically modify the annual maximum time series, resulting in underestimated estimates of uncertainty.

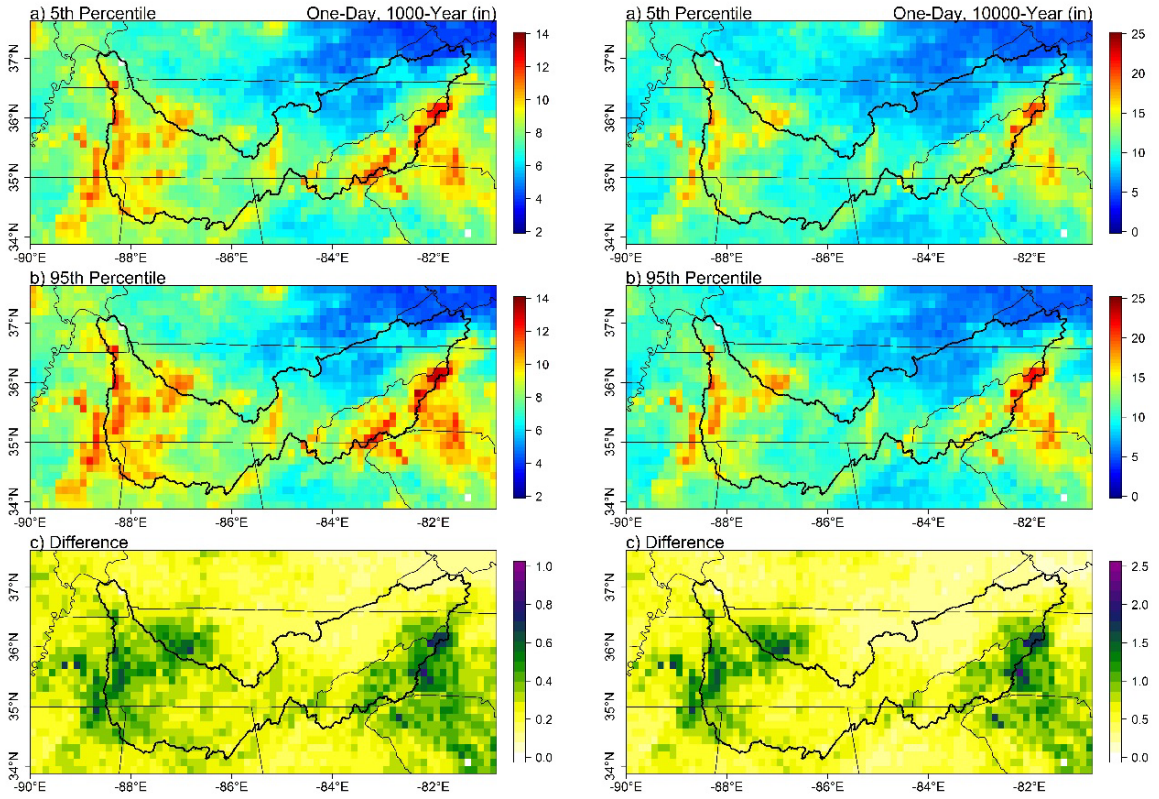


Figure 5-12 Gridded Combined Ensemble Uncertainty Estimates (inches) for the One-day (left) 1,000-year Return Level and 10,000-year Return Level (right). The Top Panel Shows the 5th Percentile, the Middle Panel Shows the 95th Percentile, and the Bottom Panel Shows the Difference, Computed as 95th Minus 5th. The TRVW is Indicated by the Thick Black Line. Note the Different Colorscales Among Plots.

As with the observation-based L-moments analysis, we demonstrate results of the gridded precipitation-frequency analysis at the two grid cells located closest to the sample sites used in section 4.3.3. These results are valid for the entire grid cell and are consequently not directly comparable to point estimates. However, these estimates can be compared to other estimates valid over a similar spatial domain. Figure 5-13 shows the distributions of GEV parameters for the two grid cells closest to these NPPs. The shape parameters are all positive, indicating that the RGCs for these two grid cells are lower bounded.

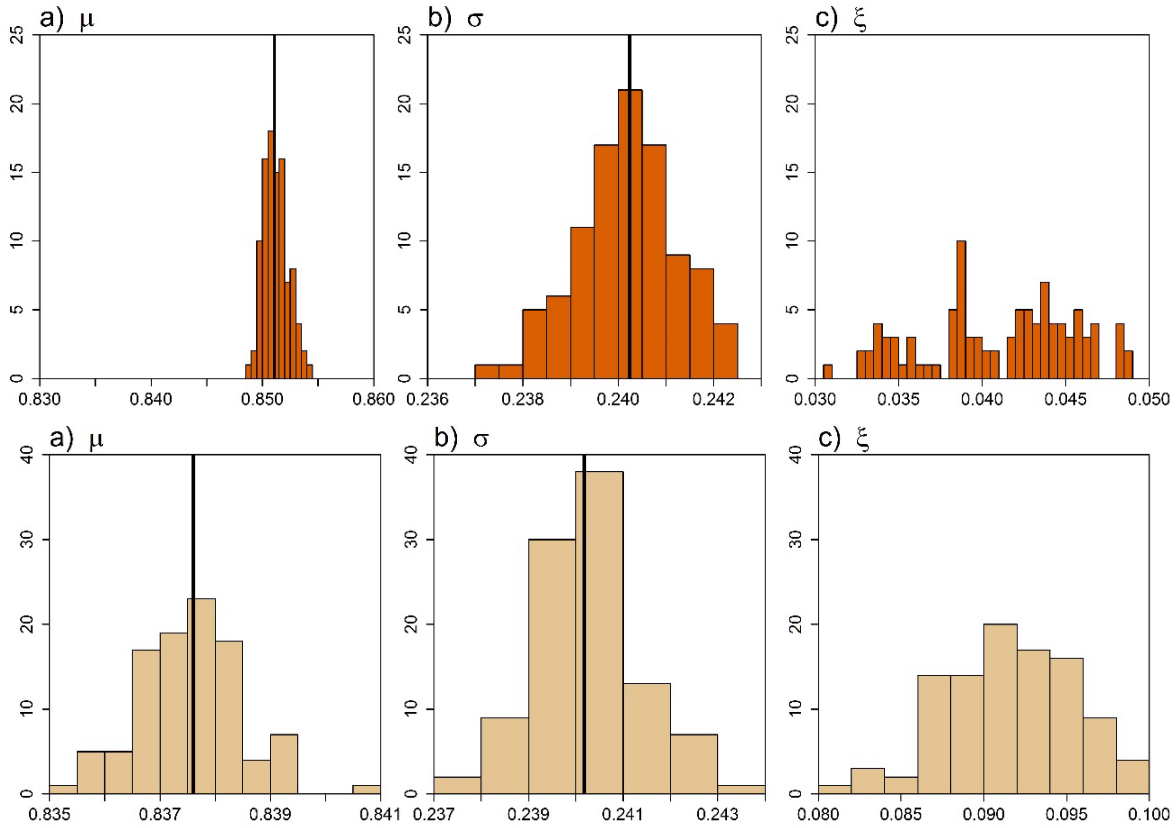


Figure 5-13 Distribution of Three GEV Parameters Based on 100 Monte Carlo Simulations for the Grid Cell Closest to the Example Site in Group 10 (top) and Group 7 (bottom). Colors Correspond to Groups from Figure 4-9. The Vertical Black Line Indicates the Sample Median.

Figure 5-14 shows the resulting precipitation-frequency results for the two grid cells closest to sample sites discussed in section 4.3. As with the observed L-moments analysis, precipitation totals for the grid cell in group 10 exceed those from the grid cell in group 7 at various return periods, including the 10,000-year return period. Uncertainty estimates between the two curves are relatively narrow and similar in magnitude. For example, at the 10,000-year return period, the range of uncertainty from the grid cell in group 10 is 0.56 inches and from the grid cell in group 7 is 0.47 inches. These particularly narrow ranges of uncertainty are likely the result of the driving data; the annual maximum time series of precipitation from individual ensemble members are likely similar in magnitude, perhaps too similar, such that removing records from individual members does not dramatically influence the remaining annual maximum time series. The narrow range of uncertainty also emphasizes the limitation of the drop-10% bootstrapping method commonly used by Reclamation for L-moments uncertainty estimates. Alternative approaches, such as Bayesian inference, may provide more realistic estimates.

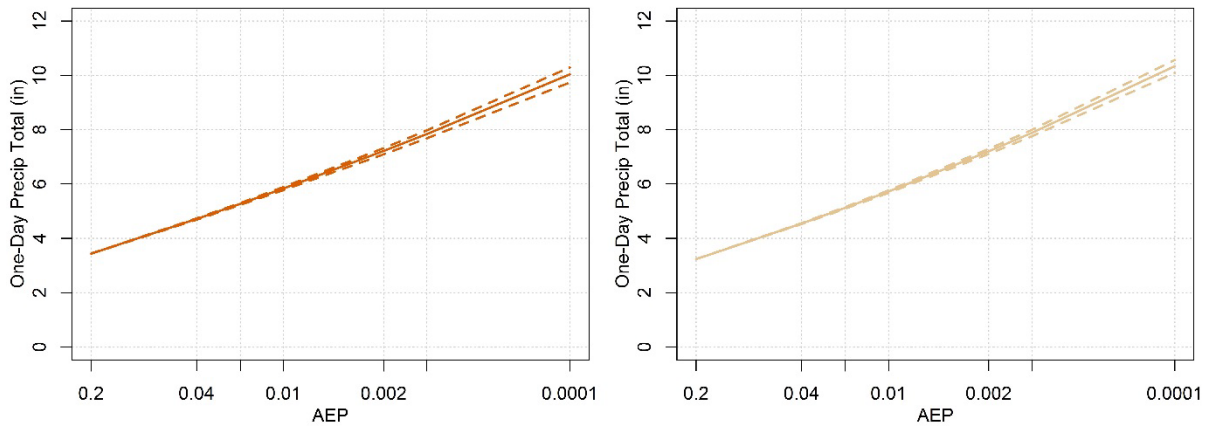


Figure 5-14 Combined Ensemble Precipitation-Frequency Results for the Grid Cells Closest to the Two Example Sites in Figure 4-18. Colors Correspond to Groups from Figure 4-9.

5.3.3 Bayesian Analysis

Consistent with the Bayesian analysis in section 4.2.4, the rstan package in the R statistical language is used to build posterior distributions of GEV parameters and, subsequently, regional growth curves and precipitation-frequency estimates at numerous return periods. One fundamental difference is the assumption that each grid cell is in itself a homogeneous region. For each grid cell, there are 33 years of annual maxima, and there are 100 ensemble members; therefore, there are 3,300 annual maxima at each grid cell. For each grid cell separately, the 3,300 annual maxima are first scaled by the "at-site mean", the mean of the 3,300 values, then are used to drive the Bayesian inference sampling routine. The prior distributions vary for each grid cell, as they are strictly informative using the MLEs to center the priors, and twice the MLE standard errors as the standard deviations of the priors. After numerous interactive model runs and sensitivity testing, we determined that the appropriate number of burn-in iterations is 5,000. Therefore, the burn-in period is defined as 5,000 iterations, with a subsequent 5,000 samples after burn-in. To be clear, at each grid cell we have 5,000 estimates of the GEV parameters. For each grid cell, and for each of the 5,000 GEV parameter sets, a regional growth curve is computed, and the at-site mean is used to scale the RGC to precipitation-frequency space. What results is a unique P-F curve for each grid cell.

We begin by analyzing the spatial distributions of the GEV parameters. To quantify the central tendency of the posterior distributions of the GEV parameters, we show the median. As a quantification of parametric uncertainty, we also show the 90% credible interval range (i.e., 95th percentile minus 5th percentile). The concept of a credible interval is akin to the traditional confidence interval, although the difference is the credible interval is empirical (calculated from the 5,000 values in the posterior distributions), while the confidence interval is theoretical (estimated using assumptions of standard error and normality). Figure 5-15 presents the spatial distributions of the GEV parameters (μ, σ, and ξ) developed using one-day annual maximum precipitation data from the combined ensemble. The left panels show the medians of the GEV parameters; the right panels show the 90% credible intervals of the GEV parameters.

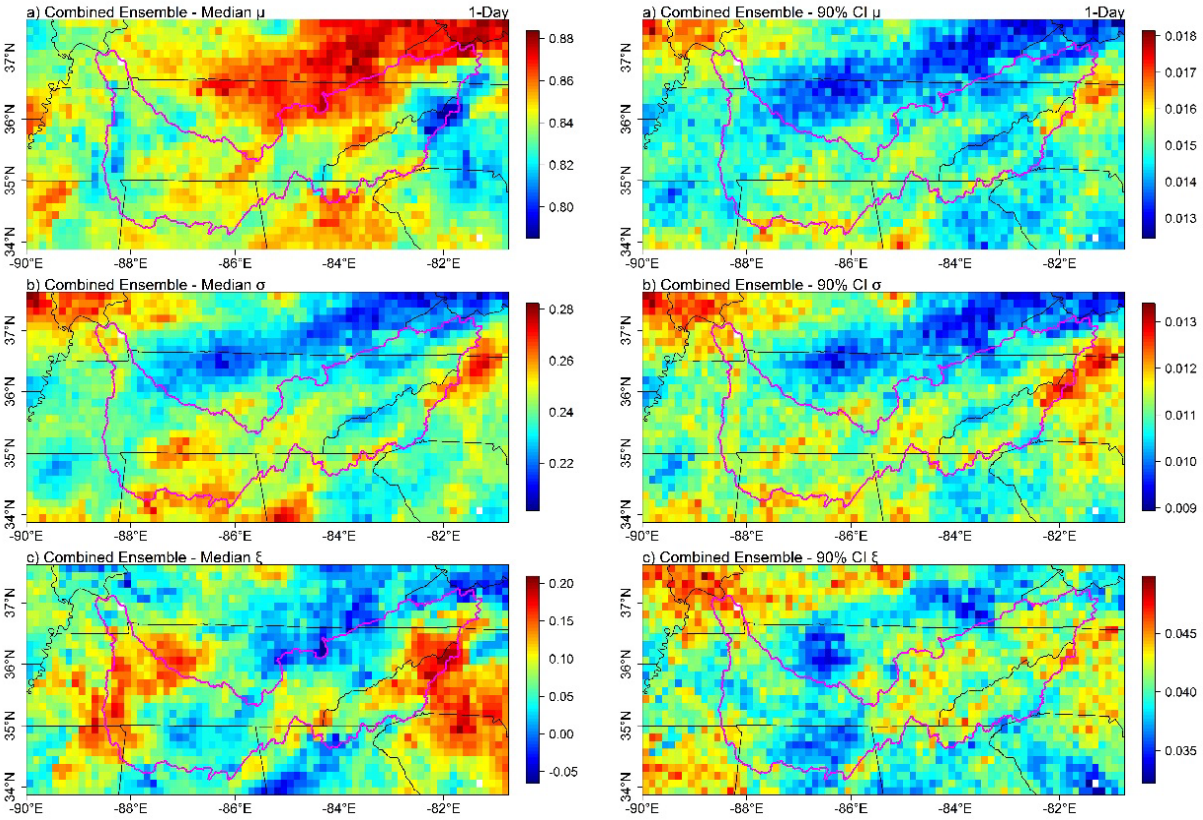


Figure 5-15 (Left) Median Spatial Distribution of the GEV Parameters Location (a), Scale (b), and Shape (c). (Right) 90% Credible Interval Spatial Distribution of the GEV Parameters Location (a), Scale (b), and Shape (c). Valid for One-day Precipitation Maxima Estimated from L-Moments.

Having 5,000 GEV parameter estimates at each grid cell enables the estimation of P-F relationships and the associated parametric uncertainty. For a best estimate, 5,000 P-F relationships are computed, one for each of the parameter sets, and a median is taken. Uncertainty at the 90% level is quantified by the 5th and 95th percentiles of the 5,000 P-F estimates. Spatial estimates of P-F magnitudes are useful to provide information where there may be none available. For example, a proposed facility may be located at a site that does not have observed data and, therefore, a site-specific P-F estimate.

We produced return period maps for one-, two-, and three-day precipitation totals, for common return periods. Rather than presenting results for each duration and return period, we focus on one-day duration at the 100- and 1,000-year return periods, commonly used metrics for hazards planning. Figure 5-16 shows the Bayesian median estimates of 100- and 1,000-year return periods. The Bayesian medians adequately capture the spatial pattern of extreme precipitation in the TVRW, with enhanced precipitation along the Appalachians, and lower magnitudes in northeastern Tennessee. There is good agreement in the western part of the watershed, albeit the Bayesian medians are a bit greater than the NA14 estimates. However, the southwestern and northeastern regions of the watershed show the Bayesian medians to be less than the NA14 estimates. Overall, there does not seem to be significant bias in the Bayesian medians with respect to the NA14 estimates.

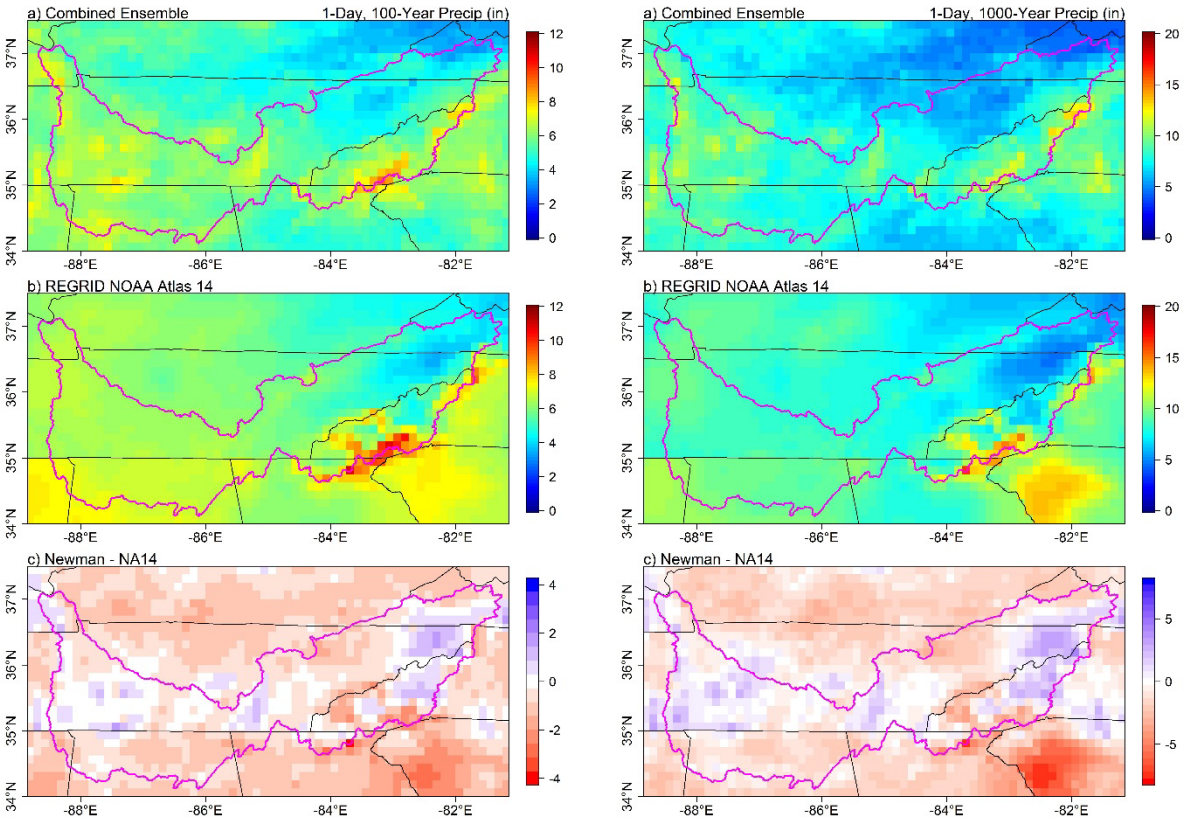


Figure 5-16 One-day, 100-year (left) and One-day, 1,000-year (right) Precipitation Totals (inches) from the Combined Bayesian Ensemble (a) and NA14 (b). The Difference Between the Two Datasets (inches), Calculated as the Combined Ensemble minus NA14 is Shown in (c). The Magenta Line Indicates the Extent of the TRVW.

The largest differences between NA14 and the Bayesian median estimates at the 1,000-year return period (the bottom right plot of Figure 5-16) are concentrated outside the TRVW, in northwestern South Carolina. The differences in this region reach approximately -8 inches, where NA14 estimates exceed those of the Bayesian estimates. The reason for these differences can be seen in the right middle panel of Figure 5-16, where NA14 estimates of the 1,000-year return period in South Carolina exceed surrounding estimates by approximately five inches. The state of South Carolina is included in NA14 Volume 2, a volume of NA14 where the probability distribution function is allowed to vary by homogeneous regions. Gauges in this region of South Carolina belong to homogeneous region 12 (see Figure 4-2), which is described by the generalized logistic (GLO) distribution (pg. 30 of Bonnin et al. 2006), a distribution that is rarely employed in practical applications (Kysely and Picek 2007). The frequency distribution used to model precipitation extremes in surrounding homogeneous regions are not GLO; rather, surrounding distributions are the generalized extreme value (GEV) and the generalized normal (GNO). The GEV and GNO distributions are closely related, and have the same weight in the upper tails, while the GLO better captures long heavy tails and outlying extreme observations (Kysely and Picek 2007). Because NA14 Volume 2 allows the underlying frequency distributions to spatially vary, there are clear discontinuities in precipitation return levels over very small distances, including those in northwestern South Carolina. Additionally, these discontinuities have potentially presented themselves as problems with the Bayesian estimates,

which may be misleading. These large differences in northwestern South Carolina are also present in the L-moments analysis (see Figure 5-8). The consistencies between the two methods (L-moments and Bayesian) suggest that the precipitation magnitudes in this region of NA14 Volume 2 could be erroneous.

We also provide an estimate of the precipitation totals for the 10,000-year return period. Although the rarity of this event puts an exceptional amount of uncertainty on the event, some design and risk analyses require estimates of these rare events. Figure 5-19 shows the Bayesian median estimate of 10,000-year precipitation totals for one-, two-, and three-day durations, all of which have similar spatial patterns. Consistent with the 100- and 1,000-year totals (Figure 5-16), there are two general regions where precipitation magnitudes are significantly greater than the rest of the study region, one in the western part of the watershed and one in the eastern part of the watershed. At the 10,000-year return period, one-, two-, and three-day precipitation totals in these regions reach beyond 20, 30, and 35 inches, respectively.

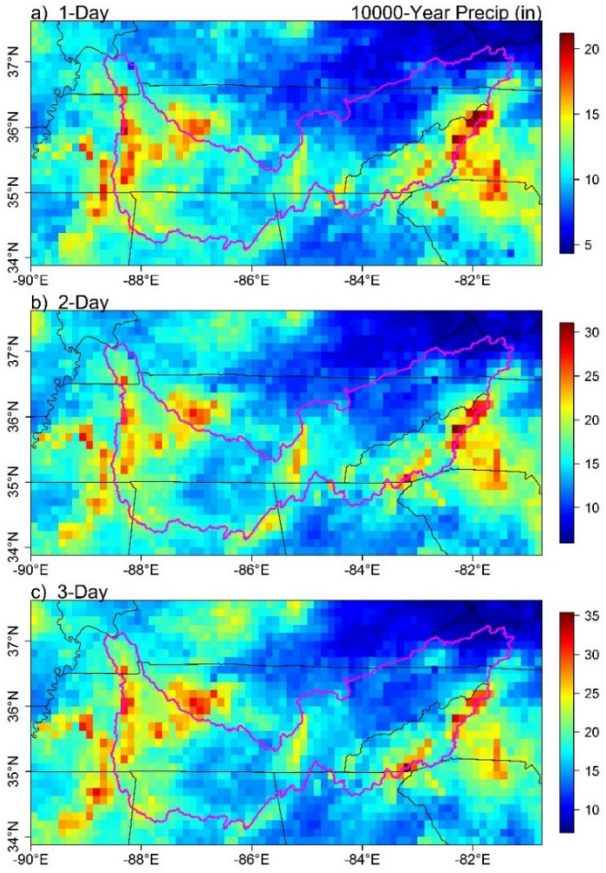


Figure 5-17 Bayesian Median Estimates of the One-day (a), Two-day (b), and Three-day (c) Precipitation Totals (inches) at the 10,000-year Return Period

A benefit to the Bayesian methodology is the built-in quantification of uncertainty. As there are 5,000 GEV parameter estimates at each grid cell, it is straightforward to calculate and quantify the uncertainty in the precipitation estimates. Note, however, that this method only accounts for uncertainty in the parameters themselves, and consequently the RGCs. This is not the total uncertainty, which can come from numerous varied sources. However, by analyzing precipitation from a 100-member ensemble, we are also accounting for uncertainty in the

historical precipitation record (Newman et al. 2015). Figure 5-18 shows the posterior distributions of the GEV parameters for only grid cells within the TRVW. The shape parameter distribution includes positive and negative values, indicating that distributions in the watershed are both upper- and lower-bounded. However, the bulk of the shape parameter distribution contains positive values, which indicates that, for the most part, there is no upper bound on the process. The physical significance of an upper-bounded distribution is merely that, given the data, the best estimates indicate that the rate of increase of precipitation depth with increasing return period is decreasing. That is, the limit of precipitation depth as return period goes to infinity is a finite value. While some people think the precipitation process is unbounded, an in-depth discussion of this is beyond the scope of this project. Additionally, and unfortunately, there will never be concrete evidence one way or another. Identifying the regions that have upper bounds can be done using the information in Figure 5-11.

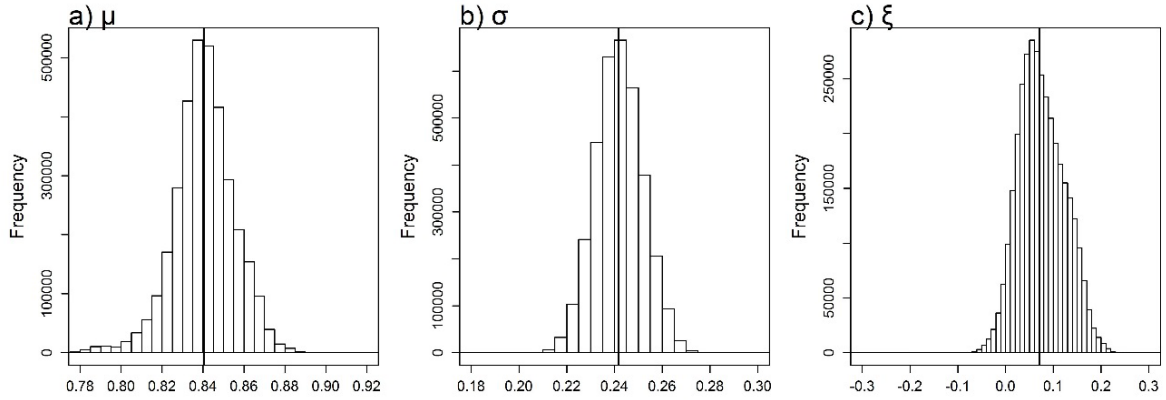


Figure 5-18 Bayesian Posterior Distributions of the GEV Parameters for Grid Cells Within the TRVW. The Vertical Black Line Indicates the Distribution Median.

Credible intervals at the 90% level are shown for each grid cell in Figure 5-19. We show the 5th and 95th percentiles, and the difference between them, for the 1,000- and 10,000-year return periods. The difference maps give a visual representation of how the uncertainty varies spatially. As expected, with greater precipitation totals comes greater uncertainty, with over five inches of uncertainty for the 10,000-year return period in the eastern part of the watershed. The range of uncertainty for both return periods is significantly smaller than what would be expected from a traditional regional frequency analysis. This reduction in uncertainty can be attributed to the fact that 3,300 values were used in the analysis: as sample size increases, the variance of the process converges on its nominal value. This can be directly seen by a comparison of Figure 4-19 and 5-21, which show the P-F curves from station- and ensemble-based analyses, respectively.

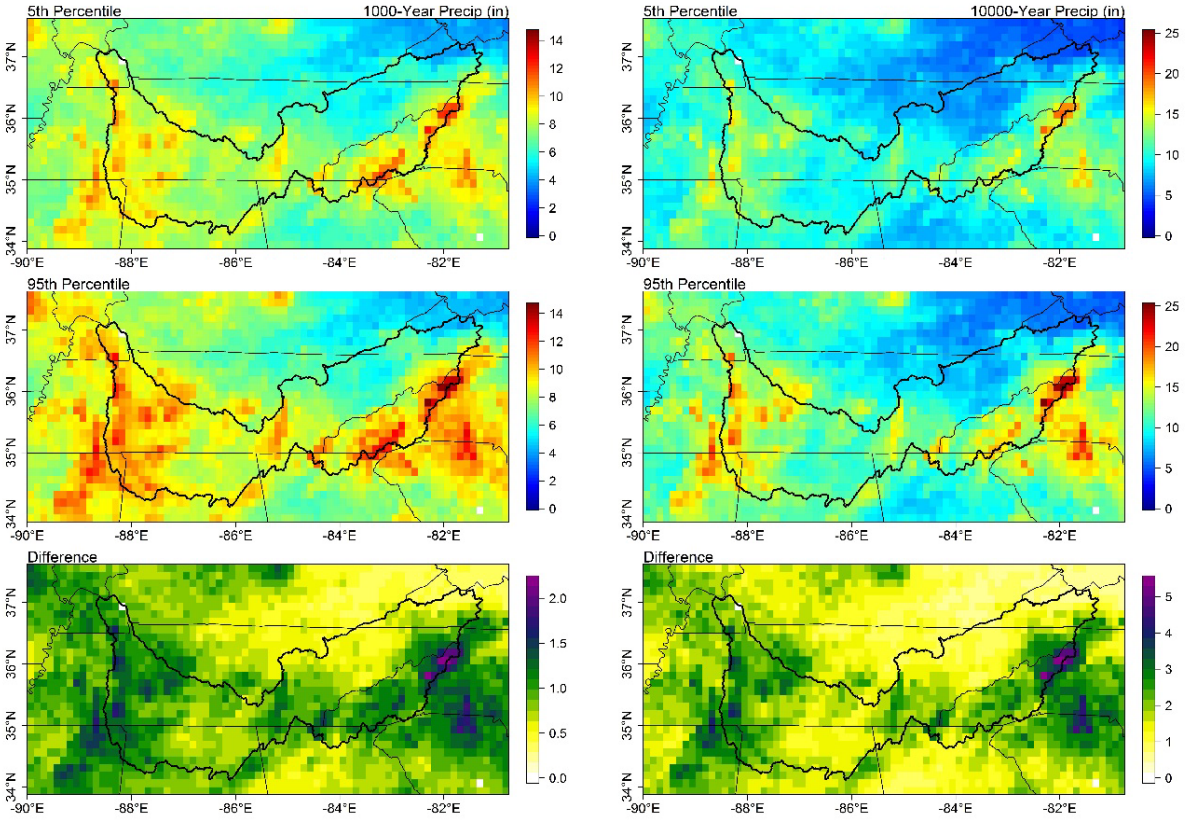


Figure 5-19 Bayesian Uncertainty Estimates (inches) for the One-Day, 1,000-Year Return Level (left) and One-Day, 10,000-year Return Level (right). The Top Panel Shows the 5th Percentile, the Middle Panel Shows the 95th Percentile, and the Bottom Panel Shows the Difference, Cmputed as 95th Minus 5th. The TRVW is Indicated by the Thick Black Line.

To illustrate a practical application of the gridded P-F estimates, such as for facilities planning, we isolate the two grid cells closest to the locations highlighted in section 4.3, a GHCN site located in group 10 and a GHCN site located in group 7. Figure 5-20 shows the posterior distributions of the GEV parameters for these grid cells. It can be inferred from the shape parameters that the P-F curve for the grid cell nearest the group 10 site is less skewed because the distribution spans zero. The grid cell nearest the group 7 site has consistently positive shape values, which indicates there is no doubt that the distribution is lower bounded and therefore has no upper limit. The P-F curves produced from these posterior distributions are shown in Figure 5-21, which confirms that the curve for the site located in group 7 is much more skewed than for the site located in group 10. As the annual exceedance probability decreases and the return period increases, the curvature of the P-F relationship for group 10 site becomes more pronounced. At the same time, there is little to no curvature in the P-F curve for the group 7 site, which is consistent with a shape parameter that is nearly zero.

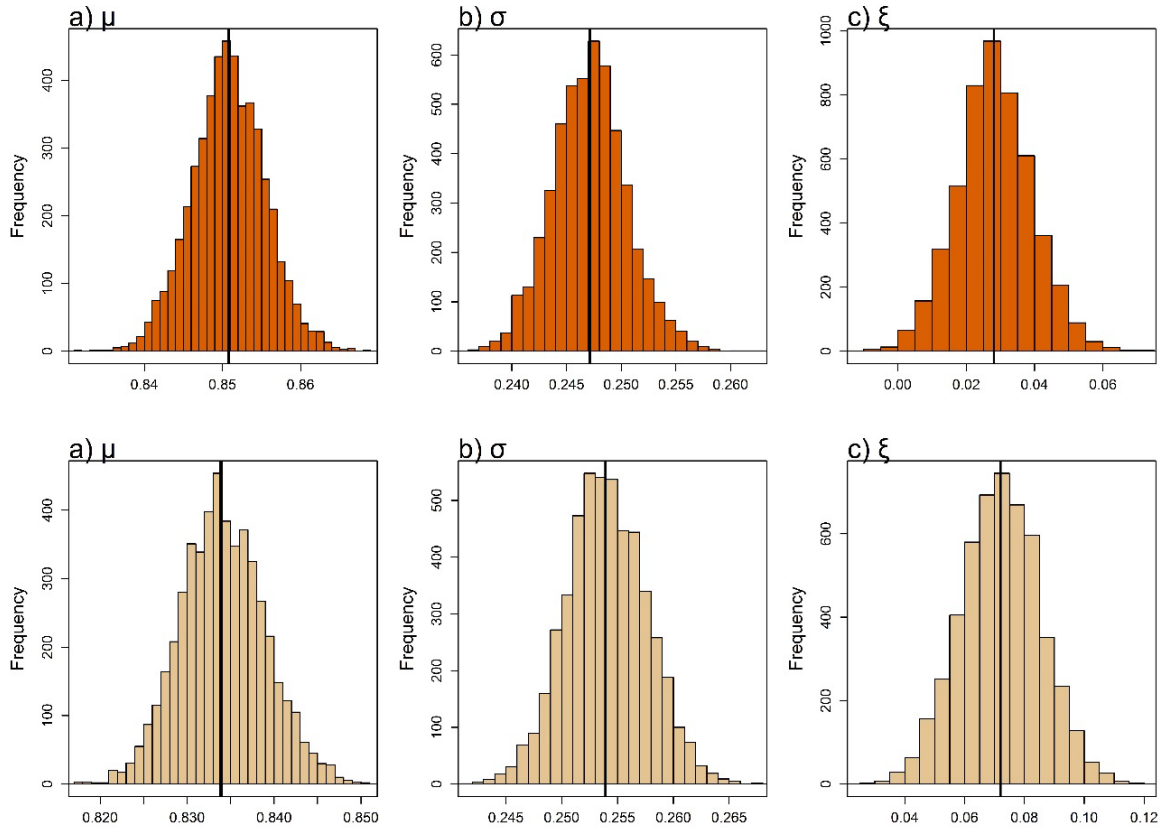


Figure 5-20 Distribution of GEV Parameters for the Grid Cells Closest to the GHCN Site Located in Group 10 (top) and in Group 7 (bottom). Colors Correspond to Groups from Figure 4-9. The Vertical Black Line Indicates the Sample Median.

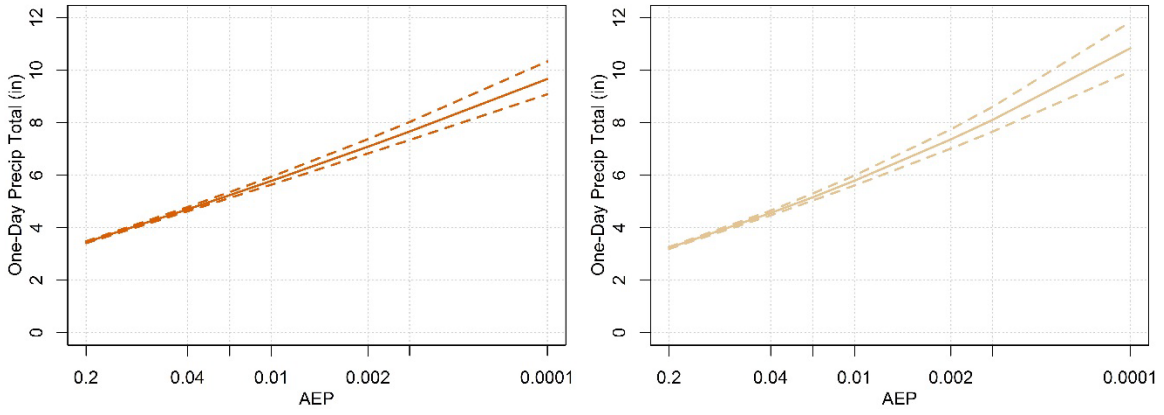


Figure 5-21 Bayesian Combined Ensemble Precipitation-Frequency Results for the Two Locations in Figure 4-18. Colors Correspond to Groups from Figure 4-9.

5.4 Comparison

We now present the similarities and differences between the L-moments results and the Bayesian inference results. Figure 5-22 shows the one-day 100- and 1,000-year precipitation totals from both methodologies, as well as the differences between them. The spatial patterns of the P-F at 100- and 1,000-year return periods are very similar between the L-moments and Bayesian median estimates, which is understandable considering the only difference is the parameter estimation methodology. Both methods produce relatively large precipitation magnitudes along the far eastern edge of the TRVW, in western North Carolina. Reduced precipitation magnitudes are present in northeastern Tennessee, in the rain shadow region. Both methods also produce relatively large precipitation totals across the eastern portion of the TRVW.

Overall, there is little difference between the results from L-moments and Bayesian inference, which is indicated by the prominence of very light and white pixels in the bottom two panels. The most notable differences between the methods are positive (e.g., blue) and are in the western part of the study region, outside the TRVW. The blue pixels indicate that the L-moments estimates are greater than the Bayesian estimates in this location. The 100-year (1,000-year) L-moments precipitation estimates exceed the Bayesian estimates by approximately 0.25 inches (1.5 inches) in this region.

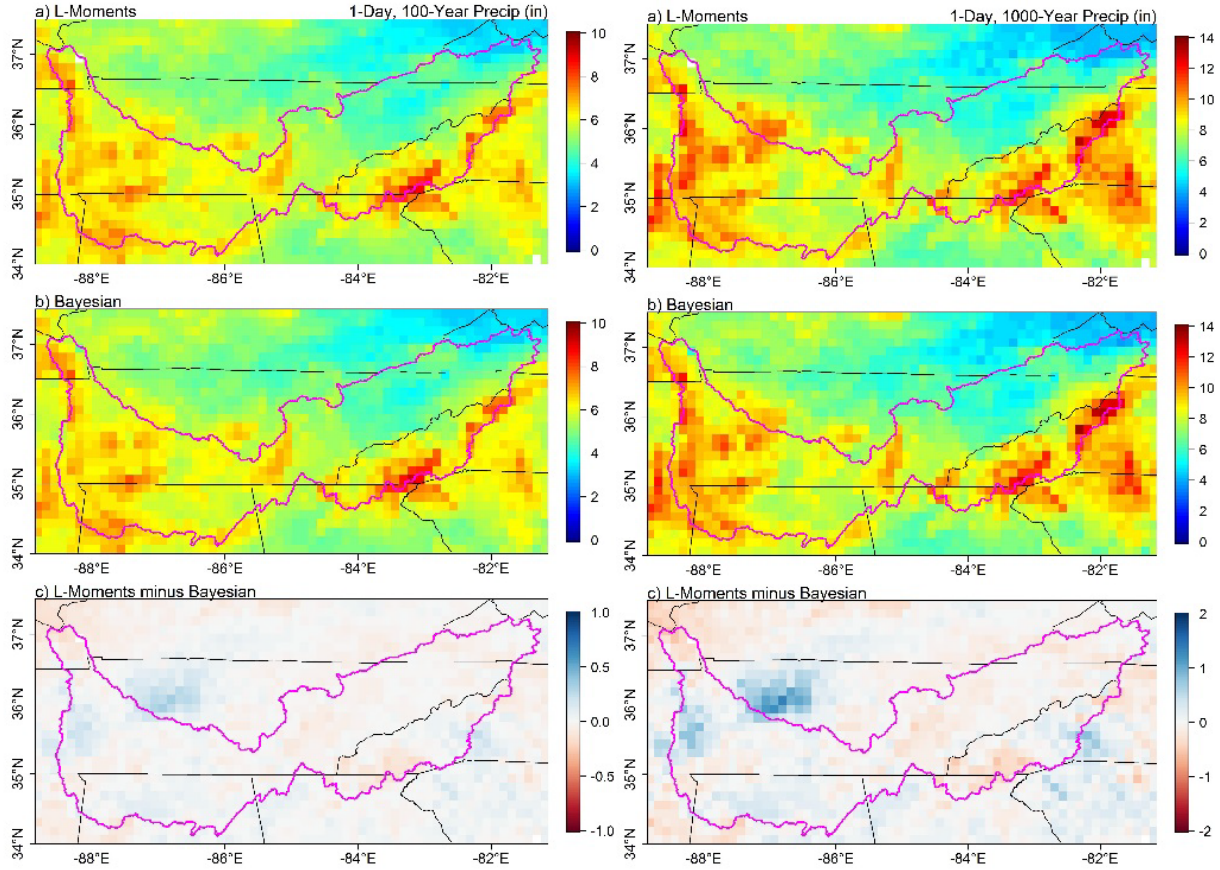


Figure 5-22 Difference Between L-Moments and Bayesian Median Estimates of One-Day Precipitation Magnitudes (inches) for 100-Year (left) and 1,000-Year (right) Return Periods

Consistent with the local-scale analyses provided in sections 5.3.2 and 5.3.3, Figure 5-23 shows precipitation-frequency results from L-moments and Bayesian methods, for two different grid cells representative of groups 10 and 7. The L-moments analyses exhibit much smaller ranges in uncertainty than the Bayesian analyses, though are contained within the Bayesian estimates, which is a finding consistent with results presented in section 5.3.2. Such reductions in uncertainty may be a product of the drop-10% bootstrap resampling method used to quantify uncertainty in the L-moments estimates. Because there are 100 ensemble members that are (assumed) independent, though likely are correlated, the drop-10% resampling technique will not provide strikingly different results for each bootstrapping sample. Conversely, the Bayesian analysis provides greater uncertainty at the rare events, though this uncertainty is likely still under-represented due to the ensemble members being very similar in magnitude. The decreased uncertainty in the extreme value parameters, as estimated by the Bayesian method, is directly due to the potential correlation between ensemble members. This sort of reduction in uncertainty can be thought of as the "right answer for the wrong reason," in that there are 100 potentially correlated extreme values per grid cell per year, which provides a false sense of confidence via the model parameter estimation methods.

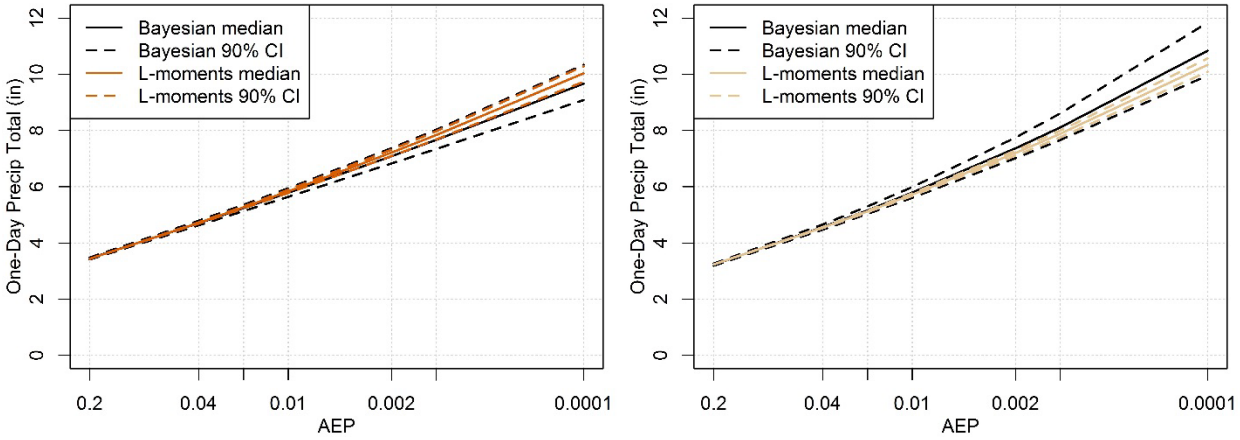


Figure 5-23 One-Day Precipitation-Frequency Curves (inches) for a Sample Grid Cell from Group 10 (left) and One from 7 (right) Based on the Regional L-Moments Approach (Color) and Regional Bayesian Approach (Black)

6 DISCUSSION AND CONCLUSIONS

This report was produced by the Department of Interior's Bureau of Reclamation (Reclamation) to develop guidance for estimating precipitation-frequency relationships in orographic regions, as requested by the Nuclear Regulatory Commission (NRC). This work is relevant to engineering design projects, as many still rely on Probable Maximum Precipitation (PMP) estimates in order to produce Probable Maximum Floods (PMF). Deterministic methods, such as PMP and PMF, are still used by many agencies to evaluate the safety of a proposed or existing facility, though these methods do not provide decision makers with information on the expected frequency of these extreme events. Additionally, no information is provided on extreme precipitation or flooding events with lesser magnitude than the PMP and PMF, which can be just as threatening to a facility. Agencies such as Reclamation and the U.S. Army Corps of Engineers have transitioned to probabilistic precipitation information for risk-based decision making.

Probabilistic precipitation information often takes the form of a precipitation-frequency relationship, which describes the depth of precipitation associated with an "average" return period. NOAA's Hydrometeorological Design Studies Center (HDSC) produces precipitation-frequency estimates across most of the U.S. These estimates are published as individual regional volumes of NOAA Atlas 14 (e.g., Bonnin et al. 2006) or NOAA Atlas 2 (depending on the region of interest), and include precipitation-frequency estimates as rare as the 1,000-year return period, which corresponds to an annual exceedance probability (AEP) of 1/1,000. However, high-hazard facilities frequently require precipitation-frequency estimates beyond the 1,000-year return period. This research provides information useful to license applicants on acceptable methods and data sources for developing and estimating precipitation-frequency analyses in orographic regions for calculating floods at critical AEPs that the applicants need to consider in evaluating siting factors, conducting Probabilistic Risk Assessments (PRAs), and in designing facilities.

The main goals of the research were to investigate the following aspects of extreme precipitation:

1. Critically review orographic extreme storm methodologies used in PMP and magnitudes less than PMP in orographic regions; and
2. Evaluate methodologies for developing precipitation-frequency estimates at AEPs significantly less than those offered in NOAA Atlas 14, in the range of 10^{-4} to 10^{-6} AEP;

To meet the objectives and goals, Reclamation performed the following five major tasks:

1. Review extreme storm precipitation techniques, precipitation-frequency methods and databases in orographic regions;
2. Develop a methodology to estimate precipitation-frequency analyses in regions of complex topography;
3. Demonstrate the precipitation-frequency methodology and provide uncertainties and confidence intervals at the regional and reactor-site scale for a pilot region in the Tennessee River Valley;
4. Transfer technology to the NRC staff via a training session, including data and software scripts; and
5. Complete a final report that conveys research findings.

The five tasks outlined above were accomplished through the use of existing and new technologies available within the scientific community. Task 1 was accomplished by reviewing a variety of techniques used in orographic storm analyses, including methods from the HMRs, private consultants, and alternative methods from the scientific community. We discussed how orographic precipitation computations vary as a function of HMR and through time. We also discussed PMP analyses performed by private companies. These studies often aim to produce site-specific PMP estimates by including datasets and methods that were not included in historical HMRs. The site-specific studies are often based on a subset of historical storm events included in the relevant HMR. Outside of coordinated and private PMP analyses, we discussed major orographic storm advancements within the scientific community, beginning in the late 1800s. The review of orographic storm techniques available within the scientific community demonstrates how limited the HMR methods were, even at the time of development.

In addition to the HMRs, we reviewed previous precipitation-frequency data and methods as part of Task 1. We reviewed federal approaches to precipitation-frequency analyses (e.g., NOAA Atlas 14), as well as approaches used by private consultants and the scientific community. We focused primarily on a discussion of regional L-moments and Bayesian inference. The regional L-moments approach has been used in conjunction with a variety of variables, including precipitation, snow water equivalent, and streamflow. The L-moments approach has also been used outside of point observations, namely a number of gridded datasets. Bayesian inference methods have been used for a wide variety of purposes, including biostatistics and medicine, real estate, paleoflood reconstructions, spatial precipitation extremes, to name a few.

Beyond a review of historical methods, we also explored recent advancements to orographic storm analyses, namely data and methodology improvements. Since the publication of nearly all HMRs, there have been significant improvements to data recording technology, observational networks, and data availability. There have also been major advancements in the type of data available. Currently, there exist a plethora of precipitation datasets from in-situ observing platforms, gridded techniques applied to point observations, and reanalysis products. There have also been significant improvements in analysis techniques. Lower computing costs have made numerical weather model simulations possible for those working outside major forecasting institutions. Improved computational abilities have also resulted in improvements to statistical techniques.

Tasks 2 and 3 included developing and demonstrating a methodology to estimate precipitation-frequency analyses in regions of complex topography. We accomplished these tasks by first combining a known objective clustering method, the SOM algorithm, with two different regional frequency methods, L-moments and Bayesian inference, across the Tennessee River Valley watershed. These frequency methods vary widely in terms of the level of complexity (and consequently effort) and in the estimation of uncertainty. These methods were applied to historical precipitation observations from 14 groups located across the Tennessee River Valley watershed. Both analyses assume the data are best described by a GEV distribution. Results suggest that the SOM algorithm is a useful tool for identifying and grouping similarly behaved point precipitation observations in regions of complex and variable terrain. Furthermore, the frequency results from these analyses indicate that uncertainty estimates from the L-moments analysis are consistently less than the uncertainty estimates from Bayesian inference. These differences are the result of estimating uncertainty differently between the two methods. The Bayesian approach uses an iterative Monte Carlo sampling scheme to estimate uncertainty in distribution parameters, while the approach used to estimate uncertainty with L-moments is based on systematically removing a fixed number of observations, estimating the distribution

parameters with the remaining observations, and repeating the process. Uncertainty estimates can play a major role in the risk-based decision-making process.

In addition to demonstrating these methods on historical point precipitation observations, we also demonstrated the application of these two frequency methods on a gridded precipitation dataset, the 100-member ensemble dataset from Newman et al. (2015). The application of these two regional frequency methods to a gridded precipitation dataset demonstrates the usefulness of these methods to a variety of data formats and structures. With some modifications, the techniques developed in this report can be applied to any gridded or point-based precipitation dataset of interest to the users.

Task 4, transfer technology to NRC staff, was accomplished through a meeting between Reclamation personnel and NRC staff that was held at the NRC headquarters in Rockville, MD on August 1-2, 2017. The event included two days of lectures, with allocated time for questions and discussion. Task 5 was accomplished through this report, which conveys all the research findings from the Reclamation team.

7 REFERENCES

Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, and A. Gruber, 2003: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeo.*, **4**(6), 1147-1167.

AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai, 2011: Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *J. Geophys. Res.*, **116** (D2).

Ahmad, M. I., C. D. Sinclair, and A. Werritty: Log-Logistic Flood Frequency Analysis, *J. Hydrol.*, **98**, 215–224, 1988

Alila, Younes, 1999: A Hierarchical Approach for the Regionalization of Precipitation Annual Maxima in Canada. *J. Geophys. Res. – Atmos.*, **104**, 31645-31655.

Alpert, P., 1989: Mesoscale Indexing of the Distribution of Orographic Precipitation over High Mountains. *J. Clim. and Applied Meteor.*, **25**, 532-545.

Arkin, P., and P. Ardanuy, 1989: Estimating Climatic-Scale Precipitation from Space: A review. *J. Clim.*, **2**(11), 1229-1238.

Banerjee, S., A. E. Gelfand, and B. P. Carlin, 2004: Hierarchical Modeling and Analysis for Spatial Data. Boca Raton, FL, USA: CRC Press.

Bárdossy, A., and G. Pegram, 2013: Interpolation of Precipitation under Topographic Influence at Different Time Scales. *Water Resourc. Res.*, **49**(8), 4545-4565.

Barnes, S. L., 1973: Mesoscale Objective Analysis using Weighted Time-Series Observations. NOAA Tech. Memo. ERL NSSL-62, National Severe Storms Laboratory, Norman, OK, 60 pp. [NTIS COM-73-10781].

Barros, A. P. and D. P. Lettenmaier, 1994: Dynamic Modeling of Orographically Induced Precipitation. *Rev. of Geophysics*, **32** (3), 265-284.

Bayes, T. and R. Price, 1763: An Essay towards solving a Problem in the Doctrine of Chances. *Phil. Trans.,* **53**, 370-418.

Bent, A. E., 1946: Radar Detection of Precipitation. *J. Meteor.,* **3,** 78–84.

Bentley, M.L., 1997: Synoptic Conditions Favorable for the Formation of the 15 July 1995 Southeastern Canada/Northeastern U.S. Derecho Event. National Weather Digest. 21 (4), 31-39.

Berrocal, V., A. Gelfand, and D. Holland, 2010: A Bivariate Space-Time Downscaler under Space and Time Misalignment. *Annals of Applied Statistics*, **4**(4), 1942-1975.

Bhargava, M. and M. Danard, 1994: Application of Optimum Interpolation to the Analysis of Precipitation in Complex Terrain. *J. Applied Meteor.*, **33**, 508-518.

Bingeman, A. K., 2001: Improving Dam Safety Analysis by Using Physically-Based Techniques to Derive Estimates of Amospherically Maximum Precipitation. Ph.D. University of Waterloo, 269 pp.

Bosilovich, M. G., 2013: Regional Climate and Variability of NASA MERRA and Recent Reanalyses: U.S. Summertime Precipitation and Temperature. *J. Applied Meteo. and Climotol.*, **52**, 1939-1951.

Bonnin, G.M., D. Martin, B. Lin, T. Parzybok, M. Yekta, and D. Riley, 2006: NOAA Atlas 14 Precipitation-Frequency Atlas of the United States Volume 2 Version 3.0: Delaware, District of Columbia, Illinois, Indiana, Kentucky, Maryland, New Jersey, North Carolina, Ohio, Pennsylvania, South Carolina, Tennessee, Virginia, West Virginia. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, Maryland.

Bracken, C., B. Rajagopalan, L. Cheng, W. Kleiber, and S. Gangopadhyay, 2016: Spatial Bayesian Hierarchical Modeling of Precipitation Extremes over a Large Domain. *Water Resources Res.*, **52** (8), 6643-6655.

Briggs, P. R. and J. G. Cogley, 1996: Topographic Bias in Mesoscale Precipitation Networks. *J. Climate*, **9**, 205-218.

Brooks, H. E. and D. J. Stensrud, 2000: Climatology of Heavy Rain Events in the United States from Hourly Precipitation Observations. *Mon. Wea. Rev.*, **128**, 1194-1201.

Browning, K. A., F. F. Hill, and C. W. Pardoe, 1974: Structure and Mechanism of Precipitation and the Effect of Orography in a Wintertime Warm Sector. *Quart. J. R. Met. Soc.*, **100**, 309-330.

Caldwell, R. J., England, J. F. Jr., Sankovich, V. L., 2011: Application of Radar-Rainfall Estimates to Probable Maximum Precipitation in the Carolinas for the Nuclear Regulatory Commission, Office of Nuclear Regulatory Research. Bureau of Reclamation, Denver, Colorado, 106 pp.

Caldwell, R. J., V. S. Bahls, R. Swain, and J. F. England, Jr., 2014: El Vado Dam Meteorology for Corrective Action Study. Bureau of Reclamation, Denver, Colorado, 123 pp.

Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M.A. Brubaker, J. Guo, P. Li, and A. Riddell, 2016: Stan: A probabilistic programming language. *J. Stat. Software,* **20.**

Castro, L. M., J. Gironás, and B. Fernández, 2014: Spatial Estimation of Daily Precipitation in Regions with Complex Relief and Scarce Data using Terrain Orientation. *J. Hydrology*, **517**, 481-492.

Cavazos, T., 1999: Large-Scale Circulation Anomalies Conducive to Extreme Precipitation Events and Derivation of Daily Rainfall in Northeastern Mexico and Southeastern Texas. *J. Clim.*, **12**, 1506-1523.

Cavazos, T., 2000: Using Self-Organizing Maps to Investigate Extreme Climate Events: An Application to Wintertime Precipitation in the Balkans. *J. Clim.*, **13**, 1718-1732.

Chilés, J. P. and P. Delfiner, 1999: Geostatistics: Modeling Spatial Uncertainty. New York, NY, USA: Wiley.

Clark, M. P., and A. G. Slater, 2006: Probabilistic Quantitative Precipitation Estimation in Complex Terrain. *J. Hydrometeor.*, **7**, 3–22, doi:10.1175/JHM474.1.

Cleveland, W. S., 1979: Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Amer. Stat. Assoc.*, **74** (368), 829-836.

Colton, D. E., 1965: Precipitation Analysis for Operational Streamflow Forecasting – The Use of Mesoscale Numerical Modelling to Enhance Estimation of Precipitation in Mountainous Areas. Precipitation Analysis for Hydrologic Modelling, University of California, Davis.

Colton, D. E., 1976: Numerical Simulation of the Orographically Induced Precipitation Distribution for Use in Hydrologic Analysis. *J. Applied Meteor.*, **15** (12), 1241-1251.

Cong, S., Y. Li, J. L. Vogel, and J. C. Shaake, 1993: Identification of the Underlying Distribution Form of Precipitation by Using Regional Data. *Wat. Resourc. Res.*, **29**(4), 1103-1111.

Coniglio, M. C., K. L. Elmore, J. S. Kain, S. J. Weiss, M. Xue, and M. L. Weisman, 2010: Evaluation of WRF Model Output for Severe Weather Forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment. *Wea. Forecasting,* **25**, 408–427, doi: 10.1175/2009WAF2222258.1.

Cooley, D., D. Nychka, and P. Naveau, 2007: Bayesian spatial modeling of extreme precipitation return levels. *J. Amer. Stat. Assoc.,* **102** (479), 824-840.

Cosgrove, B. A., and Coauthors, 2003: Real-Time and Retrospective Forcing in the North American Land Data Assimilation System (NLDAS) Project, *J. Geophys. Res.*, **108**, 8842, doi:10.1029/2002JD003118, D22.

Cressie, N., 1990: The origins of kriging. *Mathematical Geology*, **22** (3), 239-252.

Cressman, G. P., 2002: An operational Objective Analysis and Mapping Techniques for Rainfall Fields: An Objective Comparison, *Water Resourc. Res.*, **18**, 413-431.

Crochet, P., T. Jóhannesson, T. Jónsson, O. Sigurðsson, H. Björnsson, F. Pálsson, and I. Barstad, 2007: Estimating the Spatial Distribution of Precipitation in Iceland using a Linear Model of Orographic Precipitation. *J. Hydrometeo*, **8** (6), 1285-1306.

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: *A Statistical-Topographic* Model for Mapping Climatological Precipitation over Mountainous Terrain. *J. Applied Meteor.*, **33**, 140-158.

Davis, J.C., 1986: Statistics and Data Analysis in Geology. John Wiley & Sons, New York.

Diodato, N., 2005: The Influence of Topographic Co-variables on the Spatial Variability of Precipitation over Small Regions of Complex Terrain. *Int. J. Climatology*, **25** (3), 351-363.

Dominguez, F., E. Rivera, D. P. Lettenmaier, and C. L. Castro, 2012: Changes in Winter Precipitation Extremes for the Western United States under a Warmer Climate as Simulated by Regional Climate Models. *Geophys. Res. Lett.*, **39**, 1-7.

Donley, D. E. and R. L. Mitchell, 1939: The Relation of Rainfall to Elevation in the Southern Appalachian Region. *Transactions, American Geophysical Union*, 14pp.

Draxler, R. R., B. Stunder, and G. Rolph, 1999: HYSPLIT4 user's guide, *Techreport*.

Durre, I., M.J. Menne. B.E. Gleason, T.G. Houston, and R.S. Vose, 2010: Comprehensive Automated Quality Assurance of Daily Surface Observations. *Journal of Applied Moeterology and Climatology*, **49** (8), 1615-1633.

Ehret, U., 2003: Rainfall and Flood Nowcasting in Small Catchments using Weather Radar. Ph.D. Eigenverlag des Inst. für Wasserbau an der Univ. Stuttgart, Stuttgart, Baden-Württemberg, Germany.

Eliassen, A., 1954: Provisional Report on Calculation of Spatial Covariance and Autocorrelation of the Pressure Field. Inst. Weather and Climate Res., Acad. Sci. Oslo, Rep. No. 5. Reprinted in: Data Assimilation Methods. Ed. L. Bengtsson, M. Ghil, E. Kdlen), SpringerVerlag, 319-330.

Elliott, R. D. and R. W. Shaffer, 1962: The Development of Quantitative Relationships between Orographic Precipitation and Air-Mass Parameters for Use in Forecasting and Cloud Seeding Evaluation. *J. Applied Meteor.*, **1**, 218-228.

Ezer, T. and G. L. Mellor, 1994: Diagnostic and Prognostic Calculations of the North Atlantic Circulation and Sea Level Using a Sigma Coordinate Ocean Model. *J. Geophys. Res.*, **99** (7), 14159-14171.

Fabry, F., 2015: Radar Meteorology: Principles and Practice. Cambridge University Press. Cambridge, United Kingdom, pp 254.

Fisher, R. A., 1912: On an Absolute Criterion for Fitting Frequency Curves. *Messenger Math*. 41155-41160.

Flesch, T. K. and G. W. Reuter, 2012: WRF Model Simulation of Two Alberta Flooding Events and the Impact of Topography. *J. Hydrometeor.,* **13**, 695–708, doi: 10.1175/JHM-D-11-035.1.

Fletcher, Robert D., 1951: Hydrometeorology in the United States. Compendium of Meteorology.

Food and Agriculture Organization of the United Nations (FAO), 2016: About FAO. Website: http://www.fao.org/about/en/. Accessed September 1, 2016.

Fowler, H. J., and C. G. Kilsby, 2003: A Regional Frequency Analysis of United Kingdom Extreme Rainfall from 1961 to 2000, *Int. J. Climatol.*, **23**, 1313 – 1334.

Foufoula-Georgiou, E., 1989: A probabilistic storm transposition approach for estimating exceedance probabilities of extreme precipitation depths. *Water Resourc. Res.*, **25** (5), pp.799-815.

Frederick, R. H., V. A. Myers, and E. P. Auciello, 1977: Five-to 60-Minute Precipitation-Frequency for the Eastern and Central United States. NOAA, National Weather Service, Silver Spring, M.D.

Früh, B., J.Bendix, T. Nauss, M. Paulat, A. Pfeiffer, J. W. Schipper, B. Thies, and H. Wernli, 2007: Verification of Precipitation from Regional Climate Simulations and Remote-Sensing Observations with Respect to Ground-Based Observations in the upper Danube Catchment. *Meteor. Z.*, **16**, 275-293.

Früh, B., H. Feldmann, H. Panitz, G. Schädler, D. Jacob, P. Lorenz, and K. Keuler, 2010: Determination of Precipitation Return Values in Complex Terrain and Their Evaluation. *J. Climate,* **23**, 2257–2274, doi: 10.1175/2009JCLI2685.1.

Fuentes, M., and A. Raftery, 2005: Model Evaluation and Spatial Interpolation by Bayesian Combination of Observations with Outputs from Numerical Models. *Biometrics*, **61** (1), 36-45.

Fujiyoshi, Y., T. Endoh, T. Yamada, K. Tsuboki, Y. Tachibana, and G. Wakahama, 1990: Determination of a Z-R Relationship for snowfall using Radar and High Sensitivity Snow Gauges. *Journal of Applied Meteorology*, **29** (2), 147-152.

Funk, C. C., P. J. Peterson, M. F. Landsfeld, D. H. Pedreros, J. P. Verdin, J. D. Rowland, B. E. Romero, G. J. Husak, J. C. Michaelsen, and A. P. Verdin, 2014: A quasi-global precipitation time series for drought monitoring: *U.S. Geological Survey Data Series 832*, 4 p., *http://dx.doi.org/10.3133/ds832*.

Gandin, L.S. 1963: Objective Analysis of Meteorological Felds. Leningrad. Hydromet. Press. Translated from Russian by Israel Program for Scientific Translations. Jerusalem 1965.

Gomez, J.A. and Sullivan, T., 2009: Development of Site Specific PMP and PMFs for Dam Safety Evaluations. National Hydorpower Association Northeast Regional Meeting, September 17, 2009.

Goovaerts, P., 2000: Geostatistical Approaches for Incorporating Elevation into the Spatial Interpolation of Rainfall. *J. Hydrology*, **228**(1), 113-129.

Greenwood, J. A., J. M. Ladwehr, and N. C. Matalas, 1979: Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form. *Wat. Resourc. Res.*, **15**(5), 1049-1054.

Groisman, P. Y. and D. R. Legates, 1994: The Accuracy of United States Precipitation Data. *Bull. Ameri. Meteor. Soc.*, **75** (3), 215-227.

Groisman, P. Y. and D. R. Legates, 1995: Documenting and Detecting Long-Term Precipitation Trends: Where We Are and What Should be Done. *Climatic Change*, **31** (2), 601-622.

Guentchev, G., J. J. Barsugli, and J. Eischeid, 2010: Homogeneity of Gridded Precipitation Datasets for the Colorado River Basin. *J. Applied Meteo. And Climatology*, **49**, 2404-2415.

Haberlandt, U., 2007: Geostatistical Interpolation of Hourly Precipitation from Rain Gauges and Radar for a Large-Scale Extreme Rainfall Event. *J. Hydrology*, **332** (1), 144-157.

Habib, E., W. F. Krajewski, and A. Kruger, 2001: Sampling Errors of Tipping-Bucket Rain Gauge Measurements. *J. Hydrol. Eng.*, **6** (2), 159-166.

Hall, V., Whaley, S., Hayes, B., Roque-Cruz, C.; DeNeale, S., Kanney, J., Kao, S., Quinlan, K., 2015: Report for the Audit of Applied Weather Associates, LLC, Regarding Site Specific Probable Maximum Precipitation Development in Support of Near-Term Task Force Recommendation 2.1 Flood Hazard Reevaluations. Office of Nuclear Regulatory Research, Rockville, Maryland, 30 pp.

Hanel, M, and T. Adri Buishand, 2010: On the Value of Hourly Precipitation Extremes in Regional Climate Model Simulations. *J. Hydrology*, **393**, 265-273.

Hansen, E.M., Schwarz, F.K., and Riedel, J.T., 1977: Probable Maximum Precipitation Estimates, Colorado River and Great Basin Drainages Hydrometeorological Report No. 49, National Weather Service, National Oceanic and Atmospheric Administration, US Department of Commerce, Silver Spring, Maryland, 298 pp.

Hansen, J., 2001: PMPs Never Happen – Or Do They? Association of State Dam Safety Officials 2001 Annual Conference Proceedings.

Hicks, N. S., J. A. Smith, A. J. Miller, P. A. Nelson, 2005: Catastrophic Flooding from an Orographic Thunderstorm in the Central Appalachians. *Wat. Resourc. Res.*, **41**, 17pp.

Higgins, R. W., V. E. Kousky, V. B. S. Silva, E. Becker, and P. Xie, 2010: Intercomparison of Daily Precipitation Statistics over the United States in Observations and in NCEP Reanalysis Products. *J. Climate*, **23**, 4637-4650.

Hill, F. F., K. A. Browning, and M. J. Bader, 1981: Radar and Raingauge Observations of Orographic Rain over South Wales. *Quart. J. R. Met. Soc.*, 107, 643-670.

Hobbs, P. V., 1975: The Nature of Winter Clouds and Precipitation in the Cascade Mountains and their Modification by Artificial Seeding. Part I: Natural Conditions. *J. Applied Meteor.*, **14**, 783-804.

Hofstra, N., M. New, and C. McSweeney, 2010: The Influence of Interpolation and Station Network Density on the Distributions and Trends of Climate Variables in Gridded Daily Data. *Clim. Dyn.*, **35**, doi:10.1007/s00382-009-0698-1

Holman, K. D. and S. J. Vavrus, 2012: Understanding Simulated Extreme Precipitation Events in Madison, Wisconsin, and the Role of Moisture Flux Convergence during the Late Twentieth and Twenty-First Centuries. *J. Climate*, **13**, 877-894.

Holton, J. R. and Gregory J. Hakim, 2013: An Introduction to Dynamic Meteorology. 5th ed. Waltham, MA: Academic. Print.

Horel, J. D. and C. V. Gibson, 1994: Analysis and Simulation of a Winter Storm over Utah. *Weather and Forecasting*, **9**, 479-494.

Hosking, J. R. M., 1990: *L*-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics. *J. R. Stat. Soc. Ser. B. Methodol.*, **52**(1), 105–124.

Hosking, J. R. M., 2015: Regional Frequency Analysis using L-Moments. R package, Version 3.0-1. URL: http://CRAN.R-project.org/package=lmomRFA.

Hosking, J. R. M. and J. R. Wallis, 1997: Regional Frequency Analysis. Cambridge University Press, Melbourne, Australia.

Hou, D., M. Charles, Y. Luo, Z. Toth, Y. Zhu, R. Krzysztofowicz, Y. Lin, P. Xie, D. Seo, M. Pena, and B. Cui, 2014: Climatology-Calibrated Precipitation Analysis at Fine Scales: Statistical Adjustment of Stage IV toward CPC Gauge-Based Analysis. *J. Hydrometeo.*, **15**, 2542-2557.

Houze, R.A., Jr., 2012. Orographic Effects on Precipitation Clouds. *Review of Geophysics*, 50, RG101, doi:10.1029/2011RG000365.

Hughes, M., K. M. Mahoney, P. J. Neiman, B. J. Moore, M. Alexander, and F. Martin Ralph, 2014: The Landfall and Inland Penetration of a Flood-Producing Atmospheric River in Arizona. Part II: Sensitivity of Modeled Precipitation to Terrain Height and Atmospheric River Orientation. *J. Hydrometeor.,* **15**, 1954–1974, doi: 10.1175/JHM-D-13-0176.1.

Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global Precipitation at One-Degree Daily Resolution from Multisatellite Observations. *J. of Hydrometeo.*, **2** (1), 36-50.

Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4), 2013: Climate Change 2013: The Physical Science Basis.

Jin, B., Y. Wu, B. Miao, X. L. Wang, and P. Guo, 2014: Bayesian Spatiotemporal Modeling for Blending in situ Observations with Satellite Precipitation Estimates. *J. Geophys. Res.: Atmospheres*, **119**(4), 1806-1819.

Johnson, M.E., 1987: Multivariate Statistical Simulation. Wiley, 230 pp.

Joyce, R.J., Janowiak, J.E., Arkin, P.A. and Xie, P., 2004: CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *Journal of Hydrometeorology*, **5**(3), pp.487-503.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643.

Kappel, W.D., Muhlestein, G.A., Hulstrand, D.M., McGlone, D., Steinhilber, K., Lawrence, B., Rodel, J., Parzybok, T., Tomlinson, E.M., 2014: Probable Maximum Precipitation Study for Wyoming. Applied Weather Associates, Monument, Colorado, 154 pp.

Kappel, W. D., Hulstrand, D. M., Rodel, J., Muhlestein, G. A., Steinhilber, K., McGlone, D., and B. Lawrence, 2015: Probable Maximum Precipitation Study for Virginia. Applied Weather Associates, Monument, Colorado, 152 pp.

Katzfey, J. J., 1995: Simulation of Extreme New Zealand Precipitation Events. Part I: Sensitivity to Orography and Resolution. *Mon. Wea. Rev.,* **123**, 737-754.

Kaufman, L. and Rousseeuw, P., 1990: Finding Groups in Data: An Introduction to Cluster Analysis, Wiley.

Kawase, H., T. Yoshikane, M. Hara, F. Kimura, T. Yasunari, B. Ailikun, H. Ueda, and T. Inoue, 2009: Intermodel Variability of Future Changes in the Baiu Rainband Estimated by the Pseudo Global Warming Downscaling Method, *J. Geophys. Res.*, **114**, D24110, doi:10.1029/2009JD011803.

Keeney, D. P. and K. D. Holman, 2015: Island Park Dam Meteorology for Application in Hydrologic Hazard Analysis, Minidoka Project, Idaho, Pacific Northwest Region. Bureau of Reclamation, Denver, Colorado, 88 pp.

Kennedy, A. D., X. Dong, B. Xi, S. Xie, Y. Zhang, and J. Chen, 2011: A Comparison of MERRA and NARR Reanalyses with the DOE ARM SGP Data. *J. Climate*, **9**, 4541-4557.

King, A. D., Alexander, L. V. and Donat, M. G. (2013), The Efficacy of Using Gridded Data to Examine Extreme Rainfall Characteristics: A Case Study for Australia. *Int. J. Climatol.*, **33**, 2376–2387. doi:10.1002/joc.3588

Konrad, C. E., 1994: Moisture Trajectories Associated with Heavy Rainfall in the Appalachian Region of the United States. *Physical Geography*, **15**(3), 227-248. doi: 10.1080/02723646.1994.10642514

Kohonen, T., 1990: The Self-Organizing Map. Proceedings of the IEEE, 78 (9).

Koutsoyiannis, D., 1999: A Probabilistic View of Hershfield's Method for Estimating Probable Maximum Precipitation. *Water Resourc. Res.*, **35**(4), 1313-1322.

Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. and Oceanic Tech.*, **15**(3), 809-817.

Kysely, J. and J. Picek, 2007: Probability Estimates of Heavy Precipitation Events in a Flood-Prone Central-European Region with Enhanced Influence of Mediterranean Cyclones. *Adv. Geosci.*, **12**, 43-50.

Lackmann, G. M., 2013: The South-Central U.S. Flood of May 2010: Present and Future. *J. Climate,* **26**, 4688–4709, doi: 10.1175/JCLI-D-12-00392.1.

Le, N. D., and J. V. Zidek, 2006: Statistical Analysis of Environmental Space-Time Processes. New York, NY, USA: Springer.

Lenderink, G., H.Y. Mok, T.C. Lee, G.J. van Oldenborgh, 2011: Scaling Trends of Hourly Precipitation Extremes in Two Different Climate Zones – Hong Kong and the Netherlands. *Hydro. and Earth Sys. Sci.*, **15** (9), 3033-3041.

Leung, L. R. and Y. Qian, 2009: Atmospheric Rivers Induced Heavy Precipitation and Flooding in the Western U.S. Simulated by the WRF Regional Climate Model. *Geophys. Res. Lett.*, **36**, L03820, doi:10.1029/2008GL036445.

Lin, G-F. and L-H. Chen, 2006: Identification of Homogenous Regions for Regional Frequency Analysis using the Self-Organizing Map. *J. Hydrology*, **324**, 1-9.

List, R., 1988: A Linear Radar Reflectivity-rainrate Relationship for Steady Tropical Rain. *Journal of the Atmospheric Sciences*. **45** (23), 3564-3572.

Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 2013: A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions. *J. Climate*, **26**, 9384–9392, doi:10.1175/JCLI-D-12-00508.1.

Livneh B**.**, T.J. Bohn, D.S. Pierce, F. Munoz-Ariola, B. Nijssen, R. Vose, D. Cayan, and L.D. Brekke, **2015**: A Spatially Comprehensive, Hydrometeorological Dataset for Mexico, the U.S., and Southern Canada 1950-2013, *Nature Scientific Data*, 5:150042, doi:10.1038/sdata.2015.42.

Luo, L., and Coauthors, 2003: Validation of the North American Land Data Assimilation System (NLDAS) Retrospective Forcing over the Southern Great Plains, *J. Geophys. Res.*, **108,** 8843, doi:10.1029/2002JD003246.

Mahoney, K., J. Lukas, and B. McCormick, 2016: Examining Terrain Elevation Assumptions Used in Current Extreme Precipitation Estimation Practices: A Modeling Study of the 2013 Colorado Front Range Floods. Presentation at the American Meteorological Society meeting.

Manz, B., W. Buytaert, Z. Zulkafli, W. Lavado, B. Willems, L. A. Robles, and J. P. Rodríguez-Sánchez, 2016: High-Resolution Satellite-Gauge Merged Precipitation Climatologies of the Tropical Andes. *J. Geophys. Res.: Atmos.*, **121**, 1190-1207.

Masson, D. and C. Frei, 2014: Spatial Analysis of Precipitation in a High-Mountain Region: Exploring Methods with Multi-Scale Topographic Predictors and Circulation Types. *Hydro. and Earth Sys. Sci.*, **18** (11), 4543-4563.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nussen, 2002: A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States. *J. Climate*, **15**, 3237-3251.

Mead, D. W., 1919: Hydrology, the Fundamental Basis of Hydraulic Engineering, McGraw-Hill Book Company, New York.

Medina, S., B. F. Smull, R. H. Houze Jr., and M. Steiner, 2005: Cross-Barrier Flow during Orographic Precipitation Events: Results from MAP and IMPROVE. *J. Atmos. Sci.-Special Section*, **62**, 3580-3598.

Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and T.G. Houston, 2012: An Overview of the Global Historical Climatology Network-daily Database. *Journal of Atmospheric and Oceanic Technology*. **29** (7), 897-910.

Merz, B. and A. H. Thieken, 2005: Separating Natural and Epistemic Uncertainty in Flood Frequency Analysis. *J. Hydrology*, **309**, 114-132.

Mesinger, F., and Coathors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343-360.

Michalakes, J., S. Chen, J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, and W. Skamarock, 2001: Development of a Next-Generation Regional Weather Research and Forecast Model. National Center for Atmospheric Research Annual Publication.

Micovic, Z., M. G. Schaefer, and G. H. Taylor, 2015: Uncertainty Analysis for Probable Maximum Precipitation Estimates. *J. Hydrology*, **521**, 360-373.

Miller, J.F., 1964: *Technical paper No. 49 Two-to-Ten-Day Precipitation for Return Periods of 2 to 100 Years in the Contiguous United States.* U.S. Department of Agriculture, Office of Hydrology, US Weather Bureau, Washington, D.C., 29 p.

Miller, J.F., R.H. Frederick, and R.S. Tracey, 1973: *NOAA Atlas 2, Precipitation-Frequency Atlas of the Western United States.* U.S. Department of Commerce, NOAA, National Weather Service, Washington, D.C.

Milly, P.C.D., J. Betancourt, M. Falkenmark, R.M. Hirsch, Z.W. Kundzewicz, D.P. Lettenmaier, and R.J. Stouffer, 2008: Stationarity is Dead: Whither Water Management? *Science Climate Change*, **319**, 573-574.

Mladjic, B., L. Sushama, M. Khaliq, R. Laprise, D. Caya, and R. Roy, 2011: Canadian RCM Projected Changes to Extreme Precipitation Characteristics over Canada. *J. Climate,* **24**, 2565–2584, doi: 10.1175/2010JCLI3937.1.

Moeng, C-H, J. Dudhia, J. Klemp, and P. Sullivan, 2007: Examining Two-Way Grid Nesting for Large Eddy Simulation of the PBL Using the WRF Model. *Mon. Wea. Rev.*, **135**, 2295-2311.

Mueller, M. and K. M. Mahoney, 2016: Improving Extreme Precipitation Estimation using Regional, High-Resolution Model-Based Methods: Final Summary for Performance Period February 2014 - 2016 CIRES-Reclamation Cooperative Agreement R11AC81334.

Nan, Z., S. Wang, X. Liang, T. E. Adams, W. Teng, and Y. Liang, 2010: Analysis of Spatial Similarities between NEXRAD and NLDAS Precipitation Data Products. *IEEE J. Selected Top. in App. Earth Obs. And Rem. Sens.*, **3**, 371-385.

National Centers for Environmental Precipitation (NCEP), 2016: NCEP North American Regional Reanalysis (NARR), for 1979 to Present. https://data.noaa.gov/dataset/ncep-north-american-regional-reanalysis-narr-for-1979-to-present

Newman, A. J., M. P. Clark, J. Craig, B. Nijssen, A. Wood, E. Gutmann, N. Mizukami, L. Brekke, and J. R. Arnold, 2015: Gridded Ensemble Precipitation and Temperature Estimates for the Contiguous United States, *J. of Hydrometeo.*, **16**, 2481-2500.

Nicolini, M., K. M. Waldron, and J. Paegle, 1993: Diurnal Oscillations of Low-Level Jets, Vertical Motion, and Precipitation: A Model Case Study. *Mon. Wea. Rev.*, **121**, 2588-2610.

Nikulin, G., E. Kjellstrom, U. Hansson, G. Strandberg, and A. Ullerstig, 2001: Evaluation and Future Projections of Temperature, Precipitation and Wind Extremes over Europe in an Ensemble of Regional Climate Simulations. *Tellus*, **63A**, 41-55.

Norbiato, D., M. Borga, M. Sangati, and F. Zanon, 2007: Regional Frequency Analysis of Extreme Precipitation in the Eastern Italian Alps and the August 29, 2003 Flash Flood. *J. Hydro.*, **245**, 149-166.

Omre, H., 1987: Bayesian Kriging – Merging Observations and Qualified Guesses in Kriging. *Mathematic Geology*, **19** (1), 25-39.

Overeem, A., T. A. Buishand, and I. Holleman, 2009: Extreme Rainfall Analysis and Estimation of Depth-Duration-Frequency Curves using Weather Radar. *Water Resour. Res.*, **45**, W10424. doi:10.1029/2009WR007869.

Paegle, J. and D. W. McLawhorn, 1983: Numerical Modeling of Diurnal Convergence Oscillations above Sloping Terrain. *Mon. Wea. Rev.,* **111**, 67-85.

Perica, S., D. Martin, S. Pavlovic, I. Roy, M. St. Laurent, C. Trypaluk, D. Unruh, M. Yekta, and G. Bonnin, 2013: NOAA Atlas 14 Precipitation-Frequency Atlas of the United States Volume 9 Version 2.0: Southeastern States (Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi). U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, Maryland.

R Core Team, 2016: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rajagopalan, R. and U. Lall, 1998: Locally Weighted Polynomial Estimation of Spatial Precipitation. *J.Geog. Info. and Dec. Anal.*, **2** (2), 44-51.

Recherche en Prévision Numérique, 1997: PMS Model User's Manual, Prepared for BCHydro Company, Hydroelectric Engineering Division by Recherche en Prévision Numérique, Atmospheric Environment Service, Dorval, Québec.

Rennie, S. J., A. J. Illingworth, and S. L. Dance, 2010: On Differentiating Ground Clutter and Insect Echoes from Doppler Weather Radars Using Archived Data. *Atmos. Meas. Tech. Discuss.*, **3**, 1843–1860.

Rhea, J. O., 1978: Orographic Precipitation Model for Hydrometeorological Use, Atmospheric Science Paper No. 287. Department of Atmospheric Science, Colorado State University, Fort Collins, CO.

Rienecker, M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3648-3624.

Roe, G. H., 2005: Orographic Precipitation, *Annu. Rev. Earth Planet. Sci.*, **33**, 645-671.

Rozante, J. R., D. S. Moreira, L. G. G. deGoncalves, and D. A. Vila, 2010: Combining TRMM and surface observations of precipitation: Technique and validation over South America. *Weather and Forecasting*, **25** (3), 885-894.

Rudari, R., Entekhabi, D. and Roth, G., 2004: Terrain and multiple-scale interactions as factors in generating extreme precipitation events. *J. Hydrometeo.*, **5** (3), pp.390-404.

Ryberg, K.R., Emerson, D.G., and Macek-Rowland, K.M., 2009, Solid precipitation measurement intercomparison in Bismarck, North Dakota, from 1988 through 1997: U.S. Geological Survey Scientific Investigations Report 2009–5180, 24 p.

Sankovich, V., R. J. Caldwell, and K. Mahoney, 2012: Green Mountain Dam Climate Change. Technical Report. Department of the Interior, Bureau of Reclamation, Technical Service Center, Denver, CO.

Sansó, B., and L. Guenni, 1999: Venezuelan rainfall data analysed by using a Bayesian spacetime model. *J. Royal Stat. Soc.: Series C (Applied Statistics)*, **48** (3), 345-362.

Sapiano, M.R.P. and Arkin, P.A., 2009: An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *Journal of Hydrometeorology*, **10** (1), pp.149-166.

Sarangi, A., C. A. Cox, and C. A. Madramootoo, 2005: Geostatistical methods for prediction of spatial variability of rainfall in a mountainous region. *Trans. of the ASAE*, **48** (3), 943-954.

Sarker, R. P., 1965: A Theoretical Study of Mountain Waves on Western Ghats. *Indian Journal of Meteorology and Geophysics,* **16** (4), 555-572.

Sarker, R. P., 1966: A Dynamical Model of Orographic Rainfall. *Mon. Wea. Rev.,* **94** (9), 555-572.

Schaefer, M. G., 1990: Regional Analyses of Precipitation Annual Maxima in Washington State, *Water Resour. Res.*, **26**(1), 119–131, doi:10.1029/WR026i001p00119.

Schaefer, M. G., 1994: PMP and Other Extreme Storms: Concepts and Probabilities. Supervisor Dam Safety Section, Water Resources Program, Washington Department of Ecology, 61-73.

Schaefer, M. G., 2013: Regional Precipitation-Frequency Analysis and Extreme Storms including PMP Current State of Understanding/Practice. NRC Workshop – Probabilistic Flood Hazard Assessment. January 2013.

Schwarz, F. K., 1970: The Unprecendented Rains in Virginia Associated with the Remnants of Hurricane Camille. *Mon. Wea. Rev.*, **98** (11), 851-859.

Schwarz, F. K., 1973: Addendum to Hydrometeorological Report No. 45, Probable Maximum and
TVA Precipitation for Tennessee River Basins Up to 3,000 Square Miles in Are and Durations to 72 Hours. Hydrometeorological Branch, Office of Hydrology, Silver Springs, MD, 201 pp.

Schreiner, L. C. and J. T. Riedel, 1978: Hydrometeorological Report No. 51, Probable Maximum Precipitation Estimates, United States East of the 105th Meridian. Hydrometeorological Branch, Office of Hydrology, National Weather Service.

Sea, B. and W. F. Krajewski, 2010: Scale Dependence of Radar Rainfall Uncertainty: Initial Evaluation of NEXRAD's New Super-Resbaolution Data for Hydrologic Applications, *J. Hydrometeor.*, **11**, 1191-1198.

Schoelzel, C. and P. Friederichs, 2008: Multivariate Non-Normally Distributed Random Variables in Climate Research – Introduction to the Copula Approach. *Nonlin. Processes Geophys.*, **15** (5), pp.761-772.

Semmler, T. and D. Jacob, 2004: Modeling Extreme Precipitation Events – A Climate Change Simulation for Europe. *Global and Planetary Change*, **44**, 119-127.

Sevruk, B., 1982: Method of correction for systematic error in point precipitation measurement for operational use. WMO Tech. Note WMO-589, 91 pp. [Available from World Meteor. Org., Case Postale 2300, CH-1211, Geneva, Switzerland.]

Sevruk, B., 1987: Point Precipitation Measurements: Why are They Not Corrected? *Water for the Future: Hydrology in Perspective*, 164, 477-486.

Shepard, D., 1968: A Two Dimensional Interpolation Function for Irregularly Spaced Data, in *Prod. 234d National Conf. of the Association for Computing Machinery*, Prinecton, NJ, ACM, 517-524.

Shepard, D. S., 1984: Computer mapping: The SYMAP Interpolation Algorithm. *Spatial Statistics and Models,* G. L. Gaile and C. J.Willmott, Eds., D. Reidel, 133–145.

Sinclair, S., and G. Pegram, 2005: Combining radar and rain gauge rainfall estimates using conditional merging. *Atmos. Sci. Letters*, **6** (1), 19-22.

Sklar, A., 1959: Distribution functions of n dimensions and margins. *Publications of the Institute of Statistics of the University of Paris*, 8, 229-231.

Smith, R.B. and I. Barstad, 2004: A linear theory of orographic precipitation, *J. Atmos. Sci.*, **61**, 1377-1391.d

Smith, B. L., S. E. Yuter, P. J. Neiman, and D. E. Kingsmill, 2010: Water Vapor Fluxes and Orographic Precipitation over Northern California Associated with a Landfalling Atmospheric River. *Mon. Wea. Rev.*, **138**, 74-138.

Tabios III, G. Q. and J. D. Salas, 1985: A Comparative Analysis of Techniques for Spatial Interpolation of Precipitation, *Water Resources Bulletin*, **21** (3), 365-380.

Thiessen, A. H., 1911: Precipitation Averages for Large Areas, *Mon. Wea. Rev.*, **39** (7), 1082-1089.

Thornton, P. E., S. W. Running, and M. A. White, 1997: Generating Surfaces of Daily Meteorological Variables over Large Regions of Complex Terrain. *J. Hydrol.*, **190**, 214-251, doi:10.1016/S0022-1694(96)03128-9.

Tian, Y., Peters-Lidard, C.D., Eylander, J.B., Joyce, R.J., Huffman, G.J., Adler, R.F., Hsu, K.L., Turk, F.J., Garcia, M. and Zeng, J., 2009: Component analysis of errors in satellite-based precipitation estimates. *J. Geophys. Res.: Atmospheres*, **114** (D24).

Tomlinson, E.M., Henz, J.F., Williams, R.A., 2003: Technical Review for the Probable Maximum Precipitation (PMP) Site-Specific Study for Cherry Creek Reservoir. Applied Weather Associates, Monument, Colorado, 314 pp.

Tomlinson, E.M., Kappel, W.D., Muhlestein, G.A., Hulstrand, D.M., Parzybok, T., 2013: Probable Maximum Precipitation Study for Arizona. Applied Weather Associates, Monument, Colorado, 275 pp.

Tsonis, Anastasios A., 2007: An Introduction to Atmospheric Thermodynamics. 2nd ed. Cambridge, UK: Cambridge UP. Print.

Trentmann, J., and Coauthors, 2009: Multi-Model Simulations of a Convective Situation in Low-Mountain Terrain in Central Europe. *Meteorol. Atmos. Phys.*, **103**, 95-103.

Treyfry, C. M., D. W. Watkins Jr., and D. Johnson, 2005: Regional Rainfall Frequency Analysis for the State of Michigan. *J. Hydrol. Eng.*, **10**(6), 437-449.

USDA 1993: Part 630 Hydrology National Engineering Handbook, Chapter 4: Storm Rainfall Depth

U. S. Weather Bureau (USWB), 1945: Revised Report on Maximum Possible Precipitation, Los Angeles Area, California Hydrometeorological Report No. 21B, Washington D.C.

U. S. Weather Bureau (USWB), 1947: Maximum Possible Precipitation, San Joaquin Basin, California Hydrometeorological Report No. 24, Washington, D.C.

U.S. Weather Bureau (USWB), 1956: Seasonal Variation of the Probable Maximum Precipitation East of the 105th Meridian for Areas from 10 to 1,000 Square Miles and Duration of 6, 12, 24 and 48 Hours, Hydrometeorology Report No. 33, Washington, D.C.

U.S. Weather Bureau (USWB), 1961a: Generalized Estimates of Probable Maximum Precipitation and Rainfall-Frequency Data for Puerto Rico and Virgin Islands Technical Paper No. 42, Hydrometeorological Branch, Office of Hydrology, Washington, D.C., 101 pp.

U.S. Weather Bureau (USWB), 1961b: Probable Maximum Precipitation in California Hydrometeorological Report No. 36, Hydrometeorological Branch, Office of Hydrology, Washington, D.C., 226 pp.

U.S. Weather Bureau (USWB), 1961c: Rainfall-Frequency Atlas of the United States Technical Report No. 40, Hydrometeorological Branch, Office of Hydrology, Washington, D.C., 115 pp.

U.S. Weather Bureau (USWB), 1963: Probable Maximum Precipitation in the Hawaiian Islands Hydrometeorological Report Number 39, Hydrometeorological Branch, Office of Hydrology, Washington, D.C., 108 pp.

U.S. Weather Bureau (USWB), 1965: Probable Maximum and TVA Precipitation over the Tennessee River Basin above Chattanooga Hydrometeorological Report Number 41, Hydrometeorological Branch, Office of Hydrology, Washington, D.C., 157 pp.

Vasquez, T., 2013. Weather Radar Handbook. Weather Graphics Technologies. Garland, Texas. 149 pp.

Viale, M., B. A. Houze Jr., and K. L. Rasmussen, 2013: Uptream Orographic Enhancement of a Narrow Cold-Frontal Rainband Approaching the Andes. *Mon. Wea. Rev.*, 141, 1708-1730.

Verdin, A., C. Funk, B. Rajagopalan, and W. Kleiber, 2016: Kriging and Local Polynomial Methods for Blending Satellite-Derived and Gauge Precipitation Estimates to Support Hydrologic Early Warning Systems. *IEEE Trans. Geosci. and Remote Sensing*, **54** (5), 2552-2562.

Vila, D. A., L. G. G. deGoncalves, D. L. Toll, and J. R. Rozantes, 2009: Statistical evaluation of combined daily gauge observations and rainfall satellite estimates over continental South America. *J. Hydrometeo.*, **10** (2), 533-543.

Wallis, J. R., M. G. Schaefer, B. L. Barker, and G. H. Taylor, 2007: Regional Precipitation-Frequency Analysis and Spatial Mapping for 24-hour and 2-hour Durations for Washington State. *Hydrology and Earth System Sciences Discussions*, European Geosciences Union, 2007, **11** (1), pp.415-442.

Waldron, K. M., 1994: Sensitivity of Local Model Predictions to Large Scale Forcing. Ph.D. thesis, University of Utah. 150 pp.

Wang, B. M., 1984: Estimation of Probable Maximum Precipitation: Case Studies. *J. Hydraul. Eng.*, **110** (10), 1457-1472.

Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0-36-h Explicit Convective Forecasts with the WRF-ARW Model. *Wea. and Forecasting*, **23**, 407-437.

White, A. B. and Coauthors, 2012: NOAA's Rapid Response to the Howard A. Hanson Dam Flood Risk Management Crisis. *Bull. Amer. Meteor. Soc.,* **93**, 189–207, doi: 10.1175/BAMS-D-11-00103.1.

Wijngaard, J. B., Klein Tank, A. G. G., and G. P. Konnen, 2003: Homogeneity of 20th Century European Daily Temperature and Precipitation Series. *Int. J. Climatol.*, **23**, 679–692. DOI: 10.1002/joc.906

World Data Center for Meteorology (WDCMET), 2016: WDCMET Home. Website: https://www.ncdc.noaa.gov/wdcmet. Accessed September 1, 2016.

World Meteorological Organization (WMO), 1985: International Organizing Committee for the WMO Solid Precipitation Measurement Intercomparison. Final Rep. of the First Session, Norkoping, Sweden, WMO, 31 pp. [Available from World Meteor. Org., Case Postale 2300, CH-1211, Geneva, Switzerland.]

World Meteorological Organization (WMO), 2009: Manual on Estimation of Probable Maximum Precipitation. WMO-No.1045. Geneva, 293 pp.

Wright, D.B., Smith, J.A., Villarini, G. and Baeck, M.L., 2013: Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition. *J. Hydrology*, **488**, pp.150-165.

Wright, D.B., Smith, J.A., Villarini, G. and Baeck, M.L., 2014: Long-Term High-Resolution Radar Rainfall Fields for Urban Hydrology. *J. of the American Water Resources Association,* **50** (3), pp.713-734.

Wuest, M., Frei, C., Altenhoff, A., Hagen, M. Litschi, M., and C. Schar, 2010: A Gridded Hourly Precipitation Dataset for Switzerland using Rain-Gauge Analysis and Radar-Based Disaggregation. *Int. J. Climatol.*, **30**, 1764-1775.

Xia, Y., and Coauthors, 2012: Continental-Scale Water and Energy Flux Analysis and Validation for the North American Land Data Assimilation System Project Phase 2 (NLDAS-2): 1. Intercomparison and Application of Model Products. *J. Geophys. Res.*, **117**, D03109, doi:10.1029/2011JD016048.

Xie, P., and P. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78** (11) 2539-2558.

Xie, P., J. E. Janowiak, P. A. Arkin, R. F. Adler, A. Gruber, R. Ferraro, G. J. Huffman, and S. Curtis, 2003: GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *J. Clim.*, **16** (13), 2197-2214.

Xie, P., and A. Y. Xiong, 2011: A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *J. Geophys. Res.*, **116** (D21).

Yang, D., B. E. Goodison, J. R. Metcalf, et al., 1998: Accuracy of NWS 8" Standard Nonrecording Precipitation Gauge: Results and Application of WMO Intercomparison, *J. Atmos. Oceanic Technol.*, **15**, 54-68.

Yarnell, D. L., 1935: Rainfall Intensity-Frequency Data. U.S. Department of Agriculture. Miscellaneous Publication No. 204. Washington, D.C.

Zundorfer, E.A., F.K. Schwarz, E.M. Hansen, D.D. Fenn, and J.F. Miller, 1986: Probable Maximum and TVA Precipitation Estimates With Areal Distribution for Tennessee River Drainages Less Than 3,000 Mi^2 in Area, Hydrometeorological Report No. 56, National Oceanic and Atmospheric Administration, US Department of Commerce, Tennessee Valley Authority, Silver Spring, Maryland, 224 pp.

Zwiers, F. W., and V. V. Kharin, 1998: Changes in the Extremes of the Climate Simulated by CCC GCM2 under CO2 Doubling. *J. Climate*, **11**, 2200–2222.

APPENDIX A

NEWMAN ET AL. (2015) ENSEMBLE GENERATION

This appendix provides an overview of the methods used to develop the gridded ensemble dataset used in the gridded frequency analysis from section 5. We assume the reader has fundamental familiarity with spatial statistics. The authors of this report were not involved in the development of the Newman et al. (2015) gridded ensemble generation (hereafter the Newman dataset). Rather, the gridded ensemble dataset was used to illustrate the transferability of frequency analysis methods from in situ observations to a gridded dataset. If the reader requires more information to understand the methods used in developing the Newman dataset, the authors suggest direct contact with the research team responsible for developing the Newman dataset. For more information on the ensemble generation process, see Newman et al. (2015) and Clark and Slater (2006).

Overview

The general methodology presented in Clark and Slater (2006) was modified by Newman et al. (2015) to develop the gridded ensemble dataset. Briefly, locally-weighted regression models are fitted to produce "best estimate" values (i.e., the expected value) of precipitation and temperature at 1/8^{o} resolution. Regression residuals were used to perturb the best estimate values with correlated random samples. The analysis procedure resulted in a group of 100 plausible precipitation and temperature grids (each valid from 1980 through 2012) that adequately represents the uncertainty of gridded datasets with 1/8º resolution.

The following description of the gridded ensemble generation methodology has been abstracted and paraphrased from the Newman et al. (2015) document.

Data sources

Station data used to develop the Newman dataset originate from two sources: (i) the Global Historical Climatology Network (GHCN; Menne et al. 2012) and (ii) the U.S. Natural Resources Conservation Service (NRCS) Snowpack Telemetry (SNOTEL) observation network. Observations from Canada and Mexico, included in the GHCN database, are included in the Newman dataset to account for continuity across international borders. After quality control and processing of the data sources, a total of 12,153 stations were used in the precipitation ensemble generation; a total of 8,953 stations were used in the temperature ensemble generation. For more information on the data processing, see the appendix of Newman et al. (2015).

Spatial interpolation

The following four steps describe the process in generating gridded estimates of precipitation. Steps 1, 2, and 3 generate mean estimates of probability of precipitation and precipitation amounts. Step 4 is used to quantify the uncertainty in the mean estimates, which can be thought of as the standard error that is generally reported with regression model estimates. Although the steps are written with respect to precipitation, the general approaches outlined in steps 1, 3, and 4 were applied to temperature observations to generate gridded estimates of mean temperature and diurnal range (i.e., maximum minus minimum temperatures), along with

uncertainty. Steps 1-4 do not generate the ensemble members. Rather, steps 1-4 provide an outline of the regression models used to produce mean estimates of probability of precipitation, precipitation amounts, mean temperature, and diurnal range, which are subsequently used to generate the ensemble members.

1. Distance-dependent weights were computed at each grid cell for all stations in a given radius:

$$w_{i_{sta},i_{sta}} = \left[1 - \left(\frac{d_{i_{sta}}}{MAXD}\right)^3\right]^3 \quad (1)$$

These weights are used to populate the diagonal elements of the weight matrix used in local regression equations defined in Equations (3) and (6); $d_{i_{sta}}$ is the distance of the current station to the grid point being considered; and $MAXD$ is a specified maximum distance. Stations with distances greater than $MAXD$ receive zero weight. Only $n_{sta} = 30$ stations were considered for each grid cell. If all 30 stations were within a distance of 100 km, then $MAXD = 100$ km. If there were fewer than 30 stations within 100 km, then $MAXD$ was set to 1 km larger than the distance of the 30th station to the grid cell. Roughly 90% of the grid points have 30 stations within a 100 km radius. Therefore, each grid cell has a positive definite (i.e., no zero-weighted stations) weight matrix with dimensions (30 x 30).

2. Multivariate, locally weighted logistic regression was used to estimate the probability of precipitation (PoP) at each grid point using a vector of precipitation occurrence (i.e., yes/no or 1/0) from the 30 nearest stations, the weight matrix, and spatial attributes (i.e., latitude, longitude, and elevation).

$$PoP_{i_{grd}} = \frac{1}{1+\exp\left(-\boldsymbol{Z}_{i_{grd}}\boldsymbol{\beta}\right)} \quad (2)$$

In Equation (2), $\boldsymbol{Z}_{i_{grd}}$ is a row vector of spatial attributes for the target grid cell and $\boldsymbol{\beta}$ is a column vector of regression coefficients determined iteratively following Loader (1999).

$$\boldsymbol{\beta}_{new} = \boldsymbol{\beta}_{old} + (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{V}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{P}_{occur} - \boldsymbol{\pi}) \quad (3)$$

In Equation (3), $\boldsymbol{X}$ is the design matrix of spatial attributes from the 30 nearest stations, $\boldsymbol{P}_{occur}$ is the vector of precipitation occurrence (i.e., 1 = occurrence, 0 = no occurrence), $\boldsymbol{V}$ is the 30x30 variance matrix, and $\boldsymbol{\pi}$ is the vector of estimated PoP at each station. $\boldsymbol{\beta}$ is initialized as a column vector of ones.

3. Multivariate, locally weighted linear regression was used to estimate transformed precipitation (PCP) at each grid point using transformed non-zero precipitation observations ($\boldsymbol{P}'sta$), the weight matrix, and spatial attributes. The transformation used was a simple power-law, as follows.

$$\boldsymbol{P}'_{sta_{i_{grd}}} = \left(\boldsymbol{P}_{sta_{i_{grd}}}\right)^{1/Tran} \quad (4)$$

In Equation (4), $\boldsymbol{P}_{sta_{i_{grd}}}$ is the vector of non-zero precipitation observations, $\boldsymbol{P}'_{sta_{i_{grd}}}$ is the vector of transformed non-zero precipitation observations, and $Tran$ is the power-law

transformation coefficient, which was set to 4 after sensitivity testing. Then, PCP at a grid cell can be estimated using the following equation.

$$PCP_{i_{grd}} = \boldsymbol{Z}_{i_{grd}}\boldsymbol{\beta}^{a} \tag{5}$$

In Equation (5), $\boldsymbol{\beta}^{a}$ is the vector of regression coefficients estimated using the following method.

$$\boldsymbol{\beta}^{a} = (\boldsymbol{X}^{T}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^{T}\boldsymbol{W}\boldsymbol{P}'_{sta} \tag{6}$$

In Equation (6), $\boldsymbol{X}$ is the design matrix of station attributes, $\boldsymbol{W}$ is the weight matrix, and $\boldsymbol{P}'_{sta}$ is the transformed non-zero precipitation observations.

4. Uncertainty estimates (E) for the precipitation amounts were determined at each grid point by summing the regression residuals, which effectively combines spatial representativeness and individual station errors. The uncertainty estimate for a grid cell can be thought of as the reported standard error from a traditional regression model. The following equation provides the basis for calculating the uncertainty estimate for a grid cell, based on the weighted sum of the regression residuals from the 30 nearest stations to the grid cell of interest.

$$E_{i_{grd}} = \left[\frac{\sum_{i_{sta}=1}^{n_{sta}} w_{i_{sta},i_{sta}}\left(PCP_{i_{sta}} - \boldsymbol{P}'_{sta_{i_{sta}}}\right)}{\sum_{i_{sta}=1}^{n_{sta}} w_{i_{sta},i_{sta}}}\right]^{1/2} \tag{7}$$

Ensemble generation

As in Clark and Slater (2006), ensemble members were generated using spatially correlated random fields (SCRFs) sampled from the standard normal distribution. Temporal correlation (autocorrelation) is accounted for by using a conditional simulation approach (Johnson 1987), which samples new numbers that are correlated with the previous day's simulated values. Spatial correlation of the random numbers is governed by an exponential correlation structure, computed from the station data themselves.

Spatially correlated random fields

The exponential correlation structure used for the SCRFs is defined as follows.

$$c_{ij} = c_o \exp\left(-\frac{d_{ij}}{C_{len}}\right) \tag{8}$$

In Equation (8), c_{ij} is the correlation of the current grid point, c_o is the initial correlation value (estimated from station data), C_{len} is the spatial correlation length (estimated from station data), and d_{ij} is the distance (km) between the i^{th} and j^{th} grid cells. Consistent with Clark and Slater (2006), a nested grid approach was used to increase the computational efficiency of the Johnson (1987) conditional simulation method.

Spatial correlation lengths (C_{len}) are computed separately for each season, namely December-February (DJF), March-May (MAM), June-August (JJA), and September-November (SON). Temporal autocorrelation is included in the SCRFs for temperature with a 1-day lag, as follows.

$$SCRF_t = SCRF_{t-1}(\rho_{-1}) + \sqrt{1-\rho_{-1}^2}(SCRF) \tag{9}$$

In Equation (9), t is the current time step and ρ_{-1} is the lag-1 temperature autocorrelation. The precipitation SCRF is conditioned using the same approach as in Equation (9), where ρ is set to the observed cross correlation between precipitation and diurnal range; additionally, the T_{range} SCRF for time t is used in place of $SCRF_{t-1}$. These modifications to the precipitation SCRF will ensure that days with large diurnal range are less likely to have precipitation, an assumption based on the fact that non-precipitation days tend to be days with less cloud cover. The authors of Newman et al. (2015) admit this is not necessarily always true (for all seasons and all precipitation events); they will consider this for future research.

Precipitation

Precipitation will not necessarily occur at a grid cell for all ensemble members. Ensembles of precipitation occurrence were generated using the regression PoP and the cumulative probability (CP) of the precipitation SCRF. If the CP for a grid cell was greater than the PoP for the same grid cell, then precipitation occurred at that grid cell. At grid cells where precipitation occurred, precipitation amount was computed by rescaling the CP using the following formula, where CS is the rescaled CP of the precipitation SCRF.

$$CS_{i_{grd}} = \frac{CP_{i_{grd}} - \left(1 - PoP_{i_{grd}}\right)}{PoP_{i_{grd}}} \tag{10}$$

The standard normal deviate of $CS_{i_{grd}}$ is computed and used to perturb the precipitation amount at the grid cell, using the following formula.

$$Prcp'_{ens,i_{grd}} = PCP_{i_{grd}} + RN_{P,i_{grd}}\left(E_{P,i_{grd}}\right) \tag{11}$$

In Equation (11), $Prcp'_{ens}$ is the power-law transformed precipitation amount for the ensemble member at each grid cell; $RN_{P,i_{grd}}$ is the standard normal deviate of the precipitation SCRF; PCP is the regression estimated transformed precipitation amount from step 3, above; and E is the grid cell uncertainty estimate from step 4, and Equation (7). The actual precipitation amount for the ensemble member and grid cell is given by the following equation. Consistently, $Tran = 4$.

$$Prcp_{ens,i_{grd}} = \left(Prcp'_{ens,i_{grd}}\right)^{Tran} \tag{12}$$

Temperatures

Ensemble members for mean temperature and diurnal range were generated by perturbing the regression based estimate by the respective error terms and SCRFs. Note that there is no power-law transformation nor rescaling of the SCRF, as was the case for the precipitation ensemble generation. The following equation indicates that the regression based temperature estimate is simply perturbed by the SCRF, which is scaled by the uncertainty estimate, to produce the temperature estimate for a single ensemble member.

$$T_{ens} = T + RN_{T,i_{grd}}\left(E_{T,i_{grd}}\right) \tag{13}$$

In Equation (13), T_{ens} is either T_{mean} (mean temperature) or T_{range} (diurnal range) for the ensemble member at a grid point; T is the regression based estimate of either T_{mean} or T_{range};

$RN_{T,i_{grd}}$ is the standard normal deviate for the SCRF of either T_{mean} or T_{range}; and $E_{T,i_{grd}}$ is the uncertainty estimate for either T_{mean} or T_{range}.

APPENDIX B

R SCRIPT FOR SELF-ORGANIZING MAP

```
#Be aware that all scripts are run at your own risk and while every script has been
written
#with the intention of minimizing the potential for unintended consequences, the
contributors
#cannot be held responsible for any misuse or script problems.

rm(list = ls())

options(warn=-1)
library(data.table)
library(plyr)
library(dplyr)
library(ggplot2)
library(lubridate)
library(Matrix)
library(ggthemes)
library(reshape)
library(Hmisc)
library(xlsx)
library(zoo)
library(RColorBrewer)
library(grDevices)
library(sp)
library(rgdal)
library(som)

mnine = function(charr) substr(charr,1,nchar(charr)-9)
read_csv_filename <- function(filename){
  ret <- read.csv(filename)
  ret$mflagTMAX=ret$qflagTMAX=ret$sflagTMAX=NULL
  ret$mflagTMIN=ret$qflagTMIN=ret$sflagTMIN=NULL
  ret
}

files = list.files(path=datadir,pattern="\\_multivar.csv")
print(Sys.time())
df <- data.table(ldply(paste(datadir,files,sep=""), read_csv_filename))
print(Sys.time())

df$longdate = as.character(droplevels(df$datetime));
df$longdate = as.character(mnine(df$longdate))
df$datetime = NULL
df$date = as.numeric(gsub("[: -]", "" , df$longdate, perl=TRUE))
df$longdate = NULL
df$datetime = ymd(df$date)
df$date = NULL
df$month = month(df$datetime)
df$valuePRCP = df$valuePRCP/254
df$value = df$mflag = df$qflag = df$sflag = NULL
df$year = year(df$datetime)

dfsumm = dplyr::summarise(group_by(df,id),mind=min(datetime),maxd=max(datetime),
                          npr=length(which(is.na(valuePRCP)==F)),
                          ndposs=as.numeric(max(datetime)-min(datetime)))
dfsumm$pavail = round(dfsumm$npr*100/dfsumm$ndposs,digits=2)
```

```
tenyr = 365*10
iwant = subset(dfsumm,subset = pavail > 85.0 & ndposs > tenyr)
subdf = subset(df,id %in% iwant$id)

maxsm = dplyr::summarise(group_by(subdf,id,year),max=max(valuePRCP,na.rm=T))
maxsm$max[maxsm$max < 0] = 0.1
mmaxsm = dplyr::summarise(group_by(maxsm,id),mmax=mean(max))

asumm = dplyr::summarise(group_by(subdf,id,year),tpr=sum(valuePRCP,na.rm=T))
aasumm = dplyr::summarise(group_by(asumm,id),annpr=mean(tpr,na.rm=T))

mtsumm = dplyr::summarise(group_by(subdf,id,month,year),mtpr = sum(valuePRCP))

msumm = dplyr::summarise(group_by(mtsumm,id,month),monpr=mean(mtpr,na.rm=T))
msumm$mlab = month.abb[msumm$month]
msumm = msumm[order(msumm$id,msumm$month),]
msumm$month = NULL

msumm=reshape(msumm,idvar="id",timevar="mlab",direction="wide")
head(msumm)

tog = merge(aasumm,mmaxsm,by="id")
head(tog)

pdt = read.table(file="GHCN_prcp_TRVW.txt",sep=",",comment.char = "%")
colnames(pdt)=c('FID','ID','Name','Lat','Lon','Elev')
cdt = pdt[,c(4,5)]

coordinates(cdt) <- c("Lon", "Lat")
proj4string(cdt) <- CRS("+proj=longlat +datum=WGS84")
res <- spTransform(cdt, CRS("+proj=utm +ellps=WGS84"))
res = as(res, "SpatialPoints")

pdt$UDT_x=res@coords[,1]
pdt$UDT_y=res@coords[,2]

maxp1 = read.csv("GHCN_prcp_TRVW_1daymean_anntot.csv",header=T)
maxp2 = read.csv("GHCN_prcp_TRVW_2daymean.csv",header=T)
maxp3 = read.csv("GHCN_prcp_TRVW_3daymean.csv",header=T)

maxp = merge(maxp1,maxp2,by="ID")
maxp = merge(maxp, maxp3,by="ID")

colnames(maxp) = c("ID","anntot","max.1day","max.2day","max.3day")

kk=1
print(kk)
rr = merge(tog,pdt,by.x="id",by.y="ID")
rr = merge(rr, maxp, by.x="id",by.y="ID")
rr.copy = rr
rr$id = rr$annpr = rr$mmax = rr$FID = rr$Name = rr$Lat = rr$Lon = NULL

if(kk==1){
  rr$max.2day = rr$max.3day = NULL
  rr$max.1day = rr$max.1day/max(rr$max.1day)
}
if(kk==2){
  rr$max.1day = rr$max.3day = NULL
  rr$max.2day = rr$max.2day/max(rr$max.2day)
}
if(kk==3){
  rr$max.1day = rr$max.2day = NULL
  rr$max.3day = rr$max.3day/max(rr$max.3day)
```

```
}
if(kk==123){
  rr$max.1day = rr$max.1day/max(rr$max.1day)
  rr$max.2day = rr$max.2day/max(rr$max.2day)
  rr$max.3day = rr$max.3day/max(rr$max.3day)
}

rr$Elev[rr$Elev < 0] = 5

rr$Elev = rr$Elev/max(rr$Elev,na.rm=T)
rr$UDT_x = (rr$UDT_x-min(rr$UDT_x))/(max(rr$UDT_x)-min(rr$UDT_x))
rr$UDT_y = (rr$UDT_y-min(rr$UDT_y))/(max(rr$UDT_y)-min(rr$UDT_y))
rr$anntot = rr$anntot/max(rr$anntot)

xd = 5
yd = 3
som.init(rr,xdim=xd,ydim=yd,init="linear")
sout =
som(rr,xdim=xd,ydim=yd,init="linear",neigh="gaussian",topol="rect",rlen=2000000)
sout =
som.train(rr,sout$code,xdim=xd,ydim=yd,neigh="gaussian",topol="rect",rlen=2000000)
som.update(sout)

colz = colorRampPalette(c('white','firebrick3','firebrick4'))(24);
xx=as.data.frame(sout$code.sum)

bellz = c("#E69F00", "#56B4E9", "#009E73",
          "#F0E442", "#0072B2", "#D55E00",
          "darkmagenta","firebrick3",'dimgray',
          'navy','orangered','blueviolet',
          'darkgoldenrod2','aquamarine3','deeppink2')
kolors <- colorRampPalette(c("white","red3"))(12)

ggplot(xx,aes(x=x,y=y,fill=factor(nobs)))+
  geom_tile()+
  geom_text(aes(label=nobs),size=5.5)+
  scale_fill_manual(name="Frequency",values=kolors)+
  xlab("")+ylab("")+
  theme(
    panel.background = element_rect(fill="white"),
    panel.grid.major=element_line(colour = "grey"),
    axis.text=element_text(size=14),
    legend.text = element_text(size=14),
    legend.title = element_text(size=14)
  )+
  scale_x_continuous(breaks=(seq(1,xd))-1,labels=seq(1,xd))+
  scale_y_continuous(breaks=(seq(1,yd))-1,labels=seq(1,yd))

xx$sncat = NA
xx$sncat[xx$nobs>0] = 1:length(xx$nobs > 0)

myColors <-
c(brewer.pal(9,"Set2"),brewer.pal(9,"Dark2"),brewer.pal(9,"Set3"))[1:(max(xx$sncat,na.
rm=T))]
colScale <- scale_colour_manual(name = "Group",values = myColors)

ggplot(xx,aes(x=x,y=y,fill=factor(sncat)))+
  geom_tile()+
  geom_text(aes(label=nobs),size=5.5)+
  xlab("")+ylab("")+
  theme(
    panel.background = element_rect(fill="white"),
    panel.grid.major=element_line(colour = "grey"),
```

```
    axis.text=element_text(size=14),
    legend.text = element_text(size=14),
    legend.title = element_text(size=14)
    )+
  scale_fill_manual(name="Group",values=myColors)+
  scale_x_continuous(breaks=seq(1,xd)-1,labels=seq(1,xd))+
  scale_y_continuous(breaks=seq(1,yd)-1,labels=seq(1,yd))

spat=as.data.frame(cbind(rr.copy$Lon,rr.copy$Lat,sout$visual$x,sout$visual$y,
                         sout$visual$qerror))
colnames(spat) = c('UDT_x','UDT_y','x','y','qerror')
ugh = merge(spat,xx,by=c('x','y'))
colnames(ugh)=c('x','y','Lon','Lat','qerror','nobs','sncat')
outt = merge(ugh,rr.copy,by=c('Lat','Lon'))
mout = subset(outt,select = c('id','Name','Lat','Lon','Elev','sncat'))

tvw = readOGR(dsn="Shapefiles",
              layer="TennesseeRiverWatershedproj")
tvw@data$OBJECTID = rownames(tvw@data)
tvw.points = fortify(tvw)
head(tvw.points)

Sys.setenv(NOAWT=1)
library(OpenStreetMap)
library(rgdal)
library(stringr)
library(ggplot2)
library('grDevices')

ewbrks <- seq(-90,-80,2)
nsbrks <- seq(32,40,1)
ewlbls <- unlist(lapply(ewbrks, function(x) ifelse(x < 0, paste(x, "°E",sep=""),
ifelse(x > 0, paste(x, "°W",sep=""),x))))
nslbls <- unlist(lapply(nsbrks, function(x) ifelse(x < 0, paste(x, "°S",sep=""),
ifelse(x > 0, paste(x, "°N",sep=""),x))))

nm = 'stamen-terrain'
map <- openmap(upperLeft=c(37.6,-89.),lowerRight=c(33.8,-81),type=nm)
map2 <- openproj(map)
autoplot(map2)+
  geom_polygon(data=tvw.points,aes(long,lat),colour="black",fill="NA")+
  geom_point(data=ugh,aes(x=Lon,y=Lat,colour=factor(sncat)),
             size=3.5,shape=17)+colScale+
  scale_x_continuous(breaks = ewbrks, labels = ewlbls, expand = c(0, 0)) +
  scale_y_continuous(breaks = nsbrks, labels = nslbls, expand = c(0, 0)) +
  xlab("")+ylab("")+
  theme(
    panel.background = element_rect(fill="white"),
    panel.grid.major=element_line(colour = "grey"),
    panel.grid.minor=element_line(colour="grey"),
    axis.text=element_text(size=14),
    legend.title=element_text(size=15),
    legend.text = element_text(size=14)
  )

map3 <- openproj(map)
autoplot(map3)+
  geom_polygon(data=tvw.points,aes(long,lat),colour="black",fill="NA")+
  geom_point(data=ugh,aes(x=Lon,y=Lat),color='black',size=3.,shape=18)+colScale+
  scale_x_continuous(breaks = ewbrks, labels = ewlbls, expand = c(0, 0)) +
  scale_y_continuous(breaks = nsbrks, labels = nslbls, expand = c(0, 0)) +
  xlab("")+ylab("")+
  theme(
```

```
    panel.background = element_rect(fill="white"),
    panel.grid.major=element_line(colour = "grey"),
    panel.grid.minor=element_line(colour="grey"),
    axis.text=element_text(size=14)
  )

kcols <- brewer.pal(14,"YlGnBu")
kcolScale <- scale_colour_manual(name = "Group",values = kcols)

cc=colorRampPalette(c("yellow3","darkorange1","red3"))(15)

map3 <- openproj(map)
autoplot(map3)+
  geom_polygon(data=tvw.points,aes(long,lat),colour="black",fill="NA")+
  geom_point(data=rr.copy,aes(Lon,Lat,colour=rr.copy$annpr),shape=15,size=3)+
  scale_colour_gradientn(colours=cc,
                         guide=guide_colourbar(title="",
                                               barwidth=2,barheight=12,nbin=15,
                                               label.theme =
element_text(size=16,angle=0)))+
  scale_x_continuous(breaks = ewbrks, labels = ewlbls, expand = c(0, 0)) +
  scale_y_continuous(breaks = nsbrks, labels = nslbls, expand = c(0, 0)) +
  xlab("")+ylab("")+
  theme(
    panel.background = element_rect(fill="white"),
    panel.grid.major=element_line(colour = "grey"),
    panel.grid.minor=element_line(colour="grey"),
    axis.text=element_text(size=14)
  )
```

NRC FORM 335
(12-2010)
NRCMD 3.7

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

1. REPORT NUMBER
(Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)
NUREG/CR-7247

2. TITLE AND SUBTITLE
Research to Develop Guidance on Extreme Precipitation Frequency Estimates in Orographic Regions

3. DATE REPORT PUBLISHED

MONTH	YEAR
March	2026

4. FIN OR GRANT NUMBER
NRC-HQ-60-11-I-0006

5. AUTHOR(S)
Kathleen. D. Holman
Aandrew. P. Verdin
David P. Keeney

6. TYPE OF REPORT
Technical

7. PERIOD COVERED (Inclusive Dates)
09/2012-01/2017

8. PERFORMING ORGANIZATION - NAME AND ADDRESS (If NRC, provide Division, Office or Region, U. S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)
U.S. Department of the Interior, Bureau of Reclamation, Technical Service Center
Denver Federal Center
PO Box 25007
Denver, CO 80225-0007

9. SPONSORING ORGANIZATION - NAME AND ADDRESS (If NRC, type "Same as above", if contractor, provide NRC Division, Office or Region, U. S. Nuclear Regulatory Commission, and mailing address.)
Division of Risk Analysis
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission

10. SUPPLEMENTARY NOTES

11. ABSTRACT (200 words or less)
Currently, many engineering design projects use Probable Maximum Precipitation (PMP) to develop Probable Maximum Floods (PMFs), which are then used to evaluate the safety of a proposed or existing facility. Orographic methods employed in hydrometeorological reports (HMRs) used to estimate PMP vary widely through time and by region, if they were employed at all. Deterministic metrics, such as PMP, do not provide decision makers with information on precipitation events less intense than PMP, yet still extreme, nor do they provide information on the expected frequency of such events. The Nuclear Regulatory Commission (NRC) requested assistance from the Department of Interior's Bureau of Reclamation (Reclamation) in developing guidance on regional extreme precipitation analyses in orographic regions. The methods presented in this report are illustrated across the Tennessee River Valley watershed. The Tennessee River Valley is a region with pronounced orographics, and thus, serves as a suitable testbed for the methodologies outlined in this report.

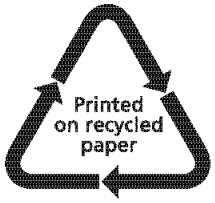
12. KEY WORDS/DESCRIPTORS (List words or phrases that will assist researchers in locating the report.)
extreme precipitation, precipitation frequency, orographic, orographic region

13. AVAILABILITY STATEMENT
unlimited

14. SECURITY CLASSIFICATION
(This Page)
unclassified
(This Report)
unclassified

15. NUMBER OF PAGES
158

16. PRICE
$0.00



Federal Recycling Program

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS

NUREG/CR-7247

Research to Develop Guidance on Extreme Precipitation Frequency Estimates in Orographic Regions

March 2026