

Developing West-WRF for Western United States Atmospheric River

Science and Impacts: Initial Implementation and Evaluation

Andrew Martin*

*Center for Western Weather and Water Extremes, Climate Atmospheric Science and Physical
Oceanography Division, Scripps Institution of Oceanography, La Jolla, CA, USA*

F. Martin Ralph

*Center for Western Weather and Water Extremes, Climate Atmospheric Science and Physical
Oceanography Division, Scripps Institution of Oceanography, La Jolla, CA, USA*

Reuben Demirdjian

*Center for Western Weather and Water Extremes, Climate Atmospheric Science and Physical
Oceanography Division, Scripps Institution of Oceanography, La Jolla, CA, USA*

Laurel DeHaan

*Climate Atmospheric Science and Physical Oceanography Division, Scripps Institution of
Oceanography, La Jolla, CA, USA*

Sam Iacobellis

*Climate Atmospheric Science and Physical Oceanography Division, Scripps Institution of
Oceanography, La Jolla, CA, USA*

18 Brian Kawzenuk

19 *Center for Western Weather and Water Extremes, Climate Atmospheric Science and Physical*
20 *Oceanography Division, Scripps Institution of Oceanography, La Jolla, CA, USA*

21 David Reynolds

22 *Cooperative Institute for Research in Environmental Sciences, Boulder, CO, USA*

23 Tamara Shulgina

24 *Climate Atmospheric Science and Physical Oceanography Division, Scripps Institution of*
25 *Oceanography, La Jolla, CA, USA*

26 John Helly

27 *San Diego Supercomputing Center, University of California, San Diego, La Jolla, CA, USA*

28 **Corresponding author address:* Andrew Martin, Climate, Atmospheric Science and Physical
29 Oceanography, Scripps Institution of Oceanography, 9500 Gilman Drive # 0224, La Jolla, CA
30 93093-0224.

31 E-mail: mc@ucsd.edu

ABSTRACT

Precipitation from Atmospheric Rivers presents a unique forecast challenge because the critical spatial scales span a few km, i.e. local scales, to thousands of km, i.e. storm scales. A compartmentalized approach is often employed to simulate local and storm scales. We evaluate forecasts of well observed Atmospheric Rivers using a modeling system designed to simulate both the storm scales and the local scales. An unprecedented number of observations from within moderate to strong atmospheric rivers including airborne, balloon-borne, ground based and remotely sensed measurements have been gathered from the CalWater 2 intensive observing periods. These observations are used to investigate whether the West-WRF modeling system adds value to global model forecasts and to a regional model employing smaller extent than West-WRF during atmospheric river events. The value added to atmospheric river water vapor transport, static stability, onshore precipitation and standard storm-scale atmospheric fields are investigated. It is found that West-WRF skill in forecasting storm-scale features is comparable to that of its parent global model. It is also found that precipitation forecast skill is improved in the West-WRF high-resolution domain. Further direct validation is presented that investigates the relative roles of storm scale forcing and local orographic response in driving errors in forecast precipitation using an observationally derived linear model. The results are used to identify best practices in NWP model design for Atmospheric River forecasting and to suggest avenues of future model improvement.

54 **1. Introduction**

55 Atmospheric Rivers (AR) play a vital role in delivering rain and snow to western North America
56 (Dettinger et al. 2011; Guan et al. 2010; Ralph et al. 2010a; Neiman et al. 2013). In some regions,
57 as much as 50% of annual precipitation falls on days when an AR is present (Rutz et al. 2014).
58 AR have been shown to be a key mechanism in regional recovery from drought (Dettinger 2013),
59 and have been linked to major flooding events in western North America and elsewhere (Neiman
60 et al. 2008, 2011; Lavers et al. 2011; Ralph et al. 2010a, 2006; Moore et al. 2012). Despite the
61 importance of AR to water supply and flood risk, the precipitation resulting from AR remains
62 poorly forecast (Lavers et al. 2016; Wick et al. 2013; Junker et al. 2009). We will argue that
63 forecast performance remains poor in part due to an insufficient approach to simulating the large
64 number of spatial scales important during an AR. We will introduce and evaluate the value added
65 by a model system designed to address this scale issue.

66 Let the parameter λ represent a scale horizon, the variable ds represent the largest spatial scale
67 at which an AR-related physical process is observable, and the parameter A represent the largest
68 spatial scale directly linked to AR. Following Zhu and Newell (1998) and others (Ralph et al. 2004;
69 Bao et al. 2006; Neiman et al. 2008; Ralph et al. 2010b), we estimate A is approximately 2500 km.
70 This scale lies between the characteristic extratropical cyclone scale and the scale characteristic
71 of polar cold fronts in their along-frontal direction. Other authors (McCabe and Dettinger 2002;
72 Dole 2008; Newman et al. 2012; Guan et al. 2012, 2013) have described indirect impacts that
73 larger scales (hemispheric, planetary) can have on AR and related extratropical systems, but we
74 will discard these scales in the current discussion.

75 To illustrate how λ impacts practical aspects of AR forecasting, we will consider only forecasts
76 of AR created by deterministic numerical weather prediction systems (NWP). To create such a

77 forecast, one needs to correctly predict a subset of storm scale ($\lambda < ds < A$) atmospheric pro-
78 cesses. These include AR location, extent and movement, vertical distributions of wind, water
79 vapor and buoyancy, and storm scale patterns of rising/sinking motion. In addition, one must cor-
80 rectly predict the local scale ($ds < \lambda$) precipitation response to the storm scale processes. These
81 local scale responses include orographic lift, terrain blocking, surface energy exchanges, turbu-
82 lence, and cloud microphysics. In practice, λ will be set by forecast scenarios and location(s). For
83 AR along the US West Coast, the local scale will be defined by the extent of orographic or convec-
84 tive clouds, mountain ridges, and valleys. Turbulence, surface energy fluxes, cloud microphysics,
85 etc. are defined by smaller scales and thus do not determine λ in this scenario and location.

86 Global numerical weather prediction (GNWP) models, those that can explicitly simulate the
87 largest weather scales on Earth, cannot yet explicitly simulate most local scale responses. The
88 challenges of predicting the local scale response to large scale weather systems (including AR) has
89 traditionally been handled one of two ways: by parameterizing physical processes at unresolved
90 scales (e.g. a cumulus model in GNWP), or by dynamical downscaling of a GNWP forecast
91 using a regional weather prediction model (RNWP). In the second case, the RNWP domain has
92 limited spatial extent and relies upon the GNWP forecast for its boundary condition, but because
93 it has finer native resolution it is able to explicitly resolve much smaller scales without the aid of
94 empirical parameterization.

95 The limited extent (X) of RNWP and their finer resolution introduces a trade-off. To create a
96 forecast at reasonable computational expense while achieving the smallest possible scale horizon,
97 X is often made smaller than A . Thus, some storm scale processes are no longer explicitly resolved
98 in order that more local scale processes become explicitly resolved. The hope in making this trade-
99 off is that the scales $X < ds < A$ will be accurately communicated to the RNWP by means of the
100 GNWP boundary condition. However, these boundary conditions will by nature be “non-native” to

101 the RNWP model and thus introduce dynamical imbalances, will evolve at timescales more coarse
 102 than the RNWP native time step, and will be adapted to the RNWP by means of interpolation.
 103 Therefore, RNWP suffer from boundary condition errors in addition to the more general initial
 104 condition, dynamics and physics errors (Cavallo et al. 2016). The impact of the trade-off on
 105 forecast error is that while dynamics errors may be decreased by explicitly resolving more local
 106 scales, the limited model extent may counteract this decrease by introducing errors at storm scales
 107 larger than X through the boundary condition error. Because the largest AR scales are greater
 108 than the domain extent in most RNWP ($A > X$) and AR precipitation is heavily influenced by
 109 very small scales not explicitly resolved at the horizontal resolution (dx) of current GNWP (i.e.,
 110 $\lambda < dx$), common NWP approaches to predicting AR precipitation suffer from forecast error at
 111 either very large or very small scales.

112 We propose that skill in deterministic forecasts of AR may be improved by simulating the largest
 113 expected scales of AR with a higher resolution and highly configurable (thus tailored to AR)
 114 RNWP model with sufficient extent and by simulating the local precipitation response in a nested
 115 RNWP at high resolution. Our RNWP system will hereafter be called “West-WRF”. West-WRF
 116 is based upon NCAR’s open-source WRF-ARW model. It is the primary NWP tool dedicated to
 117 meeting the mission and goals of the Center for Western Weather and Water Extremes (CW3E
 118 - <http://cw3e.ucsd.edu/>). These goals are motivated by special forecast requirements posed
 119 by Western US extreme weather events (see Table 1). In addition to operational forecast goals,
 120 West-WRF is designed as a platform from which to evaluate the sources of forecast error and
 121 their relationship to physical processes. West-WRF is configured with 2 domains utilizing one-
 122 way nesting. The larger parent domain is designed to forecast the largest AR storm scales, while
 123 the smaller but higher resolution nested domain explicitly resolves many local response scales.
 124 Communication of parent domain state to the local domain has been optimized by the WRF-ARW

125 developers so that boundary condition errors in the nest domain are minimal (Skamarock 2008).

126 We hypothesize that West-WRF forecasts will

- 127 • Forecast the large scale properties of atmospheric rivers with equal or greater skill than does
128 its parent GNWP model (the reduction in error due to poorly resolved dynamics will at least
129 counter the disadvantage created by boundary condition error) and
- 130 • Given equally or more skillful storm scale inputs, the local response simulated in the high-
131 resolution nested domain will incur reduced dynamic error and will more accurately reflect
132 observed local precipitation response.

133 The second hypothesis could be called the “right answer for the right reasons” hypothesis. It is not
134 considered sufficient to better forecast precipitation while simultaneously displaying larger error
135 in storm scale AR measures. Satisfying both hypothesis will establish that our proposed NWP
136 system is one for which further reduction in boundary condition, initial condition, dynamics and
137 physics errors can be reasonably expected to reduce precipitation error during AR.

138 In the remainder of this manuscript, we will demonstrate the above hypotheses by validating
139 West-WRF forecasts of AR large scale features using airborne dropwindsondes from the CalWater
140 2 research flights. We will also validate forecasts of local response using observations of precipi-
141 tation, vertical wind profiles, and vertically integrated water vapor during AR that impacted Cali-
142 fornia’s Russian River watershed (RRW). We will compare West-WRF’s performance in standard
143 skill metrics to the performance of GNWP forecasts and small-domain RNWP forecasts. We will
144 use these comparisons to demonstrate the performance by West-WRF in forecasting large scale
145 AR features (hypothesis 1) and the concomitant value added in precipitation forecasts (hypothesis
146 2). Furthermore, we will compare a decade-long observational estimate of storm-scale forcing and
147 local precipitation response to forecasts from West-WRF and GNWP. Using a linearized local re-

148 sponse function, we will investigate how best to optimize each model for orographic precipitation
149 during an AR.

150 **2. Data and Forecast Models**

151 *a. Atmospheric River Observatory*

152 The National Oceanic and Atmospheric Administration’s (NOAA) Hydrometeorological
153 Testbed has operated an “Atmospheric River Observatory” (ARO) at two sites, Bodega Bay and
154 Cazadero, CA, since 2006 (White et al. 2009, 2013). The Bodega Bay (BBY) site is situated on
155 the coast at sea level and is designed to monitor horizontal vapor flux, integrated water vapor and
156 horizontal winds in the Atmospheric River low-level jet region as they impinge upon the orograph-
157 ically productive coastal mountain ranges. The Cazadero Site is located north of Bodega Bay at
158 the top of a prominent ridge in the coastal mountain system. Cazadero drains to the South and
159 Wheatfield forks of the Gualala River. The Cazadero site reports precipitation, vertical S-Band
160 radar reflectivity and precipitation drop size distributions during AR conditions at mountaintop. In
161 this study, we will use the couplet of sites to investigate storm scale forcing near the coastal edge
162 of the Russian River Watershed (RRW) through Bulk Upslope Flux (*BUF*) measured at BBY and
163 local scale response via accumulated precipitation measured at CZC.

164 *BUF* is calculated following the methods of (Neiman et al. 2002, 2009) in which the controlling
165 layer (800 m MSL to 1200 m MSL) winds are multiplied by the local Integrated Water Vapor
166 (*IWV*) and projected onto the mountain orthogonal direction. Controlling layer wind is calculated
167 from the 449 MHz wind profiler at BBY. *IWV* is calculated via radio occultation from the BBY
168 GPS Trimble receiver. Rainfall accumulation at CZC is measured by a tipping bucket rain gauge

169 and reported to 0.1 mm precision. Hourly, quality controlled *BUF* and accumulated rainfall are
170 available from the NOAA Earth System Research Laboratory via anonymous ftp server.

171 *b. GPS-Enabled Soundings*

172 Airborne dropwindsondes from the CalWater 2 early start and CalWater 2015 intensive observ-
173 ing periods (Ralph et al. 2016) are used extensively in our analysis as a model verification tool.
174 Each sounding occurred during a Northeast Pacific AR transect performed by CalWater aircraft.
175 An example, with two transects used in this study is shown in figure 1. Transects were executed
176 to maintain a flight path perpendicular to the direction of troposphere integrated water vapor flux.
177 179 sondes from fifteen transects taken during 10 separate AR flights are used in this study (Table
178 2). We require the transect to cross the full width of the AR core. Herein, the AR core boundaries
179 are defined by the isopleth where Integrated Vapor Transport (*IVT* - Cordeira et al. (2013)) exceeds
180 $500 \text{ kg m}^{-1} \text{ s}^{-1}$. This threshold does not strictly follow from previous literature, but is designed
181 to capture the region of greatest horizontal vapor transport within the AR. If the transect includes
182 dropwindsondes which recorded $IVT \geq 500 \text{ kg m}^{-1} \text{ s}^{-1}$ and at least one poleward and one equa-
183 torward dropwindsonde which recorded $IVT < 500 \text{ kg m}^{-1} \text{ s}^{-1}$, then the transect is considered
184 to have sampled the full AR core. Lastly, terminal dropwindsondes must be greater than 100 km
185 from the West-WRF 9 km boundaries to reduce the chance that imprecise advection closure im-
186 pact the forecast soundings. Soundings were used to estimate 500 hPa geopotential height (z^{500}),
187 *IVT*, Integrated water vapor (*IWV*), partial *IVT* in a discrete layer (*dIVT*) and 925 hPa equivalent
188 potential temperature (θ_e^{925}) for the purpose of forecast verification.

189 *c. GEFS Reforecasts*

190 The National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System
191 (GEFS) model serves as the GNWP for the tests presented herein. We required that all forecast pe-
192 riods used the same deterministic model. To satisfy this requirement, we acquired control member
193 forecasts from the National Oceanic and Atmospheric Administration (NOAA) Global Ensemble
194 Reforecast Data Set (Hamill et al. (2013) - hereafter GFSRe). The GFSRe dataset is produced
195 with GEFS version 9.0.1, run at approximately 40 km native resolution. GFSRe reforecasts are
196 initialized once per day at 00 UTC and run to 192 hour lead time. GFSRe serves as the West-WRF
197 parent model in this study.

198 *d. West-WRF*

199 West-WRF is a special configuration of the WRF-ARW (Skamarock 2008) open-source numer-
200 ical weather prediction system, unified and distributed by the National Center for Atmospheric
201 Research (NCAR). West-WRF is configured with two domains utilizing horizontal resolutions
202 of 9 km (hereafter West-WRF 9 km) and 3 km (hereafter West-WRF 3 km), respectively. The
203 West-WRF 9 km domain has a much larger Earth relative footprint (Fig. 2, box “a”) and utilizes
204 interpolated forecasts from GFSRe as boundary conditions. Initial conditions for both West-WRF
205 domains are interpolated from the GFSRe analysis to the West-WRF 9 km and 3 km grids using
206 the WRF Preprocessor (Wang and coauthors 2012). The West-WRF 3 km domain Earth relative
207 footprint (Fig. 2, box “c”) covers most of the US state of California and portions of Western
208 Nevada and Southern Oregon. The RRW, where precipitation is verified in this study, lies near the
209 center of the 3 km domain. Both West-WRF 9 km and West-WRF 3 km are configured with 60
210 vertical levels with compressed spacing near the 925 hPa and the 300 hPa level in a US Standard
211 Atmospheric Sounding. West-WRF 9 km (West-WRF 3 km) uses static land surface information

212 generated from the USGS land use database (Wang and coauthors 2012) at a resolution of 5 (2)
213 arc-minutes.

214 West-WRF domains have been configured such that the parent domain simulates AR storm-scale
215 features, while the nest domain is designed to capture local scale response during heavy orographic
216 precipitation induced by AR. Figure 2 illustrates the storm-scale informed design of West-WRF.
217 The figure depicts the conditional probability $P(AR_{ij}|AR_{RRW})$, calculated using 25 years (1991 -
218 2015) of NCEP/NCAR reanalysis version 1 (Kalnay et al. 1996) and the AR catalog of Gershunov
219 et al. (2016)¹. The probability reflects the chance that AR conditions, defined as reanalysis $IVT \geq$
220 $250 \text{ kg m}^{-1} \text{ s}^{-1}$, are present at the reanalysis grid point i, j given that an atmospheric river existed
221 in the RRW according to the Gershunov et al. catalog. $IVT \geq 250 \text{ kg m}^{-1} \text{ s}^{-1}$ is here used as
222 a minimum but insufficient requirement for AR presence (Rutz et al. 2014) at each reanalysis
223 gridpoint. The measure shown in Fig. 2 was used to verify that the West-WRF 9 km domain likely
224 encloses the entire AR, and thus the storm-scale forcing, for a large percentage of AR impacting
225 the RRW.

226 The ability of West-WRF to simulate local scale response is partially accomplished by the rela-
227 tively (to current NWP systems) high horizontal and vertical resolutions of the model dynamics.
228 Additionally, this ability relies upon high resolution land surface and topographic information (2
229 arc-minute USGS data) and on sub-grid scale parameterized physics. The full list of physics op-
230 tions used with West-WRF 9 km, West-WRF 3 km and other WRF configurations tested in this
231 study are reported in Table 3. For a majority of these (e.g. radiation, boundary layer, and surface
232 layer options), the parameterized physics model was chosen according to author experience and

¹Gershunov A., T. M. Shulgina, F. M. Ralph, D. Lavers and J. J. Rutz, 2016: Assessing climate-scale behavior of Atmospheric Rivers affecting western North America. (*In-Preparation*)

233 common practice in other WRF NWP forecast efforts. We stress that the parameterized physics in
234 West-WRF 3 km have not been optimized for the purpose of orographic precipitation prediction.

235 *e. Small Extent WRF-ARW forecasts*

236 To investigate the error induced at the storm scale by restricting the parent domain extent to less
237 than the largest scales of the AR ($X < A$) we have made forecasts with a configuration of WRF-
238 ARW that mimics West-WRF in all aspects except its parent domain size. This configuration,
239 hereafter WRF-ARWS, uses the same parent 9 km horizontal resolution with the same lambert
240 conformal map projection (WRF-ARWS 9 km) and the same vertical discretization. The WRF-
241 ARWS 9 km footprint can be seen in Fig. 2 (box “b”) All parameterized physics and initial /
242 boundary conditions are likewise identical. The nest domain (WRF-ARWS 3 km) is identically
243 configured to West-WRF 3 km.

244 *f. West-WRF spectrally nudged forecasts*

245 For oceanic AR cases (section 2g), we ran a third set of West-WRF forecasts. This set utilized
246 the four-dimensional data assimilation by spectral nudging option (Miguez-Macho et al. 2004).
247 In these forecasts (hereafter West-WRFN 9 km), The West-WRF configuration was reproduced
248 exactly, except that the West-WRF solution in the 9 km domain is relaxed to the GFSRe solution
249 at scales larger than 200 km above model level 17 (approximately the free atmosphere).

250 *g. Forecast Periods*

251 To test West-WRF, GFSRe, and WRF-ARWS skill in predicting both storm scale atmospheric
252 river structure and local precipitation response, we simulated two groups of AR occurring between
253 December 2014 and March 2016. The first group, hereafter called “OCN” included fifteen moder-

ate ($IVT \geq 500 \text{ kg m}^{-1} \text{ s}^{-1}$) AR in the Northeast Pacific Ocean for which GPS-enabled dropwind-
sondes were available during a CalWater transect (see Table 2). The OCN airborne observations
represent a rich survey of the transport-normal vertical and horizontal dynamical structures of
oceanic AR. We use the 15 transects to investigate the ability of West-WRF, WRF-ARWS and
GFSRe to accurately simulate storm scale structures identified by previous authors as important to
AR orographic precipitation (e.g. vertical distribution of vapor transport and moist static stability)
at forecast lead times up to 155 hours.

The second group of forecast periods (hereafter “LND” - see table 4) was chosen to investigate
West-WRF and GFSRe quantitative precipitation forecast (QPF) skill and local scale response to
storm scale forcing during moderate AR impacting the RRW. LND AR cases were chosen using
the criteria:

1. The ARO must have recorded AR Conditions following Ralph et al. (2013) for 24 or more
hours.
2. During AR Conditions at the ARO, IVT must have exceeded $500 \text{ kg m}^{-1} \text{ s}^{-1}$, or BUF must
have exceeded 300 mm m s^{-1} . This is referred to as the “moderate AR condition”.
3. The ten strongest AR by storm-integrated BUF that occurred between December 1, 2014 and
March 31, 2016 and met criteria 1 and 2 were grouped into the LND case list.

Event start and end for LND cases were declared based upon hourly AR conditions at the ARO.
Event duration varies from 24 to 54 hours, with a median length of 32 hr (Table 4). Storm-total
precipitation accumulation ($ST \text{ } QPE$) within the RRW during LND cases is estimated by the
NCEP Stage-IV 4 km gridded product (Lin and Mitchell 2005). Stage IV contains 225 points
within the RRW.

3. Methods

a. The Verification Matrix Procedure

The variation of forecast accuracy (skill) with lead time (t_i) is estimated using the verification matrix procedure. To estimate the skill for a single event at n lead times, one needs n forecasts that verify at the time of event (t_v), each generated at a unique lead time. If this is to be done for m events with l unique observations, one can evaluate the skill at $t_i = 1, \dots, n$ by generating a set of at most $m * n$ forecasts. The most general way to accomplish this is to start n forecasts, each from a unique initial time and gather those for which lead time $t_i = t_v - t_{0i}$. Where t_{0i} represents the forecast initial time. The estimate of forecast skill at each lead time $i = 1, \dots, n$ will then be estimated from a sample of $m * l$ observation - forecast pairs. Our verification matrix is generated by nominal lead times 24 through 168 hours with an increment of 24. Our events are either the accumulation period of AR induced rainfall in the RRW, or the median time of dropwindsonde release in an OCN case AR transect. These times do not fall on 00 UTC. For this reason, we were required to bin our lead times in order to generate a full verification matrix with nominal time resolution of 24 hours. Table 5 lists the nominal lead times and their bin boundaries for the 7 lead time and 3 lead time matrices used in this study. Verification matrices containing only 3 bins were generated by combining the forecasts falling in bins 1-2, 3-4, 5-6 from the 7 lead time matrix.

b. Sonde Data Processing

Each dropwindsonde is processed by vertical smoothing onto isobaric surfaces every 25 hPa from 1000 to 300 hPa. The diagnostic variables calculated include IVT , Moist Brunt Väisälä (N_m^2) frequency following Ralph et al. (2005), equivalent potential temperature (θ_e) as approximated in Stull (2012), and partial IVT ($dIVT$) as in Cordeira et al. (2013). The dropwindsondes report an

298 observation approximately every 1 second, however we assign a static observation time for each
299 sonde that corresponds to the mean time of its transect.

300 Two of the fifteen total transect observing periods required unique processing. Those are transect
301 1 on February 7th, 2014, and transect 4 on February 11th, 2014. Both of these transects are a
302 composite of two spatially offset flight tracks.

303 *c. Transect Compositing Procedure*

304 In order to derive AR-normal composite cross-sections, it was necessary to align the dynamical
305 features within each AR transect. First, each individual dropwindsonde is linearly interpolated
306 to a gridded AR-normal transect with resolution of 50 km. Second, a common center among the
307 transects is defined as the dropwindsonde with maximum *IVT* and the endpoints defined by most
308 poleward and equatorward sondes. Finally, a composite of the transects is created by taking the
309 arithmetic mean of the interpolated transects.

310 *d. NWP Forecast to Observation Interpolation*

311 1) SPATIAL INTERPOLATION TO SONDES AND TEMPORAL INTERPOLATION

312 NWP Forecast output from West-WRF 9 km, West-WRFN 9 km, WRF ARWS and GFSRe have
313 been spatially interpolated from their native grids to the dropwindsonde Earth-relative location
314 using a bilinear method.

315 There are three sources of temporal uncertainty in our methods. First, no attempt has been made
316 to temporally interpolate NWP output to sonde report time. Because output from each model is
317 3 hourly, the maximum temporal difference from observations is 90 minutes. In order to create
318 composite dropwindsonde cross-sections, we must assume stationarity in the AR up to the longest
319 time between transect start and end. This time is 2 hr, 27 minutes. Lastly, the GFSRe duty cycle

320 is 24hr, but the valid time of any given observation may occur anywhere in the diurnal cycle. The
321 sources of temporal uncertainty above lead primarily to random temporal imprecision in model to
322 observation matching, though we cannot rule out that these sources of temporal imprecision will
323 accumulate to non-zero residual. Because their magnitude is likely to be much larger than the time
324 a dropwindsonde takes to profile the atmosphere below flight level, no effort is made to account
325 for the sonde drift in the forecast interpolation. The location used is the mean of the individual
326 sonde latitude and longitude reports. We use the Brier Skill Score (*BSS*) following Winterfeldt
327 et al. (2011) computed from West-WRFN 9 km in part to estimate the impact of this temporal
328 imprecision (see section 4a).

329 2) ATMOSPHERIC RIVER OBSERVATORY

330 NWP forecast output from all models were bilinearly interpolated to both ARO locations. Model
331 output was also linearly interpolated from the model native coordinate to the effective retrieval
332 heights (m AGL) of the 449 MHz wind profiling radar.

333 3) NCEP STAGE IV QPE

334 Model output was bilinearly interpolated from native grids to the locations of the NCEP Stage
335 IV *QPE* product. Model output was masked thereafter to exclude any points lying outside the
336 boundaries of the RRW. Watershed boundaries were defined by a geo-referenced shapefile created
337 by the US Geological Survey. Masking by the shapefile polygon was performed using NCAR
338 Command Language version 6.2 (<http://www.ncl.ucar.edu/>).

339 e. *Estimates of Deterministic Skill*

340 For atmospheric state variables at storm-scale, we define the value added of each WRF 9 km
341 forecast as the *BSS* using GFSRe as the reference forecast. For an atmospheric state variable

that is approximately Gaussian, such as $z500$, IVT , IWV , and θ_e^{925} . The BSS is equivalent to measuring the fractional reduction in mean square error by the more accurate forecast (test forecast or reference).

Qualitative and Quantitative methods are used to assess accuracy in QPF for the models considered. We compute histograms of storm-total precipitation from the NCEP Stage-IV data and from each model forecast at varying lead times according to the verification matrix procedure. For quantitative measures, we calculate the mean, standard deviation and root mean square of the quantity $QPF - QPE$ at all NCEP Stage-IV gridpoints and for all models and lead times as appropriate.

We also estimate the normalized error of West-WRF 9 km and GFSRe in reproducing the observed storm-total (ST) BUF and Precipitation (P_r) relationship at the Coastal ARO (hypothesis 2). To assess which model reproduces the area occupied in the forcing-response phase space most accurately, we define the multifactor orographic forcing-response error:

$$e_{xy} = E \left[\frac{(x - x_o)^2}{V[x_o]} + \frac{(y - y_o)^2}{V[y_o]} \right], \quad (1)$$

where x and y are the forecast storm scale forcing ($ST\ BUF$) and local response ($ST\ P_r$), respectively. The subscript o refers to the observed values of the given quantity. $E[]$ and $V[]$ refer to the expectation and variance operators, respectively.

f. Linearization of the Forcing and Response Relationships, Reduction in Error

We also assess whether reducing error in storm scale forcing or reducing error in the simulated local response, independent of forcing, lead to larger reduction of error in the forecast $ST\ P_r$. To investigate this, we note that Ralph et al. (2013) found that the relationship between $ST\ BUF$ at BBY and $ST\ Pr$ at CZC can be linearly approximated with a correlation of 0.74.

Let $e_y = \frac{E[(y-y_o)^2]}{V[y_o]}$ represent the normalized mean square error in the forecast local response compared to that measured by the ARO. We can also define

$$e_y \sim \frac{E[F(x) - y_o^2]}{V[y_o]} \quad (2)$$

where $F(x) \sim f(x)$ is the linearized approximation of the local response function described by Ralph et al. (2013). We can estimate $F(x)$ from each model and from observations by least-squares approximation. Let the quantity $F_o(x)$ represent the “perfect response” approximation to y . It is the $ST P_r$ that results from applying the linearized local response function derived from a least-squared fit of the observations to the forecast storm-scale forcing. Conversely, $F(x_o)$ represents the “perfect forcing” approximation to y . The fractional reduction in error in $ST P_r$ by perfect response approximation is then

$$\delta e_{ypr} = 1 - \frac{E[F_o(x) - y_o^2]}{V[y_o]e_y} \quad (3)$$

The fractional reduction in error in storm-total precipitation by perfect forcing is

$$\delta e_{yprf} = 1 - \frac{E[F(x_o) - y_o^2]}{V[y_o]e_y}. \quad (4)$$

F_o is estimated from the sample of historical observations at the ARO that reside within the phase space of $ST BUF$ vs. $ST P_r$ defined by the LND cases. This corresponds to a sample size of 52. $F(x)$ is estimated for each West-WRF and GFSRe by each LND case forecast binned by the three lead time verification matrix. This procedure yielded a sample size of 20 for each West-WRF and GFSRe.

4. Results

a. Deterministic Value Added in State Variables

The value added by West-WRF 9 km, West-WRFN 9km, and WRF-ARWS 9km to GFSRe for the first six lead times in the verification matrix, is displayed in Fig. 3b-d. GFSRe value added to GFSRe 25 year (Dec. - Mar. 1991 - 2015) climatology is displayed for the same lead times in Fig. 3a. Value added is estimated for z^{500} , θ_e^{925} , IWV , and IVT using BSS as described in section 3. Only sondes for which $IVT \geq 250 \text{ kg m}^{-1} \text{ s}^{-1}$ were retained for the analysis, yielding a sample of 145 dropwindsondes. Upper (lower) whiskers represent maximum (minimum) BSS , upper (lower) box bounds represent upper (lower) quartile BSS and box center line represents median BSS .

For the majority of $t_i < 108 \text{ hr}$, GFSRe forecasts of these state variables are very skillful, reducing mean square error by 50% to 95% compared to climatology. Forecasts of θ_e^{925} have the highest skill distribution and remain very high even for very long lead times, suggesting both that θ_e^{925} is very different than climatology during AR but also that GFSRe captures the departure well at the scales observable by the soundings. IWV and IVT display the least skill and most skill degradation at longer lead times, though most forecasts of IWV retain a greater than 50% error reduction. Figure 3b shows the value added by West-WRFN 9km. It is included for an estimate of the range of errors relative to the GFSRe forecast introduced by spatial and temporal representativeness and dropwindsonde spatial drift.

From Fig. 3c it is apparent that West-WRF 9 km adds significant value to GFSRe at the storm scale between $36 \text{ hr} \leq t_i \leq 131 \text{ hr}$ in oceanic AR conditions. Note that value is lost by West-WRF 9 km for early ($12 \text{ hr} \leq t_i \leq 35 \text{ hr}$) and late ($t_i > 132 \text{ hr}$) lead times for a majority of sondes and variables.

WRF-ARWS (Fig. 3d) suffers from lost value relative to GFSRe for all variables and lead times with the exception of IVT and θ_e^{925} for $60 \text{ hr} \leq t_i \leq 131 \text{ hr}$ and IWV for $84 \text{ hr} \leq t_i \leq 107 \text{ hr}$. WRF-ARWS loses value compared to GFSRe in forecasts of z^{500} with more than 75% certainty (interquartile range of BSS lies entirely below zero) for all lead times examined. In some of the early lead times, the range of value lost in forecasts of z^{500} is dramatically low, suggesting that WRF-ARWS leads to a near 90% increase in mean square error over GFSRe. The fact that these large errors in mid-tropospheric geopotential manifest early and reduce with lead time argue that the role of non-native analysis and boundary condition shock (Rodwell and Palmer 2007) is particularly dramatic in the smaller domain of WRF-ARWS.

Figures 4a and 4b show the value added by West-WRF 9 km and WRF-ARWS 9 km to West-WRFN 9 km, respectively. These may be interpreted as the relative improvement in forecast accuracy that arise from the feedback from the higher resolution dynamics to the larger scales. As in Fig. 3d, WRF-ARWS 9 km suffers from significant value lost in most fields and at most lead times, with early lead times penalized more than late lead times. West-WRF 9 km, on the other hand, displays value added over West-WRFN 9 km for a smaller range of lead times than found for value added over GFSRe. The value added is primarily for $t_i \leq 83 \text{ hr}$ and appears to be most significant in IWV and z^{500} . Interestingly, West-WRF 9 km does not ever develop consistent and significant (lower quartile $BSS > 0$) value added over West-WRFN 9 km at any point in the verification matrix when forecasting IVT . We interpret this to mean that forecasts of IVT interrogated at the scales of the dropwindsondes do not benefit from the high-resolution dynamics in West-WRF 9 km.

b. Vertical Structures of AR Static Stability

Ralph et al. (2005) demonstrated that AR core environments are unique in that the low-level jet region, which exists approximately below 3 km MSL and contains the maximum horizontal flux of water vapor, is also nearly saturated and approximately moist-neutral. Thus, relatively small vertical displacements ($\sim 10^2$ m) in the lowest 3 km of the AR core encounter minimal buoyant restoring force and nearly immediately produce condensate. Many authors (Neiman et al. 2002; Smith et al. 2010; Valenzuela and Kingsmill 2015) have argued that because it is coincident with prodigious flux of water vapor, near-saturation, and moist neutral buoyancy that the LLJ is responsible for the very efficient precipitation by AR even when encountering relatively short mountain ranges (e.g. those surrounding the RRW).

We investigated the accuracy of GFSRe, West-WRF 9 km and WRF-ARWS to reproduce observed profiles of moist static stability from the sample of dropwindsondes for which $IVT \geq 250 \text{ kg m}^{-1} \text{ s}^{-1}$. This analysis is shown in Fig. 5a-c in the form of vertically binned interquartile range of the square of N_m^2 (s^{-2}) at significant pressure levels for the retained soundings. As found in Ralph et al. (2005), N_m^2 varies about 1 or 2 s^{-2} near the Ocean's surface and increases very slightly toward 700 hPa. The measurements we report here appear to be slightly higher than those found by Ralph et al. (2005), which we attribute to different sonde selection criteria and potentially different storm-relative environments. The model forecasts are in general biased toward slightly more stable than observations at very low levels ($p > 950$ hPa) at all lead times. With the exception of an outlier near 950 hPa in the WRF-ARWS forecasts, all models additionally seem to capture the observed median stability very well at $t_i \leq 59$ hr. As lead times increase, model variability increases to more than that of observations, and all models develop a stable bias between 800 hPa $< p < 900$ hPa at $t_i \geq 108$ hr. This may be indicative of the models relaxing from an AR sounding

443 to one more typical of the mean NE Pacific wintertime boundary layer top. We will discuss in
444 section 4c the tendency for all models considered to exclude the AR from the mean forecast at
445 these lead times.

446 *c. Vertical Structures of AR Core Horizontal Vapor Transport*

447 Figure 6a displays the composite AR cross-sections of $dIVT$ from the OCN case AR core tran-
448 sects. The methodology for constructing each cross-section is discussed in section 3c. The center
449 panel in Fig. 6 displays the observed composite. In this panel, there is a strong local maximum
450 in water vapor flux near 900 hPa located at the analyzed AR core center. This water vapor flux
451 maximum is located just below a weak composite low-level jet (isotach composite, not shown).
452 Equivalent potential temperature isotherms in Fig. 6 show the composite AR core straddling a
453 baroclinic zone with temperatures decreasing poleward of the core center. This composite struc-
454 ture is consistent with AR cross-sections reported in Cordeira et al. (2013) and Ralph et al. (2016).

455 The remaining panels in Fig. 6 display the mean $dIVT$ error (forecast minus observed) com-
456 posite in West-WRF and GFSRe for the 3 lead time verification matrix. Overlaid in each is the
457 composite forecast θ_e . It is apparent that these errors in water vapor flux are spatially heteroge-
458 neous and that errors are more likely to be positive above 700 hPa and more likely to be negative
459 below 700 hPa and near the mean LLJ position for both models. Composites were made for each
460 individual variable contributing to water vapor flux (wind speed and water vapor mixing ratio -
461 not shown). It was found that the errors in each are spatially correlated with each other and with
462 $dIVT$. The upper levels of each model composite are too moist and wind speeds are too fast, while
463 the low-level jet region in each model composite is too dry and wind speeds are too slow. Sev-
464 eral authors (Thorpe and Clough 1991; Dudhia 1993; Lafore et al. 1994; Wakimoto and Murphey
465 2008) have observed a sub-geostrophic polar jet in conjunction with a super-geostrophic low-level

jet in mid-latitude cyclones containing strong cold fronts. The results reported here may therefore be consistent with the model simulations inaccurately maintaining geostrophy. The upper troposphere positive bias in $dIVT$ is significant at the $p < 0.05$ at all lead times. Lower tropospheric negative biases do not become significant at this level until the later lead times. Additionally, negative biases in the mean low-level jet position, a feature found to be critically important to driving heavy orographic precipitation (Browning et al. 1974; Bader and Roach 1977; Neiman et al. 2002; Smith et al. 2010), are quite similar in spatial extent and magnitude for both models.

Figure 7 displays the mean $dIVT$ error (forecast minus observed) composite in WRF-ARWS and GFSRe for the 3 lead time verification matrix. In contrast to the West-WRF composite cross-sections in $dIVT$, low biases in $dIVT$ grow large enough to reach the level of significance at the $p < 0.05$ level for $t_i > 60$ hr. The mean $dIVT$ error at these lead times reaches $-20 \text{ kg m}^{-2} \text{ s}^{-1}$ (leftmost panel, middle row of Fig. 7), compared to $-12 \text{ kg m}^{-2} \text{ s}^{-1}$ for both GFSRe and West-WRF. Mean errors at the longest lead times for WRF-ARWS reach $-26 \text{ kg m}^{-2} \text{ s}^{-1}$. This value is both significant according to t-test and is much greater than the corresponding mean error in West-WRF and GFSRe. Lastly, WRF-ARWS appears to lose the cross-AR baroclinicity and moist AR core in the composite transect θ_e that is maintained in GFSRe and West-WRF.

d. Errors in AR Landfall Timing

The forecast errors considered thus far include the convolution of errors in magnitude and in location or timing (phase error). Figure 8 displays error (forecast - observed) in the timing of AR condition start at the ARO during LND cases for West-WRF 3 km and GFSRe. The most notable features in Fig. 8 are the wider range of timing error calculated from the 10 LND case West-WRF 3 km AR condition onsets and the slight delay bias in the West-WRF 3 km forecasts. GFSRe in general exhibits AR onset closer to that observed with a smaller range of error. Onset bias remains

near 3 hr +/- 6 hr (median +/- upper/lower quartiles) after observed in GFSRe for $t_i \leq 59$ hr, while for the same lead times, the West-WRF 3 km forecasts contained a 9 hr +/- 15 hr bias in AR condition onset. For $60 \text{ hr} \leq t_i \leq 107 \text{ hr}$, median onset error is similar between the two models considered, but while GFSRe forecasts are closer to being equally distributed around early or late onset (with slight bias toward late), the West-WRF 3 km forecasts develop more than a 75% chance of forecasting late AR arrival. For later lead times, there is little difference to report between the two sets of forecasts. Proposing a mechanism for consistently late arrival of AR conditions is beyond the scope of this work, we present this analysis here merely for additional insight into the factors contributing to errors in forecasts of the storm scale forcing during LND cases. We will examine errors in ST Pr in the entire watershed and at the ARO next.

e. Deterministic QPF Skill During Landfalling AR

Value histograms for West-WRF 3 km and GFSRe *ST QPF* at NCEP Stage-IV grid points within the Russian River Watershed (RRW) for the 3 bin verification matrix are shown in Fig. 9. Figure 9 also displays the LND case *ST QPE* histogram. *ST QPE* from the cases investigated had median and upper (lower) quartile value of 61 and 97 (41) mm. For short lead times (Fig. 9a), the West-WRF 3 km *QPF* histogram most closely resembles the *QPE* histogram. The inset tables in Fig. 9 display mean *QPF - QPE* (bias, labelled “*AVG*”), the standard deviation of *QPF - QPE* (“*STD*”) and the root mean-square (“*RMS*”) *QPF - QPE* for both West-WRF 3 km and GFSRe. Both West-WRF 3 km and GFSRe produce low-biased *ST QPF* across lead times. For $t_i \leq 59$ hr, West-WRF 3 km accumulation bias is 5% greater in magnitude relative to median *QPE* than for GFSRe. For other lead times, bias becomes similar.

For $t_i \leq 108$ hr, West-WRF 3 km produces a smaller range in accumulation error (*STD* is 10% smaller relative to median *QPE*). For $t_i \leq 108$ hr, West-WRF 3 km *RMS* is 5-7% less than GFSRe

as a fraction of median QPE , while for $t_i > 108$ hr, GFSRe displays an advantage that is similar in magnitude. West-WRF 3 km produces total precipitation greater than 140 mm (near the upper 10% value for QPE) with non-zero frequency. GFSRe, however does not produce these upper 10% values. This last advantage West-WRF displays is likely related to higher output spatial resolution.

Analysis from Fig. 9 does not definitively answer whether West-WRF 3 km or GFSRe QPF is more accurate for RRW landfalling AR. Neither West-WRF 3 km nor GFSRe distinguished itself in a large or consistent manner over the lead times considered. Thus far, we have seen that each West-WRF and GFSRe systems display minor strengths relative to each other in AR state, structure RRW composite QPF , but on balance perform with similar accuracy. However, we cannot simply assume that the QPF performance result follows from the storm scale accuracy result. To adequately investigate the second hypothesis posed in the introduction, we must investigate each model's ability to accurately reproduce observed local precipitation response during AR.

f. Relationships Between Storm Scale Forcing and Local Scale Response

To do so, we turn to the 10 - year record of bulk upslope vapor flux and mountaintop precipitation collected at the ARO. Figure 10a shows the ARO historical (Jan 2006 - Mar 2016) $ST\ BUF - ST\ P_r$ relationship. For context, the least-square fit of $ST\ P_r$ to $ST\ BUF$ and the correlation coefficient and e_y (ratio of mean-square error to variance) that result from this fit are also displayed. Figure 10b shows the observed, GFSRe simulated and West-WRF 3 km simulated $ST\ BUF - ST\ P_r$ relationship for $60\ hr \leq t_i \leq 107\ hr$. e_{xy} (see section 3 for details) for West-WRF 3 km and GFSRe are displayed in the Fig 10b inset table. e_{xy} is also listed for the 3 lead-time verification matrix in Table 6. Note that West-WRF reduces e_{xy} by 37% to 80% compared to GFSRe. The reader can qualitatively assess the accuracy of each model in representing the forcing-response phase space by visually estimating the overlap between the GFSRe and West-WRF symbols and the

space occupied by the observational symbols in Fig. 10b. Fig. 10c-d show only cases for which West-WRF produces lower e_{xy} and those for which GFSRe performs better or there is a near-tie, respectively. Table 6 additionally presents e_y , the error in storm-total precipitation for both models normalized to the observed variance. West-WRF 3 km also outperforms GFSRe by 18% to 69% in this metric.

From Fig. 10a and Table 6 we conclude that West-WRF better predicts the combined storm-scale forcing and watershed response for the ARO. We are also interested in asking the questions of both models: “Do errors in $ST P_r$ during AR arise primarily due to errors in storm scale forcing or in the local response?” and “which model may benefit from small adjustments to accuracy in either storm scale forcing or the response to forcing?” To investigate this, we present δe_{ypf} and δe_{ypr} (see section 3e) in Table 6.

The design of the δe_y metrics leads to an issue interpreting positive values. The interpretation might be that substituting observed forcing (linearized response) into $e_{ypf(r)}$ increases error compared to the NWP modeled e_y because the error in the NWP model in predicting the storm-total precipitation is less than the error of the linear fit to the observations. The interpretation might also be that the model response function is either different enough than the observed or so non-linear that more accurate forcing produces less accurate response. To interpret the δe_y metrics, we must also compare the model’s linearized response relationship (slope - also reported in Table 6) to the observed.

For negative values, The ratio $\delta e_{ypf(r)}$ is an estimate of whether the errors in predicting storm-total precipitation arise primarily from the storm scale forcing or from the local response. West-WRF $\delta e_{ypf(r)}$ both become negative for $t_i \geq 60$ hr. At these lead times, both $ST BUF$ and $ST P_r$ are biased low compared to observations (e.g. West-WRF markers are low and to the left of observations in Fig. 10b), but the slope of the response relationship is very near that of the observed. This

suggests that if West-WRF forecast $ST\ BUF$ improves, more accurate West-WRF $ST\ P_r$ would result, as reflected by δe_{ypf} for West-WRF at $t_i \geq 60$ hr in Table 6. West-WRF δe_{ypr} suggests that $ST\ P_r$ could also improve by further tuning the precipitation response relationship, though not by as much. For West-WRF at $t_i \leq 60$ hr, imposing observational $ST\ BUF$ or response relationship does not improve forecast $ST\ P_r$. At these lead times e_y is less than 1.0, suggesting that West-WRF forecast error is smaller than the observed variance in precipitation and substituting the observed forcing (response) is not effective.

In contrast, GFSRe δe_{ypf} is non-negative and GFSRe $e_y > 1.0$ for all t_i . This suggests that the GFSRe response relationship is not accurate enough to translate improved $ST\ BUF$ into improved $ST\ P_r$. Examining Table 6 confirms that the slope of the derived linear relationship for GFSRe is not similar to the observed at any t_i . GFSRe δe_{ypr} suggests that accuracy in $ST\ P_r$ could be improved if the local response relationship is made more accurate, though not to the same degree as for West-WRF.

5. Discussion

This study is the first to investigate GNWP and RNWP skill in predicting tropospheric state during AR, AR core structure, and AR precipitation at varying spatial scales and for a comprehensive set of forecast lead times up to seven days. It was found that West-WRF is capable of adding value to its parent GNWP by means of dynamical downscaling even at storm scales for a subset of medium-range weather prediction timescales. This result suggests that assimilation of observations into the course domain at model native scales may additionally improve storm scale forecasts created by West-WRF 9 km.

It was also found that this value added is contingent upon careful construction of the most course domain so that error in boundary condition interpolation and non-native shock do not significantly

582 impact the solution at the storm scale. This was verified by examining value added in tropospheric
583 state variables during AR forecast by a WRF configuration whose boundaries do not adequately
584 contain the largest storm scales (WRF-ARWS).

585 Generally, West-WRF 9 km and GFSRe Forecasts of AR vertical and transport-normal structures
586 were found to reproduce realistic vapor transport in the low-level jet region of AR at short lead
587 times, but forecasts of LLJ water vapor flux were found to develop significant low bias by $t_i > 108$
588 hr lead times. Forecasts of moist static stability in the AR generally performed very well compared
589 to observation, especially at short lead times for all models. The GNWP considered (the Global
590 Forecast System) is capable of producing AR realistic in structure at sufficiently short lead time.
591 If well-constructed, WRF forecasts downscaled from the same GNWP can as well. These findings
592 support the first hypothesis posed in the introduction.

593 West-WRF 3 km and GFSRe displayed similar accuracy in predicting RRW storm-total precip-
594 itation during the AR considered. As lead time increased, both models produced significant dry
595 bias compared to the observed accumulated precipitation distribution. This finding may follow
596 from the inability of both models to produce a strong AR LLJ at longer lead times and from AR
597 phase errors detected at longer lead times in both models (Fig. 8). The authors note that tuning of
598 the parameterized physics sub-models in West-WRF for the purpose of accurately predicting AR
599 precipitation has not been done, and that *QPF* smoothed to lower resolutions (e.g. the difference
600 in resolution between West-WRF 3 km and GFSRe) has been shown to result in higher skill scores
601 over complex terrain (Mass et al. 2002).

602 When predicting storm-total precipitation at a mountaintop well known to be orographically
603 productive (CZC), West-WRF improves upon GFSRe mean square error by as much as 69% at
604 short lead times. It appears that this improvement occurs primarily because West-WRF better
605 reproduces the relationship between storm-scale forcing (approximated by *ST BUF*) and local

precipitation response (storm-total rainfall at CZC). This can be seen visually in Figure 10b-d and quantitatively in Table 6. West-WRF 3 km very accurately reproduces the phase space in the $ST\ BUF - ST\ P_r$ relationship found through observations during the AR cases studied. This finding supports the second hypothesis posed in the introduction.

It is found that improvement in West-WRF QPF can be expected through either more accurate storm-scale forcing (e.g. data assimilation) or local scale response (e.g. parameterized physics tuning). West-WRF Storm-scale forcing at the ARO was often low-biased, in agreement with the low-bias in low-level jet $dIVT$ (Fig. 6). Thus, the consistent underprediction of RRW $ST\ P_r$ (Fig. 9) is partially caused by storm-scale forcing that is too weak. This cause-and-effect relationship cannot be verified for GFSRe, since $ST\ BUF$ errors from GFSRe are more randomly distributed (Fig. 10b) and since the local response relationship is not similar to that observed.

The forcing-response analysis suggests that given correct forcing and improved local response (including more accurately selected parameterized physics) RNWP systems such as West-WRF can produce more accurate precipitation, even at lead times approaching 6 days. Therefore; West-WRF may be an attractive option to produce skillful QPF for regions in which heavy rain events are dominated by atmospheric rivers.

Acknowledgments. The California Department of Water Resources, the US Army Corps of Engineers Forecast Informed Reservoir Operations for Lake Mendocino Project and the National Science Foundation XSEDE award no. ATM150010 provided financial support for this research. The authors would like to thank the NOAA CalWater airborne science teams and NOAA Physical Science Division's Hydrometeorological Testbed for providing data. Additionally, the authors would like to thank Scripps Institution of Oceanography for establishing the Center for Western Weather and Water Extremes.

References

- Bader, M., and W. Roach, 1977: Orographic rainfall in warm sectors of depressions. *Quarterly Journal of the Royal Meteorological Society*, **103** (436), 269–280.
- Bao, J., S. Michelson, P. Neiman, F. Ralph, and J. Wilczak, 2006: Interpretation of enhanced integrated water vapor bands associated with extratropical cyclones: Their formation and connection to tropical moisture. *Monthly weather review*, **134** (4), 1063–1080.
- Browning, K., F. Hill, and C. Pardoe, 1974: Structure and mechanism of precipitation and the effect of orography in a wintertime warm sector. *Quarterly Journal of the Royal Meteorological Society*, **100** (425), 309–330.
- Cavallo, S. M., J. Berner, and C. Snyder, 2016: Diagnosing model errors from time-averaged tendencies in the weather research and forecasting (wrf) model. *Monthly Weather Review*, **144** (2), 759–779.
- Cordeira, J. M., F. M. Ralph, and B. J. Moore, 2013: The development and evolution of two atmospheric rivers in proximity to western north pacific tropical cyclones in october 2010. *Monthly Weather Review*, **141** (12), 4234–4255.
- Dettinger, M. D., 2013: Atmospheric rivers as drought busters on the us west coast. *Journal of Hydrometeorology*, **14** (6), 1721–1732.
- Dettinger, M. D., F. M. Ralph, T. Das, P. J. Neiman, and D. R. Cayan, 2011: Atmospheric rivers, floods and the water resources of california. *Water*, **3** (2), 445–478.
- Dole, R. M., 2008: Linking weather and climate. *Synoptic—Dynamic Meteorology and Weather Analysis and Forecasting*, Springer, 297–348.

650 Dudhia, J., 1993: A nonhydrostatic version of the penn state-ncar mesoscale model: Validation
651 tests and simulation of an atlantic cyclone and cold front. *Monthly Weather Review*, **121** (5),
652 1493–1513.

653 Guan, B., N. P. Molotch, D. E. Waliser, E. J. Fetzer, and P. J. Neiman, 2010: Extreme snow-
654 fall events linked to atmospheric rivers and surface air temperature via satellite measurements.
655 *Geophysical Research Letters*, **37** (20).

656 Guan, B., N. P. Molotch, D. E. Waliser, E. J. Fetzer, and P. J. Neiman, 2013: The 2010/2011
657 snow season in california’s sierra nevada: Role of atmospheric rivers and modes of large-scale
658 variability. *Water Resources Research*, **49** (10), 6731–6743.

659 Guan, B., D. E. Waliser, N. P. Molotch, E. J. Fetzer, and P. J. Neiman, 2012: Does the madden-
660 julian oscillation influence wintertime atmospheric rivers and snowpack in the sierra nevada?
661 *Monthly Weather Review*, **140** (2), 325–342.

662 Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau Jr, Y. Zhu, and
663 W. Lapenta, 2013: Noaa’s second-generation global medium-range ensemble reforecast dataset.
664 *Bulletin of the American Meteorological Society*, **94** (10), 1553–1565.

665 Junker, N. W., M. J. Brennan, F. Pereira, M. J. Bodner, and R. H. Grumm, 2009: Assessing the
666 potential for rare precipitation events with standardized anomalies and ensemble guidance at the
667 hydrometeorological prediction center. *Bulletin of the American Meteorological Society*, **90** (4),
668 445.

669 Kalnay, E., and Coauthors, 1996: The ncep/ncar 40-year reanalysis project. *Bulletin of the Ameri-
670 can meteorological Society*, **77** (3), 437–471.

671 Lafore, J., J. Redelsperger, C. Cailly, and E. Arbogast, 1994: Nonhydrostatic simulation of fronto-
672 genesis in a moist atmosphere. part iii: Thermal wind imbalance and rainbands. *Journal of the*
673 *atmospheric sciences*, **51 (23)**, 3467–3485.

674 Lavers, D. A., R. P. Allan, E. F. Wood, G. Villarini, D. J. Brayshaw, and A. J. Wade, 2011: Winter
675 floods in britain are connected to atmospheric rivers. *Geophysical Research Letters*, **38 (23)**.

676 Lavers, D. A., D. E. Waliser, F. M. Ralph, and M. D. Dettinger, 2016: Predictability of horizontal
677 water vapor transport relative to precipitation: Enhancing situational awareness for forecasting
678 western us extreme precipitation and flooding. *Geophysical Research Letters*, **43 (5)**, 2275–
679 2282.

680 Lin, Y., and K. E. Mitchell, 2005: 1.2 the ncep stage ii/iv hourly precipitation analyses: Devel-
681 opment and applications. *19th Conf. Hydrology, American Meteorological Society, San Diego,*
682 *CA, USA, Citeseer.*

683 Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolu-
684 tion produce more skillful forecasts? the results of two years of real-time numerical weather
685 prediction over the pacific northwest. *Bulletin of the American Meteorological Society*, **83 (3)**,
686 407.

687 McCabe, G. J., and M. D. Dettinger, 2002: Primary modes and predictability of year-to-year
688 snowpack variations in the western united states from teleconnections with pacific ocean cli-
689 mate. *Journal of Hydrometeorology*, **3 (1)**, 13–25.

690 Miguez-Macho, G., G. L. Stenchikov, and A. Robock, 2004: Spectral nudging to eliminate the
691 effects of domain position and geometry in regional climate model simulations. *Journal of Geo-*
692 *physical Research: Atmospheres*, **109 (D13)**.

693 Minder, J. R., and D. E. Kingsmill, 2013: Mesoscale variations of the atmospheric snow line over
 694 the northern sierra nevada: Multiyear statistics, case study, and mechanisms. *Journal of the*
 695 *Atmospheric Sciences*, **70** (3), 916–938.

696 Moore, B. J., P. J. Neiman, F. M. Ralph, and F. E. Barthold, 2012: Physical processes associated
 697 with heavy flooding rainfall in nashville, tennessee, and vicinity during 1-2 may 2010: The role
 698 of an atmospheric river and mesoscale convective systems*. *Monthly Weather Review*, **140** (2),
 699 358–378.

700 Neiman, P. J., D. J. Gottas, A. B. White, L. J. Schick, and F. M. Ralph, 2014: The use of snow-level
 701 observations derived from vertically profiling radars to assess hydrometeorological character-
 702 istics and forecasts over washington’s green river basin. *Journal of Hydrometeorology*, **15** (6),
 703 2522–2541.

704 Neiman, P. J., F. M. Ralph, B. J. Moore, M. Hughes, K. M. Mahoney, J. M. Cordeira, and M. D.
 705 Dettinger, 2013: The landfall and inland penetration of a flood-producing atmospheric river in
 706 arizona. part i: Observed synoptic-scale, orographic, and hydrometeorological characteristics.
 707 *Journal of Hydrometeorology*, **14** (2), 460–484.

708 Neiman, P. J., F. M. Ralph, A. White, D. Kingsmill, and P. Persson, 2002: The statistical relation-
 709 ship between upslope flow and rainfall in california’s coastal mountains: Observations during
 710 caljet. *Monthly Weather Review*, **130** (6), 1468–1492.

711 Neiman, P. J., F. M. Ralph, G. A. Wick, J. D. Lundquist, and M. D. Dettinger, 2008: Meteorolog-
 712 ical characteristics and overland precipitation impacts of atmospheric rivers affecting the west
 713 coast of north america based on eight years of ssm/i satellite observations. *Journal of Hydrom-*
 714 *eteorology*, **9** (1), 22–47.

715 Neiman, P. J., L. J. Schick, F. M. Ralph, M. Hughes, and G. A. Wick, 2011: Flooding in western
 716 washington: the connection to atmospheric rivers*. *Journal of Hydrometeorology*, **12** (6), 1337–
 717 1358.

718 Neiman, P. J., A. B. White, F. M. Ralph, D. J. Gottas, S. I. Gutman, and Coauthors, 2009: A water
 719 vapour flux tool for precipitation forecasting. *Water Management*, **162** (2), 83.

720 Newman, M., G. N. Kiladis, K. M. Weickmann, F. M. Ralph, and P. D. Sardeshmukh, 2012:
 721 Relative contributions of synoptic and low-frequency eddies to time-mean atmospheric moisture
 722 transport, including the role of atmospheric rivers. *Journal of Climate*, **25** (21), 7341–7361.

723 Ralph, F., T. Coleman, P. Neiman, R. Zamora, and M. Dettinger, 2013: Observed impacts of
 724 duration and seasonality of atmospheric-river landfalls on soil moisture and runoff in coastal
 725 northern california. *Journal of Hydrometeorology*, **14** (2), 443–459.

726 Ralph, F., and M. Dettinger, 2012: Historical and national perspectives on extreme west coast
 727 precipitation associated with atmospheric rivers during december 2010. *Bulletin of the American*
 728 *Meteorological Society*, **93** (6), 783–790.

729 Ralph, F., E. Sukovich, D. Reynolds, M. Dettinger, S. Weagle, W. Clark, and P. Neiman, 2010a:
 730 Assessment of extreme quantitative precipitation forecasts and development of regional extreme
 731 event thresholds using data from hmt-2006 and coop observers. *Journal of Hydrometeorology*,
 732 **11** (6), 1286–1304.

733 Ralph, F. M., P. J. Neiman, G. N. Kiladis, K. Weickmann, and D. W. Reynolds, 2010b: A multi-
 734 scale observational case study of a pacific atmospheric river exhibiting tropical-extratropical
 735 connections and a mesoscale frontal wave. *Monthly Weather Review*.

736 Ralph, F. M., P. J. Neiman, and R. Rotunno, 2005: Dropsonde observations in low-level jets
737 over the northeastern pacific ocean from caljet-1998 and pacjet-2001: Mean vertical-profile and
738 atmospheric-river characteristics. *Monthly weather review*, **133** (4), 889–910.

739 Ralph, F. M., P. J. Neiman, and G. A. Wick, 2004: Satellite and caljet aircraft observations of
740 atmospheric rivers over the eastern north pacific ocean during the winter of 1997/98. *Monthly*
741 *Weather Review*, **132** (7), 1721–1745.

742 Ralph, F. M., P. J. Neiman, G. A. Wick, S. I. Gutman, M. D. Dettinger, D. R. Cayan, and A. B.
743 White, 2006: Flooding on california’s russian river: Role of atmospheric rivers. *Geophysical*
744 *Research Letters*, **33** (13).

745 Ralph, F. M., and Coauthors, 2016: Calwater field studies designed to quantify the roles of
746 atmospheric rivers and aerosols in modulating u.s. west coast precipitation in a changing
747 climate. *Bulletin of the American Meteorological Society*, **97** (7), 1209–1228, doi:10.1175/
748 BAMS-D-14-00043.1, URL <http://dx.doi.org/10.1175/BAMS-D-14-00043.1>, [http://dx.doi.org/](http://dx.doi.org/10.1175/BAMS-D-14-00043.1)
749 [10.1175/BAMS-D-14-00043.1](http://dx.doi.org/10.1175/BAMS-D-14-00043.1).

750 Rodwell, M., and T. Palmer, 2007: Using numerical weather prediction to assess climate models.
751 *Quarterly Journal of the Royal Meteorological Society*, **133** (622), 129–146.

752 Rutz, J. J., W. J. Steenburgh, and F. M. Ralph, 2014: Climatological characteristics of atmospheric
753 rivers and their inland penetration over the western united states. *Monthly Weather Review*,
754 **142** (2), 905–921.

755 Skamarock, W. e. a., 2008: A description of the advanced research wrf version 3. near tech. note.
756 Tech. rep., NCAR/TN-475+ STR.

757 Smith, B. L., S. E. Yuter, P. J. Neiman, and D. Kingsmill, 2010: Water vapor fluxes and orographic
758 precipitation over northern california associated with a landfalling atmospheric river. *Monthly*
759 *Weather Review*, **138** (1), 74–100.

760 Stull, R. B., 2012: *An introduction to boundary layer meteorology*, Vol. 13. Springer Science &
761 Business Media.

762 Sukovich, E. M., F. M. Ralph, F. E. Barthold, D. W. Reynolds, and D. R. Novak, 2014: Extreme
763 quantitative precipitation forecast performance at the weather prediction center from 2001 to
764 2011. *Weather and Forecasting*, **29** (4), 894–911.

765 Thorpe, A. J., and S. A. Clough, 1991: Mesoscale dynamics of cold fronts: Structures described by
766 dropsoundings in fronts 87. *Quarterly Journal of the Royal Meteorological Society*, **117** (501),
767 903–941.

768 Valenzuela, R. A., and D. E. Kingsmill, 2015: Orographic precipitation forcing along the coast of
769 northern california during a landfalling winter storm. *Monthly Weather Review*, **143** (9), 3570–
770 3590.

771 Wakimoto, R. M., and H. V. Murphey, 2008: Airborne doppler radar and sounding analysis of an
772 oceanic cold front. *Monthly Weather Review*, **136** (4), 1475–1491.

773 Wang, W., and coauthors, 2012: Arw version 3 modeling system user’s guide. *NCAR Tech. Note*.

774 White, A., and Coauthors, 2013: A twenty-first-century california observing network for moni-
775 toring extreme weather events. *Journal of Atmospheric and Oceanic Technology*, **30** (8), 1585–
776 1603.

777 White, A. B., D. J. Gottas, A. F. Henkel, P. J. Neiman, F. M. Ralph, and S. I. Gutman, 2010: De-
778 veloping a performance measure for snow-level forecasts. *Journal of Hydrometeorology*, **11** (3),
779 739–753.

780 White, A. B., F. Ralph, P. Neiman, D. Gottas, and S. Gutman, 2009: The noaa coastal atmospheric
781 river observatory. *34th Conference on Radar Meteorology*.

782 Wick, G. A., P. J. Neiman, F. M. Ralph, and T. M. Hamill, 2013: Evaluation of forecasts of
783 the water vapor signature of atmospheric rivers in operational numerical weather prediction
784 models. *Weather and Forecasting*, **28** (6), 1337–1352, doi:10.1175/WAF-D-13-00025.1, URL
785 <http://dx.doi.org/10.1175/WAF-D-13-00025.1>.

786 Winterfeldt, J., B. Geyer, and R. Weisse, 2011: Using quikscat in the added value assessment of
787 dynamically downscaled wind speed. *International Journal of Climatology*, **31** (7), 1028–1039.

788 Zhu, Y., and R. E. Newell, 1998: A proposed algorithm for moisture fluxes from atmospheric
789 rivers. *Monthly Weather Review*, **126** (3), 725–735.

790	LIST OF TABLES	
791	Table 1.	Forecast Information Requirements to Meet Unique Challenges Posed by West-
792		ern U.S. Extreme Events. 39
793	Table 2.	OCN Cases Simulated for this Study. More Information regarding IOPs and
794		Aircraft can be found in Ralph et al. (2016). 40
795	Table 3.	Domain Attributes and Parameterized Physics Options for WRF Configurations
796		in this Study. 41
797	Table 4.	LND Cases Simulated in this Study 42
798	Table 5.	Verification Matrices Used in This Study. 43
799	Table 6.	Measures of error for storm total storm-scale forcing and local scale response,
800		as well as reduction of error in response by prescribing linearized ARO ob-
801		servated forcing and response to GFSRe and West-WRF LND case forecasts. 44

802 TABLE 1. Forecast Information Requirements to Meet Unique Challenges Posed by Western U.S. Extreme
803 Events.

Challenge	Primary Shortcoming of Current NWP	References
AR Landfall Characteristics	Location and strength of water vapor flux	Wick et al. (2013) Ralph et al. (2013)
Extreme Precipitation Skill	Overprediction of drizzle and light rain, Underprediction of heavy accumulations	Ralph et al. (2010a), Ralph and Dettinger (2012), Sukovich et al. (2014)
Snow Level	Low precision, Biases near terrain	White et al. (2010), Neiman et al. (2014), Minder and Kingsmill (2013)

804 TABLE 2. OCN Cases Simulated for this Study. More Information regarding IOPs and Aircraft can be found
805 in Ralph et al. (2016).

IOP	Aircraft	Transect Start	Transect End
CWES 1	NOAA G-IV	Feb 7, 2014 @ 2021 UTC	Feb 7, 2014 @ 2208 UTC
CWES 2	NOAA G-IV	Feb 8, 2014 @ 2050 UTC	Feb 8, 2014 @ 2146 UTC
CWES 2	NOAA G-IV	Feb 8, 2014 @ 2243 UTC	Feb 9, 2014 @ 2338 UTC
CWES 3	NOAA G-IV	Feb 11, 2014 @ 1903 UTC	Feb 11, 2014 @ 2124 UTC
CWES 4	NOAA G-IV	Feb 12, 2014 @ 1734 UTC	Feb 12, 2014 @ 1903 UTC
CWES 5	NOAA G-IV	Feb 13, 2014 @ 1833 UTC	Feb 13, 2014 @ 2058 UTC
CW2 1	NOAA G-IV	Jan 15, 2015 @ 2114 UTC	Jan 15, 2015 @ 2246 UTC
CW2 1	NOAA G-IV	Jan 15, 2015 @ 2307 UTC	Jan 16, 2015 @ 0020 UTC
CW2 2	NOAA G-IV	Jan 17, 2015 @ 2245 UTC	Jan 18, 2015 @ 0030 UTC
CW2 2	NOAA G-IV	Jan 18, 2015 @ 0130 UTC	Jan 18, 2015 @ 0307 UTC
CW2 4	NOAA G-IV	Jan 24, 2015 @ 2004 UTC	Jan 24, 2015 @ 2052 UTC
CW2 4	NOAA G-IV	Jan 24, 2015 @ 2126 UTC	Jan 24, 2015 @ 2303 UTC
CW2 6	NOAA G-IV	Feb 6, 2015 @ 2045 UTC	Feb 6, 2015 @ 2146 UTC
CW2 6	NOAA G-IV	Feb 6, 2015 @ 2159 UTC	Feb 6, 2015 @ 2259 UTC
CW2 6	NOAA G-IV	Feb 6, 2015 @ 2312 UTC	Feb 7, 2015 @ 0004 UTC

TABLE 3. Domain Attributes and Parameterized Physics Options for WRF Configurations in this Study.

Option	Outermost Domain	Nest Domain
dx	9 km	3 km
Time Step	Adaptive. ~ 45 s	Adaptive. ~ 15 s
Cumulus	Grell 3D	N/A
Land Surface	Noah	Noah
Cloud Microphysics	Thomson New	Thompson New
Planetary Boundary Layer	YSU	YSU
Surface Layer	Monin-Obukhov	Monin-Obukhov
SW Radiation	GSFC	GSFC
LW Radiation	RRTM	RRTM
Topographic Wind	No	Yes

TABLE 4. LND Cases Simulated in this Study

Case	Start at ARO	Duration	NWP Valid Start	NWP Valid End
1	Dec 10, 2014 @ 1500 UTC	32	Dec 10, 2014 @ 12 UTC	Dec 12, 2014 @ 00 UTC
2	Feb 6, 2015 @ 400 UTC	27	Feb 6, 2015 @ 00 UTC	Feb 7, 2015 @ 12 UTC
3	Feb 8, 2015 @ 900 UTC	25	Feb 8, 2015 @ 06 UTC	Feb 9, 2015 @ 12 UTC
4	Dec 9, 2015 @ 1300 UTC	26	Dec 9, 2015 @ 12 UTC	Dec 10, 2015 @ 18 UTC
5	Dec 20, 2015 @ 1400 UTC	47	Dec 20, 2015 @ 12 UTC	Dec 22, 2015 @ 18 UTC
6	Jan 17, 2016 @ 400 UTC	25	Jan 17, 2016 @ 00 UTC	Jan 18, 2016 @ 06 UTC
7	Jan 28, 2016 @ 1700 UTC	32	Jan 28, 2016 @ 18 UTC	Jan 30, 2016 @ 06 UTC
8	Mar 5, 2016 @ 2200 UTC	33	Mar 5, 2016 @ 18 UTC	Mar 7, 2016 @ 12 UTC
9	Mar 9, 2016 @ 800 UTC	42	Mar 9, 2016 @ 06 UTC	Mar 11, 2016 @ 06 UTC
10	Mar 12, 2016 @ 1500 UTC	37	Mar 12, 2016 @ 12 UTC	Mar 14, 2016 @ 06 UTC

TABLE 5. Verification Matrices Used in This Study.

Bin	1	2	3	4	5	6	7
7-bin center (hr)	24	48	72	96	120	144	168
7-bin boundaries (hr)	$12 \leq t_i \leq 35$	$36 \leq t_i \leq 59$	$60 \leq t_i \leq 83$	$84 \leq t_i \leq 107$	$108 \leq t_i \leq 131$	$132 \leq t_i \leq 155$	$156 \leq t_i \leq 179$
3-bin center (hr)	36	84	132	–	–	–	–
3-bin boundaries (hr)	$12 \leq t_i \leq 59$	$60 \leq t_i \leq 107$	$108 \leq t_i \leq 155$	–	–	–	–

806 TABLE 6. Measures of error for storm total storm-scale forcing and local scale response, as well as reduction
807 of error in response by prescribing linearized ARO observed forcing and response to GFSRe and West-WRF
808 LND case forecasts.

Lead Time (hr)		$12 \leq t_i \leq 59$	$60 \leq t_i \leq 107$	$108 \leq t_i \leq 155$
Error Measure	Model			
e_{xy}	GFSRe	4.287	4.530	7.370
	West-WRF	0.820	2.253	4.653
e_y	GFSRe	1.544	2.072	2.568
	West-WRF	0.470	1.295	2.109
δe_{ypf}	GFSRe	18.1%	14.2%	0.0%
	West-WRF	10.7%	-57.6%	-78.1%
δe_{ypr}	GFSRe	18.9%	-35.0%	-19.3%
	West-WRF	20.4	-47.2%	-51.9%
Linear Slope	Model/Obs			
$(mm\ s\ cm^{-1}\ m^{-1})$	GFSRe	0.003	-0.003	0.005
	West-WRF	0.084	0.099	0.091
	Obs	0.094	0.094	0.094

LIST OF FIGURES

- Fig. 1.** Composite IWV over the Northeast Pacific Ocean on February 8, 2014 from the Special Sensor Microwave Imager (SSM/I). Overlaid are three aircraft transects from the CalWater Early Start campaign (white lines). The Westernmost transects correspond to OCN cases 2 and 3 from Table 2. Additional annotations depict the impacts of this AR and a dynamic feature named a Mesoscale Frontal Wave. These are discussed in Ralph et al. (2016). Figure adapted from Ralph et al. (2016). 47
- Fig. 2.** Conditional probability $P(AR_{ij}|AR_{RRW})$ (%) constructed by intersecting Gershunov et al. catalog AR landfall days at the ARO with frequency of days with $IVT \geq 250 \text{ kg m}^{-1} \text{ s}^{-1}$ from NCEP/NCAR Reanalysis version 1. West-WRF 9 km (a), WRF-ARWS 9 km (b), and West-WRF 3 km (c) domain boundaries are overlaid. 48
- Fig. 3.** Value added by GFSRe over GFSRe climatology validated against CalWater 2 dropsondes for the state variables z^{500} (blue), IVT (black), IWV (green) and θ_e^{925} (red). b) as in a, except variable is West-WRFN 9 km forecast value added over GFSRe reference forecasts. c) as in b, except value added forecast is West-WRF 9 km. d) as in b, except value added forecast is WRF-ARWS 9 km. 49
- Fig. 4.** a) as in Fig. 2c, except value added is calculated using West-WRFN 9 km as the reference forecast. b) as in Fig. 2d, except value added is calculated using West-WRFN 9 km as the reference forecast. 50
- Fig. 5.** a) Predicted lower tropospheric (1000 hPa to 700 hPa) profile of N_m^2 (10^{-4} s^{-2}) for $12 \text{ hr} \leq t_i \leq 59 \text{ hr}$. Observation (black), GFSRe (red), West-WRF 9 km (blue), and WRF-ARWS 9 km (green) median value is displayed as a colored dot, while horizontal lines represent the interquartile range of values at the given pressure level. For each panel, the right split window displays the root mean-square error (10^{-4} s^{-2}) at the given pressure level for GFSRe (red), West-WRF 9 km (blue), and WRF-ARWS 9 km (green). b) As in a, except $60 \text{ hr} \leq t_i \leq 107 \text{ hr}$. c) As in a, except $108 \text{ hr} \leq t_i \leq 155 \text{ hr}$ 51
- Fig. 6.** top row) Fifteen case ensemble mean AR normal - vertical transect of $dIVT$ error (model - obs) ($\text{kg m}^{-2} \text{ s}^{-1}$ - black solid, dashed where negative), and θ_e^{925} (K - blue dashed) for $12 \text{ hr} \leq t_i \leq 59 \text{ hr}$. left panel displays mean error in West-WRF 9 km transects, right displays same quantity for GFSRe. Stipling indicates significance at $p < 0.05$ according to student's t-test. Middle row) as in a, except $60 \text{ hr} \leq t_i \leq 107 \text{ hr}$ in left and right panels. Middle panel displays the observed ensemble mean $dIVT$ ($\text{kg m}^{-2} \text{ s}^{-1}$ - black solid) and θ_e^{925} (K - blue dashed). Bottom row) as in top row, except $108 \text{ hr} \leq t_i \leq 155 \text{ hr}$ 52
- Fig. 7.** As in Fig. 4, except leftmost panels on all rows display the forecast cross-sections for WRF-ARWS 9 km. 53
- Fig. 8.** Error in timing of LND case AR condition start at the ARO for a) West-WRF 3 km and b) GFSRe forecasts verifying with $t_i \leq 179 \text{ hr}$ 54
- Fig. 9.** a) Value histogram for RRW ST QPE (NCEP Stage-IV, green bars), West-WRF (blue 'x'), and GFSRe (red '+') ST QPF calculated from $12 \text{ hr} \leq t_i \leq 59 \text{ hr}$. Inset: West-WRF and GFSRe mean (AVG), standard deviation (STD) and root mean square (RMS) $QPF - QPE$ from all RRW Stage-IV grid points. b) as in a, except $60 \text{ hr} \leq t_i \leq 107 \text{ hr}$. c) as in a, except lead times are $108 \text{ hr} \leq t_i \leq 155 \text{ hr}$ 55

851 **Fig. 10.** a) $ST P_r$ and $ST BUF$ at the ARO for 171 historical rain events (dark gray dots) and the
 852 linear trend line resulting from a least-square fit (red line). The correlation coefficient (R^2)
 853 and e_y that result from the least-square fit are provided. Also provided are one example from
 854 the LND case list of the observed (orange circle), GFSRe forecast (orange asterisk), and
 855 West-WRF forecast (orange circle and cross) position. The segments A and B represent the
 856 non-dimensional distance measured by e_{xy} . b) as in a, except shown are all LND observed
 857 (colored circles), GFSRe forecasts with $60 \text{ hr} \leq t_i \leq 83 \text{ hr}$ (colored asterisk), and West-
 858 WRF forecasts with lead times $60 \text{ hr} \leq t_i \leq 83 \text{ hr}$ (colored circle and cross). The inset table
 859 displays e_{xy} for each model. c) as in a, except only cases for which West-WRF outperformed
 860 GFSRe according to e_{xy} are displayed. d) as in b, except shown are cases for which GFSRe
 861 outperformed West-WRF or for which performance was nearly equivalent. 56

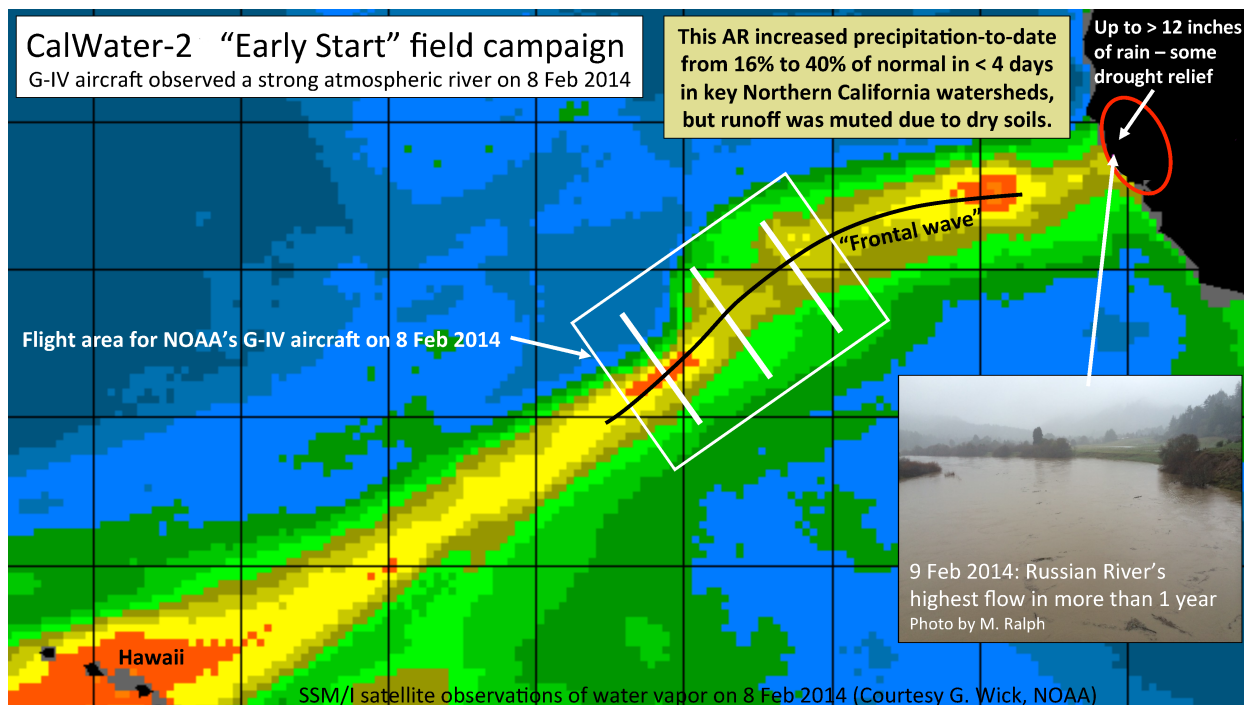
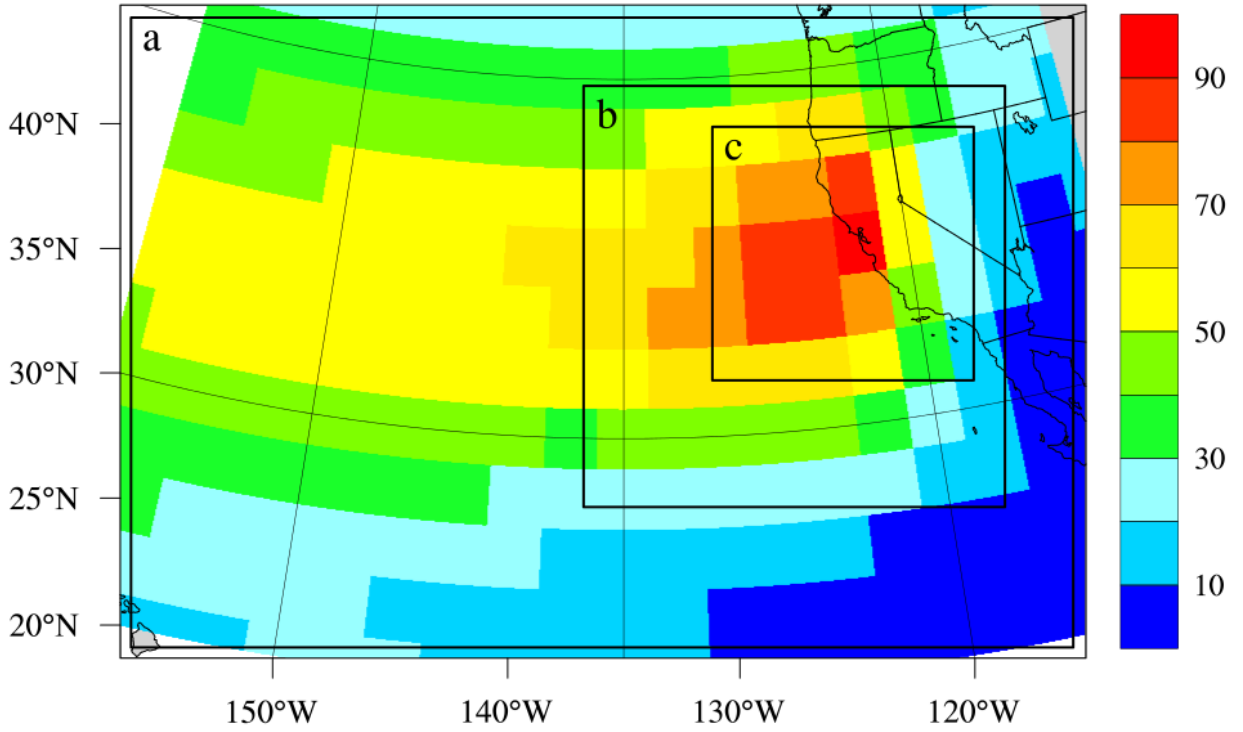


FIG. 1. Composite IWV over the Northeast Pacific Ocean on February 8, 2014 from the Special Sensor Microwave Imager (SSM/I). Overlaid are three aircraft transects from the CalWater Early Start campaign (white lines). The Westernmost transects correspond to OCN cases 2 and 3 from Table 2. Additional annotations depict the impacts of this AR and a dynamic feature named a Mesoscale Frontal Wave. These are discussed in Ralph et al. (2016). Figure adapted from Ralph et al. (2016).



867 FIG. 2. Conditional probability $P(AR_{ij}|AR_{RRW})$ (%) constructed by intersecting Gershunov et al. catalog
 868 AR landfall days at the ARO with frequency of days with $IVT \geq 250 \text{ kg m}^{-1} \text{ s}^{-1}$ from NCEP/NCAR Reanalysis
 869 version 1. West-WRF 9 km (a), WRF-ARWS 9 km (b), and West-WRF 3 km (c) domain boundaries are overlaid.

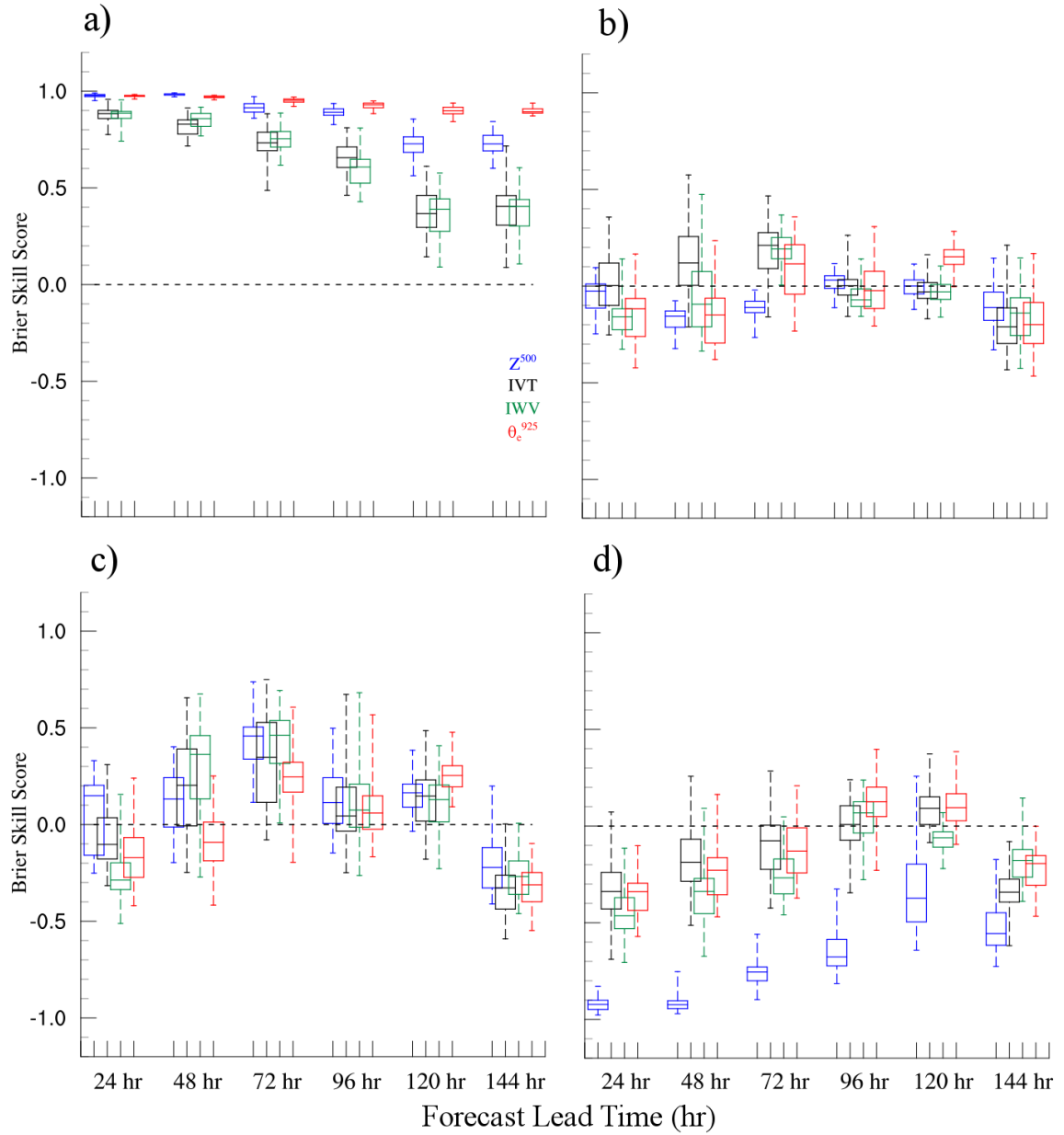
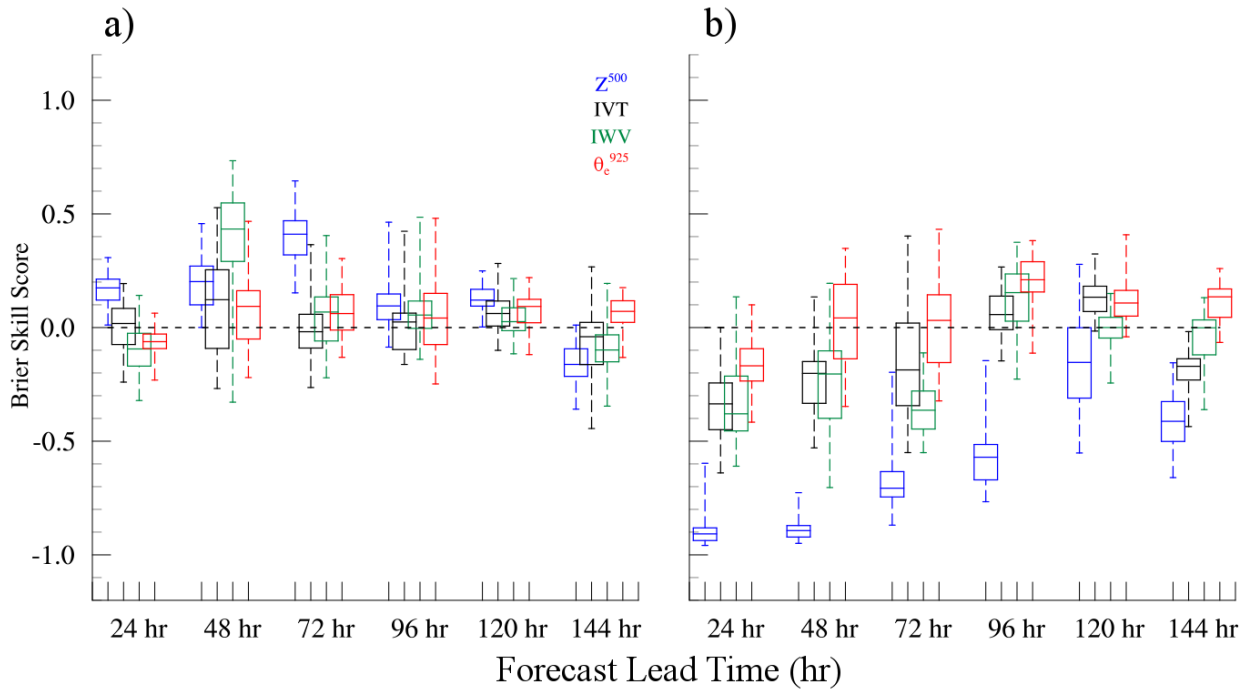


FIG. 3. Value added by GFSRe over GFSRe climatology validated against CalWater 2 dropsondes for the state variables z^{500} (blue), IVT (black), IWV (green) and θ_e^{925} (red). b) as in a, except variable is West-WRFN 9 km forecast value added over GFSRe reference forecasts. c) as in b, except value added forecast is West-WRF 9 km. d) as in b, except value added forecast is WRF-ARWS 9 km.



874 FIG. 4. a) as in Fig. 2c, except value added is calculated using West-WRFN 9 km as the reference forecast.
875 b) as in Fig. 2d, except value added is calculated using West-WRFN 9 km as the reference forecast.

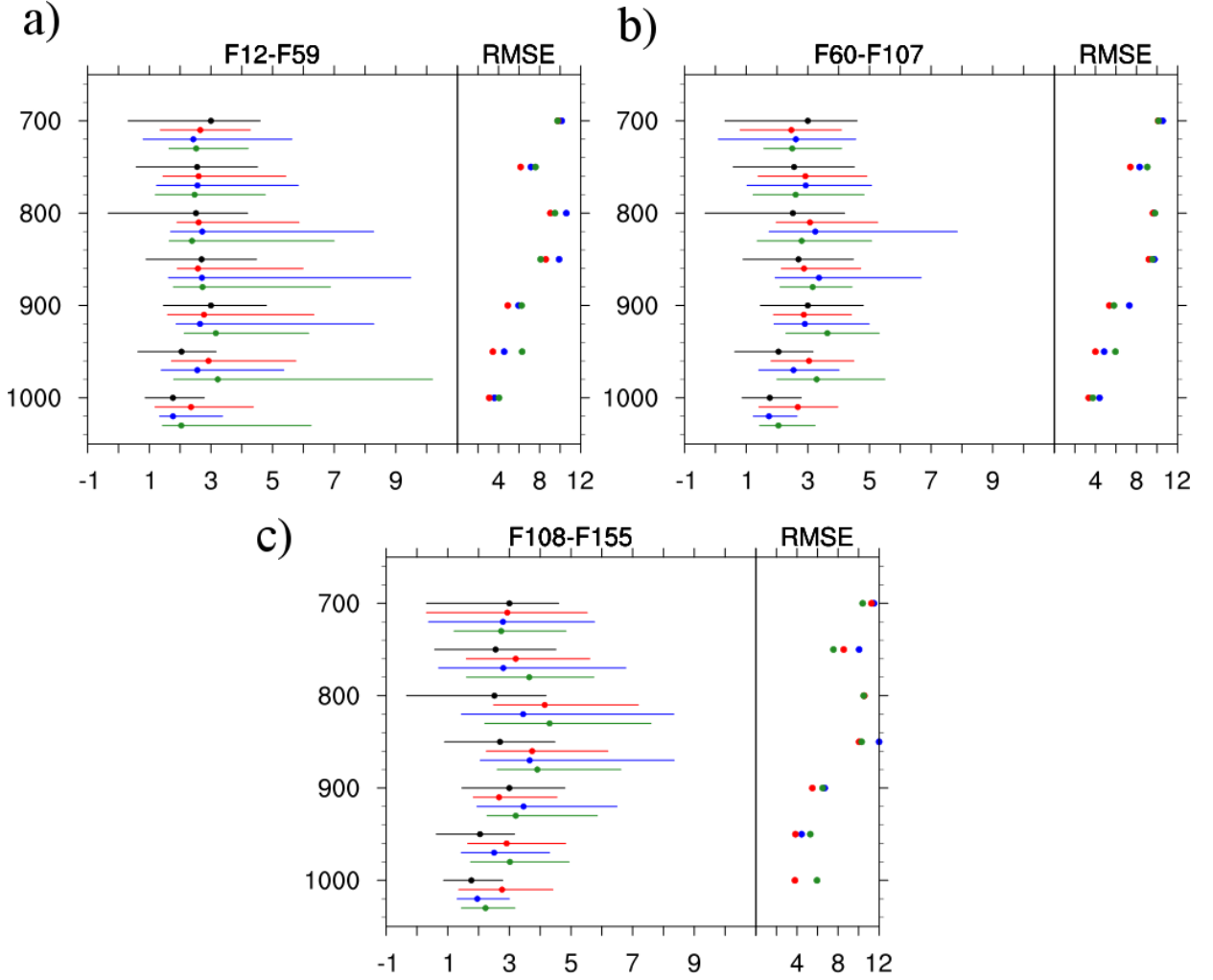


FIG. 5. a) Predicted lower tropospheric (1000 hPa to 700 hPa) profile of N_m^2 (10^{-4} s^{-2}) for $12 \text{ hr} \leq t_i \leq 59$ hr. Observation (black), GFSRe (red), West-WRF 9 km (blue), and WRF-ARWS 9 km (green) median value is displayed as a colored dot, while horizontal lines represent the interquartile range of values at the given pressure level. For each panel, the right split window displays the root mean-square error (10^{-4} s^{-2}) at the given pressure level for GFSRe (red), West-WRF 9 km (blue), and WRF-ARWS 9 km (green). b) As in a, except $60 \text{ hr} \leq t_i \leq 107 \text{ hr}$. c) As in a, except $108 \text{ hr} \leq t_i \leq 155 \text{ hr}$.

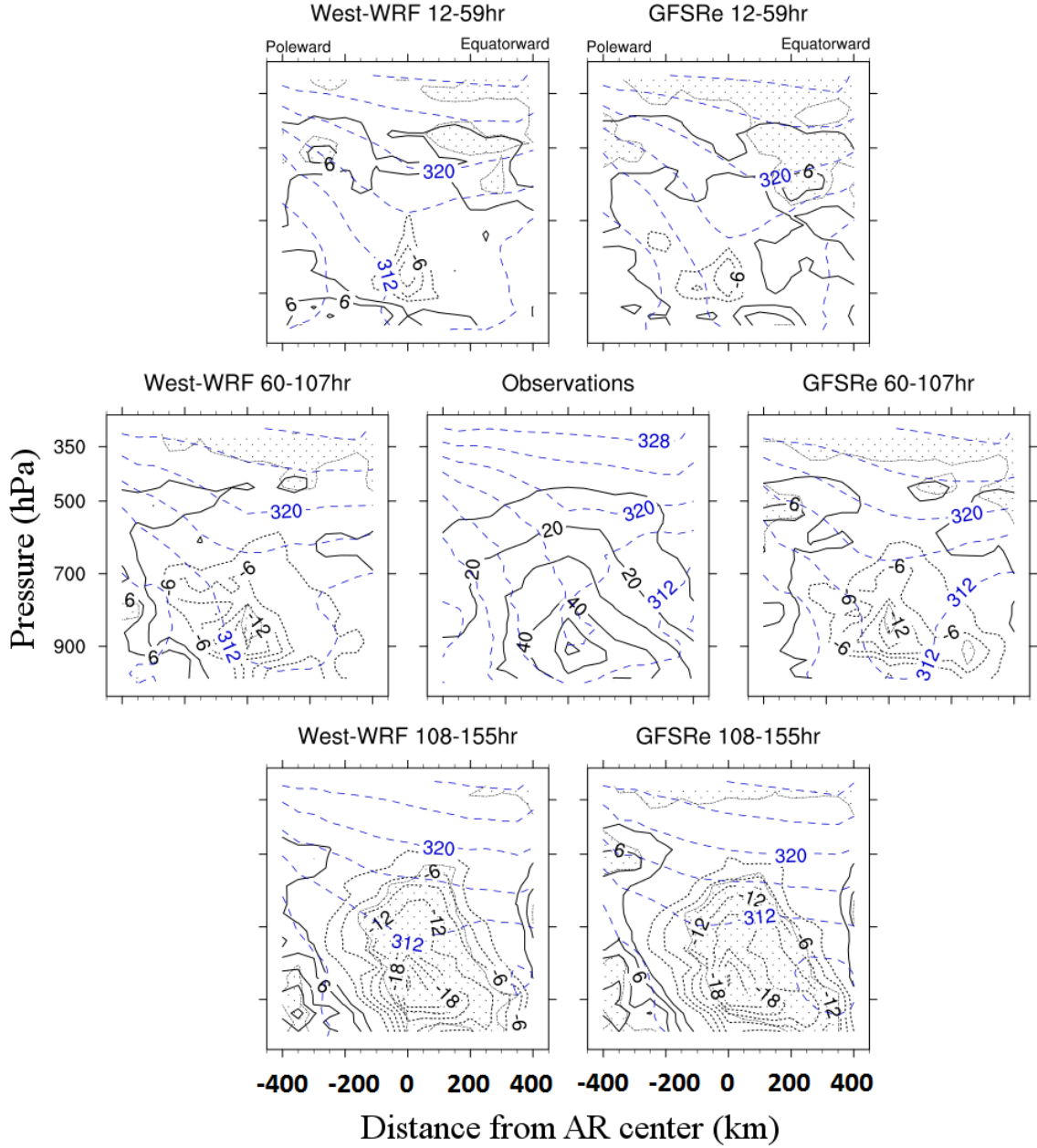


FIG. 6. top row) Fifteen case ensemble mean AR normal - vertical transect of $dIVT$ error (model - obs) ($\text{kg m}^{-2} \text{ s}^{-1}$ - black solid, dashed where negative), and θ_e^{925} (K - blue dashed) for $12 \text{ hr} \leq t_i \leq 59 \text{ hr}$. left panel displays mean error in West-WRF 9 km transects, right displays same quantity for GFSRe. Stippling indicates significance at $p < 0.05$ according to student's t-test. Middle row) as in a, except $60 \text{ hr} \leq t_i \leq 107 \text{ hr}$ in left and right panels. Middle panel displays the observed ensemble mean $dIVT$ ($\text{kg m}^{-2} \text{ s}^{-1}$ - black solid) and θ_e^{925} (K - blue dashed). Bottom row) as in top row, except $108 \text{ hr} \leq t_i \leq 155 \text{ hr}$.

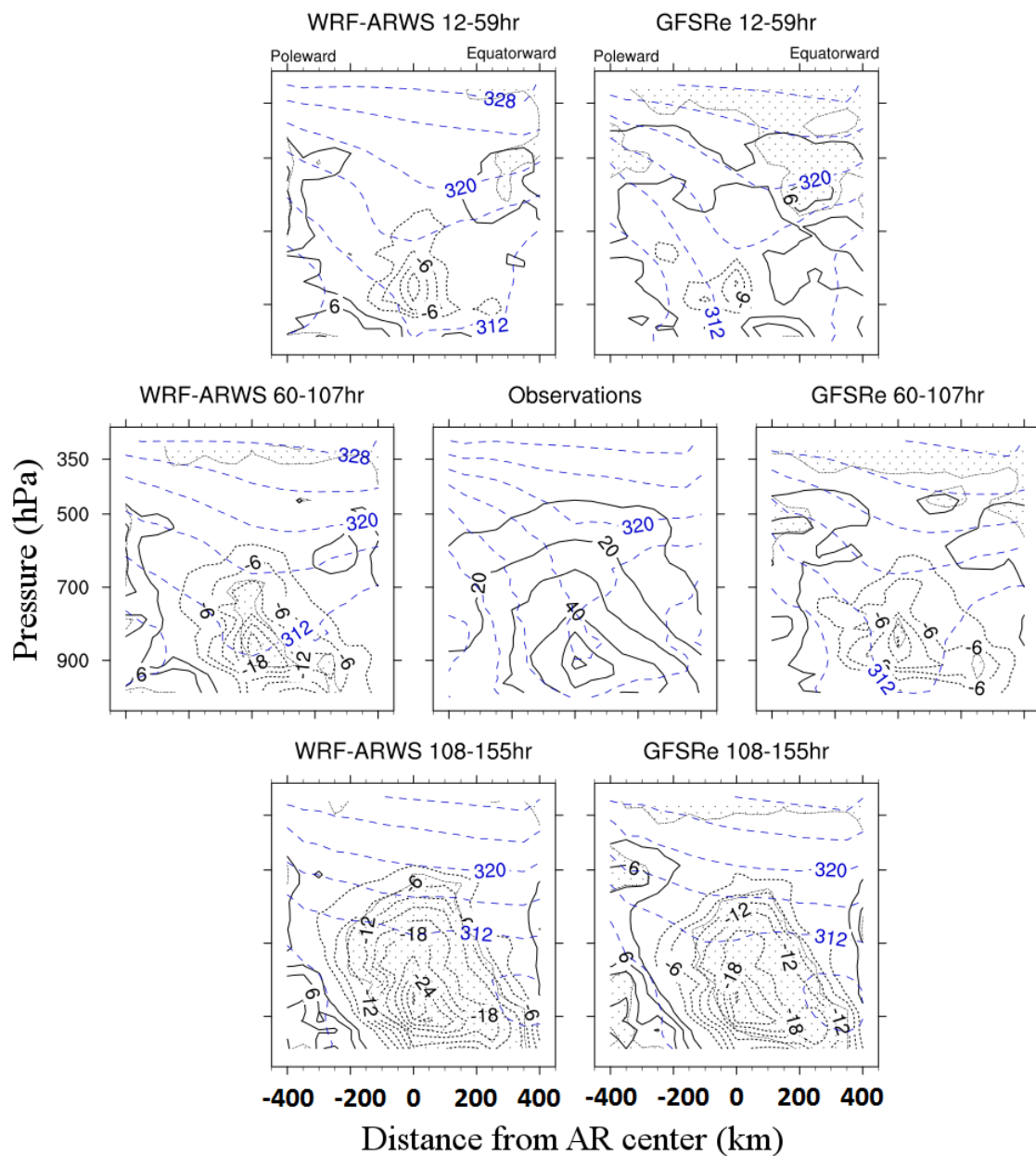
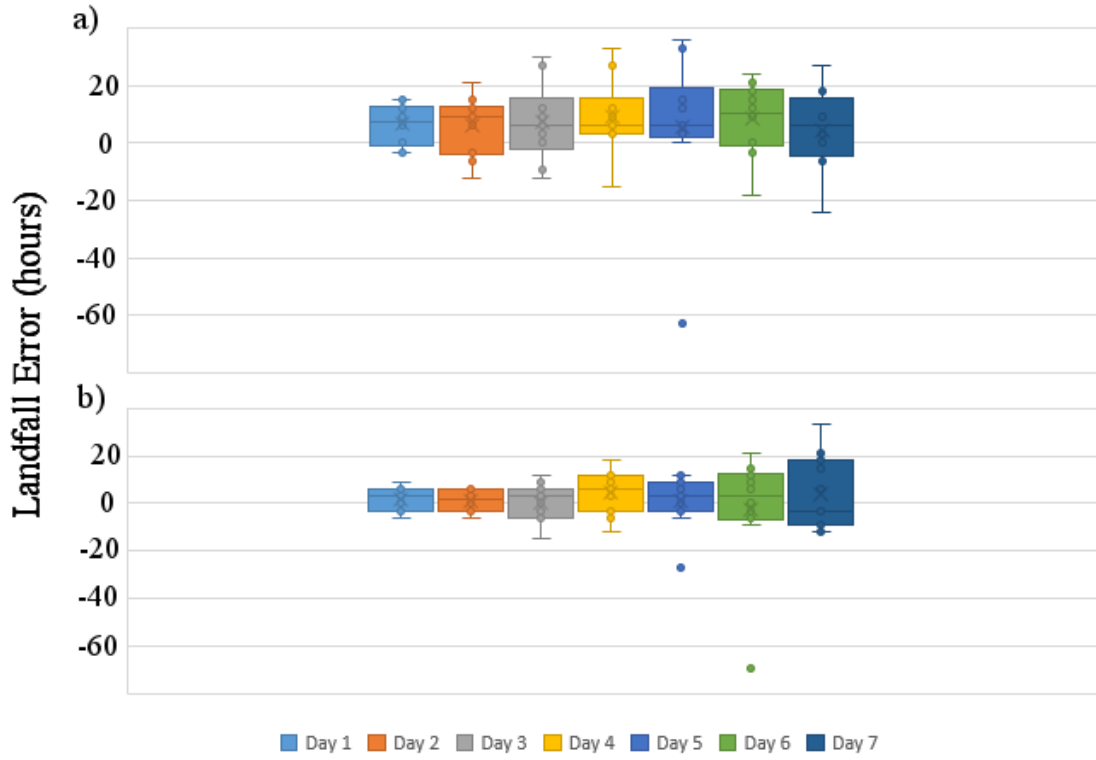


FIG. 7. As in Fig. 4, except leftmost panels on all rows display the forecast cross-sections for WRF-ARWS 9

km.



890 FIG. 8. Error in timing of LND case AR condition start at the ARO for a) West-WRF 3 km and b) GFSRe
891 forecasts verifying with $t_i \leq 179$ hr.

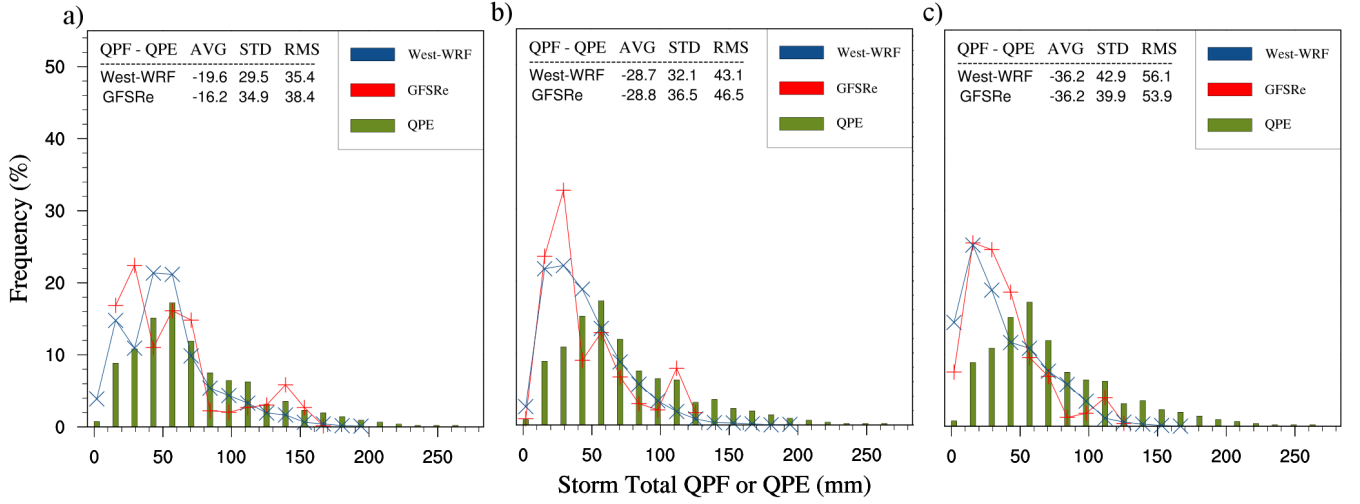


FIG. 9. a) Value histogram for RRW ST QPE (NCEP Stage-IV, green bars), West-WRF (blue 'x'), and GFSRe (red '+') ST QPF calculated from $12 \text{ hr} \leq t_i \leq 59 \text{ hr}$. Inset: West-WRF and GFSRe mean (AVG), standard deviation (STD) and root mean square (RMS) $QPF - QPE$ from all RRW Stage-IV grid points. b) as in a, except $60 \text{ hr} \leq t_i \leq 107 \text{ hr}$. c) as in a, except lead times are $108 \text{ hr} \leq t_i \leq 155 \text{ hr}$.

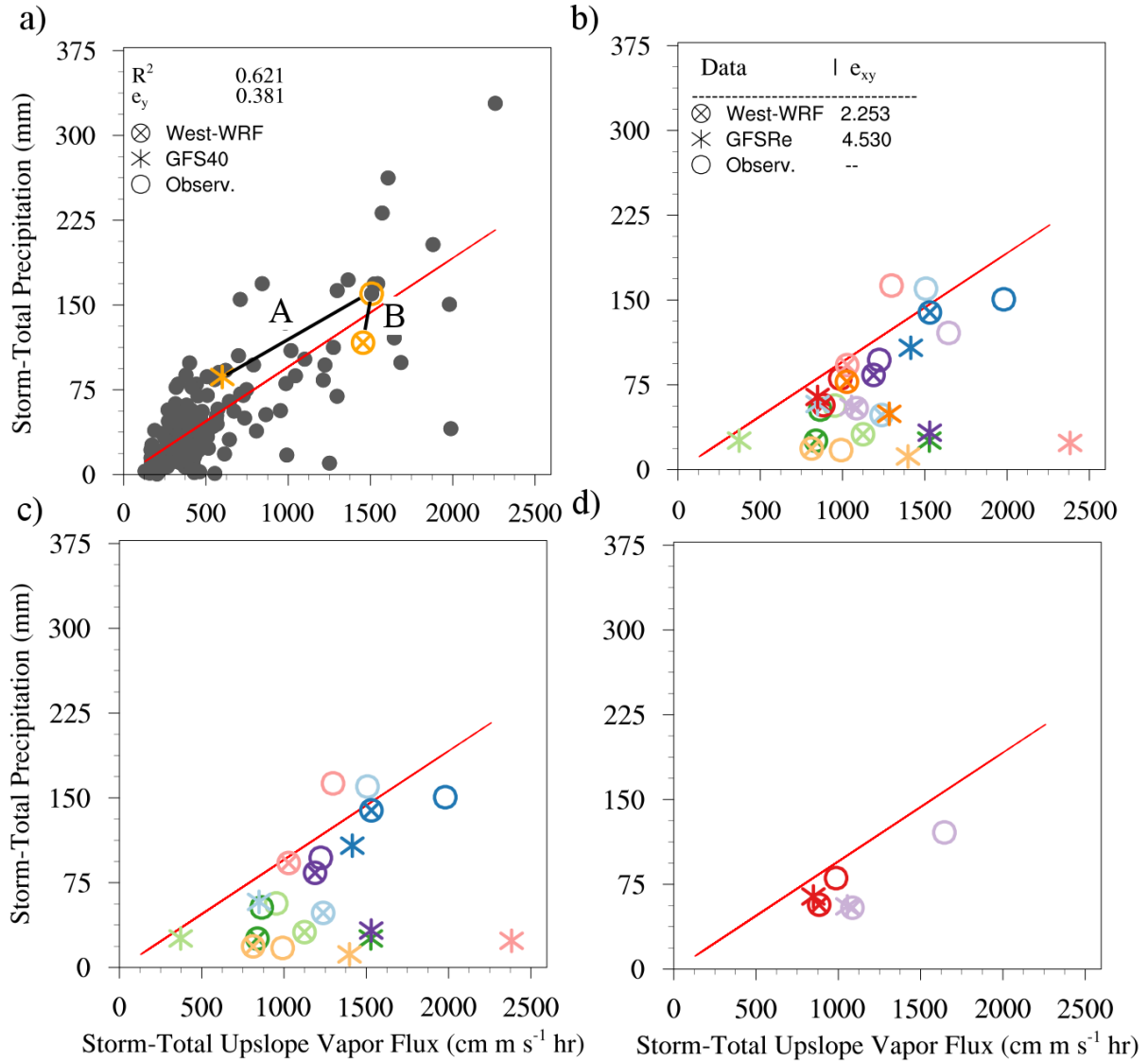


FIG. 10. a) $ST P_r$ and $ST BUF$ at the ARO for 171 historical rain events (dark gray dots) and the linear trend line resulting from a least-square fit (red line). The correlation coefficient (R^2) and e_y that result from the least-square fit are provided. Also provided are one example from the LND case list of the observed (orange circle), GFSRe forecast (orange asterisk), and West-WRF forecast (orange circle and cross) position. The segments A and B represent the non-dimensional distance measured by e_{xy} . b) as in a, except shown are all LND observed (colored circles), GFSRe forecasts with $60 \text{ hr} \leq t_i \leq 83 \text{ hr}$ (colored asterisk), and West-WRF forecasts with lead times $60 \text{ hr} \leq t_i \leq 83 \text{ hr}$ (colored circle and cross). The inset table displays e_{xy} for each model. c) as in a, except only cases for which West-WRF outperformed GFSRe according to e_{xy} are displayed. d) as in b, except shown are cases for which GFSRe outperformed West-WRF or for which performance was nearly equivalent.