2.3 Week 1 Forecast - Research and Development

2.3.1 Overview

Research to improve 0-7 day forecasts of extreme weather and water events over Lake Mendocino was a key recommendation of the PVA. At this time scale, forecasts are already quite skillful and can be used to directly inform reservoir operations. However, as evident from the research that we have performed under the FIRO Program, which is described below, there are still plenty of opportunities to further improve the week one forecast skill.

Specifically, improving precipitation predictions at this timescale for the Russian River watershed requires regional weather forecast models, forecast metrics for ARs related to water management, and integration of observational data sets. Regional weather forecast models, such as the "West-WRF" model that is developed at CW3E, is tailored toward improving forecasts of precipitation from ARs over the Russian River. Unique performance and model evaluation metrics for precipitation and landfalling ARs document trends and improvements in forecast skill in existing models that can inform and improve FIRO decision making. Assimilation of observations from existing and reconnaissance-based monitoring improve models both through providing better initial conditions and for verification.

Week-1 forecast research in support to FIRO at Lake Mendocino includes: the assessment of the skill of existing operational predictions of precipitation and streamflow in the Russian River; studies to identify key physical processes that need to be resolved in numerical simulations; the development of forecasting tools (i.e., dynamical models, assimilation of targeted observations, post-processing, and machine learning methods) tailored toward improving forecasts of precipitation and landfalling ARs over the Russian River; the design of ensemble systems to reliably quantify the prediction uncertainty; and the development of unique performance and model evaluation metrics for precipitation and landfalling ARs.

This task has produced the following accomplishments:

- Evaluation of the current skill of CNRFC forecasts of rainfall and inflow to the Lake Mendocino watershed for the cool-season indicates FIRO implementation is viable (section 2.3.2).
- Establishment of West-WRF as a skillful regional model for AR prediction and as a key tool for FIRO-related process-based research (section 2.3.3).
- More than 15 peer-reviewed publications focusing on key physical processes that impact the skill of forecasting systems for extreme weather and water events. Peer-reviewed results are essential to prove the credibility of the scientific underpinnings of FIRO, and enhance the involvement of the broader scientific community to advance understanding and modeling of ARs and extreme events in ways that benefit FIRO beyond the efforts of the FVA (section 2.3.4).
- Production of a 34-year West-WRF reforecast data set (section 2.3.5).
- Development of a novel machine-learning (ML) method that significantly improves the 0-7 day forecasts of NCEP's Global Forecast System (GFS) (section 2.3.5).

- Development of a WRF ensemble to test the impact of hydrometeorological model uncertainty in simulating recent extreme events in the Russian River (sections 2.3.6, 2.5.3).
- AR Recon data (see section 2.2.2) provide the majority of observations within the critical layer of an oceanic AR. The assimilation of these data improves the forecast skills of landfall ARs and precipitations over the U.S. West Coast (section 2.3.7).

2.3.2 Assess Quantitative Precipitation Forecast (QPF) and streamflow/inflow skill for Lake Mendocino

Assessing the current skills of quantitative precipitation forecast (QPF) and inflow forecasts over Lake Mendocino is a central component to evaluating the viability of FIRO at Lake Mendocino. Using median travel times of the water releases (i.e., flood wave) and release limits from the reservoir, the PVA established that forecasts at 1-5 day lead times are the most critical to support FIRO decision making. Based on this, CNRFC cool-season forecasts at the aforementioned lead times of precipitation from 2000-2017 and inflows from 2005-2017 have been evaluated over several accumulation periods to address flood timing-related errors. In addition, the skill in predicting extreme events, or those with a frequency of a 2-year return period or greater, have been evaluated in order to provide new insights into predictability and impacts of extreme precipitation on watershed scales.

Through this assessment, we found that the CNRFC QPF captures 50% or more of the variance (i.e., $R^2 > 0.5$) in the forecast for all lead times (1-5 days) and accumulated precipitation metrics (24 hrs, 72 hrs, and 120 hrs) except the 24 hour accumulated precipitation at 5-day lead (Table 2.3.1). The 24-hr total inflow forecast skill gradually decreases with lead time with R^2 from 0.9 for 1-day forecasts to still greater than 0.5 at 5-day leads. Other statistics, including the root mean square error of the inflow forecast skill are included in Table 2.3.1.

We further subdivided forecast errors by the magnitude of the events and focused on the events that had a climatological return period frequency greater than 2 years. Forecast errors for inflow and precipitation were loosely associated with the corresponding observed (verification) magnitudes (R^2 = 0.38 for 120-hr precipitation, R^2 = 0.44 for 120-hr total inflows). The forecast errors are insensitive to the thresholds used to identify the extreme events. On average, 72-hr and 120-hr forecasts of events that were forecasted to be smaller than an extreme (2-yr or greater return period) event were generally biased high while forecasts of events that were observed to be extreme events were biased low. Figure 2.3.1 presents the best-fit linear relation between inflow errors and precipitation errors for extreme events, which are highly correlated. Using this relationship, for the largest such forecast error (occurring on 25 January, 2008), the estimated inflow error was -9700 ac-ft (-1.2×10^7 m^3). More details on this regression and other statistics of the extreme events can be found in Weihs et al. 2020.

Table 2.3.1 Coefficient of Determination, RMSE (inches), and bias (inches) for CNRFC 24-hr, 72-hr, and 120-hr accumulated QPF for Lake Mendocino watershed as a function of lead time (days). The QPF skill is calculated during the cool season (Oct – April) from 2000 to 2017.

	Lead Time		
Accumulation Metric	**1 day**	**3 days**	**5 days**
24-hr R^2	0.83	0.65	0.39
24-hr RMSE	0.18	0.23	0.26
24-hr bias	0.02	0.01	-0.02
72-hr R^2		0.83	0.68
72-hr RMSE		0.39	0.46
72-hr bias		0.05	-0.02
120-hr R^2			0.79
120-hr RMSE			0.58
120-hr bias			0.03

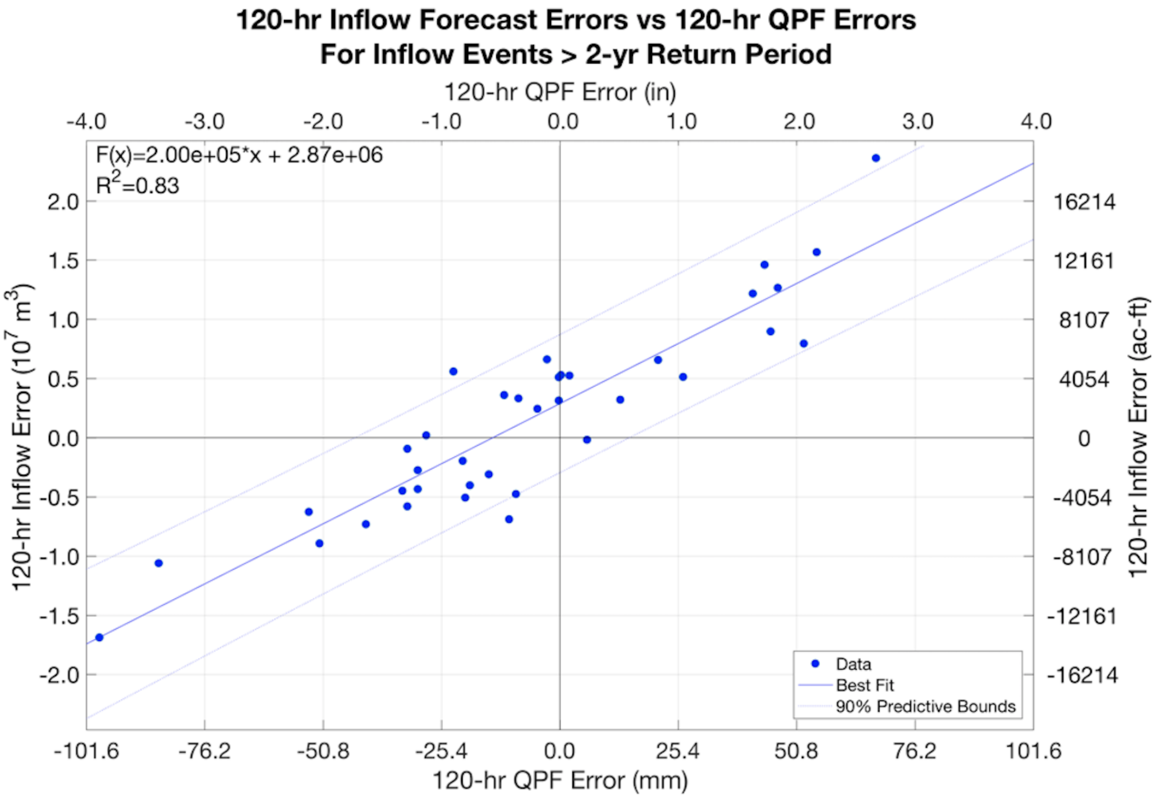

Figure 2.3.1 Relation of 120-hr forecasted inflow errors to their respective QPF errors for inflows with greater than a 2-year return period. The best fit (regression) line is drawn in dark blue and the 90% predictive bounds of the best fit are drawn in light blue. Upper left inset: coefficient of determination and equation for the linear best fit with the general form F(x) = ax+b where a is in units of mm and b is in units of m^3.

Overall, this analysis suggests that there is significant skill in current cool-season single and multi-day forecasts of total rainfall and inflow for Lake Mendocino that can be used in FIRO implementation. Twenty-four hour precipitation forecasts agree well with observations (R^2 exceeding 0.5) out to 4-day lead times, while inflow forecasts agree with observations to 5- day leads. Forecasted dry periods will also be important for reservoir operations because they may provide the basis for keeping encroached water in the reservoir for future supply. Forecasts of no significant rainfall (i.e., less than 25.4 mm,1 inch, per day) were found to be quite skillful (hit rate of 0.97). Inflow forecast errors that exceed 10,000 ac-ft (1.23×10^7 m^3) over a 5-day period were rare during the 1985-2010 period (Figure 2.3.1).

For context it is helpful to consider a FIRO-based Water Control Plan (WCP) where storage is only conditional (based on forecast) in the bottom 10,000 ac-ft (1.23×10^7 m^3) of the flood control pool ("FIRO space"). Under this condition, if inflow forecasts are too high, a portion or all the stored water in the FIRO space might be released without being replaced. If inflow forecasts are too low, flood control space above the FIRO space might be encroached. In this latter case, however, there is still more than 30,000 ac-ft of dedicated flood control space below the crest of the emergency spillway. Thus, even the largest forecast errors found in the QPF and inflow (hindcast) data appear manageable. The effect of forecast inflow errors in FIRO WCPs is fully evaluated in Section 5.

2.3.3. Establish a forecast system and evaluation methodology for West-WRF

Diagnosis of regionally important precipitation processes on scales not adequately resolved by global models is performed using regional forecast models. In turn, optimization of regional forecast models is important for accurate precipitation forecasts, through enhancements in horizontal and vertical resolution, along with the ability to explicitly represent (rather than parameterize) important physical processes associated with mesoscale meteorological features including orographic precipitation. CW3E's West-WRF (Martin et al. 2018) is a version of the Weather Research and Forecasting (WRF) model (a community model developed by the National Center for Atmospheric Research and others; Skamarock et al. 2008; Powers et al. 2017) that is run in near real time (West-WRF NRT) from December to March, and is used year-round for research purposes. West-WRF is continually being improved to optimize forecasts of extreme precipitation events, especially those associated with ARs, over the western U.S. (Martin et al. 2018). Several recent studies have demonstrated that relevant small-scale precipitation processes (e.g., frontal waves, narrow cold-frontal rainbands, and orographic convection) can be well-represented by West-WRF. In addition, West-WRF provides higher spatial resolution for longer lead times than is available from any similar mesoscale NWP model run nationally (Figure 2.3.2).

West-WRF NRT has been run every winter since 2015-2016. Forecasts created in real-time were used to add interpretive value to daily operations by several partners, including the National Weather Service, municipal water agencies, and the CA Federal-State Flood Operations Center (FOC). Notably, specialized West-WRF forecast products were provided to FOC during the emergency involving flooding on the Feather River and damage to the Lake Oroville Dam spillway. Mike Anderson, California's State Climatologist and a key individual in FOC's response to the Oroville crisis, has provided an informal letter thanking CW3E for these efforts.

Throughout the FIRO project, West-WRF was upgraded each year, and culminated in enhancements to the NRT implementation during the 2020 water year (WY). Figure 2.3.2 shows the unique configuration of West-WRF, which is designed specifically to simulate the formation and development of ARs well upwind of the western U.S., a model reach that is not available (neither in space nor time) from NCEP regional modeling systems. The current West-WRF NRT configuration includes an outer domain (with 9-km resolution) covering much of the Eastern Pacific Ocean and western U.S., and a nested 3-km domain covering California (Figure 2.3.2). For comparison, the domains of NCEP operational models, the North America Model (NAM) and the High-Resolution Rapid Refresh (HRRR) model, as well as the maximum operational-forecast lead time for each system, are shown in Figure 2.3.2. During WY 2020 West-WRF NRT simulated the atmosphere on sixty vertical levels (increased from 36 the prior water year) up to 10 hPa. Forecasts are run once daily from 0000 UTC initializations, are run out to 7-day lead times for the 9-km domain, and out to 5-day lead times for the 3-km domain. There are numerous options for parameterization schemes in WRF, and those used in West-WRF largely follow those used in the validation study by Martin et al. (2018) and in subsequent sensitivity testing by CW3E.

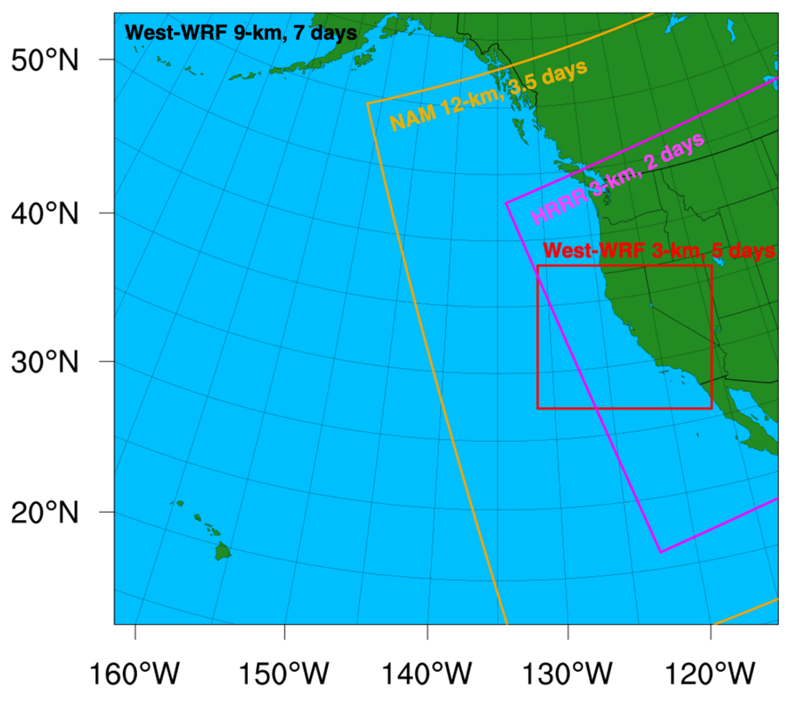

Figure 2.3.2 Computational domains and forecast maximum lead time for WY 2020: West-WRF 9-km (black) and 3-km (red), NAM 12-km (orange), and HRRR 3-km (magenta).

During WY 2020, three separate West-WRF NRT configurations were run daily, differing in input initial boundary conditions (IBCs) and lateral boundary conditions (LBCs). The first simulation uses NCEP's GFS (Environmental Modeling Center 2003) as its boundary conditions, as was done in previous years' NRT simulations. The second simulation uses the ECMWF deterministic high-resolution forecast (ECMWF 2020). We also ran an experimental third West-WRF NRT configuration, with a single-domain simulation spanning the 9-km domain, but with 4-km grid spacing and using the ECMWF forecasts as input. Using different global forecasts as input to West-WRF aids in understanding initial condition uncertainty and reveals more of the possible range of forecast solutions. The 4-km single domain simulation improves our understanding of forecast dependence of model solutions on model grid spacing, number of model vertical levels, and convective parameterization scheme.

Several automated and post-season verification tools have been developed to run in parallel with the West-WRF NRT forecasting system. The forecast accuracy and model skill in AR landfalls, and precipitation amount and spatial coverage, are evaluated on daily, event, and seasonal time scales for West-WRF and for national operational forecast models as comparison. For AR landfalls, we have applied the Method for Object-based Diagnostic Evaluation (MODE), which is part of the Model Evaluation Tools (MET) package, developed by the Developmental Testbed Center at the National Center for Atmospheric Research, to forecasted maps of IVT as a novel approach to AR forecast skill assessment. AR objects are defined based on thresholds of the IVT forecast, axis length, and other criteria. AR object forecast analyses are performed each day for the West-WRF NRT and supporting global models, wherein AR object size and shape distributions (referred to as Measure of Effectiveness), 90th percentile intensity, angle of orientation, and landfall error are assessed as a function of lead time. Since 2017, results of the object detection statistics are posted on the CW3E forecast verification viewer page here and compiled into seasonal statistics of AR performance.

West-WRF forecasts of AR frequency, landfall, and precipitation skill have been assessed over several water years. Verification results are summarized below for WY 2019 and WY 2020. Figure 2.3.3 shows a summary of statistics for West-WRF and GFS AR forecasts during WY 2019 using MODE.

- West-WRF predictions have a lower landfalling position error compared to GFS up to a 5-day lead time (Figure 2.3.3a) where all errors in both models are less than 200 km.
- The skill in predicting the AR intensity has been analyzed for weak and strong storms. For weak storms, West-WRF produced a smaller intensity error compared to GFS; however, it overpredicts the intensity of the AR object compared to GFS for strong ARs (Figure 2.3.3b).
- A new forecast metric, called Measure of Effectiveness, describes the multi-dimensional shape/size of the AR. Measure of Effectiveness is a robust metric for communicating AR skill as it provides additional information on the geospatial placement of ARs. It shows that West-WRF AR forecasts consistently improve as the AR approaches landfall (Figure 2.3.3c).
- GFS and West-WRF have similar hit rates, miss rates, and false alarm rates at 120-hr lead times over the area where AR landfalls occur most often (i.e., northward of Point Conception, CA) (Figure 2.3.3d).
- West-WRF produces a higher hit rate, which represents the fraction of forecasts of ARs that actually arrive, than GFS along the southern California coastline (Figure 2.3.3d).
- At a 5-day lead time, the hit rate can be as low as 20% in both models. West-WRF also has a higher false alarm rate along the southern California coastline. Since ARs are less common over Southern California, the landfall errors in this region may be dominated by relatively few events (Figure 2.3.3d).

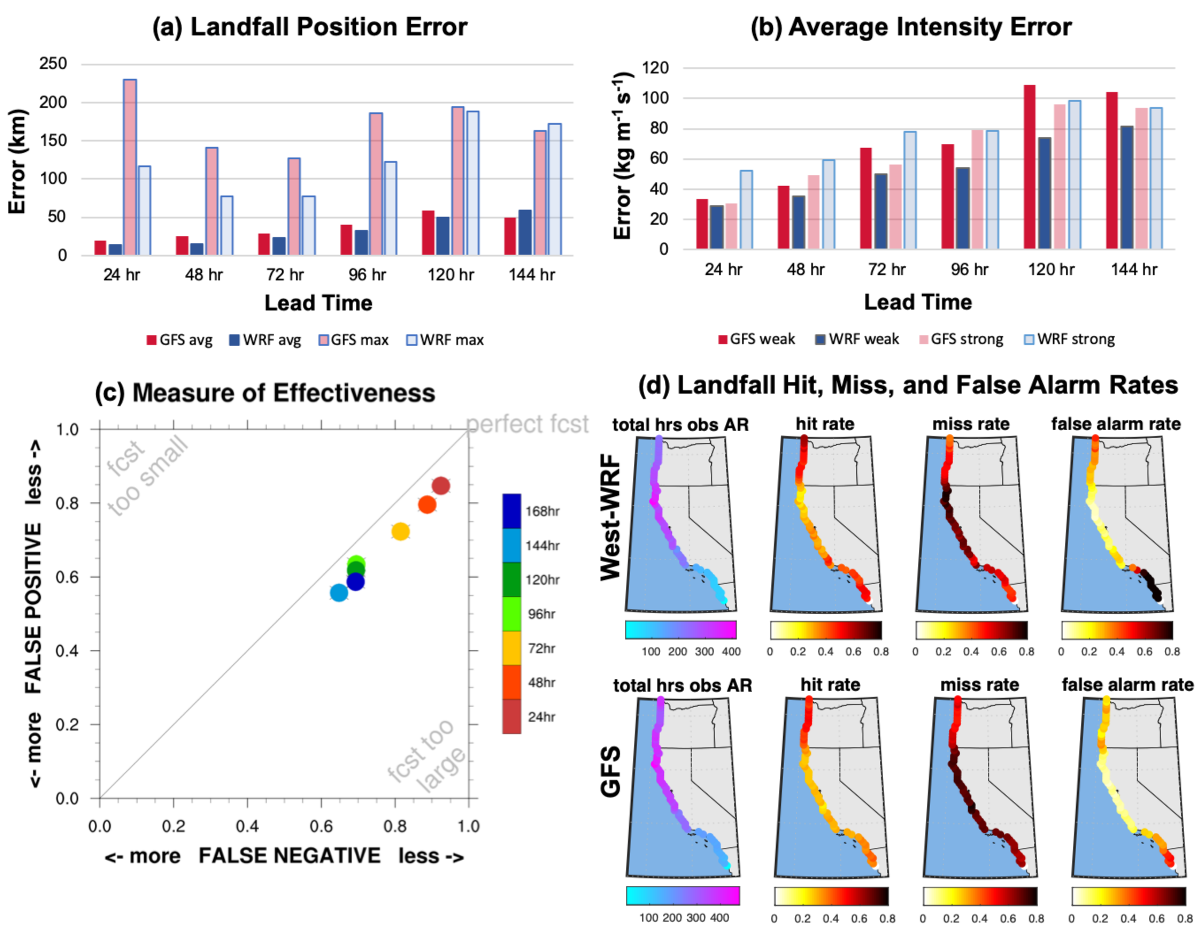

Figure 2.3.3 WY 2019 seasonal statistics of: (a) average (blue) and maximum (red) landfall position error (km) of West-WRF (light colors) and GFS (dark colors); (b) average IVT intensity error of West-WRF (light colors) and GFS (dark colors) subdivided into categories where the model is stronger (blue) / weaker (gray) than each model's analysis; (c) measure of effectiveness of West-WRF objects as a function of forecast lead time (colors); and (d) 120-hr forecasted AR landfall duration total, hit rate, miss rate, and false alarm rate for West-WRF (top row) and GFS (bottom row).

Precipitation forecast skill was primarily verified against the CNRFC's Quantitative Precipitation Estimate (QPE). West-WRF shows several areas with marked improvements over GFS in forecasts of precipitation. West-WRF has a reduced mean wet bias over the Northern and Central Sierra Nevada mountains and reduced dry mean bias over the Transverse Ranges and coastal areas of Southern California during WY 2019 (Figure 2.3.4). The forecasted precipitation over the Russian River watershed (black outline in Figure 2.3.4) and the surrounding Coastal Ranges north of the San Francisco are often too dry in both West-WRF and GFS. We found that this dry bias is persistent across all lead times, systematic across all water years where West-WRF data is available (WY 2016-WY 2020), and most pronounced over the southwestern region of the watershed near Venado, CA (38.6055° N, 123.0081° W). Given that the dry bias is systematic in both West-WRF and GFS, this finding presents an opportunity for future model developments and post processing techniques. Additionally, preliminary West-WRF 9-km forecasts using ECMWF-based initial and boundary conditions have lower centralized root mean squared error (CRMSE) in 24-hour precipitation out to 7 days' lead time in the Russian River basin (Figure 2.3.5).

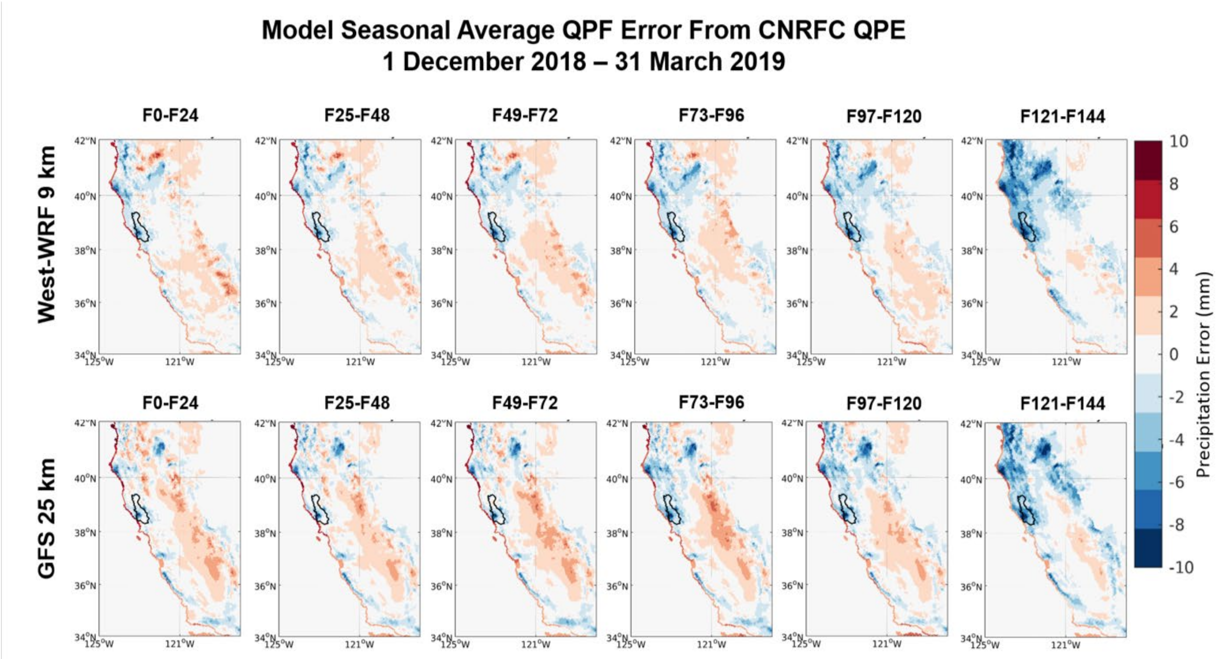

Figure 2.3.4 Mean 24-hr total QPF error (mm) computed over the period between 1 December 2018 and 31 March 2019 for (top row) West-WRF 9 km and (bottom row) GFS 25 km forecasts as a function of forecast lead time. The values above each panel represent the forecast hours used to aggregate the 24-hr QPF. The Russian River Watershed is outlined in black.

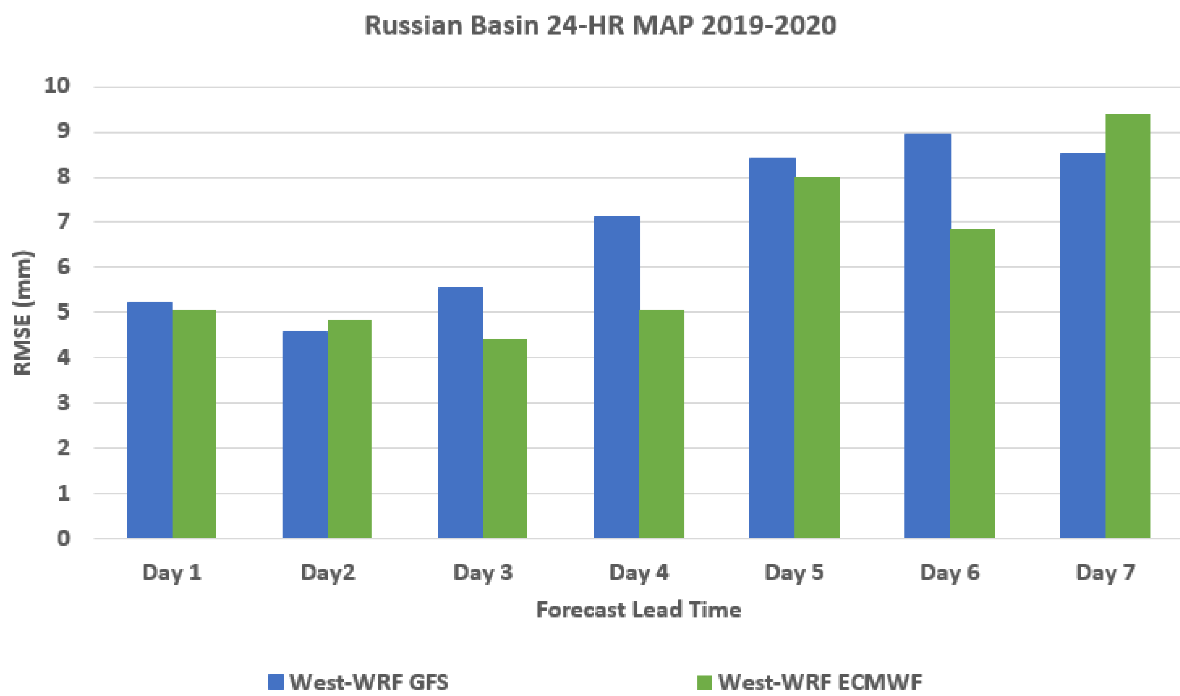

Figure 2.3.5 Centered root mean squared error (CRMSE) for 9-km West-WRF 24-hour precipitation over the Russian River basin for all matched forecast times currently available from NRT-GFS (blue) and NRT-ECMWF (green) as a function of lead time out to seven days. The sample size (days) for each lead time is overlaid in white.

Improvements to the West-WRF regional forecasting system will continue to leverage research, technology, and resources. The verification highlights the importance of multi-scale weather evolution and error propagation, where, for example, large shifts in the AR forecast location have significant impacts for watershed-level precipitation forecasts.

2.3.4. Identify key physical processes that need to be resolved in model forecasts

Improving the model representation of important physical processes is the foundation of forecast improvements. This task identifies and investigates the atmospheric processes and features that need to be better represented for forecast improvement. These include low level jets (LLJ); dynamical patterns that generate AR families; the impact of aerosols on cloud and precipitation processes; mesoscale dynamics; mesoscale frontal waves (MFWs); the interactions between extratropical cyclones and ARs; large scale circulations and their impacts on AR presence or absence; the impact of antecedent conditions and quantification of the effect of terrain and AR orientation on precipitation patterns associated with ARs; and advances in tracking ARs.

Larger scale circulations are an important factor contributing to the spatio-temporal evolution of ARs. For example, extratropical cyclone interactions with ARs can play a role in AR strength and impacts.

Zhang et al. (2019) found that 50% of cyclogenesis over the eastern Pacific involves a pre-existing AR, and 80% of ARs are associated with extratropical cyclones. A positive feedback was identified between the AR and cyclone, which intensifies the cyclone and enhances the AR IVT. This feedback is driven by water vapor provided by the AR, which enhances precipitation, resulting in latent heat release, and the stronger winds that may be associated with the cyclone. Fish et al. (2019) defined and characterized sequences of ARs occurring in quick succession as AR families (Figure 2.3.6), finding that these families are frequently associated with identifiable synoptic scale patterns. AR families can have impacts analogous to those of long duration ARs. The most extreme ARs affecting the Russian River are often the longest duration rather than the highest intensity (Figure 2.3.7, Dettinger et al., 2018; Lamjiri et al., 2017). This highlights the importance of knowing and forecasting durations and storm-total moisture transport to forecasting storm impacts and flooding and has been demonstrated for gridded precipitation data as well as point-scale in-situ observations in the Russian River. Focusing specifically on orientation of ARs affecting the Russian River, Hecht and Cordeira (2017) found that south-southwesterly oriented ARs produce, on average, more precipitation in the watershed than westerly oriented ARs. These landfalling ARs with different orientations occur in association with different synoptic-scale flow patterns. Additionally, the combination of water vapor flux direction and synoptic scale forcing for ascent can lead to greater precipitation accumulations.

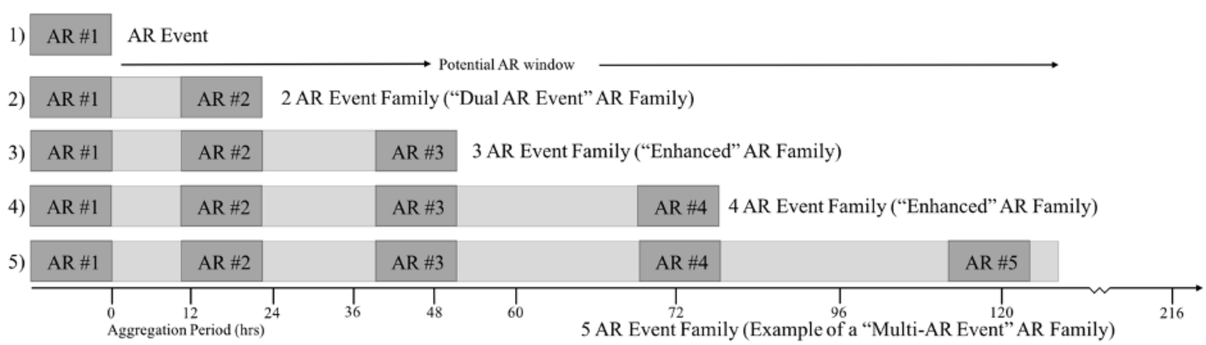

Figure 2.3.6 Schematic of the AR families definition. Adapted from Figure 2 of Fish et al.(2019).

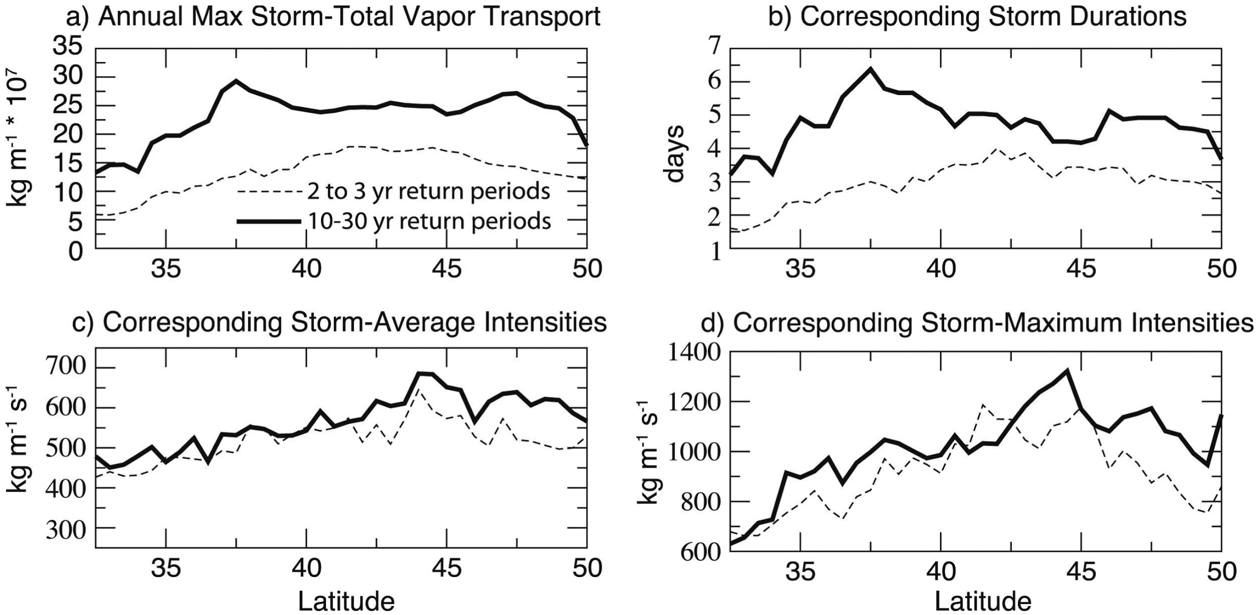

Figure 2.3.7 Latitudinal distribution of (a) the annual-maximum storm-sequence total vapor transport with return periods from 2 to 3 years and from 10 to 30 years, (b) the duration of these same storm sequences, (c) the average IVT over the entire storm sequence, and (d) the maximum instantaneous IVT during the storm sequence, for ARs or uninterrupted sequences of ARs making landfall along the U.S. West Coast, water years 1981–2016 (Figure 8 from Dettinger et al., 2018).

Mesoscale processes are another critical aspect that need to be resolved by models to accurately predict the impact associated with ARs. We have investigated key phenomena including mesoscale frontal waves (MFWs), latent heating and moist processes, and LLJs. MFWs have often been identified in ARs affecting the Russian River watershed. Multiple case studies have shown that MFWs can influence the intensity, location, and especially duration of ARs, and that they are challenging to forecast more than 1-2 days out. They often lead to serious forecast errors in landfall location and precipitation amounts. Martin et al. (2019a) highlights these issues in a December 2014 case study of an AR impacting the Russian River. Specifically, the study illustrates the need to identify offshore mesoscale frontal waves in real time and to characterize the forecast uncertainty associated with these events when creating hydrometeorological forecasts.

Using Global Precipitation Mission satellite data and West-WRF to investigate an impactful February 2017 AR, Cannon et al. (2020) illustrates a case where the precipitation processes and subsequent latent heat released offshore strongly influenced the AR's evolution. The study notes that, although these precipitation mechanisms are present in global-scale models, the difficulty that coarse-resolution models have in accurately representing the generated precipitation likely translates to uncertainty in forecasted heating tendencies, their impact on the AR evolution, and ultimately the impacts of ARs upon landfall.

Results from Demirdjian et al. (2020a) suggest that the precipitation forecast for an impactful AR that affected the Russian River in January 2017 was particularly sensitive to the initial moisture content within the AR due to its role in precipitation enhancement. Thus, the important processes that a model should represent include the evolution of moisture content and associated processes, including the latent heating. This case is representative of a typical AR evolution, which underscores the necessity of correctly modeling moist processes to improve forecasts of AR-associated precipitation.

The pre-cold-frontal LLJ is the core of the water vapor transport within ARs, though its dynamics are not completely understood. Dropsonde observations from AR Recon (see section 2.2.2) show that the LLJ typically is located at ~1-km elevation ahead of the cold front, with an average maximum wind speed of 30 m s^{-1}. The geostrophic wind is the result of the balance between the pressure gradient (flow from high to low pressure) and the Coriolis force (caused by the earth's rotation). There is an unbalanced portion of the geostrophic wind in the LLJ, known as the ageostrophic component, that occurs within the atmospheric layer from 750–1250 m known to strongly control orographic precipitation associated with ARs (Neiman et al. 2002, 2009). That means the unbalanced component of the LLJ adds vapor transport where it is most likely to enhance precipitation. Analysis in Demirdjian et al. (2020b) demonstrates that it is necessary to correctly model the LLJ, including relevant characteristics at all scales, to accurately forecast precipitation associated with ARs (Figure 2.3.8).

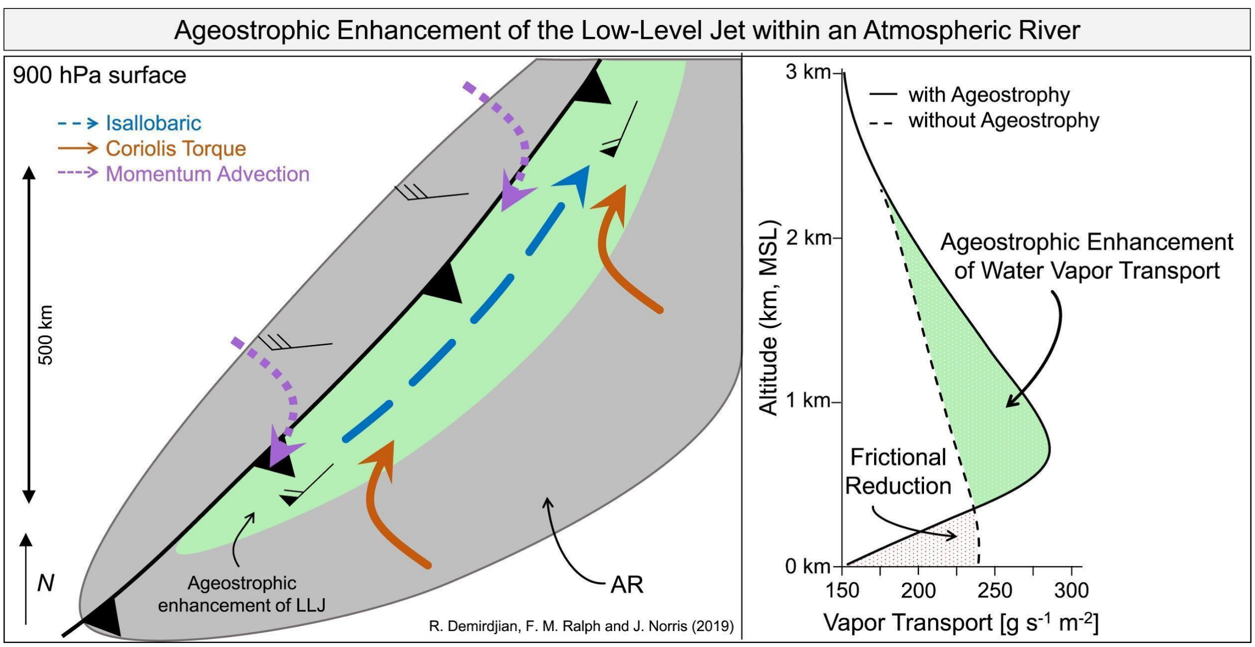

Figure 2.3.8 A conceptual representation of the enhancement (green) of water vapor transport within the low-level jet due to the ageostrophic winds and the dynamical processes serving to accelerate/decelerate the ageostrophic wind. Adapted from Figure 13 of Demirdjian et al. (2020b).

At the microphysical scale, aerosols in the atmosphere also influence AR characteristics. Thus far the focus of FIRO aerosols research at CW3E has primarily been on the impact of aerosols on cloud properties and precipitation generation. These studies use data collected in support of FIRO goals (see Section 2.2). Martin et al. (2019b) concluded that terrestrial and long-range-transported aerosols can be important sources of warm ice nucleating particles (INPs) during ARs. These warm INPs impact cloud microphysics, specifically mixed-phase cloud properties. In addition, Mix et al. (2019) investigated stable isotope compositions of surface-air water vapor along with independent meteorological factors such as temperature and relative humidity. As stable isotopes in precipitation and water vapor can help to trace rainout and post-condensation processes, they may be used to better characterize and explain macro scale structures (e.g., ARs, fronts, small-scale jets, and orography) and microphysical processes (e.g., mixed-phase hydrometeor processes, ice nucleating particles, below cloud evaporation) affecting precipitation amount and efficiency. It was clear from Mix et al. (2019) that rainout and post-condensation processes can dominate in different parts of ARs. These results and analyses contribute to answering important questions regarding the temporal evolution of AR events and the physical processes that control them at local scales, all of which affect AR impacts and predictability.

Finally, work has also been conducted to improve procedures for identifying and tracking ARs in historical data, forecasts, and future climate projections. FIRO PI F. Martin Ralph co-leads the Atmospheric River Tracking Method Intercomparison Project (ARTMIP), which is an international collaborative effort to understand and quantify the uncertainties in AR science deriving from the many detection algorithms available (Shields et al., 2018; Rutz et al., 2019). Ultimately, these uncertainties make results from most studies of the long-term properties of ARs and their impacts, and at least some short-term studies of individual ARs, somewhat detection-method dependent, in ways that can challenge the basics underlying FIRO forecasts and responses. For example, an initial study conducted for the Russian River watershed showed that tracking methodologies that are more spatially or temporally restrictive identify fewer AR events (Ralph et al. 2019). Thus, ARTMIP is needed to improve understanding and predictions of AR arrivals and impacts to better support the decision-making process in the future.

Table 2.3.2 Peer-reviewed results identifying key physical processes to be represented in models.

Reference	Primary Result
Cannon et al. 2020 (Mon. Wea. Rev.)	Precipitation processes and subsequent latent heat released offshore strongly influence AR evolution.
Demirdjian et al. 2020a (Mon. Wea. Rev.)	There is a substantial ageostrophic component of the low level jet which requires high resolution models to simulated.
Demirdjian et al. 2020b (J. Atmos. Sci.)	ARs are sensitive to moisture content because it modulates both the orographic and dynamic component of precipitation.
Fish et al. 2019 (J. Hydrometeor.)	Sequences of ARs occurring in quick succession of one another (AR families) are frequently associated with identifiable synoptic scale patterns.
Hecht et al. 2017 (Geo. Res. Lett.)	South-southwesterly oriented ARs produce, on average, more precipitation in the watershed than westerly oriented ARs.

Lamjiri et al. 2017 (Geo. Res. Lett.)	The most extreme ARs affecting the Russian River are often the longest duration rather than the highest intensity.
Lamjiri et al. 2018 (SF Estuary and Watershed Sci.)	ARs bring on average 80% of the extreme precipitation in about 100 hours per year.
Martin et al. 2018 (J. Hydrometeor.)	Improving water vapor transport accuracy can significantly reduce regional model precipitation errors during ARs.
Martin et al. 2019a (J. Hydrometeor.)	It's essential to identify offshore mesoscale frontal waves in real time and to characterize the forecast uncertainty associated with these events when creating hydrometeorological forecasts.
Martin et al. 2019b (Atmos. Chem. and Physics)	During ARs, terrestrial and long-range-transported aerosols can be important sources of warm ice nucleating particles (INPs), which can impact mixed-phase cloud properties.
Mix et al. 2019 (Atmospheres)	Rainout and post-condensation processes can dominate during different portions of ARs at local scales.
Ralph et al. 2019 (Clim. Dyn.)	19 +/- 4 ARs on average affect the Russian River each year.
Rutz et al. 2019 (JGR Atmospheres)	Methods with more restrictive AR identification and tracking criteria tend to have weaker statistics in terms of AR climatology, while the statistics for methods with less restrictive criteria are more robust.
Shields et al. 2018 (Geosci. Model Dev.)	ARTMIP is an international collaborative effort to understand and quantify the uncertainties in AR science based on detection algorithms alone to inform AR research.
Zhang et al. 2019 (Geo. Res. Lett.)	50% of cyclogenesis over the eastern Pacific have a pre-existing AR, and 80% of ARs are associated with extratropical cyclones. A positive feedback was identified between the AR and cyclone, which intensifies the cyclone deepening and also enhances the AR IVT.

The research conducted in support of this task has advanced our knowledge of the physical and dynamical processes determining the formation and evolution of ARs. Future efforts that have strong potential to further improve FIRO forecasts and benefits include:

- More accurately quantify impacts on AR precipitation from transport directions and vertical distributions of water vapor.
- Continue work on understanding MFWs and the key physical processes that drive their evolution in order to improve their representation in models.
- Machine learning algorithms deployed to exploit some really large data sets, such as the CW3E reforecast, to enhance our understanding of key physical processes.

These and other efforts will enable improving the representation of physical processes in models, and thus the accuracy of precipitation forecasts at longer and longer lead times.

2.3.5. Reforecast, machine learning, and post-processing

CW3E has developed a 34-year regional reforecast to assess the benefits of a higher resolution forecast to CNRFC operations and streamflow forecasts than is provided currently by the NCEP/CNRFC system. The CNRFC ensemble streamflow forecasts are a critical component to support FIRO, and currently these streamflow forecasts rely on the National Centers for Environmental Predictions (NCEP) 30-year GEFS V10 reforecast (~1 degree resolution) meteorology. To produce the ensemble streamflow forecasts that are used in the Ensemble Forecast Operations Model (see section 2.7), the 30-year NCEP reforecast calibrates the Meteorological Ensemble Forecast Processor (MEFP), which provides forcing for the Hydrologic Ensemble Forecast Service (HEFS) for the streamflow reforecasts. To develop its higher resolution reforecast, CW3E ran West-WRF, using the same physical packages and computational domains identical to those used for the 2019-2020 NRT season, to dynamically downscale the ~0.5 degree control member of the Global Ensemble Forecast System (GEFS) to 9 and 3 km for water years 1986-2019. All simulations were run on the U.S. Army Engineer Research and Development Center (ERDC) DoD Supercomputing Resource Center (DSRC) Onyx Cray XC40/50 system. The whole date set includes about 320 TB of data. The reforecast dataset consists of daily West-WRF initializations at 0000 UTC from 1 December to 31 March, out to 7 days for the 9-km domain and 5 days for the 3-km domain (Figure 2.3.2). Similar to the West-WRF NRT, verification metrics are being computed for the reforecast precipitation over several California watersheds, AR strength and landfall position, and near-surface meteorological conditions. To the best of our knowledge, this is the only high-resolution reforecast effort that has been undertaken at a climate time scale (30+ years) to focus on a region in the US as it requires a significant effort of individuals, computing resources, and storage.

In addition to assessing the benefits of a West-WRF high-resolution reforecast to CNRFC operations, the reforecast is also a data set that allows to develop and test post-processing techniques and machine learning methods to reduce raw model output biases and ultimately enhance predictive capabilities, as well as perform in-depth process-based studies. We are exploring =ML algorithms in a post-processing framework to try to correct some of the biases in forecasts by training ML processes with an historical data set that includes both past predictions with forecast models and corresponding observations. We have developed a convolutional neural network (CNN) to improve GFS 0-7 days forecasts of IVT (Chapman et al. 2019) using a 10-year training data set. Figure 2.3.9 shows the performance of this novel approach, called GFSnn. GFS predictions are improved by about 5% at the beginning of the forecast to over 20% seven days out, when GFS is worse than climatology. Effectively GFSnn extends GFS's period of skillful predictions, based on root-mean-squared-error (RMSE) and other metrics. The GFSnn method has been run in real-time for the latter part of the 2020 WY as part of the NRT prediction systems with results made available on the CW3E website.

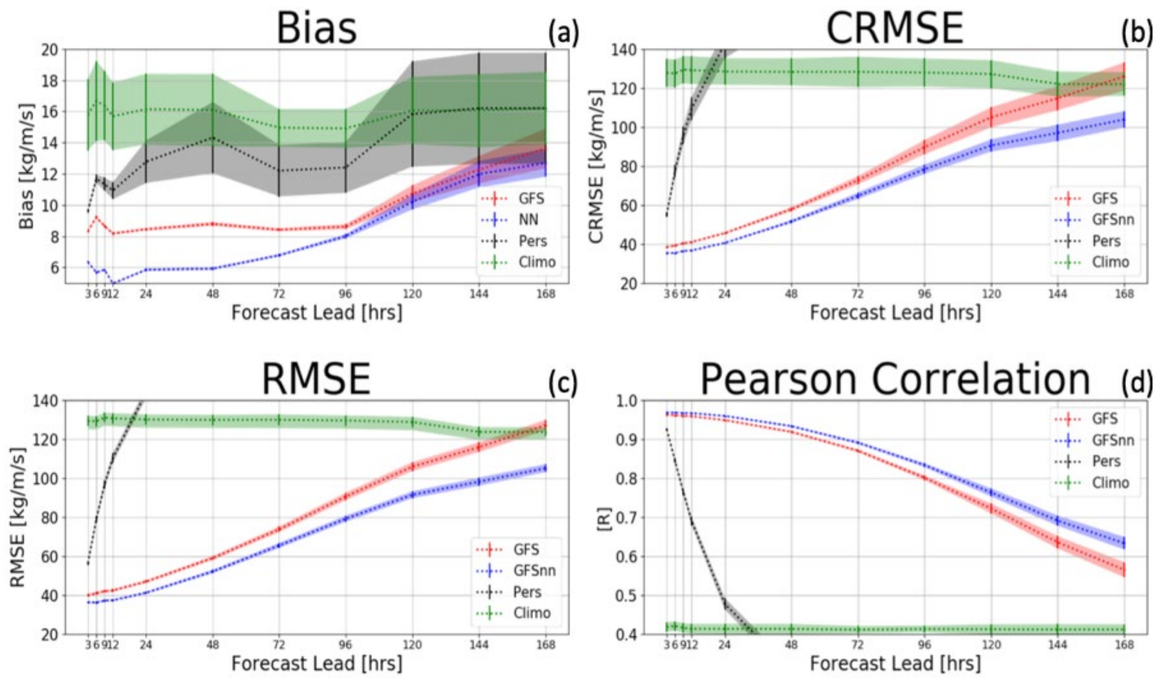

Figure 2.3.9 Performance IVT metrics computed over the 2018WY and the East Pacific / Western U.S. for the 0-168 hour prediction of IVT: (a) bias, (b) centered root mean square error (CRMSE), (c) RMSE, and (d) Pearson correlation. Shown methods include the NOAA's Global Forecast System (GFS, red), the CNN correction of GFS (GFSnn, blue), persistence (black), and climatology (red). For this study, as a ground-truth data set we used the National Aeronautics and Space Administration Modern-Era Retrospective analysis for Research and Applications version 2, MERRA-2). Adapted from Chapman et al. (2019).

Both the high resolution 34-year reforecast focused on extreme precipitation and post processing approaches that were developed during the FIRO effort offer opportunities to improve weather forecasts. The reforecast will provide collaborators at the CNRFC with a much higher resolution weather data to generate streamflow ensemble predictions, operational forecasts that for Lake Mendocino drive the Ensemble Forecast Operations Model (EFO; see section 2.7). The reforecast will also enable in-depth evaluation of West-WRF representation of key physical processes that drive extreme precipitation events. FIRO research has demonstrated that ML techniques can significantly improve forecasts from operational models. In addition, verification of the West-WRF has identified model biases (see section 2.3.4) that statistical post-processing approaches can target to improve the forecasts. The immense amount of data from the reforecast provides the training data needed to fully exploit the potential of both statistical and ML post-processing approaches to further improve the forecast.

2.3.6 Development and assessment of uncertainty quantification for large-precipitation events in West-WRF

Uncertainties regarding forecast initial conditions and the model representation of key processes limit the predictability of extreme precipitation. Efforts to understand the sources of West-WRF QPF error during ARs (Martin et al., 2018) and to generate probabilistic precipitation forcing for hydrologic modeling (e.g., Section. 2.5.3; Fig. 2.5.7) include the development of ensemble simulations that attempt to span the range of uncertainties in the model's initial state, unresolved processes, and parameterization error (Berner et al. 2014). These simulations provide robust demonstrations of QPF skill and spread that translate into confidence in the model's representation of extreme precipitation events in the Russian River watershed. Other sources of error in West-WRF simulations have been investigated by evaluating the impact of selected perturbations to the model's representation of extreme precipitation mechanisms (e.g., upslope flux, mesoscale frontal waves, and frontal rainbands). This research enables also the identification of key model deficiencies that highlight opportunities for (eventual) forecast improvements.

For example, sensitivity tests in simulations of five extreme events impacting the Russian River watershed led to significant changes in the West-WRF configuration employed in both the NRT (see section 2.3.3) and the reforecast (see section 2.3.5). These changes included an expanded model domain to better simulate offshore AR evolution and an increased number of model levels to better resolve AR interactions with terrain. Ensemble simulations were used to identify meteorological features that, when uncertain, cause divergence of simulated outcomes and forecast, such as MFWs, which in turn become targets for improved or increased observations (e.g., AR Recon, see section 2.2.2) to reduce these problematic uncertainties. Twenty ensemble members were created using 10 variations in the model physics and 10 variations in the subgrid cell processes to mimic unresolved uncertainties (i.e., via the Stochastic Kinetic Backscatter, SKEBS) for 6 total AR events. These ensembles were used to force hydrologic models to understand the relationship between atmospheric uncertainty and hydrologic sensitivity (see section 2.5.3). Increased use of ensembles in the FVA has documented West-WRF's ability to span the realistic ranges of precipitation extremes for individual events.

2.3.7. Application of advanced data assimilation methods for extreme event forecasting

The ability of an NWP model to produce accurate predictions is limited by imperfect initial conditions and representations of physical processes. Data assimilation (DA) methods blend the information provided by observations with a model forecast, resulting in an improved estimate of the current state of the atmosphere, which in turn becomes the initial condition for future simulation and forecasts. When observations are sparse, as over the Northeastern Pacific, DA techniques are needed so that the model itself can fill in conditions between observations with physically realistic estimates. Specifically, in the FIRO context, advanced DA methods are the key to improve simulations of the origins and evolution of ARs, to improve forecasts of the landfalling ARs and precipitation. Towards this end, we have performed a series of studies, including i) assessing the impacts of observation gaps in ARs over the upstream Northeastern Pacific, with particular focus on the value of AR Recon data (see section 2.2.2) and ii) comparisons of the impacts of several different DA methods on AR forecasts.

A significant data gap typically exists in the lower atmosphere (below 450 hPa) during AR events over the Northeastern Pacific. Radiance data from satellites, which is assimilated in the NCEP Global Forecast System model, doesn't address this gap. However, AR Recon data (see section 2.2.2) provide the majority of observations within the critical layer of an oceanic AR (Figure 2.3.10; Zheng et al. 2020a) during those missions. In 2016, 2018, and 2019, across 15 missions, the dropsondes provided about 99% of humidity, 78% of temperature, and 45% of wind observations over the area where an AR was present.

The impact of these AR Recon dropsonde data are examined in data denial experiments wherein simulations without the dropsondes are compared to simulations that include them and to observations. As shown in Figure 2.3.11, preliminary results show improved simulations of 24-hr precipitation totals when the dropsondes are included. Improvements in RMSE are obtained at most lead times and for most IOPs. Improvements are found for most lead times and all IOPs when correlations are used to measure success. In the case of consecutive missions that were executed every other day, the improvements resulting from dropsonde DA were consistently larger in a mission compared to the previous one (Zheng et al. 2020b). Additionally, Stone et al. (2020) demonstrated that the assimilation of AR reconnaissance soundings results in a forecast error reduction similar to the North American radiosonde network for the days when at least two flights were flown.

Analyses and forecasts of landfalling ARs are sensitive to the DA methods used. A significant finding from the AR Recon data denial experiments is that a four-dimensional DA scheme, hybrid 4D-EnVar, provides more realistic initial conditions and thereby improved downstream precipitation forecasts when compared to a three-dimensional approach, the hybrid 3D-EnVar, particularly for fast moving systems associated with complex environments (e.g., the genesis and termination stage of an AR; Zheng et al. 2020c). In applications of the hybrid 4D-EnVar method to West-WRF, AR Recon data have been shown to improve forecasts for landfalling ARs and precipitation in 12 out of 15 cases. The West-WRF hybrid 4D-EnVar system allows the assessment of the value of particular field observations, increases the skill of forecasts of extreme events, and is an important tool for improving the understanding of AR dynamics and thermodynamics. Our research shows also that the further development of data assimilation algorithms will likely improve the forecast ability to leverage the information provided by observations. For example, new algorithms can be developed to better capture error characteristics associated with different atmospheric processes (e.g., by implementing adaptive error inflation and covariance localization).

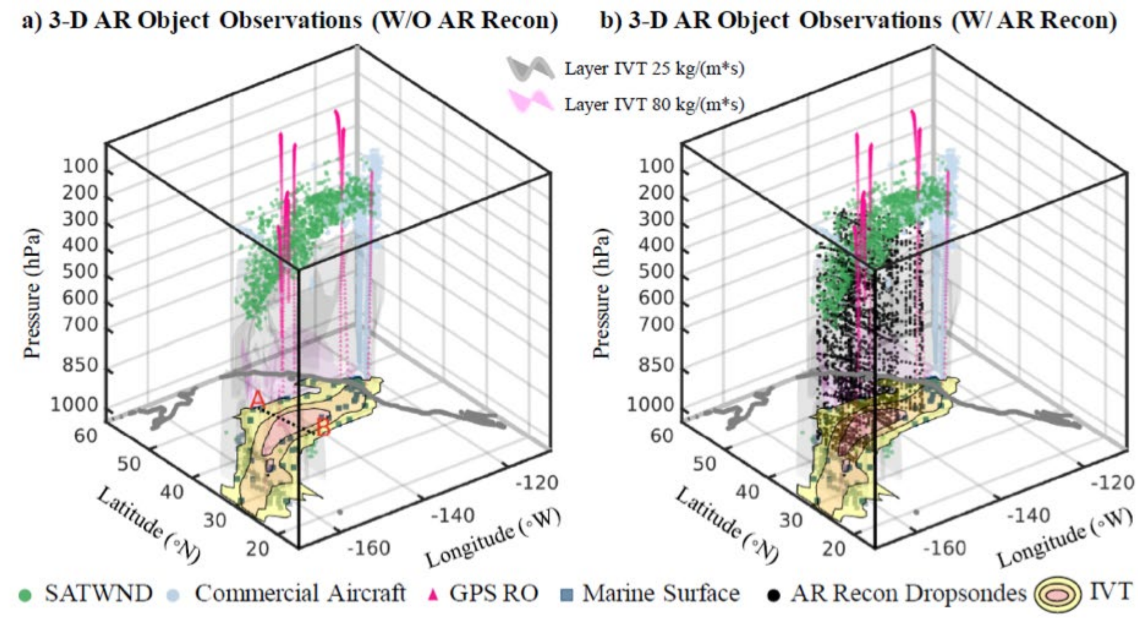

Figure 2.3.10 Three-dimensional illustration of observation distributions for non-radiance data (a) without and (b) with AR Recon flight-level and dropsonde data (black filled circles). The black dots are the AR Recon data used in the operational GFS. This figure is based on the observations for 2016 IOP1. The grey and pink shaded on (a-b) are the isosurface for 50th (25 kg m^{-1} s^{-1}) and 95th (80 kg m^{-1} s^{-1}) layer IVT values, respectively. Surface shading and contours are for the total IVT value starting from 250 kg m^{-1} s^{-1} with 250 increments. Adapted from Zheng et al. (2020).

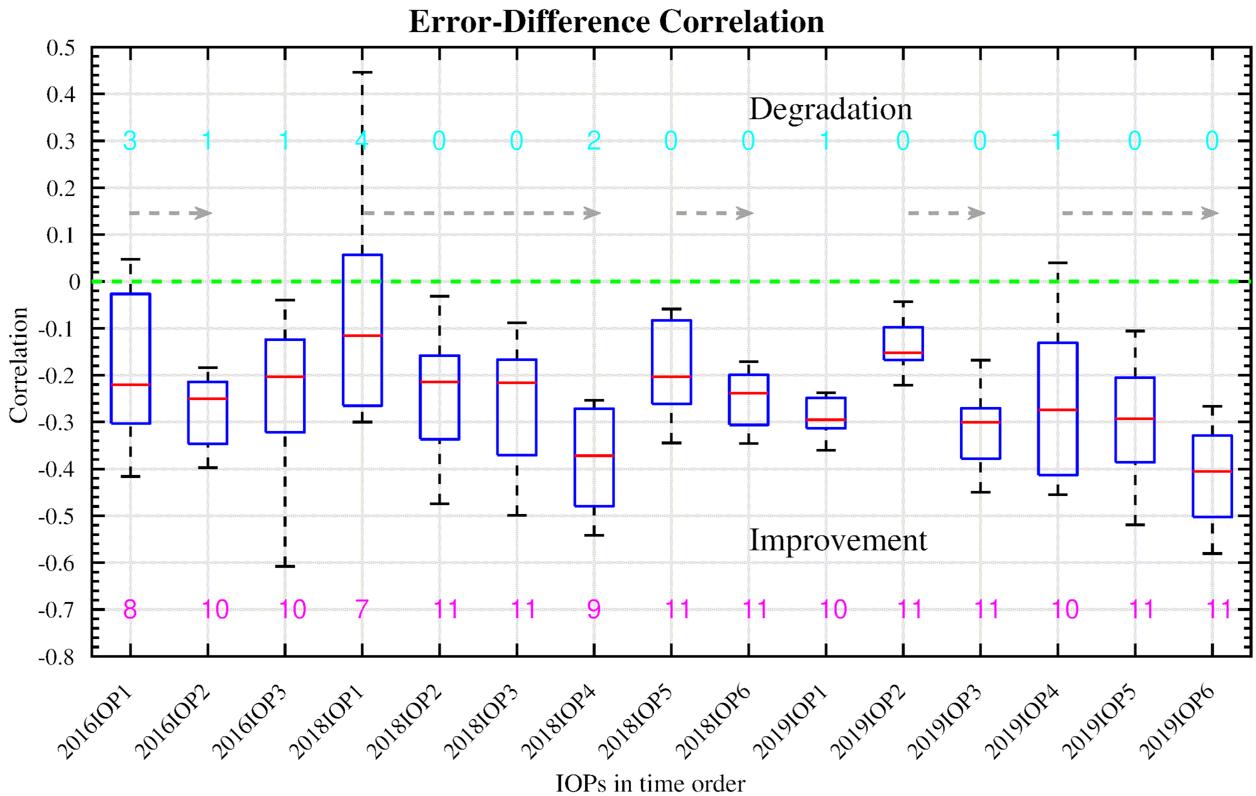

Figure 2.3.11 Improvements and degradation (box plot) of 24-h accumulated prediction forecasts from day 1 to day 6 (a total of 11 lead times) verified over the western U.S. (30°N-50°N, 125°W-110°W) resulting from the assimilation of dropsondes data collected during the 15 AR Recon IOPs, based on boxplot of spatial correlation, across 15 AR Recon IOPs. At top and bottom of each IOP is the number of lead times for which the experiment resulted in an improvement (pink) or degradation (blue). Grey arrows indicate the consecutive missions when two missions were within 3 days. Adapted from Zheng et al. (2020b).

2.3.8 Conclusions and Recommendations

By its nature, FIRO depends on the availability of skillful forecasts of storms and storm interludes. An assessment of CNRFC forecasts has shown that the forecasts are already sufficiently skillful to support FIRO viability at Lake Mendocino but improvements in forecasts are still possible and would further enhance FIRO benefits. Evaluating and improving the skill of forecasts (especially, week-one forecasts) is thus central to the viability and value of FIRO. This research highlights several processes that determine the impact of ARs. These studies provide the guidance on model processes that are important to improving future forecasts. The development of West-WRF and verification processes have shown the value of a regional model focused on extreme precipitation events, in particular ARs. Novel testing of post-processing and data assimilation approaches tested under FIRO has illustrated how new technologies can improve forecast skill. Together, the multi-pronged approach to week-1 forecasts described here has shown that forecast improvements are possible as well as providing foundational information for recommendations as to how to improve week 1 forecasts of extreme events and ARs.

The following recommendations will enhance the benefits of week-one forecasts for use in FIRO:

- Test better forcings (e.g., input boundary and initial conditions), improvement and tuning of key physical parameterizations (see section 2.3.3) of the West-WRF modeling system.
- Provide results from research in a usable way to stakeholders, e.g., via decision support and situational awareness tools (see section 2.7).
- Verify forecasts using metrics that are relevant for ARs and water management.
- Implement ensemble-prediction methods using West-WRF to provide forecast uncertainty information along with the NRT forecasts.
- Leverage the new West-WRF reforecast data to develop and test ML tools for even larger forecast improvements, including sharper and more reliable probabilistic predictions.
- Enhance algorithms to better assimilate the observations collected during AR Recon, by including high-vertical-resolution dropsondes, airborne radio occultation measurements, and buoy surface pressure.