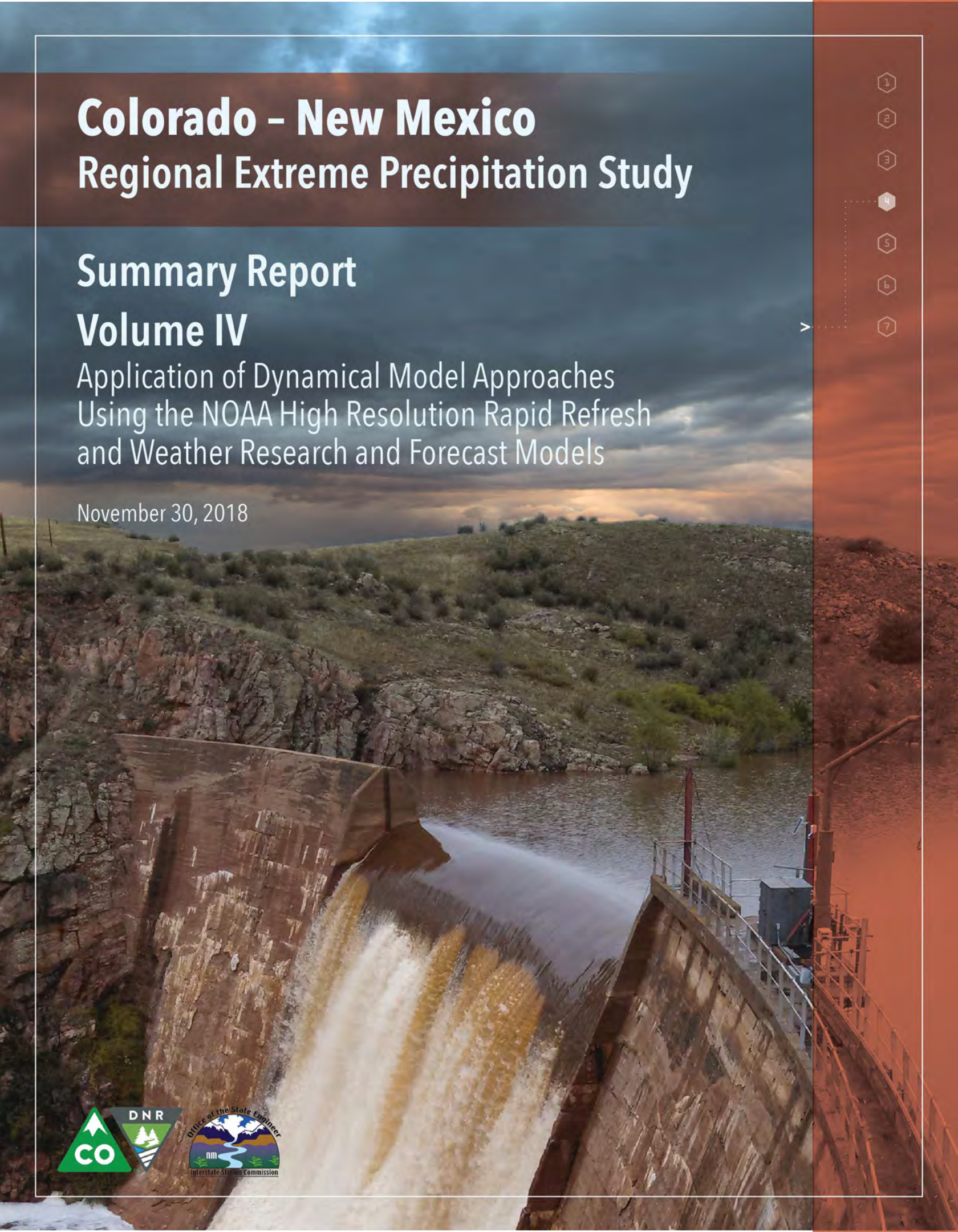
Colorado – New Mexico
Regional Extreme Precipitation Study

Summary Report
Volume IV
Application of Dynamical Model Approaches Using the NOAA High Resolution Rapid Refresh and Weather Research and Forecast Models

November 30, 2018



This page left intentionally blank

Colorado – New Mexico
Regional Extreme Precipitation Study

Summary Report

Volume IV

Application of Dynamical Model Approaches Using the NOAA High Resolution Rapid Refresh and Weather Research and Forecast Models

Prepared by:
NOAA Earth Systems Research Laboratory

November 30, 2018

COLORADO
Division of Water Resources
Department of Natural Resources

Office of the State Engineer
Interstate Stream Commission

This page left intentionally blank

Table of Contents

List of Figures

List of Acronyms

CPC: Climate Prediction Center

ESRL: NOAA Earth System Research Laboratory

GSD: Global Systems Division

GSI: Gridpoint Statistical Interpolation

HRRR: High-Resolution Rapid Refresh

HRRRX: Experimental HRRR model

MCS: mesocale convective system

MRMS: multi-radar, multi-sensor precipitation analysis

NCEP: National Centers for Environmental Prediction

NOAA: National Oceanic and Atmospheric Administration

NWP: Numerical Weather Prediction

PBL: Planetary Boundary Layer

PDF: probability distribution function

PRISM: Parameter-elevation Relationships on Independent Slopes Model

QPE: Quantitative precipitation estimation

QPF: Quantitative precipitation forecast

RAP: Rapid Refresh model

Stage-IV: "Official" precipitation analysis

SWE: Snow water equivalent

WRF: Weather Research and Forecast model

1. Introduction

A unique component of the CO-NM REPS project is the incorporation of a "dynamical modeling framework", with the goal of carrying out a preliminary proof-of-concept study to investigate how forecasts from a high-resolution, rapidly-updating numerical weather prediction (NWP) model might inform various aspects of extreme precipitation estimation. A numerical modeling framework such as this one, leveraging high-resolution model data continuous in space and time, and with considerable additional model data provided via hourly forecast cycles from a state-of-the-art NOAA weather forecast model, has not been not been employed in previous PMP studies, and its potential benefits and limitations are examined here. This component of the CO-NM REPS project is also referred to as "Task 3".

As described in earlier volumes of this report, scientific understanding of physical processes responsible for extreme precipitation events has advanced significantly since the NOAA HMRs were created. Dynamical numerical weather models in particular have served to advance the state of knowledge regarding precipitation processes; such models solve the fundamental physical equations of atmosphere and thus generate precipitation according to the environment provided, offering continuity in space and time in the model output. These fundamental aspects of dynamical models could conceivably alleviate the need for many spatial, temporal, physical assumptions (e.g., storm transposition, storm templates, moisture maximization, etc.) used in the PMP estimation process. The availability of reliable, high-resolution model data in data-sparse regions of complex, high-elevation topography is another considerable potential benefit.

The state of Colorado has considered the use of high-resolution regional modeling in PMP estimation before. As part of the State of Colorado Department of Natural Resources Project "Development of New Methodologies for Determining Extreme Rainfall," Cotton et al. (2003) details the use of a convective-storm-resolving mesoscale model for six historical heavy precipitating cases over Colorado. A total of 27 simulations were performed, varying both model parameters and physical land-surface states, and synoptic patterns were moved relative to the underlying terrain. Maximum precipitation produced by all simulations was then statistically manipulated for various depth-area-duration (DAD) events (using Hershfield parameters and kriging techniques) to make PMP estimates. Conclusions of this work were that while the modeling served as a useful demonstration of concept, simulation errors in position, timing, and event evolution were too significant, and the data produced too limited, to provide useful quantitative PMP estimates.

For this study, many of problems facing the study detailed in the Cotton et al. (2003) report have benefited from significant advances in technology and the science of forecast modeling heavy rainfall. The main resource used by CO-NM REPS Task 3 is a 5-year archive of model forecasts from a state-of-the-art, hourly-updated, 3-km

horizontal grid spacing NWP system developed by the National Oceanic and Atmospheric Administration (NOAA). Of primary interest is whether this dynamical model-based framework offers enough benefit to be considered for future iterations of extreme precipitation estimation by the dam safety and flood risk management communities. Secondary interests for the demonstration carried out in the present CO-NM REPS project include whether or not the existing 5-year model dataset can provide useful information regarding the frequency, variability, and intensity of extreme precipitation and support existing PMP estimation methods and/or specific parts of the traditional PMP estimation process.

The NWP system used in this study, the High-Resolution Rapid Refresh (HRRR) model, is run operationally for the U.S. National Weather Service (NWS), and serves a large variety of user needs. Daily users of the HRRR forecasts for decision-making include NWS forecasters, severe weather forecasters, aviation strategic and tactical planners, quantitative precipitation forecasters, renewable energy planners and market traders, and fire weather forecasters. While current high-resolution numerical model data is limited by computation costs, in the future, it is conceivable that supercomputing resources will be large enough, and high-resolution NWP systems will be sufficiently inexpensive, to allow larger numbers of simulations that might better serve PMP and regional extreme precipitation estimation.

2. Data and Methods

2.1. Hourly Updating Numerical Weather Prediction System: The High-Resolution Rapid Refresh

The NOAA Earth System Research Laboratory (ESRL) Global Systems Division (GSD) has developed a 13-km horizontal grid spacing NWP model called the Rapid Refresh (RAP; Benjamin et al. 2016). The RAP is an hourly cycled system, run every hour. The RAP first became operational at the National Centers for Environmental Prediction (NCEP) on 1 May 2012 (RAPv1, run out to 18 forecast hours), with the most recent upgrade at NCEP on 23 August 2016 (RAPv3, run out to 21 hours). An additional upgrade (RAPv4) is currently planned for spring 2018. The RAP replaced the earlier Rapid Update Cycle (RUC; Benjamin et al. 2004) as NOAA's operational rapidly-updating forecast system. Beginning in 2010, a nested version of the RAP on a three-km horizontal grid has also been run hourly over the CONUS in experimental mode at GSD; this nested version of the RAP is called the High-Resolution Rapid Refresh (HRRR) (Smith et al. 2008). The HRRR became operational at NCEP on 30 Sep 2014 (HRRRv1, run hourly out to 15h), with an upgrade to the next version of the HRRR on 23 August 2016 (HRRRv2, run hourly out to 18 h). HRRRv3 implementation is planned for spring 2018.

The RAP and HRRR model domains are shown in Figure 1. Both systems use a community-supported data assimilation system, the Gridpoint Statistical Interpolation (GSI) package.

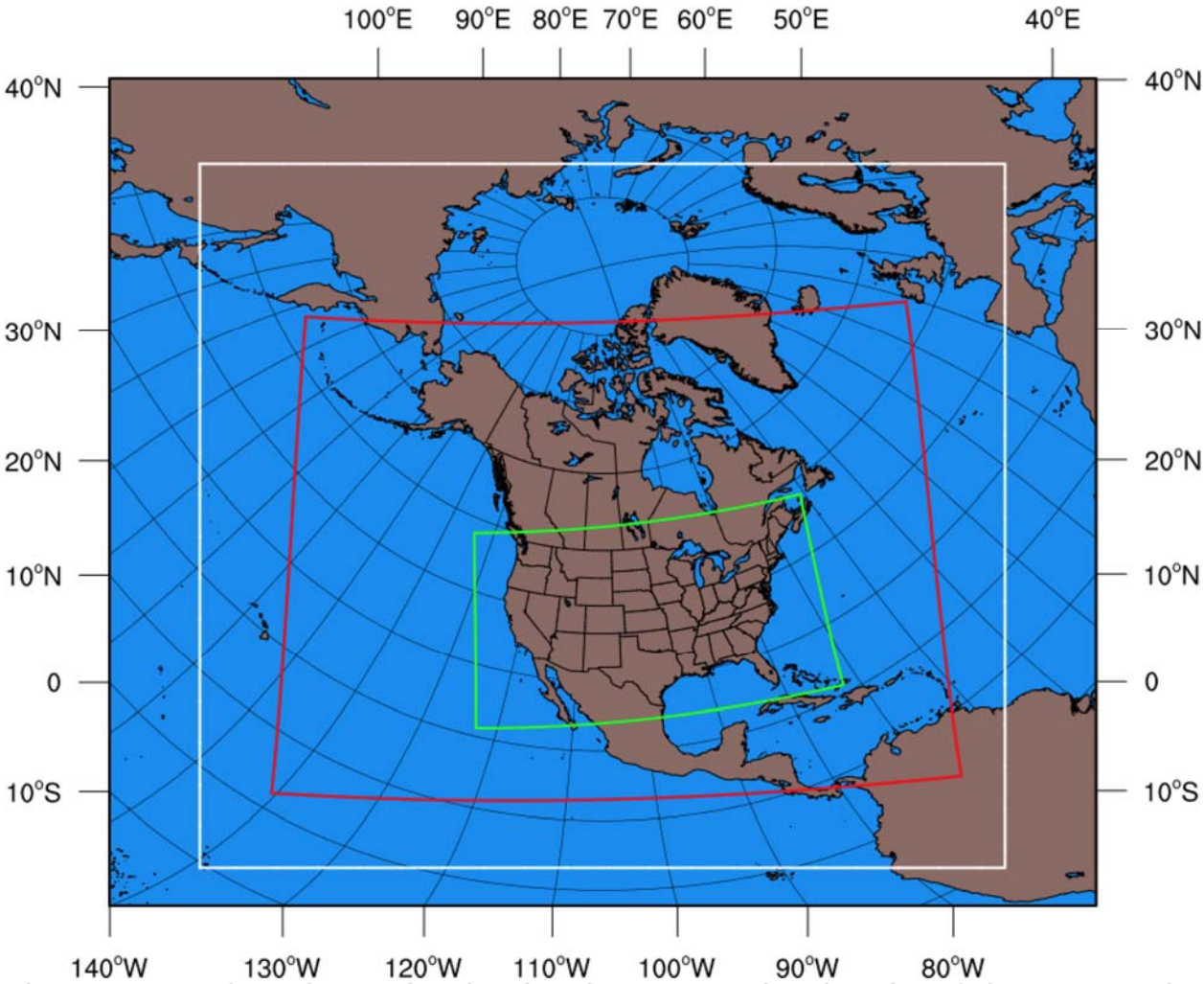


Figure 1. Map of North America showing the computation domains of the RAPv3 and RAPv4 (white), RAPv1 and RAPv2 (red), and HRRR (green) models.

The data assimilation scheme, which uses a hybrid three-dimensional ensemble/variational approach (Hu et al. 2017), uses observations from a variety of platforms to create an initial condition for the model forecast. At the end of the ensemble/variational data assimilation, a hydrometeor analysis (Benjamin et al. 2016) is carried out based primarily upon satellite and surface-based ceilometer observations. Within the RAP, radar reflectivity observations are brought in through a diabatic digital filter initialization (DDFI) procedure (Peckham et al. 2016); the diabatic portion of the filter introduces latent heating during the DFI period, with the magnitude of the latent heating related to observed radar reflectivity.

The HRRR assimilates radar reflectivity observations on its 3-km grid in a similar way, but with several important differences. First, the HRRR assimilates radar reflectivity observations over the hour leading up to the model initialization time through the execution of a one-hour “pre-forecast”. Secondly, while the HRRR uses a similarly-formulated relationship between radar reflectivity and latent heating, the latent heating varies with 15-minute reflectivity observations rather than being held

constant for the full hour. The ensemble/variational data assimilation step occurs after this one-hour "pre-forecast". This technique permits a much more realistic hydrometeor structure within the HRRR at the start of the full forecast.

The NWP component of the RAP and the HRRR is the community-supported Weather Research and Forecasting (WRF) model, in particularly the Advanced Research WRF (WRF-ARW; Skamarock et al. 2008). For more details on the RAP and HRRR configuration, including physical parameterizations used within WRF, the reader is referred to Benjamin et al. (2016).

After the initial implementation of the HRRR in 2014, forecasters noted biases in several aspects of HRRRv1 performance. Most obviously, the HRRRv1 featured a high bias in coverage of convective precipitation. It was also discovered that the HRRRv1 had a warm and dry bias in the daytime planetary boundary layer (PBL) over much of the domain, particularly in the eastern CONUS. It was hypothesized that these biases were linked by a mechanism described in the feedback cycle outlined in Fig. 10 of Benjamin et al. (2016). Insufficient cloud cover in the HRRRv1 was leading to overly deep mixing and too-deep PBLs, especially in the summertime, and an excess of incoming solar irradiance. This excessive low-level mixing tended to overcome convective inhibition too readily, producing false alarm convection in the model.

In order to alleviate the biases in the initial version of the operational HRRR, development took place particularly in the model physics parameterizations. Since the HRRR obtains its initial and boundary conditions from the RAP, development efforts within the RAP system also had an impact on HRRR forecast performance. One of the foremost changes implemented in HRRRv2 was an increase of the "wilting point" within cropland regions in the RUC LSM, effectively allowing continued transpiration from irrigated crops and increasing low-level relative humidity; the effects of this change were most pronounced over the agriculture-rich Great Plains of the US. Another major adjustment was allowing the RRTMG radiation scheme to interact with boundary layer clouds within the MYNN PBL scheme, having the net effect of increasing low-level cloudiness and reducing solar irradiance reaching the surface. A secondary low-level cooling effect comes from attenuation from climatological aerosol loading within the Thompson aerosol-aware microphysics scheme. These changes are described in more detail by Benjamin et al. (2016).

Within the HRRR DA, a number of changes were made to address the warm/dry bias in the HRRRv1. Hybrid ensemble-variational data assimilation, having been shown to greatly improve forecasts of upper-level wind and other variables within the 13km RAP (Hu et al. 2017), was implemented in the HRRR beginning with HRRRv2. Focusing more specifically on the model biases, PBL "pseudo-innovations" (Benjamin et al. 2016) were introduced for surface temperature (in addition to surface dewpoint) in order to extend the influence of surface observations in the vertical in well-mixed situations. In addition, assimilation of 2m temperature and dewpoint observations was modified to be more consistent (accounting for the difference in height between the typical 2m height of sensors and the lowest model level, near 8m AGL).

The net result of these DA and model physics modifications is shown in Figure 2. These statistics are derived from a month-long retrospective test during the period 15 July – 15 August 2014, and comparison is against the real-time HRRRX running at ESRL/GSD. Verification is against CONUS METAR observations (standard daily weather observations). Boxes on these curves represent one standard error (Weatherhead et al. 1998). As shown in Figure 2, 2m temperature RMSE is reduced by up to 40 percent during the afternoon and early evening hours, with less improvement during the late night and early morning hours. Mean 6-h forecast 2m temperature biases at 00 UTC are reduced from over +1.5 to about +0.3 degrees, with similarly large improvement earlier in the afternoon and later in the evening. Even more striking improvements are evident in the 2m dewpoint temperature verification: mean daytime dry biases of approximately -1.5 degrees are reduced to a moist bias of approximately +0.3 degrees. 10m wind speed biases are similarly reduced, particularly in the early evening period. Forecast improvements are much more muted during a month-long winter retrospective test (not shown).

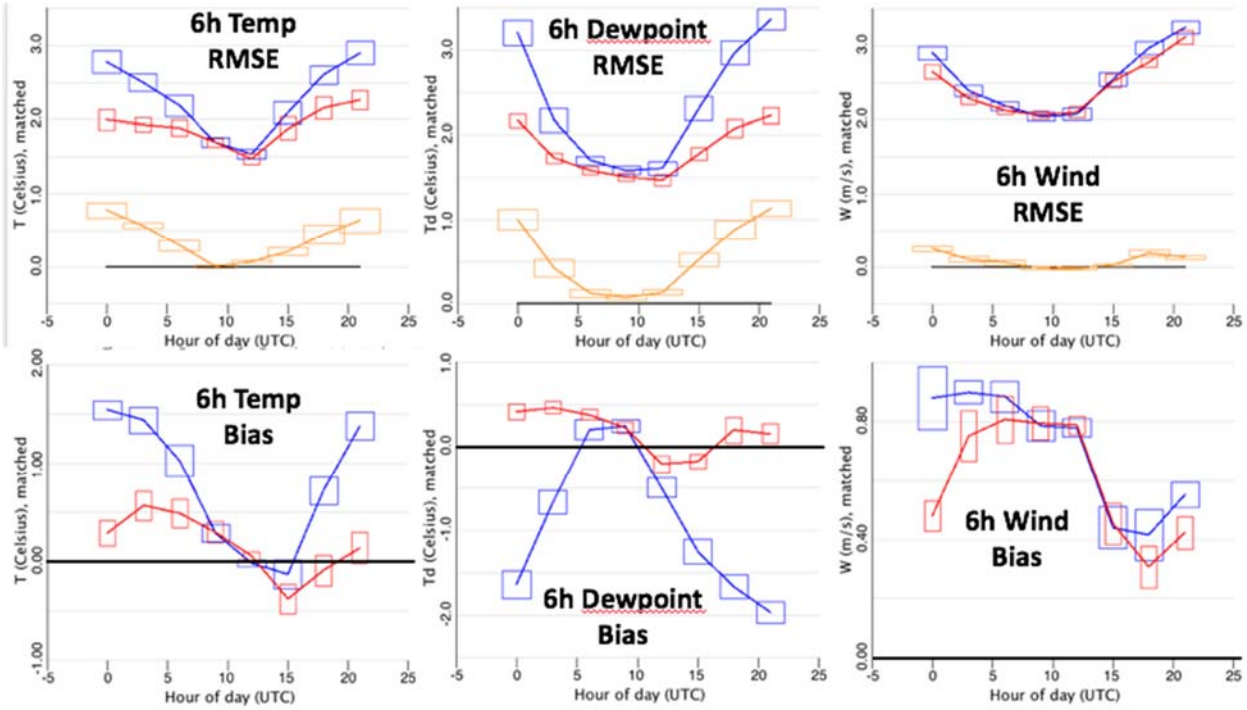


Figure 2. Verification of HRRRv1 vs. HRRRv2, against CONUS METAR observations, during the period 15 Jul - 15 Aug 2014. The blue curve indicates the HRRRv1, and the red curve indicates the HRRRv2, with one standard error indicated by the boxes.

Several cases of simulated radar reflectivity forecast differences between the HRRRv1 and HRRRv2 are shown in Figure 3. These cases are taken from the summer of 2015, when the HRRRv2 was running in the HRRR-X but the HRRRv1 was still running operationally. On 4 June 2015, the HRRR forecasted a large false alarm mesocale convective system (MCS) over central Kansas by 00 UTC 5 June, while the HRRR-X successfully forecasted only small, isolated convective cells in this region. Once

again, on 21 June 2015, the operational HRRR forecasted the development of a small MCS in Kansas, but the HRRR-X successfully forecasted the absence of storms over this region. On 12-13 July 2015, the HRRR-X produced a more realistically-organized MCS over Minnesota, with a northeast-southwest oriented leading convective line, while the operational HRRR forecasted a much more poorly organized system with extensive false alarm convection to the west of the MCS. These improvements were all implemented in the operational HRRR on 23 August 2016.

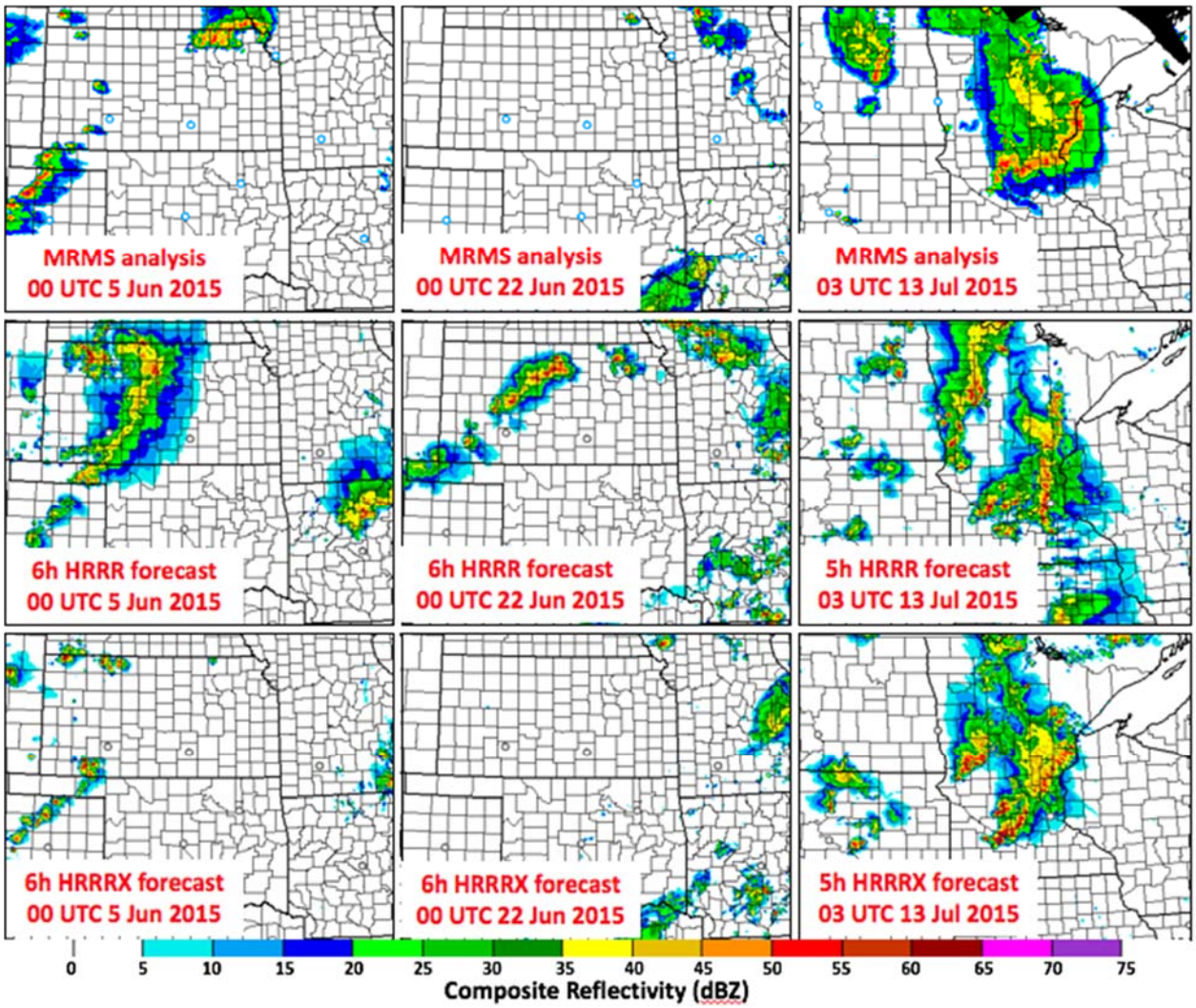


Figure 3. Simulated radar reflectivity forecasts from (bottom row) the experimental HRRRv2 and (middle row) the operational HRRRv1, compared against MRMS observations (top row) during 2015.

The second version of the HRRR exhibited a different set of biases. In particular, forecasters noted a continued tendency for the model to excessively quickly erode low-level cloud cover. A high bias in simulated radar reflectivity and precipitation in the first few hours of the HRRR forecasts was also noted. Due to these noted issues, the focus of the HRRRv3 upgrade is upon improved retention of low clouds, and

reducing a short-lead-time high precipitation / simulated radar reflectivity bias, as well as improved 2m temperature / dewpoint diurnal cycles in summertime.

Data assimilation changes in the HRRRv3 were motivated by observed short-range forecast biases present in the HRRRv2. In particular, the high bias in precipitation during the first few hours of the forecast motivated a reduction in the strength of the latent heating applied in the RAP DFI within regions of high observed three-dimensional radar reflectivity. Figure 4 shows simulated composite radar reflectivity forecasts as verified against the multi-radar, multi-sensor (MRMS) analysis product, for a month-long retrospective test during July 2016, compared against the real-time HRRRX (HRRRv2). As can be seen in Figure 4, CSI for both the 25 and 35 dBZ thresholds are increased during the first 6-8 h of the forecast, while the bias in both thresholds is drastically reduced.

Focusing on the cloud retention problem, the RAP and HRRR cloud analysis was modified to specify a high value of cloud water and cloud ice droplet number concentration (0.00001 g/kg for water, and 0.001 g/kg for ice), thereby ensuring clouds are initialized with small droplets and reducing fallout in the first few hours of the model forecast. This led to a substantial improvement in the retention of low-level clouds, particularly in the winter period (Figure 5 shows a comparison of RAP forecasts with and without this change; the same code is applied in HRRR). An additional change, arising from observation sensitivity experiments described by James and Benjamin (2017), is the consistent treatment of cloud building based on both surface METAR ceilometer data and satellite observations, limiting the height of cloud building from these observations to be below 1200 m AGL; this helps to reduce a high relative humidity bias in short-range forecasts in the low- to mid-levels (e.g., see Fig. 18 of James and Benjamin 2017).

In the realm of model physics, significant updates were made to several model physics schemes. Within the RUC LSM, a new mosaic approach to snow cover was adopted, more accurately accounting for patchy, thin snow cover, and improving low-level temperature forecasts in certain situations. In addition, real-time vegetation greenness fraction, derived from polar-orbiting satellites carrying the Visible Infrared Imaging Radiometer Suite (VIIRS) instrument, was incorporated into the RUC LSM, accounting for anomalous departures from the normal seasonal cycle of vegetation greenness. The use of VIIRS data paves the way for the use of other VIIRS products within the real-time RAP/HRRR NWP suite.

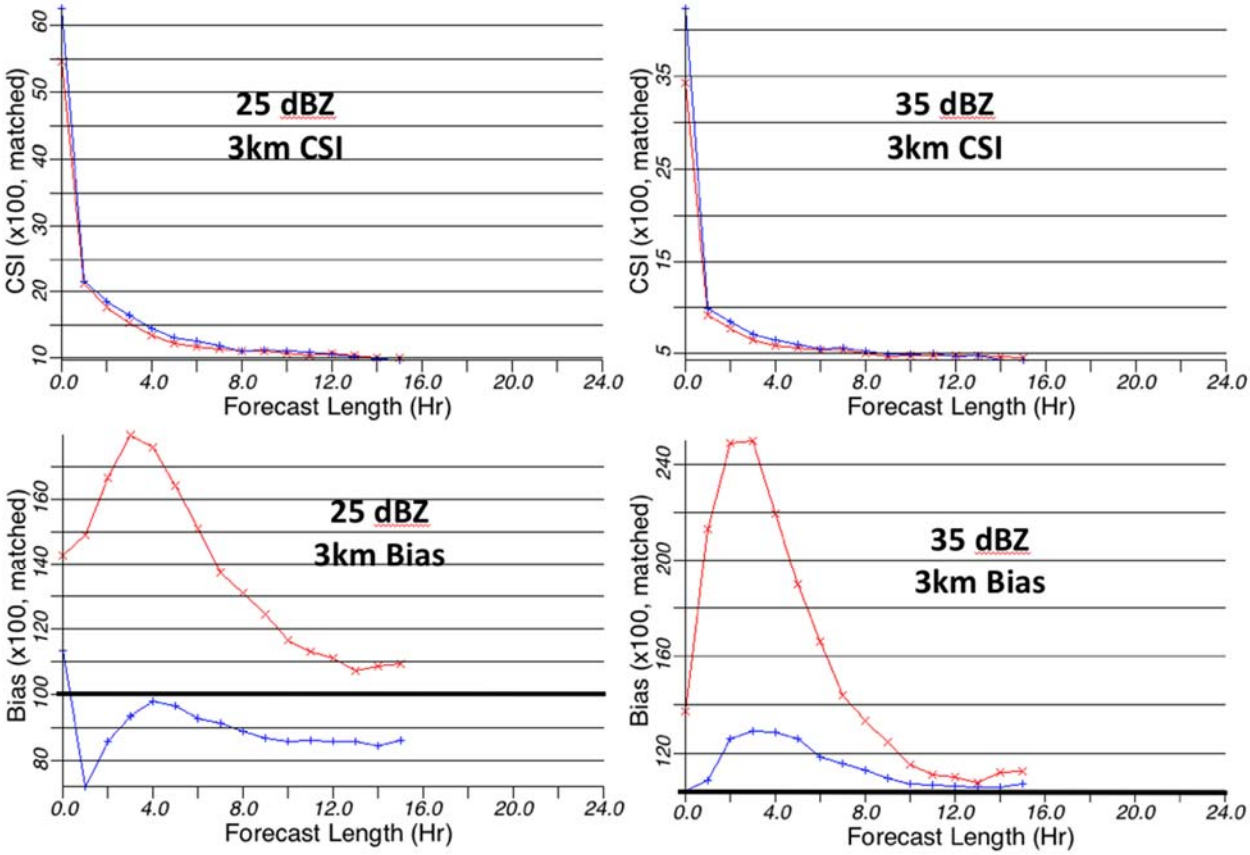


Figure 4. Composite radar reflectivity forecasts from the HRRRv2 (red) and HRRRv3 (blue) verified against MRMS observations. Results are from a month-long retrospective HRRRv2 test during July 2016, compared against the real-time HRRRX. Shown are CSI (top row) and bias (bottom row), for the 25 dBZ threshold (left) and the 35 dBZ threshold (right).

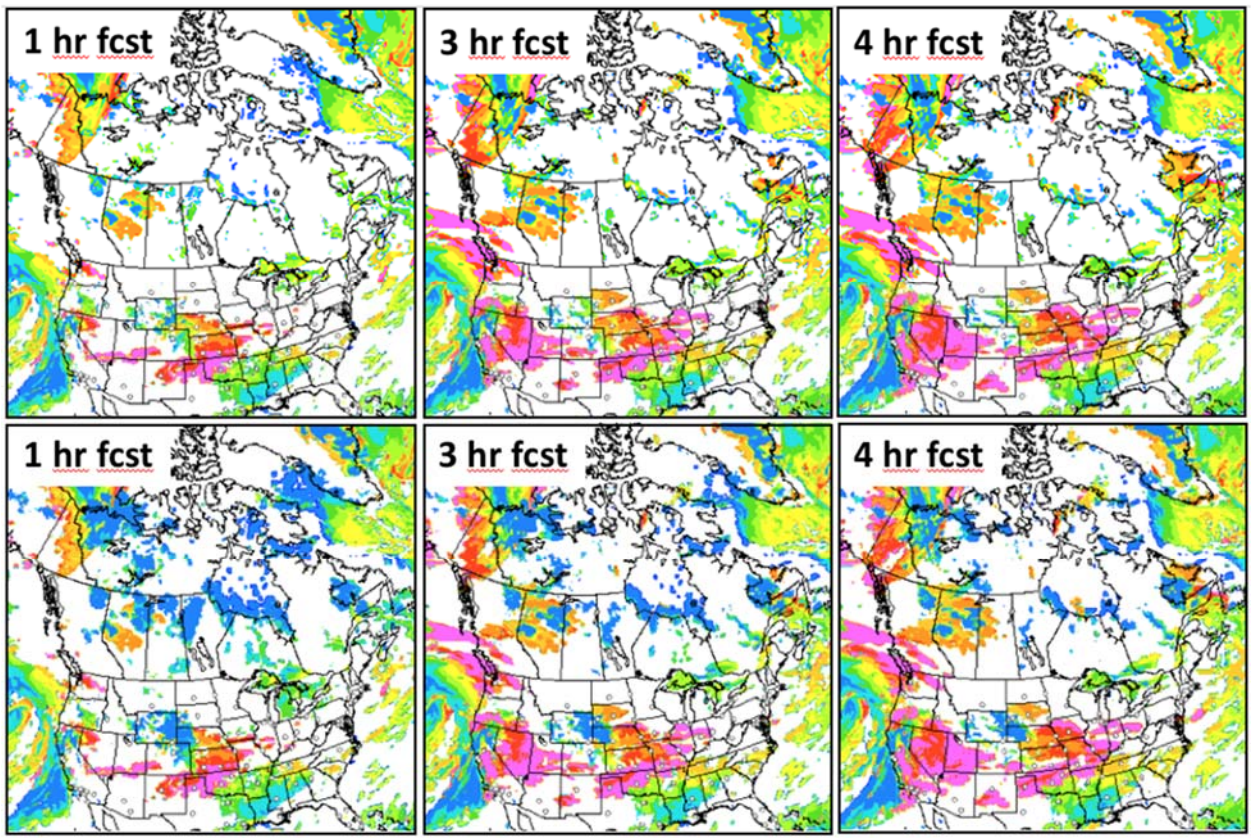


Figure 5. Cloud ceiling forecasts from the RAP, without (top row) and with (bottom row) specifying a high value of cloud water and cloud ice number concentration in the cloud analysis.

Improvements were made to several other key parameterization schemes within the HRRR for the implementation of the HRRRv3, including the Thompson microphysics scheme and the MYNN PBL scheme. An improved representation of subgrid-scale clouds is now used in MYNN, and an eddy-diffusivity mass flux scheme is employed to improve the parameterization of mixing under different stability scenarios. Many of the improvements to the MYNN PBL scheme stemmed from extensive evaluation and testing during a field campaign in the Columbia River Gorge (the Second Wind Forecast Improvement Project WFIP2; reference?). Finally, a new hybrid vertical coordinate was implemented within the RAP and the HRRR, reducing the amount of terrain-related numerical noise in the model as compared against a terrain-following coordinate (Klemp 2011).

Figure 6 and Figure 7 present verification of HRRR forecasts against METARs over the CONUS during two month-long retrospective tests of the HRRRv3 code, as compared against the HRRRv2. Improvements during the summertime (July 2016; Figure 6) are muted, but statistically significant for most variables during the morning hours. More substantial gains are seen in the winter retrospective test (January 2017; Figure 7).

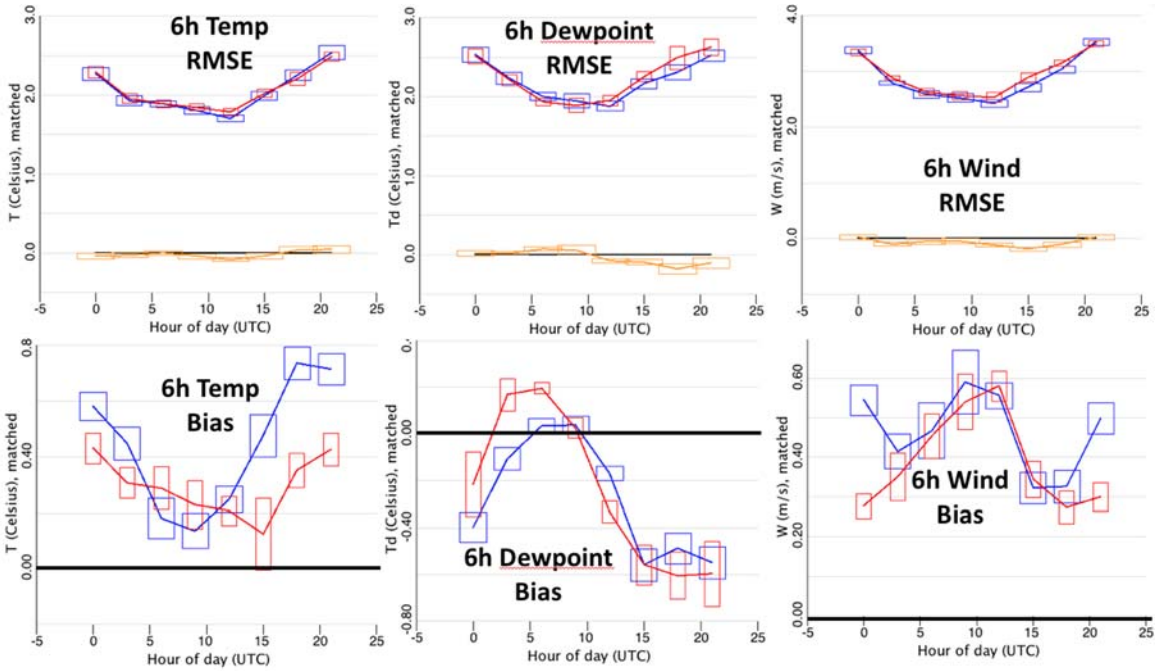


Figure 6. Verification of HRRRv2 vs. HRRRv3, against CONUS METAR observations, during the period 1 - 31 Jul 2016. The blue curve indicates the HRRRv3, and the red curve indicates the HRRRv2, with one standard error indicated by the boxes.

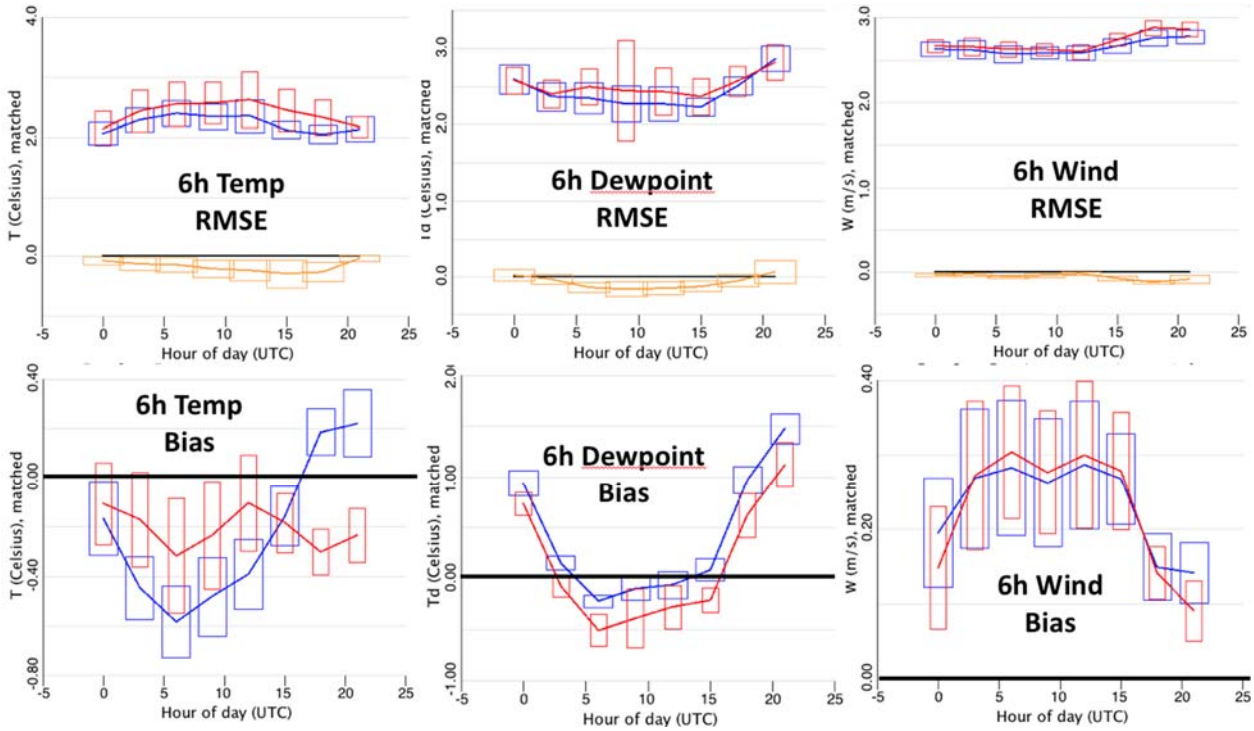


Figure 7. Verification of HRRRv2 vs. HRRRv3, against CONUS METAR observations, during the period 1- 31 Jan 2017. The blue curve indicates the HRRRv3, and the red curve indicates the HRRRv2, with one standard error indicated by the boxes.

2.2. HRRR precipitation forecasts and their use in precipitation estimation

NOAA-ESRL has performed extensive verification of HRRR precipitation forecasts with respect to the National Centers for Environmental Prediction (NCEP) Stage-IV precipitation analysis, hereafter "Stage-IV" (Nelson et al. 2016). Stage-IV is considered the "official" quantitative precipitation estimate (QPE) for validation of NOAA numerical weather prediction models. Stage-IV is produced for the contiguous United States at ~4.7-km grid spacing and accumulation periods of 1, 6 and 24 h. Forecasters at 12 River Forecast Centers (RFCs) combine precipitation estimates from NWS doppler radar, manually quality controlled point observations, and a climatological precipitation analysis to produce a best estimate of precipitation over their respective areas of responsibility. The final Stage-IV analysis is a mosaic of these 12 manually augmented grids. The use and weighting of various precipitation datasets, and the rigor with which quality control measures are applied to observations, all vary from one RFC to another, and even from one forecaster to another within an RFC. Due to inherent shortcomings with the Stage-IV methodology in complex terrain of the western US, as discussed later in this section, general conclusions about HRRR precipitation forecast performance will be based on comparison to Stage-IV east of 105 degrees west longitude.

Numerical weather prediction models, including the HRRR, provide QPF in real-time, and these forecasts can be compared against the various QPE products, including Stage IV. The HRRR is used extensively in operational QPF, particularly since its operational implementation at NCEP in Sep 2014. HRRR QPF has been evaluated in an increasing number of peer-reviewed studies (e.g., Ikeda et al. 2013, Pinto et al. 2015, Bytheway et al. 2017). As described in the previous section, a second version of the HRRR was implemented operationally in August 2016, with a third version slated for implementation in May 2018; here we provide a brief overview of QPF performance by the HRRRv3. As mentioned above, comparison is performed with Stage-IV over the eastern US, where the QPE dataset is more trusted. Figure 8 shows the variation in forecast skill with forecast lead time from HRRR forecasts during the warm season. The HRRR exhibits a near-neutral frequency bias for light to moderate precipitation amounts, and some overforecasting of heavy precipitation, particularly for thresholds exceeding 25 mm per 6 h (Figure 8). This high frequency bias (i.e., exceedances of a threshold are forecast more often than they occur in the Stage-IV analysis) is most evident at lead times of less than 6 h, and moderates to a near neutral bias by forecast hour 12. Spatial and temporal forecast errors steadily increase with increasing lead time, although subjective examination of individual cases suggests these displacement errors are more random than systematic, particularly for intense convective events. Considering these factors, it was chosen to use HRRR forecasts for lead times of 6-12 h to estimate heavy and extreme precipitation occurrence.

The spatial distribution of HRRR 6-7h forecast biases for light precipitation compared against Stage-IV during the four summers of 2014-2017 is shown in Figure 9. The evolution of the precipitation biases in the eastern US during the four summers of

2014-2017 reveals a general moist bias in 2015 and 2016, with a closer to optimal bias evident in 2017. For a heavier precipitation threshold (0.25 inch in 6 hours; Figure 10), additional regional structure is seen in the biases when the HRRR forecasts are compared against Stage-IV. In particular, a moist bias in the southeastern US and southern plains present in 2014-2016 is greatly reduced in 2017, while the HRRR continues to forecast the exceedance of this threshold less frequently than the Stage-IV over the northern high plains. Similar maps can be created for higher precipitation thresholds which are more relevant for this project; however, these maps become extremely noisy due to the relative rarity of heavy precipitation events at any single model gridpoint.

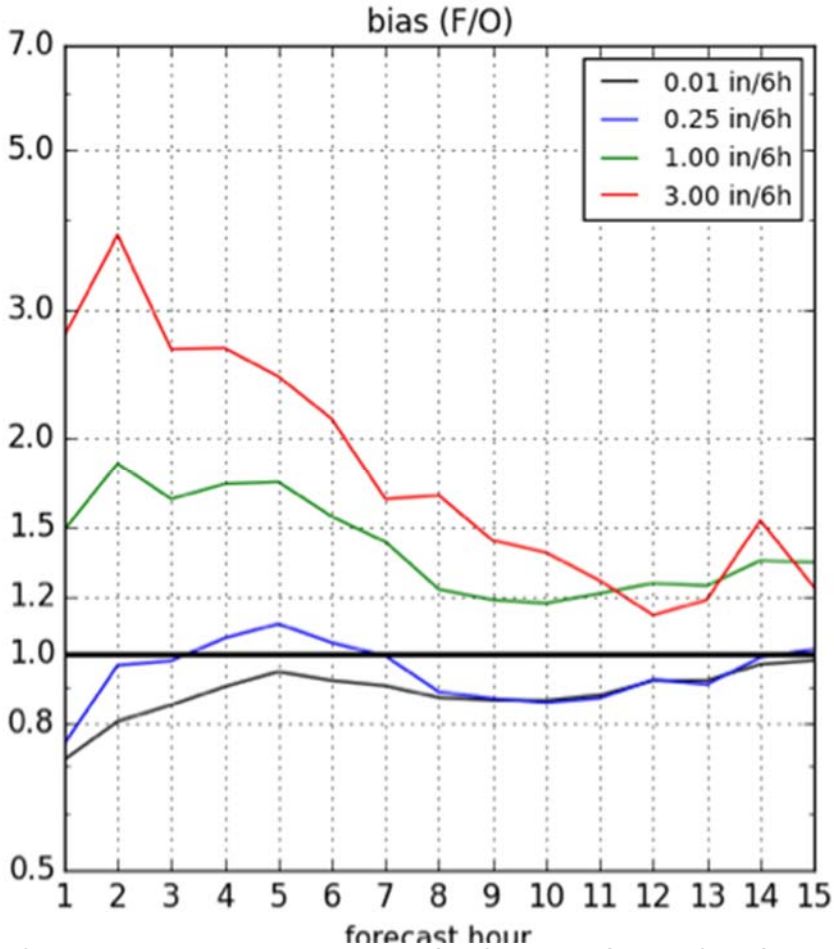


Figure 8. Bias of HRRR QPF vs. Stage-IV QPE by forecast hour, for the year 2016. Four different precipitation thresholds are shown: 0.01 inch / 6 h (black), 0.25 inch / 6 h (blue), 1 inch / 6 h (green), and 3 inch / 6 h (red).

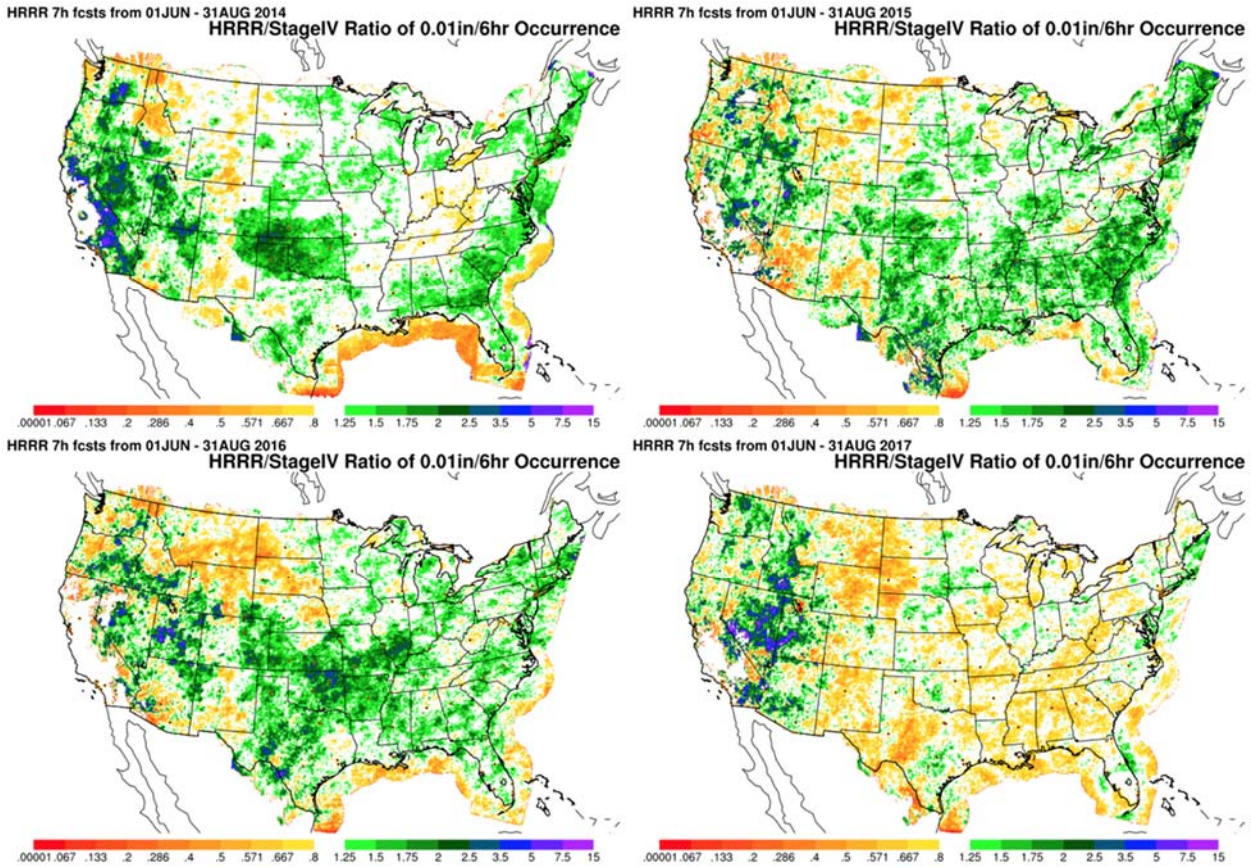


Figure 9. Ratio of frequency of 0.01 inch in 6 hours occurrence in the HRRR (lead time 6-7h) compared to the Stage-IV for (top left) summer 2014, (top right) summer 2015, (bottom left) summer 2016, and (bottom right) summer 2017. Yellow/orange shading indicates the HRRR forecasts light precipitation less frequently than it is present in Stage-IV analyses, and green/blue shading indicates the HRRR forecasts light precipitation more frequently than it is present in Stage-IV analyses.

HRRR performance for individual precipitation events, which is critical for NWS forecasters and other users, has been extensively evaluated against Stage-IV in regions where the QPE product is relatively more trusted. A recent catastrophic flooding event, Hurricane Harvey in August 2017, provided an excellent opportunity to evaluate HRRR forecasts of an extreme precipitation event. Figure 11 shows a comparison of HRRR 48-h QPF versus Stage-IV for the same period (the Experimental HRRR extends out to 48-h forecast length twice a day).

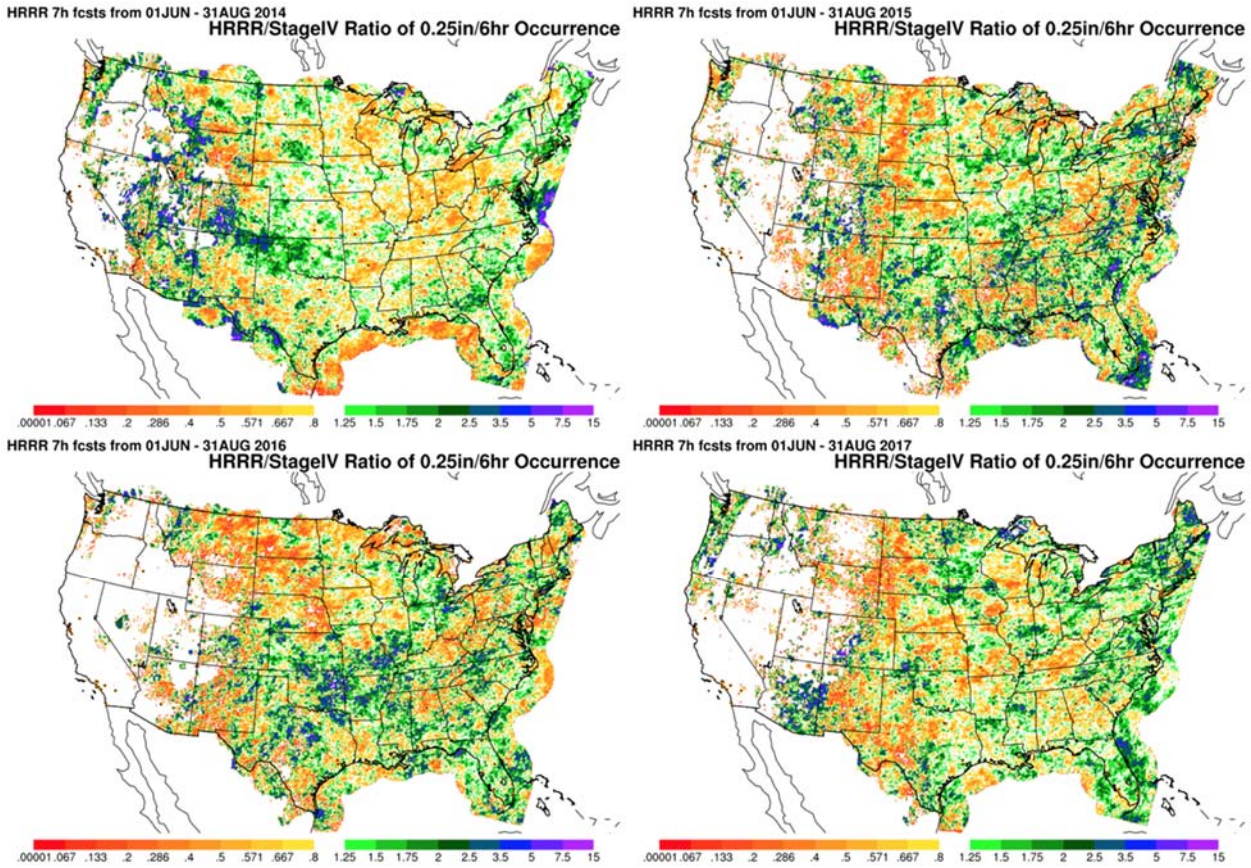


Figure 10. As in Figure 9, but for 0.25 inch in 6 hours.

As can be seen in Figure 11, the HRRR successfully forecasted the occurrence of greater than 20 inches of rain in 48 h in extreme southeastern Texas in association with this event, with excellent agreement in the location, shape, and orientation of the region of heaviest rainfall with the Stage-IV QPE. A heavy rain event in more complex terrain is shown in Figure 12, from the Sep 2013 flood along the east slope of the Colorado Front Range. As is seen in Figure 12, there is a large amount of run-to-run variability in the HRRR performance for this case. However, a key advantage of the hourly-updating HRRR is that there are multiple opportunities for the model to successfully forecast the event. This is particularly advantageous in the context of this study, where the goal is precipitation estimation instead of precipitation forecasting, and all forecast hours of all forecast cycles can be considered as physically-bounded potential precipitation scenarios.

The focus of this study is on “probable maximum precipitation”, or, in a probabilistic framework, extreme precipitation frequency. When estimating precipitation frequency over the complex terrain of the western United States, there are a number of advantages to using HRRR model forecasts as an alternative, or to augment point observations and precipitation analysis products. The spatial coverage of point observations in CO and NM is modest at best near urban centers, and poor over high terrain, and rural desert and plains regions. Some automated sites report

precipitation accumulation hourly, but most manual observations are taken at 24-h intervals ending at times that are inconsistent from one site to another.

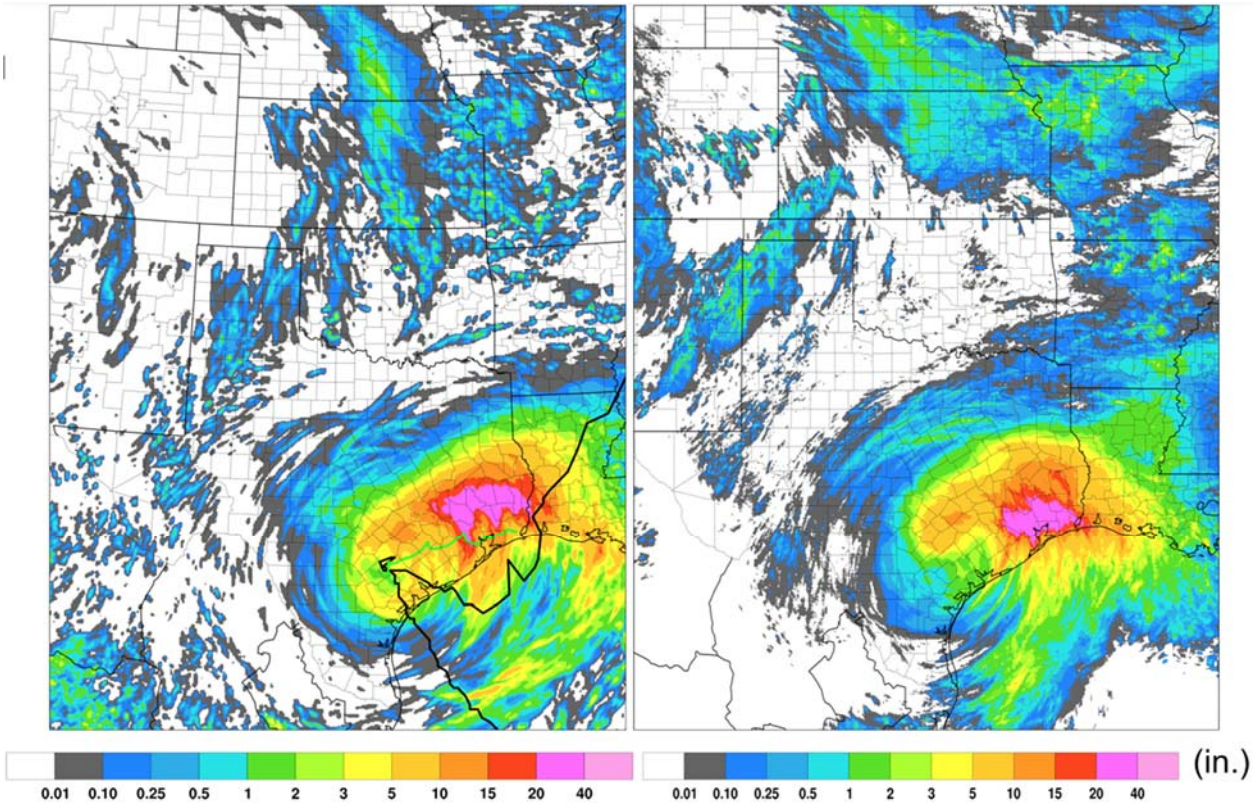


Figure 11. 48-h precipitation from 12 UTC 26 Aug - 12 UTC 28 Aug 2017, as captured by (left) the Experimental HRRR QPF initialized at 12 UTC on 26 and 27 Aug 2017, and (right) MRMS radar-only QPE for the same time period. The black line is the observed storm track of Hurricane Harvey, and the green line is the HRRR forecasted track.

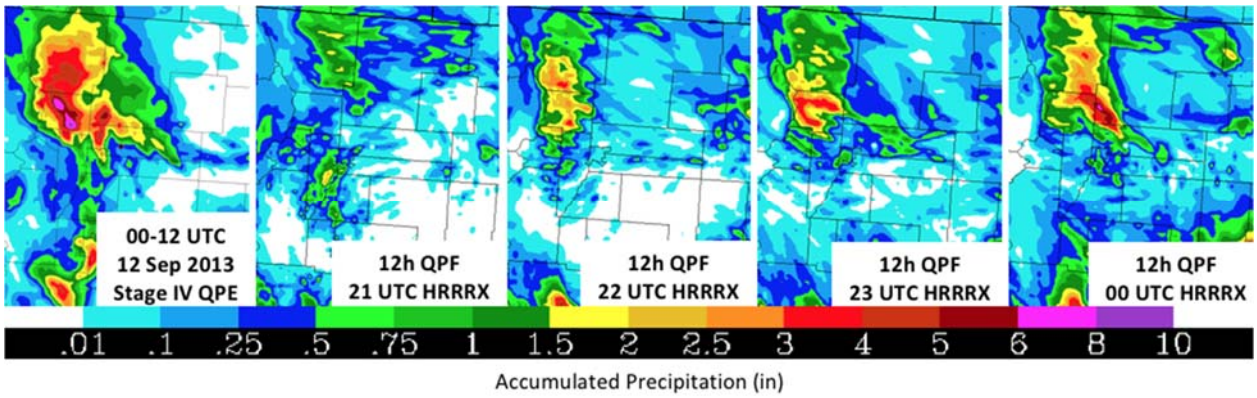


Figure 12. 12-h precipitation from 00-12 UTC 12 Sep 2013 in northern Colorado (black lines are county lines), as captured by (left) the Stage-IV QPE, and (right) four different initialization times of the Experimental HRRR (21, 22, 23, and 00 UTC from left to right).

HRRR provides precipitation forecasts on a 3x3-km grid, at intervals from 1 to 15 h (successive forecasts can also be combined to produce longer accumulation intervals, and although unavailable for use in this project, 15-minute precipitation data is also produced in real time). Despite some smoothing of steep terrain to maintain numerical stability in HRRR, forecast grids include elevations as high as 4300 m over the Colorado Front Range.

Although Stage-IV is considered an industry-standard precipitation analysis-of-record in the National Weather Service, there are a number of deficiencies in this dataset that can be mitigated or completely avoided through use of HRRR forecasts. Horizontal grid spacing in Stage-IV tails behind that of several operational numerical weather prediction models, including HRRR. Grid spacing in HRRR is 3 km, versus 4.7 km in the Stage-IV, thus use of HRRR yields a ~2.5 times increase in the number of grid points over a given area. Perhaps the largest disadvantage of Stage-IV is spatial inconsistency, which includes inconsistency in the availability, weighting and manual quality control of input datasets. The initial background for Stage-IV is typically a combination of radar-derived quantitative precipitation estimates, inverse-distance weighted analyses of point observations, and long-term climatological precipitation analyses. Radar coverage is poor over many areas of CO and NM, particularly in the San Juan Mountains of southwest CO, where the lowest tilt of the nearest radar site (Grand Junction) is centered more than 3000 m above ground level. Quality control of point observation data, although in many cases a straightforward endeavor, is a subjective process with no explicit agency-wide guidelines. Finally, the use of long-term, mean climatological precipitation analyses (e.g., the Parameter-elevation Relationships on Independent Slopes Model, or PRISM) to improve precipitation representation in complex terrain is subject to large error when all or part of a particular precipitation event exhibits an atypical precipitation-altitude relationship.

Not only are all of the Stage-IV input datasets subject to uncertainty and methodological shortcomings, but there exist differences of opinion between River Forecast Centers (RFCs) and individual forecasters regarding their use and appropriate weighting. For example, one RFC with an area of responsibility in the study region has a local policy of ignoring radar precipitation estimates during an arbitrarily defined winter season -- a policy that is not in place at neighboring offices. The resulting mosaic of quantitative precipitation estimates from multiple RFCs contains unphysical, near-zero-order discontinuities along RFC boundaries (Figure 13). HRRR forecasts contain no such discontinuities and are only subject to errors/biases in the physical parameterization of precipitation. While these biases are known to exist (e.g., Figure 8), recent subjective and objective validation of model forecasts and precipitation analyses by the National Weather Service indicates that HRRR forecasts over complex western US terrain perform nearly as well as Stage-IV (and better than some alternative precipitation analyses) relative to point observations (Figure 14).

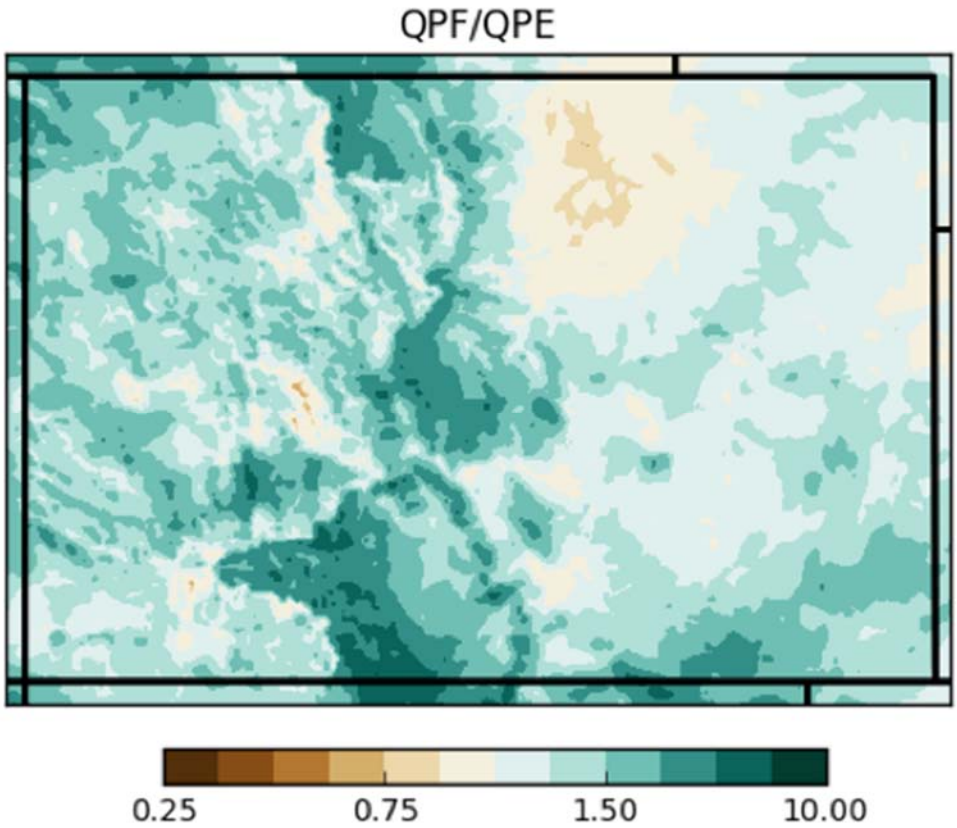


Figure 13. Ratio of HRRR 6-12-h QPF to Stage-IV QPE during the cool season 2012-2017 over Colorado. Note the major discontinuity following the continental divide in far northern and far southern Colorado.

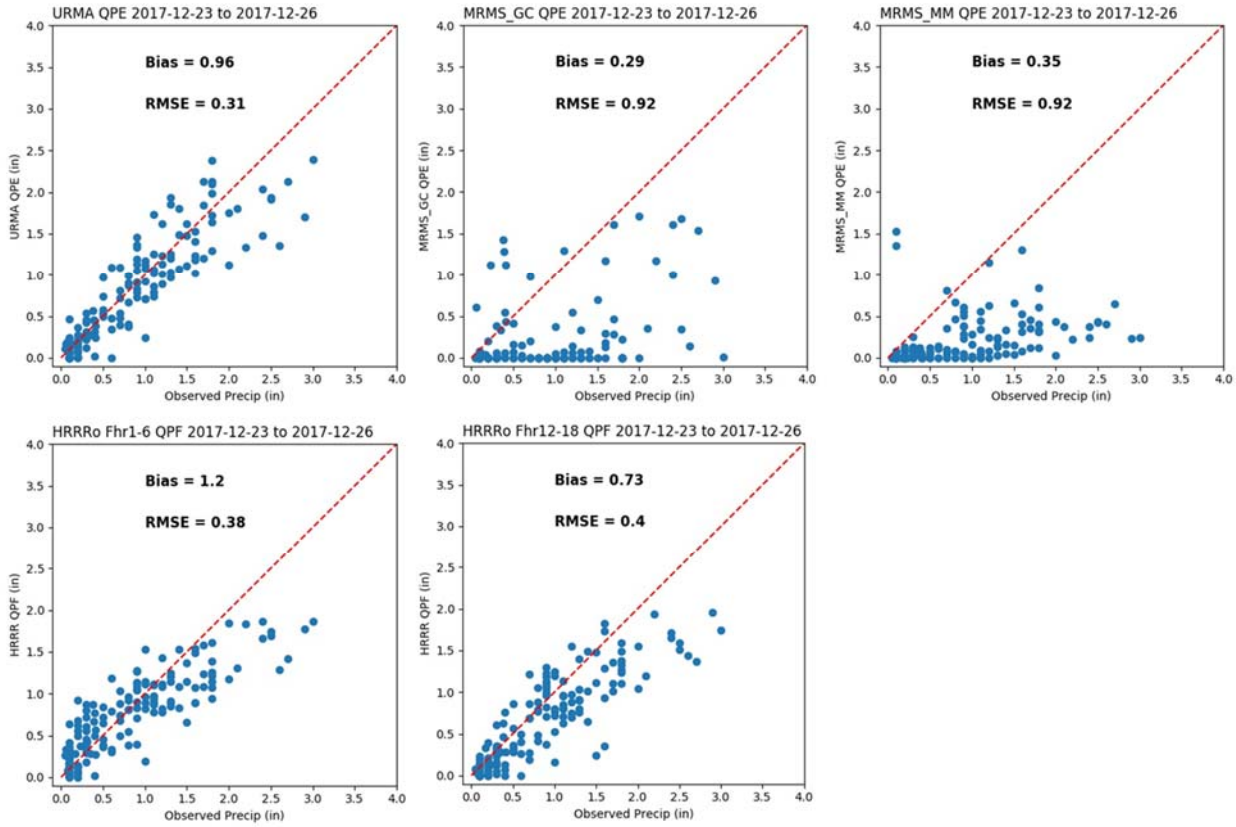


Figure 14. Scatter diagrams for (a) Stage-IV, here interpolated to 2.5-km resolution for the real-time "URMA" analysis product; (b) and (c) two types of Multi-Radar Multi-Sensor precipitation analyses; (d) HRRR 0-6-h forecasts; and (e) HRRR 12-18-h forecasts, relative to point observations for the 3-day period ending 1200 UTC 26-Dec 2017. Diagonal, dashed red line indicates a perfect forecast or analysis. Image courtesy NWS Western Region Headquarters.

2.3. Analysis of historical HRRR model data for REPS project

NOAA-ESRL maintains a local archive of several two-dimensional HRRR forecast grids, including accumulated precipitation, that begins in early 2012 and continues to the present day. For this study, we analyzed forecasts for a 5-year period from 24 January 2012 to 23 January 2017. The HRRR model data archive contains forecasts from the experimental HRRR configuration, or "HRRRX". The HRRRX is subject to frequent model changes, with model improvements occurring on a more continual basis, as opposed to the three discrete operational implementations of the HRRRv1, v2, and v3. The HRRRX is also subject to frequent irregular outages, with the dataset being about 75 percent complete. The completion rate is less for longer lead times, and for forecasts combined to derive longer precipitation accumulation intervals. Full 15h forecasts are initialized every hour when the data are available.

As is shown in Figure 15, bias characteristics vary with season and year. However, changes due to HRRR version changes appear to be of the same magnitude as seasonal variations in the bias. Differences between 6-h and 12-h bias characteristics can be

attributed to the influence of model data assimilation at the initialization time (having a stronger influence on the forecasts at 6 h than at 12 h).

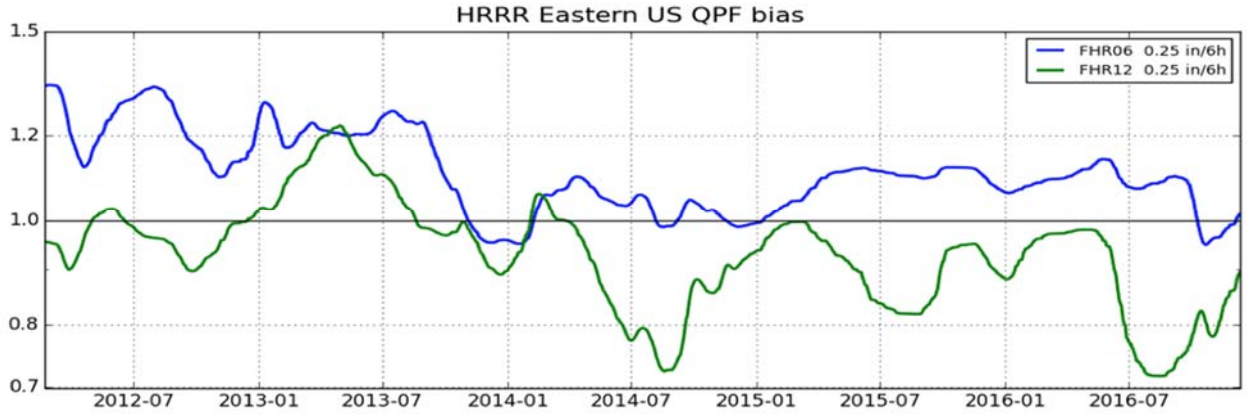


Figure 15. HRRR QPF eastern US bias as compared against Stage-IV from Jan 2012 - Jan 2017, for 0.25 inches in 6 h. The blue and green curves indicate 0-6-h and 6-12-h bias, respectively.

Evaluation of HRRR precipitation forecasts relative to Stage-IV precipitation analyses was performed using time-matched HRRR and Stage-IV grids for a subset of this period. Due to frequent data outages, the time-matched joint dataset covers only a small subset of possible forecasts. However, no systematic missing times of day or times of year were noted, and thus despite a reduced sample size, we consider this dataset representative of the full 5-year period. Although we note a number of issues with the Stage-IV dataset, our conclusions about HRRR performance relative to a Stage-IV "truth" are based on results for the eastern United States, where radar coverage is more uniform, point observations are more frequent in space and time, and thus fewer unphysical discontinuities appear in the Stage-IV grids. Long-term and case study examination of HRRR forecasts by external groups (e.g., the University of Utah) indicate that our conclusions about HRRR forecasts relative to Stage-IV analysis in the eastern United States are also broadly representative of HRRR performance in the West.

2.4. Using HRRR model data and model outputs for extreme precipitation study estimates

In this section, we provide an overview of the datasets used in this analysis, followed by a description of the methodology used to investigate the "probable maximum precipitation" question. Finally, we present our results from this study.

2.4.1 Data Descriptions and Sources

The primary dataset used in this study is the 5-year archive of HRRRX forecasts. This model dataset is currently housed on external hard drives attached to a desktop computer within NOAA/GSD, and thus public access to the HRRRX data is challenging. To this point, data requests have been handled on a case-by-case basis, but limited manpower precludes provision of model data in some cases. Public availability of

archived HRRR forecasts remains a problem for NOAA as a whole, with no publicly available archive of the operational HRRR yet hosted by NCEP due to the large file size (particularly for the 3D files) in conjunction with the hourly update cycle. An early attempt at an archive of HRRR and HRRRX forecasts has been undertaken by a small team at the University of Utah (Blaylock et al. 2017), but questions remain regarding the sustainability of the effort and quantity of data that could be hosted.

The other major dataset used within this study is the Stage-IV QPE, which is publicly available. For this project, we analyzed the Stage-IV for the same period as the HRRRX record. The Climate Prediction Center (CPC) 24-h precipitation product was also briefly examined in order to assess the representativeness of the 5-year period of study to the actual 30-year climate.

2.4.2 Methodology

For the analysis results presented in this report, we took several complementary approaches. Our primary effort was mining the HRRRX model data archive to obtain QPF from every hourly run over the 5-year period of record. A probability distribution function (PDF) of QPF for different time durations can thus be derived for each model grid point. One of our main goals within the project was to communicate the value of this derived 3-km spatial structure for the overall project goals towards the sponsors and the other tasks, and to explore ways of presenting the model data and deriving quantities that can be compared with values understood by the other two tasks.

The construction of the QPF "PDF" at each 3-km model grid point essentially represents a "poor-mans" ensemble. Since the PDF is constructed by considering HRRRX simulations initialized every hour, the model dataset contains multiple realizations of each event. Each of these realizations (varying by forecast lead time) comprise one possible outcome of the meteorological situation, similarly to the way a forecast ensemble represents a discrete number of possible outcomes.

We acknowledge that a 5-year period of record is much too short a period to expect to capture precipitation extremes at the time ranges forecasted by the HRRR model (i.e., 1 h to 15h). However, since lighter precipitation events are relatively common, the model dataset still contains valuable information about the spatial variation of the frequency of light precipitation as compared to the frequency of heavier precipitation. We might expect the PDF at each model grid point to be more reliable and representative of the longer-term climatology at lighter (i.e., more common) precipitation thresholds.

A secondary component of the project was to provide derived information from the HRRRX QPF dataset which could directly inform decisions regarding the validity of assumptions playing a role in the decision-making process for the other two tasks. This aspect of our effort was largely dictated by the interest of the other two tasks. We insisted upon first understanding (at a basic level) the intended uses of the HRRR information, and ensuring that this usage was scientifically reasonable.

3. Analysis and Findings

In this section, we present the results of our analysis of the 5-year HRRRX QPF dataset.

3.1. Raw Maximum Precipitation

Our first task within the study was to examine the maximum precipitation predicted within the HRRRX during the 5-year period of record at each grid point. Given the short period of record, it is not anticipated that the maximum precipitation occurring for a given threshold represents anywhere close to the "theoretical maximum" that PMP seeks to obtain, but it was an illustrative exercise. Figure 16 shows the 5-year maximum from the HRRRX dataset. While substantial small-scale noise is present within the model dataset, significant regional variations are also present.

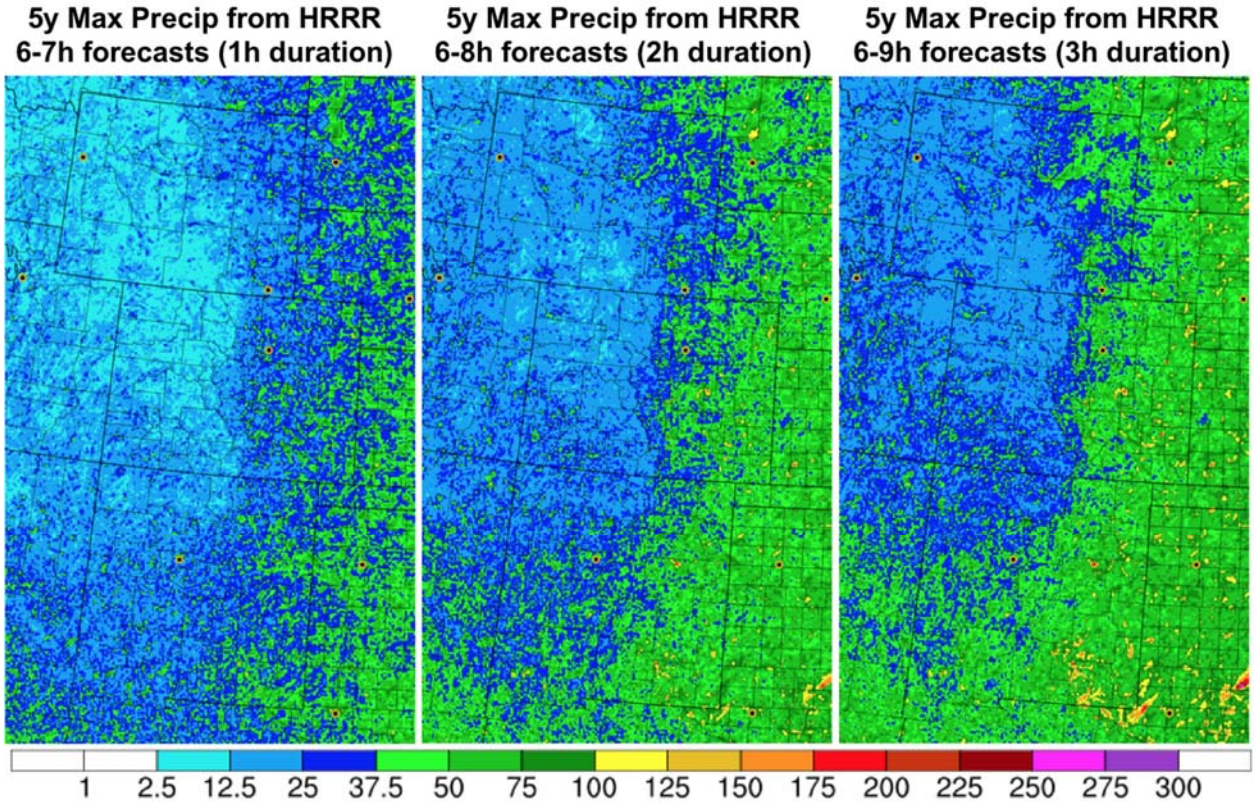


Figure 16. Maximum precipitation (mm) within the 5-year HRRRX model dataset occurring for (left) 1 h, (center) 2 h, and (right) 3 h durations.

As we would expect, higher maximum precipitation totals occur in eastern Colorado, and in eastern and southern New Mexico, where elevations are generally lower and moisture availability is greater. Five-year maximum, 1-h precipitation reaches around 30-50 mm over much of the eastern plains of Colorado and New Mexico, but closer to 10-15 mm in much of western Colorado. Maximum precipitation increases with longer time duration, as expected.

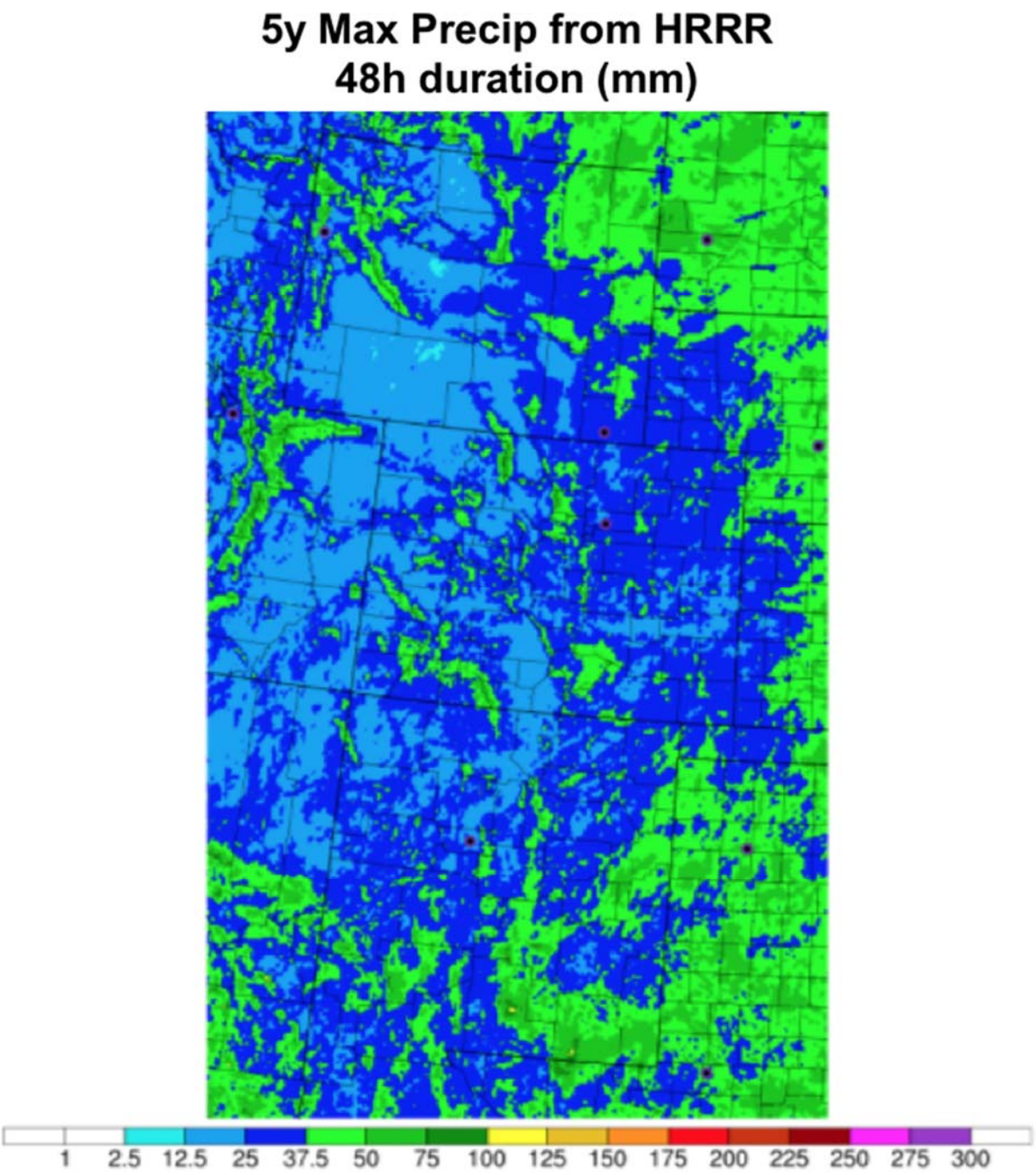


Figure 17. Maximum 48-h precipitation within the 5-year HRRRX dataset. See text for details on this calculation.

Calculations can also be performed for other time durations, although constructing maxima for durations longer than the duration of a single HRRR forecast becomes more complicated. The other tasks within the project requested maxima for the 2-day duration; Figure 17 shows this result from the HRRRX model dataset. This quantity was derived by summing up HRRRX forecasts initialized 6-h apart (at 00, 06, 12, and 18 UTC), for a total of 48 h duration. Due to the frequency of missing HRRRX forecasts, and the compounding problem of requiring 48 h of consecutive 6-h HRRR

forecasts, the record is only ~30 percent complete, thus limiting the utility of this calculation.

Given the small-scale noise evident in the 5-year maximum grid (due to the small sample size), we also investigated calculating other percentiles from the PDFs at each gridpoint in order to more accurately capture spatial patterns. Figure 18 shows the spatial smoothing that can be obtained by looking at the 99th percentile 6-h precipitation value as contrasted with the absolute maximum for the 5-year period. Much more coherent spatial structure is seen for the 99th percentile, with major terrain features emerging much more strongly, and even significant spatial structure over the high plains of eastern Colorado and New Mexico. Applying a Gaussian smoother to the 99th percentile grid produces an even more coherent grid (Figure 18c).

As referenced in the previous section, the dominant strength of the HRRR model dataset is its ability to represent realistic spatial structure in the long-term climatology. Ultimately, we believe the QPF PDFs derived from the model dataset offer much more valuable information than merely the 5-year maximum precipitation grid. To illustrate this, Figure 19 shows the frequency (in terms of probability of exceedance during the 5 year period) of two different precipitation intensity thresholds: 1 mm in 12 h (very light precipitation) and 10 mm in 12 h (moderate precipitation). The spatial structures of these probability maps (Figure 19a,b) are strikingly different from one another. In Figure 19a, it can be seen that light precipitation is very common over the higher terrain areas, particularly along the continental divide in northern Colorado, while light precipitation is relatively rare over lower elevation regions. However, Figure 19b reveals that moderate precipitation intensities are rarer over the higher terrain in northern Colorado, while precipitation of this intensity is much more common over the high plains in eastern Colorado and New Mexico. The differences can be visualized by taking a ratio of the two probabilities (Figure 19c), with a gradient apparent from higher ratios (e.g., light precipitation is common, but heavier precipitation rarer) in northern and western Colorado, to lower ratios (e.g., heavier convective precipitation is more common) in southeastern New Mexico.

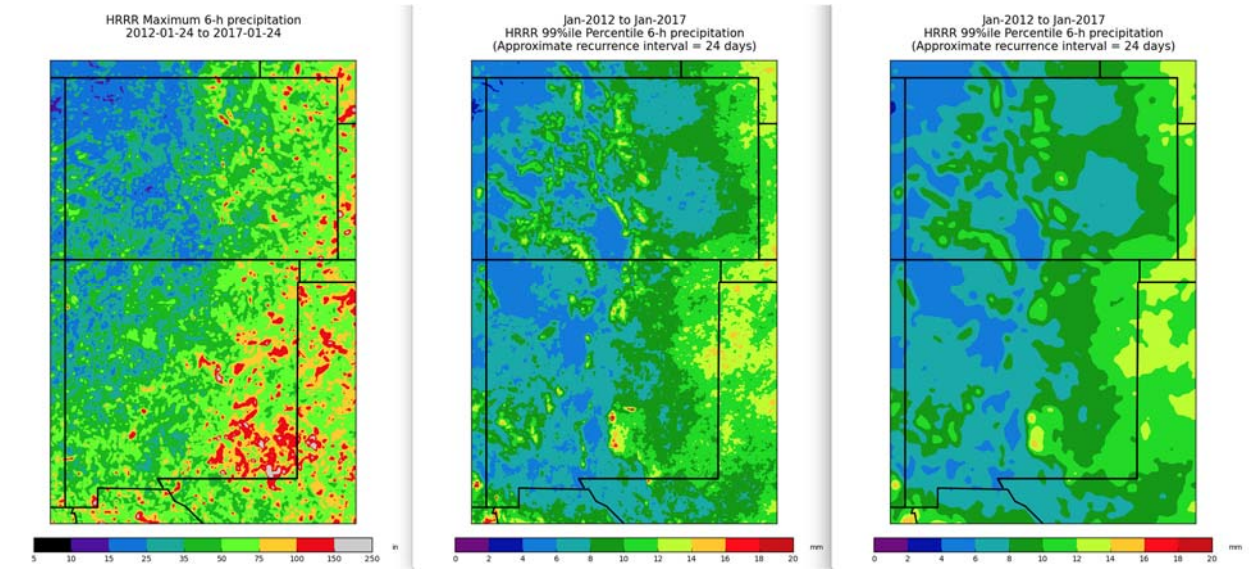


Figure 18. HRRR 3-9-h QPF (a) maximum, (b) 99th percentile, and (c) 99th percentile smoothed with a Gaussian filter for the Jan 2012 to Jan 2017 period.

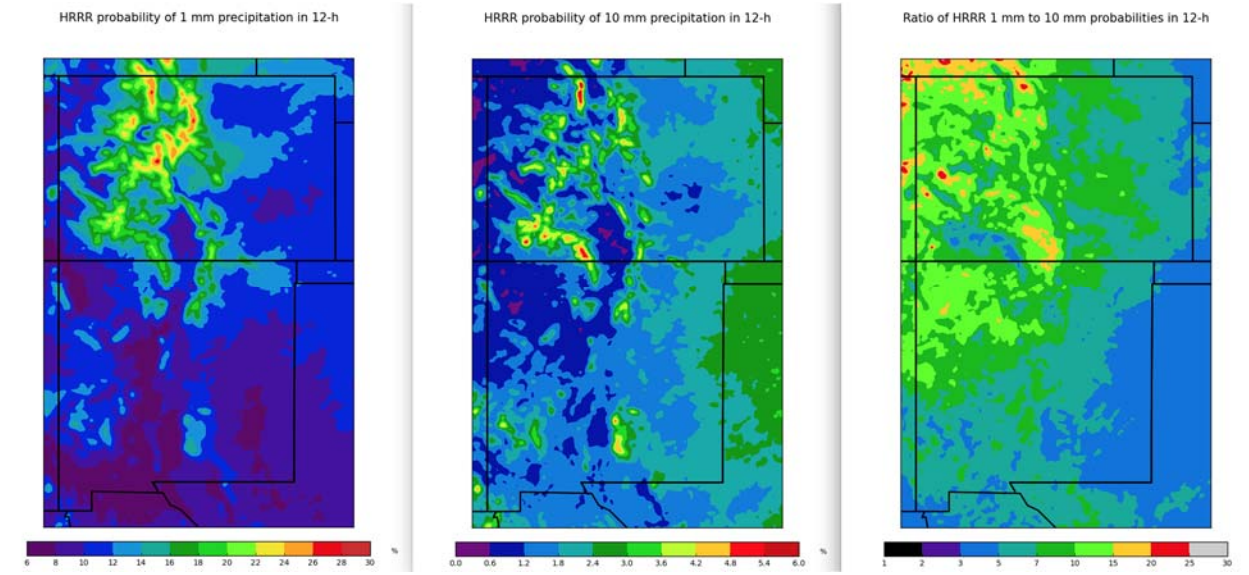


Figure 19. Probability of (a) 1 mm and (b) 10 mm precipitation, and (c) ratio of (a) to (b) over the study region, based on HRRR 3-15-h QPF from Jan 2012 to Jan 2017.

3.2. Comparison to Other Precipitation Datasets

Questions were raised early in the project regarding the degree to which the weather during the 5-year period of record resembled the longer-term climatology. To address this question, we used data from the Climate Prediction Center (CPC) daily precipitation analysis. Figure 20 shows a comparison of the annual average precipitation, as well as the 24-h maximum precipitation, during the entire 50+ year record of the CPC data against the 5 years of the HRRRX archive. As anticipated, there are minor differences in the annual average precipitation patterns when we consider only the 5 years of record of the HRRRX model dataset. The 2012-2016 period

was somewhat wetter in the higher terrain of central and northern Colorado, and somewhat drier in southern New Mexico and southeastern Colorado, than the long-term climatology indicates. More major differences are seen in the maximum 24h precipitation comparison; we would not expect the 5-year maximum to be anywhere close to the maximum from the 50+ year record. This comparison illustrates that we cannot directly use the 5-year maximum and expect it to resemble the PMP value, and also demonstrates that conclusions drawn from the 5-year period may be somewhat dependent on anomalous climatological conditions during those 5 years when compared with the longer-term average.

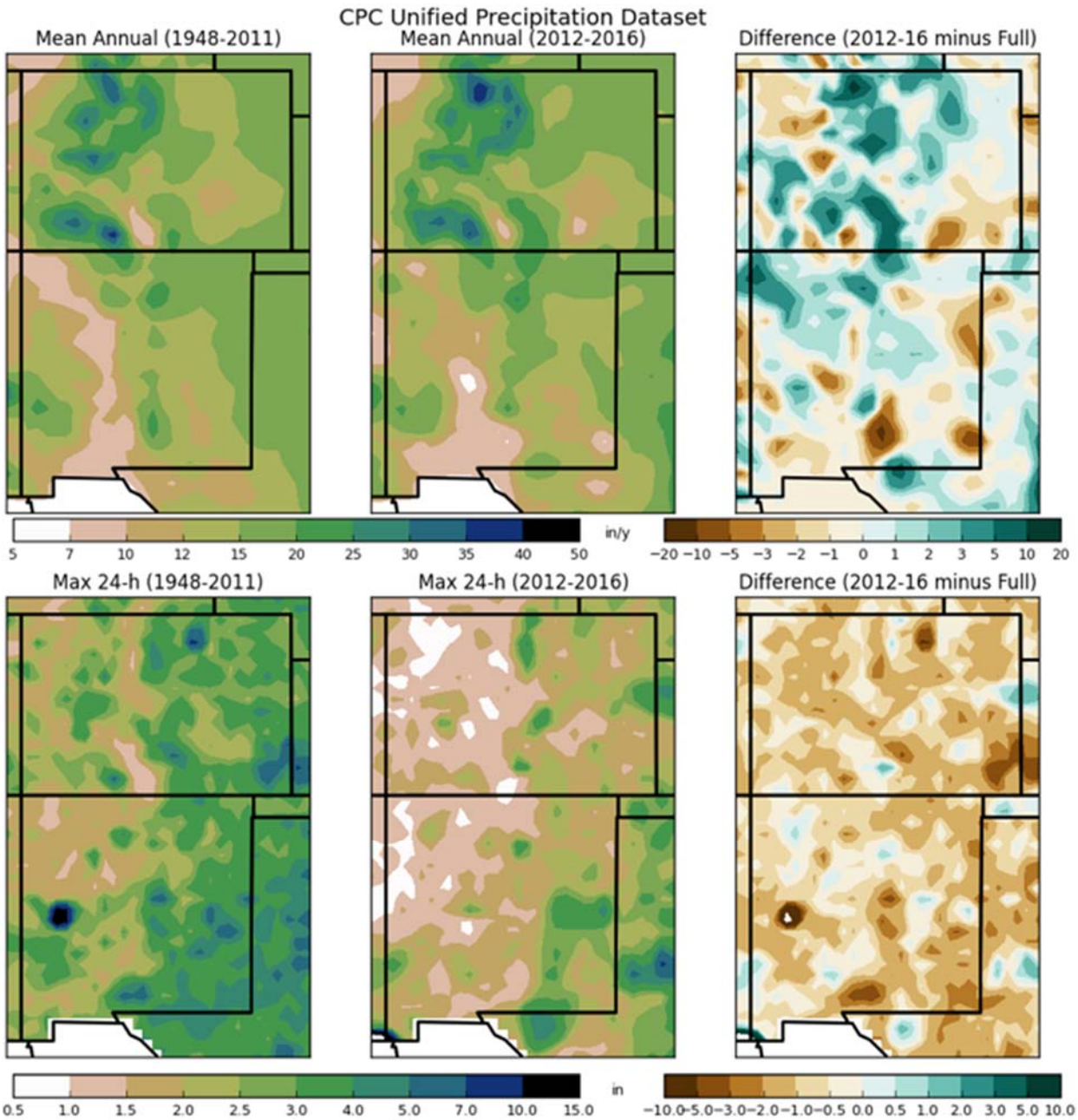


Figure 20. Comparison of mean annual precipitation and 24h max precipitation from the CPC unified precipitation dataset for (left) the entire period of record (1948-2011), (center) the 5 years of the HRRR model dataset, and (right) the difference between the two).

Figure 21 shows a comparison between the maximum 6-h precipitation from the Stage-IV analysis and the maximum from the HRRRX model dataset over Wyoming, Colorado, and New Mexico. As described in section b above, there are issues with the construction of the Stage-IV analysis over complex terrain, which likely contributes to some unrealistic boundaries in this analysis (for example, the international border). However, qualitatively the datasets agree reasonably well. At the very least, we can

conclude that the HRRRX model dataset values are not unreasonable. Comparison with other precipitation durations is not possible, since the Stage-IV analysis is a 6-h product.

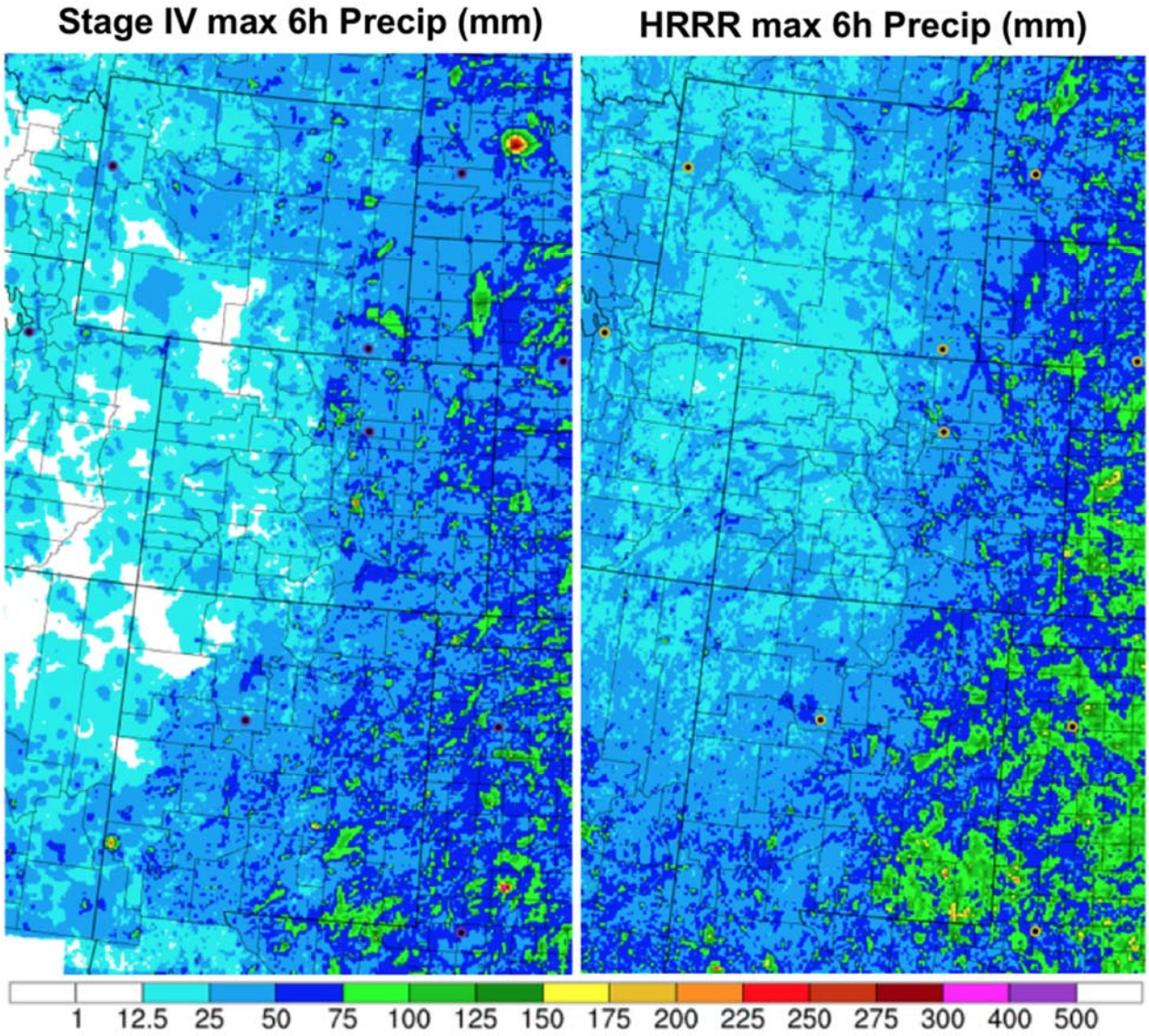


Figure 21. Maximum 6-h precipitation from (left) Stage-IV and (right) the HRRRX model dataset for the 5-year period of record. The HRRRX model data considered are accumulations between forecast hours 6 and 12. The data are time-matched (excluding times when the HRRRX was missing during the 5y period).

3.3. Climatological Analysis

In response to requests from the other tasks within the project, we undertook some pseudo-climatological analysis based upon the 5 years of HRRRX forecasts. We present these results in this section of the report. Figure 22 shows a map of the mean annual precipitation from the HRRRX model dataset, constructed from 6-12-h forecasts. Note that the estimate is somewhat low due to HRRRX outages during precipitation events.

The spatial structure of the annual mean from the 5 years is quite reasonable, and in agreement with previous studies: the heaviest annual totals are apparent in regions of high terrain. There is also a general eastward gradient in annual precipitation on the high plains, in agreement with other studies.

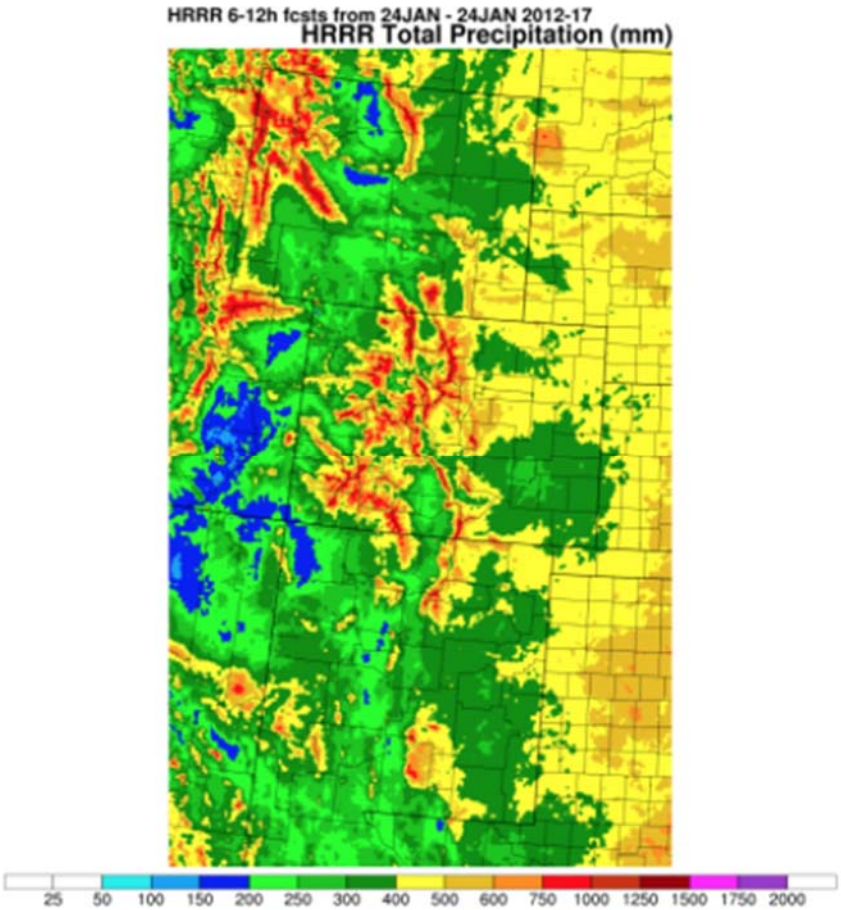


Figure 22. Mean annual precipitation from the HRRR model dataset (constructed from 6-12-h forecasts; the estimate is low due to missing data).

Task 1 was also interested in seeing the monthly variation in mean precipitation from the HRRRX model dataset, in order to inform their spatial mapping of precipitation. Figure 23 shows the 5-year mean monthly precipitation from the HRRRX model dataset. Expected seasonal patterns emerge in these maps, with orographic influences dominating wintertime precipitation, and tropical moisture availability dominating in the warm season. Precipitation in the central plains peaks in May and June, likely due to the presence of Gulf moisture in addition to synoptic scale disturbances and cold fronts. Precipitation in western New Mexico and southwestern Colorado shows a strong later summer peak in connection with the North American Monsoon.

One additional question raised by the Task 1 consultants was the dominance of rain versus snow in higher elevation regions of Colorado and New Mexico. The consultants

were seeking information on whether snow could be neglected within Colorado and New Mexico for different precipitation durations for which they were attempting to determine PMP. The HRRRX model dataset was queried to obtain guidance on this question; results are shown in Figure 24. This figure shows the ratio between the 5-year maximum snow-water-equivalent (for different durations) to the 5-year maximum total precipitation (for those same durations) from the HRRRX model dataset. Values below 1.0 (colors other than red) indicate that the maximum precipitation occurred in the form of rain at that location and for that duration. Red regions indicate that the maximum precipitation occurred in the form of snow. For the 1-h duration, there are very few regions where the maximum precipitation occurred in the form of snow; these are limited to a few isolated regions in far northern Colorado. For the 3-h and 6-h durations, occurrence of the maximum precipitation in the form of snow becomes significantly more widespread, indicating that it may be important to take snow into consideration in the derivation of a PMP value. Even for 6-h durations, though, the maximum precipitation occurs in the form of rain everywhere in the state of New Mexico.

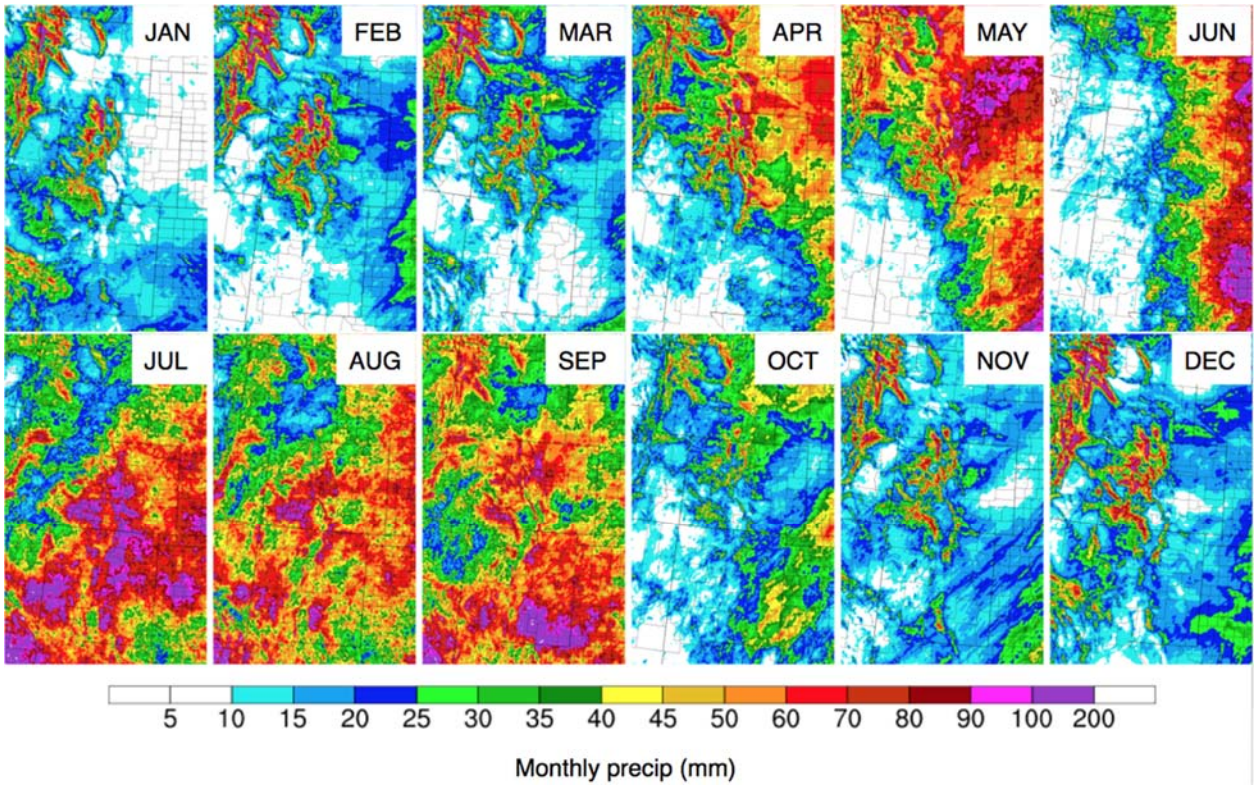


Figure 23. Mean monthly precipitation from the HRRR model dataset (constructed from 6-12-h forecasts).

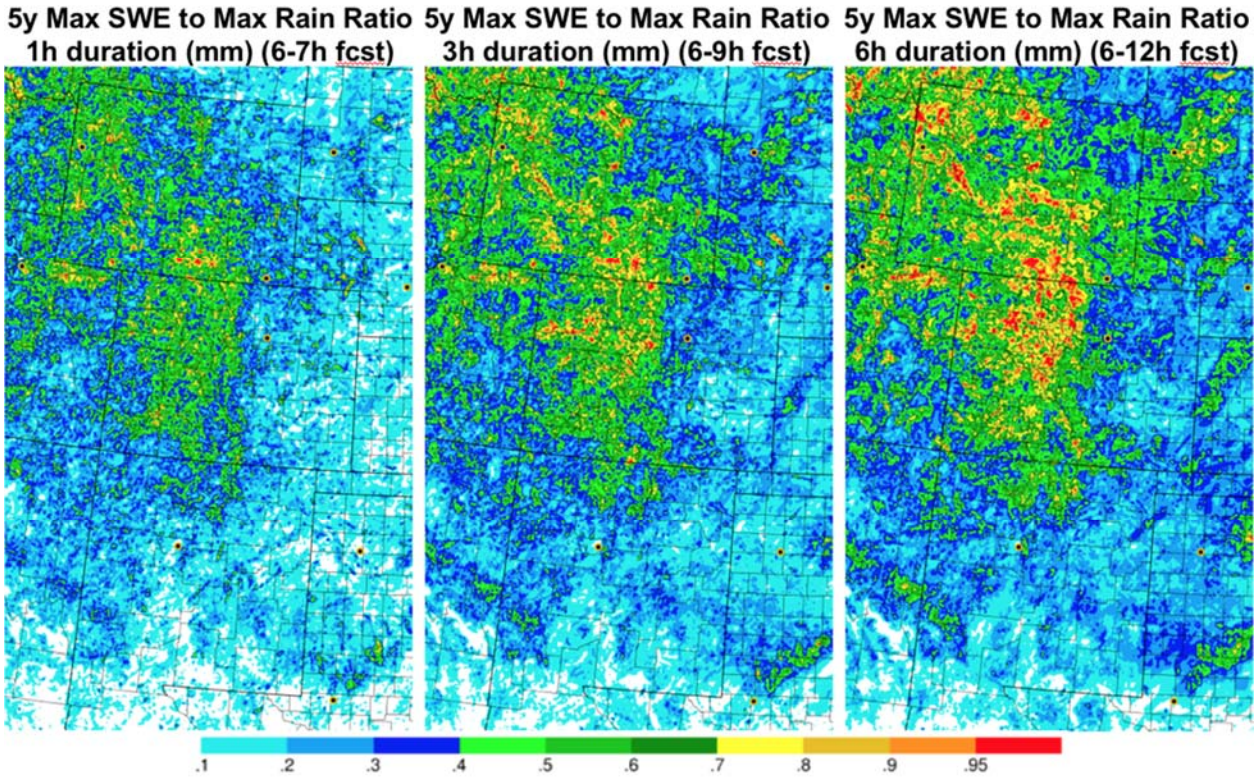


Figure 24. Ratio of 5-year maximum snow water equivalent (SWE) to 5-year-maximum precipitation from the HRRR model dataset for (left) 1-h, (center) 3-h, and (right) 6-h durations. Red colors indicate that the maximum precipitation during these 5 years occurred in the form of snow.

3.4. Elevation Analysis

One interesting avenue of research that was pursued with the HRRRX model dataset is the relationship between precipitation and topographic elevation within various precipitation datasets. A preliminary evaluation was conducted for the Stage-IV QPE and HRRRX QPF during the 5-year period of record, and the long-term Parameter-elevation Regressions on Independent Slopes Model (PRISM) dataset, over the Front Range of northern Colorado and the San Juan Range of southwestern Colorado. Figure 25 shows results from the two regions. When normalized relative to values at 2500 m (to adjust for sample size differences between the datasets), Stage-IV QPE and HRRRX QPF show very similar increases in mean cold-season precipitation with increasing elevation. The results are similar for the warm season (not shown). A key implication here is that since both a manually augmented analysis and dynamical model forecast show a similar overall precipitation structure, it is reasonable to use the model forecast to augment the analysis-based climatology in regions where Stage-IV is prone to systematic issues (e.g., RFC boundaries, limited surface observation and radar coverage). The 30-year PRISM climatology shows a much steeper increase in precipitation with respect to elevation, although this difference can be largely attributed to higher resolution (~800 m vs. 3 km) and the use of a regression model to predict high-elevation precipitation where there often insufficient nearby observing sites to properly condition the regression for local climate effects.

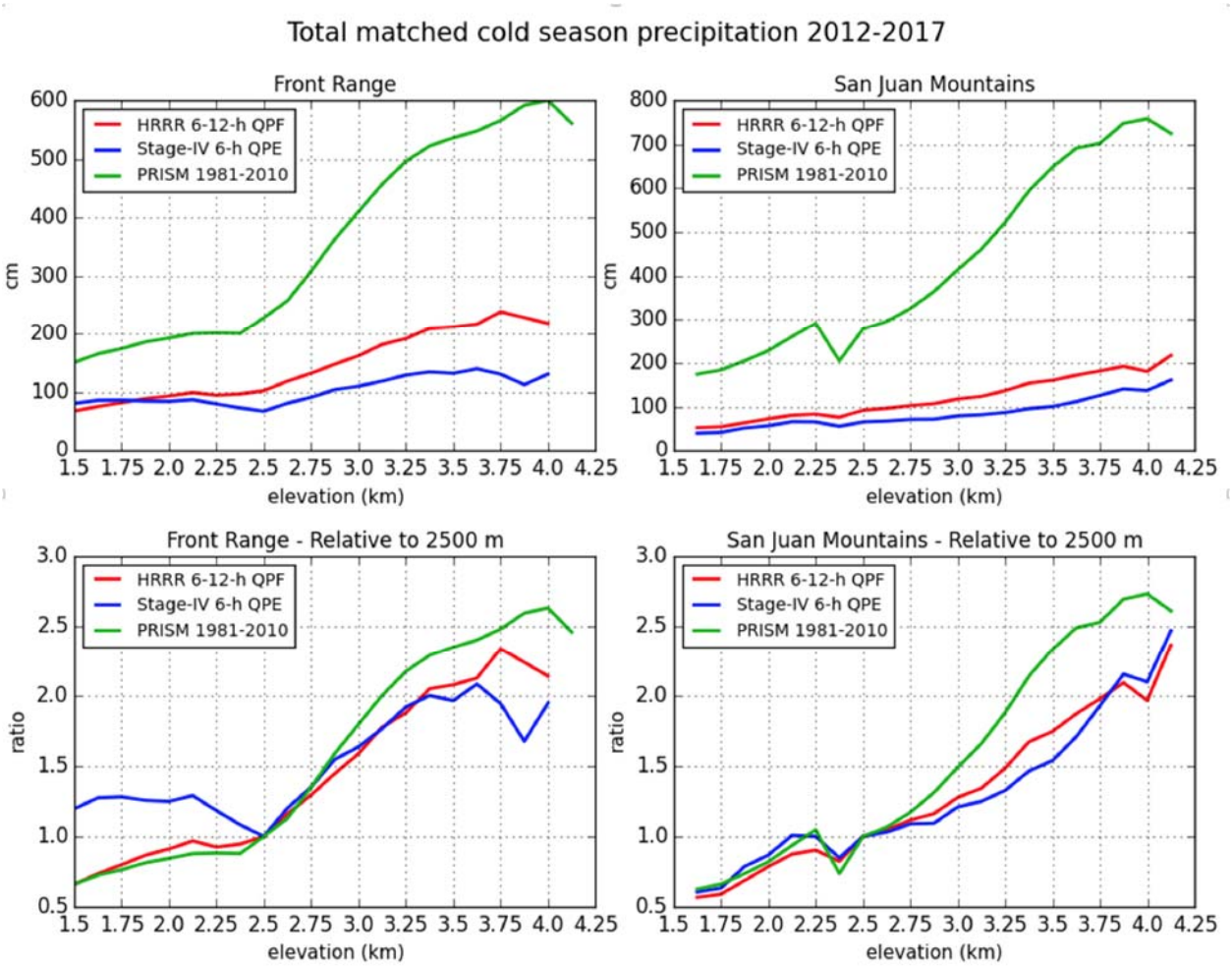


Figure 25. Precipitation-altitude relationship for the Front Range and San Juan Mountain regions of Colorado, in the HRRR 6-12-h QPF, Stage-IV QPE, and the PRISM climatological precipitation datasets, showing similar changes in total precipitation for the HRRR forecasts and Stage-IV analysis in both regions.

3.5. General Numerical Model Utility for Historical Event Analysis

While the HRRR-based analyses and precipitation estimation prototypes comprised the central activity in Task 3's primary goal to demonstrate the use of dynamical numerical weather models for the CO-NM REPS project, some additional work was performed using a more general instance of the WRF model (the same model core used by the HRRR model). This additional work was largely done in collaboration with Task 1, with the overarching goal of obtaining additional information to supplement the existing analyses of historical storms.

As brief background and context for the simulations performed here, it is worth again noting that as computational power and numerical weather forecasting models have sufficiently improved over the past decade or so, it became generally possible to simulate "PMP storms" explicitly. Convection-permitting models are necessary to

simulate heavy precipitation, especially at sub-daily scales, as sufficiently high resolution (generally ≤4 km) permits explicit simulation of deep convection (e.g., Prein et al. 2015). Studies that have taken on this general approach for the specific application of PMP estimation include Abbs (1999), who used the CSU Regional Atmospheric Modeling System (RAMS) at 7-km grid spacing in an early attempt to address PMP with a dynamic weather model by simulating an extreme storm event in Australia. Other methods have been developed such as the atmospheric boundary condition shifting (ABCS) method by Ishida et al. (2015) to ensure that strongest moisture flux impacts a given watershed directly. Many studies have sought to downscale various reanalysis datasets to reconstruct major historic storms (sometimes with moisture maximization applied) to estimate PMP (Ohara et al. 2011; Ishida et al. 2015a; Chen and Hossain 2016; Tan 2010). The historical event modeling done for the CO-NM REPS project and discussed below also falls into the general reanalysis-downscaling class of possible dynamical model methods for PMP event investigation.

Specific storms of interest for CO-NM REPS were initially selected by Task 1's assessment of:

1. Importance in existing (previous) PMP values,
2. Lack of observations from which to derive robust storm patterns and magnitudes, and
3. Uncertainty in the previous analysis results from the USACE/USBR/NWS
4. Uncertainty in the previous basemap utilized by AWA to accurately capture the spatial distribution
5. Limited surface observation data for rainfall analysis and storm maximization

Thus, four historical events were initially investigated according to the above criteria:

1. Rattlesnake, Idaho 1909
2. Savageton, WY 1924
3. Penrose, CO 1921
4. Ward District, CO 1894

And three additional cases were added at the end of the project at the request of the CO and NM project sponsors, again according to similar criteria as above:

1. Elbert/Cherry Creek, CO 1935
2. Opal, WY 1990
3. Virslyvia, NM 1922

3.5.1 Model set-up

The historical event model simulations employ the Advanced Research Weather Research and Forecasting (WRF-ARW) modeling system, Version 3.7.1 (Skamarock et al. 2008). The model domains use 3- or 4-km grid spacing and 54 vertical levels; the small grid spacing (variable between 3- and 4-km dependent upon computational resources) affords the omission of convective parameterization and sufficiently resolves flow in and around fine-scale terrain features. Initial and lateral boundary conditions are provided by 6-hourly 20th Century Reanalysis version 2c (20CRv2c) data (Compo et al. 2011; and

https://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2c.html) Other relevant model physics choices include: Thompson microphysics (Thompson et al. 2008); Yonsei University (YSU; Hong et al. 2006) planetary boundary layer, Monin-Obukhov surface layer, Dudhia shortwave radiation (Dudhia 1989), Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) longwave radiation scheme, and the Noah land surface model (Ek et al. 2003). Note that for cases involving additional experimentation and more simulations (e.g., 1924 Savageton) physical parameterization variation and sensitivity experiments were also performed.

As one might expect for very old cases, the 20CR ensemble members displayed considerable synoptic meteorological variability across the events, and the number of simulations performed for each event varied between four and sixteen simulations depending on how much spread was found between the WRF ensemble members. The original plan was to execute four separate simulations (initialized using four different members of the 20CR ensemble) for each event. Provided a strong enough signal in the large-scale initial condition environment to produce significant precipitation, this number of simulations was thought to provide some indication (though hardly exhaustive) of variability/sensitivity to the different initial conditions employed. If there a lot of model spread was found across the WRF simulations for a particular event, or if the model output looked very different from the existing Task 1 historical SPAS analysis, then additional simulations beyond the standard initial four were performed.

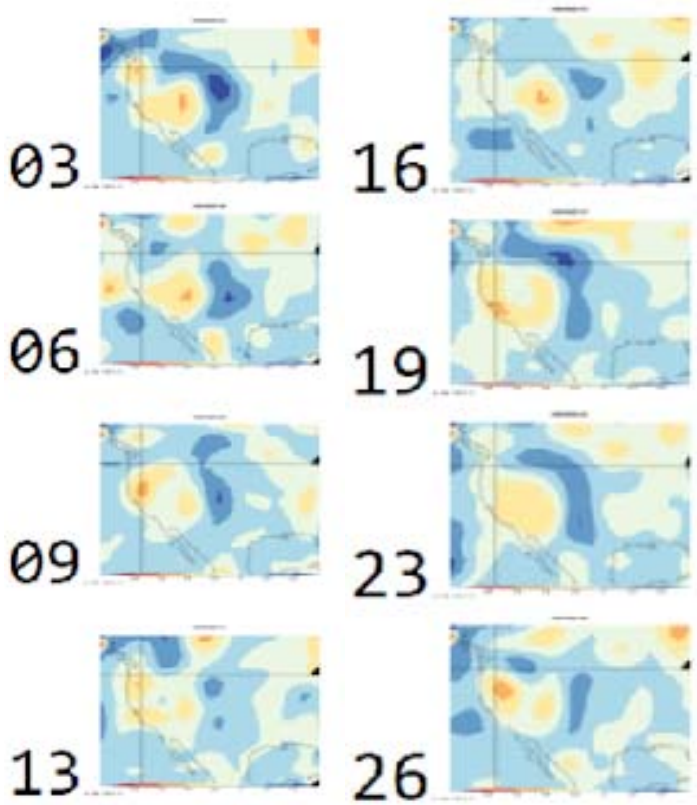


Figure 26. Example of omega (vertical velocity, shaded quantity where cool colors represent upward motion) across randomly-selected 20CRv2c (i.e., the initial

condition dataset) members (numbers denote member numbers 03, 06,…, 26 of 20CRv2c). This is an example meant to demonstrate that the 20CR dataset provides ample spread in synoptic weather conditions to account for uncertainty in historical environmental conditions.

3.5.2 Utility of Historical Simulations

A specific example of the results and ultimate utility of one historical event case study is provided below. For additional detail and the results of other simulations, please refer to Volume II (Task 1) and Mahoney et al. (2018; included as Appendix A below as well).

For each event, individual member model output was provided to AWA/Task 1, along with an "ensemble max" precipitation grid. The ensemble max grid retains the maximum event-total precipitation produced at each gridpoint and thus demonstrates how intense the event was simulated to be, gridpoint by gridpoint, across all event ensemble members.

3.5.2.1. Example case study: Rattlesnake, Idaho 1909

The Rattlesnake, Idaho 1909 flood was the result of a week-long series of inland-penetrating atmospheric rivers (ARs). ARs are synoptic-scale weather systems that are inherently more predictable and better-represented by numerical models relative to small-scale convective storms. The combination of a large-scale, intense atmospheric feature that is more likely to be well-represented in the 20CR, plus steep orography in Idaho where the AR penetrated inland offers increased potential for successful numerical model simulation of extreme precipitation relative to more weakly-forced and/or non-orographically focused precipitation events.

Indeed, for this case, a small ensemble (four simulations) offered immediate consistency among themselves and agreed closely with available historical observations (Figure 27, Figure 28). Task 1 incorporated the model output fields into their PMP calculation methods after confirming that the model output provided useful spatial pattern information when applied as an improved precipitation basemap (from which PMP estimation begins). The WRF simulations also helped to inform rain-snow delineation to identify regions with snow (which is less relevant to PMP). There was discussion of the use of single deterministic simulation data (and the benefit of internal physical consistent) vs. ensemble diagnostics (spread, ensemble max) for this case. However, since spread was generally low, these topics were left for future research consideration.

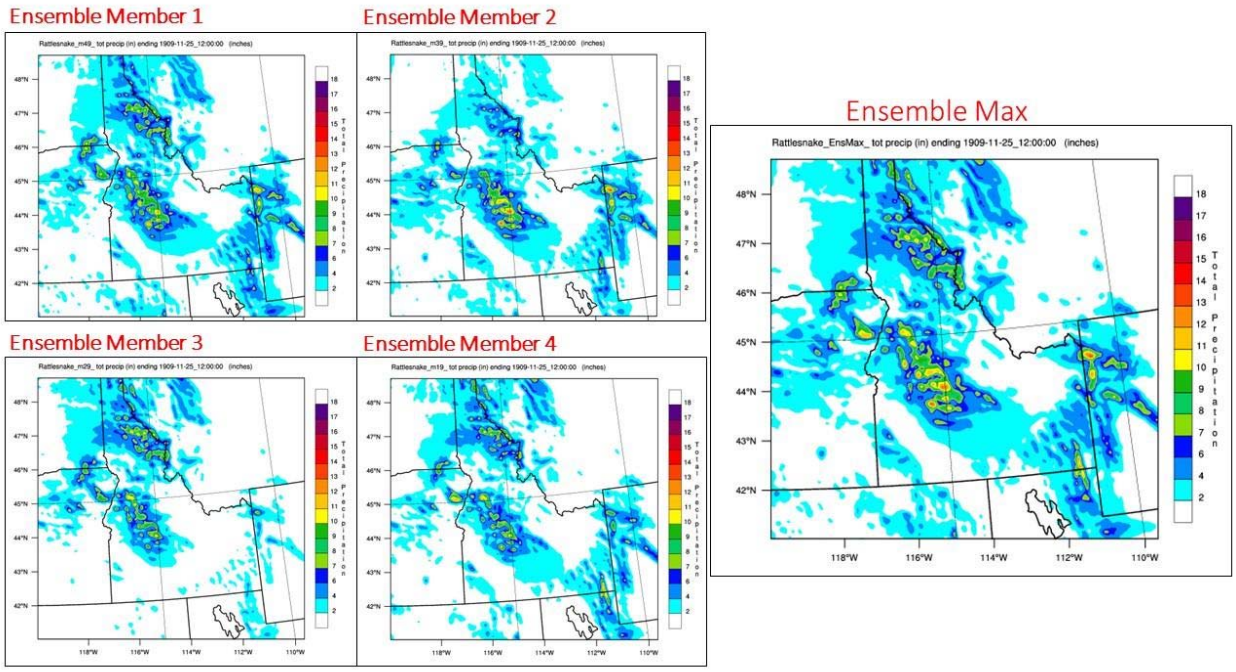


Figure 27. (left, 4-panel) WRF simulated storm-total precipitation (inches, as in color bar at right) for 1909 Rattlesnake, Idaho flood from four individual WRF simulations. (right) Ensemble maximum value from all four model runs combined precipitation (inches, as in color bar at right).

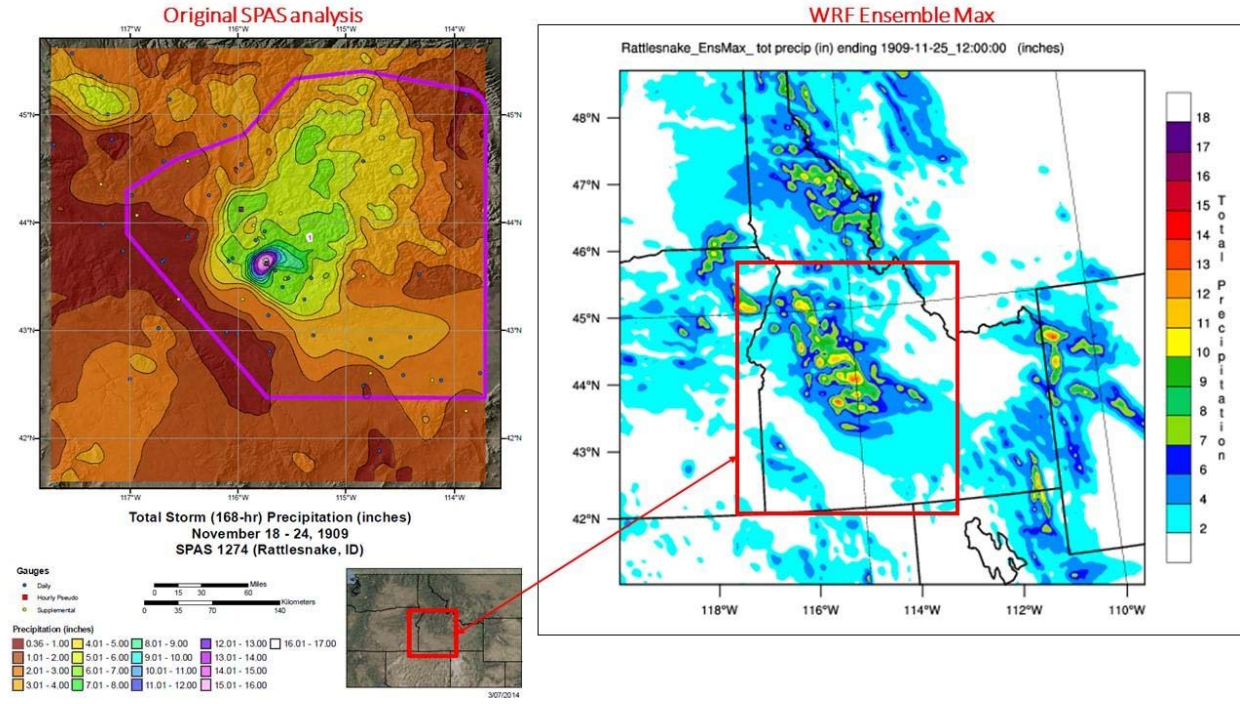


Figure 28. (left) Existing SPAS analysis produced by Task 1 prior to obtaining WRF model results (inches, as in color bar at bottom) for 1909 Rattlesnake, Idaho flood. (right) WRF ensemble maximum storm-total precipitation from all four model runs (inches, as in color bar at right). Red boxes denote approximately same area.

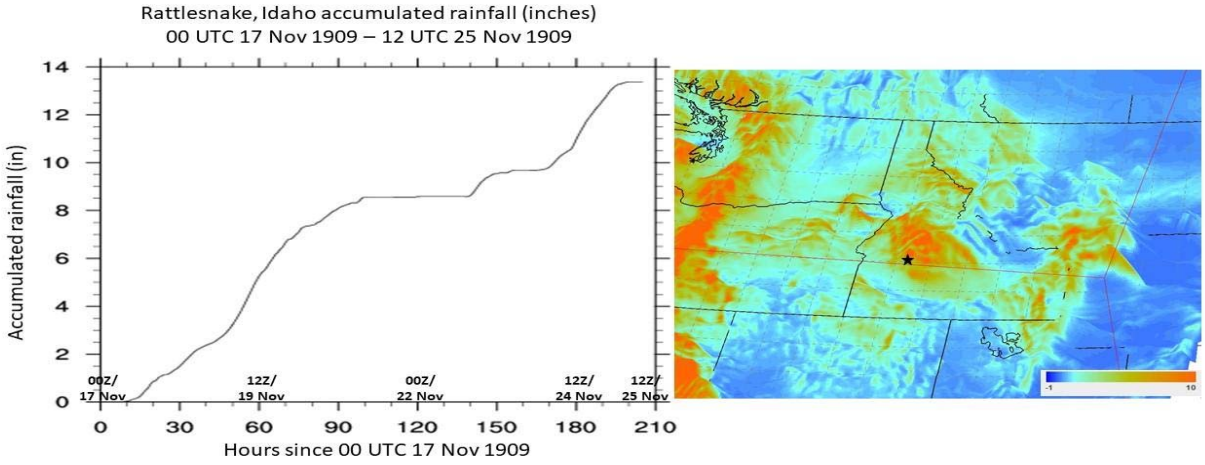


Figure 29. (left) Times series of one individual ensemble member's precipitation (inches) at a maximum grid point for 1909 Rattlesnake, Idaho flood; location denoted by star at right. (right) Storm-total precipitation from single WRF member (inches as shaded by color bar at bottom right). Time series provided merely as an example of temporal distribution information available from this type of model application.

The 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 and temporal assumptions (e.g., Fig. 29) made with very limited observational data. This event (compared with others) suggests that the utility of WRF simulations may be proportional to role of topography in controlling rainfall spatial pattern and magnitude. This hypothesis requires further testing.

Details on the rest of the historical simulations are omitted from the report at this time, as the work continues to be both a) in progress, and b) subject to outstanding important questions regarding interpretation and usability. Ongoing and future work will need to approach this specific dynamical model application of simulating historical storms with a more structured and exhaustive experimental design in order to establish the most appropriate application of results. For example, in some instances, such as the 1909 Rattlesnake example above, the most appropriate application may be to use simulated storm characteristics such as spatial and temporal distribution patterns with the aim of improving existing analyses. However, for cases where model simulations did not yield the expected, historical observation-indicated precipitation event, the model data might instead be considered as evidence of erroneous observational data. Criteria (ideally objective, and perhaps based on ensemble skill or spread statistics) should be established by more rigorous testing in future work.

Finally, it is important to emphasize that the application of dynamical model methods and resulting data is not without its own set of challenges and subjectivities. Even in the relatively straightforward reanalysis-downscaling simulations performed here, there remains substantial subjectivity in model set-up, ensemble size, analysis approach, and more. Additional work is needed to compare intrinsic uncertainties from historical point observations/data-limited observational analysis outright vs. dynamical model simulations forced by historical reanalyses.

Other dynamical model PMP applications and methods such as moisture maximization, boundary condition shifting, and climate model downscaling can offer numerically-constrained PMP estimates, but (given that model data are combined with complex and somewhat subjective moisture- and storm-maximizing PMP processes), these often introduce even more significant subjectivities, limitations, and caveats. Such approaches may offer utility for limited-area, site-specific studies in particular instances, but the development and pursuance thereof may not ultimately offer as much long-term value toward achieving an objective, NWP-generated upper-bound of precipitation. Additional possible future paths and recommendations for investment of time and NWP resources are discussed in the future work section.

4. Dynamical Model Applications Used by Tasks 1 and 2

The primary objective of Task 3 was to demonstrate ways in which a dynamical weather model-based framework might benefit future iterations of PMP, or more generally, extreme precipitation estimation. However, additional interests include whether or not the existing 5-year model dataset could provide useful information regarding the frequency, variability, and intensity of extreme precipitation in ways that might support existing PMP estimation methods and/or specific parts of the traditional PMP/extreme precipitation frequency estimation process. To this end, considerable testing, evaluation, and collaboration took place between the three Tasks.

As detailed in Volume II, Task 1 used HRRR model for the adjustment of precipitation frequency climatologies to represent rainfall-only values (Vol. II, Sect. W), and WRF dynamical model case studies for improved spatial patterns in SPAS basemaps for the Rattlesnake, ID November 1909 (SPAS 1274), Ward District, CO May 1894 (SPAS 1614) storms , and Penrose, CO 1921 (SPAS 1294) (Vol. II, Sect. P).

As detailed in Volume III, Task 2 used the HRRR model for specific storm data to supplement a variety of spatial precipitation patterns used to develop watershed precipitation-frequency relationships using stochastic storm generation methods (Vol. III, Sect 3.4). A total of 5 storm dates were provided from the HRRR forecast model dataset, from which Task 2 developed 56 MEC and 304 Local Storm spatial patterns, which helped to increase the sample size in the suite of storms used in the stochastic storm generation process.

As Task 3 was by definition a prototype demonstration of the potential utility of dynamical model data, there was a considerable learning curve between the tasks to all understand the needs and potential of the various groups and tools. As all of the consultants became increasingly familiar with one another's challenges and capabilities, exchanges and collaborations such as those listed above become possible. However, by the project's completion, model products and applications that would have potentially been useful to Tasks 1 and 2 became increasingly apparent. Tasks 1 and 2 provided specific input for how Task 3 tools and/or data could be leveraged in the future, including:

- Task 1:
    - Refine WRF reanalysis to re-create past extreme events
    - Continue analysis of HRRR output to define rain only regions and high elevation rainfall relationships
    - Investigate assumptions regarding saturated atmosphere in the storm adjustment process
    - Investigate assumptions of increasing moisture in a storm and when that changes the storm's dynamics (e.g. using a model to evaluate the 1.50 upper limit of the IPMF)
    - Utilize HRRR output to better quantify/understand application of transposition limits
    - Generate model-based storm center mass curves and/or regional storm center mass curves to aid/bolster temporal distribution analysis (5min and 1hr)
    - Generate model-based local and general storm ARFs by region to compare to SPAS and/or other studies
    - Perform additional seasonal analysis (rain/snow/total precipitation) and climatologies to compare against PRISM/NOAA
- Task 2:
    - Use a dynamical modeling framework to better quantify changes (in time and space) in the in-place moisture maximization factor (IPMF) procedure associated with the PMP methodology used by Task 1.
    - Longer duration precipitation accumulation periods (i.e. 48-hour, as opposed to HRRR's 18-hour operational forecast length) to help resolve high-elevation decreases in precipitation.
    - Creation of hourly virtual "stations" for use in observation-based storm precipitation-reconstructions for use in our Move-the-Earth process for ascertaining areal precipitation frequency estimates.
    - Storm specific temperature & freezing level height time series for use in creating representative templates for hydrologic modeling.
    - Create high-resolution, model precipitation fields with a longer period of record.

5. Conclusions

The work completed under CO-NM REPS Task 3 demonstrates the potential utility of high-resolution dynamical model data to support extreme precipitation estimation for dam safety risk assessment applications. While the short temporal period of HRRR coverage imposes a severe limitation on the degree to which maximum precipitation model output can be used outright in this study, model demonstrations summarized here highlight potential benefits to dynamical modeling frameworks, including:

- High-resolution (in space and time), spatially continuous precipitation estimation
- Data produced by solving physical equations of the atmosphere (vs. interpolation, interpretation of limited observations), reducing reliance on spatial, temporal, physical assumptions (e.g., storm transposition, storm templates, moisture maximization, etc.)
- Coverage in remote, data-sparse regions locations lacking observations (of particular benefit in regions of complex and/or high-elevation topography)
- Explicit model representation of precipitation type (snow, rain, hail)
- Reconstruction of major historical events via numerical modeling may supplement existing storm analyses, improve spatial and temporal assumptions made with very limited observational data
- Straightforward methods for quantification of uncertainty

6. Future Research and Development

The Task 3 dynamical model prototype exercises illuminated multiple potential directions for future efforts. Three "tiers" are presented in order of increasing degree of the approximate investment likely to be required. Note that all ideas presented would benefit from (i.e., their success might rely heavily on) the production of a long-term, frozen model, high-resolution dynamical-model produced reanalysis or reforecast dataset as described in more detail in "Tier III".

Tier I: Use existing model data to support specific requests for model-generated products by Task 1- and Task 2-type approaches.

Future studies could consider adding a dynamical modeling component in which the types of model data analysis and sharing that have occurred for CO-NM REPS would be part of the study from its inception (e.g., providing basemaps, storm patterns, case studies, etc.) To assess how a generic dynamical-modeling-type of effort might be improved for future studies, Task 1 and Task 2 provided input regarding what types of model capabilities, tools, or products that would have been useful to have had for CO-NM REPS. Specific responses from each respective task are documented in Section III.F, above.

Tier II: Hybrid approach using existing model data in combination with longer-term observations to maximize space-for-time relationships

The main strength of the HRRR dataset within the CO-NM REPS framework was the fine temporal and spatial resolution of the precipitation forecasts, despite the short period of "record". The limited (5-year) period of record means that the far tail of the precipitation probability density function (PDF; i.e., very rare events) is less robust. However, it is hypothesized that 5 years is sufficient to define the PDF for more frequent events. Thus, we can derive a reasonable PDF at each model gridpoint (at least for precipitation thresholds that are relatively common). For points where there exist high-quality, long-term rain gauge observations, a PDF can be derived which better represents less frequent events. By parameterizing both model and observed PDFs, regression relationships can be calculated between the parameters of observed PDFs (generally for a single accumulation interval, e.g., 24 hours), and those of model PDFs. These relationships can then be applied elsewhere in the model grid, effectively extending the model PDFs out to very rare events, and enabling calculation of extreme precipitation recurrence intervals on the entire model grid. Similar methods could be used to calculate long-term area or basin-average PDFs, by applying observation-based corrections to model forecast area- and basin-average PDFs.

Tier III: Generation of a longer-term, high-resolution dynamical model dataset in combination with advanced statistical post-processing

There is mounting desire across many user groups and sectors of the meteorological community for long-term, high-resolution reanalysis/reforecast data (e.g., Hamill et al. 2013). CO-NM REPS Task 3 work has highlighted the utility of model data at high space and time resolution for the applications of dam safety and flood risk management, but can only prototype potential capabilities given the short 5-year period of record. While deterministic, high-resolution model data and analysis are highly valued for skillful intensity estimates that are continuous in space and time, there has also been a notable recognition among some key Federal agencies of the weaknesses of a single-value, deterministic extreme precipitation estimation approach (i.e., PMP). This has led to a notable shift toward probabilistic, risk-informed decision-making frameworks for infrastructure design by some agencies and decision-making bodies in the dam safety community (e.g., Reclamation, Nuclear Regulatory Commission). NWP applications for PMP may also be best suited for a framework that more generally improves the estimation of heavy precipitation potential/likelihood at various decision-making thresholds. Therefore, an annual exceedance probability (AEP)-based framework that optimally combines high-resolution, spatially and temporally continuous data with a long period of coverage should be investigated to comprehensively address the question of how NWP can best serve the dam safety and flood risk decision-making communities.

There are many ways in which such an effort might be undertaken; we list below a collection of possible suggested starting points:

1. Survey and confirm with end users how model data would be optimally used (e.g., confirm decision-making thresholds, how model information would be brought into existing/future decision-making process, etc.).
2. Generate a historical, high-resolution model reanalysis and/or reforecast dataset. For example, one might consider:
    a. 20CRv3 forcing (1/4-degree resolution, more than 80 members. Would want to test use of ensemble mean vs. ensemble max/min/individual members).
    b. 1 cycle per day, include extra (~3 – 6) hours to allow for spin-up (e.g., Trapp et al. 2011).
    c. 1860s to present.
    d. HRRR-based compute benchmarking suggests ~0.5million core hours per year of simulation time (i.e., 150y = 75 million core hours) for a single 24-hour forecast once per day for full continental US (smaller domains could be run, e.g., just the Western US).
    e. Could also consider adaptive grid modeling framework (e.g., MPAS at https://mpas-dev.github.io/) to expend computational resources only in specific locations or instances (e.g., occurrence of significant precipitation); reduction or elimination of operational-grade data assimilation could/should also be eliminated to avoid both increased computational cost as well as inconsistent data input across a long period of record (as observational data sources change).
    f. Provided that computational resources allow, an ensemble approach is recommended with members defined by different initial condition members, stochastic perturbations to the environment and/or model physics, or a combination of the above.
3. Use advanced statistical methods to:
    a. Enhance calculation of probabilities (increase effective period of record, extend to longer return periods) based on regional/historical similarities (e.g., Scheuerer and Hamill 2015).
    b. Create point-to-areal/watershed-average values.
    c. Consider running a dynamical hydrologic modeling system (e.g., WRF-Hydro model) using historical reforecast inputs.

References

Benjamin, S. G., D. D. Devenyi, S. S. Weygandt, K. J. Brundage, J. M. Brown, G. A. Grell, D. Kim, B. E. Schwartz, T. G. Smirnova, T.-L. Smith, and G. S. Manikin, 2004: An hourly assimilation-forecast cycle: The RUC. *Mon. Wea. Rev.*, 132, 495-518.

Benjamin, S. G., S. S. Weygandt, J. M. Brown, M. Hu, C. R. Alexander, T. G. Smirnova, J. B. Olson, E. P. James, D. C. Dowell, G. A. Grell, H. Lin, S. E. Peckham, T. L. Smith, W. R. Moninger, J. S. Kenyon, and G. S. Manikin, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, 144, 1669-1694.

Blaylock, B. K., J. D. Horel, and S. T. Liston, 2017: Cloud archiving and data mining of High-Resolution Rapid Refresh forecast model output. Computers and Geosciences, 109, 43-50.

Bytheway, J. L., C. D. Kummerow, and C. Alexander, 2017: A features-based assessment of the evolution of warm season precipitation features from the HRRR model over three years of development. Wea. Forecasting, 32, 1841-1856.

Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason, R. S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R. I. Crouthamel, A. N. Grant, P. Y. Groisman, P. D. Jones, M. Kruk, A. C. Kruger, G. J. Marshall, M. Maugeri, H. Y. Mok, Ø. Nordli, T. F. Ross, R. M. Trigo, X. L. Wang, S. D. Woodruff, and S. J. Worley, 2011: The Twentieth Century Reanalysis Project. Quarterly J. Roy. Meteorol. Soc., 137, 1-28. http://dx.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 Report for Contract ENC #C154213 - State of Colorado (Dept Natural Resources, Dept. Atmospheric Science, Colorado State Univ.

Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Y. Zhu, and W. Lapenta, 2013: NOAA's Second-Generation Global Medium-Range Ensemble Reforecast Dataset. Bull. Amer. Meteor. Soc., 94, 1553–1565, doi:https://doi.org/10.1175/BAMS-D-12-00014.1.

Hu, M., S. G. Benjamin, T. T. Ladwig, D. C. Dowell, S. S. Weygandt, C. R. Alexander, and J. S. Whitaker, 2017: GSI three-dimensional ensemble-variational hybrid data assimilation using a global ensemble for the regional Rapid Refresh model. Mon. Wea. Rev., 145, 4205-4225.

Ikeda, K., M. Steiner, J. Pinto, and C. Alexander, 2013: Evaluation of cold-season precipitation forecasts generated by the hourly updating High-Resolution Rapid Refresh. Wea. Forecasting, 28, 921-939.

Mahoney, K. M., C. McColl, B. D. Kappel, and D. M. Hultstrand, 2018: New Data for Old Storms: Can New, Convection-Allowing Ensemble Simulations of Historic Storms Help Minimize Present-Day Flood Risk? 29th AMS Conference on Weather Analysis and Forecasting, 5 June 2018, Denver, CO.

Nelson, B. R., O. P. Prat, D.-J. Seo, and E. Habib, 2016: Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product intercomparisons. Wea. Forecasting, 31, 371-374.

Peckham, S. E., T. G. Smirnova, S. G. Benjamin, J. M. Brown, and J. S. Kenyon, 2016: Implementation of a digital filter initialization in the WRF model and its application in the Rapid Refresh. Mon. Wea. Rev., 144, 99-106.

Pinto, J. O., J. A. Grim, and M. Steiner, 2015: Assessment of the High-Resolution Rapid Refresh model's ability to predict mesoscale convective systems using object-based evaluation. Wea. Forecasting, 30, 892-913.

Scheuerer, M., and T. M. Hamill, 2015: Statistical post-processing of ensemble precipitation forecasts by fitting censored, shifted Gamma distributions. *Monthly Weather Review*, **143**, 4578-4596.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers, 2008: A description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp.

Trapp, R. J., E. Robinson, M. Baldwin, N. Diffenbaugh, and B. J. Schwedler, 2011: Regional climate of hazardous convective weather through high-resolution dynamical downscaling. Climate Dyn., 37, 677–688, doi:https://doi.org/10.1007/s00382-010-0826-y.

Acknowledgements

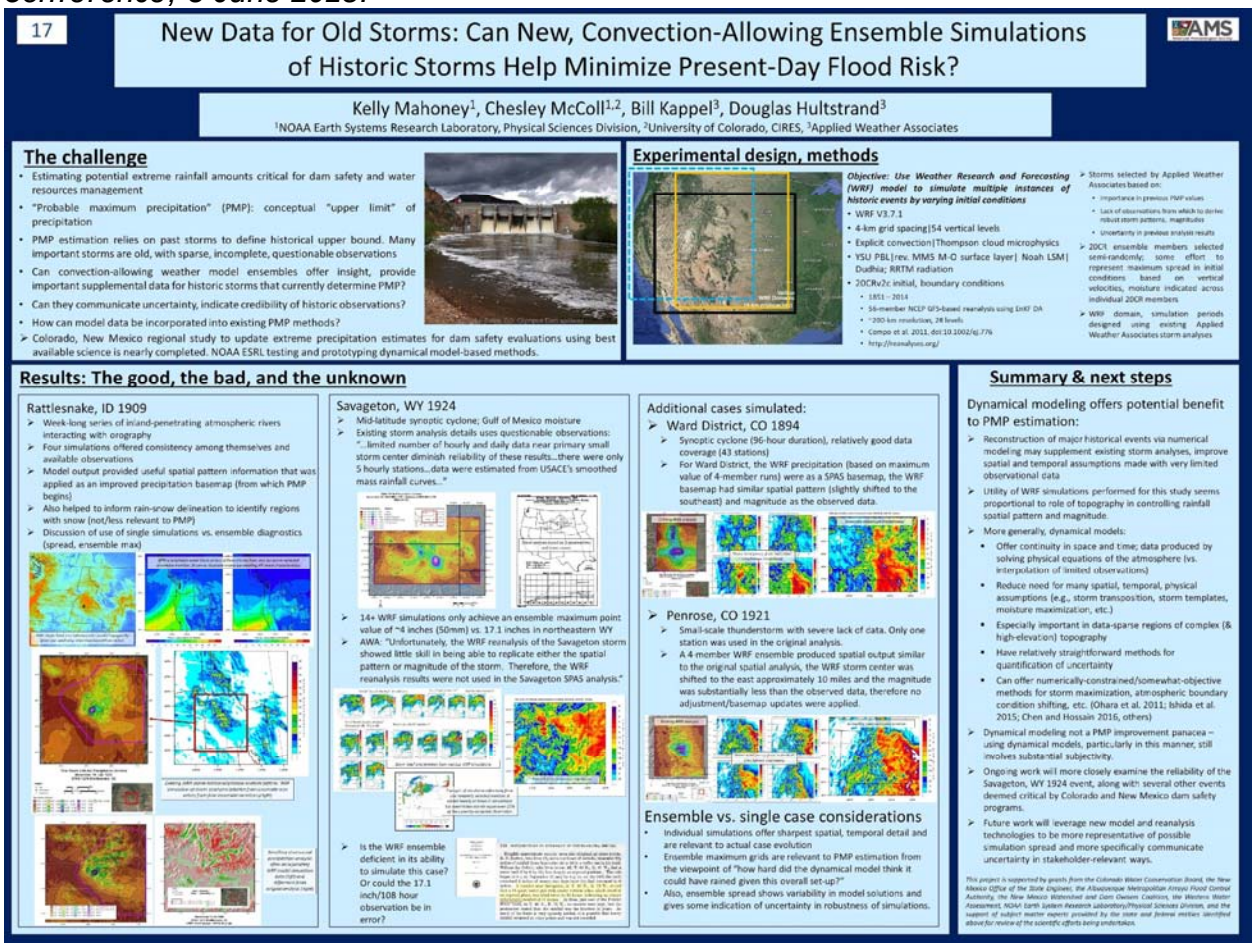
Chesley McColl (CIRES/NOAA PSD) for help obtaining 20th Century Reanalysis V2c data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/.

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.

Russ Schumacher (CSU) and David Keeney (USBR) for historical storm data.

Appendix A

Appendix A: Poster from Mahoney et al. (2018) AMS Weather Analysis and Forecasting Conference, 5 June 2018.



17

New Data for Old Storms: Can New, Convection-Allowing Ensemble Simulations of Historic Storms Help Minimize Present-Day Flood Risk?

Kelly Mahoney[1], Chesley McColl[1,2], Bill Kappel[3], Douglas Hultstrand[3]
[1]NOAA Earth Systems Research Laboratory, Physical Sciences Division, [2]University of Colorado, CIRES, [3]Applied Weather Associates

The challenge

- Estimating potential extreme rainfall amounts critical for dam safety and water resources management
- "Probable maximum precipitation" (PMP): conceptual "upper limit" of precipitation
- PMP estimation relies on past storms to define historical upper bound. Many important storms are old, with sparse, incomplete, questionable observations
- Can convection-allowing weather model ensembles offer insight, provide important supplemental data for historic storms that currently determine PMP?
- Can they communicate uncertainty, indicate credibility of historic observations?
- How can model data be incorporated into existing PMP methods?

➢ Colorado, New Mexico regional study to update extreme precipitation estimates for dam safety evaluations using best available science is nearly completed. NOAA ESRL testing and prototyping dynamical model-based methods.

Experimental design, methods

Objective: Use Weather Research and Forecasting (WRF) model to simulate multiple instances of historic events by varying initial conditions

- WRF V3.7.1
- 4-km grid spacing|54 vertical levels
- Explicit convection|Thompson cloud microphysics
- YSU PBL|rev. MM5 M-O surface layer| Noah LSM| Dudhia; RRTM radiation
- 20CRv2c initial, boundary conditions
  - 1851 – 2014
  - 56-member NCEP GFS-based reanalysis using EnKF DA
  - ~200-km resolution, 24 levels
  - Compo et al. 2011, doi:10.1002/qj.776
  - http://reanalyses.org/

➢ Storms selected by Applied Weather Associates based on:
- Importance in previous PMP values
- Lack of observations from which to derive robust storm patterns, magnitudes
- Uncertainty in previous analysis results

➢ 20CR ensemble members selected semi-randomly; some effort to represent maximum spread in initial conditions based on vertical velocities, moisture indicated across individual 20CR members

➢ WRF domain, simulation periods designed using existing Applied Weather Associates storm analyses

Results: The good, the bad, and the unknown

Rattlesnake, ID 1909

- ➢ Week-long series of inland-penetrating atmospheric rivers interacting with orography
- ➢ Four simulations offered consistency among themselves and available observations
- ➢ Model output provided useful spatial pattern information that was applied as an improved precipitation basemap (from which PMP begins)
- ➢ Also helped to inform rain-snow delineation to identify regions with snow (not/less relevant to PMP)
- ➢ Discussion of use of single simulations vs. ensemble diagnostics (spread, ensemble max)

Savageton, WY 1924

- ➢ Mid-latitude synoptic cyclone; Gulf of Mexico moisture
- ➢ Existing storm analysis details uses questionable observations: "…limited number of hourly and daily data near primary small storm center diminish reliability of these results…there were only 5 hourly stations…data were estimated from USACE's smoothed mass rainfall curves…"
- ➢ 14+ WRF simulations only achieve an ensemble maximum point value of ~4 inches (50mm) vs. 17.1 inches in northeastern WY
- ➢ AWA: "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 analysis."
- ➢ Is the WRF ensemble deficient in its ability to simulate this case? Or could the 17.1 inch/108 hour observation be in error?

Additional cases simulated:

➢ **Ward District, CO 1894**
- Synoptic cyclone (96-hour duration), relatively good data coverage (43 stations)
- For Ward District, the WRF precipitation (based on maximum value of 4-member runs) were as a SPAS basemap, the WRF basemap had similar spatial pattern (slightly shifted to the southeast) and magnitude as the observed data.

➢ **Penrose, CO 1921**
- Small-scale thunderstorm with severe lack of data. Only one station was used in the original analysis.
- A 4-member WRF ensemble produced spatial output similar to the original spatial analysis, the WRF storm center was shifted to the east approximately 10 miles and the magnitude was substantially less than the observed data, therefore no adjustment/basemap updates were applied.

Ensemble vs. single case considerations

- Individual simulations offer sharpest spatial, temporal detail and are relevant to actual case evolution
- Ensemble maximum grids are relevant to PMP estimation from the viewpoint of "how hard did the dynamical model think it could have rained given this overall set-up?"
- Also, ensemble spread shows variability in model solutions and gives some indication of uncertainty in robustness of simulations.

Summary & next steps

Dynamical modeling offers potential benefit to PMP estimation:

- ➢ Reconstruction of major historical events via numerical modeling may supplement existing storm analyses, improve spatial and temporal assumptions made with very limited observational data
- ➢ Utility of WRF simulations performed for this study seems proportional to role of topography in controlling rainfall spatial pattern and magnitude.
- ➢ More generally, dynamical models:
  - Offer continuity in space and time; data produced by solving physical equations of the atmosphere (vs. interpolation of limited observations)
  - Reduce need for many spatial, temporal, physical assumptions (e.g., storm transposition, storm templates, moisture maximization, etc.)
  - Especially important in data-sparse regions of complex (& high-elevation) topography
  - Have relatively straightforward methods for quantification of uncertainty
  - Can offer numerically-constrained/somewhat-objective methods for storm maximization, atmospheric boundary condition shifting, etc. (Ohara et al. 2011; Ishida et al. 2015; Chen and Hossain 2016, others)
- ➢ Dynamical modeling not a PMP improvement panacea – using dynamical models, particularly in this manner, still involves substantial subjectivity.
- ➢ Ongoing work will more closely examine the reliability of the Savageton, WY 1924 event, along with several other events deemed critical by Colorado and New Mexico dam safety programs.
- ➢ Future work will leverage new model and reanalysis technologies to be more representative of possible simulation spread and more specifically communicate uncertainty in stakeholder-relevant ways.

This project is supported by grants from the Colorado Water Conservation Board, the New Mexico Office of the State Engineer, the Albuquerque Metropolitan Arroyo Flood Control Authority, the New Mexico Watershed and Dam Owners Coalition, the Western Water Assessment, NOAA Earth System Research Laboratory/Physical Sciences Division, and the support of subject matter experts provided by the state and federal entities identified above for review of the scientific efforts being undertaken.