EVALUATION OF THE HIGH-RESOLUTION RAPID REFRESH AND THE COUPLED OCEAN-ATMOSPHERE MESOSCALE PREDICTION SYSTEM DURING ATMOSOPHERIC RIVER EVENTS IN CALIFORNIA

by

Kevin James Dougherty

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Atmospheric Sciences

The University of Utah

August 2020

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of **Kevin James Dougherty**

has been approved by the following supervisory committee members:

John D. Horel	, Chair	**5/27/2020** Date Approved
Jason Nachamkin	, Member	**5/20/2020** Date Approved
W. James Steenburgh	, Member	**5/18/2020** Date Approved

and by **John D. Horel** , Chair/Dean of

the Department/College/School of **Atmospheric Sciences**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

The accuracy of precipitation forecasts from the High-Resolution Rapid Refresh model (HRRR) of the National Centers for Environmental Prediction and the Navy's Coupled Ocean Atmosphere Mesoscale Prediction System (COAMPS) is examined during atmospheric river (AR) events in California. ARs are long, narrow, and transient corridors of strong horizontal water vapor transport that produce substantial amounts of precipitation and often result in flooding along the west coast of the United States and other coastal regions of the world. The initial 12-h precipitation forecasts from HRRR and COAMPS are validated relative to Stage-IV gridded precipitation analyses for an AR event during 6-8 December 2019 and during the entire 1 December 2018 – 28 February 2019 period.

The 6-8 December 2019 case study examines the integrated water vapor transport (IVT) analyzed and forecasted by the two models. COAMPS predicted weaker IVT impinging upon the California coast than HRRR, yet forecasted higher precipitation totals. Both models underestimated precipitation amounts along the coast to the north of San Francisco that may have arisen due to more frequent forecasts of winds directed slightly offshore than observed.

During the 2018-19 winter, both models had higher precipitation accuracy during AR events than during non-AR events, which are characterized by lower precipitation totals and generally less organized synoptic setting. Overall, COAMPS exhibited very

large wet biases over interior mountainous regions while HRRR biases were much smaller in those regions. Both models tended to underestimate precipitation along the coastal mountains of northern California. Based on the Fraction Skill Score (FSS) metric applied separately over specific ranges, HRRR had higher precipitation forecast skill compared to COAMPS, particularly within distances of 30-40 km and moderate 12-h precipitation totals of 10-50 mm.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would not have been able to complete my master's thesis without the support of several people. First, I would like to thank my advisor, Dr. John Horel, for his continued support and guidance throughout my entire graduate school experience. I also want to thank him for supporting me to attend the AMS Annual Meeting in Boston, MA as well as allowing me to spend a summer as an intern at the Naval Research Lab in Monterey, CA. I would also like to thank all of my fellow Mountain Meteorology group members for their help in classwork as well as programming issues. I also want to thank my committee member Dr. Jim Steenburgh for being able to answer all my very specific weather questions as well as being one of the greatest professors I have had the pleasure of having. A special thank you also to Dr. Jason Nachamkin at the Naval Research Lab for not only being a great mentor, committee member, and supplying all the COAMPS data needed to complete this research, but for supplying me with some of the best weather knowledge and hikes along the central California coast.

This work was supported by the Collaborative Science, Technology, and Applied Research (CSTAR) Program Award (55500146). Stage-IV data access was provided by the Earth Observing Laboratory web service. Computational hardware and software were provided by the University of Utah Center for High Performance Computing (CHPC) to complete this research.

CHAPTER 1

INTRODUCTION

California has the largest interannual variability in precipitation of any state, receiving almost the entirety of its yearly precipitation in the winter months, with 25-50% of the state's rainfall and snowpack being directly related to atmospheric rivers (Dettinger et al. 2011; Ralph et al. 2013; Rutz et al. 2014). Atmospheric rivers (ARs) are long, narrow, and transient corridors of strong horizontal water vapor transport that are typically associated with a low-level jet ahead of the cold front of an extratropical cyclone (American Meteorological Society 2018; Ralph et al. 2018). ARs are associated with 92% of the West Coast's heaviest 3-day rain events (Ralph and Dettinger 2012), are responsible for 40-90% of major floods in West Coast rivers (Ralph et al. 2006; Neiman et al. 2011; Konrad and Dettinger 2017), and cause 68% of postfire debris flows in Southern California (Oakley et al. 2010; Young et al. 2017; Hatchett et al. 2017). However, ARs also supply nearly half of California's water supply, end droughts in the West Coast, and provide water to wetlands, floodplains, and fisheries to allow the local ecology to thrive (Dettinger 2013; Florsheim and Dettinger 2015).

Several studies have evaluated how well operational numerical weather prediction (NWP) models forecast AR events on a synoptic scale (Wick et al. 2013; Nayak et al. 2014, Lavers et al. 2016, Nardi et al. 2018). Synoptic models have been shown to forecast

integrated water vapor transport (IVT) and AR location with good skill up to a week. As forecast lead time increases beyond that, forecast skill drops considerably. At shorter forecast lead times (1-3 days), the model's location errors reduce to roughly 100 km from actual landfall, which is important for flood forecasting. However, forecasts of the intensity and temporal evolution of AR events can change significantly in the 72-h period before landfall (Martin et al. 2019). Being able to provide accurate forecasts with as much lead time as possible is essential for public safety personnel to take the appropriate actions to protect life and property.

Although AR events are associated with widespread precipitation accumulations, flooding during these events usually takes place at local scales and weather conditions can change rapidly. An example where flooding frequently occurs is the Russian River Basin in northern California (Ralph et al. 2006; Dettinger et al. 2011; Cao et al. 2019), which is embedded within the complex terrain of the coastal mountains where landfalling ARs interact with topography, resulting in enhanced precipitation (Neiman et al. 2002; Colle 2004; Kunz and Wassermann 2011; Picard and Mass 2017). Millions of people are also at risk during high-impact precipitation events in the San Francisco Bay region immediately south of the Russian River Basin (Bridger et al. 2019).

Gowan et al. (2018) showed that high-resolution models outperformed operational models with poorer resolution for wintertime precipitation in the western contiguous United States (western CONUS). Therefore, it is important to evaluate how well high-resolution NWP models forecast these events at short time periods. The skill of high-resolution models for AR events has been of high interest, particularly since the transition of weather systems as they move onshore presents challenges for forecast models (Wick

et al. 2013; Swain 2015; Martin et al. 2019; Bridger et al. 2019).

This study addresses the ability of two mesoscale models to forecast AR and non-AR event precipitation in the California region during the 2018-19 winter season: (1) the National Centers for Environmental Prediction's (NCEP) High-Resolution Rapid Refresh model (hereafter referred to simply as HRRR) and (2) the Naval Research Lab's (NRL) Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS). Comparisons are made of forecast skill from those two models that continue to undergo improvements to the frozen operational North American Mesoscale Forecast System (NAM-3km) of the NCEP (Rogers et al. 2017).

HRRR is an hourly updated, convection-allowing model (Benjamin et al. 2016). Its 3-km grid spacing provides high-resolution detail over the CONUS at over 1.9 million grid points. HRRR analysis and selected forecast fields are available via a publicly accessible archive maintained at the University of Utah (Blaylock et al. 2017). That archive has been used in several dozen studies (e.g., McCorkle et al. 2018; Blaylock et al. 2019; Blaylock et al. 2020; Moore et al. 2020). Darby et al. (2019) suggest that HRRR has a tendency to underpredict precipitation along the west coast based on their analysis for the 2015-2016 winter.

COAMPS is the Navy's high-resolution, convection-allowing model that has 4-km grid spacing (Hodur 1997). COAMPS is primarily used operationally for defense applications pertaining to naval operations, but model output is also publicly available for a coastal California domain. COAMPS regional California domain studies have tended to focus on air-sea interactions, including the system's ability to forecast oceanic coastal circulations, offshore surface winds and low clouds, marine boundary layer structure, and

sea surface temperatures (e.g., Haack and Burk 2001; Hsu et al. 2007; Haack et al. 2008; Neveu et al. 2016). A recent study by Stone et al. (2020) focused on the improvement of forecasts during AR events along the west coast for the Navy Global Environmental Model (NAVGEM), COAMPS parent model, by adding in-situ observations from dropsondes. However, there appear to have been no published studies that validate COAMPS for regional California precipitation. COAMPS has been shown to poorly predict precipitation structure and duration compared to observations as well as errors in forecasted precipitation and wind speeds in areas of complex terrain (Nachamkin and Jin 2017; Reynolds et al. 2019; Doyle et al. 2019).

The objective of this study is to contribute to future model improvements for HRRR and COAMPS by diagnosing the ability of each modeling system to accurately forecast precipitation during AR events relative to that forecast during non-AR precipitation related events in the California region. The complexity of terrain-flow interactions and orographic enhancement in the coastal and interior regions of the state are challenging for operational models and relevant to other regions for which the models are used. To address this objective, several steps are taken:

1. Validate Stage-IV gridded precipitation analyses produced by the NCEP (Lin and Mitchell 2005) relative to observed precipitation at hundreds of locations in the region;
2. Examine in greater detail the Integrated Water Vapor Transport (IVT) analyzed and forecasted by HRRR and COAMPS during the 6-8 December 2019 AR event for which more complete sets of forecast fields are available;
3. Validate the initial 12-h precipitation forecasts (F01-F12) from HRRR and

COAMPS relative to the Stage-IV gridded precipitation analyses during the 6-8 December 2019 AR event and entire December 1, 2018 – February 28, 2019 period; and

4. Evaluate the use of a Fraction Skill Score (FSS) metric applied separately over specific ranges of precipitation values to assess the relative model performance as a function of precipitation intensity and spatial scales within which the precipitation has fallen.

The details of the precipitation observations, Stage-IV precipitation analyses, and operational models used for this research are outlined in Chapter 2. Validation of the Stage-IV gridded analyses, the 6-8 December 2019 AR case study, and cumulative statistics from the F01-F12 precipitation forecasts from HRRR, COAMPS, and NAM-3km during the 2018-2019 season follow in Chapter 3. A summary of this research and future work that could be undertaken are found in Chapter 4.

CHAPTER 2

DATA AND METHODS

2.1 Operational Models

2.1.1 HRRRv3

HRRR was developed by the Earth Systems Research Lab (ESRL) and is run operationally by the Environmental Modeling Center of the NCEP. It is a convection-allowing, hourly updated model with 3-km grid spacing over the CONUS, as shown in Figure 2.1a (Benjamin et al. 2016). The model produces forecasts from F00 to F18 every hour except at 0000 UTC, 0600 UTC, 1200 UTC, and 1800 UTC when the forecasts extend from F00 to F36. HRRR uses the NOAA Gridpoint Statistical Interpolation (GSI) data assimilation process (Wu et al. 2002; Whitaker et al. 2008; Kleist et al. 2009) as well as an assimilation of radar reflectivity every 15 minutes over 1 hour (Benjamin et al. 2016). Model characteristics are summarized in Table 2.1. HRRR output is available from the Pando archive system at the University of Utah's Center for High-Performance Computing described by Blaylock et al. (2017a).

2.1.2 COAMPS

COAMPS was developed by the Marine Meteorological Division of the NRL and is run operationally by the Fleet Numerical Meteorology and Oceanography Center. The model uses 4-km grid spacing and can be applied as a nested grid anywhere on the globe (Hodur 1997). COAMPS can be configured as standalone atmosphere or ocean forecast model or run as a coupled modeling system. This research relies on coupled model simulations run routinely over a Northern and Central California domain with forecasts extending out to 48 h (Fig. 2.1). The model is initialized daily at 0000, 0600, 1200, and 1800 UTC using the Naval Research Laboratory's Atmospheric Variational Data Assimilation System (NAVDAS; Daley and Barker 2001) and includes a vertical domain of 30 sigma-z levels from 10 m to approximately 30 km. Model characteristics are summarized in Table 2.1. COAMPS output were downloaded from NRL disk servers. However, the 0600 and 1800 UTC model runs are not saved, so only the 0000 and 1200 UTC forecasts are available for this research.

2.1.3 NAM-3km

NAM-3km is also run operationally by the NCEP and underwent its final upgrade in 2017 (Rogers et al. 2017). NAM-3km is used in this research to compare the skill of HRRR and COAMPS to a frozen model. NAM-3km has a horizontal resolution of 3km over the CONUS, matching the same domain as HRRR. It is initialized daily at 0000, 0600, 1200, and 1800 UTC and produces hourly forecasts out to 60 h. Many of the model specifications for NAM-3km are the same as HRRR as summarized in Table 2.1. Selected fields from the NAM-3km are downloaded from the NOAA Operational Model

Archive and Distribution System (NOMADS) and saved locally on servers maintained by the Center for High Performance Computing.

This study focuses on the December 2018 - February 2019 season that was a very active period for ARs in California. Within that 3-month period, there were eight AR periods ranging from weak to extremely strong events according to the AR scale proposed by Ralph et al. (2019). Figure 2.2 shows the average Stage-IV 12-h measured precipitation in the research domain from 0000 UTC 1 December 2018 to 12 UTC 28 February 2019. The periods in red indicate times when AR plumes with IVT values greater than 250 $kg\ m^{-1}\ s^{-1}$ affected the California region. AR events are extended up to 12 h if extensive precipitation continues in the region where IVT exceeded 250 $kg\ m^{-1}\ s^{-1}$ during the previous 12 h-period. The periods in blue indicate times (often preceding or following the AR periods) when precipitation did not appear to be related directly to ARs. For 12-h periods where total accumulated precipitation in the research domain was on average greater than 0.01 inches (0.254 mm), there were 53 periods where 12 h accumulated precipitation were subjectively assigned to take place during AR events and 49 periods where 12 h accumulated precipitation was not related to ARs.

2.2 Precipitation Observations and Analysis Data

Surface observations were accessed from Synoptic Data PBC, which is providing improved access to environmental observations from the approach originally developed by MesoWest (Horel et al. 2002). Synoptic Data's Mesonet Application Programming Interface provides real-time and archived data from over 50,000 surface observation stations in North America. California is an ideal place for this research as there are over

4000 active observing sites within its borders. This research relies on networks that tend to have precipitation observations appropriate for this study: National Weather Service (NWS), Remote Automated Weather Stations (RAWS), and Hydrometeorological Automated Data System (HADS). A total of 1573 stations from these three networks are available in the research domain shown in Figure 2.3. NWS sites are primarily at airports to support aviation applications and rely on heated tipping buckets supplemented by manual observations at some locations. The RAWS network is intended to monitor warm-season fire weather conditions for which heated tipping buckets are generally not required (Zachariassen et al. 2003; Horel and Dong 2010). Hence, precipitation reported during winter from RAWS stations at high elevation are not used. The HADS network supports the NWS Weather Forecast Offices (WFO) for the Flood and Flashflood Warning programs as well as River Forecast Center (RFC) operations. In addition, the network aids in fire weather support, analysis of precipitation events, hydrological and meteorological modeling, and the verification of NEXRAD precipitation estimates (Kim et al. 2009). HADS produces hourly precipitation data and uses a variety of rain gauge types depending on the contributing agency, but generally are either weighing or tipping buckets.

Automated and manual quality control steps were applied to hourly interval precipitation observations and only stations that reported all 180 12-h increments from 0000 UTC 1 December 2018 to 0000 UTC 1 March 2019 were included. Those steps led to a reduction from the 1573 available stations to a total of 864 stations, as shown in Figure 2.3: 101, 427, and 336 stations from NWS, RAWS, and HADS networks, respectively.

The 6 h Stage-IV precipitation gridded analyses produced by the NCEP (Lin and Mitchell, 2005) are a valuable resource for evaluating HRRR, COAMPS, and NAM forecasts. These precipitation analyses are a national multisensor analysis at 4-km grid spacing over the CONUS that is blended from River Forecast Center (RFC) 1-h and 6-h precipitation. The California Nevada RFC and the Northwest RFC products cover the research domain and use the PRISM/Mountain Mapper approach to produce gauge-based analyses (Hou et al. 2014).

2.3 Verification Methods

Statistical metrics based on the standard 2 × 2 contingency table (Table 2.2) are used to evaluate deterministic precipitation forecasts, including:

$$\text{Hit rate } = \frac{a}{a+c} = \frac{\text{hits}}{\text{observed events}},$$

$$\text{Frequency bias} = \frac{a+b}{a+c} = \frac{\text{forecasted events}}{\text{observed events}},$$

$$\text{False alarm ratio} = \frac{b}{a+b} = \frac{\text{false alarms}}{\text{forcasted events}},$$

$$\text{Equitable Threat Score (ETS)} = \frac{a - a_{\text{ref}}}{a - a_{\text{ref}} + b + c},$$

where

$$a_{\text{ref}} = \frac{(a+c)\times(a+b)}{n}.$$

The hit rate, which is also known as the probability of detection (POD), is the fraction of occurrences that were correctly forecasted. The frequency bias is the ratio of the total number of events forecasted to the total number of events observed. The false alarm ratio is the fraction of forecasts that were predicted, but did not occur. The ETS is the

proportion of observed and/or forecasted events that were correctly forecasted, adjusted for the frequency of hits expected by chance (climatology).

While the above statistical measures provide good insight on model skill, they are point-based statistics and do not evaluate how forecast skill varies spatially. The fractions skill score (FSS) is a verification method that measures how forecast skill varies with spatial scale for different precipitation thresholds (Roberts and Lean 2008; Blaylock and Horel 2020). The main advantage of FSS over point-based verification is that it is less sensitive to localized errors and provides information at which spatial scales model forecasts are skillful (Wolff et al. 2014).

The FSS is computed as one minus the fractional area mean square error within each neighborhood throughout the verification domain divided by the maximum possible mean square error in each of those neighborhoods. FSS is computed iteratively for increasing neighborhood sizes and calculated independently for every neighborhood throughout the entire domain. The FSS metric has an important advantage because it evaluates how forecast skill varies with neighborhood size and what neighborhood sizes provide the most skillful forecasts (Mittermaier and Roberts 2010). Traditionally, the FSS is calculated using a square neighborhood for simplicity. Relying on square neighborhoods can affect the results if features are located near the corner of a square (Skok 2016; Skok and Roberts 2016). Hence, FSS is calculated here using circular neighborhoods with radii from 3 km (a signle gridpoint) to 70 km. The FSS ranges from zero, which is a forecast with no skill, to one, which is a forecast with perfect skill. As the neighborhood size is increased, FSS increases and asymptotes for an unbiased forecast to one indicating a perfect score. When biases exist, the score will asymptote below one

(Roberts and Lean 2008).

As discussed by Blaylock and Horel (2020), FSS assesses location errors of forecasts. FSS provides no information on the spatial structure of precipitation, but it does compare the relative coverage of the precipitation observed and forecasted as a function of spatial scale—FSS increases as observed and forecasted precipitation overlap more. High FSS values indicate useful forecasts when computed within small distances. As the spatial scale is increased, it is much more likely that the predicted precipitation will overlap that observed. While the FSS is very useful to determine how forecast skill varies with neighborhood size, it is similar to the ETS metric in that a skillful forecast is judged to be one simply greater than or equal to a given threshold. That approach inhibits assessing whether the model forecast is much greater than the threshold within that neighborhood. To address this issue, FSS within threshold ranges is also used to evaluate model accuracy of precipitation intensity.

Table 2.1. Characteristics of forecast models.

Model	Domain	Grid Points	Grid Spacing (km)	Vertical Levels	Vertical Coordinate
HRRR	CONUS	1799x1059	3	50	Sigma-Isob Hybrid
COAMPS	Central and Northern California	361x301	4	60	Sigma
NAM-3km	CONUS	1799x1059	3	60	Hybrid sigma-pressure

Model	Initialized	Assimilation	Radar DA	Microphysics	Radiation
HRRR	Hourly	GSI Hybrid Ensemble to 0.85	Yes	Thompson and Eidhammer (2014)	RRTMG
COAMPS	0000, 0600, 1200, 1800 UTC	NAVDAS (Daley and Barker 2001)	No	Schmidt (2001)	Fu-Liou: Liu et al. (2003)
NAM-3km	0000, 0600, 1200, 1800 UTC	GSI Hybrid Ensemble	Yes	Ferrier-Aligo (2018)	RRTM

Table 2.2. Contingency table used for validation.

		Observed		
		Yes	No	Total
Forecast	Yes	Hit (*a*)	False Alarm (*b*)	*a* + *b*
	No	Miss (*c*)	Correct Rejection (*d*)	*c* + *d*
	Total	*a* + *c*	*b* + *d*	*n*

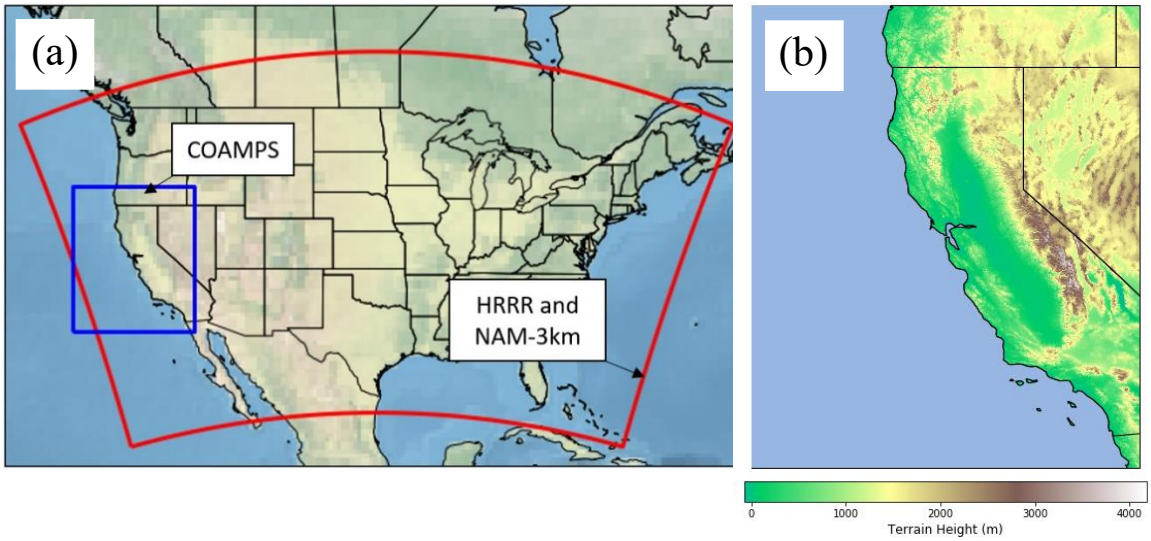

Figure 2.1. Model domains and topography of research area. (a.) HRRR and NAM-3km CONUS domains in red. COAMPS domain in blue. (b.) COAMPS domain with topography based on the color bar below.

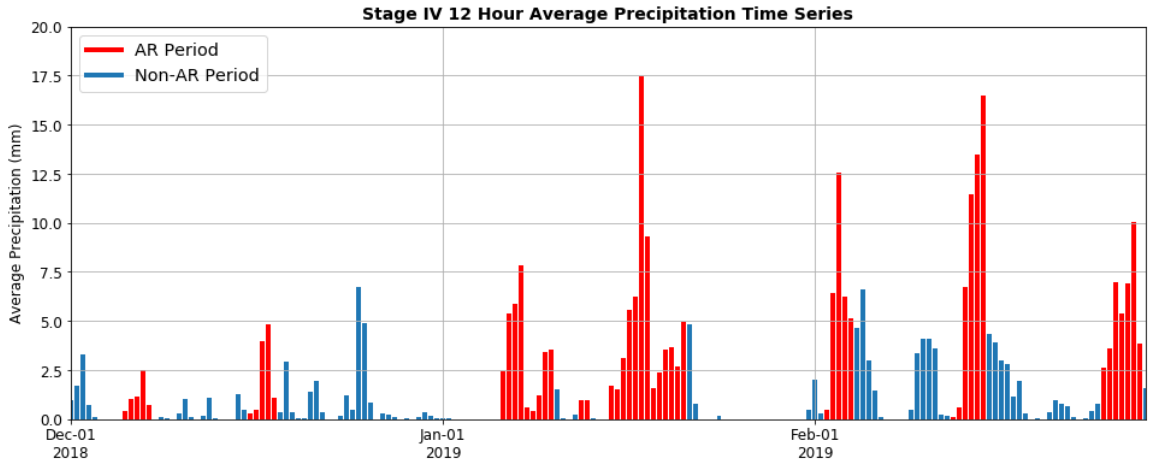

Figure 2.2. Average precipitation (mm) within the analysis domain from Stage-IV 12-h analyses from 0000 UTC 1 December 2018 to 1200 UTC 28 February 2019. Periods in red (blue) indicate precipitation related to an AR (non-AR) event.

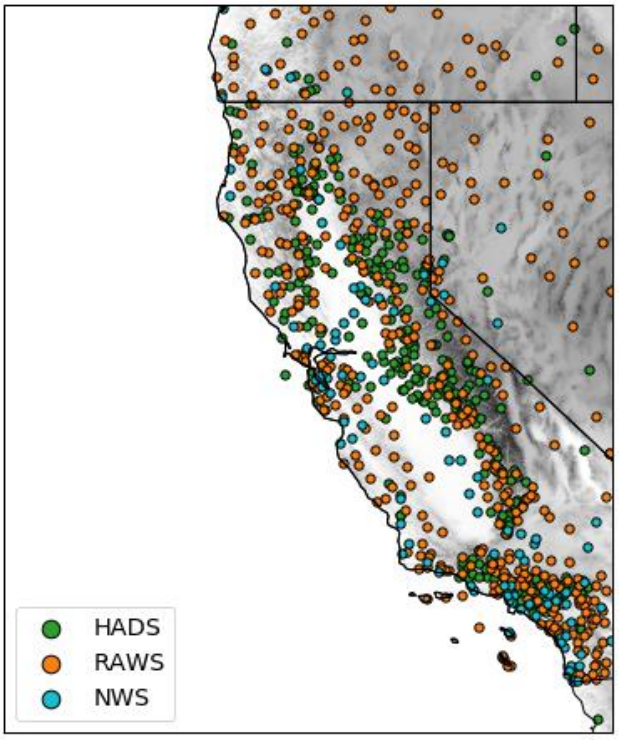

Figure 2.3. Station locations for NWS, RAWS, and HADS networks.

CHAPTER 3

RESULTS AND DISCUSSION

To assess the utility of the Stage-IV precipitation analyses for model verification in the California region, the analyses are compared in Section 3.1 to NWS, RAWS, and HADS station precipitation during the DJF 2018-19 season. A detailed analysis of an AR event during 6-8 December 2019 is then used to illustrate observed and forecasted characteristics of AR events. The case study affords a more detailed investigation of the well-recognized relationships between IVT and rainfall during AR events and the extent to which HRRR and COAMPS forecast those quantities. Finally, HRRR, COAMPS, and NAM-3km precipitation forecasts during AR and non-AR events are validated using the Stage-IV precipitation analyses during the 2018-19 DJF season.

3.1 Stage-IV Analysis Evaluation

According to the California Department of Water Resources (2019), statewide precipitation amounts during the 2018-2019 water year (1 Oct 2018 – 30 Sep 2019) were better than many other recent years leading to the state's snowpack on 1 Apr 2019 being 175 percent of average and statewide reservoir storage ending 128 percent of average. However, before 1 February 2019, the precipitation across the state was below average. Three strong AR periods during February (Fig. 2.3) reversed that situation and

precipitation increased to 124% of average for the state by the end of the month. Figure 3.1 shows the 12-h average for precipitation events during the DJF 2018-19 season omitting periods when no-little precipitation fell. Large amounts of precipitation (>5 mm per 12-h period) fell as in the coastal ranges, interior northern and Sierra Nevada mountains, and highest elevations of southern California. Some areas along the northern California and southern Oregon coast received amounts greater than 10 mm (0.394 in) per 12-h period according to the Stage-IV analyses (Fig. 3.1b).

As mentioned in the previous chapter, 864 precipitation measuring stations were judged after automated and subjective quality control to provide reasonable 12-h precipitation totals during this period. However, that number of sites is still insufficient to evaluate in great detail the model precipitation grids at 3-4 km horizontal resolution across California. Hence, the NWS, RAWS, and HADS station values are used here to corroborate that the Stage-IV analyses are reasonable for model validation with the understanding that the analyses are created in part by relying on these station reports. Of greatest interest is to determine whether there are any substantial under- or over-estimates of precipitation amounts in the Stage-IV analyses that would bias the interpretation of the verification of the model forecasts.

Figure 2.3 highlighted that precipitation in California is episodic with many 12-h periods during which little precipitation is observed across the state. The periods with largest statewide precipitation amounts are nearly all associated with AR events. Figure 3.2 shows the average bias, RMSE, and bias ratio computed from all 12-h periods during the season between the Stage-IV analysis totals interpolated to the station locations and the corresponding station observations. As shown in Figure 3.2a, the Stage-IV analyses

generally are within 1mm on average of the observations at the lower elevation locations where the station totals for the season are small (Fig. 3.1). RMSE values in those locations tend to be below 3 mm while bias ratios are often relatively close to 1 (Fig. 3.2b,c). There are fewer locations where the analyses tend to be substantially lower than the station values (Fig. 3.2a,c). However, there are more locations in the coastal ranges and higher elevations of the interior mountains where the Stage-IV analyses are greater than those observed with larger RMSEs than those at lower elevations. These results are not entirely unexpected as previous studies have shown that Stage-IV analyses tend to slightly overestimate observations in the California and Nevada RFC during DJF (Nelson et al. 2015). Concerns about the representativeness of station observations at high elevation and the many sources of uncertainty for both the observations and analyses in such regions are well recognized (Henn et al. 2017).

Not surprisingly, differences between the analyses and station observations during specific 12-h periods (Fig. 3.3) are unlikely when the statewide precipitation totals are small (Fig. 2.3). Larger discrepancies occur when there are larger precipitation totals arising from AR- or non-AR events during February 2019 (Fig. 3.3). Discrepancies between analysis values at NWS sites tend to be smaller, likely due to their locations in areas receiving less total precipitation as well as more weight given to those observations in the Stage-IV analyses. However, during some 12-h periods within AR events, the bias and RMSE values of the analyses relative to NWS stations are not substantially different from those obtained at RAWS sites, many of which are also at lower elevation or in relatively close proximity to urban areas.

The largest biases and RMSE scores are evident in Figure 3.3 between the

analyses and HADS observations during the major AR events. As shown in Figure 2.1, HADS stations are located throughout the domain for water management and flood control, some along rivers and streams at lower elevation as well as many at strategic locations at higher elevation. Considerable time was spent evaluating individual HADS stations relative to other neighboring stations and the Stage-IV analyses. In total, 250 HADS stations were eliminated through manual quality control steps (e.g., identifying very low precipitation totals when temperatures were below freezing and other nearby stations reported large snow totals). Even after these quality control steps were applied, there are several 12-h periods during AR events where the discrepancies between HADS observations and the Stage-IV analyses remain large. However, it will be shown later that the model forecasts tend to be overestimates relative to the Stage-IV analyses over high terrain. Hence, the errors exhibited by the model forecasts are even greater when compared against the HADS observations directly (not shown).

Based on the results of this subsection and additional evaluation of the Stage-IV analysis products relative to the available observations, the Stage-IV analyses are judged to be very useful for evaluating the model precipitation forecasts during AR and non-AR events in California. Hence, the Stage-IV analyses will be used for validation of model forecasts of precipitation in the remaining two sections.

3.2 6-8 December 2019 Case Study

A large midlatitude cyclone traversing the eastern Pacific Ocean led to AR conditions from 6-8 December 2019 across most of California. A large plume of moisture associated with the cyclone stretched from Hawaii to California (not shown), bringing

more than 100 mm (3.64 in) of precipitation to coastal and interior mountain ranges in northern California (Fig. 3.4). Localized areas in the northern Coast Range received more than 150 mm (5.9 in) of accumulated precipitation during the 72-h period between 1200 UTC 5 December and 1200 UTC 08 December 2019 (Fig. 3.4f). Another notable feature is the large precipitation amounts to the north of the Central Valley, which will be referred to as the Mt. Shasta area. The coastal ranges south of San Francisco also saw substantial accumulations of precipitation with lesser amounts in southern California. Areas in the Sierra Nevada mountains above 2500 m received between 30-60 cm (1-2 ft) of snow. Two Flood Advisories were issued by the Sacramento NWSWFO for Shasta and El Dorado counties during the afternoon of 07 December 2019 when the heaviest precipitation rates were observed (Castellano et al. 2019). This storm resulted in localized flooding throughout Northern California including flooded houses in a San Francisco neighborhood and rockslides along coastal highways south of San Francisco.

The 449 MHz wind profiler (BBY) maintained by the ESRL Physical Sciences Division is located near Bodega Bay, CA (red star in Fig. 3.4a). Accumulated precipitation amounts in the coastal range immediately to the north of BBY were the largest observed in the state during this event. Figure 3.5 summarizes the vertical structure of wind speed and direction based on hourly observations at BBY and hourly analyses at HRRR and COAMPS analyses available only every 12-h interpolated to that location. The median and interquartile range (25^{th} to 75^{th} percentile) wind speed and direction at each level were computed for the 72-h period from 1200 UTC 5 Dec to 1200 UTC 8 Dec 2019 for BBY and HRRR. COAMPS 25^{th} and 75^{th} percentiles were not calculated because only six analyses are included during this timeframe. On the basis of

the BBY profiler data, a low-level southerly jet in the 500 m to 1 km MSL layer was common during this event with wind speeds typically greater than 15 m s^{-1}. The sustained southerly fetch throughout the lowest 5 km during this 72-h period is remarkable and provides an indication of substantive flow impinging on the orography of the coastal ranges further north. HRRR F00 analyses are quite similar to the BBY winds, which suggests that those wind observations were being assimilated into the model data assimilation cycle. Some differences between the observed and analyzed winds are: HRRR analyzed winds tend to be weaker in the lowest 300 m than observed and HRRR wind directions tend to be more southerly than south-southeasterly throughout the lowest 5 km. COAMPS analyzed winds are even weaker than the HRRR in the lowest 300 m, and the wind directions tend to be more southeasterly than south-southeasterly.

As would be expected during an AR event (Backes et al. 2015; Ralph et al. 2017), the magnitudes of HRRR analyzed vertical moisture flux profiles each hour during this 72-h period tend to peak in the layer corresponding to the low level jet (Fig. 3.6). Hence, it should be evident that the contributions to the density-weighted IVT of this northerly-directed flow are dominated during this event by the moisture flux below 1.5 km. IVT calculated for this location from HRRR analyses was on average in excess of 317 kg m^{-1} s^{-1} over this 72-h period. COAMPS vertical moisture fluxes are much weaker at lower levels compared to HRRR fluxes.

Figure 3.7 shows HRRR IVT analyses every 12 h from 0000 UTC 06 December through 1200 UTC 08 December 2019. In the context of these 12-h snapshots, IVT peaked at 0000 UTC 7 December 2019 with values impinging on the California coast to the north of San Francisco in excess of 650 kg m^{-1} s^{-1}. Based on the real-time analyses of

AR events maintained by the Center for Western Weather and Water Extremes, this event was classified as a AR3 level event since AR conditions lasted for longer than 48 h with a maximum IVT value along the CA coast greater than 500 $kg\ m^{-1}\ s^{-1}$ (Ralph et al. 2019). Large IVT values continued for another day after 0000 UTC 7 December in northern California with extensive precipitation continuing in that region through 1200 UTC 8 December (compare Fig. 3.4e and 3.4f). For that reason, this case is an example of our approach to extend by up to 12-h the lifetime of an AR event if extensive precipitation continues in the region where IVT exceeded 250 $kg\ m^{-1}\ s^{-1}$ during the previous 12-h period.

As a summary of the IVT throughout the 72 h event, Figure 3.8a shows the average IVT based on the hourly HRRR IVT analyses. The largest IVT remained offshore and northwest of where the precipitation was greatest, which suggests that this event would have had even greater impact on northern California if the moisture plume had shifted a bit further south and east. High values of IVT also were analyzed to have penetrated inland into the northern part of the Central Valley, which is consistent with the large precipitation amounts in the Mt. Shasta area. Figure 3.8b shows the same for the 12-h COAMPS IVT analyses. Comparing to HRRR analyses, COAMPS IVT analysis is significantly weaker. However, elevated IVT values were present as well in the northern part of the Central Valley. COAMPS also had slighter more IVT >250 $kg\ m^{-1}\ s^{-1}$ off the southern coast of California.

Since the largest IVT values along the coast occurred on 0000 UTC 07 December 2019 (Figs. 3.7c and 3.9c), the ability of HRRR and COAMPS to forecast these conditions 24 h and 12 h in advance are shown in Figure 3.9. The overall orientation and

strength of IVT from both forecasts from the HRRR tend to be quite good. All panels show the dominant moisture plume paralleling the northern California coast with a weaker IVT plume impinging upon southern California. However, subtracting the forecasts from the analysis (Fig. 3.9c) highlights some important differences (Fig. 3.10). HRRR F24 has a more concentrated area of IVT making landfall north of San Francisco, with much weaker IVT values indicated by darker blue shading along the northern California coast. The F24 forecast also has much higher values of IVT penetrating inland into the Central Valley compared to HRRR analysis (Fig. 3.10a.). The F12 forecast improves upon the F24 forecast slightly by shifting the area of maximum IVT north, however the IVT forecast is still too weak along the northern coast. The F12 also extends stronger IVT values south and west over the Pacific Ocean. In addition, IVT values are still forecasted much higher inland than compared to the F00 analysis (Fig. 3.10b).

The primary IVT plume analyzed by COAMPS at 0000 UTC 7 December 2019 has a similar orientation but is much weaker than that analyzed by HRRR (compare Fig. 3.9c and 3.9f). COAMPS does not analyze values of IVT greater than 250 $kg\ m^{-1}\ s^{-1}$ making landfall across the southern California coast. Relative to the COAMPS analysis valid at that time, the COAMPS F24 forecast tended to have the IVT plume too far west while the COAMPS F12 forecast had much stronger IVT across northern California and southern Oregon and immediately offshore central California (Fig. 3.10c and 3.10d).

The analyses and forecasts of IVT from the HRRR and COAMPS provide a framework for evaluating the precipitation forecasts from those models (Figs. 3.11 and 3.12). Both models captured the general precipitation features associated with this AR event with well-defined maxima along the northern Coast Range, Mt. Shasta area, and

Sierras that correspond to the impact of the IVT plume along the northern coast and the IVT penetrating inland. Lesser precipitation amounts are forecasted to develop during the latter stages of the event in southern California. However, HRRR forecasts tended to under forecast the orographic precipitation in the coastal ranges and Mt. Shasta area as the AR event evolved (e.g., compare Figs. 3.4d and 3.11d) while COAMPS tended to overpredict precipitation amounts in those areas as well as the Sierra Nevada (e.g., compare Figs. 3.4d and 3.12d).

The differences in precipitation forecasted in the first 12 h from HRRR and COAMPS relative to the Stage-IV analyses are shown for the entire 72-h AR event in Figure 3.13. HRRR underpredicted the cumulative precipitation amounts along the northern coast where the largest totals were observed and also in the Mt. Shasta area (Fig. 3.13a). However, COAMPS exhibited a large wet bias in the interior sections of the coastal ranges, Mt. Shasta area, northern Sierras, and swaths within the northern Central Valley (Fig. 3.13b). Both models consistently underpredicted precipitation amounts immediately north of Bodega Bay and along the narrow coastal strip to the south of San Francisco. Hence, COAMPS overpredicted precipitation amounts in many areas even though the IVT predicted to impinge on the orography of northern California was weaker than that forecast by the HRRR.

In an attempt to help explain the dry bias to the north of Bodega Bay, the 12-h forecasted wind speed and direction and moisture flux at the BBY site are evaluated. HRRR F12 forecasts from 72 initializations are available from 0000 UTC 06 December 2019 to 1200 UTC 08 December 2019. Summary statistics in terms of the median and interquartile range for these F12 forecasts interpolated to the BBY site are shown in

Figure 3.14 along with the summary statistics from the BBY wind profiler and HRRR analyses previously shown in Figure 3.5. Above 2.5 km, the median wind speed values as a function of level and variability of those wind speeds from the F12 forecasts are quite similar to those observed and analyzed. However, the F12 wind speeds in the 1-2 km AGL range tended to be weaker, which leads to a more pronounced low-level jet feature at 500-800 m AGL occurring more frequently than observed during this event. The F12 wind directions below 2 km AGL (Fig. 11b) tend to be even less south-southeasterly than those observed and HRRR analyses. The median COAMPS wind speed and direction are also computed from the 6-member sample of F12 forecasts available during the 72-h event for the BBY location. COAMPS wind speeds tended to be higher than those observed above 1.5 km AGL and weaker below with less indication of a low-level jet feature (Fig. 3.14). COAMPS F12 wind directions tend to be more south-southeasterly below 1.5 km than observed as well.

Similar to Figure 3.6, the F12 moisture flux is calculated for HRRR and COAMPS at the BBY location (Fig. 3.15). HRRR F12 median moisture flux is nearly the same as to what was analyzed by the model (Fig. 3.6). Peak moisture flux is observed in the 500-800 m AGL region, similar to where the low-level jet is found in Figure 3.14. COAMPS F12 median moisture flux is stronger than what was observed (Fig 3.6), however is still considerably weaker than the HRRR.

This case study highlights that short-range HRRR and COAMPS forecasts captured well the key features of the AR event. However, differences between forecasts of IVT and those analyzed by HRRR do not align with a simplistic expectation that reduced forecasted IVT would correspond to reduced orographically-induced

precipitation. For example, COAMPS IVT forecasts were weaker than those analyzed by HRRR yet yielded excessive amounts of precipitation compared to the Stage-IV analyses. In addition, the underprediction of precipitation by both modeling systems to the north of Bodega Bay where the observed precipitation was largest has no clear-cut linkage to differences between the local forecasted wind field and those observed from the nearby wind profiler. Many of these differences will be shown in the next section to be common in many other AR events throughout a winter season.

3.3 2018-2019 Season Cumulative Statistics

The HRRR archive at the University of Utah discussed in Chapter 2 focuses on storing analyses of two-dimensional and three-dimensional fields, which are of greatest interest for most studies. When conditions warrant, e.g., for the 6-8 December 2019 AR event, the three-dimensional forecast fields are saved. Special efforts were required to obtain the COAMPS three-dimensional forecast fields for the case study as well. Hence, it is not possible to evaluate the skill of IVT and moisture flux over the 2018-2019 season as part of this study. The focus in this subsection is to validate the initial 12-h accumulated precipitation forecasted by HRRR, COAMPS, and NAM-3km for the periods from 0000-1200 UTC and 1200-0000 UTC initialized at those starting hours during the 90 days of the 2018-2019 winter season.

3.3.1 Bias

The seasonal averages of 12-h precipitation totals from the Stage-IV analyses and from the 12-h forecasts from the three models are shown in Figure 3.16. The highest 12-h average precipitation is analyzed to be in the coastal mountains of northern California and southern Oregon (Fig. 3.16a). HRRR and COAMPS 12-h forecasts (Fig. 3.16b,c) subjectively compare well with the Stage-IV analyses, with differences noticeable in the northern Sierras and along the northern coast of California. The NAM-3km average forecasted precipitation (Fig 3.16d) differs from that analyzed or forecasted by the other models with accumulations much higher throughout most of the domain. All three models also tend to forecast more precipitation in Nevada and eastern Oregon compared to the Stage-IV analyses.

The seasonal bias ratio for each model is shown in Figure 3.17 and calculated by dividing the total of the 12-h precipitation forecasts (Fig. 3.16b-d) divided by the analyzed seasonal total precipitation (Fig. 3.16a). The bias ratio is most relevant over and to the west of the Sierra Nevada where areas receive ample amount of seasonal rainfall. The large wet and dry biases in the lee of the Sierras, Nevada, and eastern Oregon result primarily from relatively small discrepancies of the model forecasts from the small seasonal precipitation totals in those regions.

Of the three models, the HRRR exhibits the smallest bias ratios in most areas. HRRR has a dry bias along the coast stretching from the Monterey Bay area north into southern Oregon, which corresponds to the region with the highest precipitation amounts analyzed (Fig. 3.16a). The HRRR also has a dry bias ratio in the northernmost sections of the Sacramento Valley and a wet bias ratio in the southern Sierra Nevada mountains and

isolated mountain peaks in northern California. These are generally high elevation areas likely influenced by substantive orographic enhancement. COAMPS has a large positive bias ratio, especially in the Central Valley that did not receive much precipitation during this season. The wet bias extends into the central and southern Sierra Nevada mountains. COAMPS also exhibits a dry bias along the northern coast of California similar to the HRRR; however, it is not as extensive. There is also a strong dry bias along the Traverse Ranges in southern California. The NAM-3km exhibits a substantial wet bias across much of the domain compared to both the HRRR and COAMPS.

The bias and RMSE averaged over the entire domain during each 12-h period during the 2018-19 season are calculated for each model relative to Stage-IV analyses (Fig. 3.18). During December, all models had small bias and RMSE values since precipitation amounts across the region tended to be low (Fig. 2.3). The biases and RMSE remained low until mid-January after which several large AR events occurred. COAMPS and NAM-3km tended to have large positive biases (overprediction) and larger RMSEs compared to the HRRR during AR episodes. The HRRR had episodes particularly during January AR events when the average bias was negative, indicative of underprediction of precipitation compared to the Stage-IV analyses.

3.3.2 Skill scores during AR versus non-AR episodes

Although a majority of the total precipitation during the study period occurred during AR events, there were roughly similar numbers of AR and non-AR episodes (49 AR events versus 33 non-AR events, respectively). A variety of deterministic skill scores that rely upon 2 x 2 contingency tables can then be applied to similar samples from all of

the models for AR versus non-AR episodes. The HRRR shows the best overall skill during AR and non-AR events independent of metric and precipitation range (Figs. 3.19 and 3.20). For example, the HRRR has the highest ETS values, frequency bias values closest to one, and lowest false alarm ratios relative to the other two models for both sets of events. The slightly-above-one HRRR frequency bias scores indicate a tendency for overprediction that likely results from the large areal coverage of the unrealistic values over Nevada and southeastern Oregon (Fig. 3.17a). The COAMPS and the NAM-3km forecasts tend to have similar skill scores for precipitation amounts up to 10 mm (0.4 in). What differences exist between the sets of skill scores for AR relative to non-AR events from those two models are not substantively different from one another.

Skill scores from all three models are slightly worse for non-AR events compared to those for AR events (compare Figs. 3.19 and 3. 20). The improved forecast skill during ARs episodes may result from better initialization and forecasts from the respective model's parent model as well as the models' handling of the larger-spatial scale of the precipitation coverage during that time of each storm's evolution compared to non-AR events that may be influence more by smaller-scale convective precipitation.

Figure 3.21 highlights the ability of the models to predict during AR events where precipitation is occurring for precipitation amounts above selected thresholds. For each precipitation threshold, FSS is calculated over the entire domain using a radial distance from 3 km (28 km^2 area) to 70 km (15394 km^2), which can be viewed as spanning over the range from local and meso- to synoptic scales. FSS values for the lowest threshold (0.254 mm) correspond to evaluating the coverage of measurable precipitation, i.e., the extent to which forecasted areas of measurable precipitation overlap with those analyzed.

While the magnitudes of FSS values in excess of 0.6 could be considered “useful” here (Roberts and Lean 2008; Blaylock and Horel 2020), the relative ability of the models to have high FSS values on the mesoscale (radial distances up to 20-30 km) is of greatest interest for this study.

All three models are able to identify during AR events areas where measurable or greater precipitation has been analyzed on the mesoscale with FSS values in excess of 0.8. For thresholds up to 25 mm (~1 in), HRRR and NAM-3km models have better correspondence between areal coverage of precipitation in excess of those thresholds than that from COAMPS. None of the models would be particularly useful at forecasting the specific locations of heavy precipitation in the Stage-IV analyses during the 12-h periods of AR episodes (areas with amounts in excess of 50 mm within 20 km radial distances). There were ~21,000 grid points that measured precipitation greater than 50 mm during the DJF period, which is only about 1% of the measured 12 h Stage-IV precipitation. Relative to the other two models, the HRRR has higher FSS values for the smaller sample (~2000) grid points where the analyzed 12-h precipitation totals exceeds 75 mm (~3 in).

The FSS scores for specific thresholds during non-AR episodes (Fig. 3.22) are lower than those during AR episodes (Fig. 3.21). Since large 12-h Stage-IV precipitation amounts are less common during non-AR episodes, FSS values for amounts in excess of 75 mm are not shown. The overall lower scores are not unexpected given the tendency for reduced coverage and less organized precipitation during the non-AR episodes. Overall, HRRR continues to have the highest overall FSS scores followed by NAM-3km and then COAMPS. HRRR exhibits useful skill for identifying areas on scales at or below 30 km when evaluating precipitation amounts at or above 10 mm. All three models

show poor skill for the limited areas analyzed to have precipitation amounts in excess of 50 mm (~ 2 in).

As with the ETS metric, the FSS counts a successful forecast as one where precipitation values forecasted and analyzed simply exceed a threshold. Hence, when a small threshold is used for FSS, a skillful forecast may have large discrepancies between what is forecasted and analyzed within the region. To address this issue, FSS is calculated limiting a skill forecast to be one in which both the forecast and analysis fields have areas in common within relatively broad precipitation ranges. Ranges include values corresponding to no measurable precipitation within the 12-h period (0 – 0.254 mm) and light (0.254 mm – 5 mm or 0.254 – 10 mm), moderate (5 – 25 mm or 10 – 50 mm) and heavy (25 – 75 mm or greater than 50 mm) precipitation.

Figure 3.23 shows the FSS within the specified ranges during AR periods. All models do better at forecasting moderate 12-h precipitation amounts (within 5 – 25 mm or 10 – 50 mm) with the HRRR having the highest FSS values. All models have less accuracy explicitly forecasting where no precipitation is analyzed or light amounts less than 10 mm or, as already evident in Figure 3.21, where precipitation was analyzed to be greater than 50 mm.

For non-AR periods (Fig. 3.24), HRRR still remains more skillful compared to the other models, but as seen before, scores are lower for non-AR compared to AR periods. All three models handle light to low-moderate ranges best (0.25 - 5 mm and 5 - 25 mm) and are less successful for the no-precipitation category or precipitation amounts greater than 25 m (the latter already evident in Fig. 3.22).

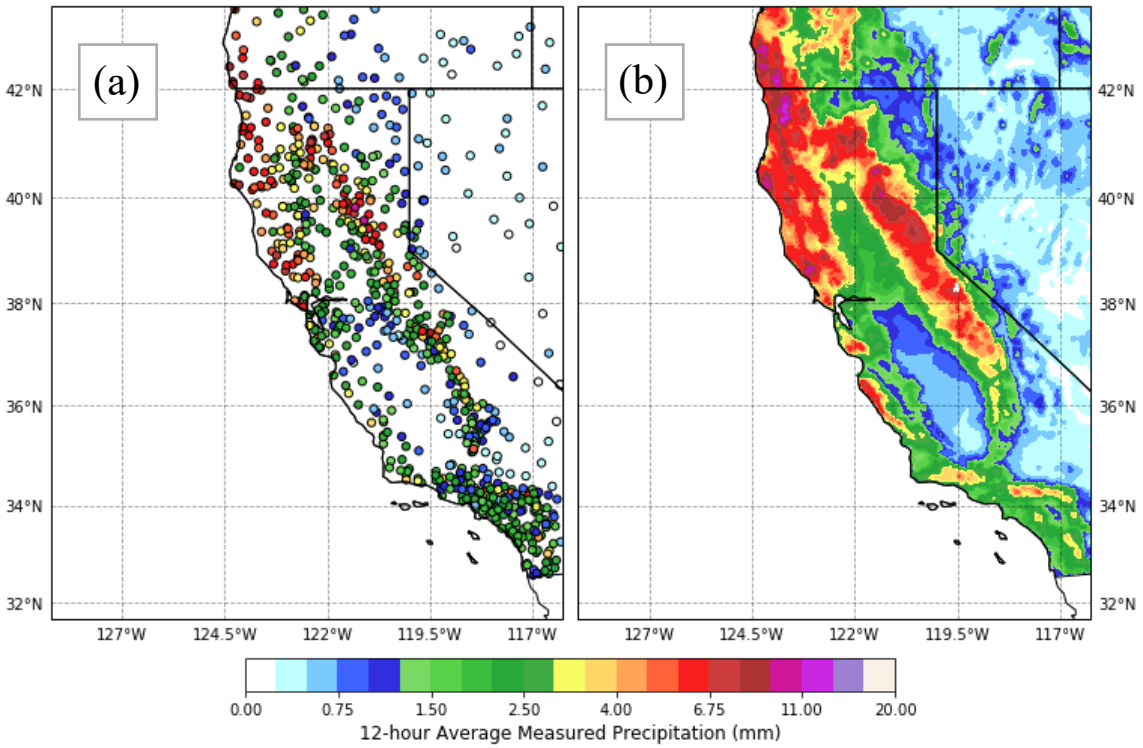

Figure. 3.1 Mean 12-h precipitation totals (mm) measured at (a) stations and for (b) Stage-IV analyses during DJF 2018-19 shaded according to the color bar.

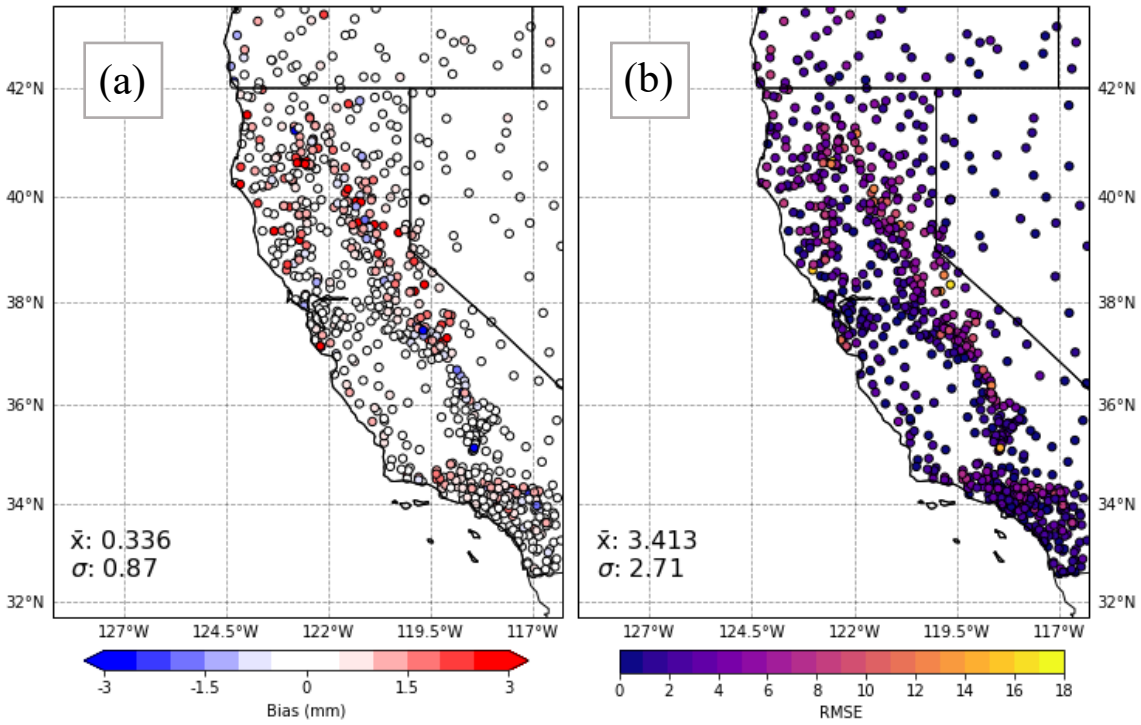

Figure 3.2. Total 12-h precipitation accumulations between Stage-IV analysis values and station observations for (a.) bias (mm), (b.) RMSE, and (c.) bias ratio. Each figure is shaded according to the scale at the bottom during DJF 2018-19. Mean and standard deviation of the bias (in mm) shown in the bottom left corner.

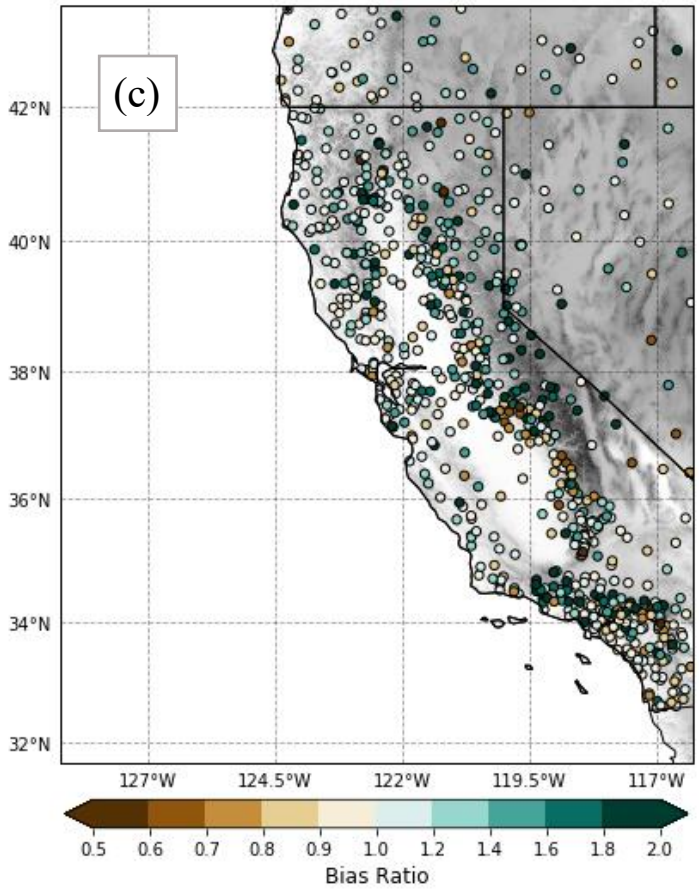

Figure 3.2 continued.

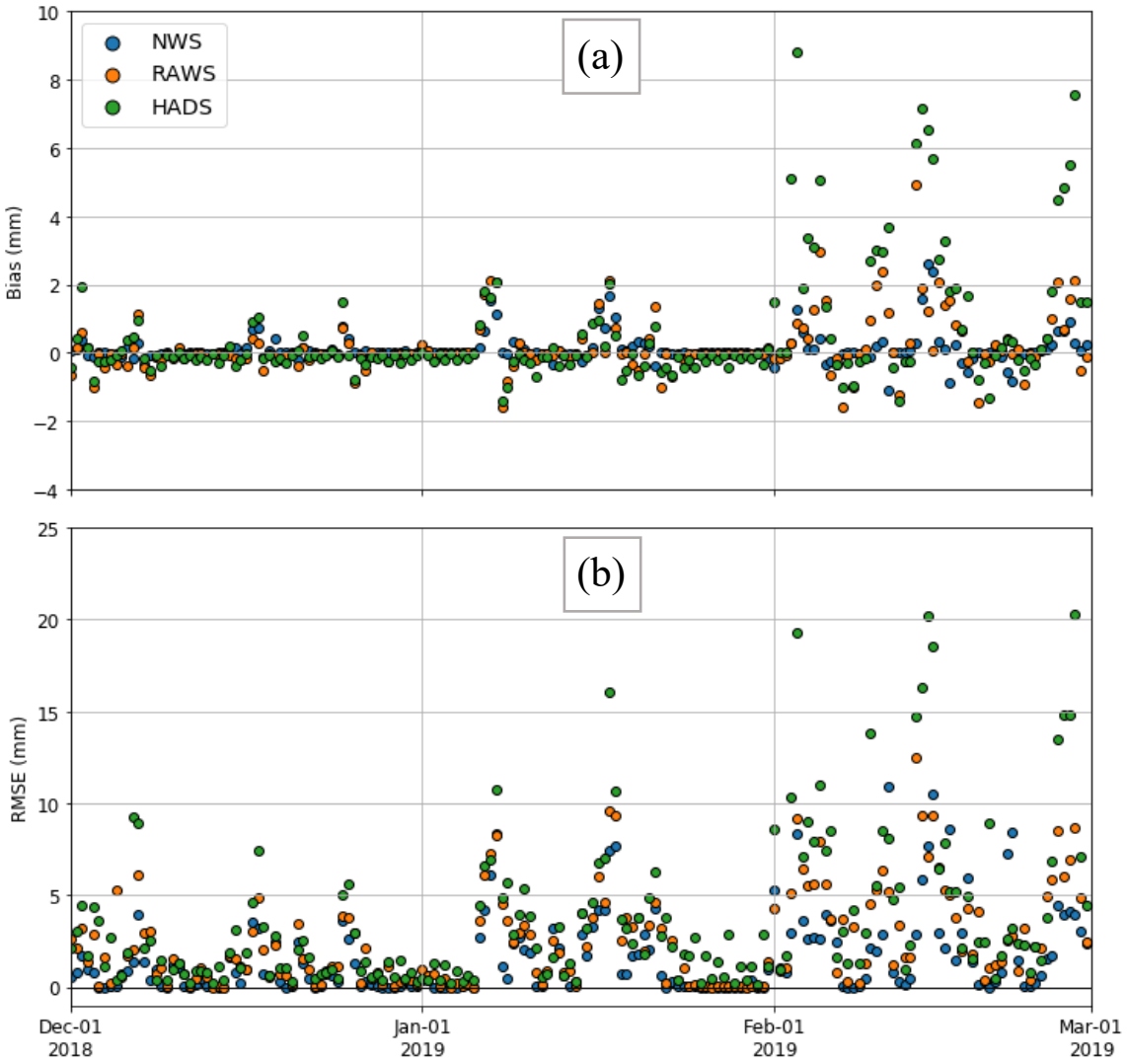

Figure 3.3. Average for each 12-h period of the (a.) bias and (b.) RMSE of Stage-IV analysis totals relative to NWS (blue), RAWS (orange), and HADS (green) station totals during DJF 2018-19.

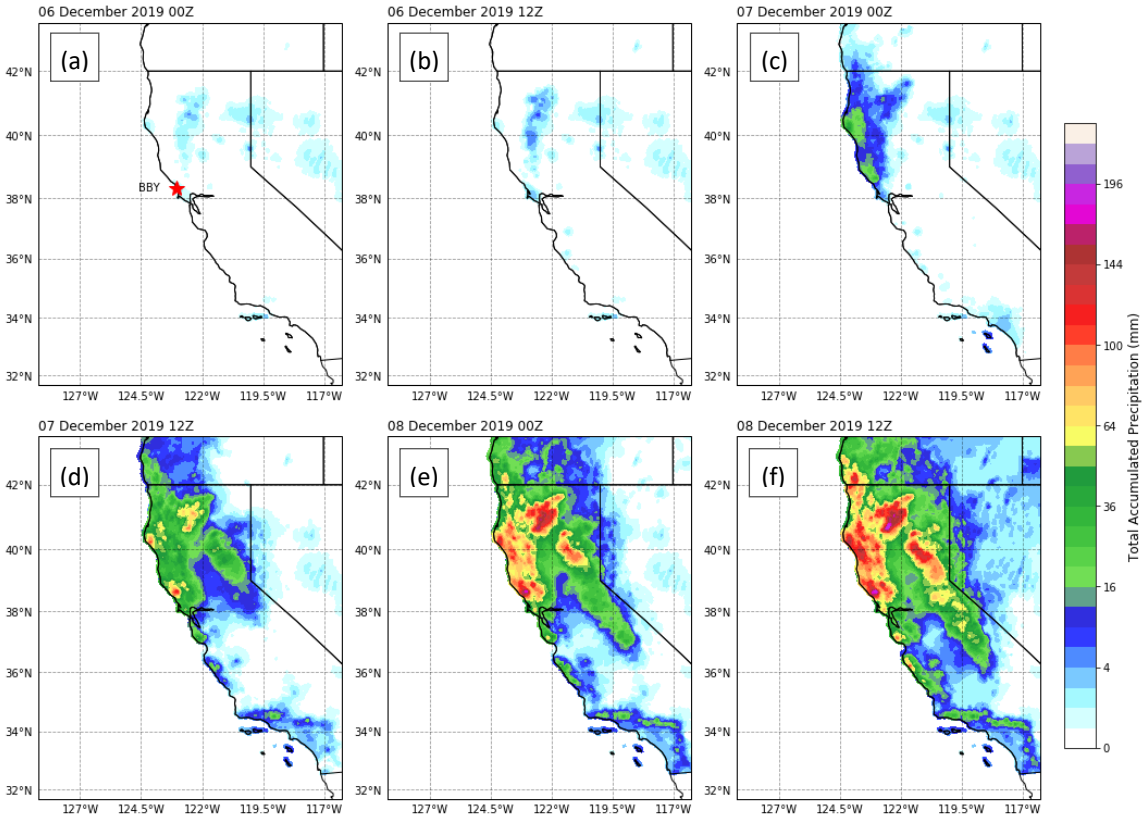

Figure 3.4. Stage-IV analyses of 12-h accumulated precipitation (mm) beginning with the 1200 UTC 05 Dec 2019 to 0000 UTC 06 Dec 2019 period and continuing until the 0000 UTC 08 Dec 2019 to 1200 UTC 08 Dec 2019 period shaded according to the color bar. Bodega Bay (BBY) wind profiler location indicated by the red star in the first panel.

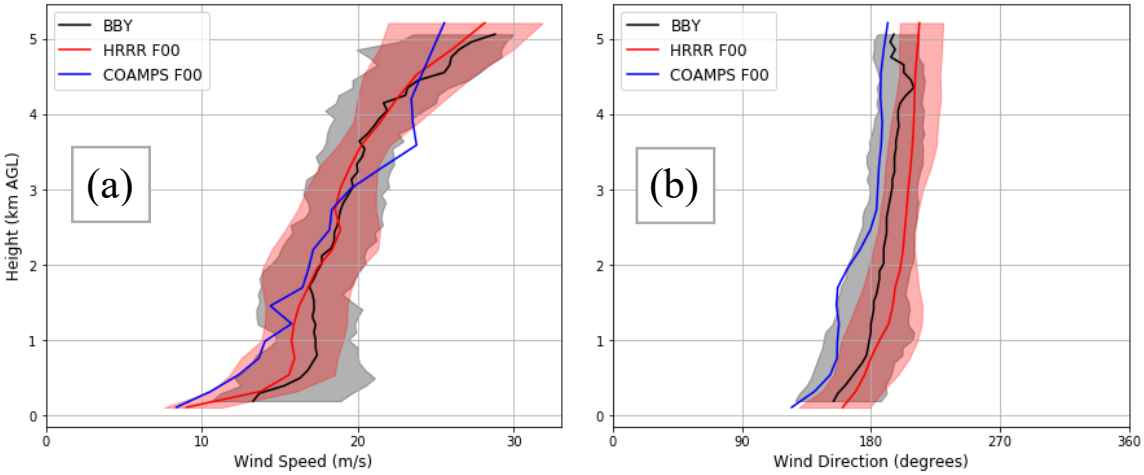

Figure 3.5. Hourly BBY wind profiles, HRRR hourly analyses, and COAMPS 12-hourly analyses near that site (black, red, and blue lines, respectively) for (a.) median wind speed (m s^{-1}) and (b.) wind direction during the 1200 UTC 05 Dec 2019 to 1200 UTC 08 Dec 2019 period. Wind speeds from the profiler and HRRR F00 analyses within the 25th to 75th percentiles are denoted by the black and red shading, respectively.

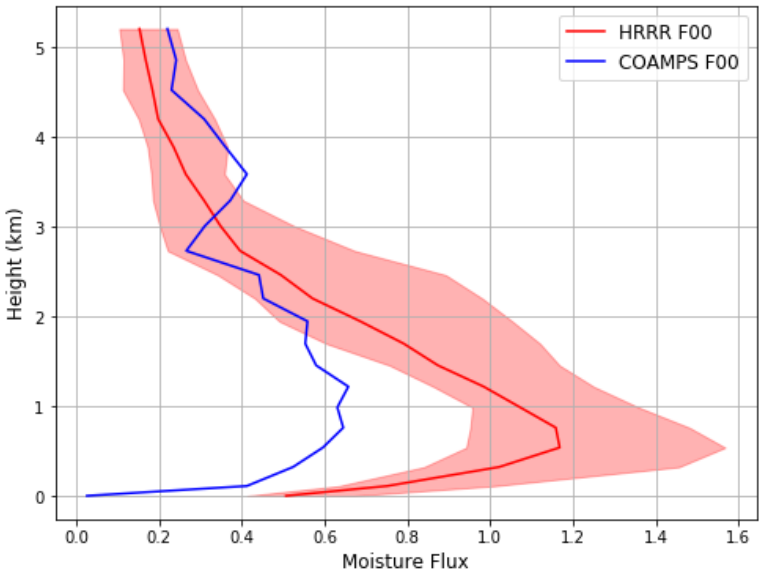

Figure. 3.6. As in Figure 3.5 except for the magnitude of the moisture flux (g kg^{-1} m s^{-1}) near BBY from HRRR F00 and COAMPS F00 analyses.

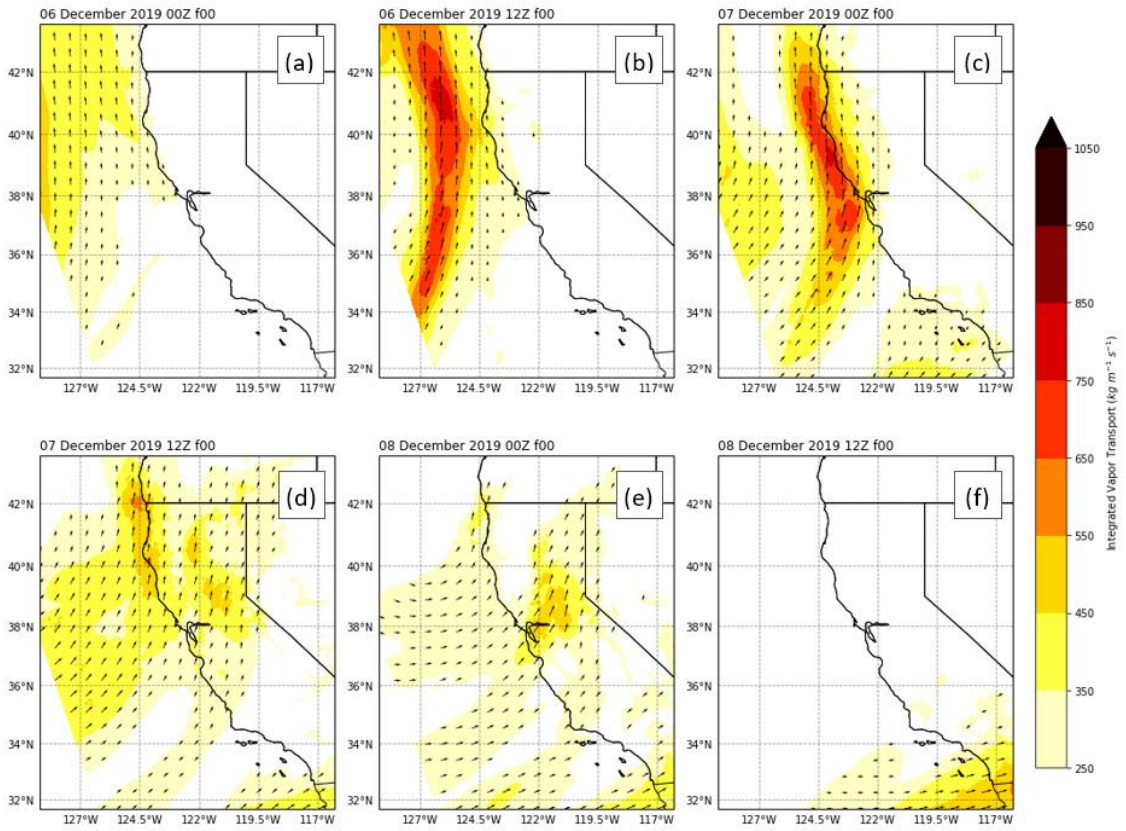

Figure 3.7. Vertically integrated water vapor transport (IVT, kg m^{-1} s^{-1}) from HRRR F00 analyses valid every 12 h from 0000 UTC 06 Dec 2019 to 1200 UTC 08 Dec 2019 shaded according to the color bar. Vectors provide a relative indication of the direction and magnitude of IVT.

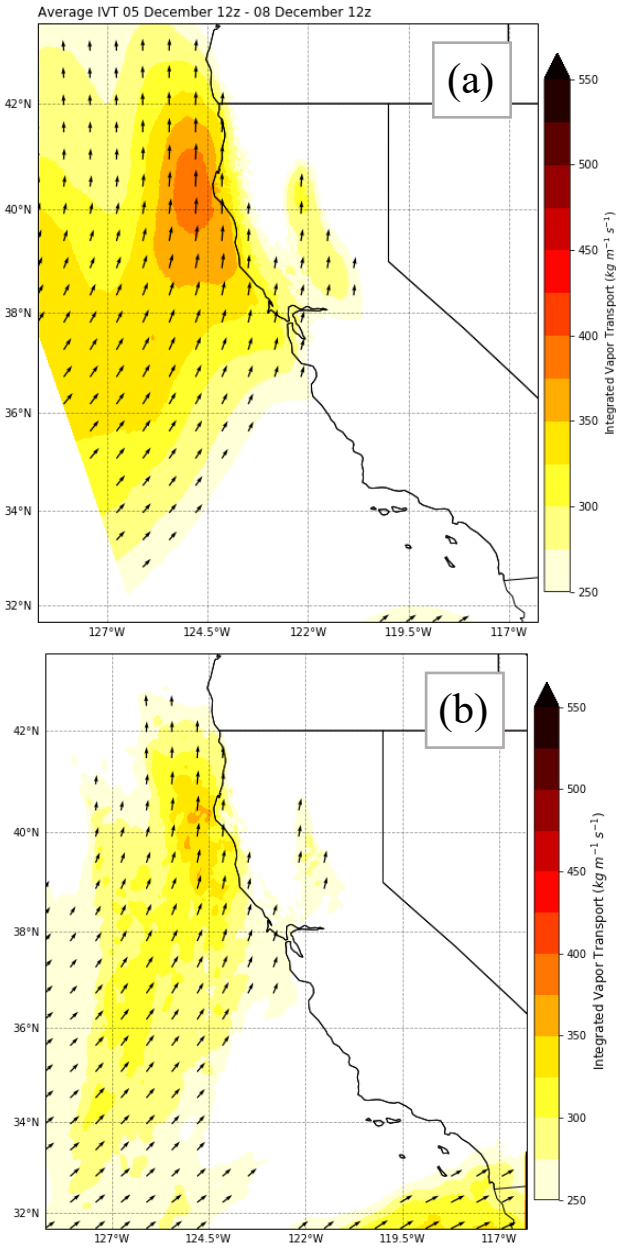

Figure 3.8. Composite IVT (kg m^{-1} s^{-1}) analyses from 1200 UTC 05 Dec 2019 to 1200 UTC 08 Dec 2019 for (a.) HRRR F00 and (b.) COAMPS F00 shaded according to the scale on the right.

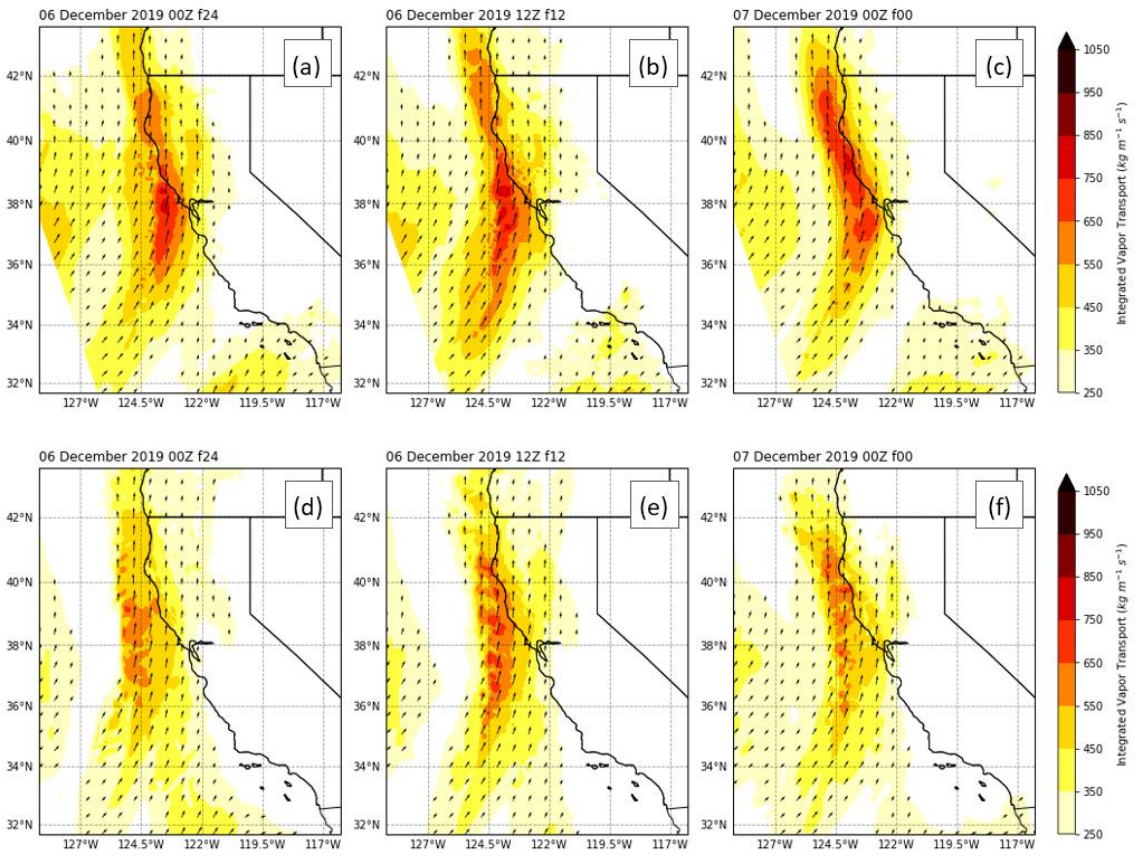

Figure 3.9. IVT (kg m^{-1} s^{-1}) valid 0000 UTC 07 Dec 2019 from (top row) HRRR: (a.) F24, (b.) F12, (c.) F00 and (bottom row.) COAMPS: (d.) F24, (e.) F12, (f.) F00.

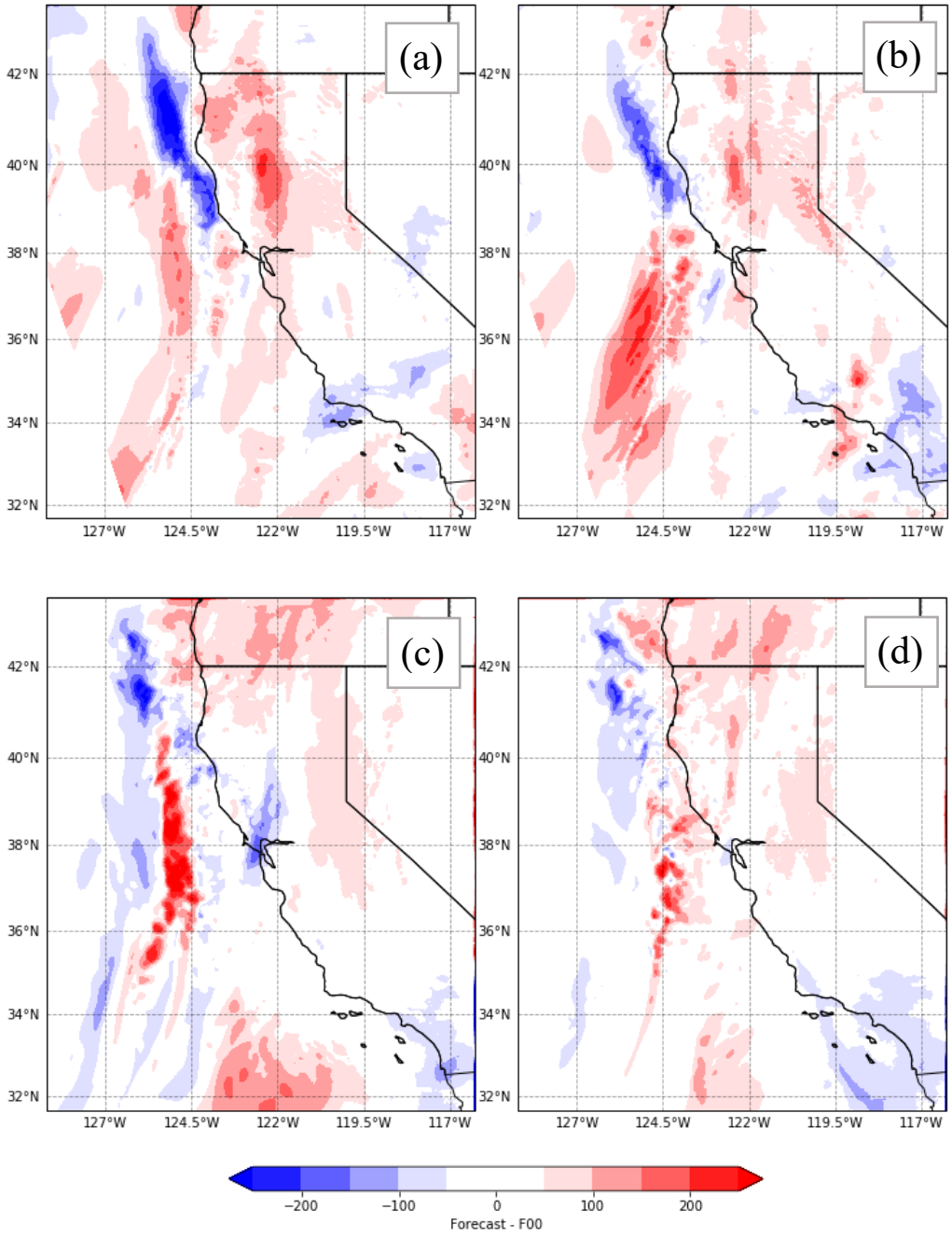

Figure 3.10. IVT difference (kg m^{-1} s^{-1}) between (a.) HRRR F24 and F00 fields valid 0000 UTC 07 Dec 2019. b.) As in (a) except for the difference between HRRR F12 and F00 fields. c.) As in (a) except for the difference between COAMPS F24 and F00 fields. d.) As in (c) except for the difference between COAMPS F12 and F00 fields.

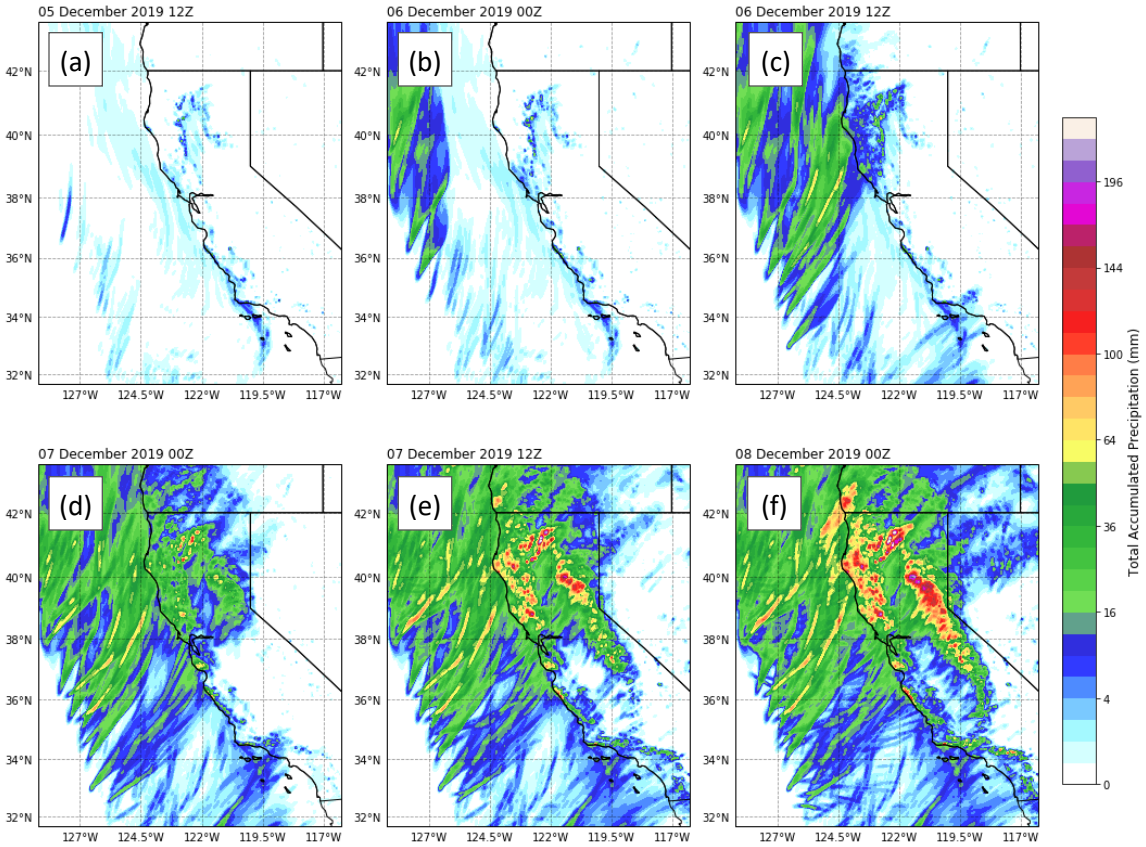

Figure 3.11. Twelve-hour accumulated precipitation (mm) beginning with HRRR forecast initialized at 1200 UTC 05 Dec 2019 for the 1200 UTC 05 Dec 2019 to 0000 UTC 06 Dec 2019 period and continuing until HRRR forecast initialized at 0000 UTC 08 Dec 2019 for the 0000 UTC 08 Dec 2019 to 1200 UTC 08 Dec 2019 period. Shading according to the color bar.

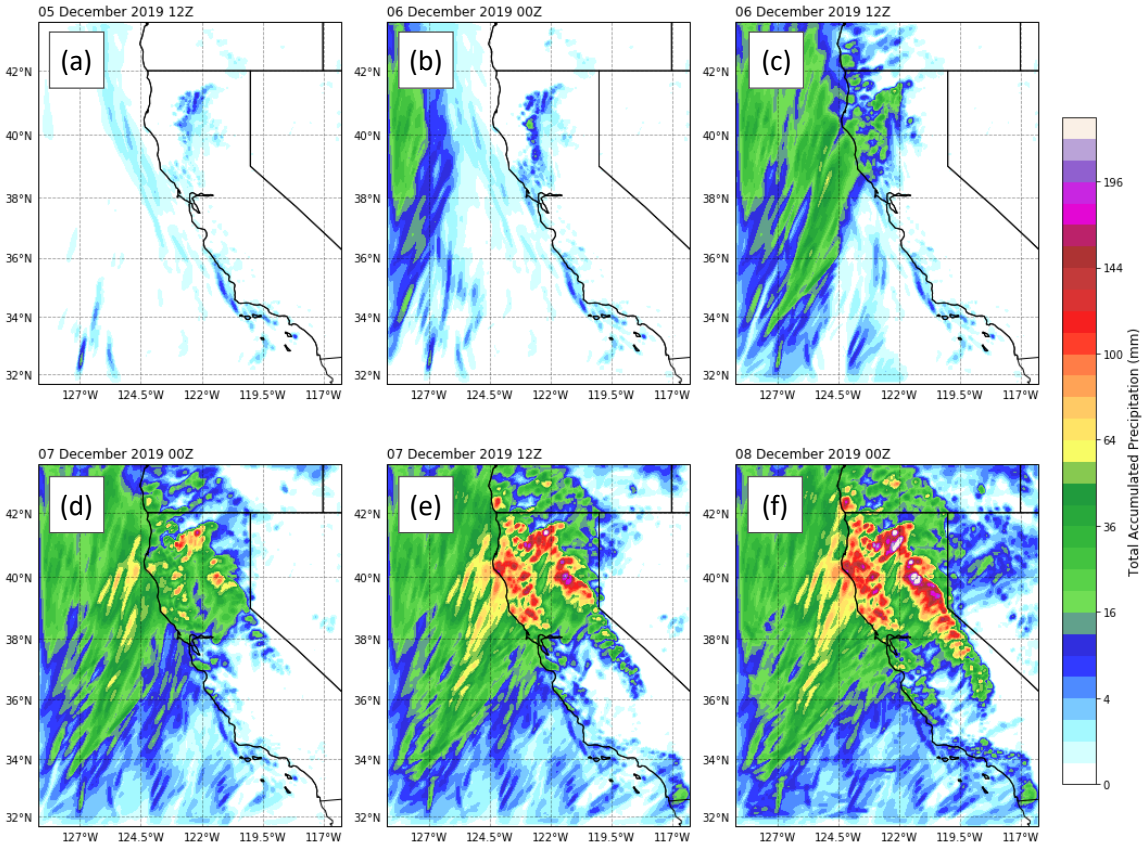

Figure 3.12. As in Figure 3.11 except for 12-h accumulated precipitation (mm) from COAMPS forecasts initialized from 1200 UTC 05 Dec 2019 to 0000 UTC 08 December 2019.

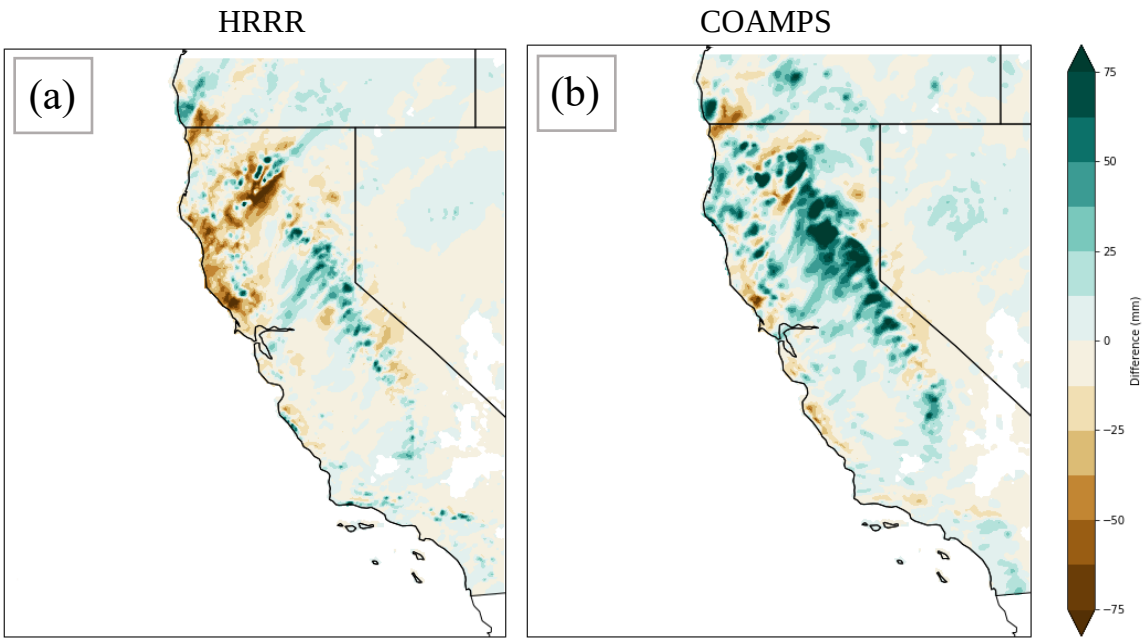

Figure 3.13. Difference in precipitation (mm) accumulated from F01-F12 for (a.) HRRR forecasts initialized from 1200 UTC 05 Dec2019 to 0000 UTC 08 Dec 2019 and Stage-IV analyses for the same 72-h period. b.) As in (a) except the differences are computed from F03-F12 COAMPS forecasts for the same 72-h period relative to the Stage-IV analyses.

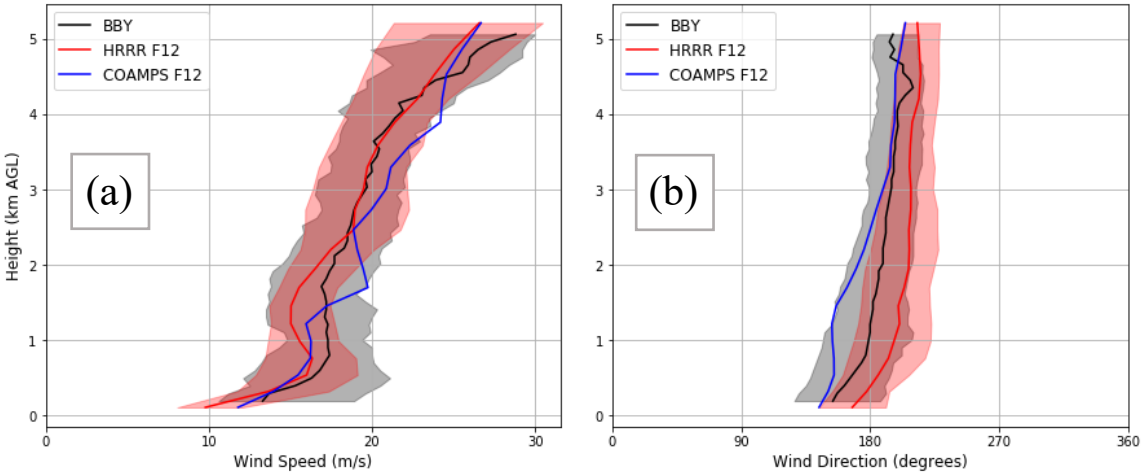

Figure 3.14. As in Figure 3.5 except the red lines and shading are from hourly HRRR F12 forecasts. The median of the COAMPS F12 forecasts is denoted by the blue line.

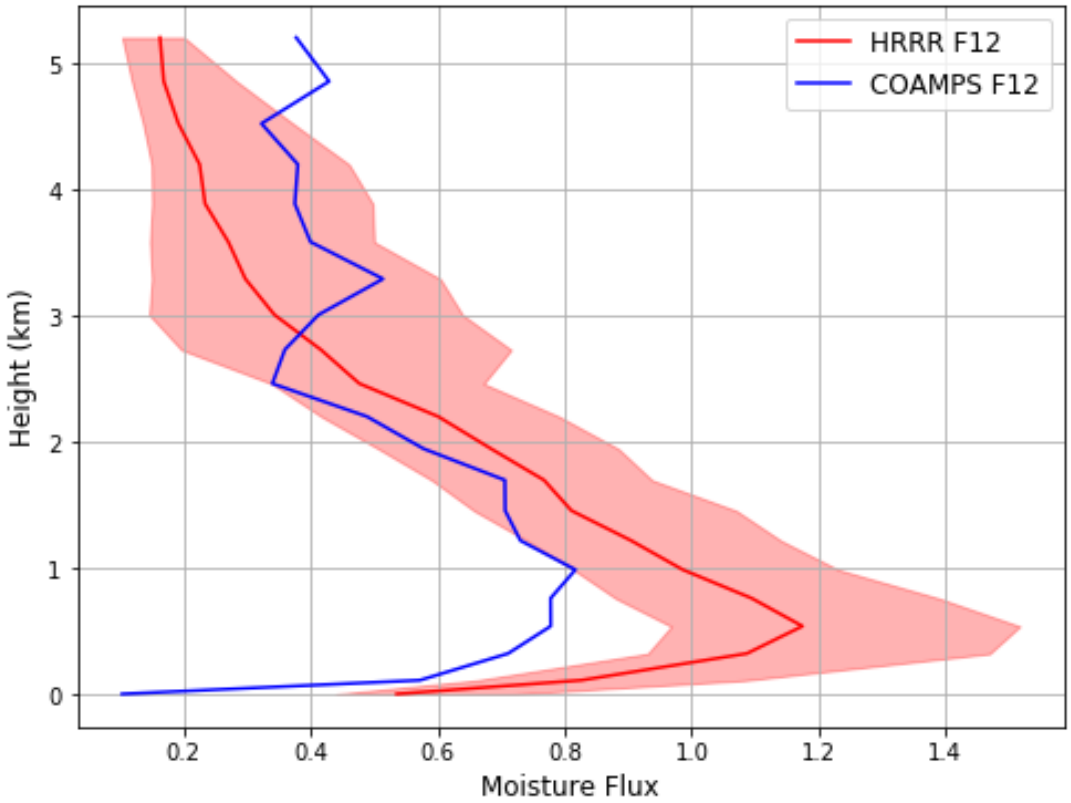

Figure 3.15. As in Figure 3.6 except the red lines and shading are from hourly HRRR F12 forecasts. The median of the COAMPS F12 forecasts is denoted by the blue line.

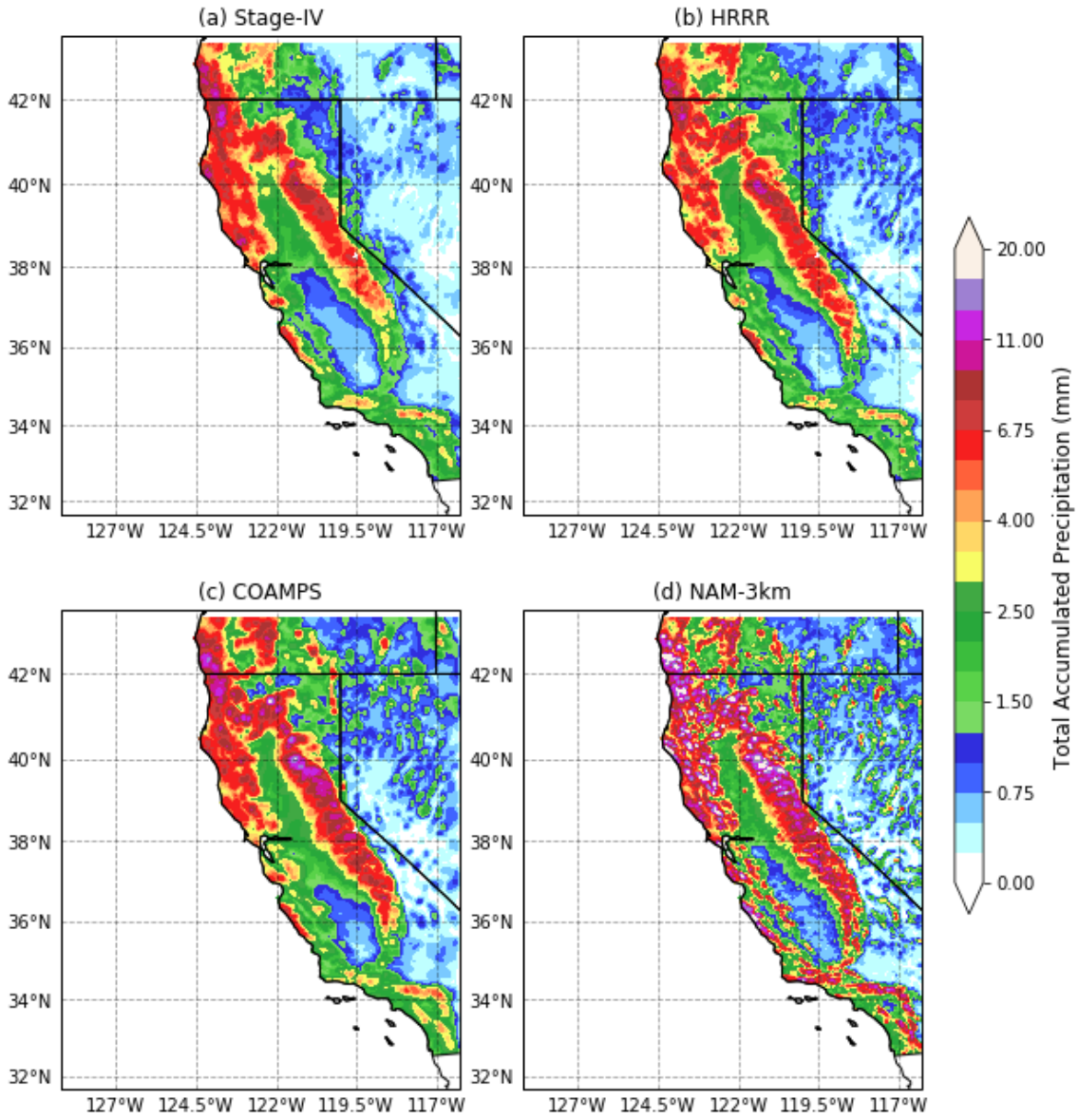

Figure 3.16. Mean 12-h precipitation totals for a.) Stage-IV, b.) HRRR, c.) COAMPS, and d.) NAM-3km during DJF 2018-19 shaded according to the color bar.

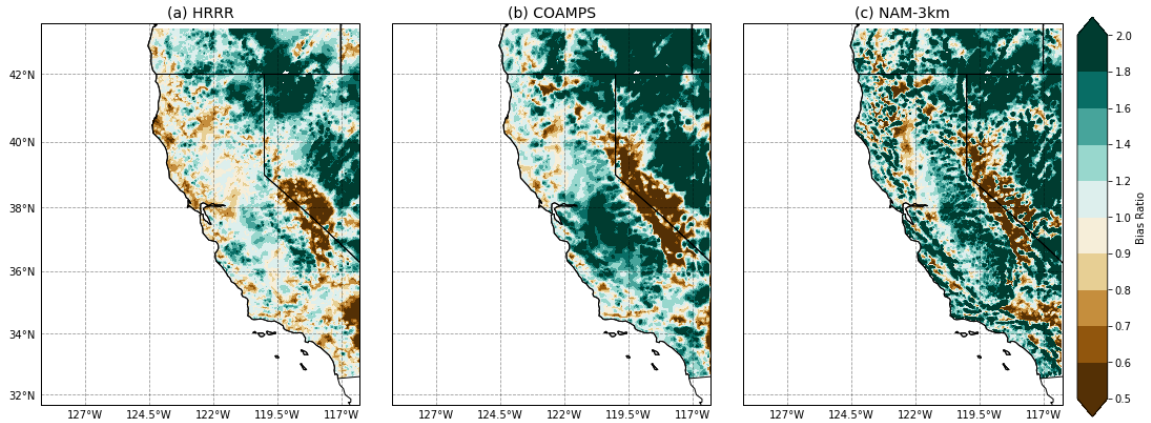

Figure 3.17. Bias ratio for 12-h precipitation totals for the (a) HRRR, (b) COAMPS, and (c) NAM-3km relative to the Stage-IV analysis totals during DJF 2018-19 shaded according to the color bar.

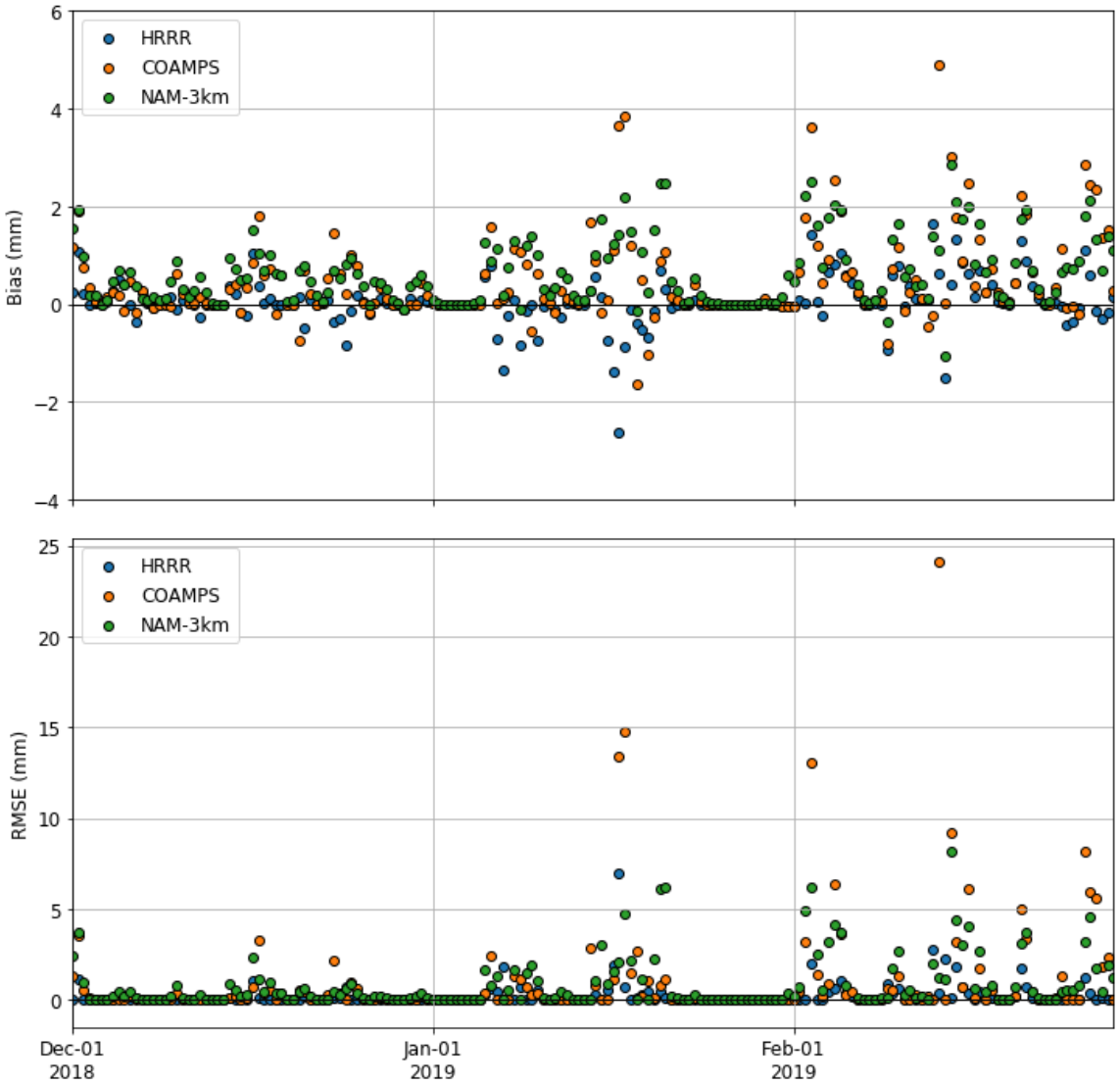

Figure 3.18. Time series of (top) mean bias scores for HRRR (blue), COAMPS (orange), and the NAM-3km (green) models during DJF 2018-19. (Bottom) Same as top row as for RMSE.

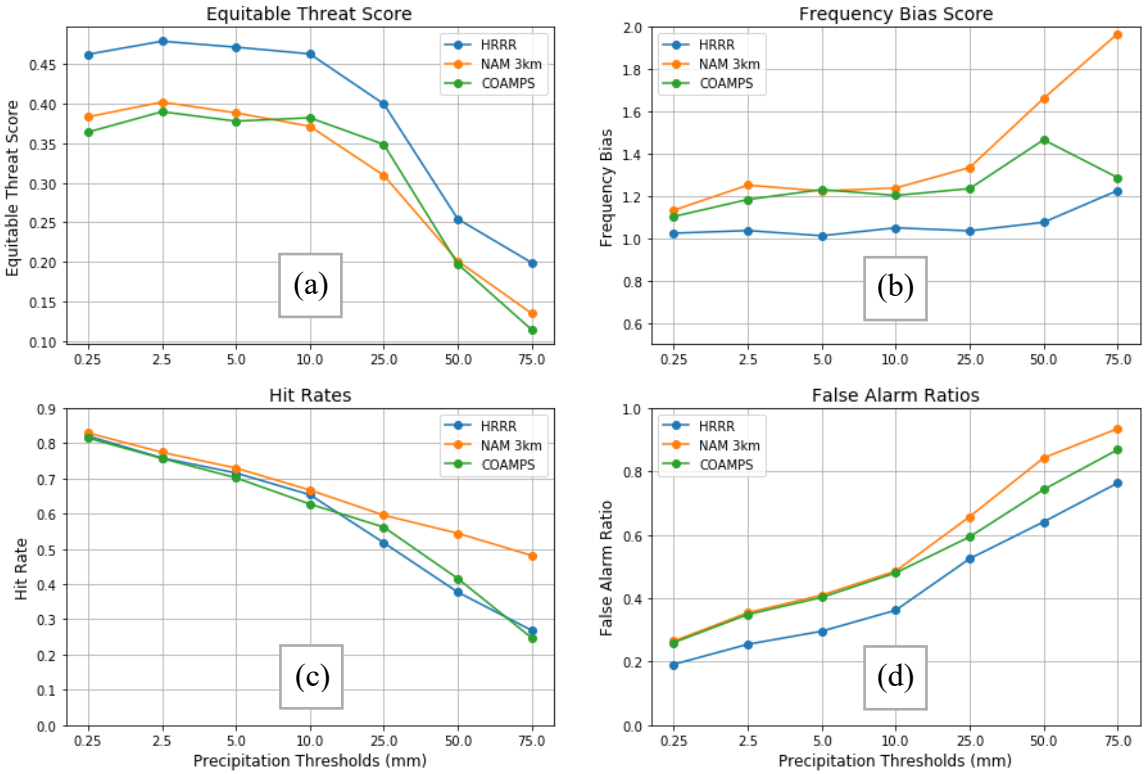

Figure 3.19. Model skill scores for 12-h model precipitation totals relative to Stage-IV analyses during AR events for (a.) equitable threat score, (b.) frequency bias, (c.) hit rate, and (d.) false alarm ratio. The x-axis indicates the starting threshold value used to calculate the skill score. Blue, green, and orange lines indicate HRRR, COAMPS, and NAM-3km, respectively.

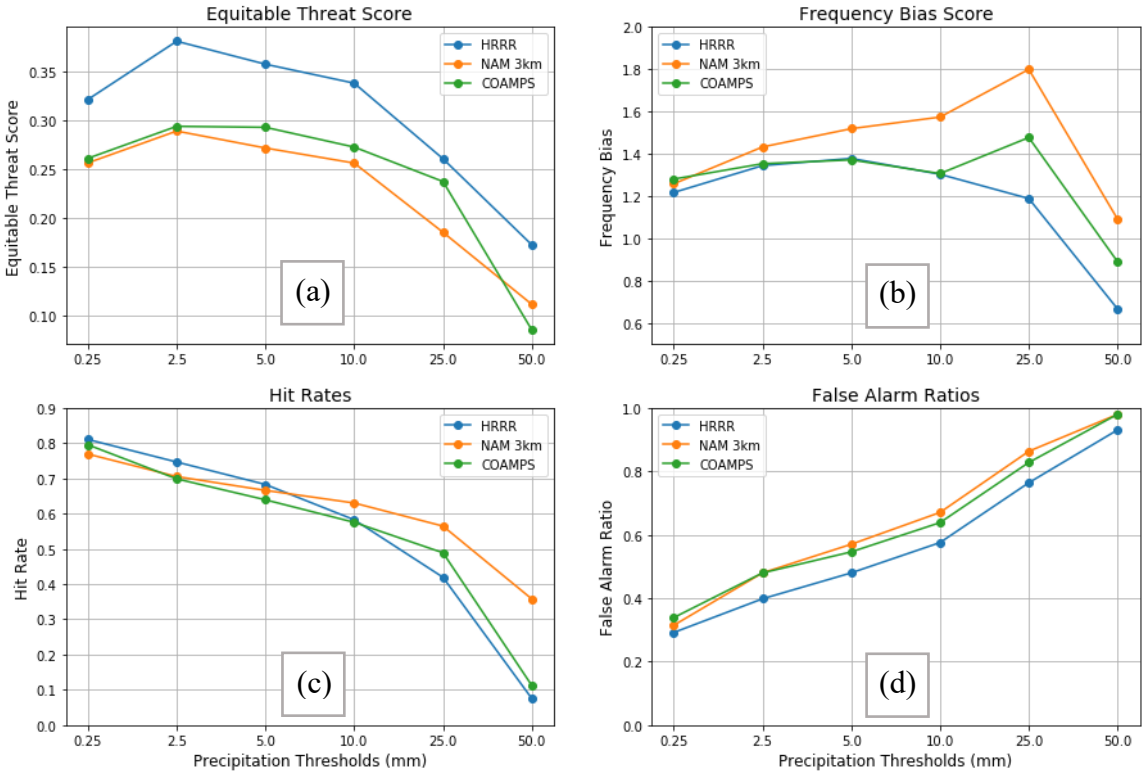

Figure 3.20. As in Figure 3.19 except for non-AR events.

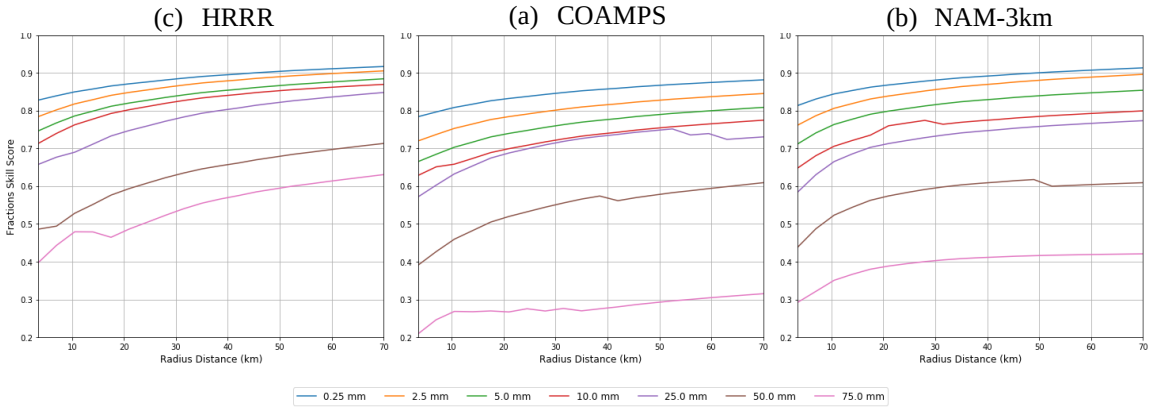

Figure 3.21. FSS computed during AR events for (a.) the HRRR, (b.) COAMPS, and (c.) the NAM-3km model. The x-axis indicates the radial distance (km). Thresholds of 12-h precipitation totals used to calculate FSS are indicated in the legend.

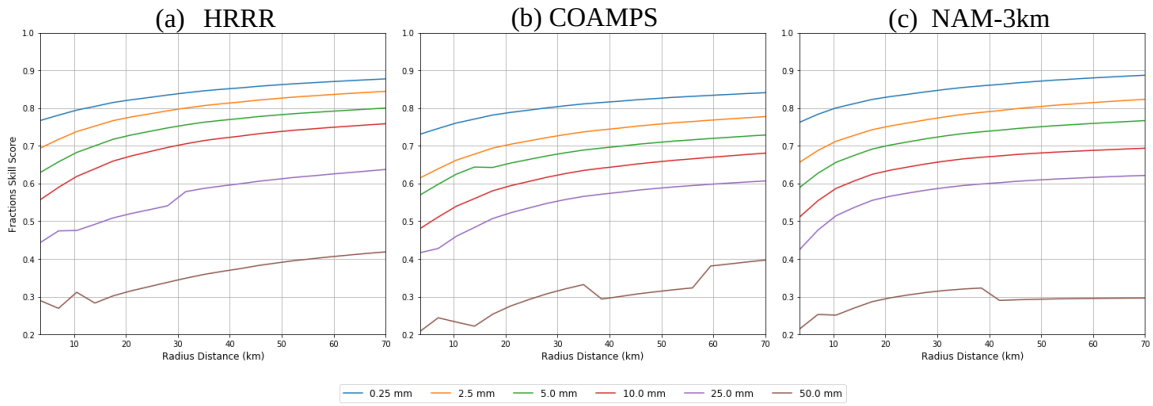

Figure 3.22. As in Figure 3.21 except for non-AR events.

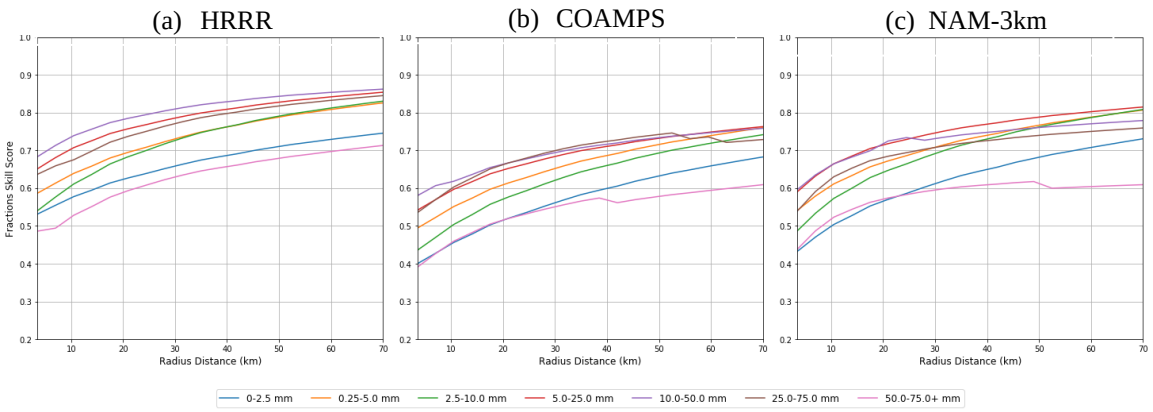

Figure 3.23. FSS computed during AR events for (a.) the HRRR, (b.) COAMPS, and (c.) the NAM-3km model using threshold ranges. The x-axis indicates the radial distance (km). Thresholds of 12-h precipitation totals used to calculate FSS are indicated in the legend.

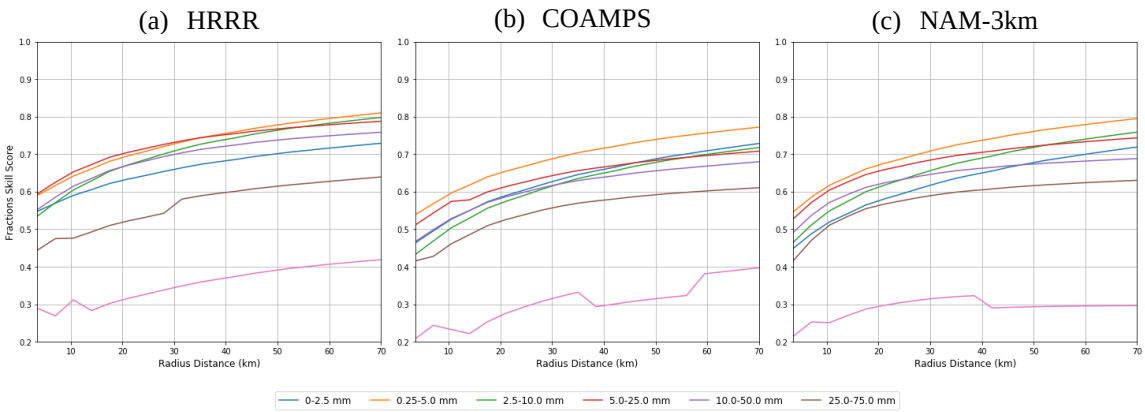

Figure 3.24. As in Figure 3.23 except for non-AR events.

CHAPTER 4

CONCLUSION

4.1 Summary

The Stage-IV gridded analyses were compared to the NWS, RAWS, and HADS station networks to determine whether the Stage-IV product was viable to use when evaluating model precipitation. There were some biases between the two at higher elevations; however, there are many uncertainties related to station measurements at high elevations. Overall, it was found that the Stage-IV analyses were appropriate to use for model validation, even if their fidelity in specific regions may be affected by radar beam blockage, lack of observations in many areas, etc.

HRRR and COAMPS were evaluated using the Stage-IV analyses during an AR case study that took place during 6-8 December 2019. Relative to their verifying analyses, HRRR and COAMPS were able to forecast the general location of IVT associated with the primary moisture plume quite well at forecast lead times of 12 and 24 h. HRRR however overpredicted IVT values inland, an error that tended to be reduced as forecast lead time decreased. COAMPS predicted much lower values of IVT compared to HRRR and did not develop the secondary IVT plume approaching southern California.

HRRR precipitation forecasts during this episode tended to be quite good;

however, those forecasts overpredicted precipitation in mountainous areas, especially in the Mount Shasta area. There was also a dry bias along the northern California coast. COAMPS overpredicted precipitation in mountainous areas as well as produced precipitation bands not analyzed in the Sacramento Valley. COAMPS overall predicted more precipitation than HRRR and yet had forecasted lower than analyzed IVT values.

To examine the dry bias along the northern California coast in both models, wind speed, direction, and moisture flux from the models were compared to the BBY wind profiler in this region. Both models predicted the median wind speeds as a function of vertical level at 12-h lead times reasonably well, although they both predicted slightly weaker speeds of the low-level jet. The forecasted median wind directions from the models differed from those observed with HRRR forecasts exhibiting more southwesterly winds (slightly greater onshore flow) while the smaller sample of COAMPS forecasts tended to have more southeasterly winds (slightly more offshore flow). HRRR 12-h forecasts of low-level moisture flux at this location were similar to its analyses. COAMPS had much weaker moisture flux values, consistent with its overall weaker IVT forecasts.

Precipitation forecasts were separated into 49 AR and 33 non-AR related 12-h periods for the cumulative precipitation statistics for the 2018-2019 season. Overall, all models performed better during AR events when the precipitation tends to be larger and more widespread compared to the weaker, more convective precipitation during the non-AR episodes. In addition, it is possible that ARs embedded within large-scale synoptic events are better initialized than the less-organized conditions associated with non-AR periods. Although the models tend to perform better during AR events, higher average

bias scores occurred during the three largest ARs during February 2019, which likely resulted from overpredictions over the interior mountainous areas. The intricacies in this model behavior could also be due to scale of the terrain in the domain. The Sierra Nevada's are much larger and steeper than the coastal mountains, which could have played a role in the dry biases found along the coastal mountains and wet bias in the interior mountains.

The accuracy of HRRR 12-h precipitation forecasts was higher than that relative to both the COAMPS and the NAM-3km model during both AR and non-AR events HRRR had smaller overall bias ratios, higher ETS scores, lower false alarm rates, and higher FSS. NAM 3-km had a stronger wet bias and a higher orographic precipitation enhancement than COAMPS. Based on FSS computed within precipitation ranges, all models tended to predict moderate amounts of precipitation (10-50 mm) better than smaller and larger amounts of precipitation.

4.2 Future Research

This study was not able to isolate the causes for the higher accuracy of 12-h precipitation forecasts from HRRR compared to the other two models. While precipitation forecasts during the first 12 h are of interest, understanding model skill at longer lead times than examined here has higher value. Although this study investigated hundreds of precipitation forecasts that are part of a 120 tera byte archive at the University of Utah, the order of magnitude more storage required to archive all of the model's two- and three-dimensional fields out to 36 h has not been available. A goal following from this work will be to shift the HRRR archive from the cloud storage

supported by the Center for High Performance Computing to that provided by Amazon Web Services (AWS). Then, it will be possible to process in the AWS cloud the grids and archive them in more practical formats for research by us and many others. Similar practical limitations impeded our analysis of three-dimensional forecast fields from COAMPS.

The importance of improving high-resolution models during AR events is vital for weather forecasters, hydrologists, and emergency managers to enhance public safety in California. ARs provide rich opportunities to understand storm-scale structure of storms making landfall, formation of low-level jets along the coast and within interior valleys, and terrain-flow interactions. Of particular interest will be to transition from examining the present deterministic framework of the operational HRRR and COAMPS to their future ensemble forecast systems in order to better understand the predictability of these storms. Further specific research that could be conducted include examining more AR events in terms of moisture flux and IVT as well as diagnosing how the model's handle how the structure of low-level jets evolve over the lifetime of the AR episodes. This research focused on the relative overall performance of the models without detailed examination using metrics to calculate statistical significance. Some further improvements of the statistical approaches taken to verify precipitation could be to include a bootstrapping technique to establish statistical significance. Also, the addition of the FSS metric for catchment areas can be determined to further understand how well the models perform in areas of complex terrain.

REFERENCES

Aligo, E.A., B. Ferrier, and J.R. Carley, 2018: Modified NAM Microphysics for Forecasts of Deep Convective Storms. *Mon. Wea. Rev.*, **146**, 4115–4153, https://doi.org/10.1175/MWR-D-17-0277.1.

American Meteorological Society, 2018: Atmospheric river. *Glossary of Meteorology*, http://glossary.ametsoc .org/wiki/Atmospheric_river.

Backes, T.M., M.L. Kaplan, R. Schumer, and J.F. Mejia, 2015: A Climatology of the Vertical Structure of Water Vapor Transport to the Sierra Nevada in Cool Season Atmospheric River Precipitation Events. *J. Hydrometeor.*, **16**, 1029–1047, https://doi.org/10.1175/JHM-D-14-0077.1.

Benjamin, S.G., and coauthors, 2016: A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Blaylock, B. and J. Horel, 2020: Comparison of Lightning Forecasts from the High-Resolution Rapid Refresh Model to Geostationary Lightning Mapper Observations. *Wea. Forecasting.* **35**, 401-416. https://journals.ametsoc.org/doi/abs/10.1175/WAF-D-19-0141.1.

Blaylock, B. K., J. D. Horel, and C. Galli, 2018: High-Resolution Rapid Refresh model data analytics derived on the open science grid to assist wildland fire weather assessment. *J. Atmos. Oceanic Technol.*, **35**, 2213–2227, https://doi.org/10.1175/JTECH-D-18-0073.1.

Blaylock, B. K., J. D. Horel, and S. T. Liston, 2017: Cloud archiving and data mining of High-Resolution Rapid Refresh forecast model output. *Comput. Geosci.*, **109**, 43–50, https://doi.org/10.1016/j.cageo.2017.08.005.

Bridger, A.F., Nguyen, D., and Chiao, S., 2019: Developing Spatially Accurate Rainfall Predictions for the San Francisco Bay Area through Case Studies of Atmospheric River and other Synoptic Events. *Atmosphere*, **10**, 541. https://doi.org/10.3390/atmos10090541.

California Department of Water Resources, 2019: Water Year 2020 Begins with Robust Reservoir Storage. https://water.ca.gov/News/News-Releases/2019/October-19/Water-Year-2020-Begins-with-Robust-Reservoir-Storage

Cao, Q., A. Mehran, F. M. Ralph, and D. P. Lettenmaier, 2019: The Role of Hydrological Initial Conditions on Atmospheric River Floods in the Russia River Basin. *J. Hydrometeor.*, **20**, 1667-1686, https://doi.org/10.1175/JHM-D-19-0030.1.

Castellano, C., C. Hecht, B. Kawzenuk, and F. M. Ralkd, 2019: CW3E Event Summary: 6-8 December 2019. https://cw3e.ucsd.edu/cw3e-event-summary-6-8-december-2019

Colle B. A, 2004: Sensitivity of Orographic Precipitation to Changing Ambient Conditions and Terrain Geometries: An Idealized Modeling Perspective. *J. Atmos. Sci.*, **61**, 588-606, https://doi.org/10.1175/1520-0469(2004)061<0588:SOOPTC>2.0.CO;2.

Daley, R. and E. Barker, 2001: NAVDAS: Formulation and Diagnostics. *Mon. Wea. Rev.*, **129**, 869–883, https://doi.org/10.1175/1520-0493(2001)129<0869:NFAD>2.0.CO;2.

Darby, L.S., A.B. White, D.J. Gottas, and T. Coleman, 2019: An Evaluation of Integrated Water Vapor, Wind, and Precipitation Forecasts Using Water Vapor Flux Observations in the Western United States. *Wea. Forecasting*, **34**, 1867–1888, https://doi.org/10.1175/WAF-D-18-0159.1.

Dettinger, M. D., 2013: Atmospheric rivers as drought busters on the U.S. West Coast. *J. Hydrometeor.*, **14**, 1721–1732, https://doi.org/10.1175/JHM-D-13-02.1.

Dettinger, M. D., F. M. Ralph, T. Das, P. J. Neiman, and D. Cayan, 2011: Atmospheric rivers, floods, and the water resources of California. *Water*, **3**, 455–478, https://doi .org/10.3390/w3020445.

Doyle, J.D., C.A. Reynolds, and C. Amerault, 2019: Adjoint Sensitivity Analysis of High-Impact Extratropical Cyclones. *Mon. Wea. Rev.*, **147**, 4511–4532, https://doi.org/10.1175/MWR-D-19-0055.1.

Florsheim, J., and M. Dettinger, 2015: Promoting atmospheric-river and snowmelt-fueled biogeomorphic processes by restoring river-floodplain connectivity in California's Central Valley. *Geomorphic Approaches to Integrated Floodplain Management of Lowland Fluvial Systems in North America and Europe*, P. Hudson and H. Middelkoop, Eds., Springer, 119–141, https://doi.org/10.1007/978-1-4939-2380-9_6.

Gowan, T. M., W. J. Steenburgh, and C. S. Schwartz, 2018: Validation of Mountain Precipitation Forecasts from the Convective-Permitting NCAR Ensemble and

Operational Forecast Systems over the Western United States. *Wea. Forecasting*, **33**, 739-765. https://doi.org/10.1175/WAF-D-17-0144.1.

Haack, T. and S.D. Burk, 2001: Summertime Marine Refractivity Conditions along Coastal California. *J. Appl. Meteor.*, **40**, 673–687, https://doi.org/10.1175/1520-0450(2001)040<0673:SMRCAC>2.0.CO;2.

Haack, T., D. Chelton, J. Pullen, J.D. Doyle, and M. Schlax, 2008: Summertime Influence of SST on Surface Wind Stress off the U.S. West Coast from the U.S. Navy COAMPS Model. *J. Phys. Oceanogr.*, **38**, 2414–2437, https://doi.org/10.1175/2008JPO3870.1.

Harshvardhan, R. Davies, D. A. Randall, and T. G. Corsetti, 1987: A fast radiation parameterization for atmospheric circulation models. *J. Geophys. Res.*,**92** (D1), 1009–1016.

Hatchett, B. J., S. Burak, J. J. Rutz, N. S. Oakley, E. H. Bair, and M. L. Kaplan, 2017: Avalanche fatalities during atmospheric river events in the western United States. *J. Hydrometeor.*, **18**, 1359–1374, https:// doi.org/10.1175/JHM-D-16-0219.1.

Henn, B., A. J. Newman, B. Livneh, C. Daly, J. D. Lundquist, 2018: An assessment of difference in gridded precipitation datasets in complex terrain. *J. Hydrometeorol.*, **556**, 1205-1219, https://doi.org/10.1016/j.jhydrol.2017.03.008.

Hodur, R.M., 1997: The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Mon. Wea. Rev.*, **125**, 1414–1430, https://doi.org/10.1175/1520-0493(1997)125<1414:TNRLSC>2.0.CO;2.

Horel, J. and coauthors, 2002: Mesowest: Cooperative Mesonets in the Western United States. *Bull. Amer. Meteor. Soc.*, **83**, 211–226, https://doi.org/10.1175/1520-0477(2002)083<0211:MCMITW>2.3.CO;2.

Horel, J.D. and X. Dong, 2010: An Evaluation of the Distribution of Remote Automated Weather Stations (RAWS). *J. Appl. Meteor. Climatol.*, **49**, 1563–1578, https://doi.org/10.1175/2010JAMC2397.1.

Hou, D. and coauthors, 2014: Climatology-Calibrated Precipitation Analysis at Fine Scales: Statistical Adjustment of Stage-IV toward CPC Gauge-Based Analysis. *J. Hydrometeor.*, **15**, 2542–2557, https://doi.org/10.1175/JHM-D-11-0140.1.

Hsu, H., L. Oey, W. Johnson, C. Dorman, and R. Hodur, 2007: Model Wind over the Central and Southern California Coastal Ocean. *Mon. Wea. Rev.*, **135**, 1931–1944, https://doi.org/10.1175/MWR3389.1.

Kim, D., B. Nelson, and D. Seo, 2009: Characteristics of Reprocessed Hydrometeorological Automated Data System (HADS) Hourly Precipitation

Data. *Wea. Forecasting*, **24**, 1287-1296, https://doi.org/10.1175/2009WAF2222227.1

Konrad, C. P., and M. D. Dettinger, 2017: Flood runoff in relation to water vapor transport by atmospheric rivers over the western United States, 1949–2015. *Geophys. Res. Lett.*, **44**, 11456–11462, https://doi .org/10.1002/2017GL075399.

Kunz, M. and S. Wassermann, 2011: Sensitivity of flow dynamics and orographic precipitation to changing ambient conditions in idealized model simulations. *Meteor. Z.*, **20**, 199–215, https://doi.org/10.1127/0941-2948/2011/0221.

Lavers, D. A., D. E. Waliser, F. M. Ralph, and M. D. Dettinger, 2016: Predictability of horizontal water vapor transport relative to precipitation: Enhancing situational awareness for forecasting western U.S. extreme precipitation and flooding. *Geophys. Res. Lett.*, **43**, 2275–2282, https://doi.org/10.1002/2016GL067765.

Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV Hourly Precipitation analyses: Development and applications, *19th Conf. on Hydrology*, San Diego, Calif.

Liu, M., J. E. Nachamkin, and D. L. Westphal, 2009: On the improvement of COAMPS weather forecasts using an advanced radiative transfer model. *Wea. Forecasting*, **24**, 286-306.

Martin, A. C., F. M. Ralph, A. Wilson, L. DeHaan, and B. Kawzenuk, 2019: Rapid Cyclogenesis from a Mesoscale Frontal Wave on an Atmospheric River: Impacts on Forecast Skill and Predictability during Atmospheric River Landfall. *J. Hydrometeor.*, **20**, 1779-1794, https://doi.org/10.1175/JHM-D-18-0239.1.

McCorkle, T. A., J. D. Horel, A. A. Jacques, and T. Alcott, 2018: Evaluating the experimental High-Resolution Rapid Refresh–Alaska modeling system using USArray pressure observations. *Wea. Forecasting*, **33**, 933–953, https://doi.org/10.1175/WAF-D-17-0155.1.

Mittermaier, M., and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Wea. Forecasting*, **25**, 343–354, https://doi.org/10.1175/2009WAF2222260.1.

Moore, B. J., Allen B. White, Daniel J. Gottas, and Paul J. Neiman, 2020: Extreme precipitation events in Northern California during Winter 2016–17: Multiscale analysis and climatological perspective. *Mon. Weather Rev,* **148** Issue 3, 1049-1074. https://journals.ametsoc.org/doi/full/10.1175/MWR-D-19-0242.1.

Nachamkin, J.E. and Y. Jin, 2017: An Eulerian Framework for Event-Based Pattern Verification. *Wea. Forecasting*, **32**, 2027–2043, https://doi.org/10.1175/WAF-D-17-0065.1

Nardi, K. M., E. A. Barnes, and F. M. Ralph, 2018: Assessment of numerical weather prediction model reforecasts of the occurrence, intensity, and location of atmospheric rivers along the West Coast of North America. *Mon. Wea. Rev.*, **146**, 3343–3362, https:// doi.org/10.1175/MWR-D-18-0060.1.

Nayak, M. A., G. Villarini, and D. A. Lavers, 2014: On the skill of numerical weather prediction models to forecast atmospheric rivers over the central United States. *Geophys. Res. Lett.*, **41**, 4354–4362, https://doi.org/10.1002/2014GL060299.

Neiman, P. J., F. M. Ralph, A. B. White, D. E. Kinsmill, and P. O. Persson, 2002: The Statistical Relationship between Upslope Flow and Rainfall in California's Coastal Mountains: Observations during CALJET. *Mon. Wea. Rev.*, **130**, 1468-1492, https://doi.org/10.1175/1520-0493(2002)130<1468:TSRBUF>2.0.CO;2.

Neiman, P. J., L. J. Schick, F. M. Ralph, M. Hughes, and G. A. Wick, 2011: Flooding in western Washington: The connection to atmospheric rivers. *J. Hydrometeor.*, **12**, 1337–1358, https://doi.org/10.1175/2011JHM1358.1.

Nelson, B.R., O.P. Prat, D. Seo, and E. Habib, 2016: Assessment and Implications of NCEP Stage-IV Quantitative Precipitation Estimates for Product Intercomparisons. *Wea. Forecasting,* **31**, 371–394, https://doi.org/10.1175/WAF-D-14-00112.1.

Neveu, E., A. M. Moore, C. A. Edwards, J. Fiechter, P. Drake, W. J. Crawford, M. G. Jacox, and E. Nuss, 2016: An historical analysis of the California Current circulation using ROMS 4D-Var: System configuration and diagnostics, *Ocean Model.*, **99**, 133-151, https://doi.org/10.1016/j.ocemod.2015.11.012.

Oakley, N. S., J. T. Lancaster, M. L. Kaplan, and F. M. Ralph, 2017: Synoptic conditions associated with cool season post-fire debris flows in the Transverse Ranges of southern California. *Nat. Hazards*, **88**, 327–354, https://doi.org/10.1007/s11069-017-2867-6.

Picard, L. and C. Mass, 2017: The Sensitivity of Orographic Precipitation to Flow Direction: An Idealized Modeling Approach. *J. Hydrometeor.*, **18**, 1673-1688, https://doi.org/10.1175/JHM-D-16-0209.1.

Ralph, F. M., and M.D. Dettinger, 2012: Historical and National Perspectives on Extreme West Coast Precipitation Associated Associated with Atmospheric Rivers during December 2010. *Bull. Amer. Meteor. Soc.*, **93**, 783-790, https://doi.org/10.1175/BAMS-D-11-00188.1.

Ralph, F. M., M. D. Dettinger, M. M. Cairns, T. J. Galarneau, and J. Eylander, 2018: Defining "atmospheric river": 288 | FEBRUARY 2019 How the Glossary of

Meteorology helped resolve a debate. *Bull. Amer. Meteor. Soc.*, **99**, 837–839, https:// doi.org/10.1175/BAMS-D-17-0157.1.

Ralph, F. M., Neiman P. J., G. N. Kiladis, K. Weickman, and D. W. Reynolds, 2011: A multi-scale observational case study of a Pacific atmospheric river exhibiting tropical-extratropical connections and a mesoscale frontal wave. *Mon. Wea. Rev.*, **139**, 1169–1189, https:// doi.org/10.1175/2010MWR3596.1.

Ralph, F. M., P. J. Neiman, G. A. Wick, S. I. Gutman, M. D. Dettinger, D. R. Cayan, and A. B. White, 2006: Flooding on California's Russian River: Role of atmospheric rivers. *Geophys. Res. Lett.*, **33**, L13801, https://doi.org/10.1029 /2006GL026689.

Ralph, F.M. and coauthors, 2017: Atmospheric Rivers Emerge as a Global Science and Applications Focus. *Bull. Amer. Meteor. Soc.*, **98**, 1969–1973, https://doi.org/10.1175/BAMS-D-16-0262.1.

Reynolds, C.A., J.D. Doyle, F.M. Ralph, and R. Demirdjian, 2019: Adjoint Sensitivity of North Pacific Atmospheric River Forecasts. *Mon. Wea. Rev.*, **147**, 1871–1897, https://doi.org/10.1175/MWR-D-18-0347.1.

Roberts, N.M. and H.W. Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Rogers, E., and Coauthors, 2017: Mesoscale modeling development at the National Centers for Environmental Prediction: Version 4 of the NAM Forecast System and scenarios for the evolution to a high-resolution ensemble forecast system. 28th Conf. on Weather and Forecasting/24th Conf. on Numerical Weather Prediction, Seattle, WA. [Available online at https://ams.confex.com/ams/97Annual/webprogram/Paper311212.html.]

Rutledge, S.A. and P. Hobbs, 1983: The Mesoscale and Microscale Structure and Organization of Clouds and Precipitation in Midlatitude Cyclones. VIII: A Model for the "Seeder-Feeder" Process in Warm-Frontal Rainbands. *J. Atmos. Sci.*, **40**, 1185–1206, https://doi.org/10.1175/1520-0469(1983)040<1185:TMAMSA>2.0.CO;2.

Rutz, J. J., W. J. Steenburgh, and F. M. Ralph, 2014: Climatological characteristics of atmospheric rivers and their inland penetration over the western United States. *Mon. Wea. Rev.*, **142**, 905–921, https://doi .org/10.1175/MWR-D-13-00168.1.

Skok, G., and N. Roberts, 2016: Analysis of fractions skill score properties for random precipitation fields and ECMWF forecasts. *Quart. J. Roy. Meteor. Soc.*, **142**, 2599–2610, https://doi.org/10.1002/qj.2849.

Skok, G., 2016: Analysis of fraction skill score properties for a displaced rainy grid point in a rectangular domain. *Atmos. Res.*, **169**, 556–565, https://doi.org/10.1016/j.atmosres.2015.04.012.

Schmidtm J. M., 2001: Moist physics development for the Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *BACIMO*, Fort Collins, CO, CD-ROM.

Stone, R.E., C.A. Reynolds, J.D. Doyle, R.H. Langland, N.L. Baker, D.A. Lavers, and F.M. Ralph, 2020: Atmospheric River Reconnaissance Observation Impact in the Navy Global Forecast System. *Mon. Wea. Rev.*, **148**, 763–782, https://doi.org/10.1175/MWR-D-19-0101.1.

Swain, D., 2015: A tale of two California droughts: lessons amidst record warmth and dryness in a region of complex physical and human geography. *Geophys. Res. Lett.*, **42**, 9999– 10,003. https://doi.org/10.1002/2015GL066628.

Thompson, G. and T. Eidhammer, 2014: A Study of Aerosol Impacts on Clouds and Precipitation Development in a Large Winter Cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, https://doi.org/10.1175/JAS-D-13-0305.1.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482, https://doi.org/10.1175/2007MWR2018.1.

Wick, G. A., P. J. Neiman, F. M. Ralph, and T. M. Hamill, 2013: Evaluation of forecasts of the water vapor signature of atmospheric rivers in operational numerical weather prediction models. *Wea. Forecasting*, **28**, 1337–1352, https://doi.org/10.1175/WAF -D-13-00025.1.

Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Wea. Forecasting*, **29**, 1451–1472, https://doi.org/10.1175/WAF-D-13-00135.1.

Wu, W. S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916, https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2.

Young, A. M., K. T. Skelly, and J. M. Cordeira, 2017: High-impact hydrologic events and atmospheric rivers in California: An investigation using the NCEI Storm Events Database. *Geophys. Res. Lett.*, **44**, 3393–3401, https://doi.org/10.1002/2017GL073077.

Zachariassen, J., K. F. Zeller, N. Nikolov, and T. McClelland, 2003: A review of the Forest Service Remote Automated Weather Station (RAWS) network. General

Tech. Rep. RMRS-GTR-119, Rocky Mountain Research Station, U.S. Forest Service, Fort Collins, CO, 153 pp. [Available online at http://www.fs.fed.us/rm/pubs/rmrs_gtr119.pdf.]