U.S.NRC

United States Nuclear Regulatory Commission

Protecting People and the Environment

Convection-Permitting Modeling for Intense Precipitation Processes

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
   Digital Communications and Administrative
    Services 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: www.ansi.org
(212) 642-4900

U.S.NRC

United States Nuclear Regulatory Commission

Protecting People and the Environment

NUREG/CR-7290

Convection-Permitting Modeling for Intense Precipitation Processes

Manuscript Completed: September 2021
Date Published: May 2023

Prepared by:
Andreas F. Prein
Jordan G. Powers
Erin L. Towler
David Ahijevych
Ryan A. Sobash
Craig S. Schwartz

National Center of Atmospheric Research (NCAR)
3090 Center Green Dr.
Boulder, CO 80301

Elena Yegorova, NRC Project Manager

Office of Nuclear Regulatory Research

ABSTRACT

This report documents work sponsored by the U.S. Nuclear Regulatory Commission (NRC) and conducted by the National Center for Atmospheric Research (NCAR) as part of the RES project, "Convection-Permitting Modeling for Intense Precipitation Processes." This project was undertaken as part of the Probabilistic Flood Hazard Assessment (PFHA) Research Program. The objective of the PFHA Research Program is to develop tools and guidance on the use of PFHA methods to risk-inform NRC's licensing of new facilities as well as the licensing and oversight of currently operating facilities as they relate to flooding hazards.

Many flooding scenarios of interest to nuclear power plant (NPP) licensing and oversight involve extreme precipitation events occurring at the plant site or within the watershed of the plant. Generating probabilistic assessments of extreme precipitation within a catchment is challenging due to typically short observational records, insufficient data coverage, and climatic variation. Furthermore, traditionally used estimators of extreme precipitation, such as Probable Maximum Precipitation (PMP), do not allow for the quantification of uncertainties in hazard estimates in either a physical or a risk sense. The application of numerical atmospheric models to the problem offers a way forward. State-of-the-art convection-permitting models (CPMs) can explicitly simulate deep convection and can accurately represent orography on fine scales, and thus they present powerful tools for investigating extreme precipitation events. Using convection-permitting models is, therefore, a valuable alternative approach in the provision of more physically-based and probabilistic flood risk assessments.

This report includes a thorough literature review and analysis that summarize the state of the science in simulating extreme precipitation events with convection-permitting models while outlining key challenges, opportunities, and promising research areas. Based on the literature review, we assess the ability of CPMs to capture extreme precipitation in recent flood events in the contiguous U.S. (CONUS) east of the Rocky Mountains by leveraging three existing convection-permitting ensemble datasets. These cover 10,570 36-hour simulations/forecasts at 3-km horizontal grid spacing (Δx=3 km) and 810 36-hour simulations at Δx=1 km spacing. Additionally, we analyze the impact of observational uncertainties on the results by using a selection of high-quality multi-sensor and gauge-based rainfall datasets.

The central finding is that numerical weather prediction models configured with convection-permitting resolutions can capture heavy precipitation events in the Eastern U.S. Many aspects of the simulations are verified by multi-sensor observational datasets, and the precipitation output from the convection-permitting (CP) model configurations shows less error than precipitation estimates based on station data. This demonstrates the potential value of incorporating CP model outputs in flood risk assessments. However, CP model configurations are not perfect and sometimes show systematic biases, as revealed in case examples of the underestimation of event peak accumulations by up to 30%. While one way to reduce these biases is by statistical post-processing, over the long-term model development to improve fidelity is the preferred way to address the issue.

The report closes with two recommendations for future work on the usage of CPM output for probabilistic flood risk assessments. First, targeted downscaling of heavy precipitation events in global climate models with CPMs would allow the building of a catalog of heavy precipitation events that are physically plausible, but unprecedented in the observational record. Second, events from this catalog and existing CPM heavy precipitation simulations could be used in

combination with statistical approaches, such as stochastic storm transposition, to generate a large set of plausible heavy rainfall events to generate input for hydrologic models (e.g., WRF-Hydro).

FOREWORD

This research is part of the NRC's PFHA Research Program and has been conducted to assist NRC in assessing potential pathways for improving flood risk estimates at current and future nuclear power plant (NPP) sites.

This report contains a review and novel research on the applicability of convection-permitting models (CPMs) in probabilistic flood hazard assessments. CPMs are configurations of weather and climate models at high resolution that allow for the explicit simulation of deep convective storms (i.e., thunderstorms) and for the more detailed representation of land surface features (e.g., orography and coastlines) compared to coarser-resolution setups. Over the past decade, the increasing usage of CPMs in weather forecasting and climate modeling has led to substantial improvements in the predictability and projections of heavy precipitation extremes.

The main objective of this project was to develop a framework to allow the incorporation of CPM simulations into probabilistic flood hazard assessments. The focus of the analysis is on the contiguous United States (CONUS) east of the Continental Divide. Heavy precipitation in this region tends to result from three forcings: tropical cyclones, mesoscale convective systems, and orographic enhancement. The presented results show that CPMs can capture observed heavy precipitation events with high accuracy, while some observed biases can be mitigated by statistical postprocessing methods (e.g., bias correction) and by model improvement. A framework for incorporating CPM output into flood hazard assessments is also presented and future research priorities are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Flooding is among the costliest and most widespread natural disaster worldwide, and economic losses from it are rising. This increase is driven by a combination of larger economic exposure, reflecting population and infrastructure growth, and increasing precipitation intensities, caused by climate variability and change. Designing and maintaining hospitals, dams, and power plants to withstand rare flood events is of high importance for society.

Currently, flood risk estimates are typically based on fitting extreme value distributions to historical precipitation records. This approach is challenging because of limitations in our observational records. First, even the longest available observational records (e.g., 100 years) are too short to estimate reliably the magnitude of rare flood events that a power plant must withstand. Second, our long-term rain gauge networks are not able to capture localized peak precipitation rates, which tends to result in an underestimation of extreme flood magnitudes. Alternatively, probable maximum precipitation (PMP) estimates are also used to assess the worst-case flood event in a region. The PMP is defined as the largest precipitation accumulation meteorologically possible for a given duration and location. A downside of PMP estimates is that they are developed from a small catalog of historical storms using deterministic heuristic techniques, and therefore they do not support probabilistic flood risk assessments.

Historically, using numerical weather models for flood risk assessment has had limitations due to the coarse resolution of the simulated rainfall fields and deficiencies in realistically simulating heavy precipitation events. Increased computer power in combination with model developments, now enables us to run regional numerical weather prediction (NWP) models and regional climate models (RCMs) at kilometer scales, improving the fidelity of simulations of heavy precipitating storms. For long simulations over large regions, though, the computational cost of convection-permitting climate modeling remains an obstacle, which can make certain uses of models for simulating rare events challenging.

Here we take advantage of three existing convection-permitting (CP) numerical weather model forecast datasets containing 10,570 simulations at 3-km horizontal grid spacing and 810 simulations at 1-km spacing (1.7 mi and 0.6 mi respectively) over regions covering the contiguous United States (CONUS) east of the continental divide. These datasets contain several recent high-impact precipitation events such as the West Virginia flooding of 2016, Hurricane Harvey (2017), and Hurricane Matthew (2016). For analysis purposes we split the CONUS into four sub-regions— southern U.S., central U.S., Appalachian Mountains, and U.S. East Coast— to account for differences in flood-producing mechanisms. In addition to the kilometer-scale forecasts, we also apply several gridded precipitation datasets, which enables us to assess the impact of observational uncertainties on forecast evaluation.

We find that numerical weather prediction models configured with convection-permitting resolutions can capture heavy precipitation events in the Eastern U.S. Many characteristics of the simulations are verified with multi-sensor observational datasets that include radar data, and the precipitation output from CP models for these types of events frequently shows less error than precipitation estimates based on station data alone. This demonstrates the potential value in incorporating CP model outputs in flood risk assessments, given that current flood standards are based on station records.

CP models can also provide information on changes in flood risks based on global warming projections. However, CP models are not perfect, and they can have systematic biases, such as underestimations of event peak accumulations (e.g., up to 30 %, as shown here). To deal with such biases, we recommend the application of statistical postprocessing, as an initial corrective, and model improvement, in the long term, before heavily relying on model-simulated precipitation for intense event analysis and flood risk assessment.

The report closes with two recommendations for future work focusing on the usage of CPMs output for probabilistic flood risk assessment. First, targeted downscaling of heavy precipitation events in global climate models with CPMs would allow the building of a catalog of heavy precipitation events that are physically plausible, but unprecedented in the observational record. Second, events from this catalog and existing CPM heavy precipitation simulations should be used in combination with statistical approaches, such as stochastic storm transposition, to generate a large set of plausible heavy rainfall events to generate input for hydrologic models (e.g., WRF-Hydro).

ACKNOWLEDGMENTS

ABBREVIATIONS AND ACRONYMS

AEP   Annual Exceedance Probability
ANOVA   Analysis Of Variance
AR   Atmospheric River
AR-WRF   Advanced Research Weather Research And Forecasting Model
ARF   Areal Reduction Factors

CAPE   Convective Available Potential Energy
CIN   Convective Inhibition
CMIP   Coupled Model Intercomparison Project Phase
CONUS   Contiguous United States
CP   Convection-Permitting
CPM   Convection-Permitting Model

DEA   Design Event Approach

ECMWF   European Centre For Medium-Range Weather Forecasts
En   Ensemble Spread
Enkf   Ensemble Kalman Filter
ERA   European Centre For Medium-Range Weather Forecasts Re-Analysis
ETS   Equitable Threat Score

FFA   Flood Frequency Analysis
FSS   Fractions Skill Score

GCM   Global Climate Model
GFS   Global Forecasting System
GMET   Gridded Meteorological Ensemble Tool
GSS   Gilbert Skill Score

Highresmip   High-Resolution Model Intercomparison Project
HPC   High-Performance Computer

IP   Intense Precipitation
IQR   Interquartile Range
IVT   Integrated Vapor Transport

LE   Large-Eddy
LENS   NCAR's Community Earth System Model Large Ensemble

LSM	Land Surface Model
MC	Monte Carlo
MCS	Mesoscale Convective System
MPEX	Mesoscale Predictability Experiment
MRMS	Multi-Radar/Multi-Sensor
MSE	Mean Squared Error
MSWEP	Multi-Source Weighted-Ensemble Precipitation
MYJ	Mellor–Yamada–Janjić
NARCCAP	North American Regional Climate Change Assessment Program
NCAR	National Center For Atmospheric Research
NCEP	National Centers For Environmental Prediction
NEXRAD	Next Generation Weather Radar
NGWOS	USGS Next Generation Water Observing System
NOAA	National Oceanic And Atmospheric Administration
Nr	Case To Case Variability
NWP	Numerical Weather Prediction
Ob	Observational Uncertainty
PBL	Planetary Boundary Layer
PDF	Probability Density Function
PGW	Pseudo Global Warming
PMP	Probable Maximum Precipitation
PW	Precipitable Water
QPE	Quantitative Precipitation Estimate
RCM	Regional Climate Model
RCP	Representative Concentration Pathway
RFA	Rainfall-Based Flood Frequency Analysis
RH	Relative Humidity
RRTMG	Rapid Radiative Transfer Model For Global Climate Models
SCS	Severe Convective Storms
Se	Seasonal Difference
SST	Stochastic Storm Transposition
TC	Tropical Cyclone

UFSS	Uniform Fraction Forecast
U.S.	United States
WMO	World Meteorological Organization
WRF	Weather Research And Forecasting Model
WSR-88D	Weather Surveillance Radar 1988 Doppler
Δx	Horizontal Grid Spacings

1   INTRODUCTION AND STATE OF KNOWLEDGE

Intense precipitation lies at the root of some of the costliest natural disasters, and, after heatwaves, precipitation-driven events are responsible for the most natural-disaster fatalities in the U.S. (https://www.weather.gov/hazstat/). The frequency of natural catastrophes related to intense precipitation has more than doubled since the 1980s (Munich Re 2017). This doubling is related to an increase of population and infrastructure in flood-prone areas and an increase in the intensity and frequency of heavy precipitation events (Groisman et al. 2005, IPCC 2013, Westra et al. 2013). The increase in such events has also been related to climate change (Min et al. 2011, O'Gorman 2015), and the trend is expected to continue over most regions of the world (e.g., Trenberth et al. 2003, Prein et al. 2017a). Climate model simulations suggest that the rarer a precipitation event, the more it intensifies due to climate change (Prein et al. 2017b, Pendergrass 2018). Lopez-Cantu et al. (2019), for example, show that precipitation intensities of the average 100-year return level event in the contiguous United States increase 10 to 50 % more than those of the 1-year event.

Most areas have limited rain gauge density, which, due to the strong spatiotemporal variability in rainfall fields (e.g., Prein and Gobiet 2017c), makes accurately observing intense precipitation events challenging (e.g., Kidd et al. 2017). Furthermore, most rain gauge records cover less than 70 years (e.g., Sun et al. 2018) and are, therefore, too short to capture major precipitation events with long return periods. Weather radar (e.g. NEXRAD) provide high-resolution observations with excellent coverage over most of the contiguous U.S. but have only been available since the early 1990s. These short observational records introduce large uncertainties into the statistical estimation of high accumulations (i.e., for the 100-year return event or larger; Coles et al. 2001). Traditionally, probable maximum precipitation (PMP; WMO 2009) estimation is used to quantify the worst-case scenario for heavy precipitation at a given location, and it assumes that the contributing factors achieve their optimum values. PMP key factors are maximized column water vapor, or precipitable water (PW; e.g., Schreiner and Riedel, 1978), low-level convergence, and upward motion. However, traditional PMP estimates are not truly physically based, given their lack of accounting for: (i) limited observational records, (ii) inaccuracies in observational records, and (iii) the spectrum and amplitude of atmospheric processes that produce large accumulations. As found in examinations of estimation methods presented below, PMP approaches also suffer from assumptions that have been contradicted by the results of model-based studies. A further shortcoming of both PMP and gauge-based intense precipitation assessments is that they rely on incomplete historical records and an assumption of climate stationarity, with the latter now appearing to be incorrect, in a changing climate, in estimating flood risks (e.g., Wright et al. 2019).

Numerical atmospheric models are proving to be a capable alternative to traditional methods for quantitative intense precipitation analysis and estimation. Until the past decade, models were applied at resolutions too coarse (e.g., horizontal grid spacings >12 km; 7.5 miles) to reliably simulate most intense precipitation events, but increasing computer power and improved numerical representations have made convection-permitting (CP) simulations much more feasible. Convection-permitting model configurations are those using horizontal grid spacings (Δx) of less than about 4 km (2.5 miles; see, e.g., Weisman et al. 1997), allowing reasonable direct simulation of convective mass transfer processes. The move to CP grid spacings has had a big impact on the simulation of convectively-forced (e.g., Weisman et al. 1997, Clark et al. 2016, Prein et al. 2019) and orographically-forced (e.g, Colle and Mass 2000, Prein et al. 2013) precipitation. Advancements in model physical process schemes have also contributed to the improvement of intense-storm simulation. Motivated by these developments, many weather

forecasting centers are now using CP configurations operationally (e.g., Smith et al. 2008, Voudouri et al. 2018). Furthermore, global forecasting systems that are nonhydrostatic and can operate efficiently down to the CP scales (e.g., MPAS, Skamarock et al. 2012, Neumann et al. 2019) are expected to be run operationally within a couple of years. Even the climate modeling community is increasingly exploring CP modeling (Prein et al. 2015), and substantial improvements in simulating heavy precipitation events in climate settings have been demonstrated (e.g., Prein et al. 2013, Ban et al. 2014, Chan et al. 2014, Kendon et al. 2014, Prein et al. 2017).

This introduction addresses the state-of-the-science for numerical simulation of heavy precipitation as reflected in the published literature and through novel research. Recent achievements in the weather and climate modeling communities in exploring the problem are highlighted. The introduction covers the spectrum of methodologies that have been used to investigate this space, including real-data v. idealized modeling, case-study v. long-term simulation, real-time forecasting v. retrospective simulation, deterministic v. ensemble modeling, and weather prediction v. climate projection settings. Our approach includes reviewing the work and findings of papers to present more fully to NRC the issues in this topic. Furthermore, we also summarize challenges that limit our ability to simulate and evaluate heavy rainfall, such as those of model physics needs and of observational uncertainties.

Section 1 starts with a discussion of the main atmospheric processes behind intense precipitation (Section 1.1). It then summarizes the types of storms yielding heavy rainfall and their simulation in models (Section 1.2), followed by a review of the challenges and opportunities in heavy precipitation modeling (Section 1.3). We end with conclusions and recommendations (Section 1.4).

1.1 Atmospheric Processes Causing Intense Precipitation and their Representation in Numerical Models

Rain accumulation can be simply summarized per the following analysis (attributed to C. F. Chappell, noted hydrologist): *the heaviest rainfall occurs where the rainfall rate is highest for the longest time* (Doswell et al. 1996). This perhaps obvious definition is universally applicable and encompasses flash floods, which reflect very high rain rates over short periods, as well as river flooding, which typically results from persistent precipitation over days or even months.[1] The definitions or thresholds for "intense" rainfall in the context of flooding depend on the catchment characteristics, local climatology, and the hydrometeorological conditions.

As summarized by Doswell et al. (1996) three main ingredients contribute to rainfall rates: (i) air ascent rates, (ii) atmospheric moisture content, and (iii) precipitation efficiency.

> (i) A high ascent rate (i.e., high vertical velocity) is essential for intense rainfall since it supports high droplet condensation rates. Thus, deep convection can be a key driver of heavy rainfall since it generally presents the strongest and deepest vertical motions in the atmosphere. High ascent rates resulting in heavy precipitation may also be forced by topography.
> (ii) Air moisture content (e.g., as represented by water vapor mixing ratio) is critical in the vertical moisture flux, defined as air moisture content $\times$ ascent rate. With rainfall rates being proportional to this flux, intense rainfall events are typically found with very moist (e.g., tropical) air masses.

---

[1] We do not consider here river flooding that results from snowmelt and its runoff.

(iii) The precipitation efficiency (e.g., Fankhauser 1988) is defined as the ratio of in-cloud vertical moisture flux to the precipitation mass flux. This ratio depends on many factors, including cloud microphysical characteristics (e.g., droplet size distribution and droplet phase– water or ice), cloud dynamics (e.g., entrainment and detrainment rates), and storm environment conditions (e.g., relative humidity, thermodynamic stability, wind shear). Typically, precipitation efficiency is about 30% (Fankhauser 1988), but it can reach up to 50% in upright convective cells (Ferrier et al. 2006).

The rainfall rate is the product of these three terms— ascent rate, moisture content, and precipitation efficiency, and the greatest rainfall rates occur when all of them approach optimum levels. Therefore, simulating intense rainfall in numerical models requires realistic representations of these conditions, conditions that span a wide range of spatiotemporal scales from the microphysical to the synoptic. The remainder of this section focuses on the essential processes that affect these ingredients for intense precipitating storms, and how well models simulate them and their effects.
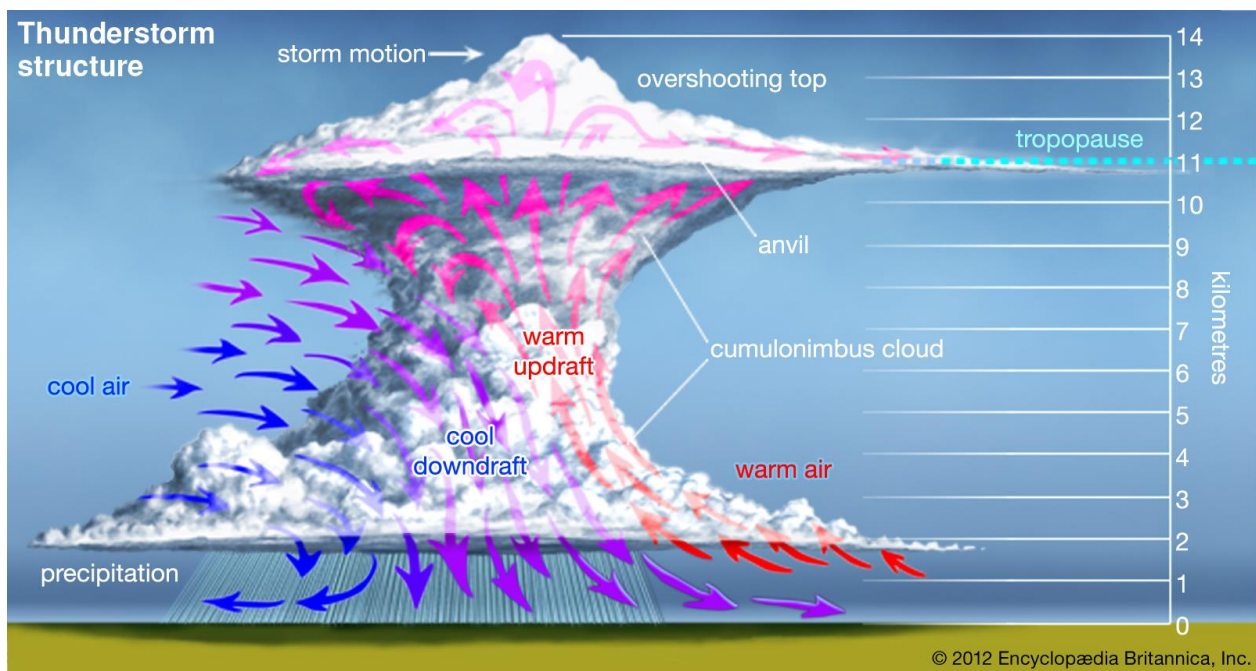
Figure 1-1   Schematic of Idealized Thunderstorm *Warm, moist, and unstable air feeds a convective cell (red arrows). This air ascends, and its vapor condenses, releasing latent and resulting in strong upward motion (updrafts). The condensation leads to raindrops and other hydrometeors (e.g., graupel, hail, and snow) that precipitate, yielding accumulations at the surface (image from Encyclopedia Britannica 2012).*

1.1.1 Deep Convection

Deep convection is thermally driven turbulent motion that moves relatively warm and moist air from the lower to the upper troposphere. As rising air saturates and condenses, latent heating drives buoyant vertical accelerations (updrafts; see Figure 1-1). The condensation process yields cloud droplets (hydrometeors), then larger particles— either frozen as snow or hail or liquid as rain— that can precipitate through the updrafts or downdrafts (downward moving air). Typically, hydrometeor formation rates are higher and precipitation is stronger for greater updraft strengths. Convection may assume the form of organized assemblies of cells (e.g., mesoscale[2] convective systems [MCSs]), deep cells embedded in tropical cyclones, or isolated thunderstorms.

The representation of deep convection in weather and climate models was historically a sub-grid-scale process and thus was parameterized. Generally, convective schemes (also called cumulus schemes or cumulus parameterizations) assume that the bulk effect of a collection of clouds within a grid cell on the larger-scale environment can be approximated by process representations using a set of parameters involved in the interactions (see, e.g., Arakawa and Schubert, 1974). This is conceptually problematic since there is no distinct spatial or process boundary between the particular grid-scale larger-scale environment and the target cloud motions that justifies such a scale separation (e.g., Skamarock et al. 2014). Rather, clouds are the product of dynamic and thermodynamic processes spanning space and time scales.

Over the years, scores of convective parameterization schemes have been developed (e.g., Manabe et al. 1965, Arakawa and Schubert 1974, Tiedtke 1989, Emanuel 1991, Kain and Fritsch 1993, Grell and Freitas 2014). While they have supported and advanced atmospheric modeling, they have also been a major source of uncertainty in weather and climate simulations, which are very sensitive to such schemes (e.g., Déqué et al. 2007, Mooney et al. 2017). Figure 1-2, for example, illustrates biases that can emerge from convective parameterizations, such as overly frequent and too-weak precipitation, with a premature diurnal onset. Other atmospheric process schemes, such as for microphysics, the planetary boundary layer (PBL), and radiation, play a secondary role in simulation uncertainties (Figure 1-2). The parameterization uncertainty collapses, however, when CP resolutions are used, as the CP grid spacing obviates the need for the convective scheme, and they are typically not engaged. In these model configurations, the moist processes produced by the microphysics scheme and model dynamics begin to represent deep convection explicitly (Figure 1-2(d)-(f)).

---

[2] Mesoscale refers to horizontal scales of a few to several hundred km.
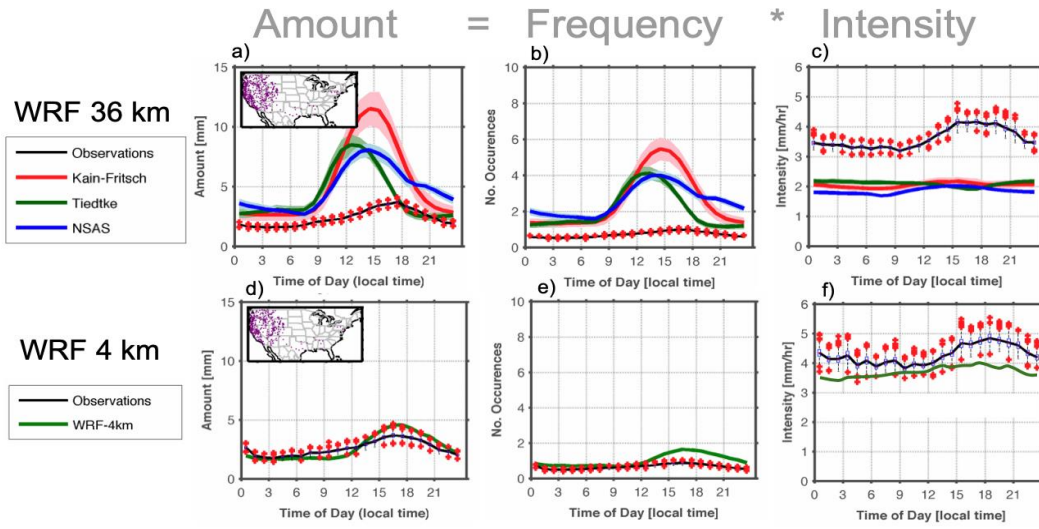
Figure 1-2   Climatological Average June, July, and August Diurnal Cycles of Precipitation Amount, Frequency, and Intensity of Hourly Precipitation in the Western U.S. *Purple dots in panels a) and d) show the locations of the rain gauges used. Shown are observed (black lines) and modeled (colored lines) hourly precipitation. Panels a–c) show modeled precipitation from a perturbed-physics Weather Research and Forecasting (WRF) model ensemble using a 36-km (22 mi) grid for the period 1991–2000. Solid, colored curves show results produced by WRF simulations using three different deep convection schemes, while the shading (spread) about the solid, colored lines represents the uncertainty from other model physics such as PBL, radiation, and microphysics (Bruyère et al. 2017, Mooney et al. 2017). CP (4-km [2.5-mi] horizontal grid spacing) WRF simulation results are shown in d)-f), covering the period 2002–2013 (Liu et al. 2017). Pannels a—c modified from Mooney et al. 2017. Pannels d—f curtesy of Andreas Prein. Results for the Eastern U.S. are similar (e.g., Prein et al. 2017).*

Furthermore, in CP configurations one sees a substantial improvement in the simulation of precipitation amount, frequency, and intensity, which is a robust result appearing across the models used and regions covered (e.g., Done et al. 2004, Schwartz et al. 2009, Clark et al. 2010, Prein et al. 2015).

There is a fundamental change in our capability to simulate heavy-precipitation storms when transitioning from grid sizes of 10 km (6.2 mi) or larger to kilometer-scale and smaller spacings (Figure 1-3). This is related to the representation of vertical motions (updrafts, downdrafts) and their attendant moist processes. The added value of further decreasing Δx to sub-kilometer scales is an area of ongoing research and will be discussed below.
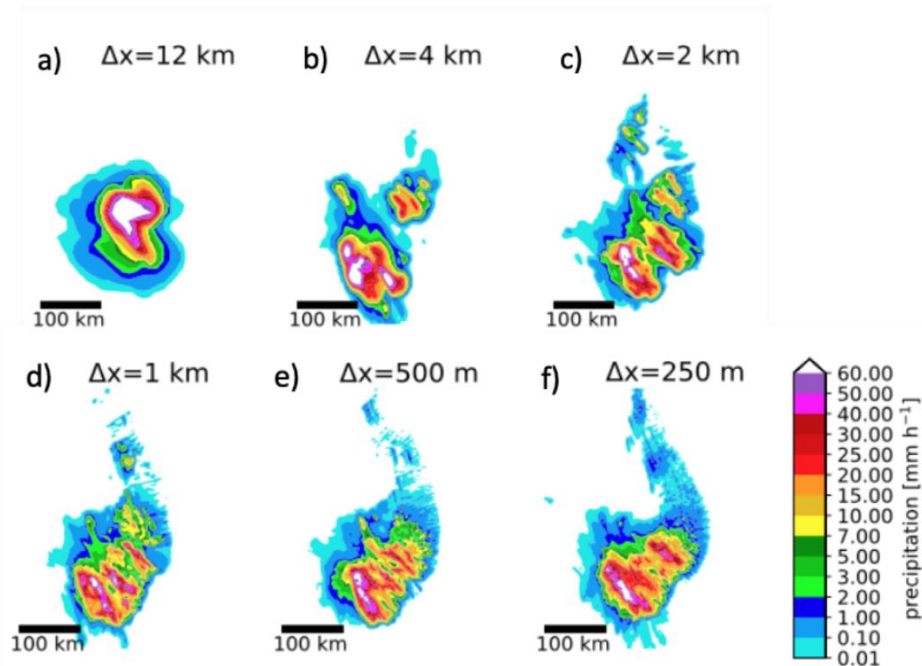
Figure 1-3   Hourly Precipitation Accumulation Simulated by an Idealized Version of the WRF Model with Horizontal Grid Spacings (Δx) Ranging From 12 km (7.5 mi) to 250 m (820 ft; Prein et al. 2019)

1.1.2   Orographic Lifting

Orographic lifting can be another driver of heavy precipitation, with two primary pathways found. First, the lifting can trigger convection, while, under certain conditions, the orography also can enhance low-level moisture inflow to the cells. A good example is that of the Colorado Big Thompson River flood event (July 31, 1976; Caracena et al. 1979). A more recent case of heavy rains from orographically-enhanced convection is that of the West Virginia flood of 2016 (Pokharel et al. 2018). Second, orographic lifting for long durations, even in non-convective, relatively stable environments, can produce very high accumulations which result in flooding. While rainfall rates are typically moderate in such situations, the extended timescales are determinative. Atmospheric rivers (ARs) that impinge on mountainous terrain, notably on the west coast of the U.S., can produce such events. ARs are often the dominant process for intense rainfall in mid- to high-latitude mountainous coastal regions on the west side of continents such as the U.S. West Coast (Lamjiri et al. 2017) and the Atlantic coast of Norway (Benedict et al. 2019).

Simulating orographically-influenced intense rainfall events demands realistic, high-resolution model terrain (see, e.g., Figure 1-4) to capture the height and steepness of the topography involved and to better reproduce the atmospheric flow interactions. The simulation of intense orographic precipitation is also sensitive to the microphysics parameterization used, while other schemes such as those for the planetary boundary layer scheme and land surface are less important (Liu et al. 2011). The optimal grid spacing and physics parameterizations depend on the specific mountain range and atmospheric process for the events under study.

a) Δx=26.4 km grid spacing

b) Δx=2.4 km grid spacing
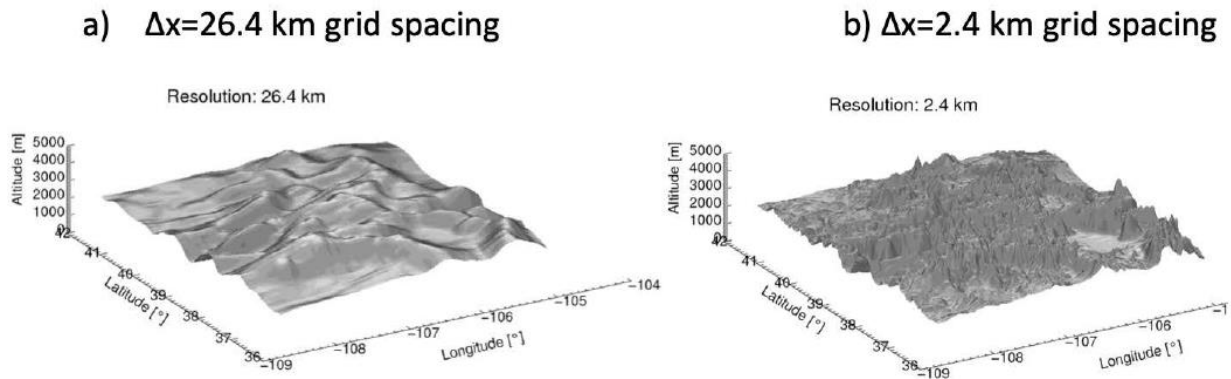
Figure 1-4   Model Representations of the Rocky Mountains in Colorado Showing Differences in Terrain Resolution with Different Horizontal Grid Spacings (Δx). (a) Δx=26.4 km (16.4 mi) and (b) Δx=2.4 km (1.5 mi)

1.1.3   Microphysics

Microphysics gain in importance in CP simulations since in these all precipitation production relies on the microphysics scheme, without any contribution from convective parameterizations. Further, more complex microphysical processes (e.g., the formation of graupel and hail) have to be considered with finer grids as cloud-scale dynamics (e.g., updrafts, downdrafts) are explicitly represented and increasingly into play. In such settings, microphysics schemes ultimately control precipitation efficiency (e.g., Weisman and Klemp 1982, Sui et al. 2005). There are sensitivities in this, however, as the simulated processes hinge on conditions that have been difficult to precisely describe, such as the fraction of solid particles in the condensate or the droplet size spectrum, and there can be large uncertainties in the parameters involved. A schematic of processes that are incorporated in a typical state-of-the-art two-moment microphysics scheme is shown in Figure 1-5. The two moments that are accounted for each hydrometeor class (e.g., rain, cloud ice) are the mass and number mixing ratio. Two-moment schemes allow the model to account for microphysical processes such as size sorting of particles (e.g., larger raindrops accumulate in the lower part of clouds), which are important for the simulation of heavy precipitation.

Modeling studies have shown that simulated heavy precipitation can be particularly sensitive to hydrometeor fall speed assumptions (Singh and O'Gorman 2014). Observations show that the terminal velocity of raindrops increases with their diameter (Foote and Du Toit 1969), increasing the potential for high accumulations when larger drops are formed. As large drops form most efficiently in warm regions of the cloud (i.e., in temperatures above freezing) by collision and coalescence— the warm rain process, clouds that have warm layers deeper than 3–4 km (1.8–2.5 mi) are potential generators of heavy rain (Doswell 2001).
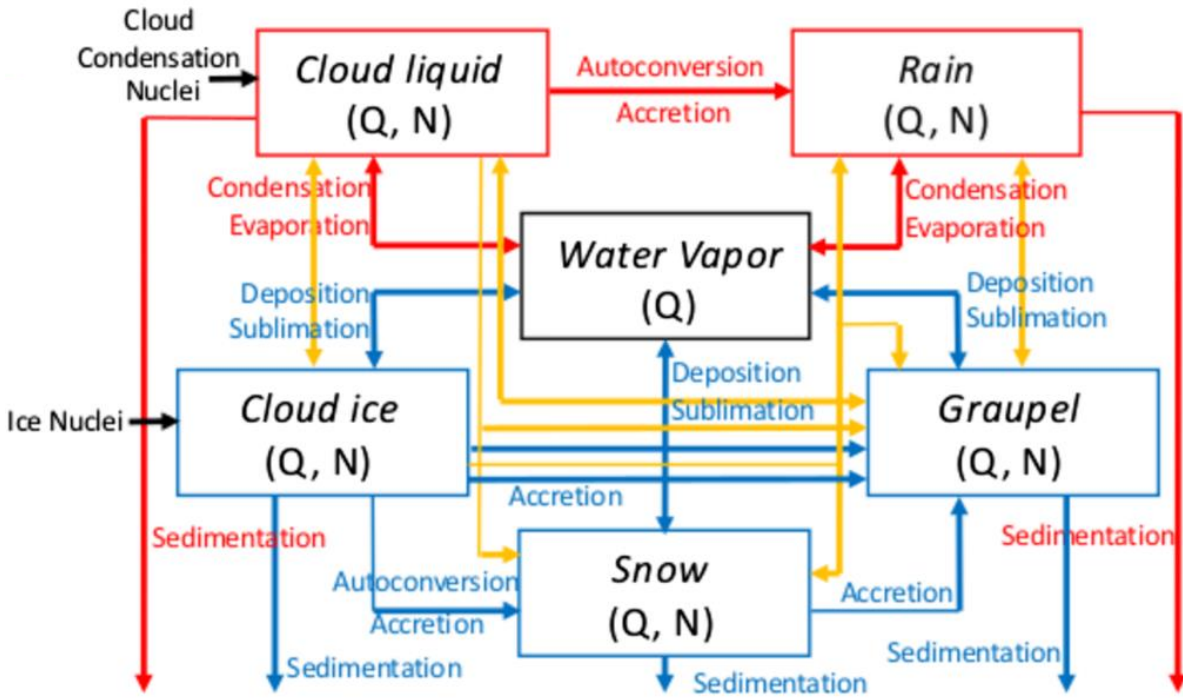
Figure 1-5   A Typical State-Of-The-Art Two-Moment Bulk Microphysics Scheme *Boxes represent different hydrometeor categories (liquid and ice) and water vapor. Q and N are the mass and number mixing ratios of a category. Arrows represent microphysical processes that convert Q and/or N between categories, as well as sedimentation (fallout from gravity). Red, yellow, and blue lines represent liquid, mixed-phase, and ice-phase processes (from Morrison et al. 2020).*

Aerosols advected into clouds are another factor affecting precipitation rates. The effects of aerosols on precipitation are complex, and in models depend on the storm type, environmental conditions (e.g., Tao et al. 2007), storm life cycle (e.g., Lynn et al. 2005, Van den Heever et al. 2006), and the microphysics scheme used (Lebo et al. 2012).

An overview of widely used microphysics schemes and their main characteristics is shown in Table 1-1 for acronyms). The schemes that represented particles such as slow falling graupel had difficulties simulating the observed, well-defined convective region and the high peak precipitation rates. A general shortcoming of all microphysics schemes appears to be an underprediction of the area of the stratiform precipitation region of an MCS. Morrison and Milbrandt (2015) also performed simulations of wintertime orographic precipitation cases and showed that the impacts of riming[3] and varying particle fall speeds on precipitation distribution are most critical and were best captured by the P3 and SBU-LIN packages. These results are consistent with findings from previous studies (e.g., Colle and Mass 2000; Lina and Colle 2011).

---

[3] Riming is the deposition and accretion of ice on hydrometeors as they fall through supercooled water droplets which freeze on contact with the particles.

Table 1-1   Widely Used Microphysics Schemes in CP Modeling

Scheme	Main Characteristics
Milbrandt–Yau two moment (MY2); Milbrandt et al. 2005a,b	***Prognostic hydrometeor categories:*** Cloud (liquid droplets), rain (precipitating drops), ice (pristine crystals), snow (large crystals/aggregates), graupel (rimed ice), and hail (high-density rimed ice and/or frozen drops) ***Particle size distribution:*** Each category is described by a complete gamma function ***Moments:*** Two-moment (with prognostic mixing ratios and total number mixing ratios)
Morrison two-moment microphysics scheme (MOR-H and MOR-G); Morrison et al. (2005,2009), Morrison and Milbrandt (2010)	***Prognostic hydrometeor categories:*** Rain, cloud ice, snow, and graupel/hail and mass mixing ratio of cloud droplets ***Particle size distribution:*** Inverse-exponential size distributions except for cloud droplets. Cloud droplet follows a gamma function with a spectral dispersion parameterization (Martin et al. 1994) ***Moments:*** Two-moment ***Difference between MOR-G and MOR-H:*** MOR-G uses a rimed-ice category that is consistent with graupel, while in MOR-H it is consistent with hail. The two settings differ in the bulk density and fall speed–size relationship for rimed ice.
NOAA/National Severe Storms Laboratory (NSSL); Ziegler (1985), Mansell et al. (2010), Mansell (2015)	***Prognostic hydrometeor categories:*** Number and mass mixing ratios of cloud droplets, rain, small ice, snow, graupel, and hail, as well as the particle volume of graupel ***Particle size distribution:*** Inverse-exponential size distributions for rain and graupel, and a gamma distribution for hail ***Moments:*** Two-moment
Stony Brook University–Lin (SBU-LIN); Lin and Colle (2011)	***Prognostic hydrometeor categories:*** Mass mixing ratios of cloud droplets, rain, cloud ice, and precipitating ice ***Particle size distribution:*** Inverse-exponential distributions are used for rain and precipitating ice particles ***Moments:*** One-moment
Thompson (THO); Thompson et al. (2008)	***Prognostic hydrometeor categories:*** Cloud droplets, snow, graupel–hail hybrid, cloud ice, and rain ***Particle size distribution:*** Snow density varies inversely with diameter, and its size distribution is a sum of two gamma functions, following observations by Field et al. (2005); in contrast, most microphysics schemes assume constant snow density. All other hydrometeors follow a generalized gamma distribution. ***Moments:*** One-moment for cloud droplets, snow, and graupel–hail hybrid. Two-moment for cloud ice and rain.
WRF single moment (WSM6); Hong et al. (2004), Hong and Lim (2006)	***Prognostic hydrometeor categories:*** Cloud droplets, rain, cloud ice, snow, and graupel ***Particle size distribution:*** inverse-exponential size distributions for rain, snow, and graupel. ***Moments:*** One-moment
WRF double moment (WDM6); Lim and Hong (2010)	***Prognostic hydrometeor categories:*** Cloud droplets, rain, cloud ice, snow, and graupel ***Particle size distribution:*** Cloud droplet and rain size distributions follow generalized four-parameter gamma functions ***Moments:*** One-moment for mixing ratios of cloud ice, snow, and graupel. Two-moment for cloud droplets, rain.
Predicted particle properties (P3); Morrison and Milbrandt (2015), Morrison (2015)	***Prognostic hydrometeor categories:*** Cloud droplets, rain, and ice. P3 avoids the arbitrary categorization of frozen particles into snow, ice, graupel, and hail by allowing for evolving frozen particles that can be any type of ice-phase hydrometeor. ***Particle size distribution:*** Three-parameter gamma distribution ***Moments:*** Two-moment
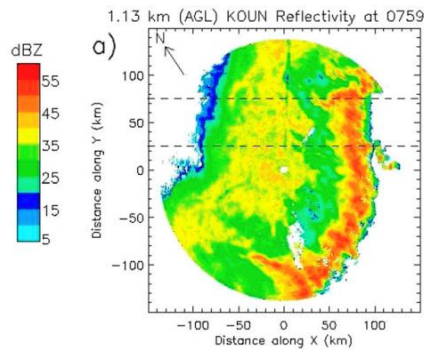
Figure 1-6   Cross Section of Observed Radar Reflectivity Through an Observed MCS From 19 June 2007 in The Central Great Plains Of The U.S. (Morrison and Milbrandt 2015)
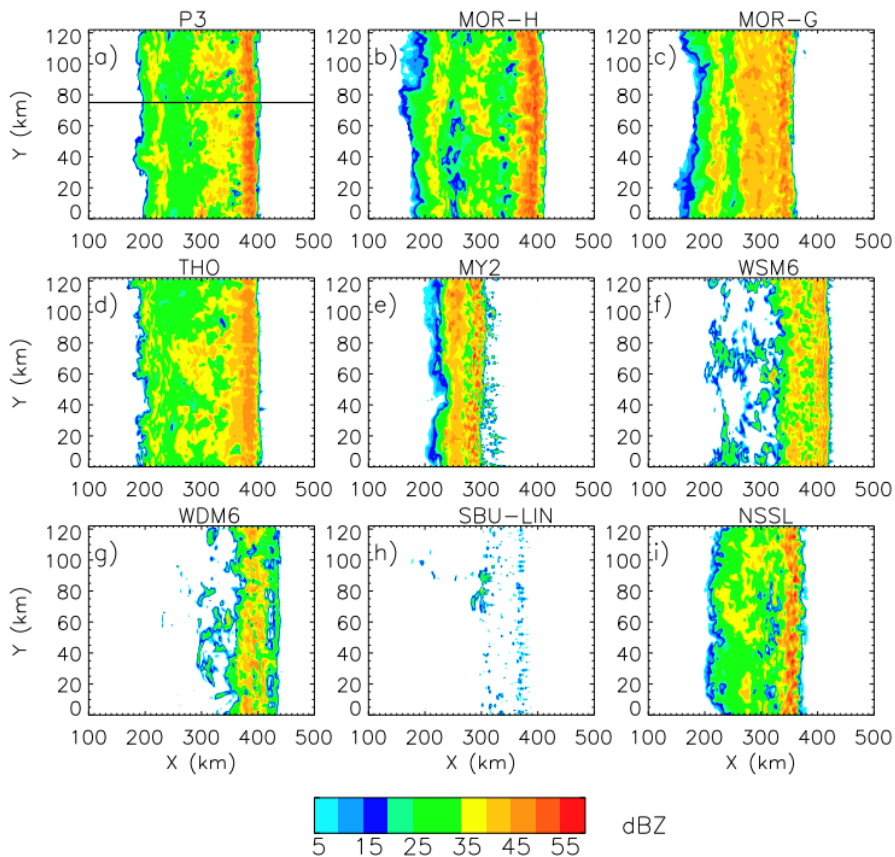
Figure 1-7   Horizontal Cross Sections at a Height of 1.1 km (0.7 mi) of Radar Reflectivity (dbz) Six Hours After Initializing an Idealized MCS Simulation Using the WRF Model *Panels show results from nine different microphysics schemes. Radar reflectivity is proportional to the particle density in the cloud and is related to the surface precipitation intensity. The model horizontal grid spacing was 1 km (0.6 mi), and 100 vertical levels were used (Morrison and Milbrandt 2015).*

1.1.4   Turbulence

Turbulent processes are another important factor that can affect the initiation, mesoscale dynamics, and rainfall efficiency of intense precipitating storms. The scale of turbulent motions spans orders of magnitude, from deep convective scales (i.e., kilometers) to energy diffusion scales (i.e., millimeters; Wyngaard 2004). Since the most energetic turbulent motions are on the scale of hundreds of meters, only parts of the turbulent energy spectra can be resolved in models applied at kilometer-scale grid spacings. This is problematic since there is no theory for representing unresolved turbulence in kilometer-scale numerical models (e.g., Wyngaard 2004, Moeng 2014). Idealized studies of convective storms show that grid spacings of less than 250 m are needed to resolve entrainment processes and updrafts (Lebo and Morrison 2015). Thus, CP model configurations yield motions that entrain too little, resulting in overestimations of maximum updraft velocities (see, e.g., Weisman et al. 1996, Fan et al. 2017, Prein et al. 2019).

Convective cell entrainment of unsaturated air facilitates evaporation and therefore affects rainfall efficiency and storm dynamics, particularly for the entrainment of dry (i.e., low relative humidity) air. A horizontal cross-section through a middle level of a simulated mature organized convective storm is shown in Figure 1-8. In this resolution-probing work (Lebo and Morrison 2015) the horizontal grid spacing (denoted as Δ_h in Figure 1-8) of 2 km (1.2 mi) simulates broad and weak updrafts, while the 250 m (880 ft) grid returns much more spatial variability, with smaller and stronger updrafts (Figure 1-8a,b). Thus, increasing model resolution yields more realism. However, even at this relatively fine spacing, updraft structures have not converged to the appearance of an actual cell, meaning that entrainment and associated processes (e.g., precipitation efficiency, draft dynamics) are still under-resolved. The consequence of inadequately simulating turbulent motions in CP models is an ongoing research problem. As Figure 1-8c shows, turbulent motions continue to change with increasing resolution, down to spacings of 33 m (Δh=33.33 m in figure). Ongoing research is focusing on how to parameterize unresolved turbulence in CP models (e.g., Muñoz-Esparza et al. 2014).
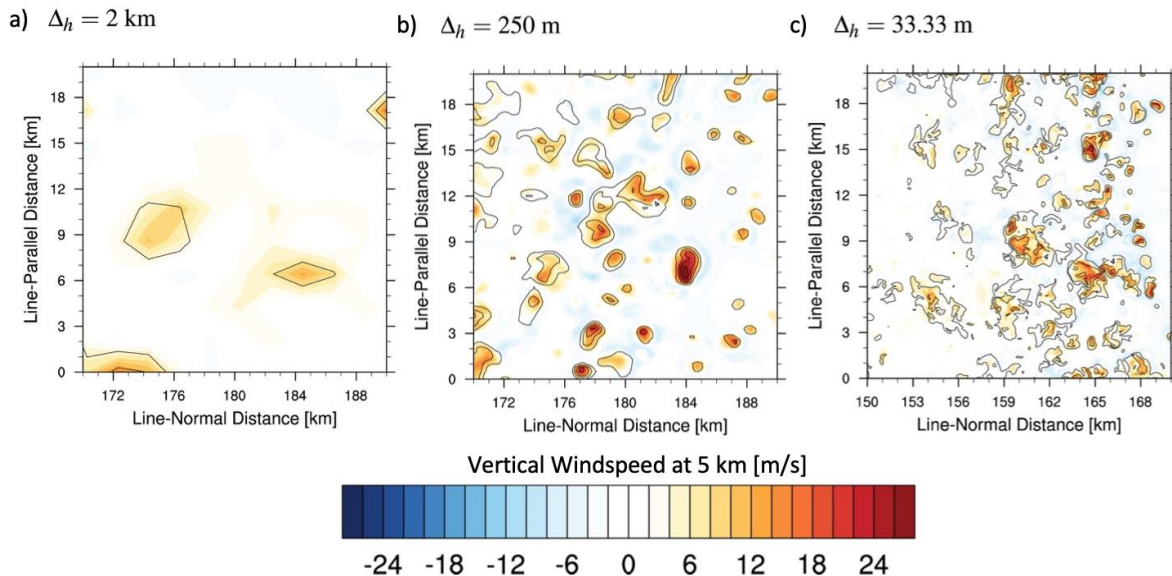
Figure 1-8   Horizontal Cross-Section at 5 km (3 mi) Through an Idealized Convective Storm System Simulated With Different Grid Spacings: (a) Δx=2 km (1.2 mi); (b) Δx=250 m (880 ft); (c) Δx=33.33 m (108 ft) *Vertical wind speed shown in filled contours. Total cloud mixing ratio is shown as black contour lines with 2 g kg−1 intervals (Lebo and Morrison 2015).*

The effect of unresolved turbulence in atmospheric models is accounted for in planetary boundary layer (PBL) schemes. These schemes parameterize the turbulent mixing of the lowest part of the atmosphere, the boundary layer, with the atmosphere above (Figure 1-9). Turbulent eddies in the PBL facilitate exchanges of moisture, heat, and momentum due to local and non-local mixing (e.g., Cohen et al. 2015). The latter process moves air parcels across large vertical distances before smaller eddies (local mixing) effectively diffuse the parcels. The accurate representation of these processes is key for the triggering and evolution of convective storms.
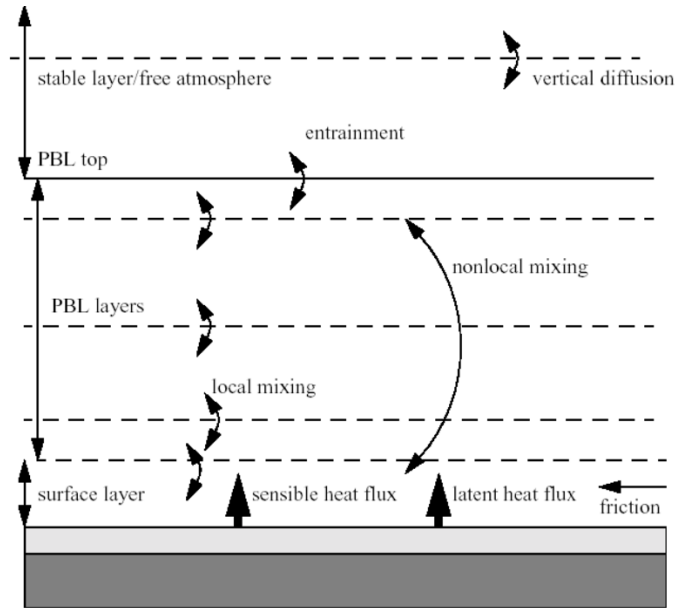
Figure 1-9   Dominant Processes in the Atmospheric Boundary Layer (From Ahasan et al. 2014)

Turbulence closures in weather prediction and climate models can be classified according to the order of the closure and whether the mixing approach is local or nonlocal (e.g., Stensrud 2009, Cohen et al. 2015). This is summarized below.

Order of PBL schemes: In PBL schemes, the equations of motion are decomposed into mean and perturbed components. The former represents the background, resolved atmospheric state, while the latter represents turbulent perturbations from the background state. The equations for turbulence contain more unknown terms than known terms, and the solution to this problem is to use a closure assumption. A first-order closure assumes that the second-order terms are functions of first-order (mean) terms, while a second-order closure uses first- and second-order terms to calculate third-order terms.

Local vs. non-local PBL schemes: Local PBL schemes only allow vertical levels that are directly adjacent to the grid cell of interest to affect that cell. Non-local schemes allow remote effects—that is, from all levels cells in the PBL— to affect the level in the grid cell of interest. Local schemes are known to have disadvantages in situations where local stability maxima exist in the PBL that are not representative of the overall stratification of the PBL (Stensrud 2009). Non-local schemes can account for large eddies that transport heat from the surface layer to the top of the PBL regardless of local stability maxima, and therefore they can represent deep PBL circulations more accurately than local schemes (Stull 1991). While local schemes can be

improved by using higher orders of closure, this increases their computational cost (e.g., Mellor and Yamada 1982).

Table 1-2 summarizes some widely used PBL schemes, with all of them are available in the WRF model, noting their advantages and disadvantages. Which scheme performs best is typically case-dependent; it also depends on interactions with other physical process schemes (e.g., interactions with microphysics; see, e.g., Mooney et al. 2017).

Table 1-2   Widely-Used Turbulence Schemes in CP Modeling (adapted from Cohen et al. 2015)

Scheme	Main Characteristics
MRF; Hong and Pan (1996)	A non-local scheme that uses a first-order closure. Can simulate deeper mixing within an unstable PBL accurately (e.g., Stull 1993). Tends to produce PBLs that are overly deep, especially in strong-wind regimes at night (Mass et al. 2002). This overmixing can suppress convective initiation (Bright and Mullen 2002).
Yonsei University Planetary Boundary Layer Scheme (YSU); Hong et al. (2006)	A non-local scheme that uses a first-order closure. Improves simulation of deeper vertical mixing in buoyancy-driven PBLs and shallower mixing in strong-wind regimes compared to the MRF scheme (Hong et al. 2006). However, it still simulates PBLs that are too deep for springtime deep convective environments, resulting in an underestimation of near-surface buoyancy (Coniglio et al. 2013).
MYJ; Janjic (1990, 1994)	Local 1.5-order closure scheme. Improves PBL simulations compared to preceding schemes (Mellor and Yamada 1982), without increases in computational costs. Tends to undermix PBLs for locations upstream of convection (e.g., Coniglio et al. 2013).
MYNN; Nakanishi and Niino (2004, 2006)	This local scheme offers both a 1.5-order (MYNN2) and a second-order (MYNN3) closure. MYNN2 improves the PBL depiction compared to non-local schemes (e.g., YSU) during springtime in environments that support deep convection (Coniglio et al. 2013). MYNN3 more accurately simulates deep PBLs, but with higher computational cost compared to MYNN2. Similar to the MYJ scheme, the local formulations of the MYNN2 and MYNN3 do not fully account for deep vertical mixing.

1.1.5   Synoptic Conditions

An NWP model's capturing of the synoptic environment is fundamental to the reproduction of intense precipitation scenarios and events. At least for non-orographically forced rainfall, the benefits of CP detail cannot be realized if the model cannot faithfully represent the large-scale environment and correctly place synoptic and mesoscale features, like cold fronts and short-wave troughs. Intense precipitation relies on significant atmospheric moisture, typically transported into the target regions by low-level advection. In both mid-latitude and high-latitude, events it is often seen that subtropical air masses feed the systems (e.g., Pfahl et al. 2014, Krichak et al. 2015, Gochis et al. 2015). To produce precipitation from the advected low-level moisture, a triggering/lifting mechanism must be present, and forcings include: orography, atmospheric boundaries (e.g., drylines, fronts), cold pool dynamics, and airmass instability.

An important synoptic element affecting the movement of precipitating systems is steering flow. For deep convective storms, mid-tropospheric winds (e.g., at ~7 km [4.3 mi] height) are most relevant (Carbone et al. 2002). However, the movement of convective storms is not only dependent on the large-scale flow, but also on the internal system's dynamics of propagation (see, e.g., Doswell et al. 1996, Houze 2004). Propagation reflects the contributions of cell generation and dissipation to the total system movement, and stationary storms, which can deposit large accumulations, develop in environments where large-scale steering flow and storm propagation dynamics are in balance. For tropical cyclones, storm motion is heavily influenced by the large-scale steering flow. Strong systems tend to be steered by winds through a deep layer of the troposphere, while movements of weak systems are most influenced by winds in the lower half of the troposphere (e.g., Chan 1985).

Bradley and Smith (1994) describe the climatology and synoptic environments of intense rain events over the U.S. southern Plains. Their review of the meteorological literature found that the dominant elements were: low-level winds and moisture, mid-level moisture, thunderstorm updraft strength, and vertical wind shear. They looked at heavy rain events with a recurrence interval of over 10 years in the study region centered on Oklahoma. Peaking in frequency from late spring to early fall, the target events relied on combinations of dynamic forcing, high moisture availability (e.g., relatively high precipitable water), and convective instability (e.g., CAPE). They found that rainstorms in this region either displayed strong dynamic forcing with vigorous ascent from a strong upper-level trough (Figure 1-10a) or weak dynamic forcing along stationary frontal boundaries or in regimes of high pressure (Figure 1-10b). The success of numerical models in capturing convectively driven events there is thus dependent on their ability to reproduce those elements. As discussed below, capturing the synoptic and mesoscale dynamical environments is something that NWP models can readily do, and high resolution is not generally required. However, the conditions of moisture availability and convective instability reflect finer-scale variability are more challenging for models to accurately depict, with their simulation sensitive to resolution and initial conditions.

(a)   (b)

Figure 1-10   Schematics of the Synoptic Environments of Intense Rainstorms over the Southern Great Plains from Bradley And Smith (1994) *(a) Strong forcing with relatively deep upper-level (500 mb) trough and strong frontal boundary. (b) Weak forcing with quasi-stationary front and lack of strong upper-level low.*

1.2   Model Simulation of Intense Precipitation: Forcing, Events, and Estimation

Several studies have examined the characteristics and distribution of intense precipitation forcing. Kunkel et al. (2012) analyzed the causes of daily precipitation events with a 1-in-5-yr

recurrence in the U.S. during the period 1908–2009 (see Figure 1-11). They found that frontal boundary forcing in extratropical cyclones was responsible for more than 50% of heavy rainfall events. This was followed by: 24% extratropical cyclone, near center of low; 13% tropical cyclone; and 5% MCS. Schumacher and Johnson (2006) found a more significant contribution from MCSs, which caused 75% of warm-season intense precipitation events in the eastern U.S. This highlights two things: (i) how difficult it is to differentiate the dominant processes that cause intense precipitation, and (ii) that most intense events are influenced by processes at multiple scales.

With respect to the forcing of intense precipitation across the CONUS, tropical cyclones are most relevant in the Southeast/Gulf Coast regions, extratropical cyclone low centers are most important along the Pacific coast, and extratropical cyclones and MCSs dominate elsewhere. A recent review of intense precipitation events and their large-scale meteorology by Barlow et al. (2019) concludes that events are often related to mesoscale processes that are triggered, enhanced, or organized by larger-scale processes. Simulating intense precipitation, therefore, requires models that can capture all scales of the relevant processes. In the following section, we discuss the sensitivities of the simulation of these various storm types to model configuration and application: grid size, physics, and setting (e.g., research case studies, weather forecasting, climate simulation).
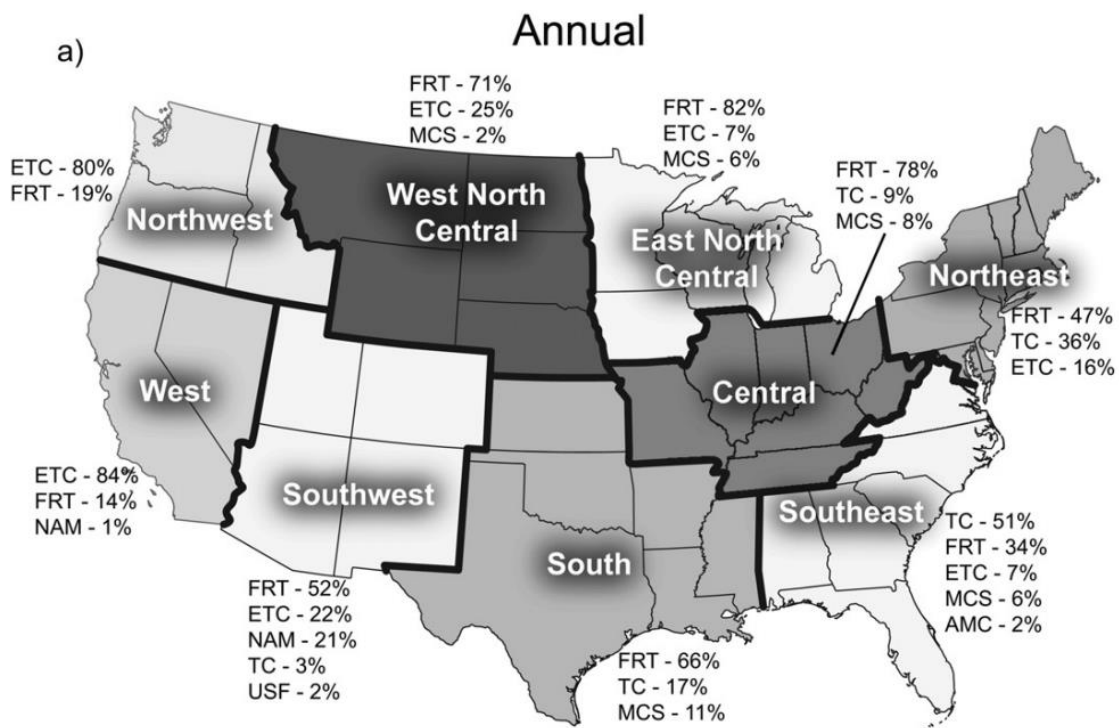
Figure 1-11 Causes of Intense Precipitation Events in Sub-Regions of the U.S. *Shown are annual contributions from: extratropical cyclones near a front (FRT), extratropical cyclones near the center of a low (ETC), tropical cyclones (TC), mesoscale convective systems (MCS), air mass (isolated) convection (AMC), the North American monsoon (NAM), and upslope flow (USF) (from Kunkel et al. 2012).*

1.2.1   Tropical Cyclones

For capturing the impact of tropical cyclone (TC) precipitation, accurate track forecasts are critical. While TC tracks are controlled mostly by synoptic conditions, there are questions about what model resolution is needed to successfully predict trajectories. For instance, while Xue et al. (2013) showed that a 4-km (2.5 mi) grid improved track forecasts compared to that of a much coarser global model, Davis et al. (2010) indicated that 12-, 4-, and 1.33-km (7.5-, 2.5-, 0.8 mi) grid spacings for forecasts of TCs had statistically indistinguishable track errors.

Given identical tracks of a landfalling TC, however, there is little question that CP configurations will produce better precipitation forecasts than coarser parameterized convection setups, as the former can explicitly represent the heavily precipitating structures (typically the eyewall and convective bands). Gentry and Lackmann (2010), for example, showed that the structure of hurricanes changes at a Δx of about 4 km (2.5 mi; Figure 1-12). At or below this scale, characteristics of the eyewall, such as its shape and embedded updrafts, are more realistically simulated. Furthermore, their storm minimum central pressure dropped by 30 hPa when Δx decreased from 8 km to 1 km (5 mi to 0.6 mi). Jin et al. (2014) and Davis and Bosart (2001) support these results, with the latter also showing that using a $\Delta x=3$ km (1.8 mi) resulted in more realistic hurricane intensification. While these studies looked at Atlantic storms, supporting results are also seen for other ocean basins (e.g., the Indian Ocean, Taraphdar et al. 2014) and from CP climate modeling (Gutmann et al. 2018).

Tropical cyclone size is another characteristic influencing storm surge and freshwater flooding, with larger cyclones typically having higher surge levels and rain accumulations. CP models better simulate cyclone size since they have a more realistic representation of the precipitating structures (e.g., spiral bands, storm core) and can more accurately capture the storm wind field. One factor in TC size is the environmental relative humidity (RH; Hill and Lackmann 2009), and precipitation outside the storm's core region can be very sensitive to the larger-scale RH.

While TC intensity and track are less sensitive to variations in model microphysics than to those in grid spacing, the cloud structure, rainfall rates, and rain areal coverage are dependent on physics scheme particulars and thus scheme choice. Wang (2002), for example, concluded that schemes' abilities to represent mixed ice-phase clouds were critical for the realistic simulation of cloud structure and precipitation amounts, in agreement with an early study of McCumber et al. (1991).

Climate change will likely increase events of intense precipitation associated with tropical cyclones (Gutmann et al. 2018, Patricola and Wehner 2018). For example, there is increasing evidence that climate change will result in a higher frequency of stronger (e.g., Gutmann et al. 2018) and slower moving (Kossin et al. 2018, Gutmann et al. 2018) systems, both of which can translate to increased rainfall accumulations. One study concluded that Atlantic tropical cyclones are moving (i.e., system translational speed) over oceanic areas 6% slower today than in the mid-20th century (Kosin 2018), with a slowdown of 20% for their movement over land. Kosin (2018) furthermore concluded that a 10% slowdown of TC speed could result in a doubling of local rain accumulation under 1°C (1.8 °F) warming. Thus, projections are that intense precipitation episodes from TCs will amplify in the future.
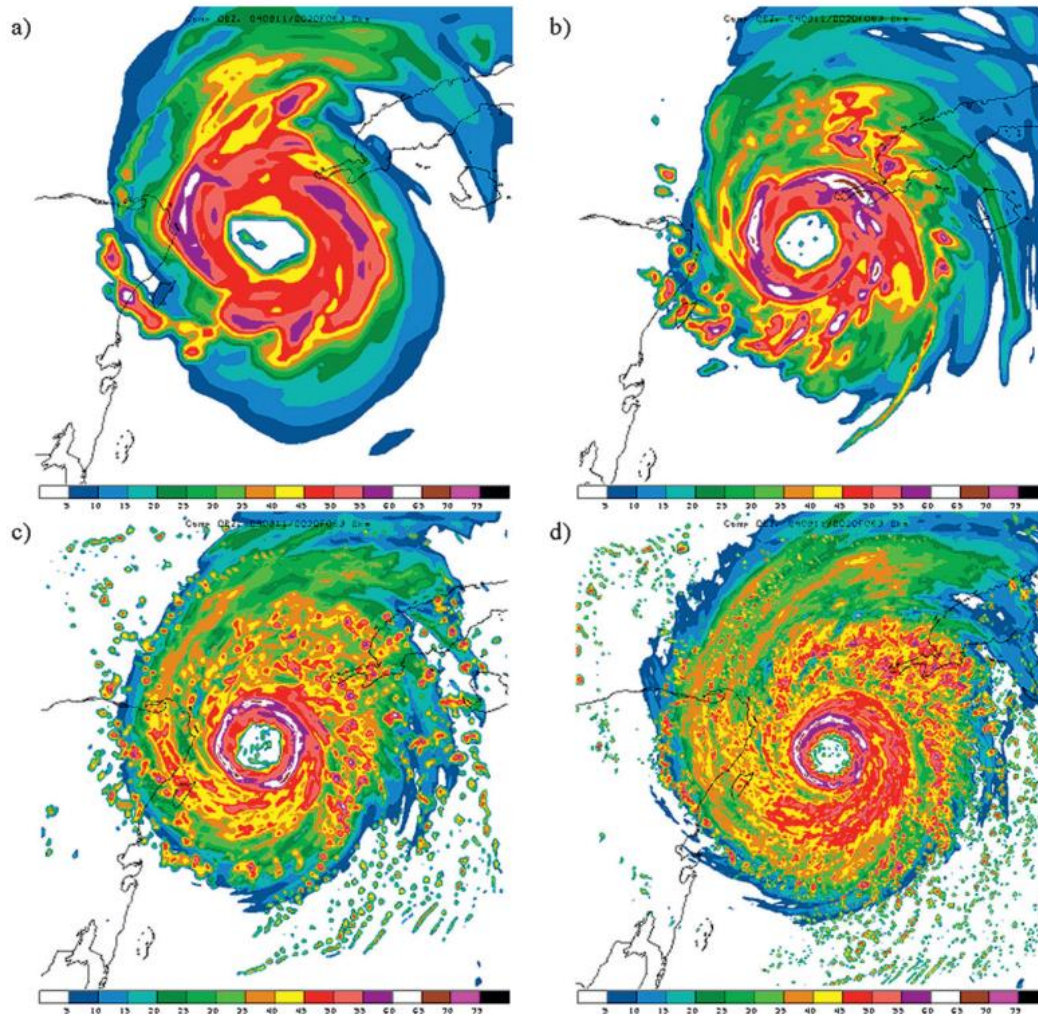
Figure 1-12   Model-Simulated Radar Reflectivity for a Simulation of Hurricane Ivan (2004) With (a) 8-, (b) 4-, (c) 2-, and (d) 1-km (5-, 2.5-,1.2-, and 0.6 mi) Horizontal Grid Spacing (Gentry And Lackmann 2010)

1.2.2   Mesoscale Convective Systems

There is little doubt that CP model configurations produce better precipitation forecasts than convection-parameterizing model configurations (e.g., Done et al. 2004; Bukovsky et al. 2006; Kain et al. 2006; Schwartz et al. 2009; Clark et al. 2016). Thus, current modeling inquiries are addressing the question of how much Δx needs to be reduced below the marginal CP scales of 3–4 km (2–2.5 mi). This has been examined in both idealized and real-data studies, summarized below.

1.2.2.1   Idealized Simulations

Idealized MCS simulations have been used since the 1990s to study sensitivities to model Δx and physics. In a seminal study, Weisman et al. (1997) simulated an idealized squall line with Δx ranging from 12 km to 1 km (7.5 mi to 0.6 mi) and concluded that Δx=4 km (2.5 mi) is

sufficient to capture most of the mesoscale structure and storm evolution compared to a 1-km (0.6-mi) simulation considered as truth. Concerning the actual precipitation from simulated storms, they found that model rainfall efficiency does increase at finer Δx, due to decreased evaporation. Using a similar experimental design, Bryan et al. (2003) produced simulations with Δx between 1 km (0.6 mi) and 125 m (410 ft) for a squall line, concluding that while 1-km (0.6 mi) forecasts could be useful for operational purposes, a Δx closer to 100 m (330 ft) was needed to resolve turbulent flows. Bryan and Morrison (2012) also used a similar model setup as Weisman et al. (1997) but included Δx=250 m (820 ft) simulations and microphysics sensitivity tests. They concluded that surface rainfall is more sensitive to varying the Δx than to varying the microphysics and further found that km-scale simulations developed more slowly and produced more precipitation than finer sub-km (e.g., 250-m; 820 ft) runs. They attributed these differences to an underestimation of dry air entrainment in convective cells at kilometer-scales. Comparing their simulations with high-resolution observations they concluded that (i) using a more complex microphysics scheme (i.e., 2-moment scheme), which represents hail as a separate hydrometeor, and (ii) applying a Δx of 250 m (820 ft) yielded the best performance.

Building on these collective results, Lebo and Morrison (2015) investigated even higher-resolution simulations and performed idealized squall line simulations with Δx down to 33.33 m and with 125-m and 250-m vertical grid spacing (410 ft and 820 ft respectively). They showed that spacings of Δx= 250 m or less are necessary to capture mid-level entrainment into convective cores and that increasing the vertical grid spacing did not significantly affect the results. This suggests that for simulations of intense precipitation events driven by convection there is greater sensitivity to, and need for, higher horizontal resolution than vertical. Recently, Prein et al. (2019) have shown that a major regime shift in simulating MCSs occurs when transitioning from Δx=12 km (7.5 mi) to Δx=4 km (2.5 mi) simulations of MCSs, in agreement with Weisman et al. (1997). Their further decreasing Δx to 1 km (0.6 mi) resulted in smaller structural changes, while the results largely did not change between Δx=1 km (0.6 mi) and Δx=250 m (880 ft). Prein et al. (2019) also have shown that Δx does not have a significant effect on climate change analyses and have concluded that CP models can provide reliable climate change signals.

In summary, these idealized studies suggest that realism in the simulation of the convective systems behind intense precipitation events is improved as horizontal grid spacing is reduced in the CP regime. Nonetheless, many of these studies also suggest that 1–3-km (0.6–1.8-mi) forecasts are adequate for forecasting purposes, which has been corroborated by real-data studies, discussed next.

1.2.2.2   Real-Data Simulations

Retrospective forecasts of heavy precipitation events are used to understand how CP forecasts are sensitive to the representation of atmospheric processes and to model configuration. MCS archetypes prone to producing intense precipitation have been documented in observations (Schumacher and Johnson 2005; Peters and Schumacher 2014) as well as simulated in case studies using CP configurations (e.g., Peters and Schumacher 2016; Nielsen and Schumacher 2018). To date, systematic real-data studies have yielded different conclusions about the advantages of decreasing grid spacing beyond 3-4 km (1.8–2.5 mi) for the successful prediction of intense-rainfall-producing MCSs and other types of convective systems across the U.S. east of the Rockies.

For example, while Kain et al. (2008) and Schwartz et al. (2009) found springtime 2- and 4-km (1.2- and 2.5-mi) model forecasts produced remarkably similar precipitation, Schwartz et al.

(2017) found that decreasing spacing to 1 km (0.6 mi) improved upon 3-km grid forecasts (1.8 mi). Interestingly, Schwartz et al. (2017) also noted 1-km (0.6 mi) forecasts produced fewer intense rainfall events (i.e., rainfall rates ≥ 20 mm/h; 0.8 in/h) than 3-km (1.8 mi) forecasts; although, the 1-km (0.6 mi) forecasts agreed better with observations. The latter result was presumably due to the finer grid's better capturing of dry air entrainment, a process that would reduce rainfall rates. Overestimating precipitation at ~3-km (1.8 mi) Δx is a finding consistent across different models (Herman and Schumacher 2016). Similar model tendencies have been documented in real-data case studies. Schumacher (2015), for instance, noted that 4-km (2.5 mi) simulations accurately produced a supercell and a subsequent flood-producing MCS, while simulations with finer Δx failed to produce the event. In contrast, Xue et al. (2013) found that a 1-km (0.6 mi) Δx was needed to capture the deep convection forcing the event studied.

The results of Schwartz et al. (2017) were partially confirmed by the recent work of Schwartz and Sobash (2019), who found that for events over the eastern CONUS and Mississippi River Basin, the modeled frequency of warm-season rainfall events with rates ≥20 mm/h (0.8 in/hr) derived from 1-km (0.6 mi) forecasts was closer to the observed frequency than that derived from 3-km forecasts (1.8–mi). Over the Great Plains, the 3-km (1.8-mi) and 1-km (0.6-mi) heavy precipitation climatologies were more similar. Schwartz and Sobash (2019) also found that 1-km precipitation forecasts were better than 3-km forecasts during the cool season and spring, but not during summer. Corroborating that summer result, Thielen and Gallus (2019) found that their 1-km (0.6 mi) forecasts of MCSs were not better than 3-km (1.8 mi) forecasts over 10 summertime cases.

Feng et al. (2018) used a 4-km (2.5 mi) model setup to study the impact of two state-of-the-art microphysics schemes— the Thompson (Thompson et al. 2008) and Morrison (Morrison et al. 2009) packages— on simulated MCSs over four months. While both schemes yielded similar precipitation amounts, hourly precipitation intensities, rainfall diurnal cycle, upper-level cloud shields, precipitation areas, and lifetimes of MCSs were better captured with the Thompson scheme.

Using a climate modeling approach, Prein et al. (2017c), through comparison with high-resolution rainfall observations, showed that a Δx=4 km (2.5 mi) model can capture MCS properties such as peak hourly precipitation, precipitation volume, system size, and translational speed. However, there was a significant underestimation of MCS frequencies in late summer in the central U.S., mainly caused by the misrepresentation of atmospheric feedbacks involving soil moisture (Barlage et al. 2018). Using the same model setup, Prein et al. (2017b) showed that a projected future climate will increase the frequency of intense precipitating MCSs in the U.S. The largest changes were found in the volume of the MCS heavy rainfall, which increased by up to 80% under the RCP8.5 scenario. Such changes are partly already detectable in observational records in the U.S. (Feng et al. 2016).

1.2.3   Extratropical Cyclones

Extratropical cyclones are another driver of intense precipitation and their simulation accuracy relies on a model's ability to represent the synoptic and mesoscale environments. Fronts in extratropical cyclones present temperature gradients along which warm air is lifted over colder air. Heavy precipitation can occur along frontal zones and near the low center.

As discussed in Secs. 2.5 and 3.5, the highest precipitation accumulations occur when fronts are nearly stationary, and intense precipitation may occur in frontal cyclones when sufficiently moist and relatively warm air ascends (Figure 1-10). The stream of warm/moist air that is

advected and lifted is sometimes called the warm conveyor belt (Figure 1-13a; e.g., Harrold et al. 1973). Warm conveyor belts can span thousands of kilometers and influence cyclone dynamics significantly from the substantial release of latent heat in their associated precipitation (e.g., Madonna et al. 2014). Pfahl et al. (2014) show that warm conveyor belts are a dominant source of intense precipitation in many mid-latitude regions, accounting for up to 80% of heavy precipitation situations in the southeastern U.S. In addition, convective cells can be embedded in the warm conveyor airstream, which can further amplify rain rates.

For realistically modeling intense precipitation events, the frontal and cyclonic processes and their interactions across a wide range of scales must be accurately simulated. While the large-scale conditions (e.g., the movement of air masses) can be captured in coarse global climate models with horizontal grid spacings on the order of 100 km (~62 mi), embedded small-scale processes and flow interactions with surface features (e.g., orography, coastlines) demand much higher-resolution configurations.
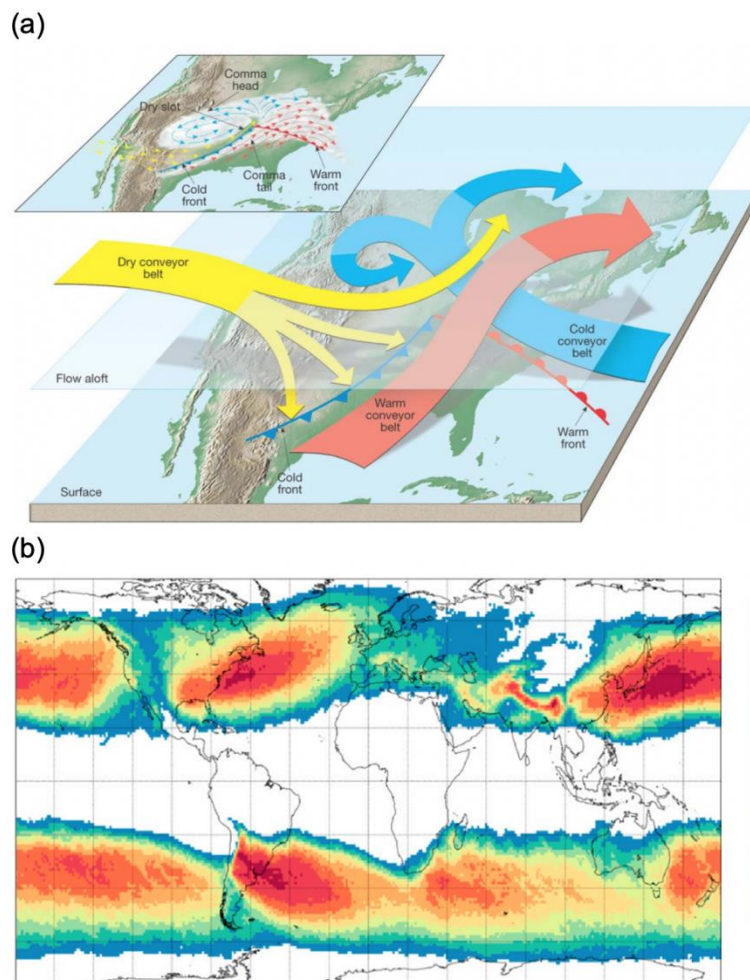
Figure 1-13   (a) Schematic of Interactions of dry, cold, and warm Conveyor Belts in an Extratropical Cyclone. (b) Analysis of Percent of Intense Precipitation that is Related to Warm Conveyor Belts (Pfahl Et Al. 2014)

1.2.4 Orographic Precipitation

As might be imagined, a model's ability to simulate intense precipitation events in which topography is a key forcing element will be strongly dependent on the model's representation of that topography. Thus, grid resolution is critical.

Case study experiments have concluded that a Δx of approximately ~1 km (0.6 mi) or finer is needed to capture orographically-forced heavy precipitation events (e.g., Colle and Mass 2000; Colle et al. 2005; Garvert et al. 2005; Hart et al. 2005; Buzzi et al. 2014; Schwartz 2014; Bartsotas et al. 2017; Gowan et al. 2018). However, there have been some exceptions (e.g., Grubisic et al. 2005).

Using a climate model setup for wintertime precipitation over the Rocky Mountains, Rasmussen et al. (2011) confirm the above-referenced case study results. They showed that Δx had to be at least 6 km (3.7 mi) to capture the observed precipitation. Apart from that of horizontal grid spacing, the applied vertical grid spacing can have significant effects on simulated accumulations in mountainous regions. In a study of a flood event in the U.S. Pacific Northwest, Colle and Mass (2000) showed that the windward-slope precipitation varied by up to 30%, while the leeside precipitation varied by up to 80%, when vertical grid spacing was changed. In this setting, the precipitation was sensitive to variations in mountain wave structures that resulted from the different vertical spacings tested.

Idealized models of intermediate complexity that use linear theory (Smith 1989) to describe airflow in complex terrain have also been used to simulate orographically-induced intense rainfall. Such models, which depend on a cumulus parameterization for convective precipitation, can account for airflow over mountain ranges and gravity wave formation, but they cannot account for gravity wave breaking or blocking, which can be important (e.g., Smith 2006, Hughes et al. 2009). Intermediate-complexity models are, however, significantly less expensive to run than full numerical models. Kunz and Kottmeier (2006), for example, used a linear model to successfully simulate two heavy rainfall events over a hilly region in southern Germany. They concluded that Δx=2.5 km (1.5 mi) performed best, while underestimations of rainfall increased with a coarsening of the grid spacing (viz., a 35% underestimation with Δx=10km; 6 mi). More recently, Horak et al. (2019) used a similar linear model approach over New Zealand at Δx=4 km (2.5 mi) and showed large improvements in the precipitation produced compared to the reanalysis dataset that was forcing the model.

Idealized approaches have also been used to investigate orographic precipitation events. In this context, Colle (2004) probed the impact of model microphysics on orographic precipitation. The study concluded that microphysics schemes that account for supercooled water and graupel are necessary to produce realistic orographic precipitation characteristics. A scheme that did not account for these effects produced 30–40% more precipitation in the immediate lee of the mountain, and a scheme that only accounted for liquid clouds underforecast leeside precipitation (Colle and Mass 2000). Liu et al. (2011) confirmed these results and showed in a regional climate model setting that orographic wintertime precipitation is highly sensitive to the model microphysics. Sensitivities to the model radiation, land surface, and planetary boundary layer schemes were small in comparison.

Using an idealized modeling setup, Siler and Roe (2014) showed that orographic precipitation increases are smaller than near-surface water vapor increases under climate change conditions. With climate change, precipitation maxima are expected to shift downstream of forcing orography, leading to larger precipitation increases in the lee of mountain ranges than on

the windward slope. On the other hand, a transition of snow to rain under climate change might intensify precipitation on the windward slopes of mountain ranges, resulting in an increased rain shadow effect on the leeside (Pavelsky et al. 2012). An increase in rainfall could also significantly increase the runoff in rain-on-snow flood events mainly due to increases in the snow-covered areas involved (Musselman et al. 2018).

1.2.5 Mixed Systems

The forcing for intense precipitation events may not necessarily fall into one of the above categories. Some cases can present a variety of processes amplifying rainfall accumulations through their interactions (Barlow et al. 2019). An example is the 2016 West Virginia Flood resulting from rainfall over June 23–24, 2016 (Figure 1-14).

This event was caused by a series of MCSs that produced torrential rains over West Virginia, with accumulations up to 9.8 in (250 mm; 9.8 in). The storms developed in a very moist and unstable air mass with precipitable water values up to 50 mm (2 in) and CAPE of about 1500 J/kg. This moist air was, at the surface, bounded by a nearly stationary frontal boundary and pushed upslope against the Appalachians, which intensified the rainfall. A series of MCSs were initiated along the frontal boundary and moved southwest into West Virginia, redeveloping over the same area. Applying a model to simulate such an event demands a realistic representation of multi-scale processes and their interactions, requiring high resolution to capture the critical topography as well as the convective dynamics and precipitation microphysics.
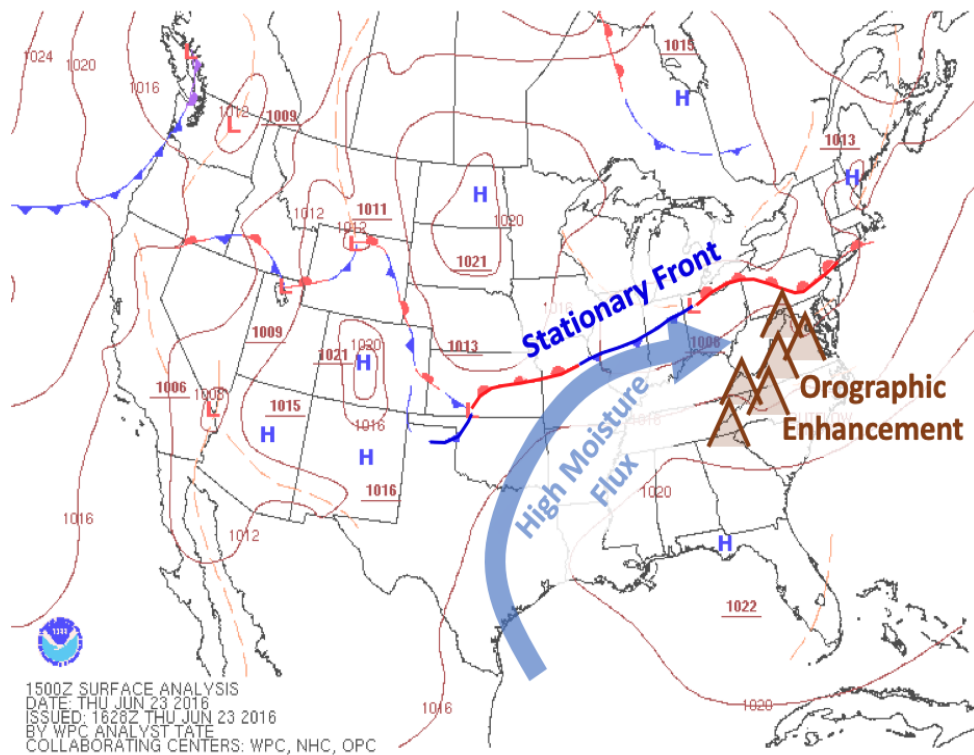
Figure 1-14   Synoptic Setting and Large-Scale Processes Involved 2016 in the West Virginia Flooding, 1500 UTC 23 June 2016 (Modified from NOAA Surface Analysis)

1.2.6 Modeling of Intense Precipitation Events for PMP Estimation

Events of extreme precipitation are inherent in the concept of Probable Maximum Precipitation (PMP). PMP has been an important input for structural and civil engineering plans and for designing systems, such as nuclear power plants, to handle extraordinary environmental conditions. The World Meteorological Organization (WMO) defines PMP as the greatest depth of precipitation for a given duration that is physically possible and reasonably characteristic over a particular geographic region at a certain time of year, with no allowance for climate trends. Thus, the notion of PMP implies the occurrence of statistically extreme precipitation events, although the most widely-used PMP methods (e.g., NWS Hydrometeorological Reports) are deterministic and do not provide statistical or probabilistic information such as average recurrence interval.

Over the past decade, there has been a growth in numerical modeling efforts in PMP to move the estimation from heuristic extrapolative methods to approaches involving full-physics atmospheric simulations. While we present a summary here, we have limited consideration of numerical modeling for PMP purposes based on feedback from NRC, whose research program is focused on probabilistic approaches.

Different methods have been used to estimate PMP, both hydrometeorological and statistical (Singh et al. 2018), although hydrometeorological methods are the most widely applied in the U.S. One hydrometeorological method often explored in NWP model investigations of PMP is the moisture maximization method. In this approach, the storm precipitation is increased to a value that is consistent with the maximum moisture in the atmosphere for the storm location and time of occurrence, based on historical records. A basic assumption in this method is that precipitation is linearly related to precipitable water. Under the moisture maximization method, PMP may be expressed as (Chen et al. 2017):

$$PMP = P \times (w_{p\_max}/w_{p\_storm}),$$

where P= observed extreme rainfall accumulation (based on a historical sample of extreme storms), w_{p_storm}= extreme storm precipitable water (based on the sample), and w_{p_max}= highest observed precipitable water (based on historical records). Problems with this formulation are the assumption of a linear relationship between PW and rainfall and the reliance on limited local historical records.

In an early NWP modeling study motivated by the procedures used for the design of high-hazard structures and PMP estimation, Zhao et al. (1997) ran the MM5 (5th-Generation Penn State−NCAR Mesoscale Model; Grell et al. 1994) to investigate the effects on rainstorms of artificially boosting atmospheric moisture. Their target warm season, central Plains event involved over 200 mm (7.9 in) of rain in a few hours and occurred in an environment of weak synoptic forcing. Moisture in the model was varied from 75−125% of observed values, and it was found that storm rainfall total, storm structure, and storm timing were strongly sensitive to such changes in water vapor. Moreover, the linear moisture maximization approach of traditional PMP analysis underestimated the total storm rainfall and gave an improper spatial distribution.

Zhao et al.'s illumination of possible deficiencies in the PMP estimation framework was reinforced by the work of Abbs (1999), who assessed the ability of models to evaluate the assumptions in the PMP calculation conceptual model, namely: (i) that the precipitation is linearly related to the precipitable water (PW); (ii) that the precipitation efficiency does not

change as the moisture availability increases; and (ii) that terrain modulates the distribution of precipitation, but does not affect the synoptic dynamics of the storm environment. Looking at simulations of various cases using the Regional Atmospheric Modeling System (RAMS; Pielke et al. 1992) down to 7 km (4.3 mi), Abbs (1999) found inaccuracies in the PMP assumptions, specifically noting that that precipitation is not necessarily linearly related to PW and that there can be terrain feedbacks to the synoptic aspects of a storm.

Ryu et al. (2016) conducted a modeling study of heavy rains over Iowa using the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2019) down to 1 km (0.6 mi). Their PMP/intense precipitation investigation probed: (i) the relations of the regional distribution of heavy rainfall events in the central U.S. to variations in precipitable water, water vapor transport, and atmospheric saturation; (ii) the effect of the diurnal variation of water cycle components on the regional distribution of heavy rainfall; and (iii) the mechanisms by which low-level moisture yields heavy rainfall. For their target region, they found that the heavy rainfall was weakly correlated with precipitable water. Neither extreme values of precipitable water or vapor flux were a necessary or sufficient condition for heavy rainfall simulation. Rather, they found that the thermodynamic properties tied to the vertical distribution of water vapor were more important.

A line of heavy precipitation modeling studies in the PMP context has focused on events forced by atmospheric rivers. Ohara et al. (2011) applied the MM5 at 3 km (1.8 mi) to an atmospheric river (AR) flood event in a Northern California watershed. They experimented with techniques later picked up by subsequent researchers, which, if desired, could be applied for future NRC precipitation studies. Their historical storm was maximized by modifying the model IC/BCs in three ways: (1) setting the relative humidity at the lateral boundaries to 100%; (2) imposing the BCs corresponding to historical peak precipitation conditions over the watershed; and (3) spatially shifting the wind field along the boundaries to reposition the incoming moisture flux. Each modification significantly increased the precipitation, demonstrating the importance of applied wind and moisture conditions at model boundaries for maximizing precipitation.

Ishida et al. (2015) built on this by using the MM5 at 3 km (1.8 mi) in simulating historic AR precipitation cases in three California watersheds, to get a bound on local rainfall as input to an estimation of extreme floods. They applied the Ohara et al. (2011) RH maximization method of setting boundary RH to 100% and also experimented with shifting the lateral boundaries to maximize the contribution to precipitation of the AR's moisture fluxes. Their simulations produced watershed precipitation amounts that were significantly larger (1.4–1.7X) than the historical maxima. This is another example of how NWP model frameworks can be manipulated to generate and analyze extraordinary precipitation.

Ohara et al. (2017) also used the MM5 to reproduce design storm events forced by ARs by modifying lateral boundary moisture. Using 3-km (1.8 mi) domains, their tests increased boundary RH to 100%. They found that while the prescribed moisture increased precipitation, the amplification was not consistent. More importantly, in some watersheds, rainfall decreased in some events, indicating that in regions of complex topography an increased synoptic moisture supply is not guaranteed to yield more local rainfall. Their results contradicted the conventional thinking of PMP moisture maximization (Schreiner and Riedel 1978) and confirmed the weakness in the traditional assumptions previously reported by Abbs (1999) and Rastogi et al. (2016). One notable conclusion was that dynamic and nonlinear interactive processes, upstream impacts, and local flow field responses make it difficult to definitively characterize the optimum atmospheric or model configurations for the physical upper bound of precipitation in intense events.

A similar approach in model-based heavy rainfall estimation— attempting to maximize precipitation via boundary RH forcing— was applied in the tropical cyclone setting by Lee et al. (2017). They used WRF at 3 km (1.8 mi) to simulate a typhoon-driven historic rainfall event in Korea (Typhoon Rusa, 2002). They ran WRF in two maximization scenarios, which in part differed from the approach of Ohara et al. (2017): (i) increasing both the lateral boundary air temperatures (keeping RH constant) and lower boundary sea surface temperatures; and (ii) setting RH=100% with temperatures as specified in (i). Only with their second approach could the model-derived PMP values agree with the official estimates. Noted advantages of their method over the traditional (non-modeling) method for heavy precipitation estimation were: (i) that the modeling approach accounts for the effects of nonlinear atmospheric processes and provides PMP estimates without cruder assumptions (e.g., the simple linearity of precipitation response with moisture content); and (ii) that it provides temporal and spatial patterns of intense precipitation based on physics.

In recent work applying moisture maximization in the simulation of intense precipitation from ARs, Toride et al. (2019) noted that the approach of Ohara et al. (2011), by systematically saturating all the model boundaries, may introduce disturbances to the atmospheric conditions beyond what is realistic. Thus, using WRF down to 4 km (2.5 mi), Toride et al. modified moisture only on the boundary segments intersecting the path of ARs and did not impose strict saturation. They found that saturation of entire model boundaries can produce unrealistic atmospheric conditions and that furthermore, given storm structure, stability, and topography, such saturation does not necessarily maximize precipitation over a watershed.

As the sampling of literature above indicates, an outstanding area of research in heavy precipitation in the PMP context is how to realistically and consistently get models to depict, with confidence, extraordinary situations. In that vein, to assist others in applying WRF for this problem, Chen et al. (2017) proposed a modeling framework. Using a rainfall event in Tennessee in 2010 with accumulations to 493 mm/48 hr (19.4 in/48hr), their analysis suggested that WRF for this purpose is most sensitive to the IC/BC conditions and that simulations benefit from finer than 5 km (3.1 mi). Certainly, the latter is something that has emerged from other, prior studies, and the need for convection-permitting grids to probe intense precipitation is given below as a finding of this review. More recent work by Toride et al. (2018), using WRF at 4 km (2.5 mi), found that model precipitation in AR events is most sensitive to microphysics, with the Goddard and the Lin-Colle schemes identified as superior in their runs. While these two studies are nowhere near enough to explore the broad parameter space of WRF and other mesoscale models for the intense precipitation problem, such work provides a point of reference for future modeling investigations.

1.3   Current Challenges and Opportunities in Modeling of Intense Precipitation

1.3.1   Observational Constraints

Evaluating intense precipitation events is challenging due to the need for high-resolution, high-quality observational datasets. Such datasets do not exist in many places, and, even in data-rich areas such as Europe and North America, observational uncertainties in intense precipitation cases are large (e.g., Prein and Gobiet 2017). Rain gauges are the backbone of precipitation observation: they typically provide the highest-quality, most reliable data and are often used to calibrate and evaluate in-situ radar and satellite observations (e.g., Crosson et al. 1996, Stocker and Wolff 2007). However, rain gauge measurements are not perfect. Their dominant error is precipitation undercatch. While the magnitude of this depends on the type of precipitation and the particular measurement array, it typically ranges from 3% to 20% for rain

and can be from 40% and 80% for snow, in case of non-shielded gauges (Førland et al. 1996; Goodison et al. 1998). Undercatch errors increase in situations with high wind speeds, which can often be the case in intense rain settings (e.g., tropical cyclones).

Another source of observation error that can be significant is sampling uncertainty. Sampling uncertainty primarily depends on the station density and the spatial heterogeneity of precipitation (i.e., stratiform vs. convective precipitation; see, e.g., Schneider et al. 2014). In one assessment, Rudolf et al. (1994) estimated that precipitation sampling errors were between $\pm$7-40% when five rain gauges were located in the relatively large area of 2.5x2.5° (~250x250 km; 155x155 mi). Using 10 stations reduced this bias to $\pm$5-20 %. Comparable results were found by Prein and Gobiet (2017).

Most model evaluation studies rely on precipitation observations that have been interpolated to a grid since that allows a direct comparison to modeled grid-average precipitation rates, simplifies the areal averaging of precipitation, and provides precipitation estimates at locations without gauges. However, interpolating station data to a grid tends to smooth gradients, which leads to an under-representation of the precipitation extrema (Haylock et al. 2008, Hofstra et al. 2010). The effect of the gridding method can alter climatological mean precipitation by as much as 50% in poorly observed regions. Moreover, the particular interpolation method applied is especially important for rainfall greater than 20 mm/d (0.8 in/d; Contractor et al. 2015). Lastly, creating reliable gridded observed precipitation analyses in mountainous regions is especially difficult due to strong spatial precipitation gradients induced by the topography, to gauge undercatch biases resulting from wind effects, to the predominance of snowfall at high elevations, and to sampling location biases, as most stations in mountainous regions are in valleys.

Some of the known sampling biases can be reduced by using multi-sensor precipitation estimates. One example is the Stage IV dataset (Crosson et al. 1996, Fulton et al. 1998), which combines observations from a nationwide radar network with surface observations. While the Stage IV dataset is often considered the gold standard for sub-daily precipitation evaluation over the United States, it also has known deficiencies, especially over complex terrain (e.g., Nelson et al. 2016). Satellite observations can provide more global coverage at high spatial and temporal resolution, but like Stage IV, satellite-based precipitation products have large uncertainties in regions with a complex topography (e.g., Derin and Yilmaz 2014, Sun et al. 2018). More recently, there has been a trend toward creating precipitation datasets by combining different methods and observation types (Zhang et al. 2018). One example is that of regional reanalyses that can improve the depicted spatial characteristics of precipitation in data-sparse regions (e.g., Dahlgren et al., 2014, Prein and Gobiet 2017). Another example is a product like the Multi-Source Weighted-Ensemble Precipitation (MSWEP; Beck et al. 2017) dataset, which combines information from rain gauges, satellites, reanalyses, and streamflow measurements.

1.3.2 Computational Needs

While it has been established that CP model configurations outperform coarser-resolution ones for simulating intense precipitation events, this comes at the price of substantially-increased computational needs. Halving the grid spacing of a model increases the computational costs by at least a factor of eight when the same area is covered. This reflects a doubling of the number of grid points in each of the two horizontal (X-Y/lat-lon) directions, compounded by a doubling of the number of model integration steps needed to produce a forecast of the same length (as the time step must be halved due to the smaller Δx). Balancing available computational resources

with necessary model resolution and simulation length/ensemble size then becomes a challenging optimization problem.

The WRF Model features linear strong scaling (Figure 1-15), which means that it takes advantage of increasing computer cores very efficiently and completes a simulation in nearly half the time when doubling the available compute resources (viz., cores). An estimate of relative computational resources needed (core-hours) when decreasing the model horizontal grid spacing compared to a 4-km (2.5-mi) simulation, leaving all other model components the same, is shown in Table 1-3. For example, running the same simulation at Δx=250 m instead of Δx=4 km (2.5-mi) requires at least 4096 times more compute resources. This is equivalent to being able to simulate a 4096-member ensemble at Δx=4 km (2.5-mi) for the same costs as one Δx=250 m (880 ft) simulation.

Table 1-3   Increase in Computational Costs (Core-Hours) When Decreasing the Horizontal Grid Spacing And Halving the Time Step, Assuming Perfect Linear Scaling, With Respect to a Δx=4 km (2.5 mi) Reference

Δx	12 km 7.5 mi	4 km 2.5 mi	2 km 1.2 mi	1 km 0.6 mi	500 m 1640 ft	250 m 880 ft
Relative Computational Resources	0.04	1	8	64	512	4096

Several studies have found that coarse CP model ensembles can provide better guidance than deterministic forecasts with higher resolution (e.g., Hagelin et al. 2017; Loken et al. 2017; Mittermaier and Csima 2017; Schwartz et al. 2017). Moreover, both Clark et al. (2011) and Schwartz et al. (2014) noted that CP ensembles with only 10 forecast members can provide adequate guidance and, furthermore, that there is a point of diminishing returns at which adding additional ensemble members is not beneficial. Ultimately, these findings could imply more-constrained computational demands for ensemble forecasting of intense precipitation.

Another resource issue in the intense precipitation event modeling is that of data volume. The tremendous outputs from km-scale runs make data archiving and sharing a challenge, especially for large ensemble systems or for long climate integrations. For such applications, it is often necessary to be selective in saving output variables. One approach to mitigate the problem is online processing and analysis during model integration; this can significantly reduce output volume and obviate the need for separate post-processing work. A strategy of simply considering reruns of simulations, via saving model input and restart files, can also be a way to reduce the long-term storage needs.
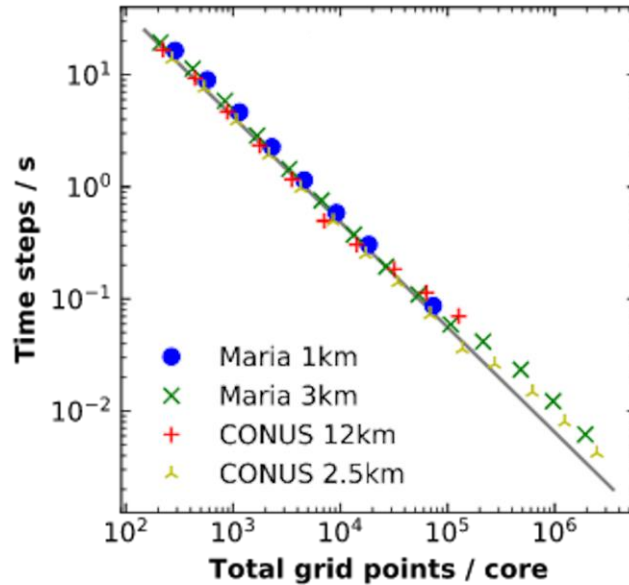
Figure 1-15 WRF Scaling Results From Four Different Simulations: Hurricane Maria at 1-km And 3-km (0.6-mi and 1.8-mi) Resolutions and the Official CONUS Benchmarks for WRF Used for Computational Estimations (http://www2.mmm.ucar.edu/wrf/wg2/bench/) at 12-km and 2.5-km (7.5-mi and 1.5-mi) Resolution

1.3.3 Model Physics

The performance of model physics packages can be critical in the simulation of intense precipitation events. As noted above, CP model configurations have the advantage of avoiding uncertainties introduced by convective parameterizations, which are a key source of error (e.g., Déqué et al. 2007, Mooney et al. 2017), especially for heavy precipitation settings (Bruyère et al. 2017). In CP simulations other model physics areas gain in importance, with microphysics (see, e.g., Schwartz et al. 2010) and PBL schemes looming large.

Generally, more complex microphysics that represent ice processes and that simulate both mass mixing ratios and number concentrations of particles (known as 2-moment schemes) outperform simpler, single-moment schemes (Wang 2002; Bryan and Morrison 2012).[4] However, the more complex schemes have more parameters, and these are often poorly constrained by observations. This makes the schemes highly tunable, introducing uncertainties in the resultant weather and climate simulations (Hourdin et al. 2017), which are sensitive to the chosen tunings. Other microphysics approaches such as spectral bin schemes (e.g., Lynn et al. 2005, Wallace and Hobbs 2006) or Lagrangian schemes (e.g., Andrejczuk et al. 2008, 2010) show promising results due to more physically based assumptions, but are computationally more expensive, which limits their application in weather forecasting and climate modeling.

Another challenge in CP modeling is the representation of turbulent processes in the boundary layer and in convective clouds (Wyngaard 2004, Prein et al. 2015). Scale-aware

---

[4] Simpler microphysics schemes only represent the mass mixing ratios of hydrometeors, such as rain, expressed in kg of precipitate in a kg mass of air (units: g/kg or kg/kg). The more complex, 2-moment schemes also carry the number concentration of ice particles per unit volume of air (no./m^3, no./cm^3).

parameterizations are under development (e.g., Shin and Hong 2015, Brast et al. 2018), but their application to intense precipitation events has not yet been systematically studied.

For operational weather forecasting and climate modeling, understanding of model physics combinations that operate well under a large range of conditions are needed. Several groups have identified model settings for skillful simulations of a variety of heavy-rainfall-producing systems such as tropical cyclones, MCSs, and atmospheric rivers. Table 1-4 summarizes model setups from CP modeling efforts over North America using the WRF Model. One can see some similarities in the physics packages used.

Table 1-4   Selected Model Physics Settings in CP Weather Prediction and Climate Simulations in North American Applications

Modeling effort	Model	Δx	Microphysics	PBL	Radiation	Land Surface	Reference
NCAR Ensemble Forecast	AR-WRF V 3.6.1	3-km (~1.8-mi)	Thompson (Thompson et al. 2008)	MYJ (Janjic 1990, 1994)	RRTMG (Mlawer et al. 1997)	NOAH	Schwartz et al. (2015, 2019)
North-American CP-climate simulations	AR-WRF V3.4.1	4-km (2.5-mi)	Thompson (Thompson and Eidhammer 2014)	YSU (Hong et al. 2006)	RRTMG (Mlawer et al. 1997)	NOAH-MP	Liu et al. 2017
High-Resolution Rapid Refresh (HRRR)	AR-WRF	3-km (~1.8-mil	Thompson (Thompson et al. 2008; Thompson and Eidhammer 2014)	MYJ (Janjic 1990, 1994)	RRTM (Mlawer et al. 1997)	RUC	Benjamin et al. 2016

Figure 1-16 summarizes the impact of model physics on uncertainties in simulating heavy precipitation events featuring embedded convection. For non-CP simulations (Δx>4 km, 2.5 mi), uncertainties originating from the convective parameterization dominate (e.g., Mooney et al. 2017). In contrast, at CP scales (4 km>Δx>500 m; 2.5 mi>Δx>1640 ft) microphysics and PBL schemes gain in importance and are the greatest sources of uncertainty. The uncertainty associated with the PBL scheme decreases when approaching large-eddy simulation scales (Δx<250 m, 880 ft) since turbulent eddies are increasingly resolved. Note that this is a schematic representation and that uncertainty sources are generally case dependent.

Other sources of uncertainty such as those stemming from model numerics and initial and boundary conditions are also important (e.g., Gallus and Bresch 2006). While numerics and physics play a major role in correctly simulating the processes of a particular precipitation system, initial condition uncertainty often dominates over physics and lateral boundary uncertainty, especially in the spatial placement of systems in short-term forecasts (Hohenegger et al. 2008; Vié et al. 2011; Peralta et al. 2012; Kühnlein et al. 2014; Romine et al. 2014; Zhang 2019). Thus, ensembles with initial condition diversity can be powerful tools for intense precipitation forecasting.
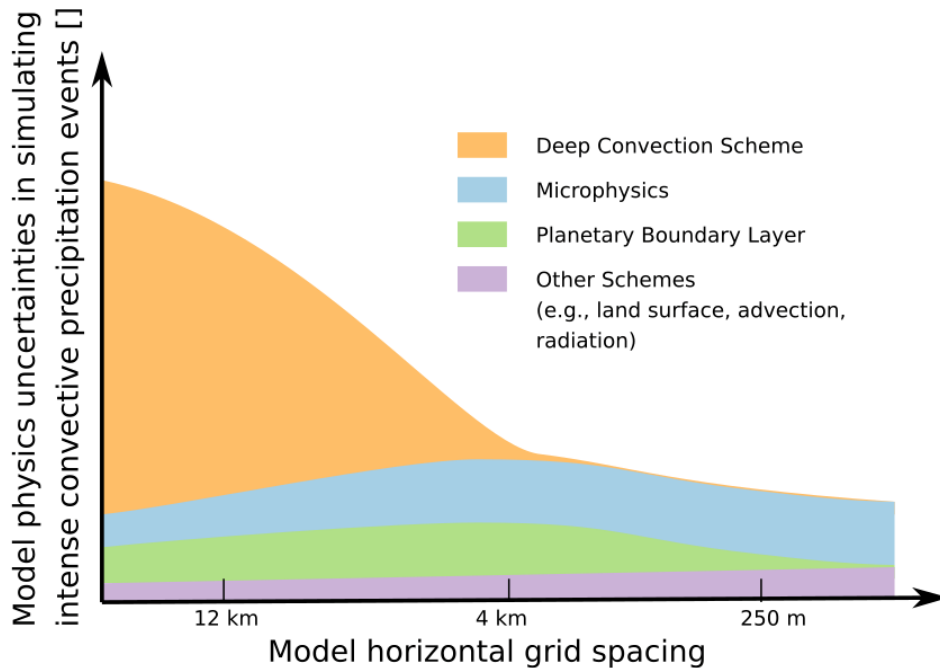
Figure 1-16 Sources of Uncertainty in the Simulation of Intense Precipitation Events Featuring Embedded Deep Convection

1.3.4 Model Numerics

The approach to the discretization of the equations of motion is another model building block that can impact the distribution and intensity of simulated precipitation. Numerical diffusion is applied to the model's prognostic variables and has a smoothing impact on atmospheric fields. This affects these quantities on scales of two to eight times the horizontal model grid spacing (Skamarock 2004, Skamarock et al. 2014). While serving a computational purpose, numerical diffusion has some undesirable effects and imposes some increase in computational cost. While the use of higher-order numerical schemes can mitigate the problems (Ghosal 1996, Weller and Weller 2008, Ogaja and Will 2014), they present their own issues and costs.

Certain specifications in model dynamics can improve numerical stability, and this is particularly important in regions with complex topography where the application of finer grids requires concomitant time step reductions. It has been found that the application of refined numerical schemes can significantly improve the numerical accuracy and stability of flows in complex terrain and can improve simulations over areas of steep slopes (e.g., Zängl 2012, Zängl et al. 2015). Likewise, the formulations for advection in models are important to forecast accuracy and realism. Skamarock and Weisman (2009), for example, showed how the implementation of positive-definite moisture advection can reduce biases in CP forecasts, illuminating the relevance of this aspect of numerics to heavy precipitation modeling.

Lastly, a model's numerics can significantly alter simulation sensitivity to model physics. Gallus and Bresch (2006), for example, found that the influence of changing model numerics on the precipitation volume and area covered by convective storms can be even larger than changes in the set of model physics schemes used.

1.3.5   Model Ensembles: Initial Conditions and Spread

Given the inherent uncertainty regarding small-scale processes and the fact that small-scale errors grow faster than large-scale errors, there is an argument that convection-allowing precipitation forecasts should be presented probabilistically. This is especially true when one is concerned with the accurate placement of precipitating systems. While postprocessing techniques can derive probabilistic information from deterministic forecast output (e.g., Theis et al. 2005), CP ensembles are a more natural way to quantify forecast uncertainty. This uncertainty is reflected in forecast spread that can originate from perturbations to initial conditions, boundary conditions, and physics, with a combination of errors from all three sources typically yielding the best forecasts (e.g., Peralta et al. 2012; Romine et al. 2014).

CP ensembles have been shown to provide valuable forecast guidance for heavy precipitation and severe weather episodes (e.g., Evans et al. 2014; Schwartz et al. 2019), and CP ensembles— despite their computational expense— are becoming operational at most major weather forecast centers. However, most CP ensembles lack spread, and often an observed precipitation event will fall outside of the ensemble envelope (Duc et al. 2013; Tennant 2015; Hagelin et al. 2017; Schwartz et al. 2014, 2015, 2017, 2019, Raynaud and Bouttier 2016). Thus, further efforts are needed to increase ensemble spread in CP ensembles, including representations of model error (e.g., Romine et al. 2014; Jankov et al. 2019) to improve spread characteristics and yield statistically reliable (i.e., well-calibrated) precipitation forecasts.

1.4   Section Summary

This section has surveyed and summarized the body of literature on the simulation of intense precipitation with numerical atmospheric models. We have examined published works across disciplines, with the studies reflecting a variety of approaches, such as idealized simulations, physical process investigations, case studies, forecast reviews, and climate projections. We have focused on the ability of models to capture atmospheric conditions and processes involved in heavy rain events. In the tropics and subtropics, intense rainfall is predominately caused by tropical cyclones, while in the middle and higher latitudes mesoscale convective systems (MCSs), frontal systems attending extratropical cyclones, and atmospheric rivers are dominant. Also, across latitudes, orographic forcing can be a factor in intense precipitation episodes.

The literature review has found that, in general, convection-permitting (CP) configurations of models improve the simulation of convective precipitation compared to coarser-resolution implementations (i.e., those that rely on convective parameterization), regardless of the atmospheric forcing. As to the upper limit on resolution for simulation success, several studies have identified ~4-km (~2.5-mi) horizontal grid spacing as the minimum needed for explicitly simulating convection and its effects (Weisman et al. 1997; Gentry and Lackmann 2010; Prein et al. 2019). Results published to date show that further decreasing model grid spacing does improve the realism of the simulated convective elements, such as updrafts/downdrafts (Wang et al. 2019) and entrainment (Lebo and Morrison 2015), but how this translates consistently into improved quantifications of the evolutions, durations, and distributions of heavy precipitation is less clear. One understanding that has emerged, however, is that grid spacings of 3–4 km (2–3 mi) tend to overestimate intense precipitation rates (e.g., Ban et al. 2014, Herman and Schumacher 2016), with the situation being improved by the application of grids with finer resolution (Schwartz et al. 2017).

For terrain-influenced events, climate simulations focusing on the Rocky Mountains suggest that heavy orographic precipitation can be captured by models with grid spacings on the order of

~10 km (6 mi) for non-convective storms (Rasmussen et al. 2011, Prein et al. 2013, Letcher and Minder 2015). However, in other regions, case studies have shown that at least 4-km (2.5-mi) grid spacing is required to capture precipitation patterns in mountainous regions (Garvert et al. 1997; Colle and Mass 2000; Colle et al. 2005; Garvert et al. 2005; Hart et al. 2005; Buzzi et al. 2014; Schwartz 2014; Bartsotas et al. 2017; Gowan et al. 2018). Moreover, high vertical resolution is also essential to capture orographically-forced gravity waves that can modulate intense precipitation (Colle and Mass 2000).

Realistic model physical process schemes are fundamental to capturing heavy precipitation events, with microphysics being particularly important in CP simulations. Microphysics schemes that account for frozen particles have repeatedly been shown to improve intense precipitation simulations (Wang 2002; Bryan and Morrison 2012; Singh and O'Gorman 2014). While there is less consensus on whether more complex microphysics schemes (i.e., two-moment schemes) outperform simpler schemes (one-moment schemes; Van Weverberg et al. 2014; Morrison et al. 2015; Feng et al. 2018), the higher-order schemes do more fully describe conditions that are better-addressed as model grid scales become finer. Although they are described as explicitly generating precipitation, microphysics schemes do, in fact, rely on numerous parameters, which introduce forecast uncertainly and error. The literature, for example, makes clear that microphysics parameters, such as those that control raindrop breakup (Van Weverberg et al. 2014) or particle fall speeds (Singh and O'Gorman 2014), can have large effects on simulated intense precipitation. Lastly, model microphysics can also influence the duration of precipitating storms (Feng et al. 2018). Other model physics components, such as PBL schemes, radiation schemes, and land surface models, tend to be less important in precipitation response than microphysics, at least in wintertime settings (Liu et al. 2011). For warm-season events, however, PBL scheme sensitivity can be larger than microphysics sensitivity (Efstathiou et al. 2013). Atmospheric models have multiple dimensions and many degrees of freedom, and the literature shows that an additional source of uncertainty stems from the interactions between model physics and numerics. Thus, changes in aspects of numerics can alter model sensitivity to physics configuration and can significantly affect the precipitation characteristics of convective storms (Gallus and Bresch 2006).

In summary, while there are sensitivities to model grid configurations, physics, and numerics, the literature makes it clear that realistic and accurate simulation of intense precipitation can be achieved with CP-configured models. Furthermore, such fine-scale modeling has been successful in capturing the recognized set of precipitation-forcing processes and conditions and has done so across varied geographic, meteorological, and climatic regimes. Thus, the science has progressed beyond merely running a model at high resolution on an intense precipitation case study, seeing that the event is simulated "reasonably well", and declaring success. The work now lies in identifying how to improve models for the intense precipitation problem, how to exploit their output for hydrometeorological and engineering needs like design storm analysis, and how to best apply them for probabilistic forecasting and risk assessment.

Based on the picture presented by the current literature and on our own atmospheric modeling experience, we see the following issues and opportunities regarding improvement of the simulation and analysis of intense precipitation with atmospheric models.

- Many studies focus on assessing uncertainties in modeling intense precipitation events but missing is a holistic analysis that compares the full range of uncertainty sources such as model numerics, resolution, physics, initial and boundary conditions, storm type, and observational error. This makes it difficult to understand the hierarchy of sensitivities and needs and, as a result, to efficiently improve current models.

- While the finer-scale model configurations of 2–4-km (1.2–2.5-mi) grid spacing that have been most written about can be successful and are deemed convection-permitting, they are not truly resolving most convective motions and processes. Thus, the next readily-reachable frontier for intense precipitation modeling is that of 1-km (0.6-mi) and sub-km grid spacings. This pushes into large-eddy-scale (LE) territory, which will present challenges in the representation and handling of turbulence and in the possible breakdown of model physics schemes, such as for the PBL, which were constructed for larger-scale representations. Thus, the development and testing of scale-insensitive model physics packages for high-resolution investigation of the intense precipitation problem would ideally be an element of the work. Scale-insensitive schemes, particularly those for microphysics and the PBL, should function seamlessly from the microscale to the LE scale. After investing in greater compute resources to generate sub-km-scale simulations of intense precipitation, systematically assessing forecast performance will be important, in part due to known deficiencies in the realization of turbulent and boundary layer processes in CP configurations.
- The creative application of ensembles is another avenue in intense precipitation prediction and analysis. For this problem, however, high-resolution ensemble modeling presents both practical and theoretical questions. One is finding a balance between high model resolution and adequate ensemble size, with each being important in either realistically capturing IP events or in optimally projecting the uncertainties of the IP situations. A challenge here, again, is presented by the compute resources demanded by covering both the resolution and ensemble size aspects. Advancements could be made by adopting alternative strategies such as targeted downscaling of high-impact events (e.g., Schwartz and Sobash 2019; Hall 2019), instead of, for example, generating a volume of daily forecasts or transient climate simulations.
- Although CP models can simulate intense precipitation events, it is not straightforward how to apply the output of these tools in the design of critical infrastructure (e.g., nuclear power plants, hospitals, airports) due to typically short observational records. Such limited records make it likely that high-return-value events (e.g., the thousand-year event) have not yet been observed. Thus, model predictions may be questioned if they significantly exceed historical records, but they may still be possible and realistic. Additionally, a return-value assessment based on historical data is not representative of future risks, due to the projected intensification of intense precipitation under climate change (e.g., Wright et al. 2019). However, numerical models are flexible and do allow to provide a best estimate (e.g., mean or median) of heavy precipitation events along with a quantitative estimate of the uncertainty surrounding that best estimate. Examples are using a storyline approach (Shepherd et al. 2018) or doing targeted downscaling of high-impact rainfall events from large-ensemble global climate model datasets (Hall 2019, Fujita et al. 2019).

2   DEMONSTRATION OF THE USE OF CONVECTION-PERMITTING NUMERICAL MODELS FOR ESTIMATING INTENSE PRECIPITATION

In this section, we summarize the forecast performance of three convection-permitting (CP) numerical weather prediction (NWP) model ensembles in simulating heavy precipitation events of recent years in the eastern United States. The analysis is performed according to the experimental design and evaluation strategy outlined in Sec. 2.1 . The primary questions are as follows.

- How well can CP models simulate heavy precipitation events, as evaluated by high-resolution, multi-sensor precipitation observations?
- How do model biases compare to the differences between multi-sensor and station-based precipitation observations?
- Can CP models capture terrain gradients driving heavy precipitation events in mountainous regions?
- How does NWP model skill depend on event intensity and seasonality?
- Is there a significant improvement with 1-km grid spacing simulations compared to 3-km runs (i.e., 1.8-mi grids v. 0.6-mi grids)?

To answer these questions, we collected a range of high-resolution gauge and multi-sensor precipitation datasets. Our primary interest is how well the simulations capture the peak accumulations and locations of maxima, and secondarily the spatial extent and total volume of precipitation. The analyses are performed in four designated regions that have varying heavy precipitation-producing processes— southern U.S., central U.S., Appalachian Mountains, and East Coast

2.1   Experimental Design and Evaluation Strategy

2.1.1   Datasets and Analysis Region

2.1.1.1   Simulations

We leverage three existing convection-permitting model forecast datasets that were produced by NCAR. These datasets capture multiple high-impact flood events in the U.S. The target datasets have been used in published work, and Table 2-1 summarizes their characteristics. These datasets represent 10,570 36-hour WRF simulations/forecasts at 3-km (1.8 mi) horizontal grid spacing (Δx=3 km; 1.8 mi) and 810 36-hour simulations at 1-km (.6 mi) horizontal grid spacing (Δx=1 km; 0.6 mi). Both of these setups employed 40 vertical levels with a model top at 50 hPa. In addition to atmospheric state variables output on discrete pressure levels, they contain hourly surface meteorological and diagnostic fields, such as accumulated precipitation, temperature, convective available potential energy (CAPE), and convective inhibition (CIN). All of these datasets are readily available to us for analysis.

As shown in Table 2-1, several published studies have evaluated precipitation in these datasets, and their performance for heavy rain events has been noted. However, the simulations have not yet been evaluated for their ability to capture high-return-period, flood-producing events, a focus of this effort.

Table 2-1   Convection-Permitting Forecast Datasets That Allow The Evaluation of Simulated Intense Flood Events

Dataset	Δx	Elements	Period	Region	References
NCAR Real-time Ensemble	3 km (1.8 mi)	10-member ensemble forecasts	5/1/2015-12/31/2017	CONUS	Schwartz et al. (2014, 2015a, 2015b), Romine et al. (2014)
NCAR MPEX Ensemble	3 km (1.8 mi) & 1 km (0.6 mi)	10-member ensemble forecasts	5/15/2013-6/15/2013	Central / eastern U.S.	Schwartz et al. (2017)
NCAR Severe Weather Study	3 km (1.8 mi) & 1 km (0.6 mi)	Deterministic forecasts; 500 cases	2010-2017	Central / eastern U.S.	Sobash et al. (2019), Schwartz et al. (2019)

2.1.1.2   Observations

While the CONUS is observationally data-rich (relative to many other regions), with this meteorological information supporting detailed CPM evaluations in intense precipitation events, there are still large observational uncertainties in the analyses produced (see, e.g., Prein and Gobiet 2017). To mitigate this, while also attempting to understand the uncertainties, we use an ensemble of observational datasets for model evaluation.

Representing multi-sensor datasets, we use the Stage-IV (Crosson et al. 1996, Fulton et al. 1998) and the Multi-Radar/Multi-Sensor (MRMS) data (Zhang et al. 2016). Both of these rely on the Weather Surveillance Doppler radar (WSR-88D) network in combination with surface precipitation measurements to derive precipitation estimates with high spatiotemporal resolution. The PIs have extensive experience using this dataset for model evaluation (e.g., Prein et al. 2017c; Schwartz et al. 2015, 2019).

In addition to these multi-sensor measurements, we use two additional station-based gridded observational datasets to estimate the uncertainties in heavy daily accumulations. The primary input for these datasets are point observations at precipitation gauges, which are interpolated onto a grid. Furthermore, digital elevation data in combination with vertical lapse rate assumptions are used to improve the spatial representation of precipitation fields in areas with complex topography.

The first station-based gridded observational datasets is the PRISM daily dataset from the University of Washington (Daly et al. 1994, 2002, 2008), which incorporates nearly 13,000 rain gauges from the CONUS over the period 1981–present. Compared to other gridded precipitation datasets, PRISM has improvements, particularly in mountainous and coastal areas of the western U.S. (Daly et al. 2008). PRISM has an advantage over the Stage-IV and WSR-88D datasets in the Rocky Mountains, as the latter observations typically have large errors in mountainous regions due to radar beam blocking. The second additional data source is the Newman precipitation dataset (Newman et al. 2015). This dataset has a similar station density

to PRISM, but it consists of a 100-member ensemble, thus allowing for the calculation of uncertainties in extreme precipitation estimations.

For a more process-based analysis, we apply the ERA-5 reanalysis to evaluate the simulations of the events' large-scale weather patterns. ERA-5 assimilates a wide range of observations to constrain a model in its depictions of atmospheric conditions. ERA-5 allows the assessment of WRF's ability to simulate quantities like CIN, CAPE, and integrated moisture flux, which are important factors in intense precipitating storms. A summary of observational datasets is shown in Table 2-2.

Table 2-2   Observational and Reanalysis Datasets Used for Model Evaluation *Footnotes provide links to download these datasets.*

Dataset	Variable	Time period	Temporal resolution	Horizontal resolution
Stage-IV[1]	precipitation	2001-present	1-hour	4 km (2.5 mi)
Multi-Radar/Multi-Sensor (MRMS)[2]	precipitation, reflectivity	2011-present	5-minutes	1 km (0.6 mi)
PRISM[3]	precipitation	1981-present	daily	4 km (2.5 mi)
GMET (100 members)[4]	precipitation	1980-2016	daily	12 km (7.5 mi)
ERA-5[5]	3D atmosphere	1979-present	hourly	31 km (19 mi)

[1] https://data.eol.ucar.edu/dataset/21.093
[2] https://mrms.ncep.noaa.gov/data/
[3] https://prism.oregonstate.edu/
[4] https://ncar.github.io/hydrology/models/GMET
[5] https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels

2.1.1.3   Analysis Regions

We have divided the eastern CONUS into four regions (Figure 2-1) that feature common climatic or topographic characteristics and in which intense precipitation events tend to reflect similar processes. The "South" includes the Gulf Coast and South Atlantic states. Storm systems, fed by tropical and sub-tropical moisture from the Gulf and Atlantic (e.g., hurricanes, mesoscale convective systems) are the main causes of intense precipitation events in this area. The "Mid CONUS" area encompasses the Great Plains and Midwest, regions in which mesoscale convective systems are the dominant rain producers. The "East Coast" region is that east of the Appalachians, and its heaviest precipitation events can result from tropical storms, MCSs, or frontal systems. The "Appalachians" region follows the main crest of the mountains and was selected to evaluate CP models' ability to capture orographically-enhanced precipitation situations.
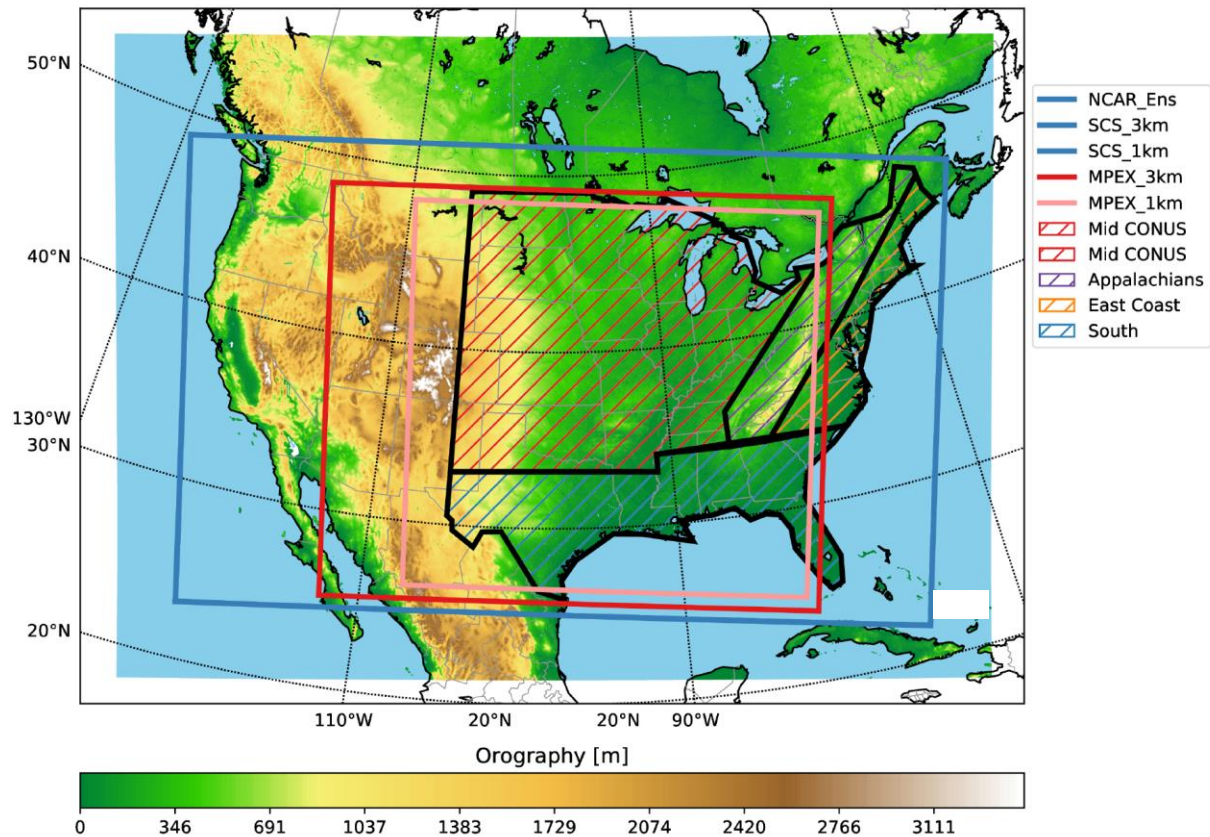
Figure 2-1 **Computational Domains (Colored Rectangles), Sub-Regions (Colored Hatched Areas Bounded by Black Lines), and Orography (m) (Shaded Background, Scale at Bottom) over the CONUS** *Note that the NCAR Ensemble, SCS 3 km (1.8 mi), and SCS 1 km (0.6 mi)simulations are performed on the same domain.*

2.1.1.4   Representativeness of Database Period

To assess the representativeness of the days where simulations exist (our "database period"; see Table 2-1) compared to the full Stage IV record (2002–2018), we calculate the peak precipitation accumulation (i.e., accumulated over storm lifetime) for each storm system in the CONUS, based on the tracking of hourly MCSs precipitation. Storm systems are defined as continuous precipitation regions (in space and time) with precipitation rates greater than 5 mm/h (0.2 in/h). Only systems that have a minimum size— defined here as the area of heavy precipitation accumulated over the storm lifetime— of 48,000 km^2 (18,533 mi^2) are considered; this is to remove small or weak systems from the analysis and to speed up the computation. This threshold was found to reliably identify mesoscale convective rainfall areas in published analyses that looked at varied precipitation thresholds and minimum rainfall areas (Clark et al. 2014, Prein et al. 2017b).

Figure 2-2a shows the peak accumulation of all tracked storms in the Stage IV dataset covering the period 2002–2018. We note that Stage IV records include some spurious data in the Western U.S. due to limitations with radar-derived estimates in complex terrain (e.g., related to

beam blocking), and therefore some of the accumulations in this region might be overestimated. Events with accumulations higher than 250 mm (10 in) can occur in any region of the eastern U.S. but are most frequent (threshold exceeded approximately 10 times a year) along the Gulf Coast and southern Atlantic coastline. There is also a high frequency of storms with large accumulations in the central U.S. east of the Mississippi.

Figure 2-2b indicates that the distribution of storms on days covered by our model datasets is similar to the one using the entire Stage-IV record, and the same hotspots appear. Many events with accumulations larger than 250 mm (10 in) are included in our forecast datasets and evaluating them will allow a statistically robust assessment of model skill in simulating heavy precipitation episodes.
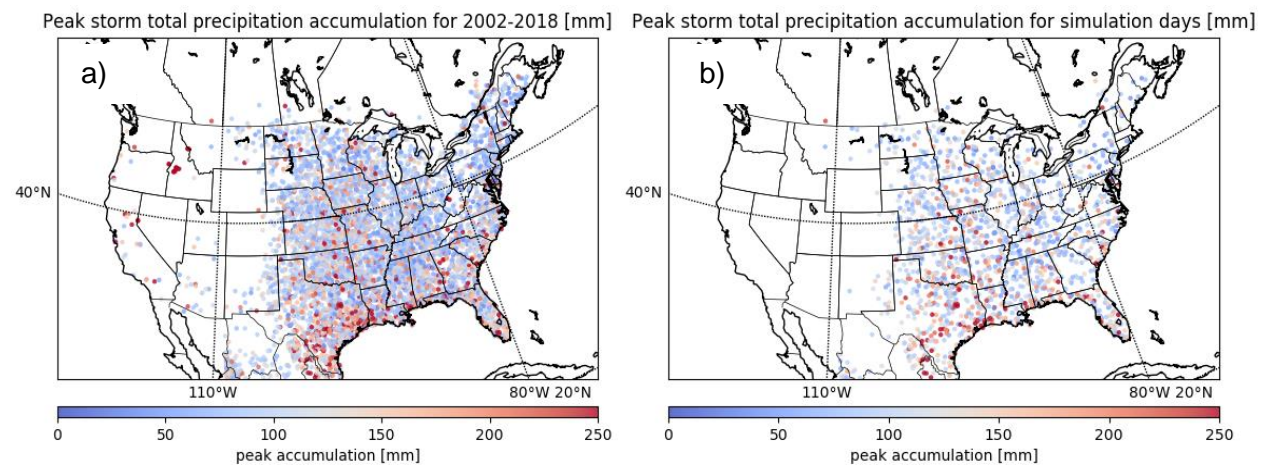
Figure 2-2   Peak Accumulation Intensity (Color) and Location for Each MCS in Stage IV for the Period 2002–2018 (a). (b) Similar to (a), but Only Showing Peak Accumulations for Storms Covered by Model Datasets to be Used (See Table 2-1)

Figure 2-3 shows the peak accumulations for the entire Stage IV record for each analysis region shown in Figure 2-1. The days covered by our database include many of the events with the highest accumulations in the South region. Also, the Mid CONUS region includes storms with peak accumulations that are representative of the entire Stage IV record. Our data coverage period, however, does not include the events with the highest accumulations in the Appalachian and East Coast regions. Still, there are many episodes in the period that exceed 150 mm (6 in) accumulation, and this set will allow for a robust assessment of model skill in these regions.
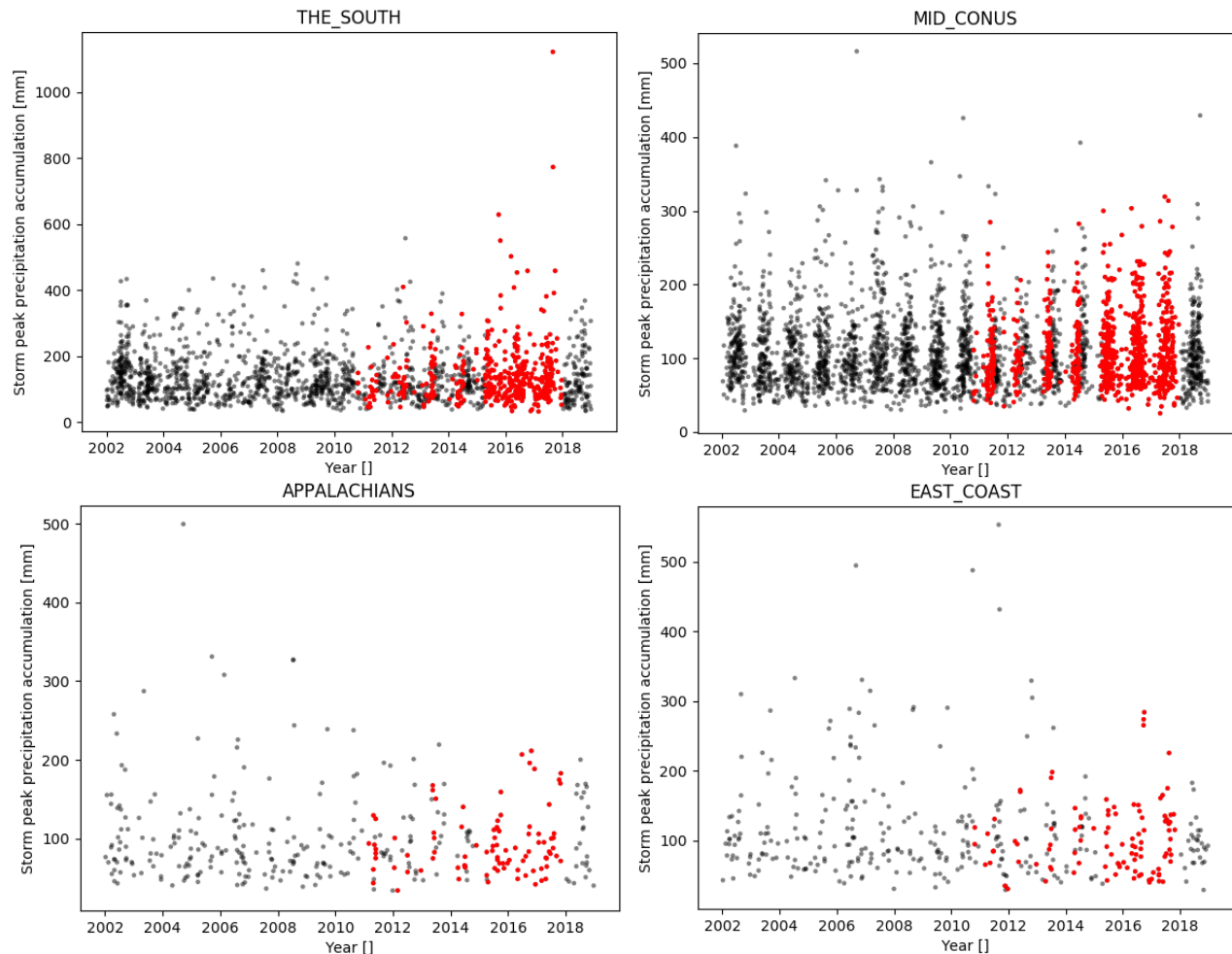
Figure 2-3   MCS Peak Accumulation in South, Mid CONUS, Appalachian, and East Coast Regions *Black dots show all MCSs in Stage IV, and red dots denote the events covered in our database of forecasts.*

2.1.2   Case Study Selection

To identify candidates for heavy precipitation case studies, we reviewed Stage IV data and noted the highest daily accumulations east of the Rockies (i.e., east of approx. 105 W) occurring in our model database period. It became apparent that the highest accumulations were caused by landfalling tropical cyclones, these favoring the South region. Thus, to cover a variety of meteorological settings and to prevent the analysis from being dominated by tropical systems, we divided the full eastern CONUS domain into four regions (Figure 2-1). The event rankings were then made separately for each area. This allowed precipitation events from a variety of forcing conditions (e.g., cold fronts, upper-level lows, MCSs, and orography) to populate the rankings.

We smoothed the Stage IV data with a 64 km x 64 km (40 mi x 40 mi) square filter to focus on mesoscale precipitation patterns. Each grid cell was averaged with its 256 neighbors. Cells with

missing accumulation (i.e., outside the coverage of the WSR-88D network) were assigned zeros before averaging. This smoothed out the largest accumulations at the convective scale and ensured that we captured events with enough rainfall volume to affect a modest river catchment. Tests using a 4-km (2.5-mi) filter length (i.e., no smoothing) and a 500-km (311-mi) filter length resulted in the selection of events similar to those identified from the mesoscale, 64-km (40-mi) smoothing. Table A-1—Table A-4 and Figure A-1—Figure A-4 show the top 20 maximum daily precipitation accumulations for each region.

2.1.3 Evaluation Strategy

Here we describe the strategies using both Eulerian and Lagrangian frameworks for model and event evaluations that are performed in Section 2.2 .

2.1.3.1   Eulerian Framework – Catchment-Scale Precipitation Characteristics

To probe the question of how well a model predicts precipitation events, an Eulerian, or fixed-point reference frame, is used to evaluate the described 1- and 3-km precipitation forecasts (1.8 mi and 0.6 mi, respectively). We define a precipitation event as an instance of a grid cell having 24-hr rainfall exceeding a given threshold. This 24-hour period corresponds to the 12- to 36-hour lead times of the forecast. Different rainfall thresholds were applied, ranging from 1 mm to 50 mm (0.04 to 2 in).

We use two primary verification metrics: Equitable Threat Score and Fractions Skill Score. The Equitable Threat Score (ETS) evaluates a dichotomous or yes/no forecast, while the Fractions Skill Score (FSS) compares the fractional coverage of events in the forecast and observations. We analyzed model skill at progressively larger neighborhoods to accommodate different tolerances for displacement errors.

Using neighborhoods has the benefit of avoiding the "double penalty problem" of traditional grid-point methods. The double penalty problem occurs due to displacements of heavy-rainfall-producing storms in the model simulations, which results in penalizing the model performance metric twice: once for missing an observed event and once for a false alarm simulating an event at a location where none was observed.

2.1.3.1.1   Equitable Threat Score

We use the Equitable Threat Score (ETS; also called Gilbert Skill Score (GSS)) to see how well the model predicts precipitation events. A forecast *hit* is when the model correctly forecasts one or more events within the specified neighborhood. A forecast *miss* is when the model has no events in the neighborhood, while the observations have one or more events in the neighborhood. A *false alarm* is when the forecast has one or more events in the neighborhood, but the observations have none. And finally, a *correct null* is when both the forecast and observations have no events in the neighborhood. The GSS is the proportion of hits to the sum of hits, misses, and false alarms, accounting for hits due to chance. The equation for calculating the GSS is given below.
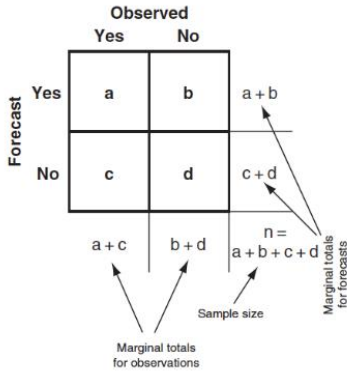
$$\text{GSS} = \frac{a/(a+b+c) - a_{\text{ref}}/(a+b+c)}{1 - a_{\text{ref}}/(a+b+c)} = \frac{a - a_{\text{ref}}}{a - a_{\text{ref}} + b + c}$$

with

$$a_{\text{ref}} = (a+b)(a+c)/n$$

Here, *a* is the number of hits, *b* is the number of misses, and *c* is the number of false alarms. *n* is the total number of forecasts, which includes hits, misses, false alarms, and correct nulls. a_{ref} is the number of hits expected due to chance, and a_{ref} increases when the event is observed with greater frequency. In other words, if the event is more common, then a forecast event will be correct by chance more often. The GSS, often used for verification of precipitation forecasts, varies between -1/3 and 1, with a larger GSS indicating better performance.

2.1.3.1.2   Fractions Skill Score

The Fractions Skill Score (FSS; Roberts and Lean, 2008) is another Eulerian verification metric, one that looks at fractional event coverage. Instead of hits, misses, and false alarms, FSS looks at the degree of agreement between the fractional coverage of observed events f_o and the fractional coverage of forecast events, f_f, within a neighborhood. Figure 2-4 shows an example of fractional coverage for hypothetical observation and forecast grids.
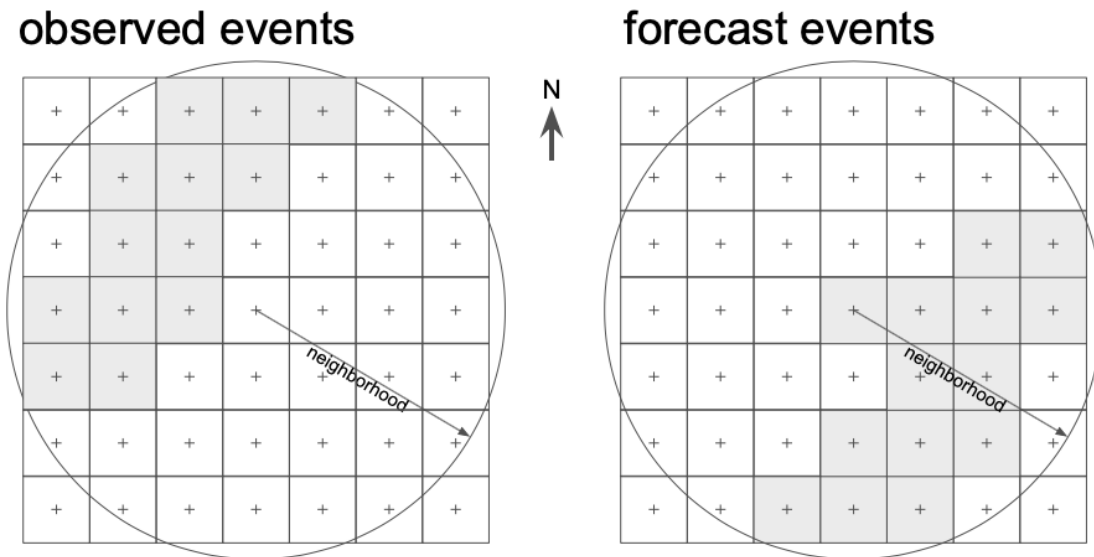
Figure 2-4   Comparison of Event Coverage (Shaded Gray) for a Hypothetical Observation (Left) and Forecast (Right). In This Example Forecast Events are Displaced to the Southeast of the Observations, but Still Reside in the Same Circular Neighborhood of Specified Radius *The fractional coverage for observations, f_o, is 13/45 (.289), and the fractional coverage for the forecast, f_f, is 14/45 (.311).*

In our example, the forecast events are too far east and south, but are still in the same neighborhood as the central grid cell. At the grid scale, there is no overlap between the observed and forecast events, but the fraction coverage error would be very low because the displacement error is smaller than the neighborhood size. As the FSS metric applied here, we utilize the *squared* fractional coverage error.

FSS is the mean squared error (MSE) of fractional event coverage normalized by the MSE of the worst-possible forecast and subtracted from 1.

$$\text{Fractions skill score (FSS):} \quad 1 - \frac{\overline{(f_f - f_o)^2}}{\overline{f_f^{\,2} + f_o^{\,2}}}$$

The mean is taken over all grid cells and events for a specific neighborhood size (see Figure 2-4). FSS can be as low as 0, for the worst-possible forecast, and as high as 1, for a perfect forecast. FSS is a function of the neighborhood size. As the neighborhood gets larger, the FSS increases.

It is easy to be better than the worst possible forecast, so a couple of reference values for FSS are used in the literature (e.g., Roberts and Lean 2008). One reference score is the FSS from a random dichotomous forecast, denoted FSS_{random}. FSS_{random} is equal to the observed base rate (i.e., the climatological occurrence frequency of events above the threshold). The other reference score, often used as a benchmark for FSS, is the uniform fraction forecast (UFSS). Once the FSS surpasses the UFSS, the forecast is said to be useful. The scale at which the FSS surpasses the UFSS is termed $scale_{min}$. This represents the smallest scale over which the forecast output contains useful information (Roberts and Lean 2008).

The FSS is similar to the GSS in several ways. Both can provide skill assessments that vary with precipitation threshold and neighborhood size (i.e., displacement of event areas). The big difference is that the GSS is based on a binary event coverage threshold. The number of events in the neighborhood is either below or above a certain value. For our GSS calculations, we use a threshold of one event. In other words, for GSS, we treat one event in the neighborhood the same as multiple events. FSS is more nuanced, quantifying differences in fractional event coverage.

2.1.3.2   Lagrangian Framework— Storms as Objects

In this study, we also use a Lagrangian framework for model evaluation. This means that we identify large-scale, coherent, hourly precipitation areas and follow them over space and time. The tracking method is similar to the method used in previous studies such as Clark et al. (2014) and Prein et al. (2017, 2020), practiced as follows.

First, we apply a spatiotemporal smoothing to hourly precipitation accumulations by using a Gaussian filter with a standard deviation of 3 in the spatial dimension and 1.5 in the temporal dimension. This is done to remove small-scale patterns (e.g., small precipitating cells, radar artifacts) from the analysis. A precipitation threshold of 5 mm h^{-1} (0.2 in h^{-1}) is applied to the smoothed precipitation data to remove weak precipitation areas. Areas with greater than 5 mm h^{-1} (0.2 in h^{-1}) precipitation rates are used to mask the original hourly precipitation rates. All grid cells with smaller 5 mm h^{-1} (0.2 in h^{-1}) are set to zero.

Grid cells that are connected (adjacent cells and diagonal cells) in space and time are identified and labeled as coherent objects. Since we are interested in heavy precipitating storms, we remove all small objects that have a spatial coverage of less than 3,000 cells (approximately 54×54 cell areas). Note that this might also remove stationary single-cell thunderstorms that can produce flash floods in small catchments, which are not the focus of this study. We use these objects to calculate storm characteristics such as translational speed, size, mean and maximum precipitation rates, and total precipitation accumulation.

An example of tracking contiguous precipitation areas is shown in Figure 2-6 for the case of the West Virginia flood event on June 23, 2016. This was produced by a long-lasting mesoscale convective system that developed along a north-west to south-east oriented frontal zone. The system became stationary (track in Figure 2-6a,b) in West Virginia causing heavy rainfall over most of the state's mountainous area for a number of hours. The severity of the event resulted from a combination of high rainfall rates, long rain duration, large spatial coverage, and orographic enhancement.

We use the Stage IV analyses as ground truth because they have high spatiotemporal resolution and cover the entire modeling period. The other observational and modeling datasets will be compared to these analyses.

We consult daily Stage IV precipitation on days with modeling output to identify the heaviest precipitation events in each of our regions (Figure 2-5). The only exception is with the MPEX simulations, which are evaluated over their entire model domain (mostly the South and Mid-CONUS regions; see Figure 2-1), since the MPEX output only covers a single month in 2013 and has a smaller domain than the other collections.

We have only considered events if they dominated the areas of precipitation accumulation on the day of occurrence (i.e., they resulted in the highest daily rainfall accumulation in the given region).

Table 2-3 presents an overview of the number of heavy precipitation events in each region and their occurrence in the observational and model datasets used. We have only considered events if they dominated the areas of precipitation accumulation on the day of occurrence (i.e., they resulted in the highest daily rainfall accumulation in the given region).

Table 2-3   Number of Heavy Precipitation Events per Region and Their Counts, per the Given Observational and Model Datasets *The first number refers to the number of events and the second to the total number of ensemble members in the case of GMET and the simulations.*

Region	Event count	STAGE IV	MRMS	PRISM	GMET	NCAR Ensemble	NCAR SCS		NCAR MPEX	
Grid spacing (km/mi)		4/2.5	1/0.6	4/2.5	12/7	3/1.8	3/1.8	1/0.6	3/1.8	1/0.6
South	32	32	22	18	22	27	8	8	0	0
Mid-CONUS	25	25	14	14	16	19	13	13	1	1
East Coast	15	15	11	10	11	13	2	2	0	0
Appalachians	13	13	8	11	10	7	6	6	0	0
MPEX	11	11	0	11	11	0	6	6	11	11
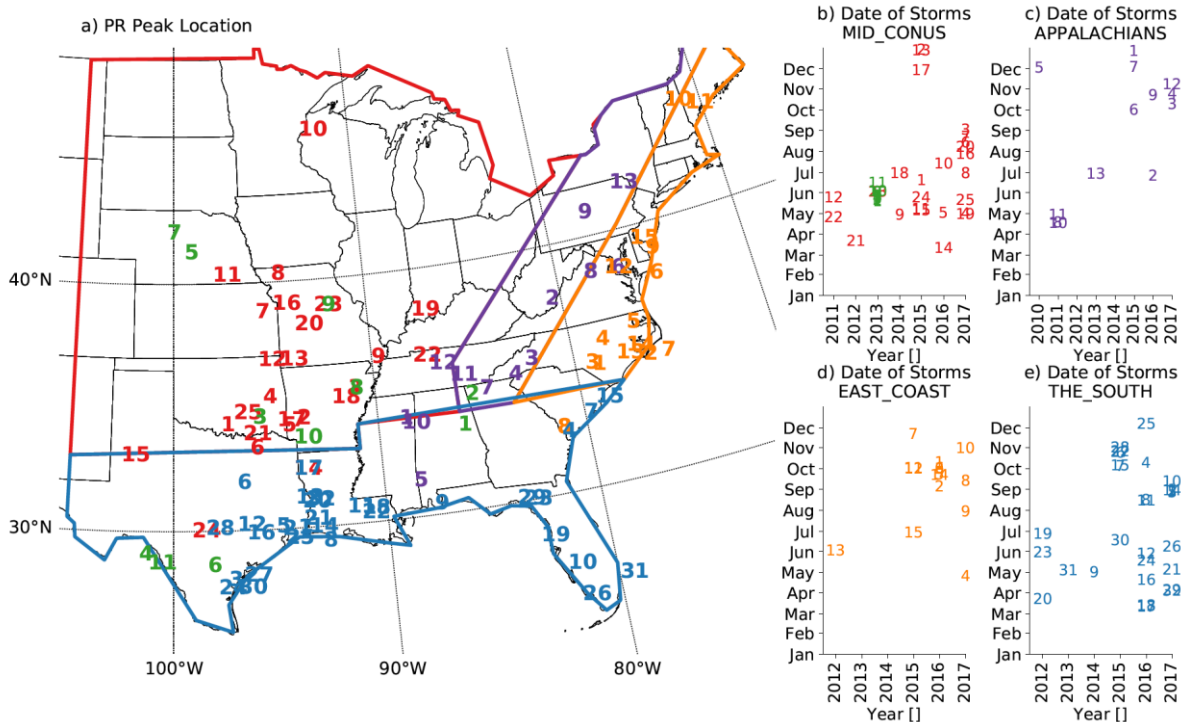
Figure 2-5   PR Peak Location (a) And Storm Dates (b-e) of Heavy Precipitation Events. Colors Show The Region That an Event Occurred in *The events are numbered according to their daily storm peak accumulation, with those denoted "1" being the heaviest.*

After selecting the events based on Stage IV observations, we conservatively regridded the other observational datasets and the simulations to the Stage IV grid. This means the 1-km (0.6 mi) simulation output was interpolated to a 4-km grid (2.5 mi), which means that the potential added value of 1-km simulations on scales smaller than 4 km (2.5 mi) cannot be assessed. However, assessing the skill on such small scales is most difficult anyway since observational datasets have large uncertainties on such small scales.

The precipitation tracking algorithm is applied to the regridded hourly precipitation of MRMS and the WRF simulations using the same setup used for tracking storms in the Stage IV dataset. The event precipitation in the daily precipitation datasets is derived by using the extent of the tracked storm from Stage IV as a mask (i.e., colored area in Figure 2-6b), setting all grid cells outside the storm area to zero value.
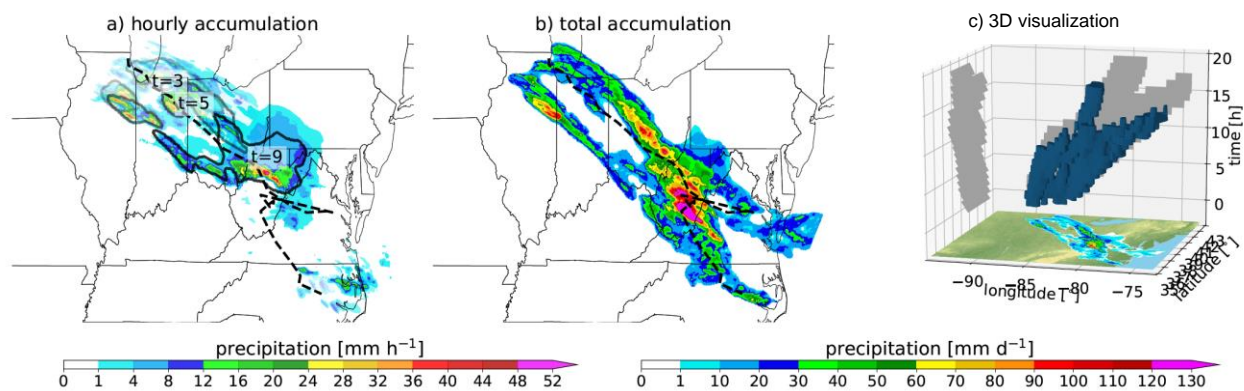
Figure 2-6   Tracking of Hourly Stage IV Precipitation Fields During the West Virginia Flood Event on June 23, 2016 *a) Hourly precipitation fields at 3, 5, and 9 hours after storm detection are shown (in increasing transparency). The perimeter of the identified precipitation object is shown in black solid contours, and the storm-track is the black dashed line. b) Event total precipitation (shaded, scale at bottom) and storm track (dashed line). c) 3D visualization of the hourly outlines of the tracked storm; the time axis is in the vertical.*

Deriving the translational speed of the tracked storm is complicated since we only have hourly precipitation accumulations from which to define the storm region, with the result that there can be substantial variations in hourly movement due to changes in the storm morphology (e.g., a disappearance of a storm component that appears to move away from the main storm). To improve the reliability of the storm speed calculation, we use three approaches and average their results: 1) speed calculated via the change in position of the center of mass of the hourly precipitation objects; 2) differences reflecting the maximization of the pattern correlation coefficient by spatially shifting the hourly precipitation patterns of two adjacent hours, with a speed corresponding to the amount of shift necessary to optimize the correlation; and 3) as in 2), but minimizing the root-mean-square error, instead of maximizing the pattern correlation.

The following storm characteristics are evaluated within the Lagrangian evaluation framework.

- **Track**: The distance between the track of the observed and simulated precipitation areas allows assessment of the skill in precipitation location.
- **Movement**: The evaluation of the movement of an area of precipitation allows one to assess if CP models can simulate stalling systems. Accurately forecasting the translation and speed of the areas is challenging since these reflect not only the synoptic wind field at mid-levels, but also mesoscale storm dynamics (e.g., cold-pool dynamics) and microphysical processes.
- **Size**: Evaluating the size of heavy precipitation areas (e.g., area of hourly precipitation >5 mm/0.2 in) allows one to assess the potential impact of the rain, which involves catchment characteristics.
- **Precipitation volume**: Correctly simulating the precipitation volume of an event is important since it is proportional to the potential runoff in a catchment. Predicting precipitation volumes requires skill in simulating precipitation rates and spatial extents.
- **Peak accumulation:** Peak accumulation is the maximum precipitation depth generated in an event. Accurately simulating the peak accumulation requires capturing the precipitating area's intensity and movement.
- **Spatial similarity**: Spatial correlation coefficients can be used to assess the quality of simulating event accumulation fields. Displacements of these simulated precipitation fields will be applied to determine the maximum possible spatial correlation. As spatial correlation coefficients are independent of mean biases in precipitation intensities, one can assess the fidelity of the simulated precipitation patterns.

These characteristics are calculated for all selected storms in each region. The presented model evaluation overview plots highlight the skill of CP models in capturing these characteristics in an ensemble framework. Forecasts found to have particularly low skill are analyzed in more detail to determine the reasons for performance deficits and the areas for model improvement.

2.1.4   Model Uncertainty Assessment

There are numerous sources of error in the model predictions of any given case, leading to uncertainty in the forecast output. These include the specifics of the model configuration and the initial conditions.

The grid spacing chosen in convection-permitting model configurations can have a pronounced impact on the simulation of intense precipitation (see, e.g., Schwartz et al. 2009). While progressively decreasing model grid spacing may better capture the regional atmospheric processes, it can also impact the simulated precipitation amounts. In this work, we will evaluate the accuracy of simulations of events where grid sizes are 3 km (1.8 mi) and 1 km (0.6 mi).

Another source of model uncertainty addressed is that of initial conditions. Thus, we use ensemble forecasts with perturbed initial conditions and assess their spread.

Furthermore, we also recognize that uncertainties in the evaluation of model results can stem from errors in the observations used for verification. That is, it is possible that error may be attributed to the model when in actuality it may stem from an error in the value used for the observational "truth".

Table 2-4 presents a summary of the uncertainty sources that are analyzed. We perform an Analysis of Variance (ANOVA) similar to Déqué et al. (2007) that provides a quantitative estimate for the sources of uncertainty in simulating intense rainfall events. This enables

insights into the reliability of CP simulations of heavy precipitation events in the four selected regions and can guide how to improve the CP configurations.

Table 2-4   Uncertainty Source Analysis

Uncertainty Source	Setting
Horizontal grid spacing (Δx)	3 km (1.8 mi), 1 km (0.6 mi)
Precipitation observations	Stage-IV (Crosson et al. 1996, Fulton et al. 1998), Mosaic WSR-88D (Zhang and Gourley 2018), PRISM (Daly et al. 1994, 2002, 2008), Newman (Newman et al 2015)
Initial Conditions	Ensemble datasets to be used reflect initial condition perturbations

2.1.5   Observational Datasets

In this study, we have combined four observational datasets and three kilometer-scale model ensemble datasets. The observational datasets are not independent, as they all are based on a set of common precipitation gauge observations. Table 2-5 provides a summary of the key characteristics of these datasets.

2.1.5.1   STAGE IV

The Stage IV dataset is a multi-sensor analysis product that provides hourly precipitation rates over the CONUS from 2001–present on a 4-km grid (2.5 mi; Crosson et al. 1996; Fulton et al. 1998). In the Stage IV product, two data sources are merged: 1) ~3,000 automated, hourly rain-gauge observations; and 2) radar estimates of hourly precipitation from the Next Generation Weather Radar (NEXRAD) system (Heiss et al. 1990). The latter estimates are derived from the Weather Surveillance Radar 1988 Doppler (WSR-88D) Radar Product Generator (Fulton et al. 1998), which creates a 131×131 grid with 4-km (2.5 mi) spacing centered on each radar location. The processed precipitation estimates going into the Stage IV dataset are bias-corrected and merged into regional values by the National Weather Service's River Forecast Centers on a 4-km (2.5 mi) CONUS-wide grid.

Table 2-5   Overview of Observational and Model Datasets Used in Analyses. ΔX Denotes the Grid Spacing of the Dataset and Δt the Temporal Resolution

Data Name	Period	Δx/Δt	Ensemble size	Source	Reference
Stage IV	2001–present	4 km 2.5 mi/ hourly	deterministic	159 stations, radar ~3,000 gauges	Crosson et al. (1996); Fulton et al. (1998)
MRMS	2014–present	1 km 0.6 mi/2.5 min.	deterministic	180 radar stations, gauges, NWP, lightning, satellite	Zhang et al. (2011)
PRISM	1982–present	4 km 2.5 mi/ daily	deterministic	13,000 gauges and radar after 2002	Daly et al. (1994, 2002, 2008)
GMET	1980–2016	12 km 7 mi/ daily	100 members	12,000 gauges	Newman et al. (2015)
NCAR Ensemble	4/7/15–12/30/2017	3 km 1.8 mi/ hourly	10 members	WRF	Schwartz (2014); Schwartz et al. (2015a,b); Romine et al. (2014)
NCAR MPEX	5/15/2013–6/15/2013	3 km 1.8 mi & 1 km 0.6 mi/ hourly	30 members at 3 km & 10 members at 1 km	WRF	Schwartz et al. (2017)
NCAR	497	3 km 1.8 mi &	deterministic	WRF	Sobash et al. (2019);
SCS	cases 2010-2017	1 km 0.6 mi/ hourly			Schwartz and Sobash (2019)

2.1.5.2   MULTI-RADAR/MULTI-SENSOR (MRMS)

The MRMS dataset is generated at the National Centers for Environmental Prediction (NCEP) and merges the output from 180 weather radars over CONUS and Canada into a 1-km (0.6 mi) 3D radar mosaic (Zhang et al. 2016). This mosaic is combined with a range of other observations such as satellite data, lightning observations, precipitation gauge data, and high-resolution numerical weather forecasts to produce a quantitative precipitation estimate (QPE, Zhang et al. 2011). The QPE product is provided on a national 1-km grid (0.6 mi) every 2.5 minutes.

2.1.5.3   PRISM

The PRISM Climate Group at Oregon State University generates a daily 4-km (2.5 mi) gridded precipitation estimate for the CONUS that covers the period 1982–present (Daly et al. 1994, 2002, 2008). The PRISM analysis includes measurements from nearly 13,000 quality-controlled precipitation gauges. PRISM uses a digital elevation model to estimate climate-elevation regressions for each grid cell, with stations being weighted according to their location, coastal proximity, and elevation. Local data and expert knowledge are also incorporated into the gridded product. Additionally, the daily precipitation estimates have taken radar measurements into account since 1 January 2002. The resulting precipitation estimates are very similar to daily

precipitation estimates from Stage IV in the eastern CONUS. The daily accumulation period used in PRISM is 1200 UTC–1200 UTC (e.g., 7 AM–7 AM EST), and this definition will be used for daily precipitation analysis throughout this report.

2.1.5.4 *GRIDDED METEOROLOGICAL ENSEMBLE TOOL (GMET)*

The GMET dataset (Newman et al. 2015) combines precipitation observations from ~12,000 gauges across the CONUS and adjacent watersheds in Canada and Mexico. GMET covers the period 1980–2016 and provides daily precipitation estimates on a ~12-km grid (7 mi). The unique aspect of this dataset is that it is probabilistic, based on a 100-member ensemble that is generated following the method of Clark and Slater (2006).

2.1.6 Model Datasets

We take advantage of three existing convection-permitting model forecast datasets that provide a total of 10,570 simulations reflecting 3-km horizontal grid spacing (1.8 mi) and 810 reflecting 1-km horizontal grid spacing (0.6 mi).

2.1.6.1 *NCAR MPEX ENSEMBLE*

In support of the Mesoscale Prediction Experiment (MPEX; Weisman et al. 2015) of May–June 2013, the National Center for Atmospheric Research (NCAR) produced real-time, 48-h, 30-member ensemble forecasts initiated at 0000 UTC daily with a 3-km (1.8 mi) grid covering two-thirds of the CONUS and running Version 3.3.1 of the Weather Research and Forecasting Model (hereafter WRF) (Skamarock et al. 2008; Powers et al. 2017). Schwartz et al. (2015a) provide a complete description of the forecasts and evaluation statistics, and the physical parameterizations used for these forecasts are noted in Table 2-6. Note that no cumulus parameterization was used in the 3-km forecasts, considered convection-permitting.

To assess whether decreasing horizontal grid spacing could improve the MPEX forecasts, 10-member ensemble forecasts with 1-km (0.6 mi) horizontal grid spacing were retrospectively produced over the MPEX period. Schwartz et al. (2017) describe this and found that 1-km forecasts of hourly-accumulated precipitation were typically better than corresponding 3-km forecasts. Both the real-time 3-km and retrospective 1-km forecasts were initialized from a 15-km (9 mi) continuously-cycling, 50-member ensemble Kalman filter (EnKF) data assimilation system, where initial conditions for the 3- and 1-km forecasts were obtained by downscaling 15-km (9 mi) EnKF analyses.

Table 2-6   WRF Physical Process Schemes for the 3-km MPEX Forecasts

Physical Process	WRF Process Scheme	References
Microphysics	Thompson	Thompson et al. (2008)
Longwave and shortwave radiation	Rapid Radiative Transfer Model for Global Climate Models (RRTMG) (including ozone and aerosol climatologies)	Mlawer et al. (1997); Iacono et al. (2008); Tegen et al. (1997)
Planetary boundary layer	Mellor–Yamada–Janjić (MYJ)	Mellor and Yamada (1982); Janjić (1994, 2002)
Land surface model	Noah	Chen and Dudhia (2001)

2.1.6.2   NCAR REAL-TIME ENSEMBLE

Based on the success of its MPEX forecasts, NCAR embarked on a longer-term forecast demonstration and produced real-time, 0000 UTC-initialized, 48-h, 10-member ensemble forecasts with 3-km (1.8 mi) horizontal grid spacing across the entire CONUS from 7 April 2015–30 December 2017 (Schwartz et al. 2015b, 2019). These ran WRF Version 3.6.1, employed identical physics to the MPEX runs, and were also initialized by downscaling 15-km EnKF (9 mi) analysis ensembles. These forecasts were skillful, valuable, and widely adopted by both the forecasting and research communities (Schwartz et al. 2019).

2.1.6.3   NCAR SCS DETERMINISTIC FORECASTS

While the MPEX ensemble dataset is useful for examining the sensitivity of warm-season precipitation forecast skill when horizontal grid spacing is reduced from 3 to 1 km (1.8 mi to 0.6 mi), NCAR also produced a larger set of 3- and 1-km WRF deterministic forecasts to examine model skill across a more diverse collection of severe convective storm (SCS) events, including cool-season severe weather events. This SCS dataset consists of 497 3- and 1-km forecasts for select events occurring between 2010–2017. Sobash et al. (2019) present specifics of the event selection, which is mostly based on the occurrence of multiple severe storm reports (e.g., tornadoes, hail, or intense wind gusts) across the central and eastern CONUS. For these runs the WRF configuration was similar to that of the NCAR real-time ensemble runs, including a full-CONUS computational domain, 0000 UTC initializations, 36-h integration length, and a similar set of physics schemes. In contrast, however, the SCS runs were deterministic forecasts employing initial and boundary conditions derived from NCEP Global Forecasting System (GFS) runs (.5 deg lat/lon grid).

2.2   Simulating Heavy Precipitation Events with Convection-Permitting Models

The Eulerian and Lagrangian model evaluations are summarized in this section. Starting with the Eulerian evaluation, we first show results based on the GSS and FSS analyses. The Lagrangian evaluation results are structured in four parts. The overall skill of all kilometer-scale simulations is first assessed and compared to observational uncertainties, and then the dependency of model skill on the seasonality of events is analyzed. We close with an assessment of the differences between the 3-km and 1-km simulations.

2.2.1   Eulerian Model Evaluation

2.2.1.1   *Gilbert Skill Score Analysis*
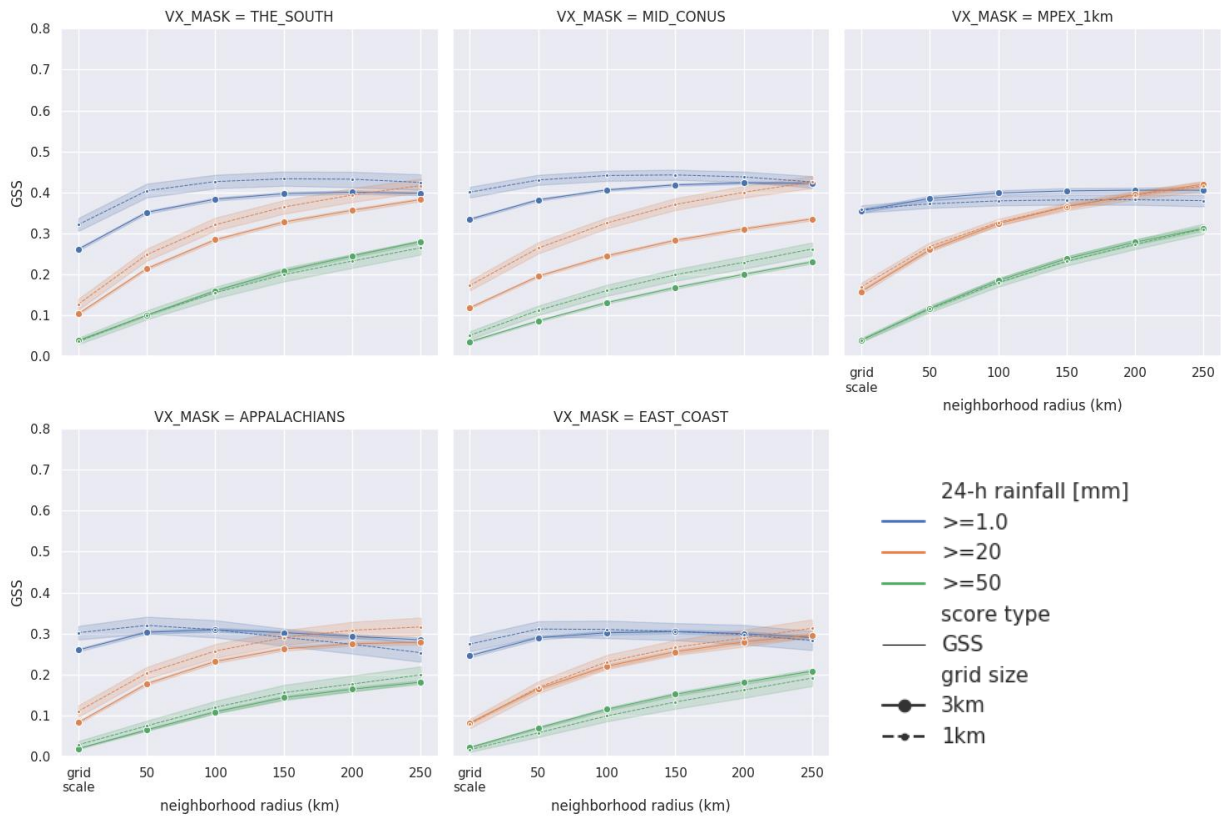
Figure 2-7   GSS for 5 Different Regions *The colors indicate different precipitation thresholds, and the line types denote the 1-km and 3-km models. The shaded bands are the 99% confidence intervals.*

It is a characteristic of the GSS that it typically increases as the verification neighborhood is enlarged. In Figure 2-7 we see that for our datasets the 1 mm/day (0.04 in/day) threshold (blue GSS lines) is always above the 50 mm/day (2 in/day) threshold (green). These thresholds' scores mostly straddle the 20 mm/day threshold (orange), but for the neighborhoods larger than 150 km (90 mi), the 1 mm/day GSS values level off or decreases, while the 20-mm/day (0.8 in/day) values approach the 1 mm (0.04 in) line. The 99% confidence intervals are wider for 1-km runs because there are fewer of them.

1-km and 3-km GSS confidence intervals typically overlap, indicating no statistical difference between the GSS values of the two sets of simulations. Where the bands don't overlap, however, the 1-km line is slightly higher. The greatest superiority of the 1-km forecast is seen in the Mid-CONUS for the 20 mm/day (0.8 in/day) threshold. The 1-km simulations are only slightly better for the 20 mm (0.8 in/day) and 50 mm (2 in/day) thresholds in the Appalachian region despite the improved representation of topography in this region. The reason for the minor improvement is not fully understood, but could be related to limitations in observing local

precipitation patterns in mountainous regions or an already adequate representation of the orographic features at 3-km spacing.

2.2.1.2   Fractions Skill Score Analysis
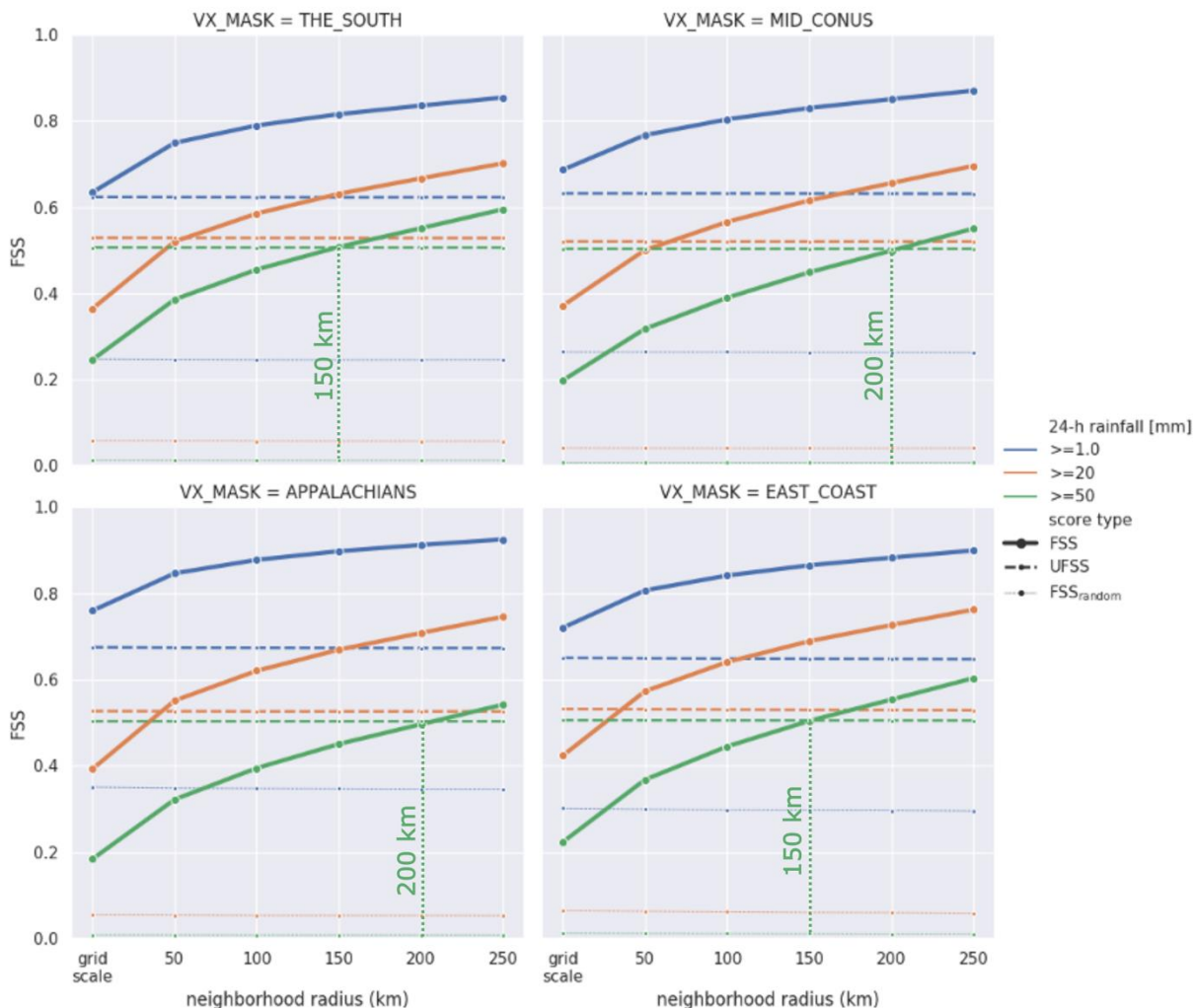
Figure 2-8   FSS Results for the NCAR Ensemble *All curves reflect 3-km forecasts. The colors signify the precipitation threshold, ranging from 1 mm/day (0.04 in/day; blue) to 50 mm/day (2 in/day; green). The thick, solid line is the FSS, while the dashed line and thin lines show reference scores. The dashed line is the Uniform Fractions Skill Score (UFSS), a reference score explained earlier in 2.1.3.1.2 , and the thinnest line is the FSS of a random binary forecast. The scale$_{min}$ parameter for the 50 mm/day (2 in/day) threshold is labeled and shown with a green dotted vertical line.*

Figure 2-8 shows the FSS for the NCAR Ensemble for four regions. The only significant difference between regions is seen in the 1-mm/day and 20-mm/day thresholds. For these low and moderate thresholds, the Appalachian and East Coast regions have higher FSS values

than the South and Mid-CONUS regions. The 50-mm/day threshold scores are similar across the four regions.

As explained earlier, the neighborhood size where the FSS matches the UFSS can be thought of as the minimum scale for which the forecast provides useful information. As the precipitation thresholds get higher, the $scale_{min}$ gets larger. For any threshold, one should be cautious about believing any forecast precipitation events with spatial scales smaller than $scale_{min}$. For the 50-mm/day threshold, $scale_{min}$ is 150 km (90 mi) for the South, 200 km (124 mi) for the Mid-CONUS, 200 km (124 mi) for the Appalachians, and 150 km (90 mi) for the East Coast.
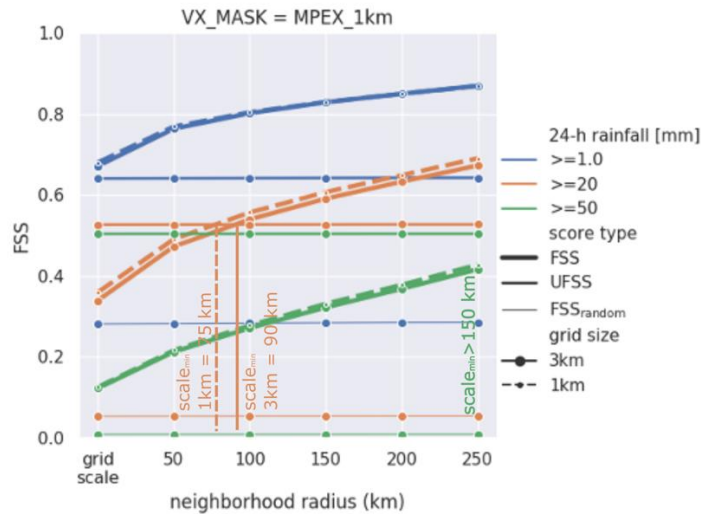
Figure 2-9   FSS Results from the NCAR MPEX Forecasts *The colors signify the precipitation threshold; the solid and dashed lines differentiate the 3-km (solid) and 1-km (dashed) forecasts, and the thinner lines show reference scores. The thickest line is the FSS. The medium-width line is the Uniform Fractions Skill Score, explained earlier in Section 2.1.3.1.2 , and the thinnest line is the FSS of a random binary forecast. The $scale_{min}$ parameter for the 50 mm/day (2 in/day) threshold is labeled in green and the 20 mm/day (0.8 in/day) parameters for the 1-km and 3-km runs are shown in orange dashed and solid vertical lines correspondingly.*

For NCAR MPEX runs, the 50 mm/day (2 in/day) FSS never surpasses UFSS in the displayed range of spatial scales (Figure 2-9). In other words, $scale_{min}$ is greater than 250 km (155 mi). For 20 mm/day (0.8 in/day), $scale_{min}$ is 75 to 90 km (47 to 56 mi), depending on the model resolution. The 1-km runs are always slightly better than the 3-km runs, but not by much.

For the NCAR SCS cases, the FSS progression is similar to that from the other datasets, but there are some differences, especially in the East Coast region (Figure 2-10). For example, the 50-mm/day FSS is particularly low on the East Coast. As with MPEX, FSS never reaches UFSS at this threshold and domain, but the low and moderate thresholds are reasonable. This results in greater separation between the FSS values for the thresholds in the East Coast than any other domain. Conversely, there is less separation between the thresholds in the South. There, the 50-mm/day FSS is the highest of all four regions. As with MPEX, the SCS 1-km runs are slightly better than the 3-km runs.
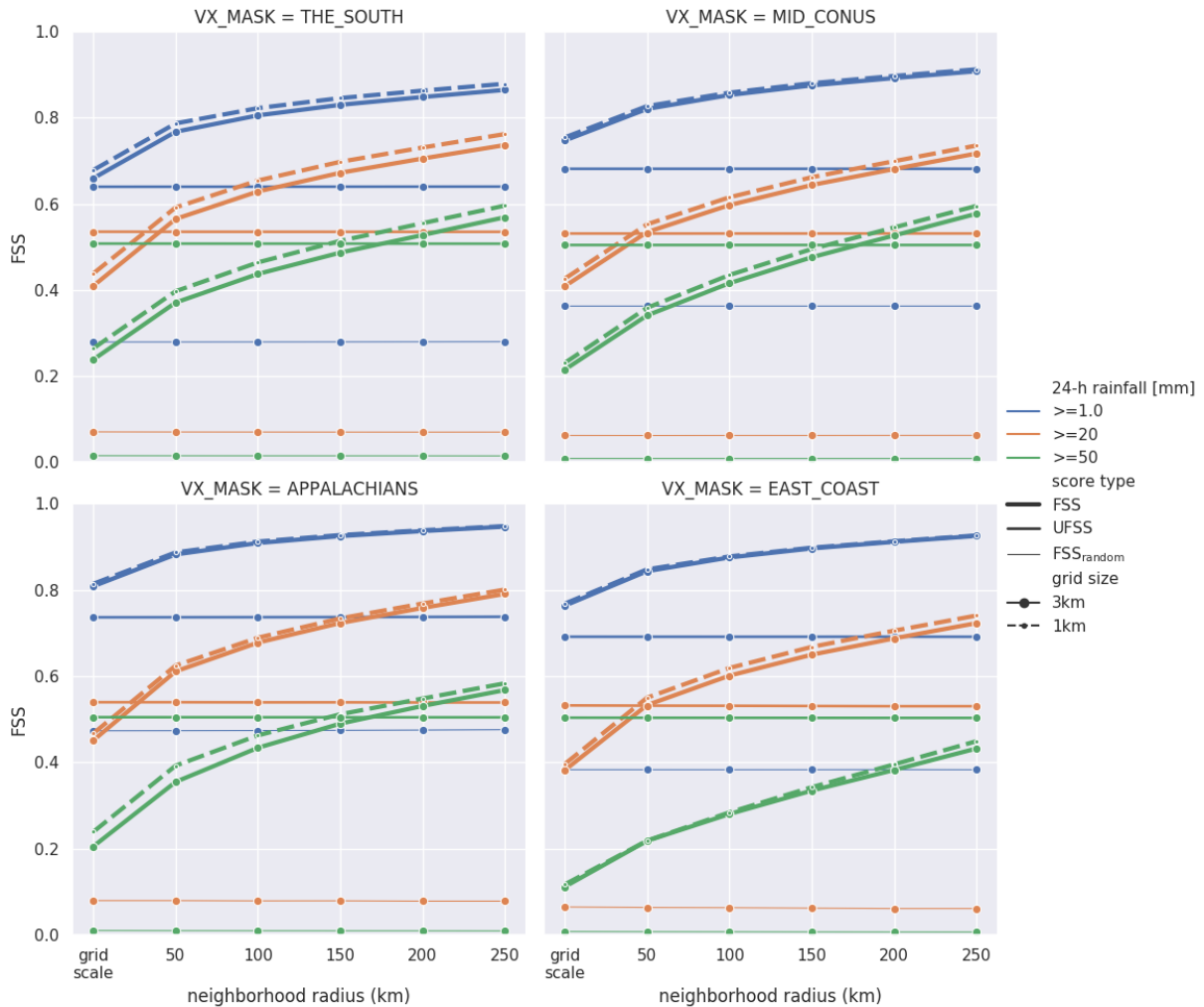
Figure 2-10 FSS Results from NCAR Severe Convective Storm Forecasts *Line types, color convention, and thresholds as in the previous two figures.*

2.2.2   Lagrangian Model Evaluation

2.2.2.1   *Peak Precipitation Displacement*

The station-based daily precipitation datasets show a small (~8 km/5 mi) eastward shift in the peak precipitation location of heavy events in the South when compared to the Stage IV data (Figure 2-11a). The median meridional location agrees very well between the observation datasets, however. The interquartile spread is approximately ~20 km (12 mi) in the zonal and meridional directions. While peak median accumulations in the simulations are shifted southward slightly (i.e., ~8 km/5 mi), the zonal location is centered well. The interquartile spread is ~40 km (24 mi) and is thereby twice as large as in the observations.

While the simulated median peak location is similarly well-captured in the Mid-CONUS region, the probability density function (PDF) is skewed towards the east. The same skew is seen in the daily observations, and while the zonal interquartile range is smaller in the observations, the

10th–90th percentile range is not. The zonal elongation of the two-dimensional PDF is likely related to the predominant eastward storm movement (i.e., reflecting prevailing mid-tropospheric westerly flow) in the Mid-CONUS.

The largest spread in simulated peak locations is found in the MPEX ensemble (Figure 2-11c). The reason for this is unclear, but it might be related to the short period considered (one month), resulting in the inclusion of events with peak accumulations that are modest compared to many of the events included in the South and Mid-CONUS analyses. We will investigate the dependence of model skill on event rarity below.

The largest observational uncertainties in the peak location are found in the Appalachian region (Figure 2-11d). The PDF from daily precipitation observations is heavily skewed towards the northeast, which is likely related to the generally northeastward orientation of the mountains chain. Surprisingly, the simulations have smaller uncertainties than the daily observations, which highlights the significant challenges in observing precipitation and the high-quality of kilometer-scale models in mountainous regions.
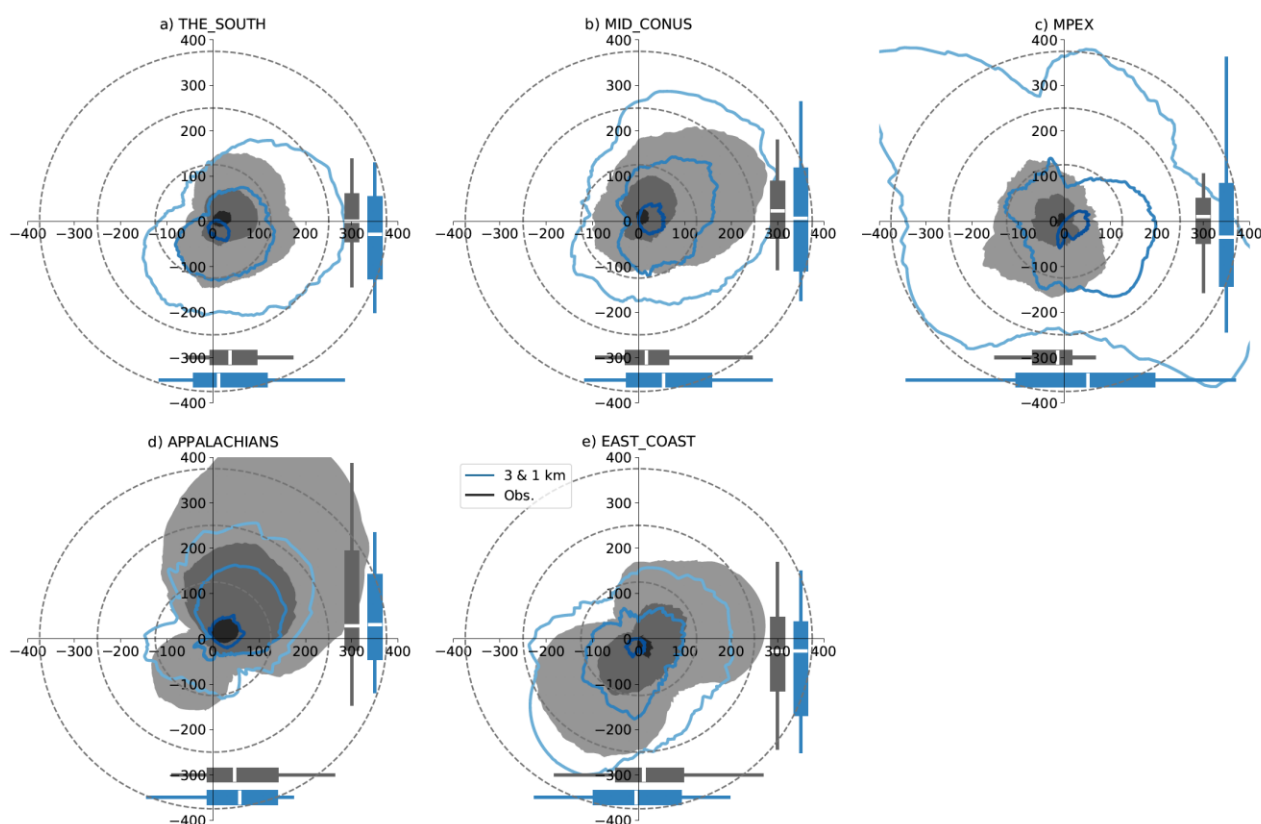
Figure 2-11   Displacement Error of Observed (Gray Shading) and Simulated (Blue Contours) Peak Precipitation Location Compared to Stage IV Observations *All datasets have been coarsened to a 20-km grid (12 mi) to remove small-scale noise. The contours show the 50, 25-75, and 10-90 percentile ranges (from dark to light). The 25-75 percent area, for example, contains 50% of all data points. Box-whisker statistics show the distribution of longitude and latitude displacement errors for the observations (gray shading) and simulations (blue contours).*

Observational and simulated uncertainties are of similar magnitude in the East Coast region (Figure 2-11e). Displacements in the meridional direction are approximately half those in the zonal direction, likely related to the predominant direction of storm movements.

2.2.2.2   Event Characteristic Evaluation

Simulating the characteristics of heavy rainfall events demands capturing event characteristics such as rain location, system translational speed, areal coverage, and intensity. We find that the various WRF simulations can capture the median speed of heavy rainfall storms in all subregions as well as the MRMS observations when compared to Stage IV data (Figure 2-12a). The main difference between the MRMS and the modeled speeds is that the latter have a greater spread. In the case of MRMS, speed variation reflects case-to-case variability, while the spread in the simulations includes the influence of initial condition uncertainty, which partly explains the larger magnitude. This larger spread is to be expected and is a common feature in the model storm characteristics (Figure 2-12a–f) and will, therefore, not be detailed further.

Differences in the median size of the hourly precipitation areas are similar between the MRMS and model datasets, except for the Appalachian region, where model precipitation areas are too small (Figure 2-12b). Mean hourly precipitation rates are also well captured in the simulations, except for the East Coast, where the model overproduces precipitation by about 5 % (Figure 2-12c). A similarly good performance is seen for the median 99th percentile of hourly precipitation (Figure 2-12d). The simulations agree very well with the MRMS observations, while they do not produce the larger intensities seen in the Stage IV data in the Mid-CONUS and Appalachian regions. The largest simulation biases, of about 15 %, are found in East Coast events.

For daily precipitation characteristics such as the 99th percentile event accumulation and the event volume, we can add the daily observational datasets to the analysis (Figure 2-12e). The median 99th percentile total event accumulation is underestimated in the simulations in most regions (cf., the East Coast), and this is most pronounced in the South (-30%). Event peak accumulation is difficult to simulate since it is an aggregated measure and depends on the correct simulation of precipitation rates, event size, and system movement. In the South, the underrepresentation of event size seems to be the dominant reason for the underestimation of peak accumulations, since the system speed and mean and extreme hourly amounts are all well simulated, while underestimated precipitation rates are also seen in the Mid-CONUS, MPEX, and Appalachian regions. The modeled small size of East Coast storms is offset by over-forecasted precipitation rates, resulting in similar peak accumulations. The MPEX accumulations are similar to those in Stage IV except in the South and Mid-CONUS regions, where peak accumulations are ~10% lower in the MPEX runs. The GMET dataset has significantly smaller median peak accumulations (-30% to -50%) than Stage IV.

Event total precipitation volume is well-simulated in the South, MPEX, and East Coast (Figure 2-12f) regions, while it is systematically underestimated in the Mid-CONUS (18%) and the Appalachians (27%). This underestimation is primarily a consequence of the sizes of the modeled precipitation areas in these regions being too small. The MRMS data has ~10% lower event precipitation volumes than the Stage IV data, except for East Coast storms. And, the GMET dataset has precipitation volumes that are below those of Stage IV in all regions, with the largest deviations in the South and MPEX regions.
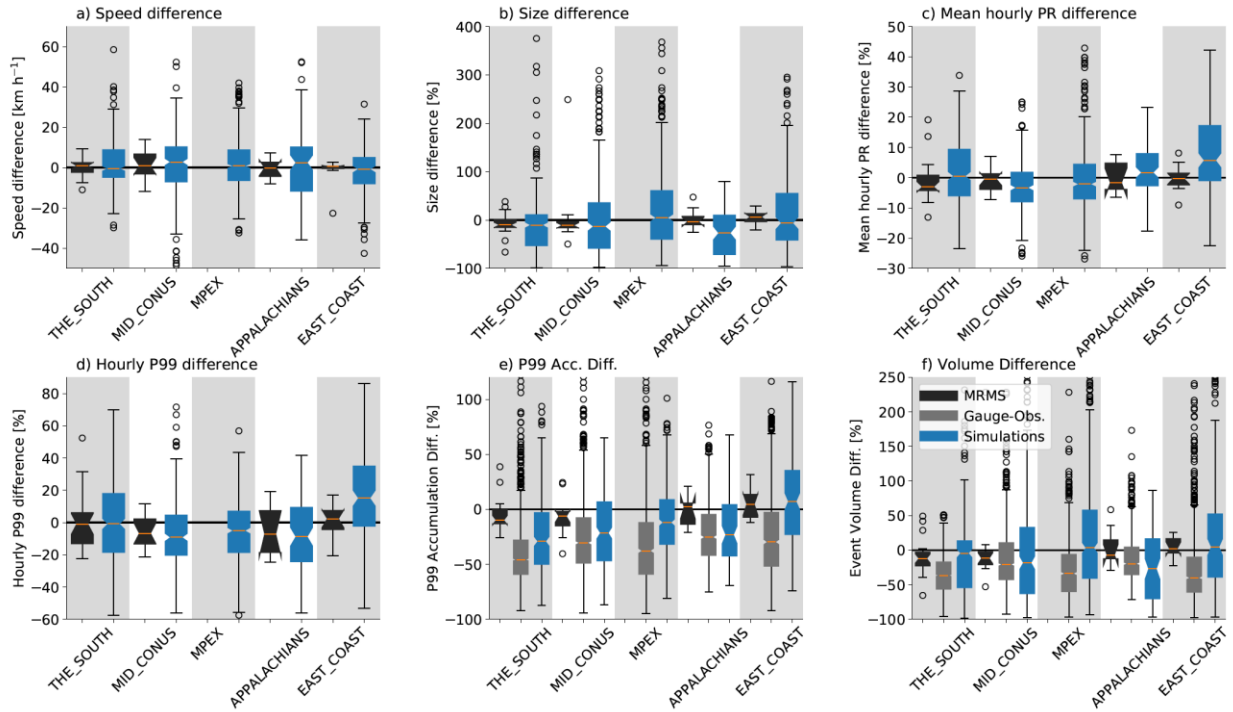
Figure 2-12 **Box-Whisker Plots of Differences in Heavy Precipitation Events Compared to Stage IV Observations. Differences in Hourly Precipitation Statistics: (a) Storm Translational Speed; (b) Precipitation Area; (c) Mean Precipitation; and (d) 99th Percentile Precipitation. (e) Daily Accumulated Differences for 99th Percentile Precipitation. (f) Storm Precipitation Volume** *Each panel shows the statistics for the five sub-regions. Differences are shown for all available simulations and observations (MRMS for hourly precipitation; MRMS and gage-based for daily precipitation).*

Observing and simulating event total peak accumulations is challenging, and we are interested in how observational uncertainties and model differences change, depending on the spatial scale. Therefore, we coarsen the datasets to increasing larger scales, comparing event peak accumulations to Stage IV observations (Figure 2-13). In general, observational uncertainties and model differences decrease on larger spatial scales. In the South region (Figure 2-13a) the MRMS data has systematically 10% lower peak accumulations than the Stage IV data. Peak accumulations in both the daily observations and the simulations become similar to those in MRMS for 1,600 km^2 areas (618 mi^2; i.e., 40×40-km boxes). Similar behavior is visible for the Mid-CONUS, MPEX, and Appalachian regions, while differences in East Coast region storms are small even on local scales.

Table 2-7   **Summary of Average Differences Between Simulated (Sim) and MRMS Event Characteristics (Rows) Compared to The Stage IV Dataset** *Shown are median (med.) differences and the interquartile range (IQR) of ensemble spread (i.e., the box length in Figure 2-12) for all five subregions in columns. Note that no MRMS observations are available during the MPEX simulation period. P99 denotes the 99 percentile.*

		South		Mid-CONUS		MPEX		Appalachians		East Coast	
		med.	IQR	med.	IQR	med.	IQR	med.	IQR	med.	IQR
Speed [km/h]	Sim	-0.5	13.6	2.8	17.4	0.8	15.1	2.4	21.9	-0.7	12.9
	MRMS	1	4.9	0.8	9.9	/	/	-0.1	7.12	0.5	1.23
Size [%]	Sim	-10.3	63.9	-13.3	94.1	4.6	100.3	-27.0	81.4	-6.4	96.3
	MRMS	-10.1	16.3	-11.0	13.2	/	/	-3.6	17.3	6.4	16.8
Mean hourly PR [%]	Sim	0.6	15.4	-3.3	9.8	-2.1	11.4	1.71	10.6	5.7	18.2
	MRMS	-2.9	5.1	-0.4	5.2	/	/	-1.6	8.65	-0.3	3.9
P99 hourly PR [%]	Sim	-0.8	36.4	-9.2	24.8	-5.1	25.6	-8.81	33.6	15.0	37.3
	MRMS	-0.9	19.2	-6.71	15.5	/	/	-7.1	25.9	2.3	11.8
P99 event accumulation [%]	Sim	-29.1	46.9	-21.5	53.5	-11.9	40.5	-23.1	46.6	7.1	58.2
	MRMS	-9.8	12.7	-6.37	13.2	/	/	2.2	19.4	4.5	20.8
Event volume [%]	Sim	-4.4	67.8	-18.3	95.8	3.3	98.6	-27.0	86.7	4.5	91.4
	MRMS	-12.1	19.7	-11.7	18.8	/	/	-6.8	27.4	2.0	20.9

Table 2-7 shows the differences of simulated and MRMS storm characteristics compared to Stage IV observations presented in Figure 2-12. The mean differences between MRMS and Stage IV in combination with the interquartile range (IQR) of differences can be used in flood risk assessment as stochastic error terms that characterize observational uncertainties. The mean difference between simulations and Stage IV reveals systematic model biases that have to be corrected before model data can be used in flood risk assessments if the differences are larger than the observational uncertainties. The simulated IQR is a representation of the range of plausible storm characteristics that might have happened. For instance, an observed heavy precipitation event could have been weaker or stronger due to the chaotic nature of the atmosphere. This range can be used in probabilistic flood risk assessments to construct precipitation fields that represent plausible realizations of historic heavy precipitation events.
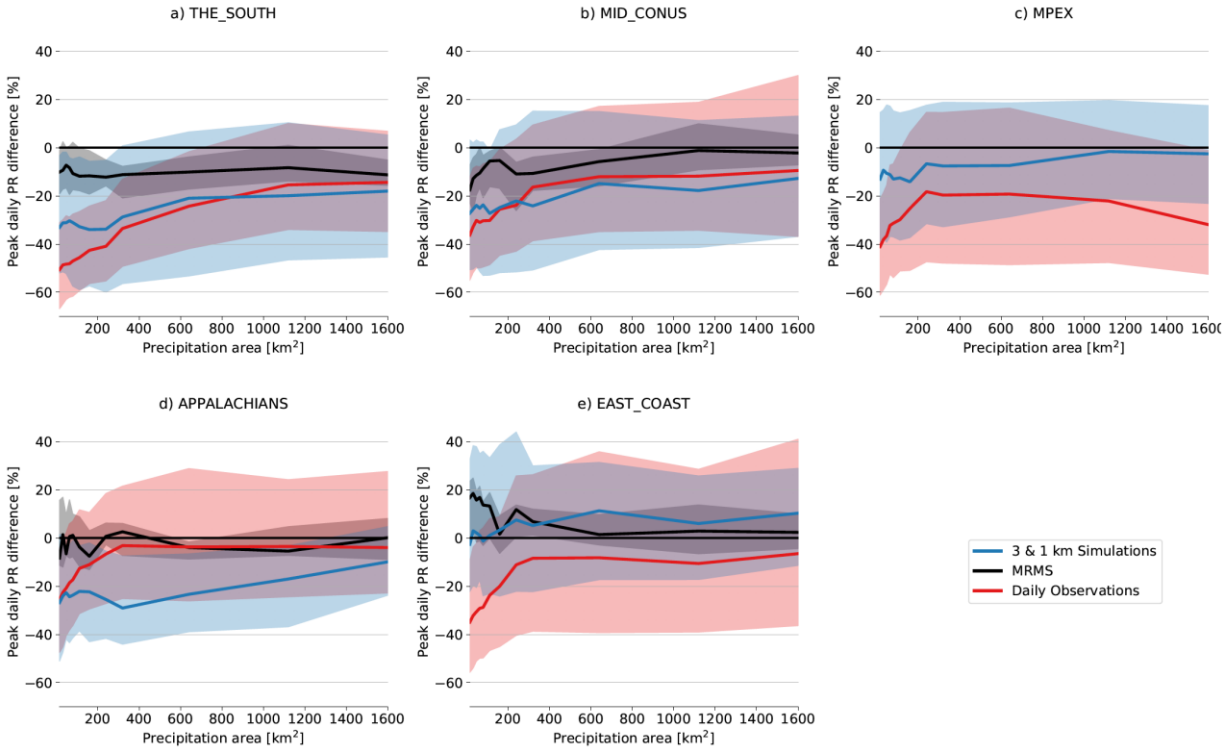
Figure 2-13 Differences in Daily Event Total Peak Precipitation Between the MRMS Dataset (Black), Daily Observations (Red), WRF Simulations (Blue), and Stage IV Data, Dependent on Precipitation Area *The precipitation area is calculated by regridding the 4-km datasets to coarser spacings. Thick lines show the ensemble average, and filled contours show the interquartile range.*

As previously discussed, orographic enhancement of precipitation can be a key process in heavy precipitation events. Therefore, we show the difference in daily event precipitation dependent on surface height above sea level (Figure 2-14). Note that in some areas, such as the South and Mid-CONUS, this might correspond more to a regionally dependent bias than to an orographic precipitation bias, since high elevation areas in those regions are in the west with generally gradual slopes towards the east. All datasets agree well below 700 m in the South, but in high elevation areas the daily observations and simulations have significantly higher precipitation accumulations than the Stage IV data (Figure 2-14a). The observational datasets also agree well with the Mid-CONUS region and MPEX results, except for GMET, which displays a significant low bias at high elevations, particularly in the Mid-CONUS region. Stage IV and the simulations agree remarkably well in the Appalachian region across all elevation levels, while the GMET dataset underestimates event mean precipitation at low levels and overestimates it at high levels (Figure 2-14d). Simulated precipitation in the East Coast region is overestimated between an elevation of 300 m and 900 m, but agrees very well with Stage IV and MRMS data below 200 m.
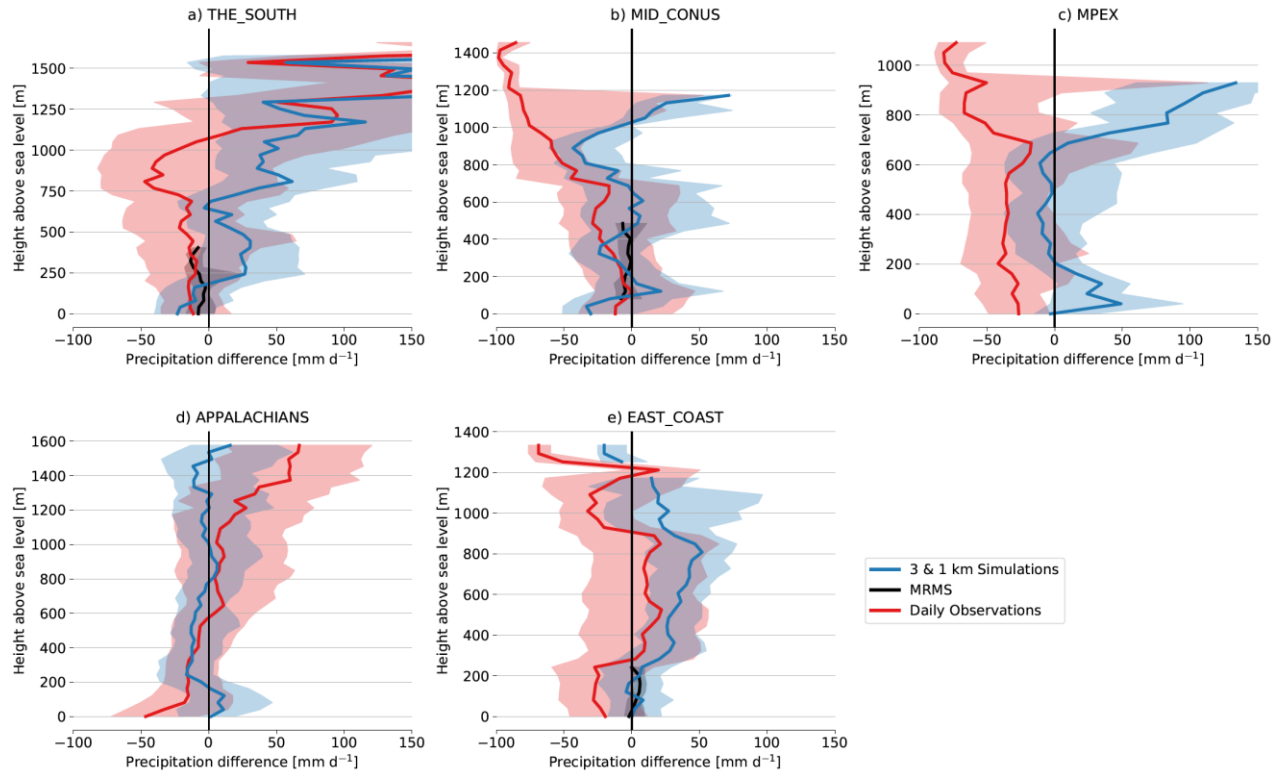
Figure 2-14   Elevation-Dependent Difference Between MRMS Data (Black), Daily Observations (Red), and the 3-km and 1-km WRF Simulations Compared to Stage IV for Heavy Precipitation Events in the Five Sub-Regions *Thick lines show the ensemble medians, and contours show the interquartile spread. Results are only shown for ensemble sizes of 20 or larger. The elevation bin size is 40 m.*

2.2.2.3   Model Skill Dependence on Seasonality and Event Intensity

The previous section focused on mean differences, averaged over all events in a region. However, model skill within a region might depend on seasonality and the rarity of an event, thus that is our focus here.

Differences in the 99th percentile event peak accumulation show complex dependencies on the rarity of events, but skill generally does not deteriorate (i.e., events with higher 99th percentile event peak accumulations in the Stage IV dataset are similarly well simulated than less heavy events: Figure 2-15a). Similarly, there is no clear dependency when using event precipitation volume as a measure for rarity (Figure 2-15b). Mean differences increase at larger precipitating volumes, especially in the Mid-CONUS region, although uncertainties in this estimate are large. The displacement error of the location of the 99th percentile precipitation shows only a weak dependency on the storm's precipitation volume (Figure 2-15c). However, there is a tendency towards smaller location biases in the South and larger biases in the Mid-CONUS as event rarity increases.
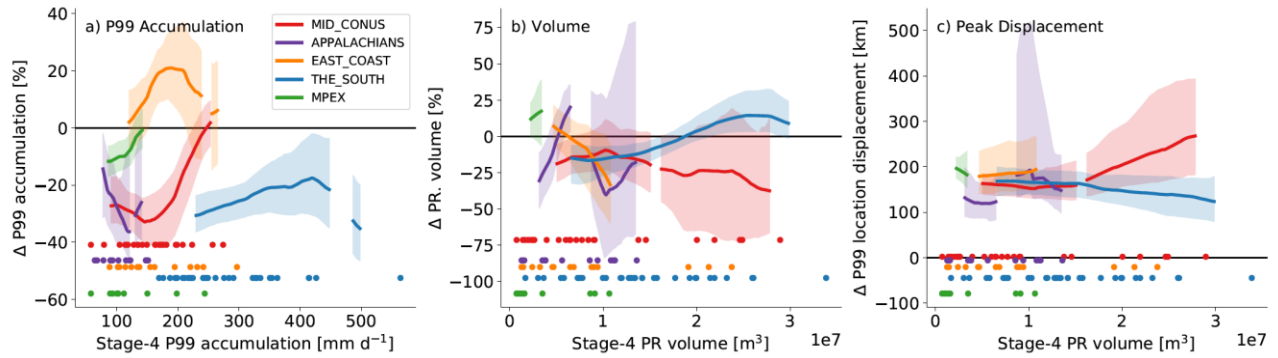
Figure 2-15 Dependence of 3-km Ensemble Skill in Simulating Event Precipitation Accumulations: (a) Peak Accumulation; (b) Total Event Volume; and (c), Displacement of Peak Accumulation Location *Differences based on comparisons to Stage IV data. Bold lines show the ensemble average. Contours show the 10–90th percentile spread of 1,000 random bootstrap samples from all 3-km simulations in a region. Dots show the Stage IV 99th percentile accumulation (a) and volume (b,c) of each event. Data is only shown for areas with four events or more. Savitzky-Golay linear filter with a window length of 20% of the data range applied.*

The precipitation accumulation differences for event 99th percentiles are largest in middle to late summer and smaller in late spring and fall in the South region (Figure 2-16a). This is different from in the Mid-CONUS region, where late spring and early summer biases are largest. The event sample size is too small and the heavy precipitation events are too seasonal to analyze the seasonal dependence of model differences in the other regions. Biases in simulated precipitation volumes show similar patterns compared to differences in peak accumulation, with the largest biases in early summer in the Mid-CONUS and in mid-summer in the South (Figure 2-16b). Biases in the simulation of the location of the 99th percentile accumulation are more coherent between subregions and show the largest displacement errors in late summer and early fall, likely due to the weak large-scale forcing during this time of year.
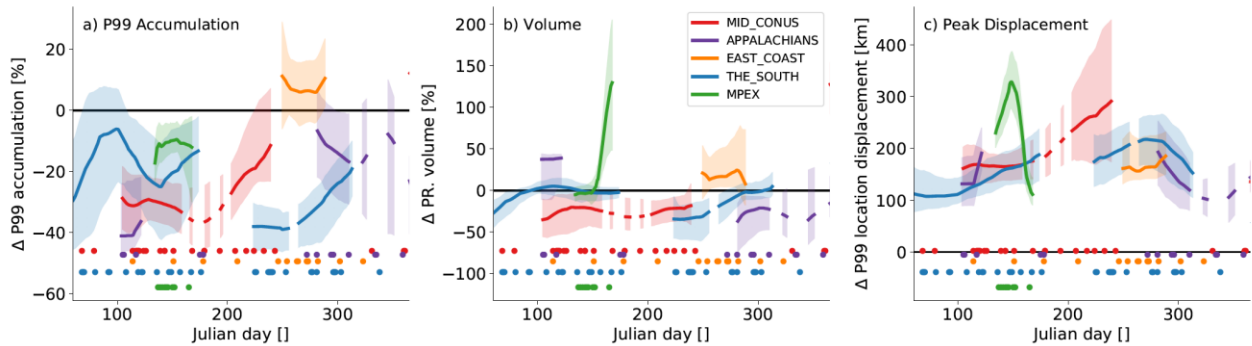
Figure 2-16 As in Figure 2-15, but with Skill Scores Shown Dependent on the Julian Day of the Year

2.2.2.4   Model Performance: 3-km v. 1-km

In this section, we explore how large and how systematic the differences between the 3-km and 1-km simulations are, through comparison with the Stage IV data. We are not able to show results for the East Coast region because the sample size of 1-km simulations is too small. Figure 2-17 shows that differences in results from the different model resolutions are small and not significant (i.e., the interquartile range of the error distributions overlaps). Some systematic improvements in the 1-km simulations can be seen in the South, however, where error distributions are typically narrower (except for the hourly precipitation areas and volumes). Part of this apparent added value might be related to the much smaller sample size of 1-km simulations (8) compared to 3-km runs (38). Noteworthy are the differences in P99 hourly precipitation, which are higher in the 1-km simulations in the South, but lower in the other regions, highlighting the regional dependence on resolution differences.
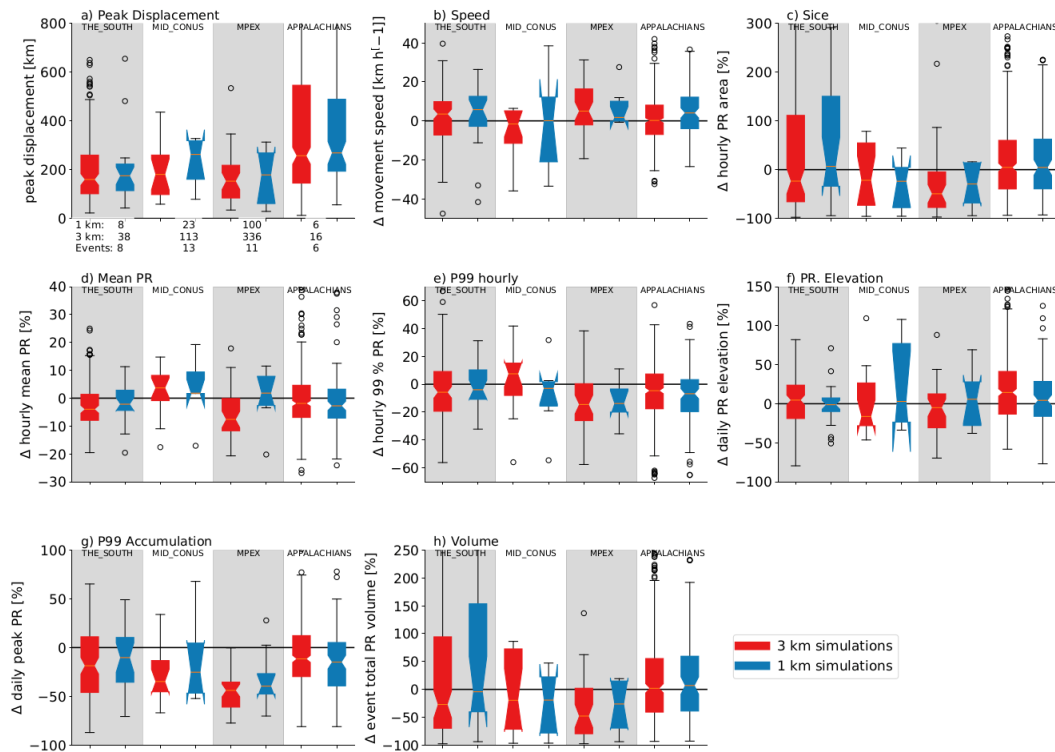
Figure 2-17   Differences in Storm Characteristics From 3-km (Blue) and 1-km (Red) Simulations Compared to Stage IV Data *(a) Differences in the peak precipitation displacement. (b) Differences in system translational speed. (c) Differences in average hourly precipitation area. (d) Differences in hourly mean precipitation. (e) Differences in 99th percentile of hourly rainfall. (f) Differences in mean precipitation averaged over terrain elevation ranges. (g) Differences in event total P99 accumulation. (h) Differences in event total precipitation volume. The number of 3-km and 1-km simulations and the number of events shown (a). The notches (area where the box becomes thinner) provide guidance concerning the significance of difference of medians (e.g., statistically significant difference between the medians occur if the notches of two boxes do not overlap). Notches can extend beyond the 1st and 3rd quartile if the confidence interval around the median is larger than the interquartile range.*

2-29
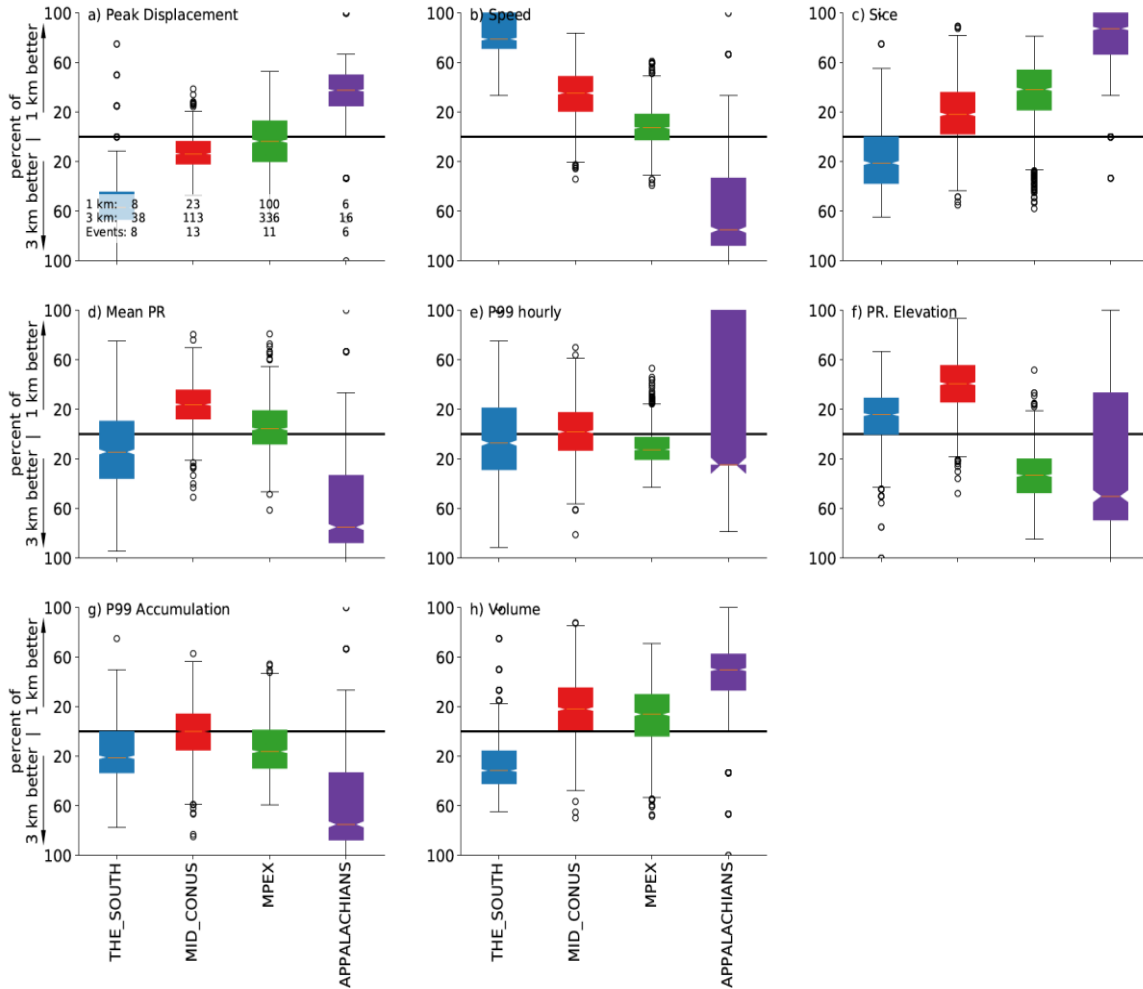
Figure 2-18 Percentage of 1-km and 3-km Simulations that Have Smaller Differences Compared to Stage IV than their 3-km/1-km Counterpart *a) Location of the peak accumulation. b) System translational speed. c) System area/size. d) Mean precipitation rate. e) 99th percentile hourly precipitation rate. f) Elevation dependence of precipitation. g) 99th percentile total accumulation. h) System total precipitation volume. The individual box/whisker elements show results in sub-regions, with the East Coast region not shown, as it did not have enough simulations for robust statistics. The box/whisker analysis spread is derived from bootstrapping storm events 1000 times. The number of 1-km and 3-km simulations and the number of events are shown in (a).*

Next, we analyze the systematic differences between simulations with different grid spacings (Figure 2-18). The differences between 3-km and 1-km grids in simulating the peak accumulation location vary strongly by region (Figure 2-18a). In the South, the 3-km simulations are better 60% of the time, while both grid spacings perform equally well in the Mid-CONUS and MPEX regions. An added value of the 1-km simulations appears in the Appalachian region, where they outperform the 3-km runs 35% of the time. The 1-km simulations better simulate the speed of storms in the South and Mid-CONUS regions 60% and 40% of the time, respectively,

but they perform worse in the Appalachian region 70% of the time (Figure 2-18b). The hourly precipitation area is better simulated in the MPEX domain runs (30% of the time) and in the Appalachian region (80 % of the time), while the mean hourly rainfall is better simulated by the 3-km simulations in the Appalachian region (80% of the time; Figure 2-18c,d) The simulation of the 99th percentile of hourly precipitation intensities is comparable between the two model grid spacings (Figure 2-18e). Elevation-dependent precipitation rates in the Appalachians are better simulated in the 3-km runs, but large uncertainties exist (Figure 2-18f). The 1-km and 3-km simulations have similar skills in simulating the 99th percentile of event peak accumulation, except for the Appalachian region, where 3-km simulations are better 80% of the time (Figure 2-18). This is different for the metric of total precipitation volume, where the 1- km runs are better than the 3-km simulations ~50% of the time (Figure 2-18h). The inconsistency of added value in the 1-km simulations compared to the 3-km runs across regions and metrics is primarily a reflection of the minor differences between the two grid spacings (see Figure 2-17).

2.3   Summary of Section 2

In this section we have analyzed the performance of three convection-permitting (CP), short-term (12-hour to 36-hour) NWP model datasets in simulating recent heavy precipitation events in the central and eastern United States. We use both Eulerian and Lagrangian evaluation frameworks for a more complete understanding and for robustness in the results. The analysis is performed over four subregions: Southern U.S., Central U.S., Eastern U.S., and Appalachians. We use the Stage IV multi-sensor precipitation observation as the reference for model evaluation (Crosson et al. 1996; Fulton et al. 1998) due to its high spatial and temporal resolution (hourly, 4-km; 2.5 mi gridded information) and the observational period (2002–present), which includes all of our simulated events. For the Lagrangian approach, we also apply three other observational datasets offering hourly (MRMS; Zhang et al. 2016) and daily precipitation accumulations (PRISM, Daly et al. 1994, 2002, 2008; GMET, Newman et al. 2015). GMET is solely based on station observations, while PRISM adds radar observations, making it more like Stage IV. The inclusion of multiple observational datasets allows us to calculate observational uncertainties and compare them to model biases. An additional difference between the Eulerian and Lagrangian model evaluations is that the former evaluates all available forecasts and focuses on heavy precipitation events by using a threshold of 50 mm/d (2 in/day), while the latter focuses on evaluating the heaviest events in a region only.

The main findings are as follows.

- CP numerical models can reliably simulate many characteristics of heavy precipitation events across various meteorological settings. Comparison of CP model results with Stage IV observations shows that CP models can accurately capture system speed and mean hourly precipitation rates. However, except for the East Coast, they systematically underestimate daily peak accumulations by 10–30% in the study region. This reflects: (i) precipitation areal extent underestimations in the South (-15%), Mid-CONUS (-20%), and Appalachian regions (-30%); and (ii) hourly peak rate underestimations in the Mid-CONUS and Appalachian regions (-15%). In terms of system positioning, the median displacement in the peak accumulation location is less than 50 km (31 mi) in all regions. The FSS analysis in the Eulerian framework shows that the simulations have useful skill in simulating 50 mm/h (2 in/day) rainfall events on spatial scales of 150 km (93 mi) in most regions. The reason why the useful scale in the FSS analysis is three times larger than the peak rainfall displacement in the Lagrangian analysis is that the FSS not only penalizes for displacement errors, but also for magnitude errors. Additionally, the ensemble spread is lowering the skillful scale as the ensemble mean might have zero

location bias, but the FSS will penalize the location biases of each ensemble member.

- Viewed with respect to Stage IV observations, CP model estimates of heavy precipitation events are better than the estimates from station-based precipitation products (i.e., GMET) for (a) daily peak accumulations and (b) event precipitation volumes. The differences between the two multi-sensor observation products (Stage IV and MRMS) are mostly smaller than the differences between the Stage IV values and the CP forecasts. This is to be expected since Stage IV and GMET use a common set of observations, and the simulations sample a larger range of uncertainties (e.g., the unpredictability of the atmosphere, biases in the input to the CP model, and biases in the CP models). Differences between Stage IV and MRMS values could be used to define stochastic transfer functions to remove systematic biases in the CP simulations.

- CP model simulations can capture orographic gradients in heavy precipitation events very well, and they outperform station-based observations (GMET), which typically underrepresent precipitation at lower elevations and overrepresent precipitation at high elevations. The models even have superior skill to the GMET dataset in their location of peak accumulations. Lastly, simulated event total accumulations are captured well, and they are within the uncertainty range of Stage IV and MRMS for accumulation areas larger than ~150 km (~93 mi).

- The relationship between model skill and event intensity is complex and depends on the region and the metric for intensity. Event peak accumulations biases tend to decrease with the rarity of an event. This might be due to the increasing impact of large-scale forcing in extreme cases. We find that simulated event volumes and peak location displacement biases decrease in the South and increase in the Mid-CONUS as event intensity increases. Model skill is also clearly dependent on the seasonality of events with lower skill during middle and late summer and higher skill in the shoulder seasons. Again, this is likely caused by weaker large-scale forcing in the summer, resulting in less-predictable meteorological conditions. There are considerable uncertainties in the magnitude of model skill dependencies on event rarity and seasonality, which could be reduced by increasing the sample size of simulated heavy precipitation events.

- Eulerian and Lagrangian model results consistently show that there is little added value in increasing the model grid spacing from 3 km to 1 km (1.8 mi to 0.6 mi), from the perspective of computational cost, given that there is a ~30-fold increase in computational resources required for the higher-resolution runs (i.e., 30 3-km simulations can be performed for the same cost of one 1-km simulation). This is in good agreement with previous results (Schwartz et al. 2009).

In summary, numerical weather prediction models configured with convection-permitting resolutions can capture heavy precipitation events in the Eastern U.S. Many characteristics of the simulated cases are verified with multi-sensor observational datasets that include radar, and the precipitation output from CP models shows less error than precipitation estimates based on station data. This demonstrates the potential value in incorporating CP model outputs in flood risk assessments since current flood standards are based on station records (e.g., NOAA Atlas 14; Bonnin et al. 2004). CP models can also provide information on historic and future changes in flood risks based on global warming projections (e.g., Prein et al. 2017). However, CP models are not perfect and can have systematic biases, such as a potential for underestimations of event peak accumulations of up to 30%, found here. To deal with such biases, we recommend the application of statistical post-processing, in the short term, and further model development, in the long term, before heavily relying on model-simulated precipitation in intense event and flood risk assessments.

3   A CONCEPTUAL FRAMEWORK FOR INTEGRATING RAINFALL SIMULATIONS INTO PROBABILISTIC FLOOD HAZARD ASSESSMENT

In this section, we build on findings from the previous sections and integrate them into a conceptual framework for probabilistic flood hazard assessments. Section 3.1 begins with background on rainfall-based flood frequency analysis, including traditional approaches that utilize point-based rainfall, as well as advances using coherent spatiotemporal rainfall fields. Section 3.2 then focuses on CPM rainfall simulations. It includes examinations of the effect of CPM ensemble size on error and variability and the role of dynamical downscaling in developing CPM datasets. In Section 3.3 , CPM criteria relevant to the conceptual framework are compared and contrasted, and we offer options on how CPMs can be integrated with flood frequency analyses. Section 3.4 concludes with overall recommendations and ideas for future work.

3.1   <u>Rainfall-based Flood Frequency Analysis</u>

Flood frequency analysis (FFA) is used to determine flood flows, typically a peak discharge or volume, associated with an annual exceedance probability (AEP). For example, a "100-year flood" has an AEP equal to 0.01. One approach to FFA is to fit a probability distribution directly to streamflow measurements, typically the annual maxima, from existing records. However, compared to streamflow measurements, precipitation observations are generally more numerous and have longer records. As such, rainfall observations have been utilized for a second approach: Rainfall-based Flood Frequency Analysis (RFA). In RFA, rainfall events are used as inputs to a calibrated rainfall-runoff hydrologic model to simulate flood events.

Traditional approaches to RFA have used representations of rainfall that are derived from station data, i.e., point-based observations. Given that there are very few extreme events recorded at a particular station, nearby rainfall stations can also be used to augment a specific site's record; this is known as Regionalized Flood Frequency Analysis. For RFA, the key rainfall characteristics to be considered are: (i) duration, (ii) intensity/volume, (iii) areal pattern, and (iv) temporal pattern. A common starting point for RFA is the Design Event Approach. A design event is statistically derived from a probability distribution fit to the intensity-duration-frequency (IDF) of the rainfall observations. The IDF curve describes rainfall rates (or depths) in terms of their duration and recurrence. Areal reduction factors (ARFs) are then used to convert extreme rainfall data at a point to estimates of areal rainfall (see, e.g., Charalambous et al. 2013). To derive a temporal pattern, rainfall can be de-aggregated in time via a hyetograph. The design rainfall input (e.g., from a 100-year precipitation event) is run through the hydrologic model with prescribed initial land surface conditions.

While the Design Event Approach (DEA) has been important for RFA, the assumptions required for the DEA have been criticized (Kucera et al. 2006; Rahman et al. 2002). A major limiting assumption is that the 100-year design precipitation event will result in a 100-year flood; in practice, however, the results are often biased or inconsistent (Rahman et al. 2002). To overcome the limitations of the DEA, Rahman et al. (2002) outline two alternative ways to approach RFA: 1) the continuous simulation approach, and 2) the joint probability approach. In the continuous simulation approach, a long period of continuous streamflows is simulated. This can be done, for example, by forcing a hydrologic model with a complete precipitation time series. The advantage of this is that the model is better able to capture antecedent conditions, which are critical to flood estimation (Ivancic and Shaw 2015). However, running the hydrologic

model continuously can be computationally expensive. The joint probability approach focuses on simulating large flood events by considering probabilistic inputs and model parameters, including their correlations. A key strategy for this is to employ the Monte Carlo (MC) simulation method, whereby input variables and parameters are sampled and used as inputs to a hydrologic model. This is repeated many times, resulting in many estimated flood peaks for use in the frequency analysis. The advantage of this is that it is less computationally intensive; the disadvantages are that it can result in unrealistic sets of input variable combinations, and it can miss the associated antecedent conditions.

Given the straightforward nature and computational advantages of MC simulation, several techniques have been advanced to move away from the traditional design event paradigm and to better capture the probabilistic nature of flooding. Several studies have demonstrated how the joint probability approach can improve upon the DEA (Rahman et al. 2002; Charalambous et al. 2013). Svensson et al. 2013 employed a Joint Probability/MC approach that run continuously, finding that flood frequency curves are most sensitive to rainfall characteristics and time between events. Despite subtle differences in how MC simulation has been applied to RFA, the approaches are statistical in nature, and they require the development of probabilistic representations of the key factors that influence runoff, such as duration and intensity/volume. Although these are the main factors, MC approaches can be even more detailed, fitting probabilistic representations of these variables and others (see e.g., Kottegoda et al. 2014).

Advances in radar and satellite observations, as well as numerical weather prediction, have allowed for better spatiotemporal measurement and simulation of rainfall. Observed or simulated precipitation fields are physically based and can be used as inputs for a distributed hydrologic model, yielding a corresponding flood hydrograph. This circumvents several limitations of RFA using point-based observations. For example, ARFs (areal reduction factors) can introduce significant errors, as rainfall characteristics at a point can be very different than those over a basin (Wright et al. 2014).

Spatiotemporal observations of precipitation extremes are often limited, motivating new techniques for their analysis. One such method is stochastic storm transposition (SST; Wright et al. 2020). SST develops a multi-event storm catalog, which is then resampled to generate many realizations of heavy precipitation events that can be shifted around a given watershed or domain. The catalog can include any type of intense episode of interest, whether from tropical storms, organized thunderstorm systems, or orographic enhancement. The main disadvantages of SST are that one is only resampling from a finite storm catalog and that the domain for doing the transposition is limited (e.g., by topography; Wright et al. 2020). Wright et al. (2020) illustrated SST in an end-to-end approach in which they first used a high-resolution, observation-based rainfall dataset (Stage IV multi-sensor quantitative precipitation estimates) to create an event catalog. They then drove the SST-shifted storms through a process-based distributed hydrologic model (WRF-Hydro) to see the differences in peak discharge. Yu et al. (2020) took the SST approach one step further, where in addition to the use of observations, they also demonstrated the use of CPM rainfall simulations from a Regional Climate Model (RCM). Yu et al. (2020) created a storm catalog from bias-corrected RCM output, which was resampled using the SST approach. That catalog was then used to drive WRF-Hydro to obtain flood quantiles for the 500-year recurrence interval. In short, CPM rainfall simulations offer a physically based alternative to limited rainfall observations and are discussed in the next section.

3.2   Convection-Permitting Model (CPM) Rainfall Event Simulations

3.2.1   Model Error and Variability in the CPM Forecast Ensembles

Given the forecast improvements from CPM configurations, weather forecasting centers are starting to apply them operationally. For short-term forecasting, the CPMs are initialized via analyses incorporating observations and run out to several days. Datasets of prior CPM forecasts have also been developed as resources for the weather forecast and research communities.

In Section 2.2 , three existing CPM forecast datasets were examined: (i) the NCAR MPEX Ensemble, (ii) the NCAR Real-Time Ensemble, and (iii) the NCAR SCS Deterministic Forecast collection. These provide (i) 10,570 simulations reflecting 3-km horizontal grid spacing (1.8 mi) and (ii) 810 simulations reflecting 1-km horizontal grid spacing (0.6 mi). These ensembles were evaluated above via comparison to observations, revealing several systematic errors, as well as substantial variability in storm characteristics within the ensembles. In this section, the effect of ensemble size on model error and variability are examined.

"Error" here is defined as the difference between the ensemble median and the observation, and this error is in large part due to the imperfect representation of rainfall processes in the model. Systematic errors must be corrected before the model output can be used as input to a flood hazard assessment (Wood et al. 2004), and statistical approaches are typically applied for this. The spread in the ensemble forecasts is a result of differences in initial conditions that lead to model integrations that diverge over time. Each member's forecast can be considered a possible outcome for an observed event. The "variability" is defined as the range of the values spanning a mean value from the set of forecasts. The range of forecasts can be used to provide examples of new storms or storm characteristics to augment the historical record used for the flood hazard assessment; Section 3.3 discusses this.

3.2.1.1   Effect of Ensemble Size on Model Error and Variability

To robustly assess the magnitude of the systematic model errors, we have examined ensemble simulations of heavy precipitation events in different regions. First, the effect of ensemble size on the mean error— the error in a variable or metric as averaged over the forecast ensemble members— has been analyzed (Figure 3-1 a,d,g). This analysis is based on 11 heavy precipitation events, identified in Section 2.2 , in the 3-km MPEX simulations; these events cover the period May 15–June 15, 2013, and the daily 30-member forecasts are analyzed. Here the ensemble mean error varies only marginally for ensemble sizes larger than 10 members for the metrics of: (i) the 99th percentile (P99) peak accumulation (Figure 3-1a), and (ii) the error in the location of the P99 peak accumulations (Figure 3-1d). We see that the mean errors in the volume of precipitation for simulated events demand a slightly larger ensemble size, of approximately 15 members, to converge (Figure 3-1g). This is broadly consistent with previous studies (e.g., Leutbecher 2019) and indicates that the model mean errors that are identified in Section 2.2 are robust, since they are mostly based on 10-member ensemble forecasts.

Second, we have analyzed the effect of ensemble size on ensemble spread. The spread of a forecast ensemble is a measure of the predictability of a heavy precipitation event, where each ensemble member can be regarded as a possible realization of the observed event. We find that the ensemble spread, here taken as the range of the 10th–90th percentiles, is significantly underestimated in small ensembles, but it begins to converge for ensembles of over 10 members. And, this is independent of the metric under investigation (Figure 3-1b,e,h). Fitting a

tangent-hyperbolic function allows us to estimate the asymptotic value for convergence of the ensemble spread. The differences in the asymptote estimation, when comparing results from ensembles of fewer members than our full set (30), show that using 10-member ensembles results in an underestimation of the ensemble spread by approximately 10% (Figure 3-1c,f,i).
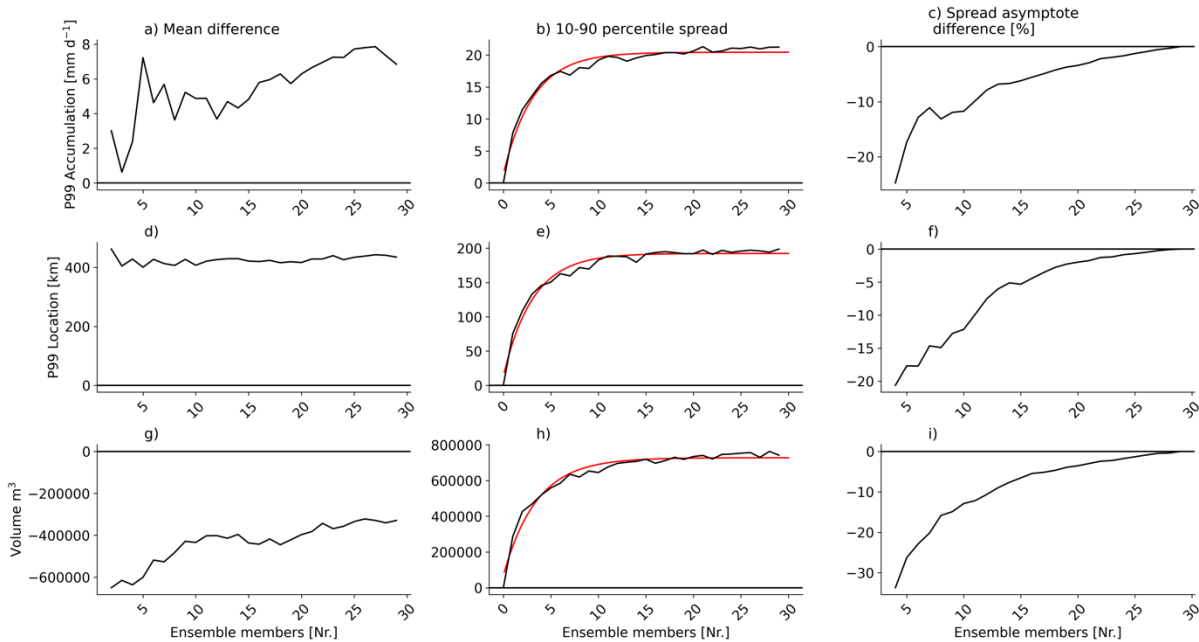
Figure 3-1 **Effect of Ensemble Size on the Estimation of P99 Accumulation, P99 Location, and Precipitation Volume: (First Column) Ensemble Mean Differences; (Second Column) 10–90th Percentile Ensemble Spread; and (Third Column) Estimate of the Spread Asymptote** *Mean differences are for 8 heavy precipitation events from the MPEX simulations (30-member, 3-km ensemble forecasts). Results are shown for the 99th percentile (P99) of daily peak accumulation (a-c), the differences in the location of the P99 accumulation (d-f), and the differences in event precipitation volume (g-i). All differences reflect a comparison of model values with those based on Stage IV observations. The red line in the ensemble spread analysis shows the best-fit estimate of a tangent-hyperbolic function to the model data, which is used to calculate the asymptote of the ensemble spread. The right column shows the difference between the asymptote derived from considering all 30 ensemble members compared to that derived from a reduced number of members. To increase statistical robustness, all results are based on the mean values of 1000 bootstrap samples.*

3.2.1.2 *Variability Sources*

To help identify foci of future CPM simulations of heavy precipitation events, a variance decomposition was performed on the 3-km NCAR ensemble dataset, calculating the contribution of four factors to the variance of forecast biases (Figure 3-2): 1) case to case variability (Nr), 2) ensemble spread (En), 3) observational uncertainties (Ob), and 4) differences

between seasons (Se). First-order mixtures of these factors are also considered. The 3-km NCAR ensemble dataset is used for this analysis due to its 10-member ensemble size and its inclusion of multiple heavy precipitation events in all regions.

The factors affecting variance differ across the geographic regions we have been considering. In the South region, the overall model variability in simulating the P99 precipitation accumulation is due to a combination of both case-to-case and seasonal variability (Figure 3-2a). This means that to better constrain model performance in terms of P99 precipitation there, it would be more effective to sample additional heavy precipitation events in different seasons rather than adding observational datasets or increasing the ensemble size. In contrast, the weights of the sources of variability are more balanced in the East Coast region. The greatest variability is found in the South region, which is mainly a reflection of the heavier precipitation events there; the least variability is in the Appalachian region (Figure 3-2d).

Uncertainties in simulating the locations of P99 events are generally attributed to a combination of case-to-case and seasonal variability, with the next-largest contribution coming from ensemble spread. The exception to this is the East Coast region, where the combination of ensemble spread, and seasonal variability is most important (Figure 3-2a). The largest uncertainty in simulating P99 location is in the Appalachian region, while the smallest is in the South (Figure 3-2e). Across most of the regions, the uncertainty in event precipitation volumes is dominated by ensemble spread and seasonal variability (Figure 3-2c). The variability in simulating precipitation volume is largest in the South and smallest in the Appalachian region, with the South's variability mainly reflecting the larger and more intense events.

Based on these results, the most effective factor in obtaining robust CPM-based estimates of the characteristics of heavy precipitation events is the capturing of events across different seasons. This is followed by increasing the forecast ensemble size, with this being especially true for estimates of the P99 location and P99 accumulation over the Appalachian and East Coast regions. Our analysis indicates that observation sampling uncertainty is of minor importance for most regions and for the given storm characteristics. However, the importance of this source of uncertainty would significantly increase if the gridded precipitation analyses used for verification were based on station data, in contrast to the Stage IV and MRMS datasets used.
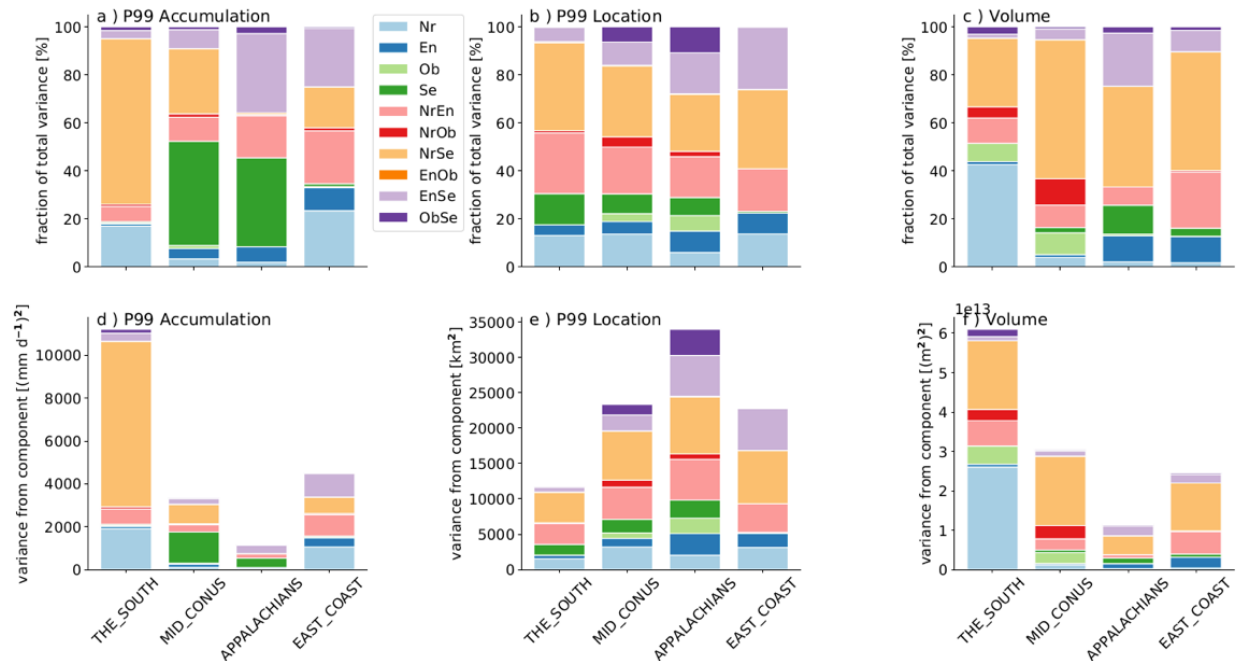
Figure 3-2 **Variance Decomposition of the Impact of the Number of Cases (Nr), Ensemble Members (En), Observational Uncertainties (Ob; Comparing Stage IV and MRMS), Seasonality (Se), and First-Order Mixture Terms on the P99 Event Accumulation (A,D), P99 Location (B,E), and Event Precipitation Volume (C,F)** *Results are based on the NCAR Ensemble simulations. The top row shows the relative contributions, and the bottom row shows the absolute contributions to the total variance. Seasonality is calculated by separating events into three periods within the year depending on their time of occurrence.*

3.2.2 Dynamically-Downscaled Datasets

CPM configurations are also being implemented in climate modeling efforts (Prein et al. 2015), including dynamical downscaling. Dynamical downscaling approaches utilize high-resolution, limited-area models that are forced by coarser-resolution boundary conditions, but otherwise freely evolve and develop their own weather and climate. In this methodology boundary conditions are available from observational reanalyses and from global climate model (GCM) output.

3.2.2.1 Observational Reanalyses

Given the relative sparsity of direct precipitation observations, climate reanalyses can provide effective boundary forcings for RCMs run with CPM configurations. Climate centers have developed a variety of reanalysis products for the 20th century, such as the NCEP/NCAR Reanalysis Project (Kalnay et al. 1996) and the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA; Uppala et al. 2005). ECMWF also has climate reanalysis based on a 10-member ensemble (CERA-20C; https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/cera-20c).

The disadvantage of downscaling reanalyses using CPMs is similar to that of using forecast datasets, in that although the CPM simulations provide information about uncertainty for observed events, they do not yield "new" events.

3.2.2.2 GCMs

The primary purpose of GCMs has been to investigate future climate. As GCMs are typically run at coarse resolutions, their configurations are not suited to the accurate simulation of precipitation extremes (Diffenbaugh et al. 2005). However, GCM output can be used as boundary conditions to drive higher-resolution models.

In the climate community, there has been a proliferation of GCM output for climate research and projection applications. For example, centers have provided GCM output for model comparison purposes (e.g., the Coupled Model Intercomparison Project Phase 3, 5, and 6; CMIP3, CMIP5, and CMIP6). In particular, the CMIP6 effort has a High-Resolution Model Intercomparison Project (HighResMIP; Haarsma et al. 2016) that is relevant to the investigation of small-scale processes, including extreme precipitation. Furthermore, there have been efforts to dynamically downscale GCM output using RCMs to 50-km or finer grids (31 mi), with a prominent example being the North American Regional Climate Change Assessment Program (NARCCAP; Mearns et al. 2011). In theory, any or all of these models could provide boundary conditions for downscaling to a CPM application. However, because of the computational effort required, it is not possible to so downscale every GCM or RCM. To address this issue, event-based approaches have been adopted. As one prominent example, Mahoney et al. (2013) conducted an event-based experiment in which three independent RCMs, each with a 50-km (31 mi) mesh, were downscaled to 1.3 km (0.8 mi) with the Weather Research and Forecasting (WRF) Model. In the study, Mahoney et al. (2013) identified the 60 largest 24-hr precipitation events from both the past and future in the RCM output to downscale. Results showed a large spread in projected changes, with extreme event intensities either staying the same or increasing.

GCM large-ensemble experiments have been conducted to investigate variability associated with the chaotic nature of the climate system. For example, NCAR's Community Earth System Model Large Ensemble (LENS) is a community resource (Kay et al. 2015) that offers a 40-member ensemble, with each member having a resolution of 1-degree latitude/longitude. Analysis of LENS has shown that there is considerable internal variability in the climate system, whether past or future (Deser et al. 2014). While the LENS resolution is too coarse to directly examine precipitation extremes, LENS and other large ensembles can be used with targeted event-based approaches to downscaling. To this end, extreme precipitation events, their or environments, can be identified in time series extracted from the ensemble. Huang et al. (2020), for example, detected extreme atmospheric river (AR) events in LENS by identifying environments conducive to ARs. Specifically, they identified the highest 5-day running mean values of integrated vapor transport (IVTs) in LENS. From this, they selected the 60 rarest/most intense events in the dataset, which they re-simulated with WRF at 3 km (1.8 mi). With this setup, Huang et al. (2020) found increases in total accumulated precipitation and in hourly maximum precipitation intensity. Recently, additional efforts to generate a collection of large (i.e., many-member) ensembles from multiple models ("multi-model large ensembles") have been realized (Deser et al. 2020). Multi-model large ensembles are a powerful new resource that can yield robust climate risk assessments, including assessments involving extremes, by considering uncertainties due to both initial conditions and model differences.

The main advantage of the use of GCMs in intense precipitation analysis is that they can provide boundary conditions to CPMs that can result in new, plausible storms for both historical

and future periods. The main disadvantage of using GCMs is that the approach requires identifying extreme events or environments in coarse-resolution model setups, which may be difficult and may be sensitive to the region or season of interest.

3.3   <u>Conceptual Framework Criteria and Integration</u>

This section has reviewed rainfall-based flood frequency analysis and CPM rainfall simulations towards the development of a conceptual framework for their integration. For their application in such a framework, CPM rainfall simulations can be assessed with the following criteria.

i.   Realism: CPM rainfall simulations must be realistic. This involves minimizing, correcting, or at least understanding systematic errors (i.e., biases) in the output rainfall fields.
ii.   Variability: For critical infrastructure applications such as nuclear power plant design and operation and for flood probability analysis, CPM simulations must capture the full range of variability of plausible intense rainfall scenarios.
iii. Computational Cost: Given that the goal is to develop a framework for analyzing events with return periods of 10,000 years or longer, the computational costs of CPM simulations and applications must be considered, and ideally, minimized.

Table 3-1 presents assessments of these three criteria in the application of CPMs for intense rainfall. We stress that these are based on the professional experience and judgment of the authors, with collectively many years working in the application and development of fine-scale atmospheric models. In terms of CPM forecast realism, although it has been established that CPM configurations can accurately simulate intense rainfall events, initial and lateral boundary conditions do have a direct and significant effect on the model error. As such, properly using input observations, as is done in forecasts and reanalyzes, and limiting forecast lengths to constrain error growth (e.g., doing shorter-term predictions) allow for higher confidence in the realism of the output. This confidence is somewhat reduced for the output of GCMs, as they are not constrained directly by observations and their projections cannot be verified in the same way as those of CPMs. Nonetheless, it has been shown that GCMs can skillfully reproduce the large-scale circulation patterns (Flato et al. 2013) that are used as the driving boundary conditions for CPMs; thus, in Table 3-1GCMs still are given a relatively high rating for realism.

In Table 3-1 the variability criteria show differences across the CPM rainfall simulation sources. Although CPM forecasts reflect perturbed initial conditions, the large-scale forcings and general weather patterns during the relatively short forecast cycle do not usually change sufficiently to allow for large divergence in the model solutions. As such, there are no "new" events. This is similar to the application of downscaled reanalyses, which are in large part constrained to the input observations. GCMs provide higher variability because they have less-constrained environments that can evolve to yield new conditions and events that are plausible, but that have not been captured in the record. The use of ensembles to create simulation datasets increases the variability across the board.

Computational cost is another important criterion. On the one hand, the rating of cost could be considered fixed for the different sources, since the CPM computational effort should be the same regardless of the source of the driving conditions. However, here it is acknowledged that there are existing operational forecasts that could be exploited as input for analysis, where the computation has already been performed, resulting in a lower cost of a study; thus, we have rated it more favorably. To produce new downscaled simulations from observational reanalyses or from GCMs would require similar computational resources. However, downscaling a

continuous record would require more computational power than doing so for only selected events.

Table 3-1   **CPM Rainfall Simulation Ratings for the Criteria of Realism, Variability, and Computational Cost by Driving Boundary Condition Source. Ratings are Color Coded Where Green Indicates the Highest/Best Rating, Orange is the Lowest/Worst Rating; and Yellow is in Between**

Source	CPM Rainfall Simulations					
	Realism	Variability		Cost		
Operational Forecasts	High	Lower	Higher (if ensembles)	Lesser (if existing)	Greater (continuous)	Lesser (event-based)
Downscale Reanalysis	High	Lower		Greater (if new)		
Downscale GCMs	Med/High	Higher		Greater (if new)		

Although the ratings in Table 3-1 are subjective, they provide guidance for the conceptual framework for how CPMs can be integrated into flood hazard assessment. Next, we present the conceptual frameworks for the joint probability, SST, and continuous approaches, with a focus on how CPMs can be used to mitigate drawbacks in each approach.

Joint probability: The main disadvantage of the joint probability approach is that it can present unrealistic parameter combinations, and therefore yield unrealistic floods. Adding data from CPM simulations of heavy rainfall events can be used to improve fits of the probability distributions and correlations, thus better constraining the parameter combinations and dependencies. Although the CPM information is spatially explicit, the observed record could be augmented by examining rainfall characteristics within a prescribed "neighborhood" (i.e., radius around a station of interest); this is possible by using either the neighborhood maximum or the neighborhood average. While joint probability approaches are often tailored to a particular application, at a minimum they all require fitting probability distributions to the characteristics of rainfall duration, intensity, and temporal evolution— all of which would be available from CPMs. Adding to the distributions using events from CPM simulations with higher variability would be ideal. This could be achieved using CPM ensembles from any source, with those from GCM ensembles offering the most variability (Table 3-1). Based on the variance decomposition above, we find that the most effective means of increasing the variability is to simulate additional events in different seasons; the next most effective means is increasing the forecast ensemble size. In terms of the information on frequency used to develop the IDF curves, only GCMs could offer new input on this; otherwise, frequency information would be limited to the available record.

SST: The main disadvantage of the SST approach is that it relies on a finite storm catalog that can only be shifted within a limited domain. A dataset of CPM rainfall simulations, however, can

be readily mined to increase the storm catalog size. Further, the greater the storm variability in the catalog, the less it matters that the storms can only be shifted within a limited domain. Thus, capitalizing on greater variability in CPM output would be ideal. Greater variability can be achieved from the use of CPM ensembles, especially those driven from GCM output, as well as by using CPM simulations from different seasons.

Continuous Approaches: The main disadvantage of continuous approaches is the greater computational effort needed both to develop the hydrometeorological inputs and then to use them to drive a hydrologic model. The computational demand of adopting a continuous approach is more of an issue for the use of complex hydrologic models, such as distributed hydrologic models, than it is for simpler rainfall-runoff models. Nevertheless, given the importance of antecedent conditions on flooding (Ivancic and Shaw 2015, Sharma et al. 2018; Berghuijs et al. 2016) and general increases in available computing power, continuous approaches, especially those using observational reanalyses, are being tried more often. For example, a new partnership between NCAR and the USGS will develop a 40-year (1980–2019) high-resolution (4-km; 2.5 mi) retrospective of both WRF-simulated rainfall and WRF-Hydro-simulated streamflows across the contiguous United States (CONUS). There will also be a 20-year projection using the pseudo global warming (PGW) approach under the RCP8.5 (Representative Concentration Pathway) greenhouse gas emission scenario[5]. Additional 1-km (0.6-mi) grid simulations will be performed in USGS Next Generation Water Observing System (NGWOS) focus basins, including the Delaware River basin, the upper Colorado River basin, and the Illinois River basin. Although it is not being developed explicitly for flood frequency analysis, this dataset will provide a valuable high-resolution reanalysis that could be leveraged in such an effort. Furthermore, it provides a concrete reference for how CPM rainfall simulations can be integrated into hydrologic modeling in a continuous framework.

3.4   Recommendations

Based on the analyses and findings above we present two main recommendations for future work with CPMs for intense precipitation analysis and simulation.

Recommendation 1. Perform targeted downscaling of GCM large ensembles. Regardless of the RFA approach used, there is a need for the rainfall dataset to contain a wide range of storm events affecting a region or watershed, especially plausible episodes that are not part of the record. As such, targeted downscaling of GCM ensembles is a valuable, productive approach, and we believe this would result in a step-change in the variability of storm events considered.

Figure 3 offers an illustration of the recommendation. The first step would be to analyze NCAR's Community Earth System Model Large Ensemble (LENS) to identify the environments for intense precipitation events, which would vary with region and season, for contemporary and future periods. This work would have to rely on meteorological expertise for dataset interpretation, analysis, and evaluation. In the second step, a model in a convection-permitting configuration, such as WRF at 3 km, would be used to simulate new precipitation events in a target region. As has been done in several recent studies (e.g., Mahoney et al. 2013, Huang et al. 2020), this could be done for both past and future periods. This downscaling effort would demand significant compute resources. To provide some context, one CONUS-scale 3-km, 36-hr WRF forecast takes approximately 2000 core-hours on the NCAR supercomputer. Running 10-member ensembles would thus take about 20,000 core-hours of HPC time per case. And,

---

[5] RCP8.5 is the current emissions level/"business as usual" assumption for CO_2 in climate modeling.

running 100 cases for the 10-member ensemble (i.e., 1,000 simulations), would require 2.0M (million) core-hours. While this is not a trivial resource need, it is within the level of allocation that high-performance computers (HPCs) like NCAR's can provide.

In the third step, the 3-km WRF output could drive WRF-Hydro to simulate new flood events. It is estimated that driving CPM data (from WRF) through WRF-Hydro carries about a 15% computational overhead. Using this estimate, to run 100 cases (1,000 simulations), the compute resources needed to run WRF and WRF-Hydro would be ~2.3M core-hours.

Recommendation 2. Utilize CPM rainfall simulations to expand the storm catalog for SST. Building on the previous recommendation, the rainfall simulations from Step 2 (Figure 3-3) could be used to increase and diversify the event population for SST analysis. However, even in the absence of new simulations, the previously described 40-year, 4-km (2.5-mi) retrospective modeling study could be mined to develop a storm catalog and flood peak database for the CONUS[6]. As mentioned, WRF-Hydro will have already been calibrated to ingest the retrospective WRF output, which could be utilized for event-based SST. This could simplify flood risk assessments at specific sites.

Step 1. Scan NCAR's Community Earth System Model Large Ensemble (LENS) to identify the top extreme environments for current and future periods.

Step 2. Use 3-km WRF to simulate new precipitation events in target region.

Step 3. Run 3-km WRF output through WRF-Hydro to simulate new flood events.
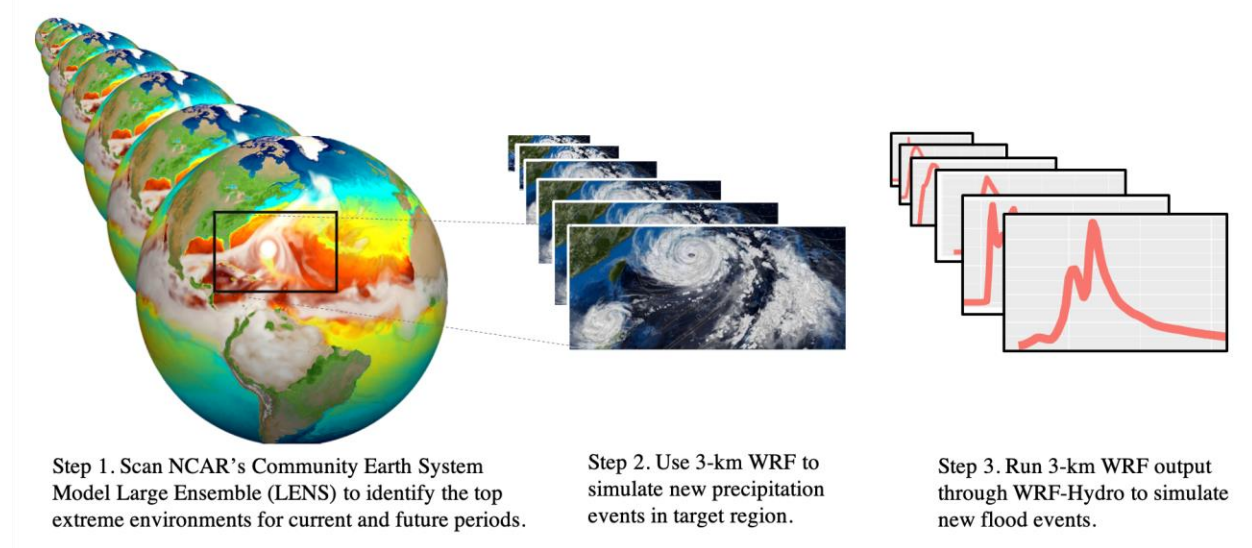
Figure 3-3   Recommended Steps to Perform Targeted Downscaling of GCM Large Ensembles. Targeted Downscaling Allows for the Creation of Physically Plausible, but So-far Unseen, Heavy Precipitation Events for Past, Current, and Future Conditions that can be Used in Probabilistic Flood Risk Assessments

Finally, we close with two additional remarks.

Integrating CPM rainfall simulations is not straightforward for RFA approaches that have traditionally used point-based rainfall. The spatially explicit nature of CPM simulations is somewhat mismatched with RFA approaches that have traditionally utilized point-based rainfall. Though it would be possible for maximum or average rainfall characteristics within a

---

[6] This will likely be publicly available in 2022.

neighborhood to be applied to improve the fits of the probability distributions, there would be some subjectivity in this process, and the authors are not aware of any studies that have demonstrated it to date. As such, the approach would need to be tested and refined in terms of its utility for improving/constraining the distributional fits and the resulting estimated floods. If desired, NRC could select several case study locations to modify the framework to demonstrate how this could be done. Starting with the existing forecast ensembles, for example, the sensitivity of the input probabilistic distributions to the new data could be tested. Once the framework had been established, it could be extended for use with other CPM datasets (e.g., reanalysis, ensembles, etc.).

The sensitivity of flooding to land surface conditions should be further investigated. While this report focuses on rainfall characteristics as a key driver of flooding, land surface conditions, such as antecedent soil moisture, are also critical factors and have been less explored. Continuous approaches using coupled models, such as the noted 40-year, high-resolution retrospective of WRF and WRF-Hydro, could provide valuable datasets that could be examined for this purpose. However, undertaking a continuous dynamical run for many years and ensembles will be computationally prohibitive for some time. As an alternative, hybrid statistical-dynamical approaches could be investigated. These could be used to complement dynamical event-based approaches to better account for the role of antecedent conditions.

As one strategy, sensitivity analyses could first be conducted to see how long of a simulation prior to an event is needed to improve flood predictions. Second, the coarse-resolution boundary conditions of the preceding days could be examined, and from this, statistical relationships between the large-scale forcings and local conditions could be developed to determine the likelihood of wet or dry conditions. This could be used in a probabilistic framework to estimate the relevant land surface parameters. Another important consideration is that flooding can result from multiple moderate rainfall events in a sequence or from a combination of event types (e.g., rain on snow), and recent work has shown how such compound events need to be considered in future hazard projections (Zscheischler et al. 2018). Thus, it would also be valuable to explore the role of compound events in the context of flood hazard assessment.

4 REFERENCES

Abbs, D. J. (1999), A numerical modeling study to investigate the assumptions used in the calculation of probable maximum precipitation. Water Resour. Res., 35(3), 785–796, doi:10.1029/1998WR900013.

Ahasan, M.N., Chowdhury, M.A.M. and Quadir, D.A., 2014. Sensitivity test of parameterization schemes of MM5 model for prediction of the high impact rainfall events over Bangladesh. Journal of Mechanical Engineering, 44(1), pp.33-42.

Andrejczuk, M., Grabowski, W.W., Reisner, J., and Gadian, A., 2010. Cloud-aerosol interactions for boundary layer stratocumulus in the Lagrangian Cloud Model. Journal of Geophysical Research: Atmospheres, 115(D22).

Andrejczuk, M., Reisner, J.M., Henson, B., Dubey, M.K. and Jeffery, C.A., 2008. The potential impacts of pollution on a nondrizzling stratus deck: Does aerosol number matter more than type?. Journal of Geophysical Research: Atmospheres, 113(D19).

Arakawa, A. and Schubert, W.H., 1974. Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. Journal of the Atmospheric Sciences, 31(3), pp.674-701.

Archfield, S. A., Hirsch, R. M., Viglione, A., & Blöschl, G. 2016. Fragmented patterns of flood change across the United States. Geophysical Research Letters, 43, 10,232–10, 239. https://doi.org/10.1002/2016GL070590

Ban, N., Schmidli, J. and Schär, C., 2014. Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. Journal of Geophysical Research: Atmospheres, 119(13), pp.7889-7907.

Barlage, M., Chen, F., Miguez-Macho, G., Liu, C., Liu, X., and Niyogi, D., 2018. Enhancing Hydrologic Processes in the Noah-MP Land Surface Model to Improve Seasonal Forecast Skill. AMS Annual Meeting 2018, https://ams.confex.com/ams/98Annual/webprogram/Paper334298.html (accessed on Oct. 2nd, 2019)

Barlow, M., Gutowski, W.J., Gyakum, J.R., Katz, R.W., Lim, Y.K., Schumacher, R.S., Wehner, M.F., Agel, L., Bosilovich, M., Collow, A. and Gershunov, A., 2019. North American extreme precipitation events and related large-scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. Climate Dynamics, pp.1-41.

Bartsotas, N. S., E. I. Nikolopoulos, E. N. Anagnostou, S. Solomos, and G. Kallos, 2017: Moving toward subkilometer modeling grid spacings: Impacts on atmospheric and hydrological simulations of extreme flash flood–inducing storms. J. Hydrometeor., 18, 209–226, https://doi.org/10.1175/JHM-D-16-0092.1.

Beauchamp, J., R. Leconte, M. Trudel, and F. Brissette, 2013: Estimation of the summer-fall PMP and PMF of a northern watershed under a changed climate. Water Resour. Res., 49, 3852–3862, doi:10.1002/wrcr.20336.

Beck, H. E., and Coauthors, 2019: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. Hydrol. Earth Syst. Sci, 23, 207–224.

Beck, H.E., Van Dijk, A.I., Levizzani, V., Schellekens, J., Gonzalez Miralles, D., Martens, B. and De Roo, A., 2017. MSWEP: 3-hourly 0.25 global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. Hydrology and Earth System Sciences, 21(1), pp.589-615.

Benedict, I., Ødemark, K., Nipen, T. and Moore, R., 2019. Large-scale flow patterns associated with extreme precipitation and atmospheric rivers over Norway. Monthly Weather Review, 147(4), pp.1415-1428.

Benjamin, S.G., Weygandt, S.S., Brown, J.M., Hu, M., Alexander, C.R., Smirnova, T.G., Olson, J.B., James, E.P., Dowell, D.C., Grell, G.A. and Lin, H., 2016. A North American hourly assimilation and model forecast cycle: The Rapid Refresh. Monthly Weather Review, 144(4), pp.1669-1694.

Berghuijs, W. R., Woods, R. A., Hutton, C. J., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. Geophysical Research Letters, 43(9), 4382–4390. https://doi.org/10.1002/2016GL068070

Bonnin, G.M., Martin, D., Lin, B., Parzybok, T., Yekta, M. and Riley, D., 2004. NOAA Atlas 14: Precipitation-frequency atlas of the United States. US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, Maryland.

Bradley, A.A. and J.A. Smith, 1994: The hydrometeorological environment of extreme rainstorms in the Southern Plains of the United States. J. Appl. Meteor., 33, 1418–1431, https://doi-org.cuucar.idm.oclc.org/10.1175/1520-0450(1994)033<1418:THEOER>2.0.CO;2

Brast, M., Schemann, V. and Neggers, R.A., 2018. Investigating the scale adaptivity of a size-filtered mass flux parameterization in the gray zone of shallow cumulus convection. Journal of the Atmospheric Sciences, 75(4), pp.1195-1214.

Bright, D.R. and Mullen, S.L., 2002. The sensitivity of the numerical simulation of the southwest monsoon boundary layer to the choice of PBL turbulence parameterization in MM5. Weather and Forecasting, 17(1), pp.99-114.

Bruyère, et al., 2017. Impact of climate change on Gulf of Mexico hurricanes. Technical Note NCAR/TN-535+ STR.(NCAR, 2017).

Bryan, G. H., and H. Morrison, 2012: Sensitivity of a simulated squall line to horizontal resolution and parameterization of microphysics. Mon. Wea. Rev., 140, 202–225, https://doi.org/10.1175/MWR-D-11-00046.1.

Buzzi, A., S. Davolio, P. Malguzzi, O. Drofa, and D. Mastrangelo, 2014: Heavy rainfall episodes over Liguria of autumn 2011: Numerical forecasting experiments. Nat. Hazards Earth Syst. Sci., 14, 1325–1340, https://doi.org/10.5194/nhess-14-1325-2014.

Caracena, F., Maddox, R.A., Hoxit, L.R. and Chappell, C.F., 1979. Mesoanalysis of the Big Thompson storm. Monthly Weather Review, 107(1), pp.1-17.

Carbone, R.E., Tuttle, J.D., Ahijevych, D.A. and Trier, S.B., 2002. Inferences of predictability associated with warm season precipitation episodes. Journal of the Atmospheric Sciences, 59(13), pp.2033-2056.

Chan, J.C., 1985. Identification of the steering flow for tropical cyclone motion from objectively analyzed wind fields. Monthly weather review, 113(1), pp.106-116.

Chan, S.C., Kendon, E.J., Fowler, H.J., Blenkinsop, S., Roberts, N.M. and Ferro, C.A., 2014. The value of high-resolution Met Office regional climate models in the simulation of multi hourly precipitation extremes. Journal of Climate, 27(16), pp.6155-6174.

Charalambous, J., Rahman, A., & Carroll, D. (2013). Application of Monte Carlo Simulation Technique to Design Flood Estimation: A Case Study for North Johnstone River in Queensland, Australia. Water Resources Management, 27(11). https://doi.org/10.1007/s11269-013-0398-9

Chen, X. and F. Hossain, 2018: Understanding model-based probable maximum precipitation estimation as a function of location and season from atmospheric reanalysis. J. Hydrometeor., 19, 459–475, https://doi.org/10.1175/JHM-D-17-0170.1

Chen, X., F. Hossain, and L. R. Leung, 2017: Establishing a numerical modeling framework for hydrologic engineering analyses of extreme storm events. J. Hydrologic Engr., 22(8 https://doi.org/10.1061/(ASCE)HE.1943-5584.0001523.

Clark, A. J., and Coauthors, 2011: Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. Mon. Wea. Rev., 139, 1410–1418, https://doi.org/10.1175/2010MWR3624.1.

Clark, A.J., Bullock, R.G., Jensen, T.L., Xue, M. and Kong, F., 2014. Application of object-based time-domain diagnostics for tracking precipitation systems in convection-allowing models. Weather and Forecasting, 29(3), pp.517-542.

Clark, A.J., Gallus Jr, W.A. and Weisman, M.L., 2010. Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF model simulations and the operational NAM. Weather and Forecasting, 25(5), pp.1495-1509.

Clark, M. P., and A. G. Slater, 2006: Probabilistic quantitative precipitation estimation in complex terrain. Journal of Hydrometeorology, 7 (1), 3–22.

Clark, P., Roberts, N., Lean, H., Ballard, S.P. and Charlton-Perez, C., 2016. Convection-permitting models: a step-change in rainfall forecasting. Meteorological Applications, 23(2), pp.165-181.

Cohen, A.E., Cavallo, S.M., Coniglio, M.C. and Brooks, H.E., 2015. A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern US cold season severe weather environments. Weather and forecasting, 30(3), pp.591-612.

Coles, S., Bawa, J., Trenner, L. and Dorazio, P., 2001. An introduction to statistical modeling of extreme values (Vol. 208, p. 208). London: Springer.

Colle, B. A., and C. F. Mass, 2000: The 5–9 February 1996 flooding event over the Pacific Northwest: Sensitivity studies and evaluation of the MM5 precipitation forecasts. Mon. Wea. Rev., 128, 593–618, https://doi.org/10.1175/1520-0493(2000)128<0593:TFFEOT>2.0.CO;2.

Colle, B. A., J. B. Wolfe, W. J. Steenburgh, D. E. Kingsmill, J. A. W. Cox, and J. C. Shafer, 2005: High-resolution simulations and microphysical validation of an orographic precipitation event over the Wasatch Mountains during IPEX IOP3. Mon. Wea. Rev., 133, 2947–2971, https://doi.org/10.1175/MWR3017.1.

Colle, B.A., 2004. Sensitivity of orographic precipitation to changing ambient conditions and terrain geometries: An idealized modeling perspective. Journal of the Atmospheric Sciences, 61(5), pp.588-606.

Coniglio, M.C., Correia Jr, J., Marsh, P.T. and Kong, F., 2013. Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. Weather and Forecasting, 28(3), pp.842-862.

Contractor, S., Alexander, L.V., Donat, M.G. and Herold, N., 2015. How well do gridded datasets of observed daily precipitation compare over Australia?. Advances in Meteorology, 2015.

Crosson, W. L., C. E. Duchon, R. Raghavan, and S. J. Goodman, 1996: Assessment of rainfall estimates using a standard ZR relationship and the probability matching method applied to composite radar data in central Florida. Journal of Applied Meteorology, 35 (8), 1203–1219.

Dahlgren, P., Kållberg, P., Landelius, T. and Undén, P., 2014. EURO4M Project Report, D 2.9 Comparison of the Regional Reanalyses Products with Newly Developed and Existing State-of-the Art Systems. Technical Report, available at: http://www. euro4m. eu/Deliverables. html (last access: 10 June 2018).

Daly, C. et al, 2002: A knowledge-based approach to the statistical mapping of climate. Climate Res., 22, 99–113.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J. and Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology: a Journal of the Royal Meteorological Society, 28(15), pp.2031-2064.

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. Journal of Applied Meteorology, 33 (2), 140–158.

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

Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. Climate Research, 22 (2), 99–113.

Davis, C.A. and Bosart, L.F., 2001. Numerical simulations of the genesis of Hurricane Diana (1984). Part I: Control simulation. Monthly Weather Review, 129(8), pp.1859-1881.

Déqué, M., Rowell, D.P., Lüthi, D., Giorgi, F., Christensen, J.H., Rockel, B., Jacob, D., Kjellström, E., De Castro, M. and van den Hurk, B.J.J.M., 2007. An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. Climatic Change, 81(1), pp.53-70.

Derin, Y. and Yilmaz, K.K., 2014. Evaluation of multiple satellite-based precipitation products over complex topography. Journal of Hydrometeorology, 15(4), pp.1498-1516.

Deser, C., Lehner, F., Rodgers, K.B., Ault, T., Delworth, T.L., DiNezio, P.N., Fiore, A., Frankignoul, C., Fyfe, J.C., Horton, D.E. & Kay, J.E. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. Nature Climate Change, pp.1-10.

Diffenbaugh, N.S., Pal, J.S., Trapp, R.J. & Giorgi, F. (2005). Fine-scale processes regulate the response of extreme events to global climate change. Proceedings of the National Academy of Sciences, 102(44), pp.15774-15778.

Do, H. X., Westra, S., & Leonard, M. (2017). A global-scale investigation of trends in annual maximum streamflow. Journal of Hydrology, 552, 28–43. https://doi.org/10.1016/j.jhydrol.2017.06.015

Done, J., Davis, C.A. and Weisman, M., 2004. The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. Atmospheric Science Letters, 5(6), pp.110-117.

Doswell III, C.A., H. E. Brooks, H.E. and R. A. Maddox, R.A., 1996:. Flash flood forecasting: An ingredients-based methodology. Weather and Forecasting, 11(4), pp.560-581.

Doswell, C.A., 2001. Severe convective storms—An overview. In Severe Convective Storms (pp. 1-26). American Meteorological Society, Boston, MA.

Efstathiou, G.A., Zoumakis, N.M., Melas, D., Lolis, C.J. and Kassomenos, P., 2013. Sensitivity of WRF to boundary layer parameterizations in simulating a heavy rainfall event using different microphysical schemes. Effect on large-scale processes. Atmospheric Research, 132, pp.125-143.

Emanuel, K.A., 1991. A scheme for representing cumulus convection in large-scale models. Journal of the Atmospheric Sciences, 48(21), pp.2313-2329.

Fan, J., Han, B., Varble, A., Morrison, H., North, K., Kollias, P., Chen, B., Dong, X., Giangrande, S.E., Khain, A. and Lin, Y., 2017. Cloud-resolving model intercomparison of an MC3E squall line case: Part I—Convective updrafts. Journal of Geophysical Research: Atmospheres, 122(17), pp.9351-9378.

Fankhauser, J.C., 1988. Estimates of thunderstorm precipitation efficiency from field measurements in CCOPE. Monthly Weather Review, 116(3), pp.663-684.

Feng, Z., Leung, L.R., Hagos, S., Houze, R.A., Burleyson, C.D. and Balaguru, K., 2016. More frequent intense and long-lived storms dominate the springtime trend in central US rainfall. Nature Communications, 7, p.13429.

Feng, Z., Leung, L.R., Houze Jr, R.A., Hagos, S., Hardin, J., Yang, Q., Han, B. and Fan, J., 2018. Structure and evolution of mesoscale convective systems: Sensitivity to cloud microphysics in convection-permitting simulations over the United States. Journal of Advances in Modeling Earth Systems, 10(7), pp.1470-1494.

Ferrier, B.S., Simpson, J. and Tao, W.K., 1996. Factors responsible for precipitation efficiencies in midlatitude and tropical squall simulations. Monthly Weather Review, 124(10), pp.2100-2125.

Field, P.R., Hogan, R.J., Brown, P.R.A., Illingworth, A.J., Choularton, T.W. and Cotton, R.J., 2005. Parametrization of ice-particle size distributions for mid-latitude stratiform cloud. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131(609), pp.1997-2017.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C. & Coauthors. (2013). Evaluation of climate models. In: Stocker, T.F., Qin, D., Plattner, G-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (eds), Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Foote, G.B. and Du Toit, P.S., 1969. Terminal velocity of raindrops aloft. Journal of Applied Meteorology, 8(2), pp.249-253.

Førland, E.J., Allerup, P., Dahlström, B., Elomaa, E., Jónsson, T., Madsen, H., Perälä, J., Rissanen, P., Vedin, H. and Vejen, F., 1996. Manual for operational correction of Nordic precipitation data. Klima Report, 24, p.96.

Fujita, M., Mizuta, R., Ishii, M., Endo, H., Sato, T., Okada, Y., Kawazoe, S., Sugimoto, S., Ishihara, K. and Watanabe, S., 2019. Precipitation Changes in a Climate With 2-K Surface Warming From Large Ensemble Simulations Using 60-km Global and 20-km Regional Atmospheric Models. Geophysical Research Letters, 46(1), pp.435-442.

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

Gallus Jr, W.A. and Bresch, J.F., 2006. Comparison of impacts of WRF dynamic core, physics package, and initial conditions on warm season rainfall forecasts. Monthly Weather Review, 134(9), pp.2632-2641.

Garvert, M. F., B. A. Colle, and C. F. Mass, 2005: The 13–14 December 2001 IMPROVE-2 event. Part I: Synoptic and mesoscale evolution and comparison with a mesoscale model simulation. J. Atmos. Sci., 62, 3474–3492, https://doi.org/10.1175/JAS3549.1.

Garvert, M.F., Smull, B. and Mass, C., 2007. Multiscale mountain waves influencing a major orographic precipitation event. Journal of the Atmospheric Sciences, 64(3), pp.711-737.

Gentry, M.S. and Lackmann, G.M., 2010. Sensitivity of simulated tropical cyclone structure and intensity to horizontal resolution. Monthly Weather Review, 138(3), pp.688-704.

Ghosal, S., 1996. An analysis of numerical errors in large-eddy simulations of turbulence. Journal of Computational Physics, 125(1), pp.187-206.

Gochis, D., Schumacher, R., Friedrich, K., Doesken, N., Kelsch, M., Sun, J., Ikeda, K., Lindsey, D., Wood, A., Dolan, B. and Matrosov, S., 2015. The great Colorado flood of September 2013. Bulletin of the American Meteorological Society, 96(9), pp.1461-1487.

Goodison, B.E., Louie, P.Y. and Yang, D., 1998. WMO solid precipitation measurement intercomparison (p. 212). Geneva, Switzerland: World Meteorological Organization.

Gowan, T. M., W. J. Steenburgh, and C. S. Schwartz, 2018: Validation of mountain precipitation forecasts from the convection-permitting NCAR ensemble and operational forecast systems over the western United States. Wea. Forecasting, 33, 739–765, https://doi.org/10.1175/WAF-D-17-0144.1.

Grell, G., J. Dudhia, and D. R.Stauffer, 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech. Note. NCAR/TN-398+STR. 138pp. 10.5065/D60Z716B.

Grell, G.A. and Freitas, S.R., 2014. A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. Atmos. Chem. Phys, 14(10), pp.5233-5250.

Groisman, P. Y., Knight, R. W., & Karl, T.R. (2001). Heavy precipitation and high streamflow in the contiguous United States: Trends in the twentieth century. Bulletin of the American Meteorological Society, 82(2), 219–246. https://doi.org/10.1175/1520-0477(2001)082<0219:HPAHSI>2.3.CO;2

Groisman, P. Y., R. W. Knight, D. R. Easterling, T. R. Karl, G. C. Hegerl, & Razuvaev, V.N. (2005). Trends in intense precipitation in the climate record, J. Clim., 18(9), 1326–1350, doi:10.1175/JCLI3339.1.

Grubišić, V. , R. K. Vellore , and A. W. Huggins , 2005: Quantitative precipitation forecasting of wintertime storms in the Sierra Nevada: Sensitivity to the microphysical parameterization and horizontal resolution. Mon. Wea. Rev., 133, 2834– 2859, https://doi.org/10.1175/MWR3004.1.

Gutmann, E.D., Rasmussen, R.M., Liu, C., Ikeda, K., Bruyere, C.L., Done, J.M., Garrè, L., Friis-Hansen, P. and Veldore, V., 2018. Changes in hurricanes from a 13-yr convection-permitting pseudo–global warming simulation. Journal of Climate, 31(9), pp.3643-3657.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., & von Storch, J.-S. (2016). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geosci. Model Dev., 9, 4185-4208, doi:10.5194/gmd-9-4185-2016.

Hagelin, S., J. Son, R. Swinbank, A. McCabe, N. Roberts, and W. Tennant, 2017: The Met Office convective-scale ensemble, MOGREPS-UK. Q.J.R. Meteorol. Soc., 143, 2846–2861, https://doi.org/10.1002/qj.3135.

Hall, A., 2019, Why changes in extreme precipitation are different upon downscaling: a case study in California, 2019 Latsis Symposium, https://ethz.ch/content/dam/ethz/special-interest/conference-websites-dam/latsis-2019-dam/documents/abstracts_summary_final.pdf (10.23.2019)

Harrold TW. Mechanisms influencing the distribution of precipitation within baroclinic disturbances. Quarterly Journal of the Royal Meteorological Society. 1973 Apr;99(420):232-51.

Hart, K. A. , W. J. Steenburgh , and D. J. Onton , 2005: Model forecast improvements with decreased horizontal grid spacing over finescale intermountain orography during the 2002 Olympic Winter Games. Wea. Forecasting, 20, 558–576, https://doi.org/10.1175/WAF865.1.

Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D. and New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. Journal of Geophysical Research: Atmospheres, 113(D20).

Heiss, W. H., D. L. McGrew, and D. Sirmans, 1990: NEXRAD: next generation weather radar (WSR-88D). Microwave Journal, 33 (1), 79–89.

Herman G. R. and R. S. Schumacher 2016: Extreme precipitation in models: An Evaluation, Wea. Forecasting, 31, 1853-1879.

Hill, K.A. and Lackmann, G.M., 2009. Influence of environmental humidity on tropical cyclone size. Monthly Weather Review, 137(10), pp.3294-3315.

Hodgkins, G. A., Whitfield, P. H., Burn, D. H., Hannaford, J., Renard, B., Stahl, K., Fleig, A. K., Madsen, H., Mediero, L., Korhonen, J., Murphy, C., & Wilson, D. (2017). Climate-driven variability in the occurrence of major floods across North America and Europe. Journal of Hydrology, 552, 704–717. https://doi.org/10.1016/j.jhydrol.2017.07.027

Hofstra, N., New, M. and McSweeney, C., 2010. The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. Climate Dynamics, 35(5), pp.841-858.

Hong, S.Y. and Lim, J.O.J., 2006. The WRF single-moment 6-class microphysics scheme (WSM6). Asia-Pacific Journal of Atmospheric Sciences, 42(2), pp.129-151.

Hong, S.Y. and Pan, H.L., 1996. Nonlocal boundary layer vertical diffusion in a medium-range forecast model. Monthly weather review, 124(10), pp.2322-2339.

Hong, S.Y., Dudhia, J. and Chen, S.H., 2004. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. Monthly Weather Review, 132(1), pp.103-120.

Hong, S.Y., Noh, Y. and Dudhia, J., 2006. A new vertical diffusion package with an explicit treatment of entrainment processes. Monthly weather review, 134(9), pp.2318-2341.

Horak, J., Hofer, M., Maussion, F., Gutmann, E., Gohm, A. and Rotach, M.W., 2019. Assessing the added value of the Intermediate Complexity Atmospheric Research (ICAR) model for precipitation in complex topography. Hydrology and Earth System Sciences, 23(6), pp.2715-2734.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y. and Rauser, F., 2017. The art and science of climate model tuning. Bulletin of the American Meteorological Society, 98(3), pp.589-602.

Houze Jr, R.A., 2004. Mesoscale convective systems. Reviews of Geophysics, 42(4).

Huang, X., Swain, D., & Hall A. (2020). Future precipitation increase from very high resolution ensemble downscaling of extreme atmospheric river storms in California, Science Advances, 6(29), DOI: 10.1126/sciadv.aba1323.

Hughes, M., Hall, A. and Fovell, R.G., 2009. Blocking in areas of complex topography, and its influence on rainfall distribution. *Journal of Atmospheric Sciences*, *66*(2), pp.508-518.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, doi:10.1017/CBO9781107415324.

Ishida, K., M. L. Kavvas, S. Jang, Z. Q. Chen, N. Ohara, and M. Anderson: Physically based estimation of maximum precipitation over three watersheds in Northern California: Atmospheric boundary condition shifting. J. Hydrologic Engr., 0401-5014, doi:10.1061/(ASCE)HE.1943-5584.0001175.

Ivancic, T. J., & Shaw, S. B. (2015). Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge. Climatic Change, 133(4), 681–693. https://doi.org/10.1007/s10584-015-1476-1

Janjić, Z.I., 1990. The step-mountain coordinate: Physical package. Monthly Weather Review, 118(7), pp.1429-1443.

Janjić, Z.I., 1994. The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. Monthly weather review, 122(5), pp.927-945.

Jankov, I., J. Beck, J. Wolff, M. Harrold, J. B. Olson, T. Smirnova, C. Alexander, and J. Berner, 2019: Stochastically perturbed parameterizations in an HRRR-based ensemble. Mon. Wea. Rev., 147, 153–173, https://doi.org/10.1175/MWR-D-18-0092.1.

Jin, H., Peng, M.S., Jin, Y. and Doyle, J.D., 2014. An evaluation of the impact of horizontal resolution on tropical cyclone predictions using COAMPS-TC. Weather and Forecasting, 29(2), pp.252-270.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. Wea. Forecasting, 23, 931–952, https://doi.org/10.1175/WAF2007106.1.

Kain, J.S. and Fritsch, J.M., 1993. Convective parameterization for mesoscale models: The Kain-Fritsch scheme. In The representation of cumulus convection in numerical models (pp. 165-170). American Meteorological Society, Boston, MA.

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

Kaplan, J., DeMaria, M. and Cione, J., 2011. Improvement in the rapid intensity index by incorporation of inner core information. JHT Final Report.

Kendon, E.J., Roberts, N.M., Fowler, H.J., Roberts, M.J., Chan, S.C. and Senior, C.A., 2014. Heavier summer downpours with climate change revealed by weather forecast resolution model. Nature Climate Change, 4(7), p.570.

Kidd, C., Becker, A., Huffman, G.J., Muller, C.L., Joe, P., Skofronick-Jackson, G. and Kirschbaum, D.B., 2017. So, how much of the Earth's surface is covered by rain gauges?. Bulletin of the American Meteorological Society, 98(1), pp.69-78.

Kossin, J.P., 2018. A global slowdown of tropical-cyclone translation speed. Nature, 558(7708), p.104.

Kottegoda, N. T., Natale, L., & Raiteri, E. (2014). Monte Carlo Simulation of rainfall hyetographs for analysis and design, 519, 1–11. https://doi.org/10.1016/j.jhydrol.2014.06.041

Krichak, S.O., Barkan, J., Breitgand, J.S., Gualdi, S. and Feldstein, S.B., 2015. The role of the export of tropical moisture into midlatitudes for extreme precipitation events in the Mediterranean region. Theoretical and Applied Climatology, 121(3-4), pp.499-515.

Kuczera, G., Lambert, M.F., Heneker, T.M., Jennings, S., Frost, A., & Coombes, P. (2006). Joint probability and design storms at the Crossroads. Australian Journal of Water Resources 10(1):63–79.

Kunkel, K. E., T. R. Karl, D. R. Easterling, K. Redmond, J. Young, X. Yin, and P. Hennon, 2013: Probable maximum precipitation and climate change. Geophys. Res. Lett., 40, 1402–1408, doi:10.1002/grl.50334.

Kunkel, K.E., Easterling, D.R., Kristovich, D.A., Gleason, B., Stoecker, L. and Smith, R., 2012. Meteorological causes of the secular variations in observed extreme precipitation events for the conterminous United States. Journal of Hydrometeorology, 13(3), pp.1131-1141.

Kunz, M. and Kottmeier, C., 2006. Orographic enhancement of precipitation over low mountain ranges. Part II: Simulations of heavy precipitation events over southwest Germany. Journal of applied meteorology and climatology, 45(8), pp.1041-1055.

Lamjiri, M.A., Dettinger, M.D., Ralph, F.M. and Guan, B., 2017. Hourly storm characteristics along the US West Coast: Role of atmospheric rivers in extreme precipitation. Geophysical Research Letters, 44(13), pp.7020-7028.

Lebo, Z.J. and Morrison, H., 2015. Effects of horizontal and vertical grid spacing on mixing in simulated squall lines and implications for convective strength and structure. Monthly Weather Review, 143(11), pp.4355-4375.

Lee, J., J. Choi, O. Lee, J. Yoon, and S. Kim, 2017: Estimation of probable maximum precipitation in Korea using a regional climate model. Water, 9, 240, doi:10.3390/w9040240.

Letcher, T.W. and Minder, J.R., 2015. Characterization of the simulated regional snow albedo feedback using a regional climate model over complex terrain. Journal of Climate, 28(19), pp.7576-7595.

Leutbecher, M. (2019). Ensemble size: How suboptimal is less than infinity?. Quarterly Journal of the Royal Meteorological Society, 145, pp.107-128.

Lim, K.S.S. and Hong, S.Y., 2010. Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. Monthly weather review, 138(5), pp.1587-1612.

Lin, Y. and Colle, B.A., 2011. A new bulk microphysical scheme that includes riming intensity and temperature-dependent ice characteristics. Monthly Weather Review, 139(3), pp.1013-1035.

Lins, H. F., & Slack, J. R. (1999). Streamflow trends in the United States. Geophysical Research Letters, 26(2), 227–230. https://doi.org/10.1029/1998GL900291

Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A.J., Prein, A.F., Chen, F., Chen, L., Clark, M., Dai, A. and Dudhia, J., 2017. Continental-scale convection-permitting modeling of the current and future climate of North America. Climate Dynamics, 49(1-2), pp.71-95.

Liu, C., Ikeda, K., Thompson, G., Rasmussen, R. and Dudhia, J., 2011. High-resolution simulations of wintertime precipitation in the Colorado Headwaters region: Sensitivity to physics parameterizations. Monthly Weather Review, 139(11), pp.3533-3553.

Loken, E. D., A. J. Clark, M. Xue, and F. Kong, 2017: Comparison of next-day probabilistic severe weather forecasts from coarse- and fine-resolution CAMs and a convection-allowing ensemble. Wea. Forecasting, 32, 1403–1421, https://doi.org/10.1175/WAF-D-16-0200.1.

Lynn, B.H., Khain, A.P., Dudhia, J., Rosenfeld, D., Pokrovsky, A. and Seifert, A., 2005. Spectral (bin) microphysics coupled with a mesoscale model (MM5). Part II: Simulation of a CaPE rain event with a squall line. Monthly Weather Review, 133(1), pp.59-71.

Lynn, B.H., Khain, A.P., Dudhia, J., Rosenfeld, D., Pokrovsky, A. and Seifert, A., 2005. Spectral (bin) microphysics coupled with a mesoscale model (MM5). Part I: Model description and first results. Monthly Weather Review, 133(1), pp.44-58.

Madonna, E., Limbach, S., Aebi, C., Joos, H., Wernli, H. and Martius, O., 2014. On the co-occurrence of warm conveyor belt outflows and PV streamers. Journal of the Atmospheric Sciences, 71(10), pp.3668-3673.

Mahoney, K., Alexander, M., Scott, J.D. & Barsugli, J. (2013). High-resolution downscaled simulations of warm-season extreme precipitation events in the Colorado Front Range under past and future climates. Journal of Climate, 26 (21), pp.8671-8689.

Manabe, S., Smagorinsky, J. and Strickler, R.F., 1965. Simulated climatology of a general circulation model with a hydrologic cycle. Mon. Wea. Rev, 93(12), pp.769-798.

Mansell, E.R., 2010. On sedimentation and advection in multimoment bulk microphysics. Journal of the Atmospheric Sciences, 67(9), pp.3084-3094.

Mansell, E.R., Ziegler, C.L. and Bruning, E.C., 2010. Simulated electrification of a small thunderstorm with two-moment bulk microphysics. Journal of the Atmospheric Sciences, 67(1), pp.171-194.

Mass, C.F., Ovens, D., Westrick, K. and Colle, B.A., 2002. Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest. Bulletin of the American Meteorological Society, 83(3), pp.407-430.

McCumber, M., Tao, W.K., Simpson, J., Penc, R. and Soong, S.T., 1991. Comparison of ice-phase microphysical parameterization schemes using numerical simulations of tropical convection. Journal of Applied Meteorology, 30(7), pp.985-1004.

Mearns, L. and Coauthors. (2011). The North American Regional Climate Change Assessment Program dataset. National Center for Atmospheric Research Earth System Grid data portal. [Available online at http://dx.doi.org/10.5065/D6RN35ST.]

Mellor, G.L. and Yamada, T., 1982. Development of a turbulence closure model for geophysical fluid problems. Reviews of Geophysics, 20(4), pp.851-875.

Min, S.K., Zhang, X., Zwiers, F.W. and Hegerl, G.C., 2011. Human contribution to more-intense precipitation extremes. Nature, 470(7334), p.378.

Mittermaier, M., and G. Csima, 2017: Ensemble versus deterministic performance at the kilometer scale. Wea. Forecasting, 32, 1697–1709, https://doi.org/10.1175/WAF-D-16-0164.1.

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J. and Clough, S.A., 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. Journal of Geophysical Research: Atmospheres, 102(D14), pp.16663-16682.

Moeng, C.H., 2014. A closure for updraft–downdraft representation of subgrid-scale fluxes in cloud-resolving models. Monthly Weather Review, 142(2), pp.703-715.

Mooney, P.A., Broderick, C., Bruyère, C.L., Mulligan, F.J. and Prein, A.F., 2017. Clustering of observed diurnal cycles of precipitation over the United States for evaluation of a WRF multiphysics regional climate ensemble. J. Clim., 30(22), 9267-9286.

Morrison, H. and Milbrandt, J.A., 2015. Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: Scheme description and idealized tests. Journal of the Atmospheric Sciences, 72(1), pp.287-311.

Morrison, H., Milbrandt, J.A., Bryan, G.H., Ikeda, K., Tessendorf, S.A. and Thompson, G., 2015. Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part II: Case study comparisons with observations and other schemes. J. Atmos. Sci., 72(1), 312-339.

Morrison, H., Thompson, G. and Tatarskii, V., 2009. Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes. Mon. Wea. Rev. 137(3), 991-1007.

Munich Re, 2017. TOPICS Geo - Natural catastrophes 2017. https://www.munichre.com/site/touch-publications/get/documents_E711248208/mr/assetpool.shared/Documents/5_Touch/_Publications/TOPICS_GEO_2017-en.pdf, accessed 09/19/2019

Muñoz-Esparza, D., Kosović, B., Mirocha, J. and van Beeck, J., 2014. Bridging the transition from mesoscale to microscale turbulence in numerical weather prediction models. Boundary-layer Meteorology, 153(3), pp.409-440.

Mure-Ravaud, M., A. Dib, M. Kavvas, and E. Yegorova, 2018: Maximization of the precipitation from tropical cyclones over a target area through physically based storm transposition. Hydrology and Earth System Sciences Disc., 1-39. 10.5194/hess-2017-665.

Musselman, K.N., Lehner, F., Ikeda, K., Clark, M.P., Prein, A.F., Liu, C., Barlage, M. and Rasmussen, R., 2018. Projected increases and shifts in rain-on-snow flood risk over western North America. Nature Climate Change, 8(9), p.808.

Nakanishi, M. and Niino, H., 2004. An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification. Boundary-layer meteorology, 112(1), pp.1-31.

Nakanishi, M. and Niino, H., 2006. An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. Boundary-Layer Meteorology, 119(2), pp.397-407.

Neumann, P., Düben, P., Adamidis, P., Bauer, P., Brück, M., Kornblueh, L., Klocke, D., Stevens, B., Wedi, N. and Biercamp, J., 2019. Assessing the scales in numerical weather and climate predictions: will exascale be the rescue?. Philosophical Transactions of the Royal Society A, 377(2142), p.20180148.

Newman, A. J., and Coauthors, 2015: Gridded ensemble precipitation and temperature estimates for the contiguous United States. Journal of Hydrometeorology, 16 (6), 2481–2500.

Newman, A.J., Clark, M.P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., Mizukami, N., Brekke, L., and Arnold, J.R., 2015. Gridded ensemble precipitation and temperature estimates for the contiguous United States. J. Hydrometeorology, 16(6), 2481-2500.

Nielsen, E. R., and R. S. Schumacher, 2018: Dynamical Insights into Extreme Short-Term Precipitation Associated with Supercells and Mesovortices. J. Atm. Sci., 75, 2983-3009.

O'Gorman, P.A., 2015. Precipitation extremes under climate change. Current climate change reports, 1(2), pp. 49-59.

Ogaja, J. and Will, A., 2016. Fourth order, conservative discretization of horizontal Euler equations in the COSMO model and regional climate simulations. Met. Z., DOI, 10.

Ohara, N., M. L. Kavvas, S. Kure, Z.Q. Chen, and E. Tan, 2011: Physically based estimation of maximum precipitation over American River watershed, California. J. Hydrologic Engr., 16, 351–361, doi:10.1061/(ASCE)HE.1943-5584.0000324.

Ohara, N., M.L. Kavvas, M.L. Anderson, Z.Q. Chen, and K. Ishida, 2017: Characterization of extreme storm events using a numerical model–based precipitation maximization procedure in the Feather, Yuba, and American River Watersheds in California. J. Hydrometeor., 18, 1413–1423, https://doi-org.cuucar.idm.oclc.org/10.1175/JHM-D-15-0232.1

Pasteris, P. P., 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology: a Journal of the Royal Meteorological Society, 28 (15), 2031–2064.

Patricola, C.M. and Wehner, M.F., 2018. Anthropogenic influences on major tropical cyclone events. Nature, 563(7731), p.339.

Pavelsky, T.M., Sobolowski, S., Kapnick, S.B. and Barnes, J.B., 2012. Changes in orographic precipitation patterns caused by a shift from snow to rain. Geophysical Research Letters, 39(18).

Pendergrass, A.G., 2018. What precipitation is extreme?. Science, 360(6393), pp.1072-1073.

Peters, J. M., and R. S. Schumacher, 2014: Objective categorization of heavy-rain producing MCS synoptic types by rotated principle component analysis. Mon. Wea. Rev., 142, 1716-1737.

Peters, J. M., and R. S. Schumacher, 2015: Mechanisms for organization and echo training in a flash-flood-producing mesoscale convective system. Mon. Wea. Rev., 143, 1058-1085.

Pfahl, S., Madonna, E., Boettcher, M., Joos, H. and Wernli, H., 2014. Warm conveyor belts in the ERA-Interim dataset (1979–2010). Part II: Moisture origin and relevance for precipitation. Journal of Climate, 27(1), pp.27-40.

Pielke, R. A., W. R. Cotton, R. L. Walko, C. J. Tremback, W. A. Lyons, L. D. Grasso, M. E. Nicholls, M. D. Moran, D. A. Wesley, T. J. Lee, and J. H. Copeland, 1992: A comprehensive meteorological modeling system: RAMS. Meteor. Atmos. Phys., 49, 69–91.

Pokharel, B., Wang, S.Y.S., Lin, Y.H., Zhao, L. and Gillies, R., 2018. Diagnosing the Atypical Extreme Precipitation Events Under Weakly Forced Synoptic Setting: The West Virginia Flood (June 2016) and Beyond. Climate Prediction S&T Digest, p.8.

Potvin, C. K., and M. L. Flora, 2015: Sensitivity of idealized supercell simulations to horizontal grid spacing: Implications for Warn-on-Forecast. Mon. Wea. Rev., 143, 2998–3024, https://doi.org/10.1175/MWR-D-14-00416.1.

Prein A.F., R. M. Rasmussen, D. Wang, S. E. Giangrande (2019) Sensitivity of Mesoscale Convective Systems to Model Grid Spacing in Current and Future Climates, Journal of Geophysical Research - Atmosphere, submitted

Prein, A. F., C. Liu, K. Ikeda, R. Bullock, R. M. Rasmussen, G. J. Holland, and M. Clark, 2017: Simulating North American mesoscale convective systems with a convection-permitting climate model. Climate Dynamics, 55, 95–110.

Prein, A. F., R. M. Rasmussen, D. Wang, and S. Giangrande, 2020: Sensitivity of Organized Convective Storms to Model Grid Spacing in Current and Future Climates. Philosophical Transactions of the Royal Society A. (in press).

Prein, A.F. and Gobiet, A., 2017. Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. International Journal of Climatology, 37(1), pp.305-327.

Prein, A.F., Gobiet, A., and Truhetz, H., 2011. Analysis of uncertainty in large scale climate change projections over Europe. Meteorologische Zeitschrift, 20(4), 383-395.

Prein, A.F., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N.K., Keuler, K. and Georgievski, G., 2013b. Added value of convection permitting seasonal simulations. Climate Dynamics, 41(9-10), pp.2655-2677.

Prein, A.F., Holland, G.J., Rasmussen, R.M., Done, J., Ikeda, K., Clark, M.P. and Liu, C.H., 2013. Importance of regional climate model grid spacing for the simulation of heavy precipitation in the Colorado headwaters. Journal of Climate, 26(13), pp.4848-4857.

Prein, A.F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F. and Brisson, E., 2015. A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. Reviews of Geophysics, 53(2), pp.323-361.

Prein, A.F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., van Lipzig, N.P.M., & Leung, R. (2015). A review on regional convection-permitting climate modeling: de-monstrations, prospects, and challenges. Rev. Geophys. 53, 323–361. https://doi.org/10.1002/2014RG000475.

Prein, A.F., Liu, C., Ikeda, K., Bullock, R., Rasmussen, R.M., Holland, G.J. and Clark, M., 2017b. Simulating North American mesoscale convective systems with a convection-permitting climate model. Climate Dynamics, pp.1-16.

Prein, A.F., Liu, C., Ikeda, K., Trier, S.B., Rasmussen, R.M., Holland, G.J. and Clark, M.P., 2017b. Increased rainfall volume from future convective storms in the US. Nature Climate Change, 7(12), p.880.

Prein, A.F., Rasmussen, R., and Stephens, G., 2017a. Challenges and advances in convection-permitting climate modeling. Bull. Amer. Meteor. Soc., 98(5), 1027-1030.

Prein, A.F., Rasmussen, R.M., Ikeda, K., Liu, C., Clark, M.P. and Holland, G.J., 2017a. The future intensification of hourly precipitation extremes. Nature Climate Change, 7(1), p.48.

Rahman, A., Weinmann, P. E., Hoang, T. M. T., & Laurenson, E. M. (2002). Monte Carlo simulation of flood frequency curves from rainfall, 256, 196–210.

Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., Tewari, M., Barlage, M., Dudhia, J., Yu, W. and Miller, K., 2011. High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: a process study of current and warmer climate. Journal of Climate, 24(12), pp.3015-3048.

Rastogi, D., S.-C. Kao, M. Ashfaq, R. Mei, E. D. Kabela, S. Gangrade, B. S. Naz, B. L. Preston, N. Singh,and V. G. Anantharaj, 2017: Effects of climate change on probable maximum precipitation: A sensitivity study over the Alabama-Coosa-Tallapoosa River Basin, J. Geophys. Res. Atmos.,122, 4808–4828, doi:10.1002/2016JD026001.

Roberts, N. M., and H. W. Lean, 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon. Wea. Rev., 136, 78–97.

Romine, G. S., C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. Mon. Wea. Rev., 142, 4519–4541, doi:https://doi.org/10.1175/MWR-D-14-00100.1.

Rotunno, R., Klemp, J.B., and Weisman, M.L., 1988. A theory for strong, long-lived squall lines. J. Atmos. Sci., 45(3), 463-485.

Rudolf, B., Hauschild, H., Rueth, W. and Schneider, U., 1994. Terrestrial precipitation analysis: Operational method and required density of point measurements. In Global precipitations and climate change (pp. 173-186). Springer, Berlin, Heidelberg.

Ryu, Y., J.A. Smith, M.L. Baeck, L.K. Cunha, E. Bou-Zeid, and W. Krajewski, 2016: The regional water cycle and heavy spring rainfall in iowa: Observational and modeling analyses from the IFloodS campaign. J. Hydrometeor., 17, 2763–2784, https://doi-org.cuucar.idm.oclc.org/10.1175/JHM-D-15-0174.1

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M. and Rudolf, B., 2014. GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. Theoretical and Applied Climatology, 115(1-2), pp.15-40.

Schreiner, L.C. and Riedel, J.T., 1978. Probable maximum precipitation estimates, United States east of the 105th meridian.

Schumacher, R. S., 2015: Resolution dependence of initiation and upscale growth of deep convection in convection-allowing forecasts of the 31 May–1 June 2013 supercell and MCS. Mon. Wea. Rev., 143, 4331–4354, https://doi.org/10.1175/MWR-D-15-0179.1.

Schumacher, R. S., and R. H. Johnson, 2005: Organization and environmental properties of extreme-rain-producing mesoscale convective systems. Mon. Wea. Rev., 133, 961-976.

Schumacher, R. S., and R. H. Johnson, 2006: Characteristics of US extreme rain events during 1999–2003. Weather and Forecasting, 21 (1), 69–85.

Schwartz, C. S., 2014: Reproducing the September 2013 record-breaking rainfall over the Colorado Front Range with high-resolution WRF forecasts. Wea. Forecasting, 29, 393–402, https://doi.org/10.1175/WAF-D-13-00136.1.

Schwartz, C. S., and Coauthors, 2009: Next-day convection-allowing WRF Model guidance: A second look at 2-km versus 4-km grid spacing. Mon. Wea. Rev., 137, 3351– 3372, https://doi.org/10.1175/2009MWR2924.1.

Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. Wea. Forecasting, 25, 263–280, https://doi.org/10.1175/2009WAF2222267.1.

Schwartz, C. S., and R. A. Sobash, 2019: Revisiting sensitivity to horizontal grid spacing in convection-allowing models over the central–eastern United States. Mon. Wea. Rev., In press, https://doi.org/10.1175/MWR-D-19-0115.1.

Schwartz, C. S., G. S. Romine, K. R. Fossell, R. A. Sobash, and M. L. Weisman, 2017: Toward 1-km ensemble forecasts over large domains. Mon. Wea. Rev., 145, 2943–2969, https://doi.org/10.1175/MWR-D-16-0410.1.

Schwartz, C. S., G. S. Romine, K. R. Smith, and M. L. Weisman, 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. Wea. Forecasting, 29, 1295–1318, https://doi.org/10.1175/WAF-D-13-00145.1.

Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015a: NCAR's experimental real-time convection-allowing ensemble prediction system. Weather and Forecasting, 30 (6), 1645–1654.

Schwartz, C. S., Romine, G. S., Weisman, M. L., Sobash, R. A., Fossell, K. R., Manning, K. W. and Trier, S. B., 2015. A real-time convection-allowing ensemble prediction system initialized by mesoscale ensemble Kalman filter analyses. Weather and Forecasting, 30(5), 1158–1181.

Schwartz, C.S. and Sobash, R.A., 2019. Revisiting sensitivity to horizontal grid spacing in convection-allowing models over the central and eastern United States. Monthly Weather Review, 147(12), pp.4411-4435.

Schwartz, C.S., G.S. Romine, M.L. Weisman, R.A. Sobash, K.R. Fossell, K.W. Manning, and S.B. Trier, 2015b: A real-time convection-allowing ensemble prediction system initialized by mesoscale ensemble Kalman filter analyses. Wea. Forecasting, 30, 1158–1181, https://doi.org/10.1175/WAF-D-15-0013.1.

Schwartz, C.S., G.S. Romine, R.A. Sobash, K.R. Fossell, and M.L. Weisman, 2015a: NCAR's experimental real-time convection-allowing ensemble prediction system. Wea. Forecasting, 30, 1645–1654, https://doi.org/10.1175/WAF-D-15-0103.1.

Schwartz, C.S., Kain, J.S., Weiss, S.J., Xue, M., Bright, D.R., Kong, F., Thomas, K.W., Levit, J.J., and Coniglio, M.C., 2009. Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. Mon. Wea. Rev., 137(10), 3351-3372.

Schwarz, F. K., 1970: The unprecedented rains in Virginia associated with the remains of Hurricane Camille. Mon. Wea. Rev., 98, 851–859, https://doi-org.cuucar.idm.oclc.org/10.1175/1520-0493(1970)098<0851:TURIVA>2.3.CO;2

Sharma, A., Wasko, C., & Lettenmaier, D. P. (2018). If Precipitation Extremes Are Increasing, Why Aren't Floods? Water Resources Research, 54, 8545–8551. https://doi.org/10.1029/2018WR023749

Shepherd, T.G., Boyd, E., Calel, R.A., Chapman, S.C., Dessai, S., Dima-West, I.M., Fowler, H.J., James, R., Maraun, D., Martius, O. and Senior, C.A., 2018. Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. Climatic Change, 151(3-4), pp.555-571.

Shin, H.H. and Hong, S.Y., 2015. Representation of the subgrid-scale turbulent transport in convective boundary layers at gray-zone resolutions. Monthly Weather Review, 143(1), pp.250-271.

Siler, N. and Roe, G., 2014. How will orographic precipitation respond to surface warming? An idealized thermodynamic perspective. Geophysical Research Letters, 41(7), pp.2606-2613.

Singh A., V. P. Singh, and A. R. Byrd, 2018: Computation of probable maximum precipitation and its uncertainty. Int'l. J. Hydrol., 2(4), 504-514, doi:10.15406/ijh.2018.02.00118.

Singh, M.S. and O'Gorman, P.A., 2014. Influence of microphysics on the scaling of precipitation extremes with temperature. Geophysical Research Letters, 41(16), pp.6037-6044.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, Z. Liu, J. Berner, W. Wang, J. G. Powers, D. M. Barker, and X.-Y. Huang, 2019: A description of the Advanced Research WRF Model Version 4. NCAR Tech. Note, NCAR/TN-5565+STR, 162 pp. doi: http://dx.doi.org/10.5065/1dfh-6p97.

Skamarock, W.C., 2004. Evaluating mesoscale NWP models using kinetic energy spectra. Monthly Weather Review, 132(12), pp.3019-3032.

Skamarock, W.C., Park, S.H., Klemp, J.B. and Snyder, C., 2014. Atmospheric kinetic energy spectra from global high-resolution nonhydrostatic simulations. Journal of the Atmospheric Sciences, 71(11), pp.4369-4381.

Smith, R.B., 1989. Hydrostatic airflow over mountains. In Advances in geophysics (Vol. 31, pp. 1-41). Elsevier.

Smith, R.B., 2006. Progress on the theory of orographic precipitation. *Special Papers-Geolocial Societey of America*, *398*, p.1.

Smith, T.L., Benjamin, S.G., Brown, J.M., Weygandt, S., Smirnova, T. and Schwartz, B., 2008. 11.1 Convection forecasts from the hourly updated, 3-km High Resolution Rapid Refresh (HRRR) Model.

Sobash, R.A., C.S. Schwartz, G.S. Romine, and M.L. Weisman, 2019: Next-day prediction of tornadoes using convection-allowing models with 1-km horizontal grid spacing. Wea. Forecasting (online), https://doi.org/10.1175/WAF-D-19-0044.1.

Stensrud, D.J., 2009. Parameterization schemes: keys to understanding numerical weather prediction models. Cambridge University Press.

Stocker, E.F. and Wolff, D.B., 2007. The TRMM multi-satellite precipitation analysis: quasi-global, multi-year, combined-sensor precipitation estimates at finescale. J. Hydrometeorol., 8, 3355.

Stull, R.B., 1991. Static stability—An update. Bulletin of the American Meteorological Society, 72(10), pp.1521-1530.

Stull, R.B., 1993. Review of non-local mixing in turbulent atmospheres: Transilient turbulence theory. Boundary-Layer Meteorology, 62(1-4), pp.21-96.

Sui, C.H., Li, X., Yang, M.J. and Huang, H.L., 2005. Estimation of oceanic precipitation efficiency in cloud models. J. Atmos. Sci., 62(12), 4358-4370.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S. and Hsu, K.L., 2018. A review of global precipitation data sets: Data sources, estimation, and intercomparisons. Reviews of Geophysics, 56(1), pp.79-107.

Svensson, C., Kjeldsen, T. R., Jones, D. A., Svensson, C., Kjeldsen, T. R., Flood, D. A. J. Jones, D. A. (2013). Flood frequency estimation using a joint probability approach within a Monte Carlo framework. Hydrological Sciences Journal, 58(1). https://doi.org/10.1080/02626667.2012.746780

Tao, W.K., Li, X., Khain, A., Matsui, T., Lang, S. and Simpson, J., 2007. Role of atmospheric aerosol concentration on deep convective precipitation: Cloud-resolving model simulations. Journal of Geophysical Research: Atmospheres, 112(D24).

Taraphdar, S., Mukhopadhyay, P., Leung, L.R., Zhang, F., Abhilash, S. and Goswami, B.N., 2014. The role of moist processes in the intrinsic predictability of Indian Ocean cyclones. Journal of Geophysical Research: Atmospheres, 119(13), pp.8032-8048.

Thielen, J.E. and W.A. Gallus, 2019: Influences of horizontal grid spacing and microphysics on WRF forecasts of convective morphology evolution for nocturnal MCSs in weakly forced environments. Wea. Forecasting, 34, 1495–1517, https://doi.org/10.1175/WAF-D-18-0210.1.

Thompson, G. and Eidhammer, T., 2014. A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. Journal of the atmospheric sciences, 71(10), pp.3636-3658.

Thompson, G., Field, P.R., Rasmussen, R.M. and Hall, W.D., 2008. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. Monthly Weather Review, 136(12), pp.5095-5115.

Tiedtke, M., 1989. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. Monthly Weather Review, 117(8), pp.1779-1800.

Toride, K., Y. Iseri, A. M. Duren, J. F. England, M. L. Kavvas, 2019: Evaluation of physical parameterizations for atmospheric river induced precipitation and application to long-term reconstruction based on three reanalysis datasets in Western Oregon. Sci. Total Env., 658, 570–581, https://doi.org/10.1016/j.scitotenv.2018.12.214.

Toride, K., Y. Iseri, M. D. Warner, C. D. Frans, A. M. Duren, J. F. England, and M. L. Kavvas, 2019: Model-based Probable Maximum Precipitation estimation: How to estimate the worst-case scenario induced by atmospheric rivers? J. Hydrometeor. (accepted), https://doi.org/10.1175/JHM-D-19-0039.1.

Trenberth, K. E., A. Dai, R. M. Rasmussen, & Parsons, D.B. (2003). The changing character of precipitation, Bull. Am. Meteorol. Soc., 84,1205–1217, doi:10.1175/BAMS-84-9-1205.

Trier, S. B., 2015: A real-time convection-allowing ensemble prediction system initialized by mesoscale ensemble Kalman filter analyses. Weather and Forecasting, 30(5), 1158–1181.

Uppala, S. M., and Coauthors. (2005). The ERA-40 Re-Analysis. Quart. J. Roy. Meteor. Soc., 131, 2961–3012.

Van den Heever, S.C., Carrió, G.G., Cotton, W.R., DeMott, P.J. and Prenni, A.J., 2006. Impacts of nucleating aerosol on Florida storms. Part I: Mesoscale simulations. Journal of the Atmospheric Sciences, 63(7), pp.1752-1775.

Van Weverberg, K., Goudenhoofdt, E., Blahak, U., Brisson, E., Demuzere, M., Marbaix, P. and van Ypersele, J.P., 2014. Comparison of one-moment and two-moment bulk microphysics for high-resolution climate simulations of intense precipitation. Atmospheric Research, 147, pp.145-161.

Verrelle, A., D. Ricard, and C. Lac, 2015: Sensitivity of high-resolution idealized simulations of thunderstorms to horizontal resolution and turbulence parameterization. Quart. J. Roy. Meteor. Soc., 141, 433–448, https://doi.org/10.1002/qj.2363.

Voudouri, A., Khain, P., Carmona, I., Avgoustoglou, E., Kaufmann, P., Grazzini, F. and Bettems, J.M., 2018. Optimization of high resolution COSMO model performance over Switzerland and Northern Italy. Atmospheric Research, 213, pp.70-85.

Wagner, P.D., Fiener, P., Wilken, F., Kumar, S. and Schneider, K., 2012. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. Journal of Hydrology, 464, pp.388-400.

Wallace, J.M. and Hobbs, P.V., 2006. Atmospheric Science: An Introductory Survey (Vol. 92). Elsevier.

Walsh, J. and Coauthors. (2014). Ch. 2: Our Changing Climate. Climate Change Impacts in the United States: The Third National Climate Assessment. In: Melillo J M, Terese T C, Richmond, Yohe GW (eds) U.S. Global Change Research Program, 19–67, doi:10.7930/J0KW5CXT

Wang, D., Giangrande, S.E., Feng, Z., Hardin, J.C., Prein, A.F., 2019. Updraft and Downdraft Core Size and Intensity as Revealed by Radar Wind Profilers: MCS Observations and Idealized Model Comparisons. Submitted to JGR: Atmospheres

Wang, Y., 2002. An explicit simulation of tropical cyclones with a triply nested movable mesh primitive equation model: TCM3. Part II: Model refinements and sensitivity to cloud microphysics parameterization. Monthly Weather Review, 130(12), pp.3022-3036.

Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. Mon. Wea. Rev., 125, 527–548, https://doi.org/10.1175/1520-0493(1997)125<0527:TRDOEM>2.0.CO;2.

Weisman, M.L. and Klemp, J.B., 1982. The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. Monthly Weather Review, 110(6), pp.504-520.

Weisman, M.L., Skamarock, W.C. & Klemp, J.B. (1997). The resolution dependence of explicitly modeled convective systems. Monthly Weather Review, 125(4), pp.527-548.

Weller, H. and Weller, H.G., 2008. A high-order arbitrarily unstructured finite-volume model of the global atmosphere: Tests solving the shallow-water equations. International journal for numerical methods in fluids, 56(8), pp.1589-1596.

Westra, S., Alexander, L.V. and Zwiers, F.W., 2013. Global increasing trends in annual maximum daily precipitation. Journal of Climate, 26(11), pp.3904-3918.

WMO, World Meteorological Organization, 2009. Manual on estimation of probable maximum precipitation (PMP). World Meteorological Organization

Woldemichael, A.T., F. Hossain, and R. Pielke, 2014: Impacts of postdam land use/land cover changes on modification of extreme precipitation in contrasting hydroclimate and terrain features. J. Hydrometeor., 15, 777–800, https://doi-org.cuucar.idm.oclc.org/10.1175/JHM-D-13-085.1

Wood, A. W., Leung, L. R., Sridhar, V., & Lettenmaier, D. P. (2004). Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs. Climatic Change, 62(1–3), 189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9e

Wright, D. B., Smith, J. A., & Baeck, M. L. (2014). Critical Examination of Area Reduction Factors. J. Hydrol. Eng., 19(APRIL), 769–776. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000855.

Wright, D. B., Yu, G., & England, J. F. (2020). Six decades of rainfall and flood frequency analysis using stochastic storm transposition: Review, progress , and prospects. Journal of Hydrology, 585(February), 124816. https://doi.org/10.1016/j.jhydrol.2020.124816

Wright, D.B., Bosma, C.D. and Lopez-Cantu, T., 2019. US hydrologic design standards insufficient due to large increases in frequency of rainfall extremes. Geophysical Research Letters, 46(14), pp.8144-8153.

Wyngaard, J.C., 2004. Toward numerical modeling in the "Terra Incognita". Journal of the Atmospheric Sciences, 61(14), pp.1816-1826.

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, and K. K. Droegemeier, 2013: Prediction of convective storms at convection-resolving 1-km resolution over continental United States with radar data assimilation: An example case of 26 May 2008 and precipitation forecasts from spring 2009. Adv. Meteor., 2013, 259052, https://doi.org/10.1155/2013/259052.

Yu, G., Wright, D. B., & Li, Z. (2020). The Upper Tail of Precipitation in Convection-Permitting Regional Climate Models and Their Utility in Nonstationary Rainfall and Flood Frequency Analysis. Earths Future, https://doi.org/10.1029/2020EF001613

Zängl, G., 2012. Extending the numerical stability limit of terrain-following coordinate models over steep slopes. Monthly Weather Review, 140(11), pp.3722-3733.

Zängl, G., Reinert, D., Rípodas, P. and Baldauf, M., 2015. The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. Quarterly Journal of the Royal Meteorological Society, 141(687), pp.563-579.

Zhang, J., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans. Bulletin of the American Meteorological Society, 92(10), 1321–1338.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. Bulletin of the American Meteorological Society, 97(4), 621–638.

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., Grams, H., Wang, Y., Cocks, S., Martinaitis, S. and Arthur, A., 2016. Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. Bull. Amer. Meteor. Soc., 97(4), 621-638.

Zhang, X., Anagnostou, E. and Schwartz, C., 2018. NWP-based adjustment of IMERG precipitation for flood-inducing complex terrain storms: Evaluation over CONUS. Remote Sensing, 10(4), p.642.

Zhao, W., J. A. Smith, and A. A. Bradley, 1997: Numerical simulation of a heavy rainfall event during the PRE-STORM experiment. Water Resour. Res., 33, 783–799, doi:https://doi-org.cuucar.idm.oclc.org/10.1029/96WR03036.

Ziegler, C.L., 1985. Retrieval of thermal and microphysical variables in observed convective storms. Part 1: Model development and preliminary testing. Journal of the atmospheric sciences, 42(14), pp.1487-1509.

Zscheischler, J. and Coauthors. (2018). Future climate risk from compound events. Nat. Clim. Change 8, 469–477.

APPENDIX A
HEAVY PRECIPITATION CASES FOR MODEL EVALUATION

In this appendix, the cases that were used for model evaluation are described.
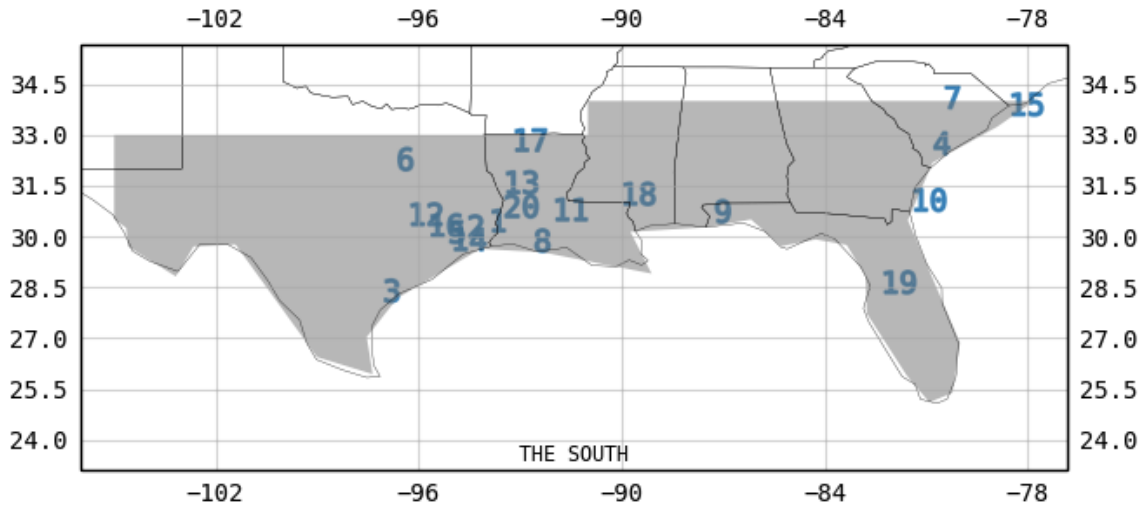
A.1   Gulf Coast

Figure A-1   Locations of the Daily Peak Accumulations of the top 20 Heaviest Precipitation Events in the South Region, with Numbers Indicating Event Ranks

Table A-1   Event Ranks and Precipitation Accumulations in the South Region. Numbers Correspond to Event Locations Mapped in Figure A-1.

| rank	date	lat	lon	3km	1km	peak daily accum [mm/d]	name
1	08/30/2017	30.14	-93.93	10	0	443	Hurr Harvey
2	08/28/2017	29.97	-94.57	10	0	385	Hurr Harvey
3	08/26/2017	28.07	-97.06	10	0	372	Hurr Harvey
4	10/08/2016	32.42	-80.81	10	0	364	Hurr Matthew
5	08/27/2017	29.81	-95.13	10	0	363	Hurr Harvey
6	10/24/2015	31.96	-96.66	10	0	329	TX MCSs
7	10/04/2015	33.78	-80.49	10	0	293	SC front+Joaquin
8	08/14/2016	29.55	-92.61	10	0	292	
9	04/30/2014	30.42	-87.28	1	1	291	MS-FL coast MCSs
10	09/11/2017	30.75	-81.45	10	0	287	Hurr Irma
11	08/13/2016	30.47	-92.05	10	0	285	stalled MCS LA
12	05/27/2016	30.33	-96.33	11	1	285	TX MCS
13	03/10/2016	31.27	-93.5	11	1	278	TX LA serial MCS
14	08/29/2017	29.59	-95.06	10	0	274	Hurr Harvey
15	10/05/2015	33.57	-78.57	10	0	255	
16	04/18/2016	30.00	-95.74	10	0	248	TX MCS
17	03/09/2016	32.52	-93.24	11	1	247	TX LA serial MCS
18	03/11/2016	30.93	-90.04	10	0	245	
19	06/25/2012	28.32	-82.34	1	1	228	TS Debby FL
20	03/21/2012	30.56	-93.5	1	1	219	Stalld cold frnt LA

A.2 Atlantic Coast
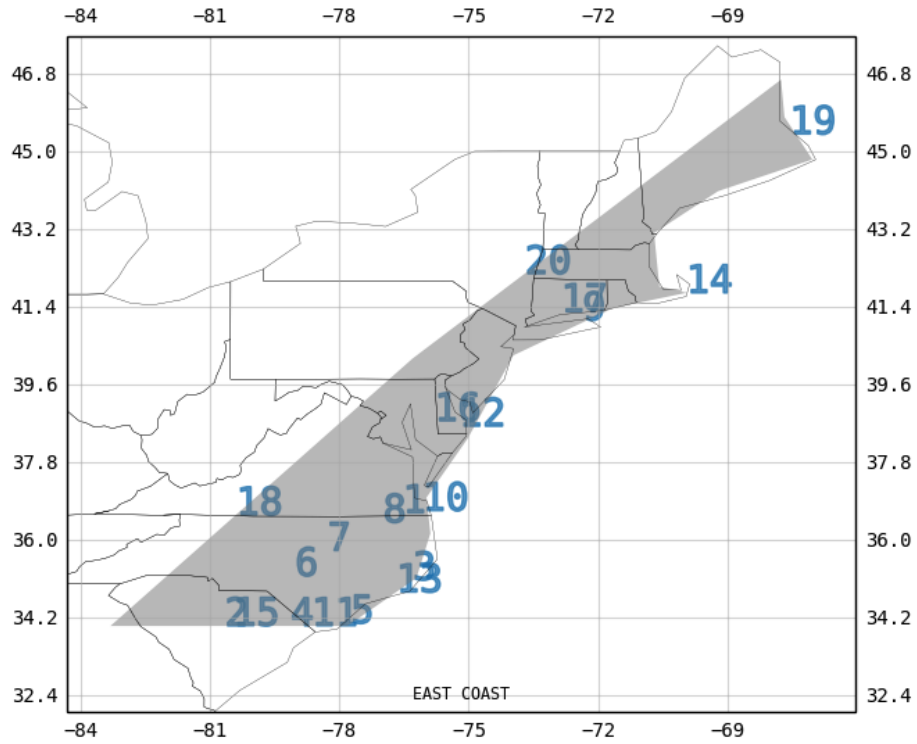
Figure A-2 **Locations of the Daily Peak Accumulations of the top 20 Heaviest Precipitation Events in the Atlantic Coast Region, with Numbers Indicating Event Ranks**

Table A-2   Event Ranks and Precipitation Accumulations in the Atlantic Coast Region. Numbers Correspond to Event Locations Mapped in Figure A-2.

| rank	date	lat	lon	3km	1km	peak daily accum [mm/d]	name
1	10/09/2016	36.62	-76.53	10	0	267	Hurr Matthew
2	10/04/2015	34.00	-80.65	10	0	262	SC front+Joaquin
3	09/03/2016	35.07	-76.3	10	0	211	Hurr Hermine NC/SC
4	10/08/2016	34.00	-79.13	10	0	201	Hurr Matthew
5	10/05/2015	34.07	-77.73	10	0	179	
6	9/29/2016	35.16	-79.03	10	0	150	
7	04/25/2017	35.75	-78.27	10	0	150	
8	09/22/2016	36.40	-76.99	10	0	139	
9	06/08/2013	41.09	-72.37	11	11	137	
10	09/21/2016	36.67	-76.08	10	0	135	
11	10/03/2015	34.01	-78.64	10	0	128	SC front+Joaquin
12	09/30/2016	38.63	-75.23	10	0	116	
13	11/20/2015	34.79	-76.68	10	0	109	
14	08/19/2017	41.72	-69.97	10	0	106	
15	09/12/2017	34.01	-80.46	10	0	104	Hurr Irma
16	07/29/2017	38.75	-75.8	10	0	100	
17	10/30/2017	41.28	-72.86	10	0	96	warm front NJ
18	08/04/2016	36.57	-80.39	10	0	96	
19	10/01/2015	45.40	-67.57	10	0	94	
20	09/30/2015	42.16	-73.7	10	0	93	

A.3 Central U.S.
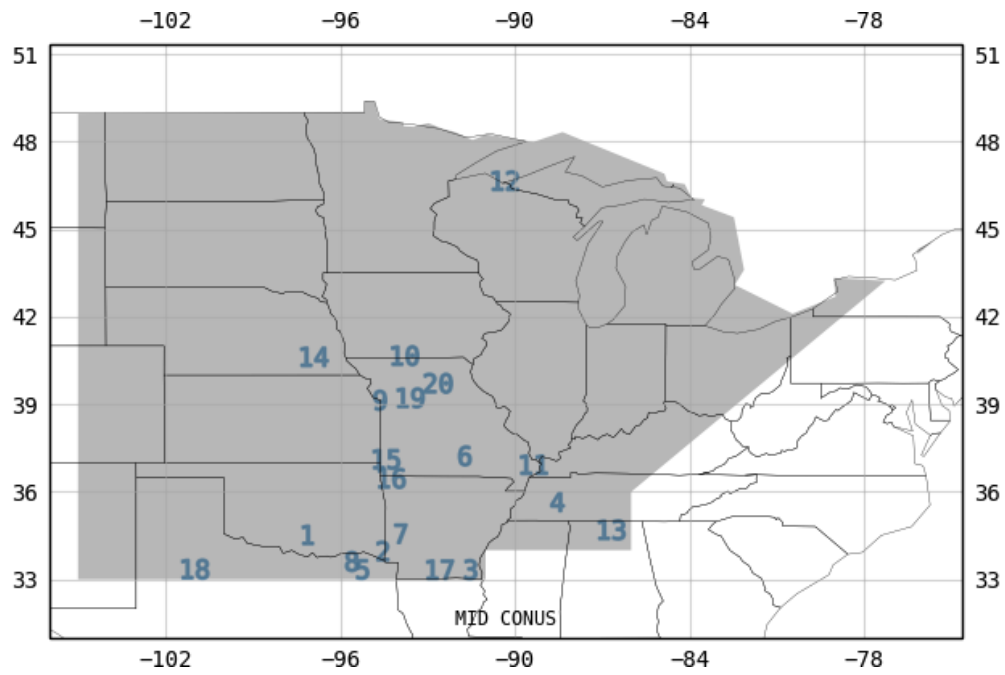
Figure A-3 **Locations of the Daily Peak Accumulations of the top 20 Heaviest Precipitation Events in the Central U.S. Region, with Numbers Indicating Event Ranks**

Table A-3   Event Ranks and Precipitation Accumulations in the Central U.S. Region. Numbers Correspond to Event Locations Mapped in Figure A-3.

| rank	date	lat	lon	3km	1km	peak daily accum [mm/d]	name
1	06/18/2015	34.19	-97.4	11	1	223	TS Bill TX/OK
2	12/28/2015	33.64	-94.8	11	1	183	
3	03/10/2016	33.01	-91.78	11	1	182	TX LA serial MCS
4	09/01/2017	35.33	-88.8	10	0	181	
5	10/24/2015	33.02	-95.49	10	0	181	TX MCSs
6	4/30/2017	36.90	-91.98	10	0	173	slow low OK AR
7	04/30/2016	34.23	-94.18	11	1	172	AR MCSs
8	08/13/2017	33.28	-95.86	10	0	171	MCSs OK/TX border
9	08/22/2017	38.81	-94.89	10	0	166	
10	06/29/2017	40.34	-94.33	10	0	165	Trng suprcls N. MO
11	04/28/2014	36.60	-89.9	1	1	157	
12	07/12/2016	46.33	-90.88	11	1	157	
13	12/26/2015	34.37	-87.23	10	0	157	
14	05/07/2015	40.30	-97.45	11	1	150	
15	05/24/2011	36.79	-94.96	1	1	149	MCS E. OK
16	12/27/2015	36.13	-94.78	11	1	149	
17	03/09/2016	33.00	-93.12	11	1	149	TX LA serial MCS
18	05/05/2015	33.02	-101.54	11	1	145	
19	07/27/2017	38.89	-94.14	10	0	143	
20	06/30/2017	39.39	-93.17	10	0	140	

A.4 Appalachians
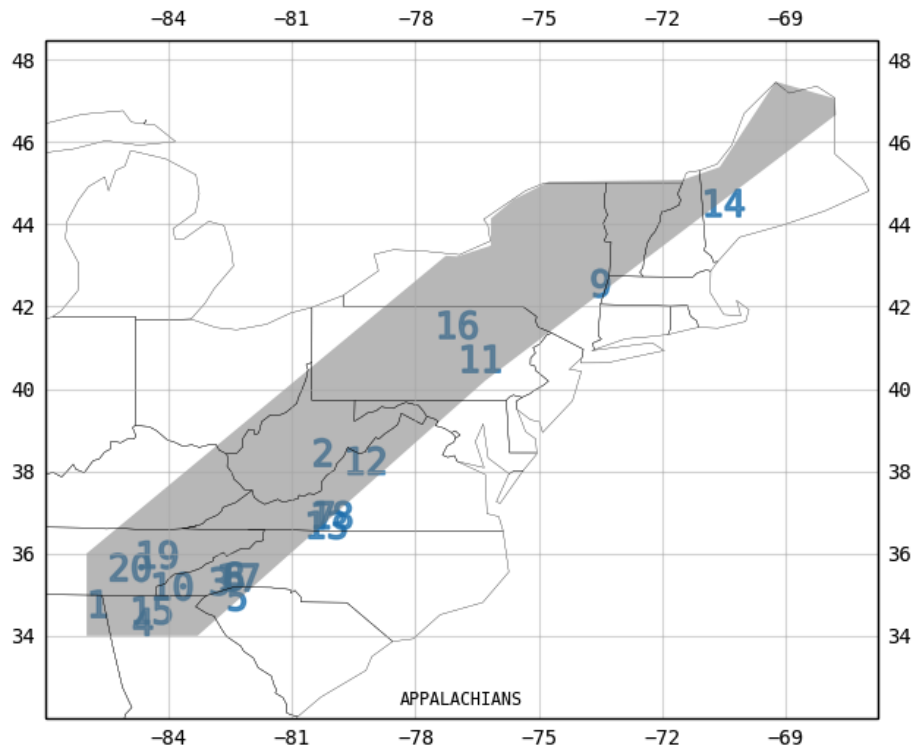
Figure A-4 Locations of the Daily Peak Accumulations of the top 20 Heaviest Precipitation Events in the Appalachian Region, with Numbers Indicating Event Ranks

Table A-4   Event Ranks and Precipitation Accumulations in the Appalachian Region. Numbers Correspond to Event Locations Mapped in Figure A-4.

rank	date	lat	lon	3km	1km	peak daily accum [mm/d]	name
1	12/26/2015	34.45	-86	10	0	132	MCS stratiform
2	06/24/2016	38.12	-80.52	11	1	120	WV orographic
3	10/09/2017	35.00	-83.03	10	0	117	
4	09/12/2017	34.01	-84.9	10	0	104	Hurr Irma
5	10/04/2015	34.62	-82.6	10	0	102	SC front+Joaquin
6	10/24/2017	35.12	-82.69	10	0	100	
7	08/04/2016	36.60	-80.47	10	0	98	orographic NC VA
8	12/01/2010	35.12	-82.69	1	1	96	
9	09/30/2015	42.24	-73.76	10	0	94	
10	12/02/2015	34.87	-84.46	10	0	93	
11	04/17/2011	40.40	-76.96	1	1	92	
12	09/30/2016	37.92	-79.76	10	0	91	
13	04/24/2017	36.37	-80.71	10	0	91	
14	10/30/2017	44.18	-71.04	10	0	90	
15	04/07/2014	34.29	-84.96	1	1	88	
16	10/21/2016	41.22	-77.52	10	0	87	
17	04/16/2011	35.09	-82.84	1	1	86	
18	04/20/2015	36.60	-80.58	11	1	85	
19	07/07/2013	35.64	-84.83	1	1	84	
20	04/28/2011	35.33	-85.48	1	1	84	

2. TITLE AND SUBTITLE

Convection-Permitting Modeling for Intense Precipitation Processes

3. DATE REPORT PUBLISHED

MONTH	YEAR
May	**2023**

4. FIN OR GRANT NUMBER

5. AUTHOR(S)

Andreas F. Prein, Jordan Powers, Erin Towler, David Ahijevych, Ryan Sobash, Craig S. Schwartz

6. TYPE OF REPORT

Technical

7. PERIOD COVERED (Inclusive Dates)

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.)

National Center of Atmospheric Research (NCAR)
3090 Center Green Dr.
Boulder, CO 80301

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.)

DRA
RES
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555-0001

10. SUPPLEMENTARY NOTES

11. ABSTRACT (200 words or less)

The resilient design of critical infrastructure such as roads, dams, and power plants is essential for human safety. Designed standards are traditionally based on observational records, which can be problematic since structures, such as nuclear power plants, should withstand very rare extreme events such as flood return periods of up to one million years. Comparatively short observational records, sampling, and measurement biases create substantial uncertainties in return period estimates of rare flood events. The here presented research assess if simulated precipitation from kilometer-scale atmospheric models can be used to improve flood risk estimates of critical infrastructure. Simulated heavy precipitation events from three kilometer-scale 36-hour weather forecast datasets that cover the central and eastern U.S. are compare to high-resolution multi-sensor and station-based precipitation observations. We show that kilometer-scale models can accurately simulate extreme storm characteristics such as movement speed, orographic precipitation gradients, mean and extreme precipitation intensities, and the location of peak precipitation accumulations. The simulations can outperform gridded precipitation observations that solely rely on gauges in capturing extreme accumulations. These findings are incorporated into a conceptual framework for integrating kilometer-scale rainfall simulations into probabilistic flood hazard assessments.

12. KEY WORDS/DESCRIPTORS (List words or phrases that will assist researchers in locating the report.)

Heavy rainfall, convection-permitting modeling, flood hazard assessments, meteorology, hydrology

13. AVAILABILITY STATEMENT
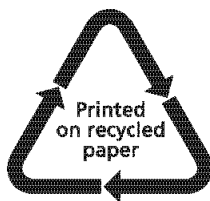
unlimited

14. SECURITY CLASSIFICATION

(This Page)

unclassified

(This Report)

unclassified

15. NUMBER OF PAGES

16. PRICE

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS

@NRCgov

Convection-Permitting Modeling for Intense Precipitation Processes