

## Blasts from the Past

### Reimagining Historical Storms with Model Simulations to Modernize Dam Safety and Flood Risk Assessment

Kelly Mahoney, Chesley McColl, Douglas M. Hultstrand, William D. Kappel, Bill McCormick, and Gilbert P. Compo

**ABSTRACT:** Accurate estimation of the potential “upper limit” for extreme precipitation is critical for dam safety and water resources management, as dam failures pose significant risks to life and property. Methods used to estimate the theoretical upper limit of precipitation are often outdated and in need of updating. The rarity of extreme events means that old storms with limited observational data are often used to define the upper bound of precipitation. Observations of many important old storms are limited in spatial and temporal coverage, and sometimes of dubious quality. This reduces confidence in flood hazard assessments used in dam safety evaluations and leads to unknown or uncertain societal risk. This paper describes a method for generating and applying ensembles of high-resolution, state-of-the-art numerical model simulations of historical past extreme precipitation events to meet contemporary stakeholder needs. The method was designed as part of a research-to-application-focused partnership project to update state dam safety rules in Colorado and New Mexico. The results demonstrated multiple stakeholder and user benefits that were applied directly into storm analyses utilized for extreme rainfall estimation, and diagnostics were developed and ultimately used to update Colorado state dam safety rules, officially passed in January 2020. We discuss how what started as a prototype research foray to meet a specific user need may ultimately inform wider adoption of numerical simulations for water resources risk assessment, and how the historical event downscaling method performed offers near-term, implementable improvements to current dam safety flood risk estimates that can better serve society today.

**KEYWORDS:** Ensembles; Mesoscale models; Emergency preparedness; Flood events; Policy; Risk assessment

<https://doi.org/10.1175/BAMS-D-21-0133.1>

Corresponding author: Kelly Mahoney, [kelly.mahoney@noaa.gov](mailto:kelly.mahoney@noaa.gov)

In final form 16 September 2021

©2022 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

**AFFILIATIONS:** Mahoney—NOAA/Physical Sciences Laboratory, Boulder, Colorado; McColl and Compo—NOAA/Physical Sciences Laboratory, and Cooperative Institute for Research in the Environmental Sciences, University of Colorado Boulder, Boulder, Colorado; Hultstrand and Kappel—Applied Weather Associates, Monument, Colorado; McCormick—Colorado Division of Water Resources, Denver, Colorado

**A**ccurate assessment of flood risk is critical to protecting lives and property worldwide. The design and safe operation of dams, levees, culverts, bridges, storm drainage infrastructure, and many nuclear facilities are informed by estimates of an “upper bound” of possible precipitation. In particular, dams and nuclear facilities in populated areas are often referred to as “critical” or “high hazard” due to the risk to life and property a failure presents. These structures must be built to withstand the most extreme storm or flood considered possible at that location. In engineering practice, this concept is called Probable Maximum Precipitation (PMP), and it is defined as the “theoretical maximum precipitation for a given duration under modern meteorological conditions” (WMO 2009). In the United States, PMP is generally estimated using a deterministic “moisture maximization method” (also referred to as the storm-based approach), which combines observations of historical extreme precipitation events in regions relevant to the location of interest with storm maximization assumptions. In an attempt to facilitate and standardize PMP estimation, NOAA published a series of regional “manuals” known as Hydrometeorological Reports (HMRs) beginning in the 1940s ([https://www.weather.gov/owp/hdsc\\_pmp](https://www.weather.gov/owp/hdsc_pmp)). HMRs, and many of the approximations at the core of PMP and federally published extreme precipitation estimation concepts in general, are frequently cited as being in need of update and improvement (NRC 1994; Tomlinson and Kappel 2009; England et al. 2020; Wright et al. 2021).

Due to recent advancements in high-resolution weather modeling, particularly the ability to simulate convection explicitly, numerical weather models are able to simulate “PMP storms” (i.e., replicate past extreme events, which historically “control” PMP estimates), along with generating reliable model simulation output over longer, continuous periods. Dynamical weather models produce spatially and temporally continuous precipitation estimates, often at considerably higher resolution than observations or historical reanalysis datasets. Because these data are produced by solving physical equations of the atmosphere (in contrast to interpolation methods historically employed to make up for limited observations), dynamical model representation of storm physics and evolution also reduces reliance on spatial, temporal, and physical assumptions that currently underpin PMP estimation (e.g., storm transposition, storm templates, moisture maximization; see also WMO 2009; Mukhopadhyay and Kappel 2016). Dynamical model output also provides coverage in remote, data-sparse regions (e.g., complex and/or high-elevation topography). Furthermore, the explicit model representation of precipitation can resolve precipitation type (snow, rain, hail) as opposed to approximations based on algorithms using surface temperature or similar. Numerical models on both weather and climate time scales are also likely to be critical to informing updates to PMP, which incorporate the role of climate change in the anticipated increase of extreme precipitation (e.g., Mahoney et al. 2018b; McCormick et al. 2020). Finally, assessing uncertainty in PMP estimation is a considerable challenge (e.g., Micovic et al. 2015); dynamical model output offers relatively straightforward methods for the quantification of uncertainty (e.g., ensemble diagnostics and analytics such as spread, sensitivities relative to stochastic perturbation impacts, and information theory.)

Studies investigating numerical model approaches for the specific application of PMP estimation extend back over two decades, including Abbs (1999) who made an early attempt to estimate PMP by simulating an extreme storm event in Australia. Other methods have been developed such as the “atmospheric boundary condition shifting” method by

Ishida et al. (2015a), aimed at maximizing moisture flux over a given watershed. Other studies have sought to downscale various reanalysis datasets to reconstruct major historic storms, sometimes with moisture maximization applied, to estimate PMP (Tan 2010; Ohara et al. 2011; Ishida et al. 2015b; Chen and Hossain 2016).

Outside of the realm of PMP specifically, there have been a handful of other studies to examine the possibility of high-resolution historical event modeling, even for events that occurred so long ago that there are extremely limited, if any, observations available (e.g., Hart 2010; Becker et al. 2010). Specifically, Michaelis and Lackmann (2013) performed a downscaling study of the Twentieth Century Reanalysis (20CR) ensemble dataset (Compo et al. 2011), dynamically downscaling the 20CR ensemble mean using a 6-km horizontal grid within the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008) for the New England Blizzard of 1888. Stucki et al. (2015) performed a similar dynamical downscaling of the 20CR, using individual 20CR ensemble members to downscale a severe foehn storm to a 3-km horizontal grid in order to further examine loss modeling.

The existing charge to update and improve PMP estimation (e.g., NRC 1994; ESEWG 2018), in combination with significant gains in computational power and improvements in numerical weather forecasting models over the past several decades, present opportunities to reconsider improvements in existing PMP estimation methods (e.g., Cotton et al. 2003; Mahoney et al. 2018a; ESEWG 2018; Toride et al. 2019). A recent project focused on PMP estimation for the application of improving dam safety in Colorado and New Mexico (United States) applied numerical weather modeling to supplement existing data and advance both deterministic and probabilistic PMP estimation methods. Model-based historical event reconstruction was one way that numerical model data were directly incorporated into updated PMP estimates; it is this method that we describe here.

The purpose of this manuscript is to demonstrate the potential utility of high-resolution mesoscale model simulations of historical extreme precipitation events that currently control PMP estimates in the United States, but around which there may be great uncertainty due to extremely limited observations. Specifically, we describe a high-resolution, convection-permitting dynamical downscaling ensemble modeling method used to simulate seven historic storms that are important for current PMP estimates in the western United States. We also document how results have been incorporated into the 2016–18 Colorado–New Mexico Regional Extreme Precipitation Study (CO-NM REPS; CODNR and NMOSE 2018) and led to updated and improved state dam safety rules. A forward-looking aim of this paper is also to more generally document the strengths and weaknesses of model approaches for specific application to PMP, and to establish broader context for future opportunities to improve extreme precipitation estimation and flood risk assessment.

## Data and methods

***The Twentieth Century Reanalysis project.*** Upper-air data are particularly important to initial and lateral boundary conditions in a numerical weather simulation, but are exceedingly rare prior to radiosonde launches becoming routine in the 1940s (Durre et al. 2006). Thus, many historical reanalysis products such as National Center for Atmospheric Research (NCAR) reanalysis (Kalnay et al. 1996) often begin during or after the 1940s, when rawinsonde data coverage became more established. This presents a particular problem for investigating extreme events that occurred prior to the 1940s. Therefore, where does one obtain observational data for extreme events that occurred long ago?

Advances in data assimilation and the innovative use of historic surface observations, have allowed reconstruction of three-dimensional atmospheric states in products such as the National Oceanic and Atmospheric Administration (NOAA)/Cooperative Institute for Research in Environmental Sciences (CIRES) Twentieth Century Reanalysis (20CR) project

as far back as the mid-nineteenth century (Compo et al. 2011). Reanalyses are of great value in their own right (e.g., Maddox et al. 2013; Slivinski 2018; Slivinski et al. 2021), providing the ability to examine long-past events, but their potential to serve as initial and boundary conditions also make possible high-resolution mesoscale model simulations of historic storms (e.g., Hart 2010; Becker et al. 2010; Michaelis and Lackmann 2013; Stucki et al. 2015). The use of the 20CR for such purposes has been described as realizing “the potential to enter an era that has hitherto been the province of environmental historians” (Stucki et al. 2015).

The 20CR version 2c (20CRv2c) was used for most of the simulations conducted in this study, as it was the most recent version of the 20CR at the time of the CO-NM REPS study (2016–18). The 20CRv2c is a 56-member reanalysis product utilizing the NCEP Global Forecast System modeling framework in combination with ensemble Kalman filter data assimilation techniques to incorporate surface observations from 1851 to 2014. The horizontal grid spacing is effectively ~200 km, with 28 vertical levels. More information can be found in Compo et al. (2011), Slivinski et al. (2019), and at [www.esrl.noaa.gov/psd/data/gridded/data.20thC\\_ReanV2c.html](http://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2c.html).

That the 20CRv2c contains 56 members (with later versions now containing more than 80 members; see Slivinski et al. 2019) is a key advantage relative to deterministic reanalysis datasets. Such a large ensemble imparts the benefit of representing uncertainty and spread in the reanalysis solutions, which provides critical context regardless of whether high-resolution downscaling applications ultimately select a smaller subset of members due to computational constraints.

Methods to select ensemble members vary across past studies. For example, Stucki et al.’s (2015) foehn wind study selected 20CR members based on pressure gradient magnitude to infer wind speed potential, and the objectives of that study advocate for selecting a more probable versus an outlier simulation. For other objectives, the 20CR ensemble mean may provide adequate initial condition data for further downscaling (e.g., Michaelis and Lackmann 2013). For the present study’s goal of representing a spectrum of possible rainfall scenarios, 20CRv2c ensemble members were selected as initial conditions based on two fundamental heavy precipitation ingredients: omega (atmospheric vertical motion) and precipitable water (PW). The specific selection of individual members was performed first by randomly selecting an initial, more manageable, subset of 20–30 members, and then visually assessing moisture and vertical motion to inform an ultimately heuristic smaller subset of 20CRv2c members. Members were also chosen to include extrema in the spread in the driving synoptic environment (e.g., the members with highest and lowest PW or omega maxima); in order to include as much of the 20CR range of possible environments in which the storm could have evolved.

**Cases examined.** Extreme precipitation events were selected based on the needs of the CO-NM REPS study, aimed at updating and improving PMP estimates for Colorado and New Mexico. The storms chosen are listed in Table 1, and were chosen based on

- 1) importance in existing and previous PMP values,
- 2) lack of observations from which to derive robust storm patterns and magnitudes,
- 3) uncertainty in the previous analysis results, and
- 4) limited surface observation data for rainfall analysis and storm maximization.

**Ensemble modeling framework.** All historical event model simulations employ the Advanced Research Weather Research and Forecasting (WRF-ARW) modeling system, version 3.7.1 (Skamarock et al. 2008). Convection-permitting models are necessary to simulate heavy precipitation with acceptable fidelity, especially at subdaily scales, as sufficiently high model resolution (generally  $\leq 4$  km) permits explicit simulation of deep convection

**Table 1. The seven selected historical events, identified by their year of occurrence, location (city, state), maximum precipitation record, and accumulation duration.**

Year of occurrence	Location	Max total storm precip from SPAS analysis	Precip accumulation duration
1909	Rattlesnake, Idaho	17.20 in./436.88 mm	72 h
1923	Savageton, Wyoming	17.65 in./446.02 mm	72 h
1921	Penrose, Colorado	12.19 in./309.63 mm	18 h
1894	Ward District, Colorado	11.15 in./283.21 mm	72 h
1935	Elbert/Cherry Creek, Colorado	24.00 in./609.60 mm	12 h
1990	Opal, Wyoming	7.16 in./181.64 mm	2 h
1922	Virsylvania, New Mexico	7.53 in./191.62 mm	4 h

(e.g., Prein et al. 2015), which is often critical to generating the types of extreme rainfall that define PMP-type events. Table 2 details the WRF Model setup: 54 vertical levels were used, and the small innermost nest grid spacing affords the omission of convective parameterization and sufficiently resolves flow in and around fine-scale terrain features. As detailed above, initial and lateral boundary conditions are provided by the 20CRv2c. While specific grid spacings, nesting options, and model physics combinations were tested and evaluated in initial method development work, the relevant selected model physics choices are detailed in Table 2.

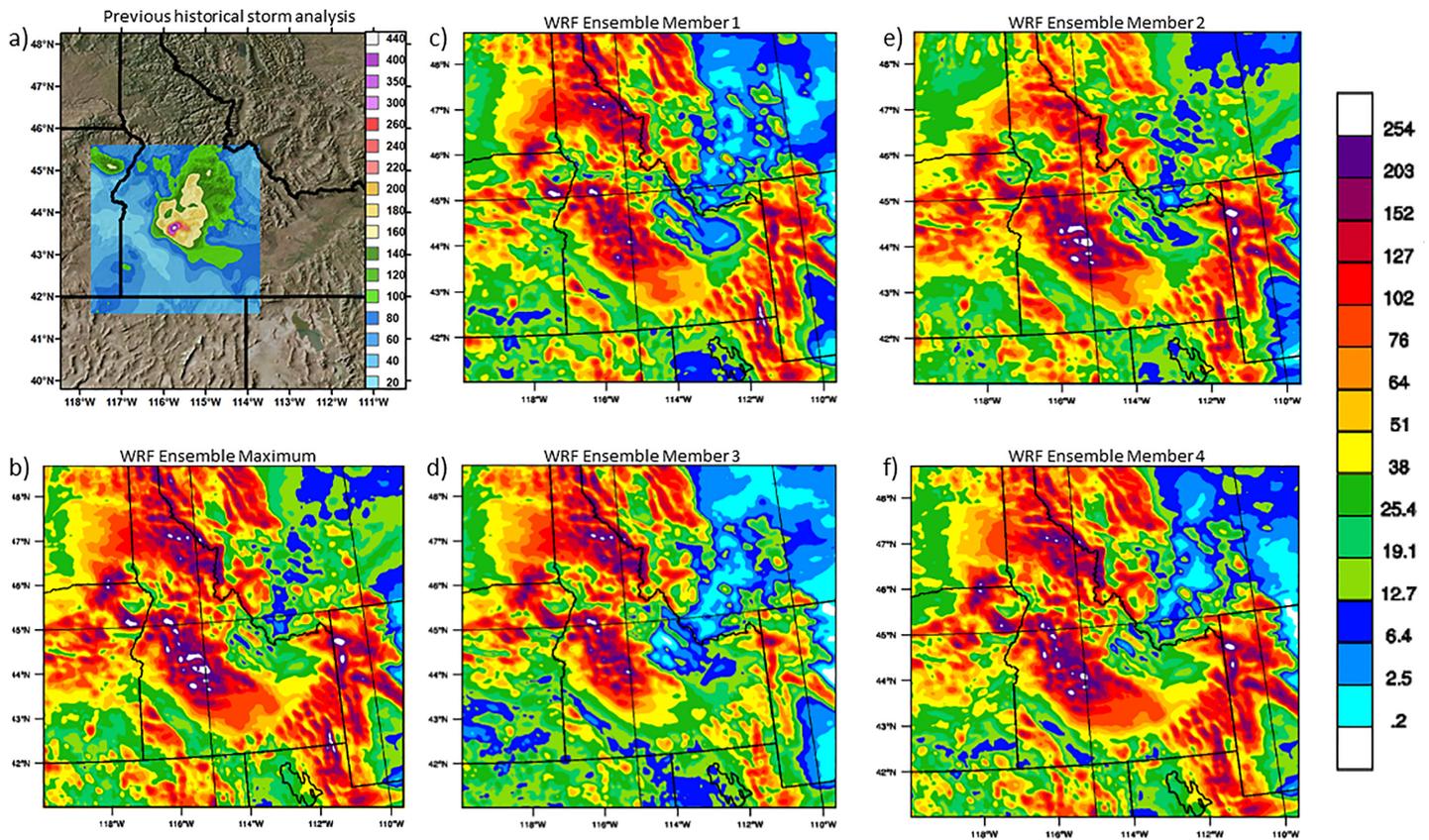
The original CO-NM REPS plan scoped four high-resolution simulations (initialized using four different members of the 20CR ensemble) for each historical event. However, initial WRF ensemble results sometimes raised more questions than answers; for example, impractically large downscaled ensemble spread, or WRF simulations so starkly different from the existing historical analysis so as to be deemed unusable by practitioners. In these situations, additional simulations beyond the standard initial four were performed, the implications of which are discussed below.

### Historical storm simulation results

**Example 1: Strongly forced, orographically focused event.** An example of a “successful” historical event model study, the 1909 Rattlesnake, Idaho, record rainfall event was the result of a week-long series of inland-penetrating atmospheric rivers, which produced a reported 16.12 in. (409 mm) of precipitation over the period 18–24 November 1909 (Fig. 1a; CODNR and NMOSE 2018, Vol. II, appendix M). Atmospheric rivers are synoptic-scale weather systems that are inherently more predictable and better represented by numerical models relative to small-scale convective storms (e.g., Moore et al. 2015). As such, the combination

**Table 2. WRF Model specifications used for historical simulations.**

Characteristic	Details
Model version	WRF (ARW), version 3.7.1
Grid	4-km horizontal grid spacing, 54 vertical levels
Initial and boundary conditions	56-member NCEP GFS-based reanalysis using EnKF DA; [20CRv2c; Compo et al. (2011), doi:10.1002/qj.776; <a href="http://reanalyses.org/">http://reanalyses.org/</a> ]
Model physics	Explicit convection Thompson microphysics YSU planetary boundary layer (PBL) scheme Unified NOAA land surface model, revised MM5 Monin–Obukhov surface layer Dudhia, RRTM radiation physics



**Fig. 1.** (a) Total storm precipitation analysis using Applied Weather Associates Storm Precipitation Analysis System (SPAS; Hultstrand and Kappel 2017) data and estimation methods for the 1909 Rattlesnake, Idaho, event. (b) WRF ensemble maximum precipitation from members 1–4 (mm, shaded, see color bar at right). (c) Event total WRF-simulated precipitation for ensemble member 1 (mm, shaded, color bar at right). (d) As in (c), but for ensemble member 2. (e) As in (c), but for ensemble member 3. (f) As in (c), but for ensemble member 4.

of large-scale, intense atmospheric features more likely to be well represented in the 20CR, plus steep orography in the Sawtooth Mountain Range of Idaho, where the storm center occurred, suggests enhanced potential for successful numerical model simulation of extreme precipitation relative to more weakly forced, small-scale, and/or nonorographically focused precipitation events.

Indeed, for this case, a small ensemble of four WRF simulations produced notable internal consistency and agreed closely with available historical observations (Fig. 1). The WRF Model precipitation output fields were incorporated into PMP calculation methods first as an improved precipitation “base map” (from which PMP estimation begins; see also CODNR and NMOSE 2018, Vol. II and Fig. 2). The precipitation base map offers a starting point for spatial distribution of precipitation values between observational data points that inform the ensuing PMP estimation process. WRF simulation output was next used to inform a more robust method to delineate regions of rain versus snow (relative to coarser, temperature-based approximations). Snow is less relevant to direct surface runoff and is thus omitted from PMP calculations. High-resolution, convection-resolving simulations allow the benefit of relatively sophisticated cloud microphysics representation of precipitation type (rain vs snow), allowing for further refinement of flood-relevant precipitation. Updating the preexisting storm analysis generated by interpolating sparse observations with high-resolution, rainfall-only WRF information made significant differences (upward of ~75 mm, or ~3 in.), distributed across the domain (Fig. 2). And because this storm controls PMP depths in many locations, this difference in precipitation depths directly affects flood-runoff and hydrologic design parameters.

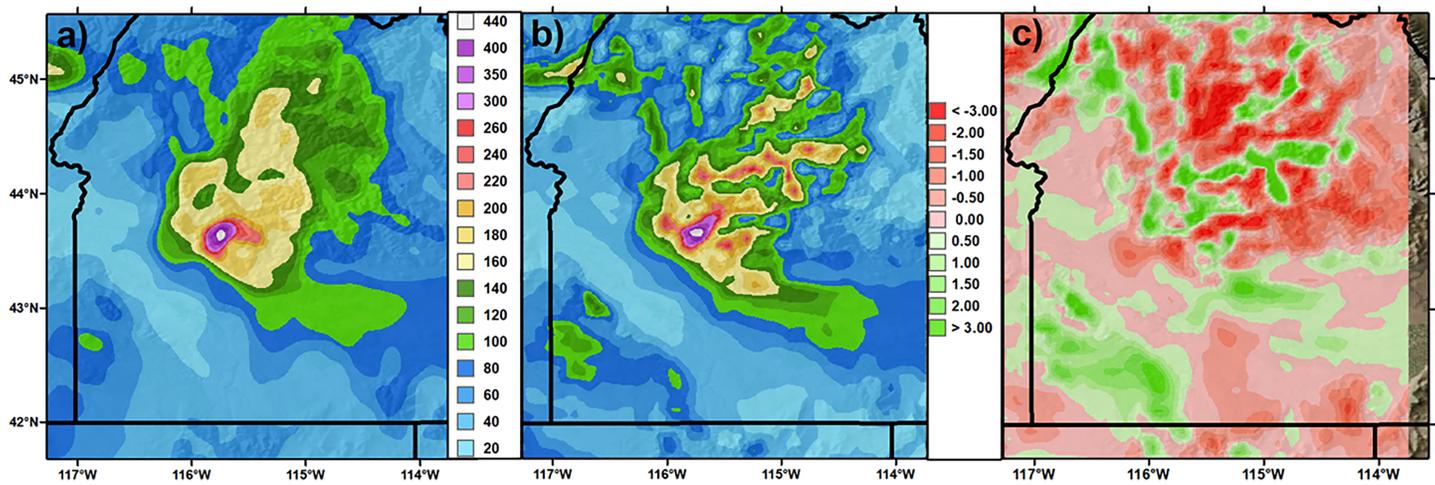


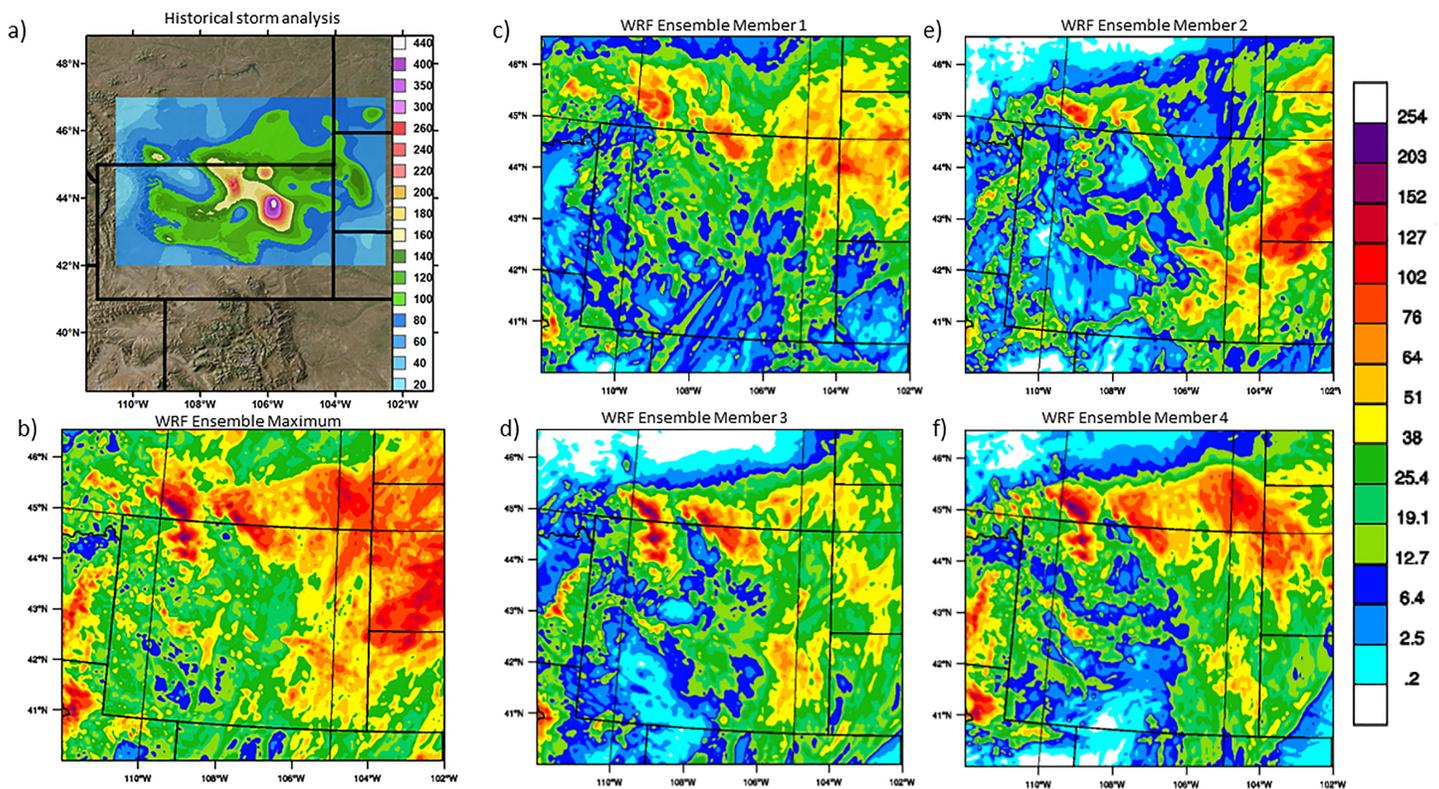
Fig. 2. Historical storm analyses for 1909 Rattlesnake, Idaho, extreme precipitation event. (a) Original storm total precipitation analysis (no incorporation of WRF simulation information; mm, see color bar to right of this panel). (b) As in (a), but updated using WRF spatial, temporal, and microphysical (precipitation-type) information (see color bar to left of this panel). (c) Difference between original historical analysis and updated WRF-informed storm reanalysis (see color bar to left of this panel). Storm analyses performed by Applied Weather Associates for the CO-NM REPS using SPAS.

With respect to the utility of the WRF simulations for this event, ensemble diagnostics such as “ensemble maximum precipitation” (Fig. 1b) were deemed to be of greatest use, versus the selection of a “best” individual simulation to be used in isolation. Along with using ensemble maxima and averages to improve the estimated historical spatial distribution of precipitation, temporal storm patterns were also useful to more accurately quantify the accumulation characteristics through time. Even considering the uncertainties implicit in the historical reanalysis data, using a dynamical weather model initialized and periodically updated on its boundaries more strongly bases the storm analysis in physical, dynamical evolution as opposed to stochastic, synthetic storm time series templates, which bear no individual connection to specific cases.

In summary, a modest ensemble of four high-resolution dynamical model simulations of the November 1909 Rattlesnake, Idaho, record rainfall event was judged by practitioners and a large team of subject-matter experts recruited for the CO-NM REPS Project Review Board to demonstrate value in improving PMP estimates for a critical PMP-controlling event. For this storm in particular, the most useful aspects for application to current PMP estimation practice included the simulated spatial and temporal distributions of precipitation, as well as explicitly predicted rain-snow delineation.

**Example 2: Weakly forced, nonorographic, observation-limited event.** A rain event with a reported storm center near Savageton, Wyoming, occurred during September 1923. This storm has historically controlled PMP depths for many regional studies (e.g., Schreiner and Riedel 1978; Hansen et al. 1988) but contains tremendous uncertainty related to the storm center rainfall amount and spatial accumulation patterns. The case has remained a PMP-controlling event in many recent PMP studies (e.g., Tomlinson et al. 2008; Kappel et al. 2014; CODNR and NMOSE 2018; Kappel et al. 2021), despite known shortcomings in the number and reliability of available historical observations. It was therefore selected as an important historical dynamical model downscaling candidate based on its critical role in determining PMP values.

The case is described in past PMP studies as having produced 17.1 in. (434 mm) of precipitation in 108 h (Fig. 3a); however, the source of the singular maximum precipitation observation was a rancher who recorded the precipitation by twice emptying a “14-quart water pail” (Fig. 4) (Follansbee and Hodges 1925; Grover 1925). Aside from the potential uncertainty



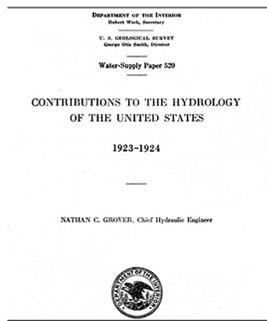
**Fig. 3.** (a) Total storm precipitation analysis using Applied Weather Associates SPAS data and estimation methods for 1923 Savageton, Wyoming, event. (b) WRF ensemble maximum precipitation from all WRF ensemble members (mm, see color bar to right of this panel). (c) Event total WRF-simulated precipitation for ensemble member 1 (mm, see large color bar to right of figure). (d) As in (c), but for ensemble member 2. (e) As in (c), but for ensemble member 3. (f) As in (c), but for ensemble member 4.

implicit in such an observation, the case is also characterized by having a very limited number of hourly and daily data near the relatively small storm center, and the existing data are largely estimated from generalized assumptions used by the U.S. Army Corps of Engineers and National Weather Service to convert limited point data into smooth mass rainfall curves (U.S. Army Corps of Engineers 1962).

The meteorological description for this case provided by historical reports classifies the event as a midlatitude synoptic cyclone with moisture sourced from the Gulf of Mexico. The initial four downscaled WRF simulations for the Savageton, Wyoming, 1923 event yielded very little precipitation in the vicinity of northeastern Wyoming. WRF configurations were accordingly adjusted and new physics, initialization times, and additional 20CR members were used in an attempt to simulate precipitation totals even in the ballpark of the extreme historical precipitation observation. After more than 15 WRF configurations and initializations were tested, the most precipitation generated was an ensemble maximum single gridpoint value of ~50 mm (~4 in.) versus 17.1 in. in northeastern Wyoming.

To respect time constraints and not impede the larger CO-NM REPS process, practitioners chose to move on from this case, concluding: “Unfortunately, the WRF reanalysis of the Savageton storm showed little skill in being able to replicate either the spatial pattern or magnitude of the storm. Therefore, the WRF reanalysis results were not used in the Savageton SPAS [Applied Weather Associates’ Storm Precipitation Analysis System] analysis” (CODNR and NMOSE 2018, Vol. II). A more comprehensive exploration of the role of historical observational uncertainty relative to increasing ensemble size and spread characteristics is recommended.

Though practical project requirements, resource constraints, legacy, and institutional inertia collectively resulted in the Savageton, Wyoming, 1923 remaining as a PMP-controlling



## 118 CONTRIBUTIONS TO HYDROLOGY OF UNITED STATES, 1923–1924.

Roughly approximate records were also obtained at other points. B. F. Horton, who lives  $4\frac{1}{2}$  miles northeast of Arvada, measured  $8\frac{1}{2}$  inches of rainfall from September 20 to 30 in a coffee can in his yard. William La Follett, who lives in sec. 23, T. 45 N., R. 87 W., had a water tank 3 by 8 by  $2\frac{1}{2}$  feet deep in an exposed position. The rain began at 4 a. m. September 27, and by 4 p. m. on the 29th the tank contained 5 inches of water; two days later this had increased to 6 inches. A rancher near Savageton, in T. 45 N., R. 76 W., stated that a 14-quart water pail with nearly vertical sides, which stood in an exposed place, was filled twice in 48 hours, indicating an almost unbelievable rainfall of 17 inches. At Ross, just east of the Powder River basin, in T. 40 N., R. 75 W., no records were kept, but the postmaster stated that the rainfall was the heaviest in years. As much of the basin is very sparsely settled, it is possible that heavy rainfall occurred at other points and was not recorded.

Fig. 4. Excerpt from Grover's (1925, p. 118) "Contributions to the Hydrology of the United States 1923–1924," summarizing a rancher's precipitation observation of 17 in. (~432 mm) over 48 h via a 14-quart (~13.2-L) water pail measurement.

event, we highlight this case as an example where numerical experiments may provide insight into situations of potentially questionable historical observational veracity. Though not carried out here, one might imagine that in situations such as this, historical downscaling model ensembles may also aid in deselecting events and associated limited observations, which may not be justified to include in modern-day PMP estimation. In other words, if a sufficient number of numerical simulations, proven in advance to be of high-quality experimental design and proven skillful for other simulations, cannot simulate a reasonable proportion of historical precipitation known to be of a dubious observational nature, then the event may be considered to be in need of additional review.

**Additional storms studied.** Five additional historical extreme precipitation cases were also downscaled by the WRF high-resolution ensemble method to further evaluate case-to-case relative utility of the approach. The perceived applicability of the ensemble simulation for these cases varied; for more synoptically driven cases, WRF simulation output was more likely to be deemed useful in updating spatial and temporal storm patterns. For cases considered to be driven more by convective, isolated-thunderstorm type activity, the WRF simulations were more likely to be regarded as lacking adequate correspondence to the available observations upon which preexisting estimates and current practices are based, and thus not included (see CODNR and NMOSE 2018, Vol. II, appendix L, for additional case studies).

Improving, further testing, and ultimately standardizing the experimental design for a modeling component of any specific study is needed in order to establish the most robust methods and appropriate application of results. Recommended criteria for more rigorous testing in future work are described below.

### Application of results and implications for future work

**Open questions and opportunities.** High-resolution, convection-resolving downscaling of extreme precipitation events using coarse, observation-limited historical reanalysis ensemble members suggests promise for modernizing and improving PMP and extreme precipitation estimation. Innovative but provisional, this exploratory effort to update official state dam safety rules by combining traditional historical event data with state-of-the-art modeling capabilities also illuminated opportunities and development needs in the following areas.

First, running an ensemble of simulations offers a more physically based and internally consistent method to estimate extreme precipitation from historical events. However, these advances come with increased computational and overall labor/effort costs relative to following current PMP estimation procedures. A possible reduction in computational cost may be found in understanding whether, and in what situations, the 20CR ensemble mean may be used as initial and boundary conditions to force a single downscaled simulation, which still retains the salient aspects of an event. While some studies have used this approach with relative success (e.g., Michaelis and Lackmann 2013), preliminary testing in this study confirms that the success of this method is generally inversely proportional to the amount of spread in the 20CR for a given event. That is, the more spread found in the 20CR for a given event, the less useful an ensemble mean-forced downscaling is likely to be. Given that extreme precipitation is often characterized by a rare combination of environmental properties, which lead to nonlinear precipitation processes, it seems unlikely that downscaling the ensemble mean, with all the fields essentially averaged together so as to smooth potentially important environmental extrema, will reliably represent an extreme event. Indeed, tests using the 20CR ensemble mean for initial conditions for the cases chosen in this study resulted in less intense downscaled outcomes. Finally, a single deterministic outcome relative to an ensemble of multiple simulations also reduces uncertainty information and valuable context regarding confidence.

Second, the 20CRv3 has even more ensemble members from which to choose (80 members vs 20CRv2c's 56 members). The sensitivity of using this dataset instead was evaluated for one case in this study. Though results were unchanged in the test, using additional 20CR ensemble members fundamentally increases the potential to represent greater spread in downscaled ensembles. Therefore, particularly in the face of growing 20CR ensemble size, exploring methods for strategic, targeted 20CR member selection is recommended. Future work could also establish a standardized method to evaluate 20CR spread to contextualize initial condition uncertainty. Relevant measures of spread and variability in the 20CR membership could also be compared to the spread and variability of those same metrics in the downscaled ensemble. This would be an insightful marker of what has been gained (or lost) in the exercise of high-resolution downscaling to better understand historical extreme event potential.

Finally, the ensemble downscaling framework yields critical uncertainty information, but also new questions regarding optimal use of the multimember output. The incorporation of individual model member fields versus ensemble diagnostics (e.g., mean, max, spread) was explored via ongoing collaborative discussion with CO-NM REPS practitioners. Individual simulations retain the model-derived benefit of internal physical consistency, while ensemble diagnostics provide useful analytic insight. In this study, it was ultimately uncommon for data from a single, individual model simulation to be deemed robust, reliable, or as useful (relative to the entire ensemble) in isolation. Instead, ensemble diagnostics and intra-ensemble, member-to-member comparisons were key to gaining use and comfort with PMP practitioners. For the unique challenge of PMP, the ensemble maximum ("ensemble max") product in particular seems to be an appropriate diagnostic selection. The ensemble max grid retains the maximum event-total precipitation produced at each grid point and thus demonstrates how intense the event was simulated to be, grid point by grid point, across all event ensemble members. In the CO-NM REPS project, individual member model output for each event, along with an ensemble max precipitation grid, was provided to be considered for possible input into the PMP analysis. The ensemble max grid was ultimately selected as the product from which to reevaluate or modify existing precipitation base maps, but exploration of more sophisticated ensemble postprocessing strategies is recommended.

***Improving future project design and research integrations.*** Going forward, dynamical model approaches to simulating historical storms for applications such as PMP should

establish, a priori, a more structured and exhaustive experimental design, which includes clear standards for the governance of possible application of model results. The November 1909 Rattlesnake, Idaho, WRF simulations demonstrated that reconstruction of major historical events via numerical modeling may beneficially supplement existing storm analyses and also improve spatial, temporal, and physical assumptions (e.g., precipitation type) made with very limited observational data. This event (in combination with and compared with others) highlights the role of topography in producing more constrained simulations that may be deemed more valuable to practitioners. This hypothesis requires further testing in a research realm but may offer guidance in the planning stages of emerging studies.

Conversely, for cases where model simulations did not yield the expected, historical observation-indicated precipitation, model data might instead be considered as a tool in flagging potentially erroneous, or at least unacceptably uncertain, observational data. It should also be noted that additional sources of uncertainty are introduced via model study methodological choices: domain configuration, simulation duration, ensemble size, spinup time, model physics, and more. Care should be taken to minimize the degree to which these additional subjectivities may compound existing uncertainties in the PMP estimation process (e.g., storm-maximizing PMP processes). Collectively, the above considerations again advocate for the a priori establishment of project design and model data incorporation criteria that is ideally objective, perhaps based on ensemble skill or spread statistics and integrates a quantitative measure of observational uncertainty.

### ***The future of dynamical weather modeling in extreme precipitation estimation.***

Combining old PMP methods with new model data offers incremental improvements for limited-area, site-specific studies in particular instances. However, adding long-term value toward achieving an objective, NWP-generated upper bound of precipitation will require additional work. There is mounting desire across many user groups and sectors of the hydrometeorological and hydrologic communities for the application of advanced dynamical model methods and output in extreme precipitation estimation. The utility of high-resolution model data has been demonstrated for dam safety and flood risk management applications in case-specific efforts through exploratory prototypes using longer-term, continuous model data output. While, to date, only short-term (~5 years) prototype capabilities have been possible due to computing requirements (e.g., CODNR and NMOSE 2018, Vol. IV), there is increasing acknowledgment (Mahoney et al. 2018a; Mahoney 2021; Prein et al. 2021) of the opportunity to amass high-resolution data through data mining of existing model output.

Recognizing the considerable potential of diverse dynamical modeling methods, of which historical extreme event downscaling is just one example, we emphasize the responsibility to comprehensively reexamine the challenge of PMP estimation and consider the full scope of future potential improvements. We specifically recommend an inclusive and thorough National Academies of Science, Engineering, and Medicine (NASEM) study of the current state of the practice and options for extreme rainfall estimation.

### **Summary**

A method to generate and apply high-resolution, state-of-the-art numerical model simulations of historical extreme precipitation events has demonstrated potential to benefit dam safety. Developed, tested, and evaluated as part of the CO-NM REPS project, the historical event downscaling simulations presented here provided multiple avenues for updating previous extreme storm data and PMP depths. The results ultimately informed updates to the State of Colorado Dam Safety Rules associated with allowable rainfall

estimates used to develop inflow design floods for design of safe spillways at high and significant hazard dams (Colorado Division of Water Resources 2020). Lessons learned from different types of event simulations are demonstrated through two example cases. Despite a relatively small sample of cases, we identify areas of relative robustness of results: for example, strongly forced, orographically controlled cases produced numerical simulations that practitioners and experts identified as sufficiently matching historical observations, and thus were incorporated in updated PMP estimates. We also document strengths, weaknesses, and future opportunities to improve this approach for use in current PMP estimation, and establish broader motivation for the use of dynamical models in future PMP estimation improvements.

The results of this study corroborate prior historical weather event “reconstruction” work such as those by Stucki et al. (2015), who have advocated for a complementary approach in which traditional and numerical methods are combined. As further posited by Stucki et al. (2015), the introduction of gridded, small time-step numerical model data may well alter our foundational understanding of, and perspectives on, historical extreme precipitation events. The approach illuminates new potential for the so-called “trading of space for time” in which statistical and dynamical analyses are combined to synthetically increase sample sizes (which is always a challenge in studies of rare events). Though case study-focused herein, the approach further informs future applications requiring high-resolution, spatially consistent, and/or long-term gridded data. Decision-maker acceptance of the historical downscaling approach demonstrates an increasing appetite for including dynamical modeling more broadly, for example, with respect to addressing nonstationarity in PMP estimation (e.g., Mahoney et al. 2018b; McCormick et al. 2020).

The historical event dynamical model downscaling approach described here is but one effort among a larger call to improve PMP estimation, a concept and quantity that many argue is in need of fundamental reimagining versus the application of state-of-the-art “band aids.” However, as the practice of PMP underpins present-day dam safety principles, its criticality to maintaining safe and usable estimates renders it, for now, embedded in our collective societal well-being. Complementing the larger, forward-looking movement to improve extreme event risk assessment for hydro-engineering applications, opportunities such as historical event downscaling can offer improvements to current estimates, and in turn better serve society *now*.

**Acknowledgments.** This project was completed under the CO-NM REPS effort, supported by the Colorado Water Conservation Board, the New Mexico Office of the State Engineer, and other state entities; support from the NOAA Physical Sciences Laboratory made the completion of this study possible. Applied Weather Associates provided their time, expertise, and SPAS software, and have allowed use of their graphics herein. Support for the Twentieth Century Reanalysis Project version 2c dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office and Physical Sciences Laboratory. Computer simulations were accomplished using the NOAA Theta Supercomputer, WRF (NCAR), and the 20CR (NOAA PSL). We appreciate the thoughtful discussions throughout the project with Mark Perry (Colorado Department of Water Resources), Trevor Alcott (NOAA GSL), Eric James (NOAA GSL), Rob Cifelli (NOAA PSL), and Laura Slivinski (NOAA PSL). This manuscript was improved by the thoughtful comments of Prof. Russ Schumacher and two anonymous reviewers. We also gratefully acknowledge the Department of Commerce’s support of the allowance to operate the Children’s Commerce Center during the COVID-19 pandemic, without which this work would also not have been possible.

**Data availability statement.** The model simulations performed here are initialized using publicly available NOAA–CIRES–DOE 20th Century Reanalysis (<https://psl.noaa.gov>). The numerical model

simulation output is too large to host via URL but we instead provide all the information needed to replicate the simulations (see “Data and methods” section and Table 2). The model code, compilation script, initial and boundary condition files, and the namelist settings are all also archived and can be made available from the NOAA Hera High Performance Storage System.

## 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**, 785–796, <https://doi.org/10.1029/1998WR900013>.
- Becker, M., M. S. Gilmore, J. Naylor, J. K. Weber, R. A. Maddox, G. P. Compo, J. S. Whitaker, and T. M. Hamill, 2010: Simulations of the supercell outbreak of 18 March 1925. *25th Conf. on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., P8.15, [https://ams.confex.com/ams/25LS/techprogram/paper\\_176071.htm](https://ams.confex.com/ams/25LS/techprogram/paper_176071.htm).
- Chen, X., and F. Hossain, 2016: Revisiting extreme storms of the past 100 years for future safety of large water management infrastructures. *Earth's Future*, **4**, <https://doi.org/10.1002/2016EF000368>.
- CODNR and NMOSE, 2018: Colorado–New Mexico Regional Extreme Precipitation Study. Summary Rep., Colorado Division of Water Resources and New Mexico Office of the State Engineer, 7 volumes, <https://spl.cde.state.co.us/artemis/nrmonos/nr5102p412018internet/>.
- Colorado Division of Water Resources, 2020: Rules and Regulations for Dam Safety and Dam Construction, 2-CCR 402-1. 40 pp., [www.sos.state.co.us/CCR/GenerateRulePdf.do?ruleVersionId=8426&fileName=2%20CCR%20402-1](http://www.sos.state.co.us/CCR/GenerateRulePdf.do?ruleVersionId=8426&fileName=2%20CCR%20402-1).
- Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, <https://doi.org/10.1002/qj.776>.
- Cotton, W. R., R. L. McAnelly, and T. Ashby, 2003: Development of new methodologies for determining extreme rainfall. Final Rep. for Contract ENC #C154213, State of Colorado Dept. of Natural Resources, 143 pp., [https://rams.atmos.colostate.edu/precip-proj/reports/022003/DNR\\_Final\\_report.pdf](https://rams.atmos.colostate.edu/precip-proj/reports/022003/DNR_Final_report.pdf).
- Durre, I., R. S. Vose, and D. B. Wuertz, 2006: Overview of the integrated global radiosonde archive. *J. Climate*, **19**, 53–68, <https://doi.org/10.1175/JCLI3594.1>.
- England, J. F., V. Sankovich, and R. J. Caldwell, 2020: Review of probable maximum precipitation procedures and databases used to develop hydrometeorological reports. Rep. NUREG/CR-7131, U.S. Nuclear Regulatory Commission, 104 pp., [www.nrc.gov/docs/ML2004/ML20043E110.pdf](http://www.nrc.gov/docs/ML2004/ML20043E110.pdf).
- ESEWG, 2018: Extreme Rainfall Product Needs. Extreme Storm Events Work Group, 36 pp., [https://acwi.gov/hydrology/extreme-storm/product\\_needs\\_proposal\\_20181010.pdf](https://acwi.gov/hydrology/extreme-storm/product_needs_proposal_20181010.pdf).
- Follansbee, R., and P. V. Hodges, 1925: Some floods in the Rocky Mountain region. USGS Water Supply Paper 520, 26 pp., <https://pubs.usgs.gov/wsp/0520g/report.pdf>.
- Grover, N. C., 1925: Contributions to the hydrology of the United States, 1923–1924. USGS Water Supply Paper 520, 129 pp., <https://doi.org/10.3133/wsp520>.
- Hansen, E. M., D. D. Fenn, L. C. Schreiner, R. W. Stodt, and J. F. Miller, 1988: Probable maximum precipitation estimates, United States between the continental divide and the 103rd Meridian. NOAA Hydrometeorological Rep. 55A, 242 pp., <https://repository.library.noaa.gov/view/noaa/7154>.
- Hart, R. E., 2010: Simulation of historical hurricane events using 20th Century Reanalysis. *29th Conf. on Hurricanes and Tropical Meteorology*, Tucson, AZ, Amer. Meteor. Soc., P2.128, [https://ams.confex.com/ams/29Hurricanes/techprogram/paper\\_169111.htm](https://ams.confex.com/ams/29Hurricanes/techprogram/paper_169111.htm).
- Hultstrand, D. M., and W. D. Kappel, 2017: The Storm Precipitation Analysis System (SPAS) Report. Nuclear Regulatory Commission (NRC) Inspection Rep. 99901474/2016-201, Enercon Services, Inc., 95 pp.
- Ishida, K., M. L. Kavvas, S. Jang, Z. Q. Chen, N. Ohara, and M. L. Anderson, 2015a: Physically based estimation of maximum precipitation over three watersheds in Northern California: Atmospheric boundary condition shifting. *J. Hydrol. Eng.*, **20**, 04014052, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001026](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001026).
- , ———, ———, ———, ———, and ———, 2015b: Physically based estimation of maximum precipitation over three watersheds in Northern California: Relative humidity maximization method. *J. Hydrol. Eng.*, **20**, 04015014, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001175](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001175).
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kappel, W. D., and Coauthors, 2014: Probable Maximum Precipitation Study for Wyoming. Wyoming Water Development Office, 154 pp., <https://www.wy.us/PMP/PMP.html>.
- , D. M. Hultstrand, J. T. Rodell, G. A. Muhlestein, and K. Steinhilber, 2021: Statewide probable maximum precipitation for North Dakota. North Dakota State Water Commission.
- Maddox, R. A., M. S. Gilmore, C. A. Doswell III, R. H. Johns, C. A. Crisp, D. W. Burgess, J. A. Hart, and S. F. Piltz, 2013: Meteorological analyses of the Tri-State tornado event of March 1925. *Electron. J. Severe Storms Meteor.*, **8** (1), 1–27.
- Mahoney, K. M., 2021: What can dynamical weather modeling offer the next generation of probable maximum precipitation (PMP) estimation? *35th Conf. on Hydrology*, Online, Amer. Meteor. Soc., 603, <https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/382195>.
- , E. James, T. Alcott, and R. Cifelli, 2018a: Application of dynamical model approaches using the NOAA High Resolution Rapid Refresh (HRRR) and Weather Research and Forecast (WRF) models. Colorado–New Mexico Regional Extreme Precipitation Study Summary Report, Vol. IV, Colorado Division of Water Resources and New Mexico Office of the State Engineer, 54 pp., <https://spl.cde.state.co.us/artemis/nrmonos/nr5102p412018internet/>.
- , J. Lukas, and M. Mueller, 2018b: Considering climate change in the estimation of extreme precipitation for dam safety. Colorado–New Mexico Regional Extreme Precipitation Study Summary Report, Vol. VI, Colorado Division of Water Resources and New Mexico Office of the State Engineer, 65 pp. <https://spl.cde.state.co.us/artemis/nrmonos/nr5102p412018internet/>.
- McCormick, W., J. Lukas, and K. Mahoney, 2020: 21st century dam safety rules for extreme precipitation in a changing climate. *J. Dam Saf.*, **17**, 29–42.
- Michaelis, A. C., and G. M. Lackmann, 2013: Numerical modeling of a historic storm: Simulating the Blizzard of 1888. *Geophys. Res. Lett.*, **40**, 4092–4097, <https://doi.org/10.1002/grl.50750>.
- Micovic, Z., M. G. Schaefer, and G. H. Taylor, 2015: Uncertainty analysis for probable maximum precipitation estimates. *J. Hydrol.*, **521**, 360–373, <https://doi.org/10.1016/j.jhydrol.2014.12.033>.
- Moore, B. J., K. M. Mahoney, E. M. Sukovich, R. Cifelli, and T. Hamill, 2015: Climatology and environmental characteristics of extreme precipitation events in the southeastern United States. *Mon. Wea. Rev.*, **143**, 718–741, <https://doi.org/10.1175/MWR-D-14-00065.1>.
- Mukhopadhyay, B., and W. D. Kappel, 2016: Probable maximum precipitation. *Handbook of Applied Hydrology*, 2nd ed. V. P. Singh, Ed., McGraw-Hill Education, 1382 pp.
- NRC, 1994: *Estimating Bounds on Extreme Precipitation Events: A Brief Assessment*. The National Academies Press, 29 pp.
- Ohara, N., M. Kavvas, S. Kure, Z. Chen, S. Jang, and E. Tan, 2011: Physically based estimation of maximum precipitation over American River Watershed, California. *J. Hydrol. Eng.*, **16**, 351–361, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000324](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000324).
- Prein, A. F., and Coauthors, 2015: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Rev. Geophys.*, **53**, 323–361, <https://doi.org/10.1002/2014RG000475>.
- , D. Ahijevych, J. Powers, R. Sobash, C. Schwartz, and E. Towler, 2021: On the applicability of kilometer-scale heavy precipitation simulations in flood risk assessments. *5th Annual Probabilistic Flood Hazard Assessment Workshop*, Rockville, MD, Nuclear Regulatory Commission Office of Nuclear Regulatory Research, 22 pp., [www.nrc.gov/docs/ML2106/ML21064A424.pdf](http://www.nrc.gov/docs/ML2106/ML21064A424.pdf).
- Schreiner, L. C., and J. T. Riedel, 1978: Probable maximum precipitation estimates, United States east of the 105th Meridian. NOAA Hydrometeorological Rep. 51, 87 pp., [www.nrc.gov/docs/ML0901/ML090150038.pdf](http://www.nrc.gov/docs/ML0901/ML090150038.pdf).
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- Slivinski, L. C., 2018: Historical reanalysis: What, how, and why? *J. Adv. Model. Earth Syst.*, **10**, 1736–1739, <https://doi.org/10.1029/2018MS001434>.

- , and Coauthors, 2019: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quart. J. Roy. Meteor. Soc.*, **145**, 2876–2908, <https://doi.org/10.1002/qj.3598>.
- , and Coauthors, 2021: An evaluation of the performance of the Twentieth Century Reanalysis version 3. *J. Climate*, **34**, 1417–1438, <https://doi.org/10.1175/JCLI-D-20-0505.1>.
- Stucki, P., and Coauthors, 2015: Dynamical downscaling and loss modeling for the reconstruction of historical weather extremes and their impacts: A severe Foehn storm in 1925. *Bull. Amer. Meteor. Soc.*, **96**, 1233–1241, <https://doi.org/10.1175/BAMS-D-14-00041.1>.
- Tan, E., 2010: Development of a methodology for probable maximum precipitation estimation over the American river watershed using the WRF model. Ph.D. dissertation, University of California, Davis, 194 pp.
- Tomlinson, E. M., and W. D. Kappel, 2009: Revisiting PMPs. *Hydro Rev.*, **28**, 10–17.
- , ——, T. W. Parzybok, D. Hultstrand, G. Muhlestein, and P. Sutter, 2008: State-wide Probable Maximum Precipitation (PMP) Study for the state of Nebraska. Nebraska Dam Safety, Lincoln, NE, 26 pp., [www.appliedweatherassociates.com/uploads/1/3/8/1/13810758/nebraska-pmp-paper-7-30-08.pdf](http://www.appliedweatherassociates.com/uploads/1/3/8/1/13810758/nebraska-pmp-paper-7-30-08.pdf).
- 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.*, **20**, 2383–2400, <https://doi.org/10.1175/JHM-D-19-0039.1>.
- U.S. Army Corps of Engineers, 1962: Storm rainfall in the United States, depth-area-duration data. Office of Chief of Engineers, USACE, 102 pp.
- WMO, 2009: Manual on Estimation of Probable Maximum Precipitation (PMP). 3rd ed. WMO-1045, Secretariat of the WMO, 257 pp., [https://library.wmo.int/doc\\_num.php?explnum\\_id=7706](https://library.wmo.int/doc_num.php?explnum_id=7706).
- Wright, D. B., C. Samaras, and T. Lopez-Cantu, 2021: Resilience to extreme rainfall starts with science. *Bull. Amer. Meteor. Soc.*, **102**, E808–E813, <https://doi.org/10.1175/BAMS-D-20-0267.1>.