Improved Forecast Skill Through the Assimilation of Dropsonde Observations From the Atmospheric River Reconnaissance Program

Minghua Zheng[1] [iD], **Luca Delle Monache[1]** [iD], **Bruce D. Cornuelle[2]** [iD], **F. Martin Ralph[1]** [iD], **Vijay S. Tallapragada[3]** [iD], **Aneesh Subramanian[4]** [iD], **Jennifer S. Haase[2]** [iD], **Zhenhai Zhang[1]** [iD], **Xingren Wu[5]** [iD], **Michael J. Murphy[2]** [iD], **Timothy B. Higgins[4]** [iD], and **Laurel DeHaan[1]** [iD]

[1]Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA, [2]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA, [3]National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP)/Environmental Modeling Center (EMC), College Park, MD, USA, [4]University of Colorado, Boulder, CO, USA, [5]I. M. Systems Group Inc. at NOAA NCEP/EMC, College Park, MD, USA

Abstract Landfalling atmospheric rivers (ARs) over the western US are responsible for ∼30%–50% of the annual precipitation, and their accurate forecasts are essential for aiding water management decisions and reducing flood risks. Sparse coverage of conventional observations over the Pacific Ocean, which can cause inadequate upstream initial conditions for numerical weather prediction models, may limit the improvement of forecast skill for these events. A targeted field program called AR Reconnaissance (Recon) was initiated in 2016 to better understand and reduce forecast errors of landfalling ARs at 1–5 days lead times. During the winter seasons of 2016, 2018, and 2019, 15 Intensive Observation Periods (IOPs) sampled the upstream conditions for landfalling ARs. This study evaluates the impact on forecast accuracy of assimilating these dropsonde data. Data denial experiments with (WithDROP) and without (NoDROP) dropsonde data were conducted using the Weather Research and Forecasting model with the Gridpoint Statistical Interpolation four-dimensional ensemble variational system. Comparisons between the 15 paired NoDROP and WithDROP experiments demonstrate that AR Recon dropsondes reduced the root-mean-square error in integrated vapor transport (IVT) and inland precipitation for more than 70% of the IOPs, averaged over all forecast lead times from 1 to 6 days. Dropsondes have improved the spatial pattern of forecasts of IVT and precipitation in all 15 IOPs. Significant improvements in skill are found beyond the short range (1–2 days). IOP sequences (i.e., back-to-back IOPs every other day) show the most improvement of inland precipitation forecast skill.

Plain Language Summary Atmospheric rivers (ARs) are long and narrow corridors of water vapor transport in the atmosphere from the tropics to higher latitudes. Up to half of the annual precipitation in the western US states is related to landfalling ARs. Therefore, accurate forecasts of these events are important to reduce socio-economic damage from the associated high-impact weather and to better manage water resources over the western US. However, the forecast accuracy can be limited by sparse observations upstream over the Pacific Ocean. To fill observation gaps and improve forecast skills, a weather reconnaissance mission called Atmospheric River Reconnaissance was initiated in 2016 to release dropsondes from aircraft in key upstream sensitive regions. This study assesses the value of these observations in improving the initial conditions and the forecast skill of a regional model operated at the Center for Western Weather and Water Extremes for near real-time weather forecasts. We found that the use of these additional observations collected during 2016, 2018, and 2019 has a positive impact on reducing the forecast error of water vapor transport within the Northeastern Pacific domain and the inland precipitation skill over the western US. These observations have more impact in poorly forecasted cases.

1. Introduction

Forecast skill of a numerical weather prediction (NWP) model depends on many factors, including observations of the initial atmospheric state and the ocean/land surface, data assimilation techniques, and model

Formal analysis: Minghua Zheng, Zhenhai Zhang

Funding acquisition: F. Martin Ralph

Investigation: Minghua Zheng

Methodology: Minghua Zheng, Luca Delle Monache, Bruce D. Cornuelle, Vijay S. Tallapragada, Aneesh Subramanian, Xingren Wu, Michael J. Murphy

Project Administration: Luca Delle Monache, F. Martin Ralph

Software: Bruce D. Cornuelle, Aneesh Subramanian

Supervision: Luca Delle Monache, F. Martin Ralph, Vijay S. Tallapragada

Validation: Minghua Zheng, Zhenhai Zhang, Timothy B. Higgins, Laurel DeHaan

Visualization: Minghua Zheng, Luca Delle Monache

Writing – original draft: Minghua Zheng

Writing – review & editing: Minghua Zheng, Luca Delle Monache, Bruce D. Cornuelle, F. Martin Ralph, Vijay S. Tallapragada, Aneesh Subramanian, Jennifer S. Haase, Zhenhai Zhang, Xingren Wu, Michael J. Murphy, Timothy B. Higgins, Laurel DeHaan

dynamics and physics. In particular, the forecast is sensitive to the availability and quality of observations in the upstream regions and the initial analysis that assimilates those observations (Bengtsson et al., 2005; Froude et al., 2007; Langland et al., 1999; Peevey et al., 2018). For example, the western United States (US) can suffer from forecast skill degradation due to a lack of high-quality observations over the Northeastern Pacific Ocean (Charles & Colle, 2009; McMurdie & Casola, 2009; Wedam et al., 2009). High-accuracy data that can resolve the detailed vertical structure of the low-level circulation is of key importance over data-sparse regions (Agustí-Panareda et al., 2010; Zheng, Delle Monache, Wu, et al., 2021). This study focuses on improving the forecasting of atmospheric rivers (ARs), which are narrow corridors that transport water vapor in the atmosphere from the tropics to higher latitudes and play a key role in regional weather and hydrology (Guan & Waliser, 2019; Ralph et al., 2004, 2006; Waliser & Guan, 2017; Zhang et al., 2019; Zhu & Newell, 1998). Landfalling ARs originating from the Pacific Ocean are responsible for up to 50% of the annual precipitation over the western US (Dettinger et al., 2011), thereby contributing to both water supply and flood risks (Corringham et al., 2019; Dettinger, 2013; Neiman et al., 2008; Rutz et al., 2014). Khouakhi and Villarini (2016) found that 22%–65% of annual storm surge maxima along the US West Coast were also associated with ARs.

Despite the growing awareness of the key roles of landfalling ARs in extreme weather and water events, operational NWP models still have a significant margin for improvement when forecasting the landfalling metrics (Martin et al., 2019; Wick et al., 2013). To improve the sampling of ARs and the forecast skill of landfalling events over the western US, the Atmospheric River Reconnaissance field program (AR Recon, Ralph et al., 2020) was undertaken over several winter seasons to collect more observational data within and near ARs over the Northeastern Pacific. AR Recon collected dropsonde and ancillary data in dynamically active regions where moisture convergence, precipitation, and latent heating influence frontogenesis (Lackmann, 2002) and cyclogenesis (Davis, 1992; Zhang & Ralph, 2021) which eventually modify downstream precipitation. Since 2018, observation targeting based on sensitivity analysis methods has been implemented using adjoint techniques (Demirdjian et al., 2020; Doyle et al., 2014, 2019; Errico, 1997; Reynolds et al., 2019) and ensemble-based methods (Ancell & Hakim, 2007; Chang et al., 2013; Torn & Hakim, 2008; Zhang et al., 2007; Zheng et al., 2013). Ensemble-based sensitivity analyses emphasize synoptic-scale differences between ensemble members that are associated with significant weather features at the initial time or an earlier forecast time. This technique has been used successfully in past field campaigns, particularly for hurricanes and severe weather events (Majumdar, 2016; Romine et al., 2016). The moist adjoint sensitivity calculations available from the Naval Research Laboratory (NRL) Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS) are demonstrated to be robust in identifying valid areas for targeted observations of ARs (Doyle et al., 2014; Reynolds et al., 2019).

Most of the previous targeted missions showed overall positive impact on forecast skill for a variety of high impact weather events (Aberson, 2010; Feng & Wang, 2019; Joly et al., 1999; Langland, 2005; Langland et al., 1999; Pu et al., 2008; Romine et al., 2016; Schindler et al., 2020; Szunyogh et al., 2000; Weissmann et al., 2011) specifically in short-range (<3 days) forecasts. For example, dropsonde observations collected from the North Atlantic Waveguide and Downstream Impact Experiment (NAWDEX) campaign reduced mean forecast error in the 500-hPa geopotential height by 1%–3% over Europe and the North Atlantic domain with the European Centre for Medium-Range Weather Forecast (ECMWF) global model (Schindler et al., 2020). Other missions showed neutral impact on model forecast skill (Hamill et al., 2013; Tong et al., 2018). Lavers et al. (2018) used the independent AR Recon dropsonde data to show that the water vapor flux forecast errors in ARs using the ECMWF Integrated Forecasting System (IFS) are most strongly associated with the initial conditions in 850 hPa wind, which is typically near the top of the planetary boundary layer, and the amplitude of the model errors is about 20% of the mean observed water vapor flux. Stone et al. (2020) found that AR Recon soundings collected during the winter of 2018 have a significant positive impact in the Navy Global Environmental Model (NAVGEM) using a method called the Forecast Sensitivity Observation Impact (e.g., Lorenc & Marriott, 2014), and found that the per observation impact of AR Recon soundings in NAVGEM is more than double that of the North American radiosonde network. A recent study by Zheng, Delle Monache, Wu, et al. (2021) found that AR Recon can fill observation gaps from near the surface to the middle troposphere by providing high-quality in-situ observations within and near ARs, where all-sky radiances either have degraded quality or have been rejected by the data assimilation system due to heavy precipitation. This enhances observations over regions with the highest adjoint

sensitivity for downstream 1–2 days precipitation forecasts (Reynolds et al., 2019). These works argue that high-vertical resolution profiles from AR Recon dropsondes can fill critical data gaps and bring new information to operational model analyses, and have the potential to improve forecasts of landfalling ARs.

The goal of this work is to investigate the impact of assimilation of these AR Recon data on the model forecast skill for landfalling ARs and their associated inland precipitation in a regional implementation of the Weather Research and Forecasting (WRF, Skamarock et al., 2008) model called West-WRF that is tailored to predictions over the western US (Martin et al., 2018). A set of data denial experiments based on three winter seasons of observations for 15 IOPs (Table S1 in Supporting Information S1) from AR Recon (2016, 2018, and 2019) were conducted to examine the impacts of this data set (Zheng, Delle Monache, Wu, et al., 2021). Specifically, this study explores how AR Recon data ingested by the model modify the initial conditions in the Northeastern Pacific Ocean and affect forecasts of ARs and precipitation over the western US. Below are the three motivational questions:

- How do AR Recon data modify initial conditions in the presence of ARs in the regional West-WRF model?
- What is the impact of AR Recon data on the model forecast skill for ARs?
- Have the AR Recon data improved the overall precipitation forecast skill over the US West?

Section 2 describes the model configuration and experiment setup. Section 3 investigates the data impact in a case study in February 2018. Section 4 quantifies the overall data impact on the forecasts of ARs for the complete data set using several forecast and precipitation metrics. The discussions in Section 5 examine the variation of impact across cases and offer interpretations of the results. Section 6 summarizes the main conclusions from the study and discusses future work.

2. Model Configuration, Experimental Design, and Verification Data

2.1. Model Configuration

The model we used for this study is a research version of the Center for Western Weather

and Water Extremes (CW3E) operational West-WRF model, which is tuned to optimize predictions for the western US region (Martin et al., 2018). West-WRF is based on the Advanced Research WRF (WRF ARW) model, version 3.9.1.1. West-WRF adds value to forecast fields compared to the global forcing model, especially for the orographic precipitation and horizontal transport of moisture that are frequently associated with landfalling ARs (see Martin et al., 2018 for more details). Two domains were configured using horizontal resolutions of 9 and 3 km (denoted D01 and D02 in Figure 1a). Both domains are configured with 48 layers in the vertical with a model top at 10 hPa. The main configuration is as follows: (a) new Thompson microphysics scheme (Thompson et al., 2008); (b) Yonsei University planetary boundary layer scheme (Hong et al., 2006); (c) Rapid Radiative Transfer Model for long-wave and short-wave radiation (Chou & Suarez, 1994; Mlawer et al., 1997; Matsui et al., 2018); (d) Noah-MP land surface model (Niu et al., 2011); and (e) Grell-Freitas scheme (Grell & Freitas, 2014) for cumulus parameterization (only for the outer 9-km domain). The National Center for Environmental Prediction (NCEP) Global Forecast System (GFS) operational analysis/forecast products at 0.25° × 0.25° lat-lon grids are used for initial and boundary conditions of the outer domain D01. The inland precipitation for the 3-km domain is validated across all AR Recon Intense Observation Periods (IOPs) over a common domain as shown in Figure 1b. The verification region for integrated water vapor (IWV) and vapor transport (IVT) extends from 105°–165°W to 15°–60°N. IVT and IWV are calculated following the definitions proposed by Zhu and Newell (1998):

$$IVT = -\frac{1}{g} \int_{p_{sfc}}^{300\,hPa} \left(q V_{h} \right) dp, \tag{1}$$

$$IWV = -\frac{1}{g} \int_{p_{sfc}}^{300\,hPa} q\, dp, \tag{2}$$

where g is the gravitational constant, q is the specific humidity, $\mathbf{V}_{h}$ represents the horizontal wind vector, and p_{sfc} is the surface pressure.
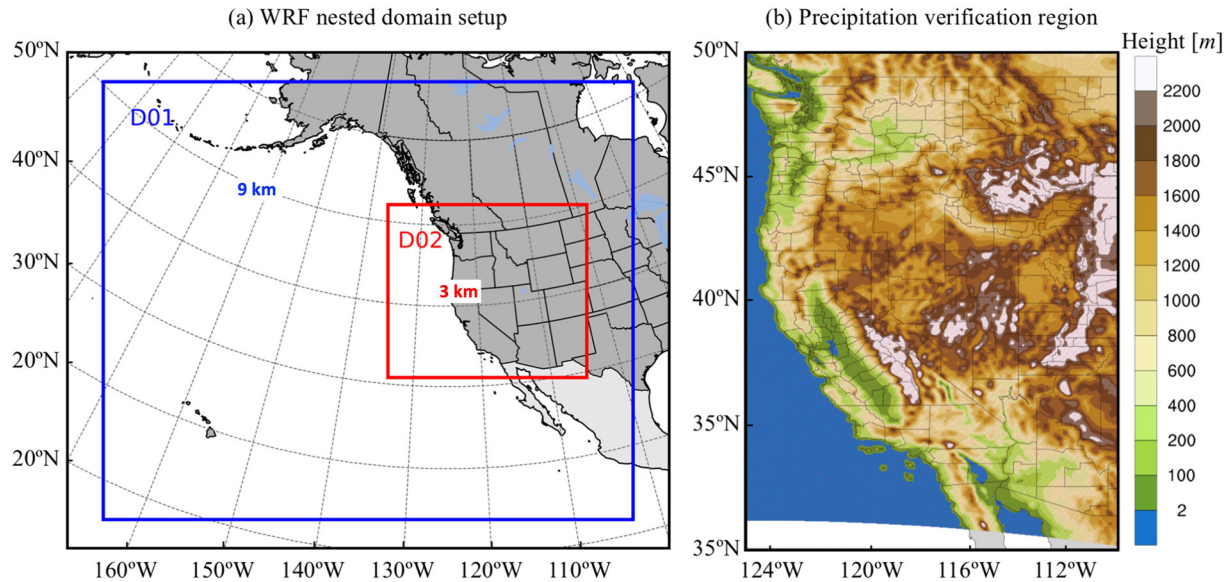
Figure 1. Weather Research and Forecasting (WRF) domain configuration (a) and a common verification region (b) for inland precipitation (110°–125°W and 30°–50°N).

2.2. Data Assimilation System and the Experiment Design

For this study, we used the Gridpoint Statistical Interpolation (GSI) hybrid four-dimensional ensemble-variational (4D-EnVar) data assimilation system (Kleist & Ide, 2015; Wang & Lei, 2014). This hybrid method combines the flow-dependent background error-covariance matrix with the static error-covariance matrix, and 4D indicates that it can generate a 4D analysis trajectory to allow for the appropriate assimilation of observations that are not taken at the same time. The ensemble input for the hybrid 4D-EnVar utilizes dynamically downscaled forecasts at a 9-km horizontal resolution forced by the 20 member NCEP Global Ensemble Forecast System (GEFS) and the 20 member Canadian Ensemble Forecasts from the Meteorological Service of Canada (MSC), both of which are available at 0.5-degree horizontal grid spacing. Data assimilation has been performed for a 6-h time window centered at 0000 UTC for each AR Recon IOP (Table S1 in Supporting Information S1). The first guess data are the hourly outputs from West-WRF initialized from 1800 UTC on the previous day. Seven separate instances in time are used for the first guess and ensemble input during the 6-h assimilation window to capture the fast-moving nature of ARs. The ensemble input from the aforementioned global models were initialized from 1200 UTC on the previous day, which allows for the growth of perturbations as a reliable representation of flow-dependent forecast errors.

The observations assimilated for each IOP include conventional observations in the PrepBUFR file from the NCEP Global Data Assimilation System (GDAS), remotely sensed refractivity from GNSS Radio Occultation (RO), and atmospheric motion vector winds (AMV, Velden et al., 2005). The observations of temperature, humidity, and horizontal wind from the AR Recon flight-level data and dropsondes are included in the GDAS PrepBUFR file. Note that satellite radiances are not assimilated in this analysis. Similar observational network was used in previous studies to investigate data impact for other campaign programs such as the Mesoscale Predictability Experiment (MPEX, Romine et al., 2016).

Two sets of parallel experiments were performed: (a) WithDROP, which assimilated the PrepBUFR data, that included AR Recon flight-level and dropsonde profiles, GPS RO refractivity, and AMVs, and (b) No-DROP, which assimilated the same observations as WithDROP with the exception of AR Recon data, which are removed from the PrepBUFR data. Assimilation occurs on the 9-km domain only, which will provide the initial conditions for the 3-km domain, and subsequent forecasts are initialized on both the 9-km and 3-km grid from that assimilation procedure. Initial condition and boundary condition data for both experiments can be accessed from Zheng, Delle Monache, Cornuelle, et al. (2021). Extended 144-h forecasts are produced on both the 9-km and 3-km grids beginning at the 0000 UTC initialization for each IOP. Note that cycling of data was not performed in these numerical experiments so that a clean impact from each

individual AR Recon flight can be seen. In other words, any differences in these two parallel simulations are predominantly attributable to the ingestion of the AR Recon data. While cycling allows observations from prior cycles to inform the current analysis, these regional modeling experiments do not assimilate data from the entire global observing system. Therefore, boundary condition errors, particularly errors over the southern boundary, can interact with the signals (i.e., dropsonde impacts) after a few cycles and interfere with the interpretation of signals from the dropsonde and make it hard to separate signals from these model errors. Hodyss and Majumdar (2007) specifically discussed the contamination of data impacts in a global model by initially small mesoscale instabilities. With the non-cycled runs, we are trying to avoid contamination with accumulated boundary errors that might complicate the conclusions of observation impacts.

2.3. Verification Data

To verify the initial conditions and the forecasts of different experiments from the West-WRF/GSI system, the fifth ECMWF Reanalysis (ERA5) product (Hersbach et al., 2020; Simmons et al., 2020) is chosen. IVT and IWV are analyzed to characterize the kinematic and moisture signatures of ARs. Both variables are integrated vertically from the surface to 300 hPa (Cordeira et al., 2013). Additionally, key meteorological variables such as temperature, horizontal winds, and moisture are also investigated.

The comparisons with ERA5 will be presented for forecast validation. The overall verification results with the NCEP GFS Final (FNL) Analysis (NCEP, 2015) are generally consistent with that of ERA5 (not shown). Though ERA5 is not the "truth," the model and assimilation system that was used to generate this product is acknowledged to be the state of the science and has outperformed other reanalysis products overall for other weather phenomena (e.g., Mahto & Mishra, 2019). Note that the ERA5 product assimilated temperature and wind collected during AR Recon in 2018 and 2019. The Stage-IV hourly precipitation product (Lin & Mitchell, 2005) at 4 km horizontal resolution was utilized to validate the precipitation forecast skill over the verification domain (110–125°W, 30–50°N; Figure 1b).

3. A Case Study of Data Impact in IOP7

### 3.1. Synoptic Overview of IOP7

The case we are investigating in detail is an AR that made landfall over the Pacific Northwest from February 3 to 4, 2018. This case was chosen for three reasons: (a) it was a classic landfalling AR that brought precipitation to the Pacific Northwest, where the close relation between a landfalling AR and inland precipitation is well established in the literature (e.g., Rutz et al., 2014); (b) this IOP sampled most of the sensitive targets detected by the NRL adjoint model, whereas in some other cases the targets were too far away for an aircraft to sample or only a limited number (one or two) of aircraft were available; (c) this IOP had one NOAA G-IV and two Air Force C-130 aircraft so that impacts for observations from all three aircraft could be represented. Note that this case is a prototype IOP that represents positive impact IOPs. Results for all IOPs with positive impacts and neutral/negative impacts are summarized and discussed in Sections 4 and 5.

At 1200 UTC February 2 (Figure 2a), a large broad AR was located between a deep Aleutian low centered near 167°W, 43°N and a high pressure system near 131°W, 38°N. The maximum IVT was ~1,100 kg m^{-1} s^{-1} near 146°W, 43°N, with the upper-level North Pacific jet core ~5° longitude west of the IVT core. At this time, the leading edge of this AR reached Vancouver Island (VI). A nearby low-pressure system was located off the coast of British Columbia (BC) with its associated frontal system extending further inland. 12 hours later (Figure 2b), the magnitude of the AR over the primary IVT core had reduced by ~100 kg m^{-1} s^{-1}. In contrast, on the northeastern side, the AR had strengthened from a peak value of 450 to 800 kg m^{-1} s^{-1}, forming a secondary AR core offshore of VI. The leading edge of the AR had a plume of IVT extending further inland over Oregon and Washington. Its associated extratropical cyclone moved to the north of VI with a central pressure of 1,008 hPa. A total of 86 dropsondes were deployed for assimilation at 0000 UTC February 3. The dropsonde flights transected the AR structure six times (Figure 2b). At 1200 UTC February 3 (Figure 2c), the AR continued making landfall and the IVT orientation became more northwesterly directing more moisture further inland from the northeastern Pacific. A short-lived trough of low pressure
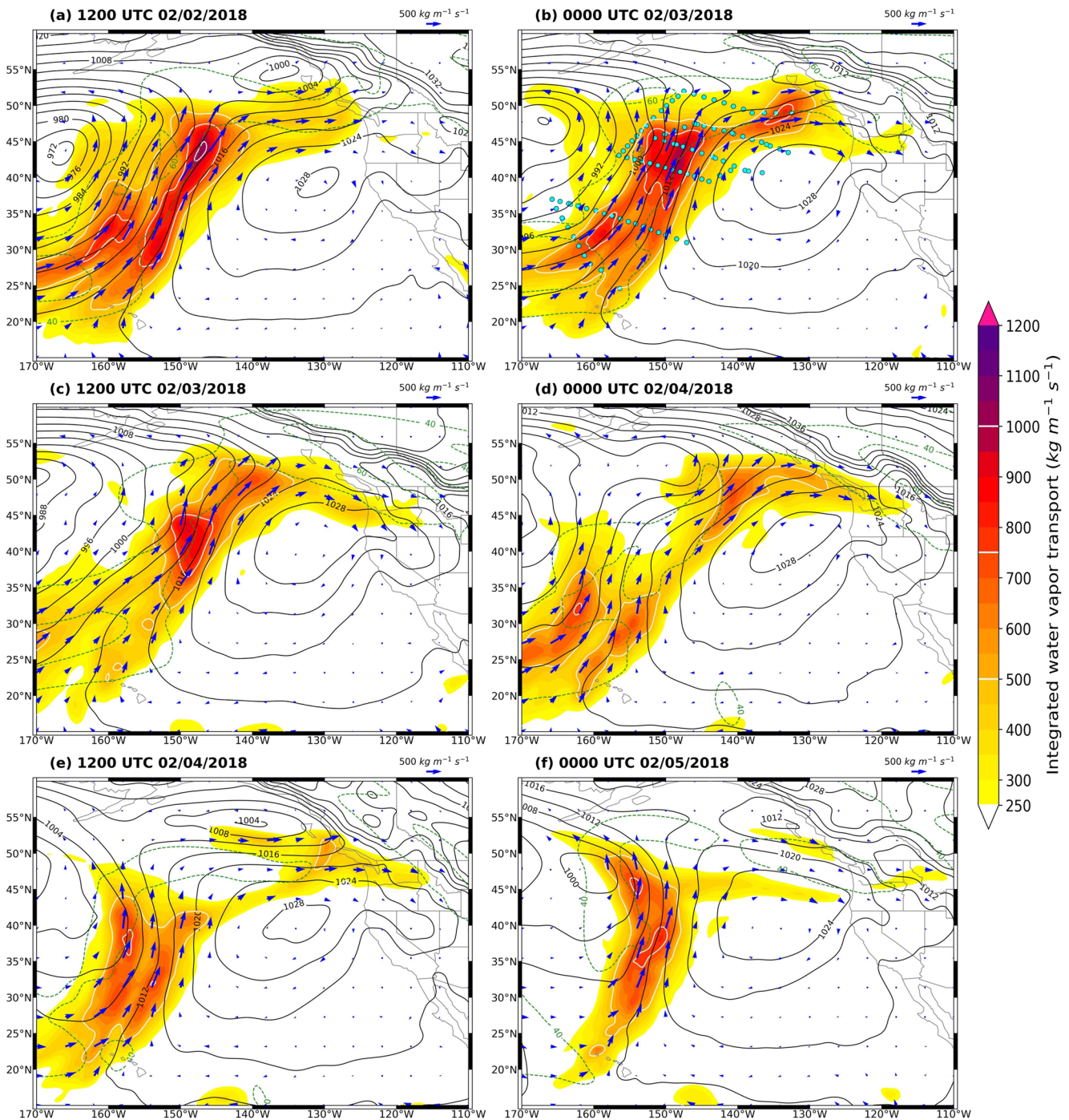
Figure 2. IVT amplitude (shaded, kg m^{-1} s^{-1}) and vectors (blue arrows, kg m^{-1} s^{-1}), MSLP (black contours, hPa), and 300 hPa wind speed (green dashed contours, m s^{-1}). Analyzed period starts from (a) 1200 UTC February 2, 2018 to (f) 0000 UTC February 5, 2018 with a 12-h interval. The analyses are the ERA5 reanalysis data. Cyan filled circles on panel (b) represent the locations of AR Recon dropsondes assimilated in WithDROP during the 6-h time window centered at 0000 UTC February 3, 2018.

formed in the Pacific Northwest, and together with the enhanced IVT, created moisture convergence in the lower troposphere that was favorable for inland precipitation. Secondary frontal waves started to develop downstream of the Aleutian low at 0000 UTC February 4 (Figure 2d). Even though the intensity of the broad AR weakened overall, the leading edge of the AR over the Pacific Northwest remained similar in amplitude and extension. After 12 h (Figure 2e), the AR split into two parts, the main part had a core IVT of ~800 kg

Model Versus Observations
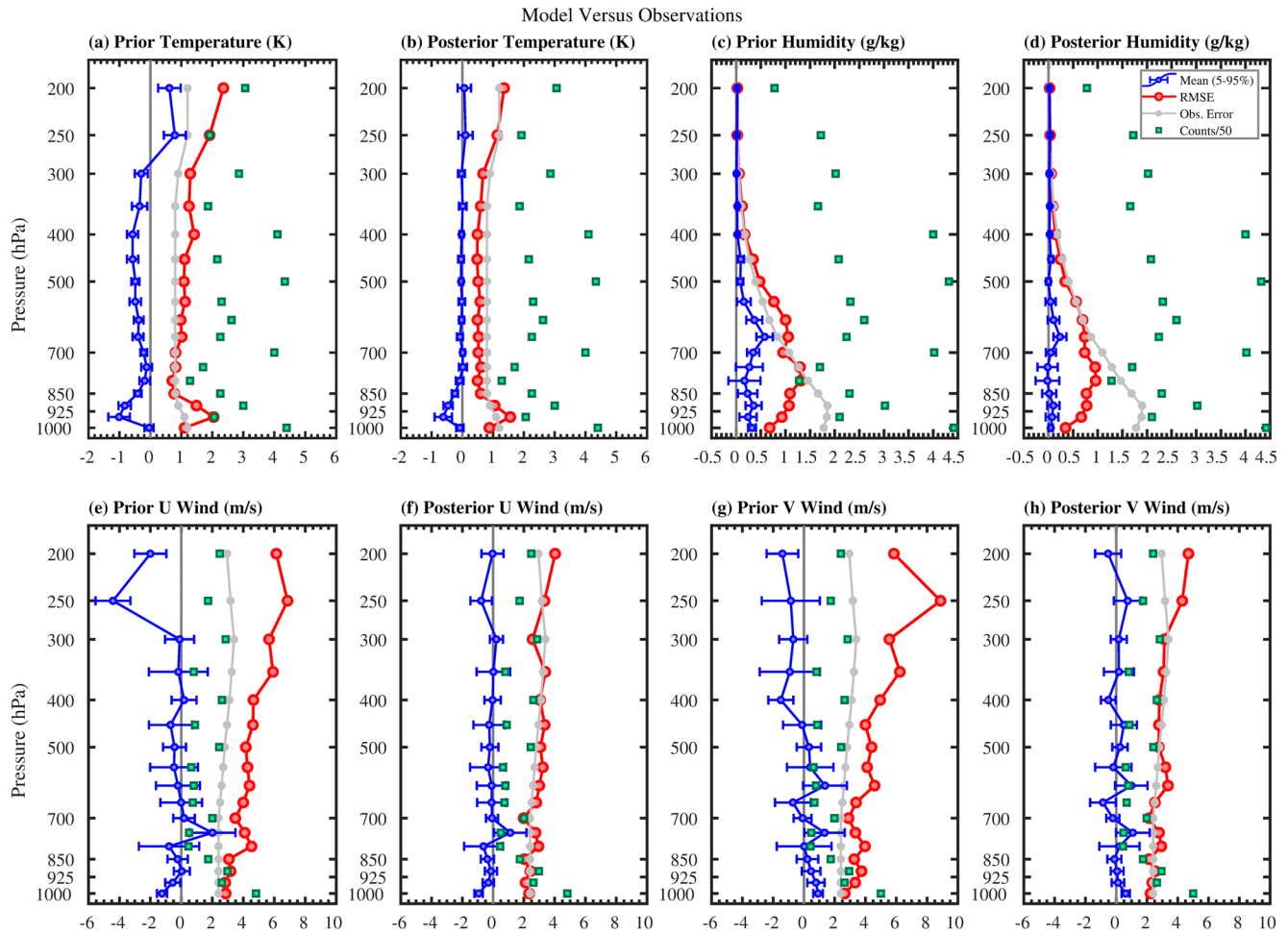
Figure 3. Innovation (Prior) and analysis residual (Posterior) of temperature T (a and b), specific humidity Q (c and d), zonal wind u (e and f) and meridional wind v (g and h). Profiles displayed on each panel are mean difference (blue line with filled circles and horizontal bar), RMSE (red line with filled circles), assimilated observation count divided by 50 (green squares), and observation errors for each variable (gray line with filled circles), respectively. The data assimilation was performed during 2100 UTC February 2 to 0300 UTC February 3, 2018. The dropsonde observations are binned every 50 hPa on the vertical.

$m^{-1} s^{-1}$ near 158°W, 37°N, and the landfalling part had a core IVT of ~550 kg $m^{-1} s^{-1}$ near 130°W, 50°N. At 0000 UTC February 5 (Figure 2f), the landfalling AR dissipated, leaving only one small filament of elevated IVT over eastern Washington and eastern Oregon.

3.2. Dropsonde Impact on Initial Conditions

3.2.1. Innovation and Analysis Increment

The impact of a given observation type depends on the observation errors and the difference between the given observations and the first guess. The latter is often referred as the "innovation." In this study, we defined the "innovation" as the differences between the first guess and an observation to show the model errors. Accordingly, we define "analysis residual" as the difference between the model analysis and an observation. Figure 3 shows the innovation and analysis residual for the dropsonde observations used for this case in the WithDROP experiment. Model state vectors were converted to dropsonde observation space so that the model first guess and analysis can be compared directly to the observations. Compared to the dropsondes, the first guess was colder below the 300 hPa pressure level, with the mean difference of −1K near 950 hPa (Figure 3a); while a warm difference of ~0.5K was observed above the 300 hPa pressure level. The largest root-mean-squared error (RMSE) was also found near 950 hPa and showed greater values in the upper levels than the assigned observation errors, suggesting the dropsondes contained new information for

the model background. After assimilating the observations that included AR Recon data, the temperature mean difference from 200 to 800 hPa was completely removed and the maximum bias near 950 hPa was reduced by 40% (Figure 3b). The RMSE was reduced by ~40%–50% from 200 to 700 hPa and ~15%–30% below 700 hPa pressure level (Figure 3b). The specific humidity showed that the first guess was moister than the observation below the 500 hPa pressure level and the assimilation process removed most of the bias except for the layer at 600–700 hPa where the maximum difference existed (Figures 3c and 3d). The first guess RMSE for humidity peaked from 550 to 850 hPa and decreased by ~20%–40% after the assimilation (Figures 3c and 3d). After assimilation, over 75% of the wind differences (Figures 3e–3h) were removed except for the southeasterly difference at the lowest levels (950–1,000 hPa). The RMSE of wind was maximized near 800 hPa in the lower troposphere and near 250 hPa in the upper troposphere (Figures 3e and 3g). It is noteworthy that the RMSE was much larger than the observation error throughout the troposphere for horizontal winds, indicating that there was valuable new information contained in the wind observations from dropsondes. The largest RMSE reduction attributable to assimilation was 40%–60% from 200 to 500 hPa for horizontal winds (Figures 3f and 3h). The lower troposphere exhibited 15%–30% reduction of RMSE for winds below 500 hPa (Figures 3f and 3h).

### 3.2.2. Impact on Thermodynamic Variables and IVT

To examine how the dropsonde observations modify the initial conditions, we have analyzed the difference between WithDROP and NoDROP. Figure 4 shows a representative cross-section along a flight track made from the cold side of the AR to the warm side to sample the AR sector (Cobb et al., 2021; Ralph et al., 2017; Zheng, Delle Monache, Wu, et al., 2021). The NoDROP experiment showed a moist error overall with respect to ERA5 from 700 to 900 hPa, specifically near regions with large vertical and horizontal humidity gradients such as 150°–152°W and 145°–146°W (Figure 4a). Note that ERA5 and the model output have different resolutions, which might have a slight impact on how gradients are represented at each grid. The lowest level (~975 hPa) in NoDROP showed weak dry bias west of 147°W and moist bias east of it. The assimilation of dropsonde data corrected the initial conditions by reducing the moist error overall from 600 to 975 hPa (Figures 4b and 4c). For example, the correction near 151.5°W at 850 hPa was ~60% of the initial moist error in NoDROP and near 145°W at 600 hPa was ~50% (Figures 4a and 4c). However, the initial weak moist error near 148.5°W from 500 to 600 hPa, where the dry intrusion exists, was amplified in WithDROP by 40%–50% (Figures 4a and 4b).

The NoDROP experiment overall showed weaker wind speed than the ERA5 near the upper-level jet (ULJ) core and the lower-level jets (LLJs), but it showed stronger wind between 145°W and 150°W near 275 hPa (Figure 4d). The assimilation of dropsonde data reduced the positive wind error right below the ULJ core by ~30% and the positive error near the LLJs by ~20% (Figures 4e and 4f). However, the stronger wind error near 275 hPa was not significantly modified (Figures 4e and 4f). The most striking impact from dropsonde data was seen in temperature (Figures 4g–4i). A coherent region of sloping cold bias of 1–4K in the NoDROP run prevailed along the path from 275 to 400 hPa while a region of warm error of a comparable amplitude was just above it at 200 hPa (Figure 4g). The cold error was also seen in the lower troposphere near 600 and 900 hPa (Figure 4g). The assimilation of dropsondes corrected ~50% of the bias in the upper level from near 200 to 500 hPa, ~30%–70% of the bias from 550 to 700 hPa, and ~20%–35% of the bias from 750 to 925 hPa.

In addition to the standard meteorological variables, the layer IVT cross-section was also made along the same path to represent the dynamical and moisture structure of the AR. Figure 5a shows the layer IVT difference between NoDROP and ERA5. For this case in the ERA5 reanalysis, 20 kg m^{-1} s^{-1} of layer IVT represents the threshold for the 20th percentile of the AR moisture and wind (Figure 5a) and is used to define the limits of the AR. The layer IVT was highest in one branch of the LLJ around 148.5°W at 800 hPa and a secondary IVT LLJ was near 152.5°W. The NoDROP run underestimated IVT in both LLJs around 800 hPa (Figure 5a) and overestimated the layer IVT between the two jets near 151°W at the same level. The largest overestimation for layer IVT was over the southern flank of the AR near 750 hPa and above the stronger jet in the AR core between 500 and 600 hPa (Figure 5a). The dropsonde data corrected both the underestimated IVT near the AR core and the overestimated IVT above the southern flank by ~20%–30% (Figures 5b and 5c).

The total IVT amplitude in NoDROP was generally overestimated within the domain except near the areas with magnitude greater than 750 kg m^{-1} s^{-1} (Figure 5d). The dropsondes reduced the IVT amplitude along
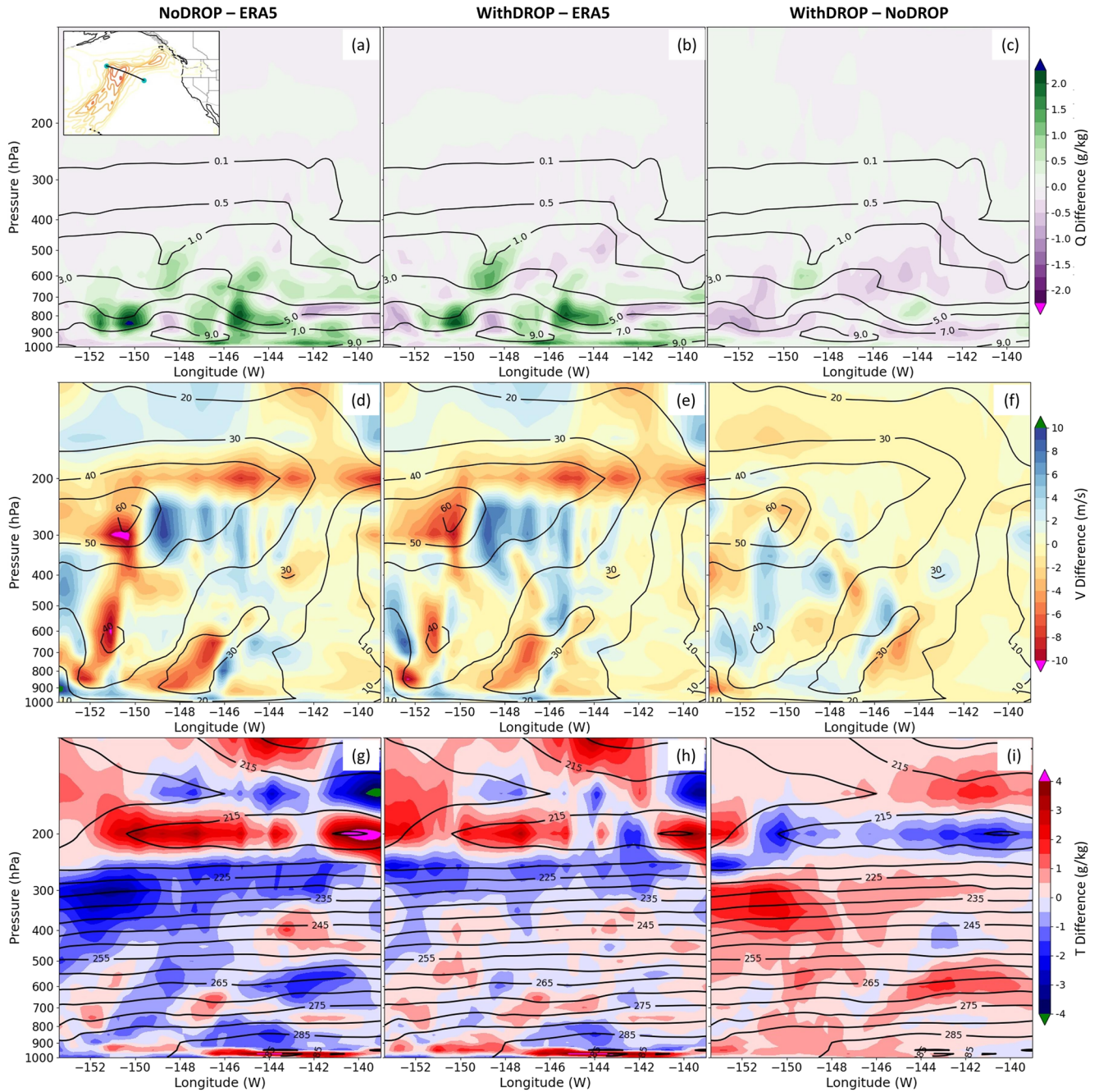
Figure 4. Vertical cross-section along the path from (153.4°W, 46.5°N) to (139.0°W, 41.0°N) for (a–c) specific humidity (g kg⁻¹), (d–f) wind speed (m s⁻¹), and (g–i) temperature (K). Black contours for each row represent ERA5 specific humidity, wind speed, and temperature, respectively. The shading on the left, middle, and right column represents NoDROP minus ERA5, WithDROP minus ERA5, and WithDROP minus NoDROP, respectively. The analysis is valid at 0000 UTC February 3, 2018.
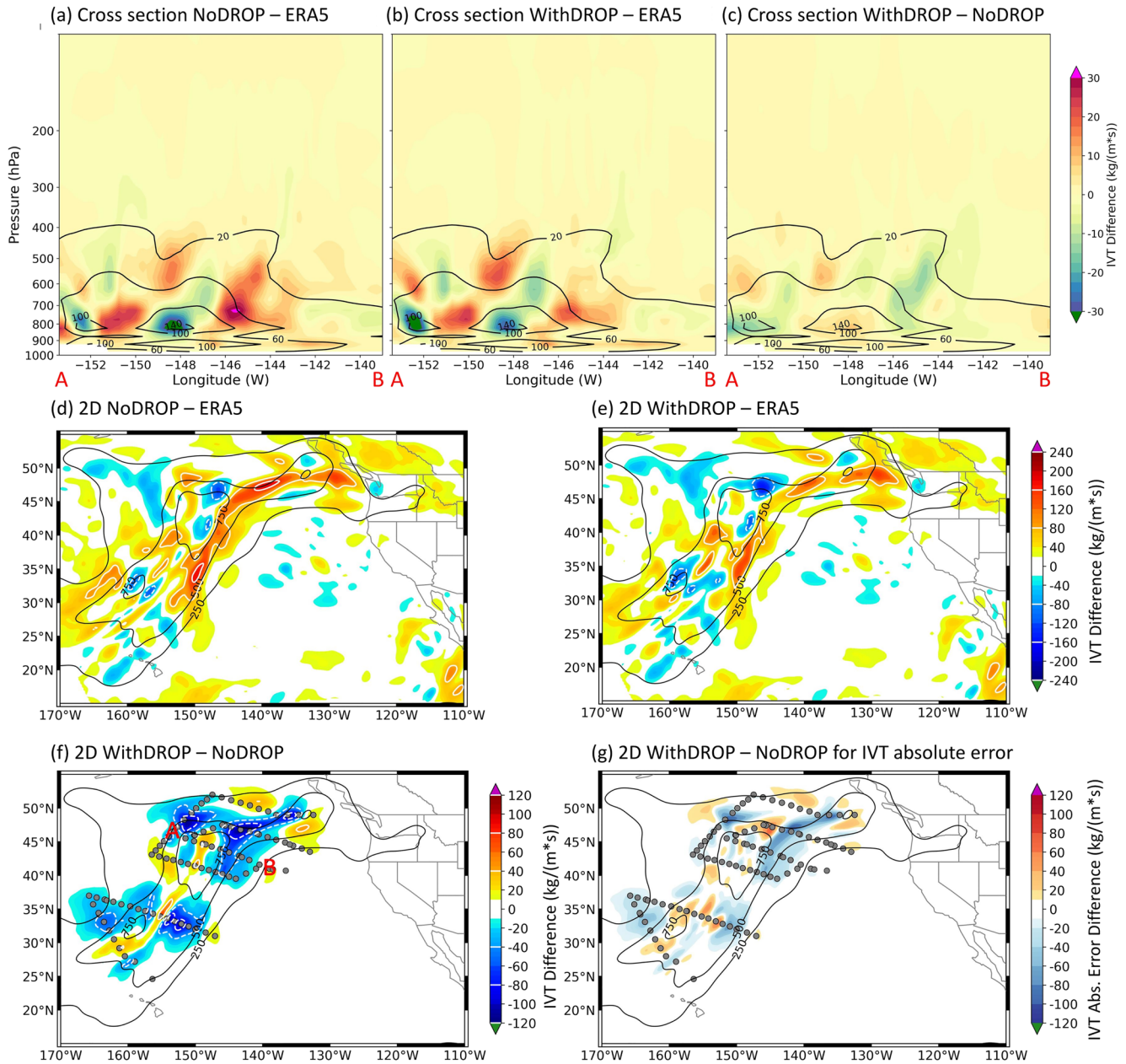
Figure 5. Dropsonde impact on layer IVT and total IVT. (a) Difference (shaded, [kg m^{-1} s^{-1}]) in layer IVT between NoDROP and ERA5 along the same path shown in Figure 4 from A (153.4°W, 46.5°N) to B (139.0°W, 41.0°N); (b) same as (a) but for WithDROP; (c) same as (a) but for the difference between WithDROP and NoDROP. The black contour is for the ERA5 analyzed layer IVT. (d) 2-D distribution of the difference for total IVT amplitude (shaded with white contours, [kg m^{-1} s^{-1}]) between NoDROP and ERA5; (e) same as (d) but for WithDROP; (f) same as (d) but for the difference between WithDROP and NoDROP overlayed with assimilated dropsonde locations (gray filled circles); (g) difference (shaded, [kg m^{-1} s^{-1}]) in IVT absolute error between WithDROP and NoDROP. The absolute error is the absolute value of a model run minus ERA5. The black contours in (d, e, f, g) are for the ERA5 IVT starting from 250 kg m^{-1} s^{-1}. The analysis is valid at 0000 UTC February 3, 2018.

the AR flanks and increased the IVT value near the AR core (Figures 5e and 5f). The absolute error predominantly decreased by ~30%–75% along both northern and southern flanks of the AR, attributable to the assimilation of AR Recon dropsondes (Figure 5g).

3.3. Impact on Forecasts

3.3.1. Impact on Temperature and IVT

The forecast of this case was initialized at 0000 UTC February 3, 2018. During the AR Recon operations, the major targeting guidance for this IOP was to improve the forecasts of the landfalling AR and its associated precipitation over the Pacific Northwest from 1200 UTC February 3 to 1200 UTC February 4 based on the areas of highest sensitivity identified using the NRL adjoint sensitivity tool. Therefore, we assessed the impact of the dropsondes on the 24-h forecast for IVT valid at 0000 UTC February 4, which is in the middle of the targeted precipitation window. At the initial time, the 300-hPa temperature in the NoDROP experiment showed a cold bias above the AR region (Figures 4c and 6a) that seemed to be associated with the uncertainty in simulating the upper-level clouds (Figure S1 in Supporting Information S2). The dropsonde data helped reduce the cold error by ~50% near 151°W, 45°N and over the eastern part of the AR (Figures 6c and 6e). The dropsondes from the western flight track created a dipole structure of temperature anomaly (Figure 6e). In the 24-h forecast, the cold error in NoDROP was amplified and moved eastward close to the North American coast (Figure 6b). The WithDROP run showed ~30%–50% less cold error than the NoDROP run (Figures 6d and 6f). Over Washington, the cold error of ~1.5K was reduced to ~0.7K (Figures 6b, 6d, and 6f), where the precipitation impact was also clearly observed (see next section).

The initial temperature difference between the NoDROP experiment and the ERA5 at 850-hPa was not as large as that at 300 hPa (Figure S2a in Supporting Information S2) over the oceanic AR region. A cold error of ~0.5–2K was seen in NoDROP near 152°W, 48°N and 142°W, 43°N (Figure S2a in Supporting Information S2) where the WithDROP run showed 30%–50% less error (Figure S2c in Supporting Information S2). After 24 h, the NoDROP run continued to show a cold error over the Northeastern Pacific Ocean (Figure S2b in Supporting Information S2). Note that a warm error extended from VI to western Washington over the landfalling part of the AR shown in Figure 2c. The WithDROP run had half of this warm error reduced due to the assimilation of dropsondes at the initial time (Figures S2d and S2f in Supporting Information S2), leading to less warm air convergence over western Washington and increased stability. Consequently, this contributed to reduce precipitation in the WithDROP run around this forecast time.

For IVT, the NoDROP run showed large forecast errors around both cold and warm sides of the landfalling AR (Figure 7a). The IVT amplitude over the inland penetration of the AR was over-forecasted in NoDROP (Figures 7a and S3a in Supporting Information S2). The WithDROP run had ~50% less error over VI and the west of Washington (Figures 7b, 7c, S3a, and S3b in Supporting Information S2). The WithDROP run also reduced the negative bias near 147°W, 52°N by ~25% and the overprediction errors near 143°W, 46°N by ~50% (Figures 7b–7d), where a cold front can be identified (not shown).

#### 3.3.2. Impact on Precipitation

The maximum precipitation associated with the landfalling AR occurred between 1200 UTC February 3 and 1200 UTC February 4 over northwestern Washington in both the Stage-IV data and the forecasts. The maximum precipitation in Stage-IV was about 65 mm near 121.6°W, 47.7°N (Figure 8a). A second precipitation center was further south near 121.9°W, 46.9°N (Figure 8a). The NoDROP experiment generally overpredicted the amount of precipitation within the domain, and displaced the center of the heaviest precipitation seen in Stage-IV ~100 km southward, creating a peak in the precipitation difference (shown in pink in Figure 8d) with overestimation error of ~40 mm near 122.0°W, 47.0°N (Figures 8b and 8d). The precipitation in the WithDROP experiment was more realistic and closer to Stage-IV than that in NoDROP in both intensity and distribution (Figures 8c and 8e). The overprediction error was almost completely corrected near the peak in precipitation, which can be attributed to the assimilation of dropsonde data. The RMSE averaged over the inland precipitation common domain (Figure 1b) was reduced by 12.8% during 12–36 h.

The overprediction of precipitation in the NoDROP run was associated with the colder cloud tops at the upper levels (Figure 6b) and the enhanced water vapor transport from offshore over the leading edge of the
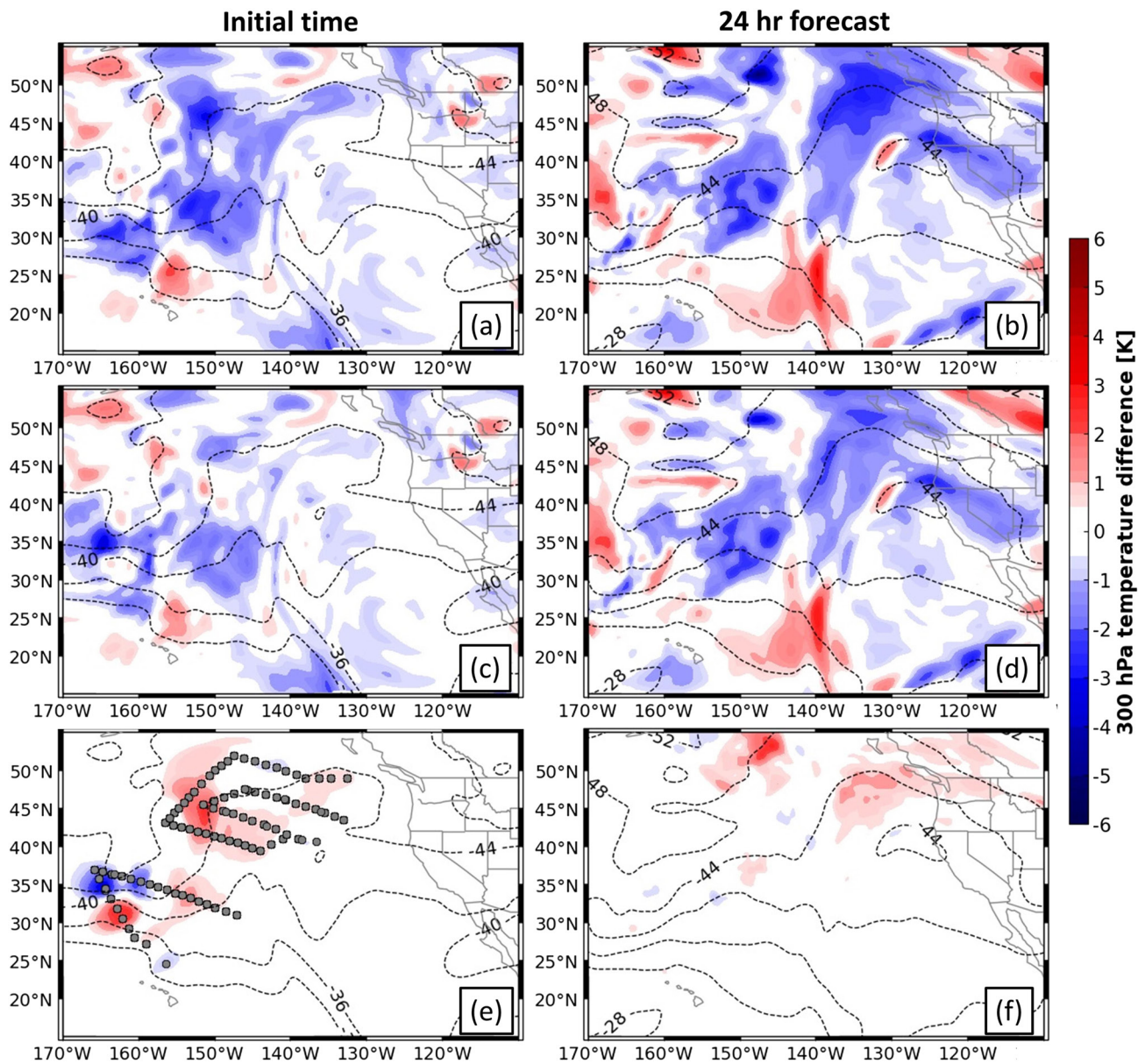
Figure 6. Dropsonde impact on 300 hPa temperature. The left column (panels a, c, e) is for the initialization time of 0000 UTC February 3, 2018 and the right column (panels b, d, f) is for the 24-h forecast. (a and b) are for the NoDROP 300-hPa temperature error (shaded, [°C]); (c and d) are same as (a and b) but for the WithDROP run; (e and f) are the difference between WithDROP and NoDROP. The dashed black contours on the left and right columns are the ERA5 300-hPa temperature (units: [°C]) at 0000 UTC February 3 and 4, respectively.

AR (Figures 7a and S3a in Supporting Information S2) that can be traced back to initial condition errors (Figures 6a and S2c in Supporting Information S2). Dropsonde observations reduced the cold bias in the cloud top and corrected the IVT overestimation at the initial time, leading to a significant improvement in the regional precipitation forecast. Zheng, Delle Monache, Wu, et al. (2021) showed that the operational observation system has difficulty in measuring the lower to middle troposphere within an AR, indicating that a greater impact from dropsondes would be expected in the lower to middle troposphere. The analysis here demonstrated that these data can also significantly correct the upper-level temperature error specifically near the cloud tops over an AR, and this correction is important for the accurate prediction of precipitation as well. Hanna et al. (2008) found that lower cloud-top temperature is associated with increasing rainfall intensity. In this case, the cold bias of cloud-top temperature, extending inland from offshore the Pacific coast, partly caused an overestimation of the precipitation over western Washington, and after reducing
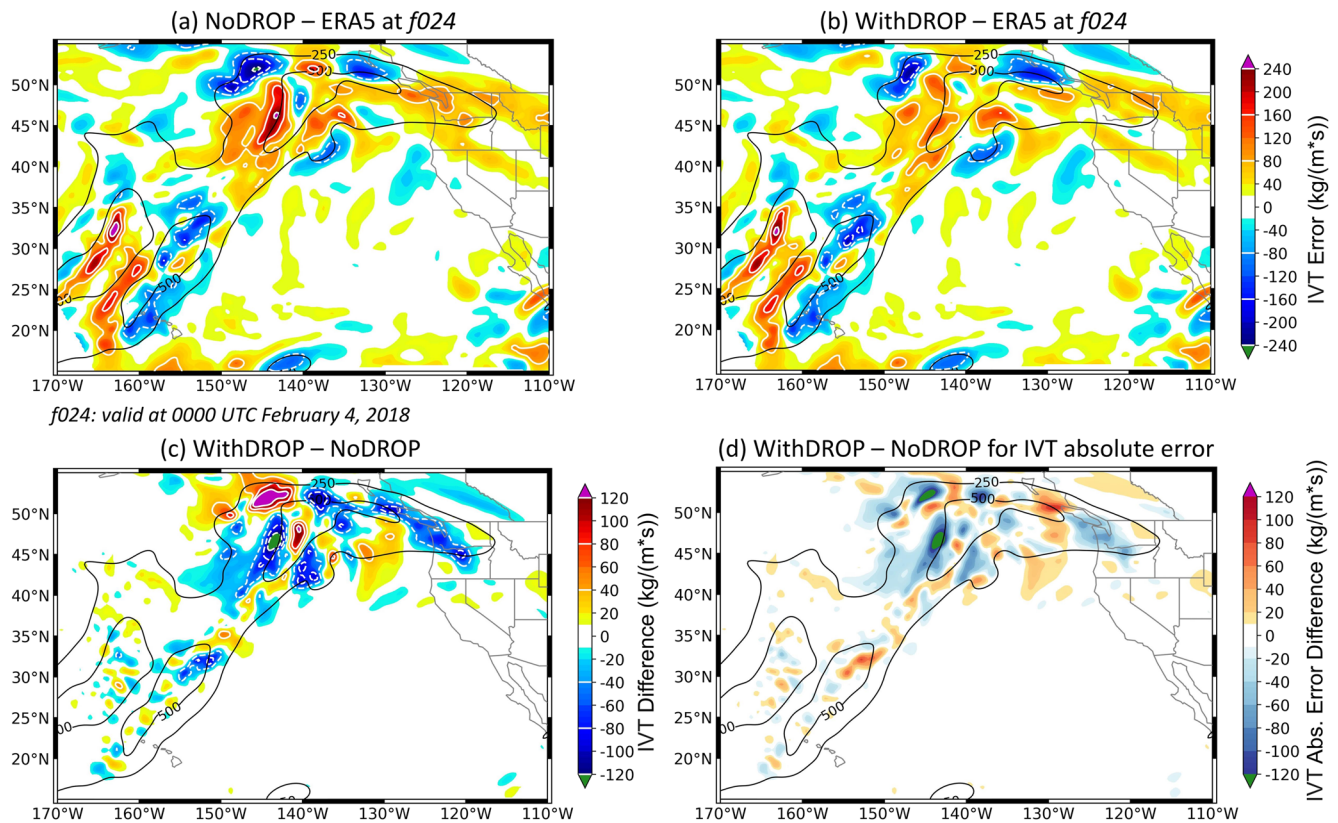
Figure 7. Impact on IVT forecast at 24 h. (a) The difference for IVT amplitude (shaded with white contours, [kg m^{-1} s^{-1}]) between NoDROP and ERA5; (b) same as (a) but for WithDROP; (c) same as (a) but for the difference between WithDROP and NoDROP; (d) same as (c) but for the IVT absolute error. The black contours for each panel are the IVT amplitude in ERA5. See Figure S3 in Supporting Information S2 for the zoomed-in plots over Pacific Northwest.

the cold bias by assimilating the dropsonde data, the overpredicted precipitation was reduced by ~50% near Mount Rainier.

4. Data Impact Results for All 15 IOPs

4.1. Impact on IVT and IWV

To systematically investigate the data impact of AR Recon data, we computed the differences in forecast skill of water vapor transport and inland precipitation for all IOPs between WithDROP and NoDROP experiments. At the initial time, all of the 15 IOPs showed improved IVT representation based on the RMSE reduction metrics with respect to ERA5 reanalysis with the median value ~3.8% and the 75th percentile improvement was ~8.2% (Figure 9a). The median value for the RMSE reduction was a maximum of ~4.8% for the 12-h forecast and continued to be positive but decreased through 72 h. The RMSE reduction was overall neutral from 78 to 102 h, with the 90–96 h forecast showing small negative median values and more degraded IOPs than improved IOPs. Beyond 108 h, the RMSE reduction metrics were positive, with over two-thirds of the IOPs improving from 126 to 144 h. The data impact on the spatial pattern of IVT, represented by the correlation between NoDROP error and the difference of NoDROP and WithDROP, was positive across all the 15 IOPs at all forecast times (Figure 9b), indicating that the dropsondes were improving the initial condition in all IOPs, even if the improvement size was too small to change the RMSE by much in some IOPs. The error-difference correlation was maximized during the 6–12 h forecasts with the value between 0.2 and 0.5. The trend of impact based on the error-difference correlation was generally consistent with that for the RMSE reduction: with a decreasing positive impact from the initial time to ~60 h, an impact close to neutral from 90 to 102 h, and a positive impact again from 114 to 144 h (Figure 9b). Figures 9a and 9b showed that the impact on IVT did not linearly decrease with forecast time. Instead, a second positive impact emerged
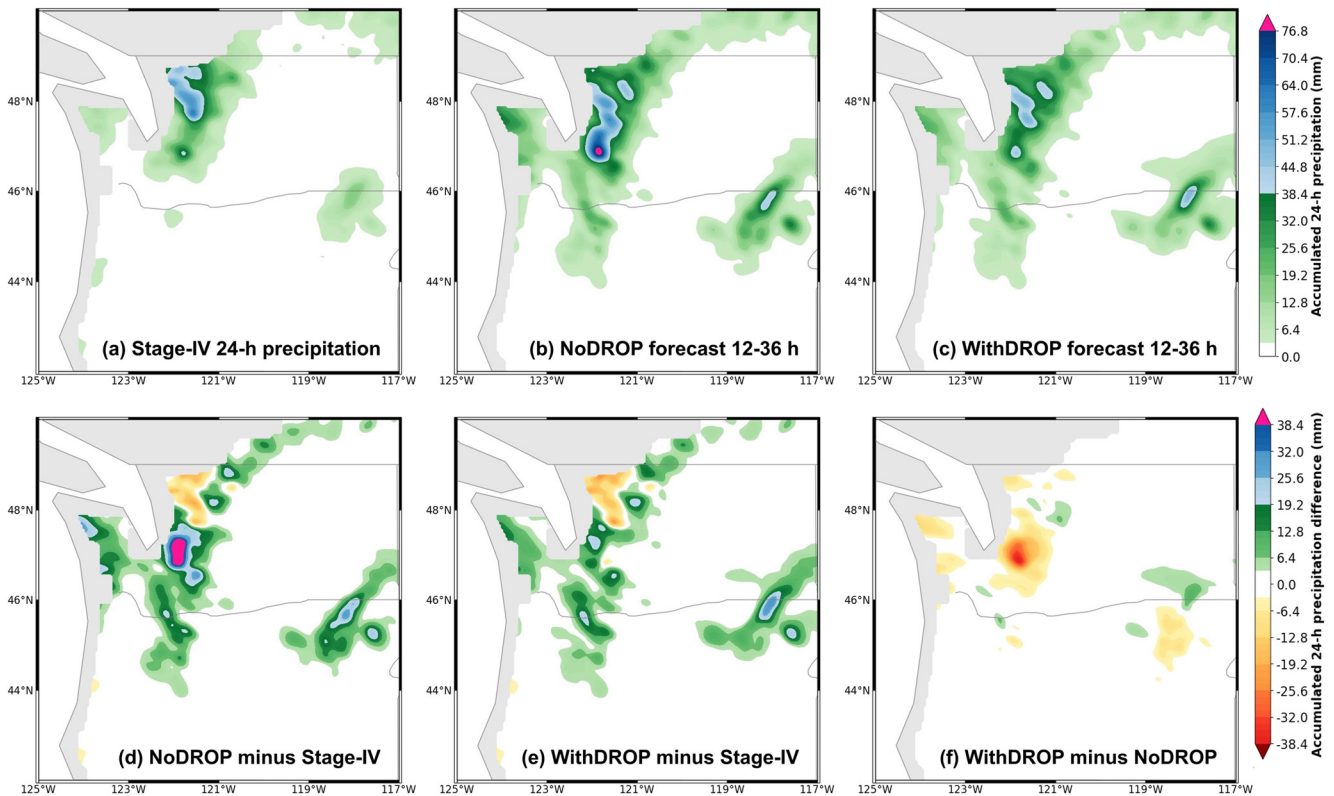
Figure 8. The 24-h accumulated precipitation (mm) from 1200 UTC 02/03 to 1200 UTC 02/04, 2018 for (a) Stage-IV, (b) the NoDROP 12–36 h forecast, and (c) the WithDROP 12–36 h forecast. Panels (d–f) are the 24-h accumulated precipitation difference (mm) between (d) NoDROP and Stage-IV, (e) WithDROP and Stage-IV, and (f) WithDROP and NoDROP, respectively. The model runs are initialized at 0000 UTC February 3, 2018.

beyond day 4. Looking at each of the 15 IOPs (Figure 9c), 12 out of the 15 IOPs showed generally improved forecasts using RMSE with 11 of them having positive values in two-thirds of lead times. IOPs 3, 7, and 14 showed negative median values, however, 43% of the forecast lead times for these three cases still show positive RMSE impact (Figure 9c). The error-difference correlation shows positive values for all 25 lead times with the IOP-based groupings (Figure 9d).

For IWV, the median values across the 15 IOPs for all lead times were positive (Figure 10a). This is different from the IVT validation, which only showed a positive impact in 12 out of the 15 IOPs throughout the 6-day forecasts (Figures 9c and 9d). Since IVT is a combination of humidity and wind while IWV is a pure humidity variable, improvements at more lead times in IWV indicate that the dropsondes may have a larger positive impact on humidity than on wind for 2–4-day forecast. This can be explained by the results found by Zheng, Delle Monache, Wu, et al. (2021), which demonstrates that AR Recon dropsondes provide a higher percentage of available direct humidity observations than that wind observations. The significant data gaps for direct humidity set up the dropsonde to play a more important role in correcting the errors and improving the humidity forecasts. The spatial error-difference correlations for IWV are positive at all time steps for all the IOPs, again suggesting the improvement of spatial pattern error is more robust than the intensity error for IWV (Figure 10).

4.2. Impact on Precipitation Forecasts

One important goal for assimilating AR Recon data is to improve the precipitation forecast over the US West Coast. Figure 11 shows the data impact on 24-h accumulated precipitation for all IOPs over the common inland precipitation verification domain (Figure 1b) with Stage-IV precipitation as the ground truth for RMSE. For the first 24 h, improvement was found in the precipitation RMSE over the domain in 13 out of 15 IOPs with the median value of 3.2% (Figure 11a). Positive impacts on IVT/IWV may not always produce
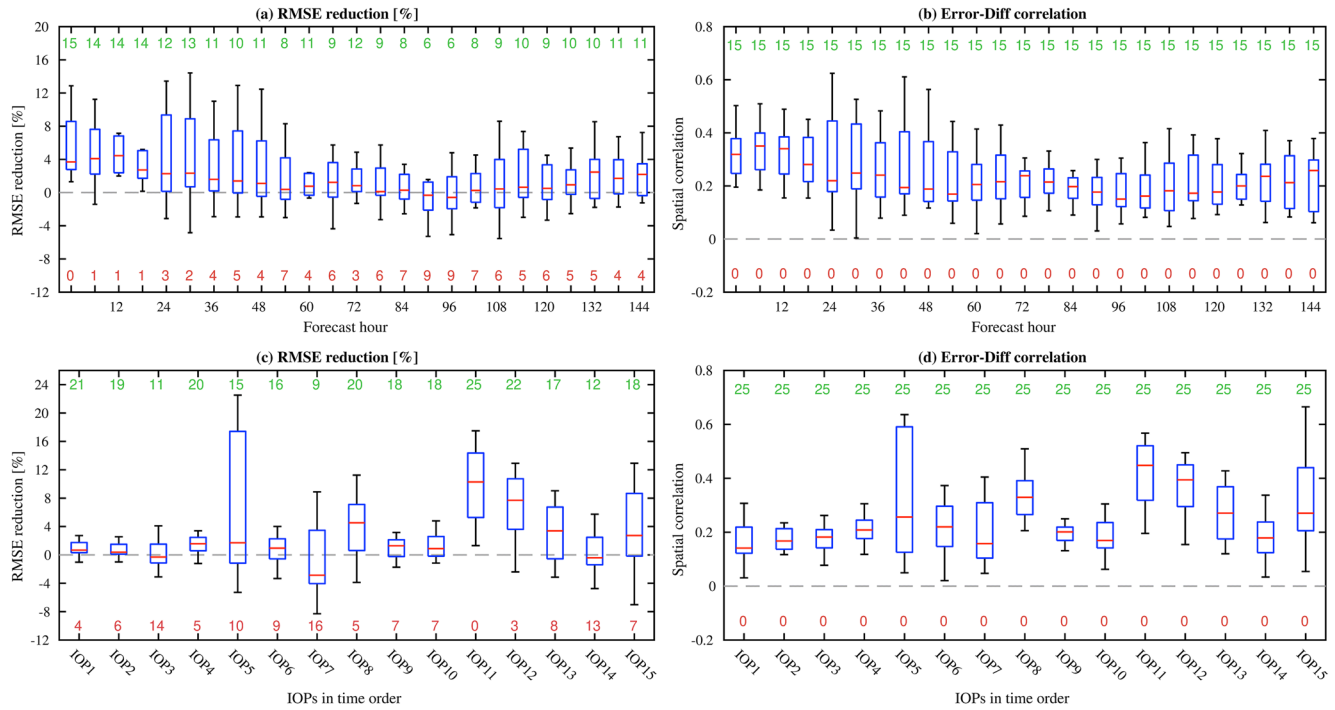
Figure 9. Boxplot of RMSE (a and c) and error-difference correlation (b and d) for IVT w.r.t. the ERA5 data over the common verification area (105°–165°W, 15°–60°N) based on the forecast hour (a and b) and the individual IOPs (c and d). The box plot graphically depicts the range of values for the 15 IOPs in (a and b) and the 25 lead times in (c and d) as follows: the top and bottom edges of the box indicate the top and bottom quartiles, the centerline in the box denotes the median and the whiskers at the top and bottom extend to the most extreme data points, which are no more than 1.5 times the interquartile range from the box. The green and red numbers on each panel represent improved IOPs or lead times and degraded IOPs or lead times, respectively.
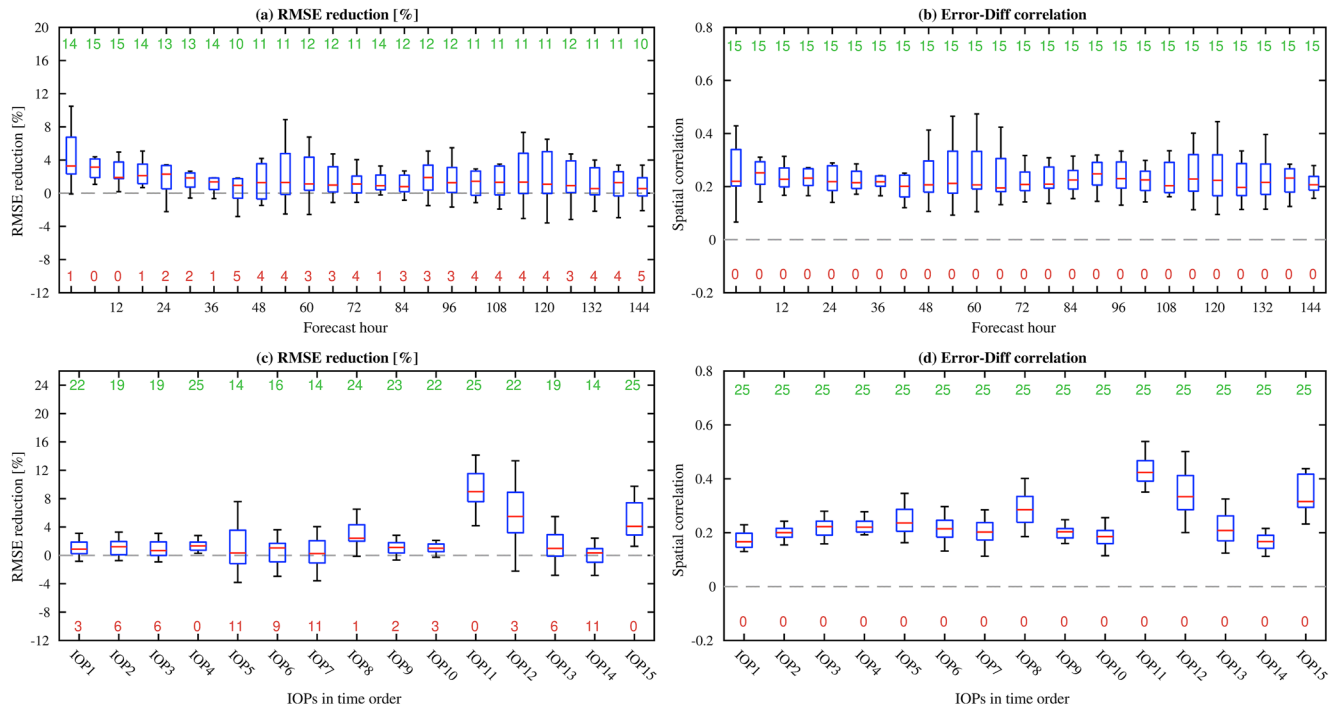
**Figure 10.** Same as Figure 9 but for the IWV forecasts.
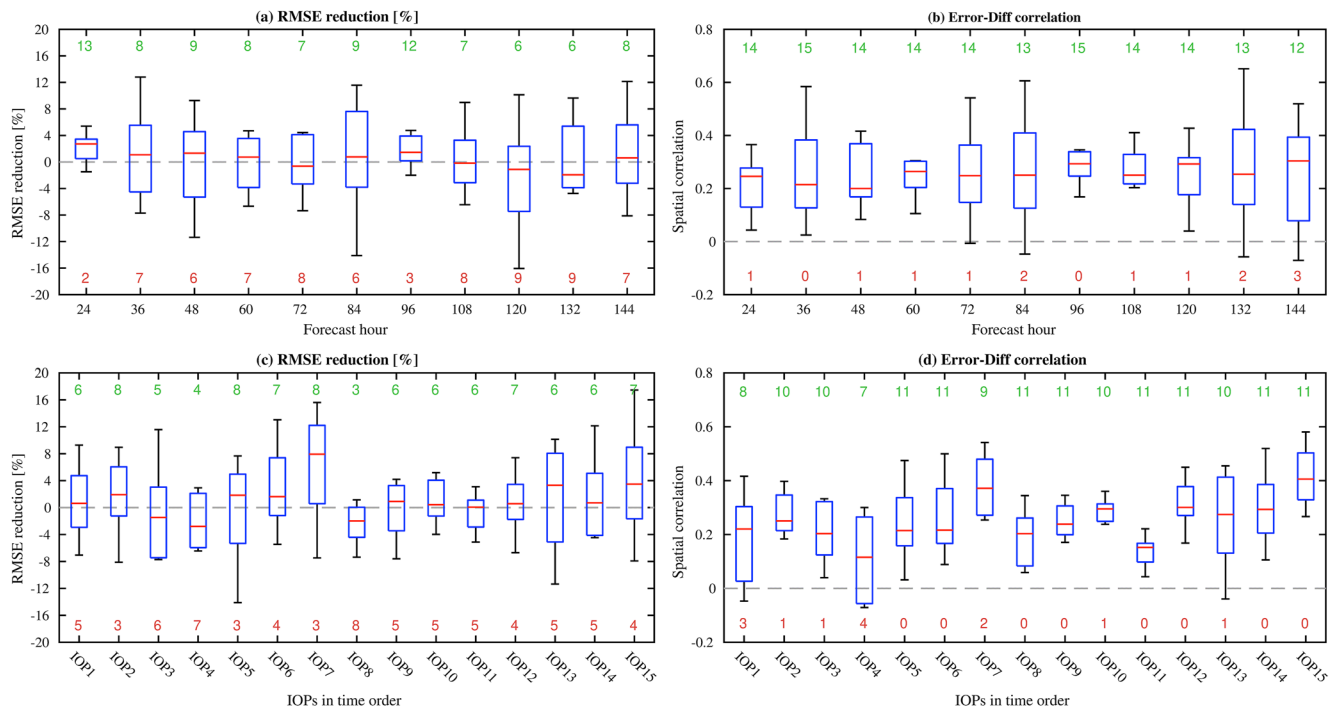
Figure 11. RMSE and error-difference correlation for common area verification for precipitation based on the forecast hour (a and b) or IOPs (c and d).

positive impacts on precipitation. Other factors, such as model microphysics, can be important. Beyond 24 h, the results for precipitation are more neutral than for IVT and IWV. For instance, 8 out of the 15 IOPs showed improved precipitation skill for the forecast ending at 36 h (Figure 11a) while 13–14 IOPs showed improved skill for IWV forecast from 24 to 36 h (Figure 10a). The forecasts ending at 96 h show a reemergence in positive impact. The median values were positive for RMSE improvement in 7 out of the 11 forecast lead times while negative values are seen in two lead times ending at 120 and 132 h (Figure 11a). For the error-difference correlation, the median value was positive for all the lead times, demonstrating that the improvement of the spatial correlation was greater than the RMSE improvement (Figure 11b) with the impact being consistently positive for both metrics. The median value was positive in 11 out of 15 IOPs using RMSE metrics while 3 IOPs showed negative values (Figure 11c). Although positive median values are observed in the majority of cases, it is noteworthy that negative values exist in some forecast times for all the IOPs. The spatial correlation generally showed improved skill at more lead times than that of the RMSE metric based on the IOP grouping (Figure 11d).

The Fractions Skill Score (FSS), a neighborhood method, was also employed to objectively assess the spatial skill at different scales of precipitation forecasts for comparing the NoDROP and WithDROP experiments. Unlike traditional methods which only compare forecast precipitation to observed precipitation in the same grid box, neighborhood methods match forecasted precipitation that is nearby the grid point of observed precipitation (Ebert, 2009; Roberts & Lean, 2008; Schwartz et al., 2009). The calculations of FSS require that the predicted and observed precipitation be on the same grid. Specified thresholds (P, e.g., $P = 10$ mm) are selected to define the precipitation events and a radius of influence (R, e.g., $R = 20$ km) is chosen. FSS is computed for both NoDROP and WithDROP outputs with Stage-IV precipitation as the observed precipitation fields. The forecast is considered perfect if FSS is 1 and no skill if FSS is zero.

Using a 5-mm minimum precipitation threshold, the FSS values from the NoDROP and WithDROP runs are very close in the 24-h forecast (Figure 12a). However, a small (<5%) but significant positive impact was seen in 11 out of the 15 IOPs. Only three IOPs showed a very small (<1%) degraded impact. The IOP3 showed a very small FSS because it did not rain much over the region and the sample size is too small to compute a meaningful FSS value. Using a 10 mm threshold, the FSSs in both runs are smaller than that for 5 mm threshold (Figure 12c). IOPs 7, 9, and 15 show more FSS improvement in WithDROP than in
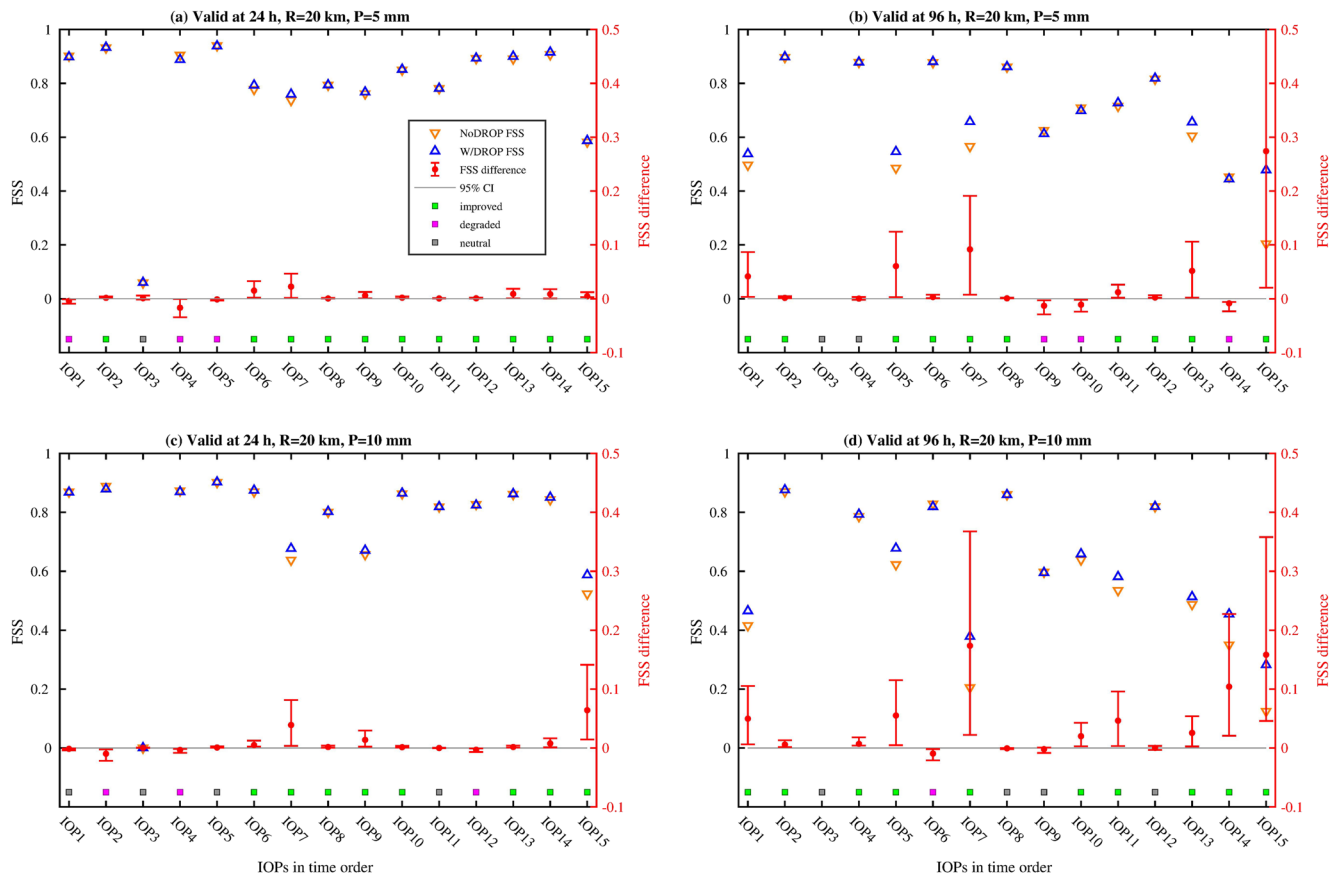
Figure 12. FSS of precipitation forecasts in NoDROP and WithDROP and their difference for each of the 15 IOPs. (a) FSS of NoDROP (orange upward-pointing triangle, left axis) and WithDROP (blue downward-pointing triangle, left axis) with precipitation threshold P of 5 mm and radius of influence R of 20 km for the accumulated 24-h precipitation ending at 24-h forecast time, and the FSS difference between WithDROP and NoDROP (red filled circles and vertical bars, right axis); (b) same as (a) but for the precipitation ending at 96 h; (c) same as (a) but for $P = 10$ mm; (d) same as (b) but for $P = 10$ mm. Horizontal gray dashed lines are the "zero lines" for both y-axes on each panel. The vertical bar on the red line represent the bounds of the 95% bootstrap confidence interval (CI) based on FSS differences between WithDROP and NoDROP. For the differences between WithDROP and NoDROP to be statistically significant, the bounds of 95% CI must not contain zero, otherwise, the differences are insignificant. Green, magenta, and gray filled squares are the statistically significantly improved, significantly degraded, and neutral IOPs, respectively.

NoDROP (Figure 12c). FSS for the accumulated precipitation from 72 to 96 h was also investigated. With a 5-mm threshold, NoDROP showed an overall decrease in skill when compared to the 0–24 h forecast, particularly for IOPs 1, 5, 7, 9, 13–15 (Figure 12b). However, FSS was improved by ∼10%–60% with dropsondes assimilated in IOPs 1, 5, 7, 13, and 15 (Figure 12b), indicating that dropsondes are particularly important in poorly forecasted AR cases. With the 10 mm threshold, FSS in NoDROP decreased significantly when compared to that with a 5-mm threshold in IOPs 7, 11, and 14–15 (Figure 12d). The difference in FSS between WithDROP and NoDROP for precipitation with threshold greater than 10 mm shows an even greater positive impact from dropsondes than was seen using a 5-mm threshold, specifically in IOPs 7, 10–11, and 14 (Figure 12d).

The FSS results highlight that dropsondes are more important when the precipitation forecast has poorer skills specifically with higher precipitation threshold. Note that for all FSS shown for the "degraded" cases, the differences in FSS between WithDROP and NoDROP are typically negligible (i.e., less than 2% of the NoDROP FSS).

5. Discussion

5.1. Impact of Consecutive Flights and/or Multiple Flight Paths

Results from Section 4 demonstrated an overall positive impact on AR-specified variables and precipitation from assimilating AR Recon dropsonde data, even though the positive impact is not uniform. The dependence of dropsonde impact on the weather situation and the lead time is consistent with the conclusions from the literature (e.g., Langland et al., 1999; Romine et al., 2016; Stone et al., 2020). Table S1 in Supporting Information S1 summarizes the synoptic conditions for each IOP and the overall impact on IVT and precipitation.

We found that during a sequence of consecutive flights, which are defined as two IOPs less than 3 days apart, the positive impact on precipitation over the US West Coast was larger in the subsequent missions and was maximized in the last mission of a sequence (IOPs 2, 7, 12, and 15 in Figures 11c, 11d, 12c, and 12d, Table S1 in Supporting Information S1). In contrast, the first missions in a sequence had a relatively larger impact on IVT and smaller impact on precipitation over West Coast, particularly if only one aircraft was employed or flew over a remote area that is far away from the coast (IOPs 5, 8, and 11 in Figures 12c and 12d, Table S1 in Supporting Information S1).

The enhanced impact on precipitation from consecutive flights could be associated with the first guess and boundary conditions that inherited accumulated benefits from pre-existing AR Recon missions in the forcing data. We hypothesize that assimilating the targeted observations in the current cycle amplifies the positive impact from the first guess, which has information from upstream sampling in the previous cycle. The accumulated impact is greatest toward the end of the sequential flight period. The effect is strong enough to dominate despite the cold start for each IOP. Consistent results were implied in previous studies (e.g., Schindler et al., 2020; Stone et al., 2020). Specifically, Stone et al. (2020), who assimilated the dropsondes in a global model along with data from the entire global observing system, reported that the largest total observational impact was from observations collected on February 3, 2018, which is the IOP7 in this work and the last mission of a sequence of flights from January 27 to February 3 during 2018 AR Recon. Ongoing research is focused on examining this hypothesis with varying flight scenario experiments. In addition, multiple aircraft in a single IOP also increase the positive impact because the first-order impacts from dropsondes, like forecast errors, are flow-dependent. For example, assuming there are two flight paths along 130°W and 150°W (e.g., the IOP on January 29, 2018 in Figure 13), the impact from the eastern flight path reaches the West Coast first and then decays after 1 day. Depending on the phase speed of the weather system, the impact from the western flight path may advect toward the coast, bringing an interval of higher dropsonde impact (e.g., Figure 12b, IOP5) on subsequent days. When a Rossby wave packet is present over the western flight path, the impact could also propagate at the group velocity of the wave packet (Majumdar et al., 2010; Zheng et al., 2013). Nevertheless, cycled data denial experiments based on a large number of AR Recon cases are needed to confirm the impact of consecutive flights and multiple aircraft.

5.2. Targeting Tools, Data Assimilation Techniques, and Model Physics

The impact of targeted observations on model analysis and forecasts can depend on the detection of targeted regions during the flight planning stage (Ancell & Hakim, 2007; Bishop et al., 2001; Buizza et al., 2007; Majumdar et al., 2002; Wu et al., 2009). For example, if target areas are not correctly selected, the sampling of the real-time initial condition sensitivity areas will not be enhanced, thereby leading to negligible impact on downstream forecasts. The primary tools employed during AR Recon operations are the adjoint sensitivity tool from COAMPS (Errico, 1997; Doyle et al., 2014, 2019; Reynolds et al., 2019) and ensemble sensitivity analysis (Ancell & Hakim, 2007; Chang et al., 2013; Torn & Hakim, 2008; Zhang et al., 2007; Zheng et al., 2013) based on the ECMWF, GEFS, and MSC ensembles. These tools quantify how a forecast at a valid time over a verification domain is dependent on the initial state variables or the atmospheric state at an earlier time.

Both methods have their strengths and limitations. Neither the COAMPS adjoint nor the ensemble analysis employs the same forecast system as the WestWRF/GSI system used in this study. The adjoint sensitivity was based on analyses and forecasts from COAMPS, which provides an accurate, efficient, and physically
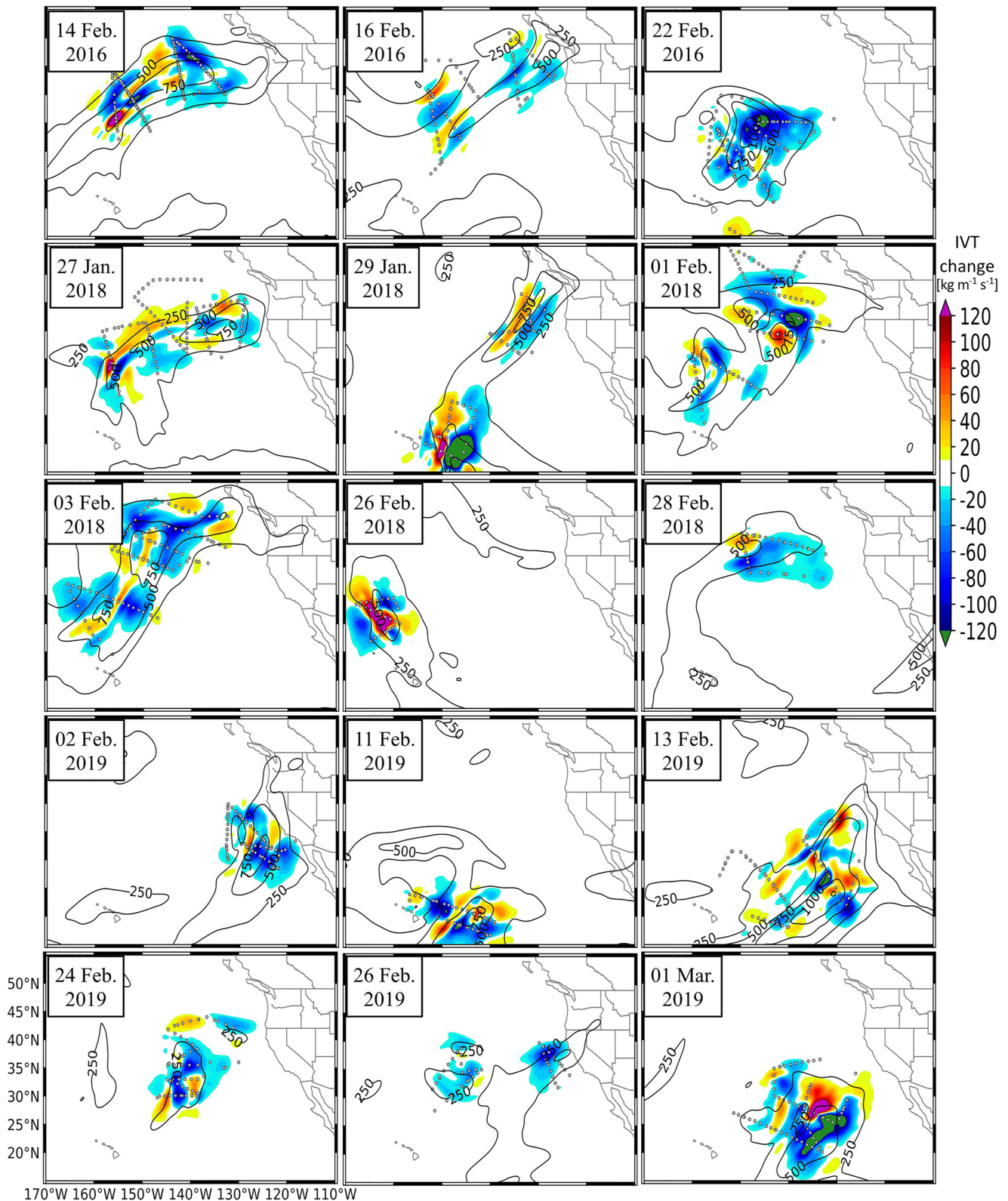
Figure 13. Difference in IVT amplitude between WithDROP and NoDROP in all 15 IOPs. Shades are the amplitude change (kg m^{-1} s^{-1}). Black solid contours are IVT (kg m^{-1} s^{-1}) amplitude in ERA5 starting from 250 kg m^{-1} s^{-1}. Gray filled dots are the locations of AR Recon flight-level and dropsonde data.

meaningful method to identify the most sensitive areas. The initial condition perturbations that will influence the downstream precipitation the most are not necessarily the same perturbations needed in West-WRF. Another limitation of adjoint sensitivity is that it is a deterministic method and therefore does not take into account the probabilistic guidance for forecast uncertainty. On the other hand, the ensemble sensitivity analysis considers the forecast uncertainty across ensemble members and provides a flexible and probabilistic tool to identify the sensitive areas but is limited by the quality of the ensemble. Ideally, if the ensemble is calibrated, the spread would reliably characterize the uncertainty over a number of cases. However, the ensemble spread in current operational ensemble products is far from perfect, due to ensemble size, observation constraints, and ensemble generation methods (e.g., Zheng et al., 2019). The limitations of the targeting tools mean that the AR Recon dropsondes are not guaranteed to be in the optimal locations for improving forecast skill in a particular model. For this reason, a variety of tools were used in flight planning to maximize confidence in identifying the sensitive areas.

Another contributing factor for dropsonde impact is the data assimilation technique (Majumdar, 2016). The dropsonde data used for this study are similar to what has been used for the NCEP GFS model, which are quality controlled data at reduced levels and only include a small portion of the raw dropsonde profiles particularly near the PBL (see Figure 14d of Zheng, Delle Monache, Wu, et al., 2021 for more details). Preliminary tests (not shown) using data assimilation techniques tailored to assimilate dropsondes suggested that if the dropsonde release time information was not considered during assimilation, for example using the 3DVar and hybrid 3DVar, results show degraded impact from dropsonde assimilation. This is particularly important for fast-moving atmospheric flows, which is also consistent with the findings from Zhang and Pu (2020) for hurricane forecasts. Parameter settings such as the localization scales and the observation error assignment will also influence the data impact. The choice of ensemble input, used to derive the flow-dependent error covariance, is also critical to the data impact. One strength for this set of experiments is the use of the combined GEFS and CMC ensemble. AR Recon data typically modified the IVT amplitude by 10%–40% (Figures 13 and S4 in Supporting Information S2). We think the overall positive impact we saw from the initial time is partly attributed to the use of the multi-ensemble that improved the ensemble error-spread skills compared with an individual ensemble (Zheng et al., 2019). In follow-up work, we will investigate whether the EnKF will further improve the data impact compared with the hybrid 4DEnVar.

Errors in the NWP model employed is another limiting factor for dropsonde impact, including the errors in model first guess and the model physics. On one hand, if the model is perfect and the initial condition has little deviation from the dropsondes, the room for improvement from dropsonde assimilation is limited. For example, Hamill et al. (2013) showed that the Winter Storm Reconnaissance data did not significantly improve forecast skill and considered that limited impact could be due to the advanced assimilation system employed as well as a denser observation network, rendering little room to improve upon the numerical model. On the other hand, if the deficiency in the model is too big and dominates the error growth with increasing forecast time, the correction of initial conditions might be negligible beyond certain forecast times or for certain weather systems. IOPs 4 and 13 appear to be such examples (Figures 13 and S4 in Supporting Information S2, Table S1 in Supporting Information S1). Although both IOPs showed improvement in initial conditions and short-term forecast (<18 h), the prediction for 24–48 h precipitation was not improved, partly due to the complex topography and cloud physics. Lavers et al. (2018) found that wind errors could be the major source of the IVT uncertainty in the ECMWF IFS and implied that the wind component of IVT might be more challenging to correct than the humidity component in NWPs, which might partly explain why we see more improvements in IWV than in IVT. To what degree the practical predictability can be improved requires further investigation through, for example, perturbed initial conditions, multi-physics experiments, and studies of more cases to separate initial-condition dominated errors and model-physics dominated errors.

One caveat is that the assimilated data did not include satellite radiances. Zheng, Delle Monache, Wu, et al. (2021) showed that most of the clear-sky radiance data were not used under AR conditions, leaving data gaps in the lower to middle troposphere. The all-sky radiance data for the GFS operational system were rejected as well if precipitating clouds are present, leaving observing gaps over the precipitating areas. Furthermore, assimilated microwave radiances were assigned large observation errors and often have limited impact on influencing forecasts over the Northeastern Pacific (Zhu et al., 2016, 2019). Extensive coverage

of AMV winds was assimilated in our experiments to reflect a portion of the information from satellite observations. Nevertheless, future work will be focused on the assessment of dropsonde impact in the context of all-sky radiance data with the regional model and the improvement of data assimilation techniques for data-sparse regions.

6. Conclusions

Landfalling ARs have a large impact on both hazardous weather and managing valuable water resources over the western US. Significant observational gaps still exist over the upstream Northeastern Pacific, leading to uncertainty in initial conditions that impact forecast skill for these AR events. Atmospheric River Reconnaissance (AR Recon, Ralph et al., 2020) was initiated in 2016 as a way to improve the forecast skill for the western US. The analysis presented here investigated the overall impact of the dropsonde observations collected during AR Recon in 2016, 2018, and 2019 using a GSI 4DEnVar system for implementation of the WRF model tailored for the western US named West-WRF (Martin et al., 2019).

Paired data denial experiments were performed, including WithDROP runs that assimilated dropsondes with other conventional and remotely sensed data, and NoDROP runs with the same assimilated observations as the WithDROP experiments except removing AR Recon dropsonde data. The difference between these paired runs showed a statistically significant impact from AR Recon flight-level and dropsonde data, especially in cases that were found after the fact to have lower-skill forecasts. In a representative case study initialized on 0000 UTC February 3, 2018, the assimilation of dropsonde data significantly reduced the moisture bias in the critical layer of the AR ($\sim$700–950 hPa) and the cold bias in the cloud top aloft around 300 hPa, and thereby reduced the IWT transport (IVT) bias by 30%–75% along the flanks of the AR at the initial time. The benefit of initial condition corrections in the WithDROP experiment propagated to the downstream areas, which clearly eliminated a localized precipitation error centered near Mount Rainier. The precipitation forecast error over western Washington was improved by $\sim$50% at a lead time of 12–36 h. The case study demonstrated that the dropsonde assimilation can not only improve the lower-level water vapor transport within the AR but can also improve the cloud temperature in the upper levels, both contributing significantly toward improving precipitation predictions.

Results based on the 15 Intensive Observation Periods (IOPs) of AR Recon showed that dropsondes improved the spatial pattern forecast for both IWV and IVT in all 15 IOPs for all forecast times out to day 6 over the extended Northeastern Pacific and Western US domain. Improvement for IVT and IWV based on the reduction of the median RMSE value was seen in over 80% of the IOPs and 85% of the forecast times out to day 6. Consistent forecast improvement in IVT and IWV was found in the first 3 days with the peak impact occurring at 6–18 h. The impact on accumulated 24-h precipitation over the US West Coast was generally consistent with the IVT/IWV but slightly smaller. It was found that the improvement for precipitation forecasts was maximized during subsequent consecutive flights, suggesting that consecutive missions can enhance the impact on forecasting downstream West Coast precipitation. The precipitation improvement based on the Fractional Skill Score tended to be greater in cases with large forecast error, such as IOPs 5, 7, and 15 for the forecast times of 72–96 h, with threshold greater than 10 mm showing an even greater positive impact from dropsondes than was seen using a 5-mm threshold, which suggest that AR Recon data are more important in poorly forecasted medium to heavy precipitation events than for light precipitation and good forecasts.

This study sheds light on how initial condition corrections associated with AR Recon dropsonde assimilation can improve forecast skill over the Northeast Pacific and the US West Coast with the advanced 4DEnVar method and a well-tuned ensemble as flow-dependent background errors. Given these promising results, future work will include assessing the impact in experiments with the full observational system, concentrating on events with poor forecast skill.

Data Availability Statement

The experiments were completed at the San Diego Supercomputer Center under the research allocation ATM150010. The West-WRF data can be accessed via UC San Diego Library Digital Collections from https://doi.org/10.6075/J0445MMG (Zheng, Delle Monache, Cornuelle, et al., 2021). The AR Recon dropsonde profiles have been made available publicly and interactively via the web interface at https://cw3e.ucsd.edu/arrecon_data. ERA5 verification data can be retrieved from https://cds.climate.copernicus.eu/#!/search?text=ERA5ERA-5&type=dataset. Stage-IV precipitation data can be accessed via https://doi.org/10.5065/D6PG1QDD.

References

Aberson, S. D. (2010). 10 years of hurricane synoptic surveillance (1997–2006). *Monthly Weather Review*, *138*(5), 1536–1549. https://doi.org/10.1175/2009MWR3090.1

Agustí-Panareda, A., Beljaars, A., Cardinali, C., Genkova, I., & Thorncroft, C. (2010). Impacts of assimilating AMMA soundings on ECMWF analyses and forecasts. *Weather and Forecasting*, *25*(4), 1142–1160. https://doi.org/10.1175/2010WAF2222370.1

Ancell, B., & Hakim, G. J. (2007). Comparing adjoint-and ensemble-sensitivity analysis with applications to observation targeting. *Monthly Weather Review*, *135*(12), 4117–4134. https://doi.org/10.1175/2007MWR1904.1

Bengtsson, L., Hodges, K. I., & Froude, L. S. (2005). Global observations and forecast skill. *Tellus A: Dynamic Meteorology and Oceanography*, *57*(4), 515–527. https://doi.org/10.3402/tellusa.v57i4.14723

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, *129*(3), 420–436. https://doi.org/10.1175/1520-0493(2001)129<0420:aswtet>2.0.co;2

Buizza, R., Cardinali, C., Kelly, G., & Thépaut, J. N. (2007). The value of observations. II: The value of observations located in singular-vector-based target areas. *Quarterly Journal of the Royal Meteorological Society*, *133*(628), 1817–1832. https://doi.org/10.1002/qj.149

Chang, E. K., Zheng, M., & Raeder, K. (2013). Medium-range ensemble sensitivity analysis of two extreme Pacific extratropical cyclones. *Monthly Weather Review*, *141*(1), 211–231. https://doi.org/10.1175/mwr-d-11-00304.1

Charles, M. E., & Colle, B. A. (2009). Verification of extratropical cyclones within the NCEP operational models. Part I: Analysis errors and short-term NAM and GFS forecasts. *Weather and Forecasting*, *24*(5), 1173–1190. https://doi.org/10.1175/2009WAF2222169.1

Chou, M.-D., & Suarez, M. J. (1994). An efficient thermal infrared radiation parameterization for use in general circulation models. *NASA Technical Memorandum*, *3*, 85.

Cobb, A., Michaelis, A., Iacobellis, S., Ralph, F. M., & Delle Monache, L. (2021). Atmospheric river sectors: Definition and characteristics observed using dropsondes from 2014-2020 CalWater and AR Recon. *Monthly Weather Review*, *149*, 623–644. https://doi.org/10.1175/MWR-D-20-0177.1

Cordeira, J. M., Ralph, F. M., & Moore, B. J. (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. https://doi.org/10.1175/MWR-D-13-00019.1

Corringham, T. W., Ralph, F. M., Gershunov, A., Cayan, D. R., & Talbot, C. A. (2019). Atmospheric rivers drive flood damages in the western United States. *Science Advances*, *5*(12), eaax4631. https://doi.org/10.1126/sciadv.aax4631

Davis, C. A. (1992). A potential-vorticity diagnosis of the importance of initial structure and condensational heating in observed extratropical cyclogenesis. *Monthly Weather Review*, *120*(11), 2409–2428. https://doi.org/10.1175/1520-0493(1992)120<2409:apvdot>2.0.co;2

Demirdjian, R., Doyle, J. D., Reynolds, C. A., Norris, J. R., Michaelis, A. C., & Ralph, F. M. (2020). A case study of the physical processes associated with the atmospheric river initial-condition sensitivity from an adjoint model. *Journal of the Atmospheric Sciences*, *77*(2), 691–709. https://doi.org/10.1175/JAS-D-19-0155.1

Dettinger, M. D. (2013). Atmospheric rivers as drought busters on the US West Coast. *Journal of Hydrometeorology*, *14*(6), 1721–1732. https://doi.org/10.1175/JHM-D-13-02.1

Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J., & Cayan, D. R. (2011). Atmospheric rivers, floods and the water resources of California. *Water*, *3*(2), 445–478. https://doi.org/10.3390/w3020445

Doyle, J. D., Amerault, C., Reynolds, C. A., & Reinecke, P. A. (2014). Initial condition sensitivity and predictability of a severe extratropical cyclone using a moist adjoint. *Monthly Weather Review*, *142*(1), 320–342. https://doi.org/10.1175/mwr-d-13-00201.1

Doyle, J. D., Reynolds, C. A., & Amerault, C. (2019). Adjoint sensitivity analysis of high-impact extratropical cyclones. *Monthly Weather Review*, *147*, 4511–4532. https://doi.org/10.1175/MWR-D-19-0055.1

Ebert, E. E. (2009). Neighborhood verification: A strategy for rewarding close forecasts. *Weather and Forecasting*, *24*(6), 1498–1510. https://doi.org/10.1175/2009WAF2222251.1

Errico, R. M. (1997). What is an adjoint model? *Bulletin of the American Meteorological Society*, *78*(11), 2577–2591. https://doi.org/10.1175/1520-0477(1997)078<2577:wiaam>2.0.co;2

Feng, J., & Wang, X. (2019). Impact of assimilating upper-level dropsonde observations collected during the TCI field campaign on the prediction of intensity and structure of Hurricane Patricia (2015). *Monthly Weather Review*, *147*(8), 3069–3089. https://doi.org/10.1175/2008WAF2222161.1

Froude, L. S., Bengtsson, L., & Hodges, K. I. (2007). The predictability of extratropical storm tracks and the sensitivity of their prediction to the observing system. *Monthly Weather Review*, *135*(2), 315–333. https://doi.org/10.1175/MWR3274.1

Grell, G. A., & Freitas, S. R. (2014). A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, *14*, 5233–5250. https://doi.org/10.5194/acp-14-5233-2014

Guan, B., & Waliser, D. E. (2019). Tracking atmospheric rivers globally: Spatial distributions and temporal evolution of life cycle characteristics. *Journal of Geophysical Research: Atmospheres*, *124*(23), 12523–12552. https://doi.org/10.1029/2019JD031205

Hamill, T. M., Yang, F., Cardinali, C., & Majumdar, S. J. (2013). Impact of targeted winter storm reconnaissance dropwindsonde data on midlatitude numerical weather predictions. *Monthly Weather Review*, *141*(6), 2058–2065. https://doi.org/10.1175/mwr-d-12-00309.1

Hanna, J. W., Schultz, D. M., & Irving, A. R. (2008). Cloud-top temperatures for precipitating winter clouds. *Journal of Applied Meteorology and Climatology*, *47*(1), 351–359. https://doi.org/10.1175/2007JAMC1549.1

Hersbach, H., Bill, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hodyss, D., & Majumdar, S. J. (2007). The contamination of 'data impact'in global models by rapidly growing mesoscale instabilities. *Quarterly Journal of the Royal Meteorological Society*, *133*(628), 1865–1875. https://doi.org/10.1002/qj.157

Hong, S. Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*(9), 2318–2341. https://doi.org/10.1175/MWR3199.1

Joly, A., Browning, K. A., Bessemoulin, P., Cammas, J. P., Caniaux, G., Chalon, J. P., et al. (1999). Overview of the field phase of the Fronts and Atlantic Storm-Track EXperiment (FASTEX) project. *Quarterly Journal of the Royal Meteorological Society*, *125*(561), 3131–3163. https://doi.org/10.1002/qj.49712556103

Khouakhi, A., & Villarini, G. (2016). On the relationship between atmospheric rivers and high sea water levels along the US West Coast. *Geophysical Research Letters*, *43*(16), 8815–8822. https://doi.org/10.1002/2016GL070086

Kleist, D. T., & Ide, K. (2015). An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Monthly Weather Review*, *143*(2), 452–470. https://doi.org/10.1175/MWR-D-13-00350.1

Lackmann, G. M. (2002). Cold-frontal potential vorticity maxima, the low-level jet, and moisture transport in extratropical cyclones. *Monthly Weather Review*, *130*(1), 59–74. https://doi.org/10.1175/1520-0493(2002)130<0059:cfpvmt>2.0.co;2

Langland, R. H. (2005). Issues in targeted observing. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3409–3425. https://doi.org/10.1256/qj.05.130

Langland, R. H., Toth, Z., Gelaro, R., Szunyogh, I., Shapiro, M. A., Majumdar, S. J., et al. (1999). The North Pacific Experiment (NOR-PEX-98): Targeted observations for improved North American weather forecasts. *Bulletin of the American Meteorological Society*, *80*(7), 1363–1384. https://doi.org/10.1175/1520-0477(1999)080<1363:tnpent>2.0.co;2

Lavers, D. A., Rodwell, M. J., Richardson, D. S., Ralph, F. M., Doyle, J. D., Reynolds, C. A., et al. (2018). The gauging and modeling of rivers in the sky. *Geophysical Research Letters*, *45*(15), 7828–7834. https://doi.org/10.1029/2018GL079019

Lin, Y., & Mitchell, K. E. (2005). *1.2 the NCEP stage II/IV hourly precipitation analyses: Development and applications* (Vol. *10*). Proceedings of the 19th Conference Hydrology, American Meteorological Society.

Lorenc, A. C., & Marriott, R. T. (2014). Forecast sensitivity to observations in the Met Office global numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, *140*(678), 209–224. https://doi.org/10.1002/qj.2122

Mahto, S. S., & Mishra, V. (2019). Does ERA-5 outperform other reanalysis products for hydrologic applications in India? *Journal of Geophysical Research: Atmospheres*, *124*(16), 9423–9441. https://doi.org/10.1029/2019JD031155

Majumdar, S. J. (2016). A review of targeted observations. *Bulletin of the American Meteorological Society*, *97*(12), 2287–2303. https://doi.org/10.1175/bams-d-14-00259.1

Majumdar, S. J., Bishop, C. H., Buizza, R., & Gelaro, R. (2002). A comparison of ensemble-transform Kalman-filter targeting guidance with ECMWF and NRL total-energy singular-vector guidance. *Quarterly Journal of the Royal Meteorological Society*, *128*(585), 2527–2549. https://doi.org/10.1256/qj.01.214

Majumdar, S. J., Sellwood, K. J., Hodyss, D., Toth, Z., & Song, Y. (2010). Characteristics of target areas selected by the ensemble transform Kalman filter for medium-range forecasts of high-impact winter weather. *Monthly Weather Review*, *138*(7), 2803–2824. https://doi.org/10.1175/2010MWR3106.1

Martin, A. C., Ralph, F. M., Demirdjian, R., DeHaan, L., Weihs, R., Helly, J., et al. (2018). Evaluation of atmospheric river predictions by the WRF Model using aircraft and regional Mesonet observations of orographic precipitation and its forcing. *Journal of Hydrometeorology*, *19*(7), 1097–1113. https://doi.org/10.1175/JHM-D-17-0098.1

Martin, A. C., Ralph, F. M., Wilson, A., DeHaan, L., & Kawzenuk, B. (2019). Rapid cyclogenesis from a mesoscale frontal wave on an atmospheric river: Impacts on forecast skill and predictability during atmospheric river landfall. *Journal of Hydrometeorology*, *20*(9), 1779–1794. https://doi.org/10.1175/JHM-D-18-0239.1

Matsui, T., Zhang, S. Q., Lang, S. E., Tao, W. K., Ichoku, C., & Peters-Lidard, C. D. (2018). Impact of radiation frequency, precipitation radiative forcing, and radiation column aggregation on convection-permitting West African monsoon simulations. *Climate Dynamics*, *55*(1), 193–213. https://doi.org/10.1007/s00382-018-4187-2

McMurdie, L. A., & Casola, J. H. (2009). Weather regimes and forecast errors in the Pacific Northwest. *Weather and Forecasting*, *24*(3), 829–842. https://doi.org/10.1175/2008WAF2222172.1

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*(D14), 16663–16682. https://doi.org/10.1029/97JD00237

NCEP. (2015). *NCEP GDAS/FNL 0.25 degree global tropospheric analyses and forecast grids, research data archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, NCEP*. National Weather Service, NOAA, U.S. Department of Commerce. https://doi.org/10.5065/D65Q4T4Z

Neiman, P. J., Ralph, F. M., Wick, G. A., Kuo, Y. H., Wee, T. K., Ma, Z., et al. (2008). Diagnosis of an intense atmospheric river impacting the Pacific Northwest: Storm summary and offshore vertical structure observed with COSMIC satellite retrievals. *Monthly Weather Review*, *136*(11), 4398–4420. https://doi.org/10.1175/2008MWR2550.1

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*(D12). https://doi.org/10.1029/2010JD015139

Peevey, T. R., English, J. M., Cucurull, L., Wang, H., & Kren, A. C. (2018). Improving winter storm forecasts with observing system simulation experiments (OSSEs). Part I: An idealized case study of three US storms. *Monthly Weather Review*, *146*(5), 1341–1366. https://doi.org/10.1175/MWR-D-17-0160.1

Pu, Z., Li, X., Velden, C. S., Aberson, S. D., & Liu, W. T. (2008). The impact of aircraft dropsonde and satellite wind data on numerical simulations of two landfalling tropical storms during the tropical cloud systems and processes experiment. *Weather and Forecasting*, *23*(1), 62–79. https://doi.org/10.1175/2007WAF2007006.1

Ralph, F. M., Cannon, F., Tallapragada, V., Davis, C. A., Doyle, J. D., Pappenberger, F., et al. (2020). West Coast forecast challenges and development of Atmospheric River Reconnaissance. *Bulletin of the American Meteorological Society*, *101*(8), E1357–E1377. https://doi.org/10.1175/BAMS-D-19-0183.1

Ralph, F. M., Dettinger, M., Lavers, D., Gorodetskaya, I. V., Martin, A., Viale, M., et al. (2017). Atmospheric rivers emerge as a global science and applications focus. *Bulletin of the American Meteorological Society*, *98*(9), 1969–1973. https://doi.org/10.1175/BAMS-D-16-0262.1

Ralph, F. M., Neiman, P. J., & Wick, G. A. (2004). Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the winter of 1997/98. *Monthly Weather Review*, *132*(7), 1721–1745. https://doi.org/10.1175/1520-0493(2004)132<1721:sacaoo>2.0.co;2

Ralph, F. M., Neiman, P. J., Wick, G. A., Gutman, S. I., Dettinger, M. D., Cayan, D. R., & White, A. B. (2006). Flooding on California's Russian River: Role of atmospheric rivers. *Geophysical Research Letters*, *33*(13), L13801. https://doi.org/10.1029/2006GL026689

Reynolds, C. A., Doyle, J. D., Ralph, F. M., & Demirdjian, R. (2019). Adjoint sensitivity of North Pacific atmospheric river forecasts. *Monthly Weather Review*, *147*(6), 1871–1897. https://doi.org/10.1175/MWR-D-18-0347.1

Roberts, N. M., & Lean, H. W. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, *136*(1), 78–97. https://doi.org/10.1175/2007MWR2123.1

Romine, G. S., Schwartz, C. S., Torn, R. D., & Weisman, M. L. (2016). Impact of assimilating dropsonde observations from MPEX on ensemble forecasts of severe weather events. *Monthly Weather Review*, *144*(10), 3799–3823. https://doi.org/10.1175/MWR-D-15-0407.1

Rutz, J. J., Steenburgh, W. J., & Ralph, F. M. (2014). Climatological characteristics of atmospheric rivers and their inland penetration over the western United States. *Monthly Weather Review*, *142*(2), 905–921. https://doi.org/10.1175/MWR-D-13-00168.1

Schindler, M., Weissmann, M., Schäfler, A., & Radnoti, G. (2020). The impact of dropsonde and extra radiosonde observations during NAWDEX in autumn 2016. *Monthly Weather Review*, *148*(2), 809–824. https://doi.org/10.1175/MWR-D-19-0126.1

Schwartz, C. S., Kain, J. S., Weiss, S. J., Xue, M., Bright, D. R., Kong, F., et al. (2009). Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. *Monthly Weather Review*, *137*(10), 3351–3372. https://doi.org/10.1175/2009MWR2924.1

Simmons, A. J., Soci, C., Nicolas, J., Bell, B., Berrisford, P., Dragani, R., et al. (2020). *Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1* (ECMWF Technical Memorandum 859, pp. 40). Retrieved from https://www.ecmwf.int/node/19362

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., & Powers, J. G. (2008). *A description of the Advanced Research WRF version 3*. NCAR Technical note-475+ STR.

Stone, R. E., Reynolds, C. A., Doyle, J. D., Langland, R. H., Baker, N. L., Lavers, D. A., & Ralph, F. M. (2020). Atmospheric River Reconnaissance observation impact in the Navy global forecast system. *Monthly Weather Review*, *148*(2), 763–782. https://doi.org/10.1175/MWR-D-19-0101.1

Szunyogh, I., Toth, Z., Morss, R. E., Majumdar, S. J., Etherton, B. J., & Bishop, C. H. (2000). The effect of targeted dropsonde observations during the 1999 Winter Storm Reconnaissance Program. *Monthly Weather Review*, *128*(10), 3520–3537. https://doi.org/10.1175/1520-0493(2000)128<3520:teotdo>2.0.co;2

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, *136*(12), 5095–5115. https://doi.org/10.1175/2008MWR2387.1

Tong, M., Sippel, J. A., Tallapragada, V., Liu, E., Kieu, C., Kwon, I. H., et al. (2018). Impact of assimilating aircraft reconnaissance observations on tropical cyclone initialization and prediction using operational HWRF and GSI ensemble–variational hybrid data assimilation. *Monthly Weather Review*, *146*(12), 4155–4177. https://doi.org/10.1175/MWR-D-17-0380.1

Torn, R. D., & Hakim, G. J. (2008). Ensemble-based sensitivity analysis. *Monthly Weather Review*, *136*(2), 663–677. https://doi.org/10.1175/2007MWR2132.1

Velden, C., Daniels, J., Stettner, D., Santek, D., Key, J., Dunion, J., et al. (2005). Recent innovations in deriving tropospheric winds from meteorological satellites. *Bulletin of the American Meteorological Society*, *86*(2), 205–224. https://doi.org/10.1175/BAMS-86-2-205

Waliser, D., & Guan, B. (2017). Extreme winds and precipitation during landfall of atmospheric rivers. *Nature Geoscience*, *10*(3), 179–183. https://doi.org/10.1038/ngeo2894

Wang, X., & Lei, T. (2014). GSI-based four-dimensional ensemble–variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP Global Forecast System. *Monthly Weather Review*, *142*(9), 3303–3325. https://doi.org/10.1175/MWR-D-13-00303.1

Wedam, G. B., McMurdie, L. A., & Mass, C. F. (2009). Comparison of model forecast skill of sea level pressure along the east and west coasts of the United States. *Weather and Forecasting*, *24*(3), 843–854. https://doi.org/10.1175/2008WAF2222161.1

Weissmann, M., Harnisch, F., Wu, C. C., Lin, P. H., Ohta, Y., Yamashita, K., et al. (2011). The influence of assimilating dropsonde data on typhoon track and midlatitude forecasts. *Monthly Weather Review*, *139*(3), 908–920. https://doi.org/10.1175/2010MWR3377.1

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

Wu, C. C., Chen, J. H., Majumdar, S. J., Peng, M. S., Reynolds, C. A., Aberson, S. D., et al. (2009). Intercomparison of targeted observation guidance for tropical cyclones in the northwestern Pacific. *Monthly Weather Review*, *137*(8), 2471–2492. https://doi.org/10.1175/2009mwr2762.1

Zhang, F., Bei, N., Rotunno, R., Snyder, C., & Epifanio, C. C. (2007). Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *Journal of the Atmospheric Sciences*, *64*(10), 3579–3594. https://doi.org/10.1175/JAS4028.1

Zhang, S., & Pu, Z. (2020). Evaluation of the four-dimensional ensemble-variational hybrid data assimilation with self-consistent regional background error covariance for improved hurricane intensity forecasts. *Atmosphere*, *11*(9), 1007. https://doi.org/10.3390/atmos11091007

Zhang, Z., & Ralph, F. M. (2021). The influence of antecedent atmospheric river conditions on extratropical cyclogenesis. *Monthly Weather Review*, *149*, 1337–1357. https://doi.org/10.1175/MWR-D-20-0212.1

Zhang, Z., Ralph, F. M., & Zheng, M. (2019). The relationship between extratropical cyclone strength and atmospheric river intensity and position. *Geophysical Research Letters*, *46*, 1814–1823. https://doi.org/10.1029/2018gl079071

Zheng, M., Chang, E. K., & Colle, B. A. (2013). Ensemble sensitivity tools for assessing extratropical cyclone intensity and track predictability. *Weather and Forecasting*, *28*(5), 1133–1156. https://doi.org/10.1175/WAF-D-12-00132.1

Zheng, M., Chang, E. K., & Colle, B. A. (2019). Evaluating US East Coast winter storms in a multimodel ensemble using EOF and clustering approaches. *Monthly Weather Review*, *147*(6), 1967–1987. https://doi.org/10.1175/MWR-D-18-0052.1

Zheng, M., Delle Monache, L., Cornuelle, B. D., Ralph, M. F., Tallapragada, V. S., Subramanian, A., et al. (2021). *Data from: Improved forecast skill through the assimilation of dropsonde observations from the Atmospheric River Reconnaissance program*. UC San Diego Library Digital Collections. https://doi.org/10.6075/J0445MMG

Zheng, M., Delle Monache, L., Wu, X., Ralph, F. M., Cornuelle, B., Tallapragada, V., et al. (2021). Data gaps within atmospheric rivers over the Northeastern Pacific. *Bulletin of the American Meteorological Society*, *102*(3), E492–E524. https://doi.org/10.1175/BAMS-D-19-0287.1

Zhu, Y., Gayno, G., Purser, R. J., Su, X., & Yang, R. (2019). Expansion of the all-sky radiance assimilation to ATMS at NCEP. *Monthly Weather Review*, *147*(7), 2603–2620. https://doi.org/10.1175/MWR-D-18-0228.1

Zhu, Y., Liu, E., Mahajan, R., Thomas, C., Groff, D., Van Delst, P., et al. (2016). All-sky microwave radiance assimilation in NCEP's GSI analysis system. *Monthly Weather Review*, *144*(12), 4709–4735. https://doi.org/10.1175/MWR-D-15-0445.1

Zhu, Y., & Newell, R. E. (1998). A proposed algorithm for moisture fluxes from atmospheric rivers. *Monthly Weather Review*, *126*(3), 725–735. https://doi.org/10.1175/1520-0493(1998)126<0725:apafmf>2.0.co;2