An Assessment of Dropsonde Sampling Strategies for Atmospheric River Reconnaissance🖉

MINGHUA ZHENG🆔,[a] RYAN TORN,[b] LUCA DELLE MONACHE,[a] JAMES DOYLE,[c] FRED MARTIN RALPH,[a]
VIJAY TALLAPRAGADA,[d] CHRISTOPHER DAVIS,[e] DANIEL STEINHOFF,[a] XINGREN WU,[d,f] ANNA WILSON,[a]
CAROLINE PAPADOPOULOS,[a] AND PATRICK MULROONEY[a]

[a] Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California
[b] University at Albany, State University of New York, Albany, New York
[c] U.S. Naval Research Laboratory, Monterey, California
[d] NOAA/NCEP/Environmental Modeling Center, College Park, Maryland
[e] National Center for Atmospheric Research, Boulder, Colorado
[f] Axiom at EMC/NCEP/NOAA, College Park, Maryland

ABSTRACT: During a 6-day intensive observing period in January 2021, Atmospheric River Reconnaissance (AR Recon) aircraft sampled a series of atmospheric rivers (ARs) over the northeastern Pacific that caused heavy precipitation over coastal California and the Sierra Nevada. Using these observations, data denial experiments were conducted with a regional modeling and data assimilation system to explore the impacts of research flight frequency and spatial resolution of dropsondes on model analyses and forecasts. Results indicate that dropsondes significantly improve the representation of ARs in the model analyses and positively impact the forecast skill of ARs and quantitative precipitation forecasts (QPF), particularly for lead times > 1 day. Both reduced mission frequency and reduced dropsonde horizontal resolution degrade forecast skill. On the other hand, experiments that assimilated only G-IV data and experiments that assimilated both G-IV and C-130 data show better forecast skill than experiments that only assimilated C-130 data, suggesting that the additional information provided by G-IV data is necessary for improving forecast skill. Although this is a case study, the 6-day period studied encompassed multiple AR events that are representative of typical AR behavior. Therefore, the results indicate that future operational AR Recon missions incorporate daily mission or back-to-back flights, maintain current dropsonde spacing, support high-resolution data transfer capacity on the C-130s, and utilize G-IV aircraft in addition to C-130s.

KEYWORDS: Atmospheric river; Dropsondes; Data assimilation; Regional models

1. Introduction

Weather reconnaissance aims to collect accurate meteorological observations in data-sparse areas, improve numerical weather forecasts of high-impact weather events, and advance the understanding of certain weather phenomenon (e.g., Burpee et al. 1996; Langland et al. 1999; Aberson 2010; Weisman et al. 2015; Ralph et al. 2020; Wick et al. 2020; McMurdie et al. 2022). Observations collected during weather reconnaissance campaigns are often referred to as "targeted observations" [see Majumdar (2016) for an overview of these observations].

Data from operational campaigns are typically assimilated into operational numerical weather prediction (NWP) models to improve the forecast accuracy of a specific high-impact weather type and to mitigate the associated social and economic impacts. One such impactful weather type is atmospheric rivers (ARs), a global weather phenomenon that can transport large amounts of water vapor from the tropics to mid- and high latitudes (Zhu and Newell 1998; Waliser and Guan 2017). In recent years, ARs have been increasingly recognized as key drivers of high-impact weather and hydrological events in many regions of the world (Waliser and Guan 2017; Zhang et al. 2019; Payne et al. 2020; Zhang and Ralph 2021). Landfalling ARs contribute up to 50% of the annual precipitation over states along the U.S. West Coast (Guan et al. 2010; Dettinger et al. 2011; Rutz et al. 2014). This can be beneficial as a solution to water sparsity (Dettinger 2013) but can also be hazardous as a cause of major flooding events (Ralph et al. 2006; Ralph and Dettinger 2011; Henn et al. 2020). To better observe ARs and to improve the forecast skill over the U.S. West, Atmospheric River Reconnaissance (AR Recon) campaigns have been conducted to collect observations, including dropsondes from 1 to 3 aircraft (Ralph et al. 2020), extra drifting buoys relative to the existing network of ocean buoys (Reynolds et al. 2023), additional radiosonde launches (Ralph et al. 2021; Cobb et al. 2024), and airborne GPS radio occultation (Haase et al. 2021). These collaborative efforts are led by the Center for Western Weather and Water Extremes (CW3E) of Scripps Institution of Oceanography at University of California San Diego in collaboration with NOAA and the Air Force (Ralph et al. 2020). AR Recon has been included in the National Winter Season Operations Plan [NWSOP; Office of the Federal Coordinator for Meteorology (OFCM); OFCM 2019, 2022] since 2019. More details about this collaborative campaign can be found in Ralph et al. (2020).

The impacts of observations (e.g., dropsonde data), collected from historical reconnaissance missions, on the NWP forecast skill have been extensively investigated, with varying results. Several studies have shown an overall positive impact of targeted observations on forecast skill, including for hurricanes (e.g., Burpee et al. 1996; Pu et al. 2008; Weissmann et al. 2011; Majumdar et al. 2013; Feng and Wang 2019), winter storms (e.g., Langland et al. 1999; Szunyogh et al. 2000; Schindler et al. 2020), and mesoscale weather (e.g., Romine et al. 2016), though some studies have reported neutral (e.g., Hamill et al. 2013) or negative impacts (e.g., Aberson 2008; Keclik et al. 2017) from targeted observations. The inconsistent findings across studies result from a complex array of factors, including sample size, phenomenology, observational deployment strategy (e.g., Majumdar et al. 2002a,b), and possibly the data assimilation methods (e.g., Bergot 2001) and numerical model characteristics. Zheng et al. (2021a) demonstrated the vital role of AR Recon observations in filling observational spatial gaps from near surface to the middle troposphere within and around ARs in the northeast Pacific Ocean. Zheng et al. (2021b) showed that the assimilation of dropsondes with the Weather Research and Forecasting (WRF; Skamarock et al. 2019) model and the hybrid four-dimensional ensemble-variational (4DEnVar; Wang and Lei 2014; Kleist and Ide 2015) method improves the forecast of water vapor transport out to a lead time of three days, and improves precipitation forecast in the targeted time window (nominally a lead time of 12–36 h). Other studies demonstrate a greater positive dropsonde impact, on a per-observation basis, relative to the North American radiosonde network (Stone et al. 2020), and a comparable positive impact with that from microwave satellite radiance types (Sun et al. 2022). Recent research conducted by Lord et al. (2023a,b) demonstrated enhanced forecast accuracy for U.S. West Coast precipitation in the medium range and for dynamical fields in the short range, utilizing the Global Forecast System (GFS) at the National Centers for Environmental Prediction (NCEP).

While many studies investigated the impact from all reconnaissance observations, very few have investigated the sensitivity of forecast skill to different sampling strategies by subsampling observations in space and/or time. Kren et al. (2020) found that different flight paths can change the downstream forecast uncertainty by up to 8% under the framework of the observing system simulation experiment (OSSE). They suggested that field missions should take the uncertainty in flight path design, including the orientation, pattern, sensitivity regions, and meteorological features, into consideration. To date, there are still fundamental questions regarding the sampling strategy that remain unanswered, such as determining the optimal temporal spacing between aircraft missions and spatial resolution for dropsondes. Answering these questions will allow us to maximize the potential benefits to the forecast and demonstrate the need for appropriate resources.

Over the past couple of years, the AR Recon team has continued to expand the sample size with an increasing number of intensive observation periods (IOPs), due to demonstrated need for this information (e.g., Zheng et al. 2021a) and positive impacts on precipitation forecasts (e.g., Lord et al. 2023a). This expansion has included a focus on flight sequences, that is, a series of missions flown on consecutive days (e.g., Cobb et al. 2024). These missions have allowed for the design of a variety of full data denial experiments (e.g., Masutani et al. 2013) using subsets of dropsondes to represent different flight scenarios. As a result, improved strategies could be identified by comparing the overall skill of different experiments.

The main objective of this study is to explore the impact of different temporal sampling, spatial sampling (horizontal and vertical), and aircraft used on the forecast skill of an AR-related heavy precipitation event that was sampled over a 6-day sequence of IOPs during AR Recon 2021.

Specifically, we are focused on the following research and operational questions:

(i) How does the time frequency of flights and the number of flights impact the forecast skill of landfalling ARs and the associated precipitation?
(ii) How does the spatial resolution of dropsondes influence the forecast skill of landfalling ARs and the associated precipitation?
(iii) What are the added benefits of using high-vertical-resolution observations compared with the operational practice of including only mandatory and significant-level data?

These questions will be explored in this paper.

2. Methodology

a. Experiment design

1) PREDICTION SYSTEM CONFIGURATION

(i) Forecasting model

The model configuration employed for this study is based on West-WRF (Martin et al. 2018; Zheng et al. 2021b), using WRFv4.0 (Skamarock et al. 2019) with a focus on performing data impact studies and improving forecast skills over the western United States. The model had a domain extended to the central Pacific Ocean and the East Coast of the United States. (Fig. 1a). A total of 80 levels are configured with a model top at 10 hPa (Fig. 1b). The vertical resolution is enhanced near the altitudes where low-level and upper-level jets frequently occur to better capture sharp vertical gradients that can exist in these layers (e.g., Ralph et al. 2004, 2005). The model physics schemes are listed in Table 1. Model initial and boundary conditions (ICBCs) for D01 are forced by the analysis and forecast products (0.25° × 0.25° latitude–longitude grids) from the operational GFS at NCEP. D01 provides ICBCs for D02 and the one-way nesting option is employed.

(ii) Data assimilation system

The data assimilation system used for this study is based on the Gridpoint Statistical Interpolation (GSI) hybrid four-dimensional ensemble-variational (4DEnVar; Wang and Lei 2014) data assimilation technique (Kleist and Ide 2015). The background error-covariance matrix for 4DEnVar combines both the static and ensemble error-covariance matrices. The ensemble error-covariance matrix was calculated from a 30-member 9-km ensemble generated by the West-WRF Model using the NCEP Global Ensemble Forecast System (GEFS) forcing dataset. The 4DEnVar allows for the assimilation of observations at the appropriate assimilation time window around the observed time.
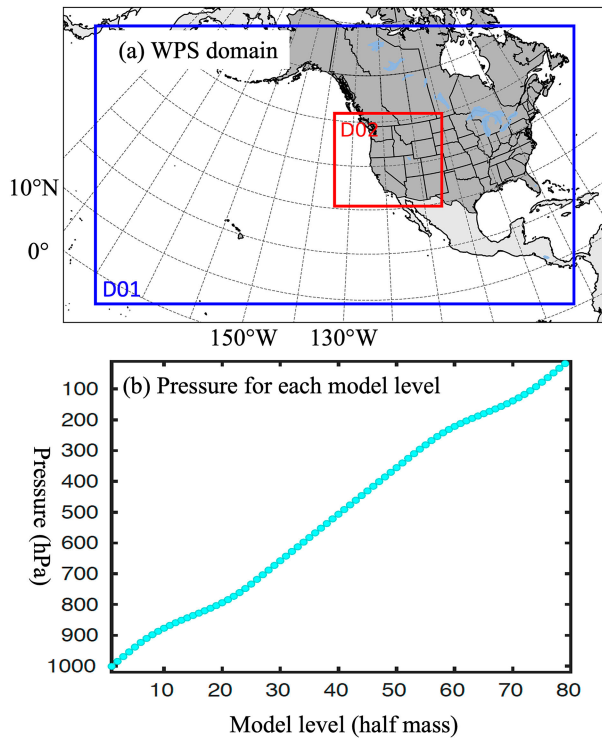
FIG. 1. WRF (a) Preprocessor System (WPS) domain and (b) the vertical-level configuration. D01 is the outer domain at 9-km grid spacing and D02 is for the nested domain at 3-km grid spacing.

The observations assimilated by the West-WRF GSI system include the following: 1) conventional observations included in the PrepBUFR files from the Global Data Assimilation System (GDAS) of NCEP; 2) the operational version of atmospheric motion vectors (AMVs; Velden et al. 2005; Santek et al. 2019); 3) The GNSS Radio Occultation (RO) refractivity (e.g., Healy 2011); 4) microwave radiance observations from Advanced Microwave Sounding Unit-A (AMSU-A), Advanced Technology Microwave Sounder (ATMS), Microwave Humidity Sounder (MHS), and Special Sensor Microwave Imager/Sounder (SSMI/S); and 5) infrared radiance observations from High-Resolution Infrared Radiation Sounder (HIRS/4) and Infrared Atmospheric Sounding Interferometer (IASI). Data assimilation was conducted only within the outer domain, and the inner domain was initialized by interpolating from the outer domain. This means that assimilation occurred over the region where the inner domain existed, but it was not conducted at the resolution of the inner domain.

The cycled data assimilation was performed every 6 h centered at 0000, 0600, 1200, and 1800 UTC of each day, where the model background data were the hourly outputs from the previous cycle using West-WRF. The model background and the ensemble perturbation input during each of the 6-h assimilation time window was based on hourly forecasts between 3 and 9 h from the previous forecast. Observations were sorted to each hourly time interval of this 6-h window according to the time when each observation occurred.

(iii) Dropsonde data for assimilation

We focus primarily on the dropsonde observations collected from the NOAA G-IV and Air Force C-130s (Table S1, Figs. S1 and S2 in the online supplemental material) during AR Recon, which represent the key dataset of interest in this study. Note that the operational PrepBUFR data includes the AR Recon dropsondes at a lower vertical resolution (i.e., mandatory and significant levels), which is used in the operational NCEP analyses and hereafter is referred to as the operational version. The operational version with reduced vertical resolution typically includes ~20–40 pressure levels while the raw dropsonde profiles can typically have ~2000–4000 vertical levels, depending on the aircraft and the flight altitudes. To create an observation dataset with an improved vertical resolution of the AR Recon dropsondes over the operational version, the raw dropsonde files were processed through the superobbing method (e.g., van Leeuwen 2015). The superobbed observations were obtained by averaging observations whose observed pressure values fell within half of the model layer above and below a model level for temperature, humidity, horizontal wind, and pressure. The superobbed data, created through this method, can condense information obtained from raw observations, while also providing a greater number of levels compared to the operational reduced version of dropsonde data. These superobbed dropsondes data were utilized to replace the lower-vertical-resolution dropsonde data in the operational PrepBUFR files to create new PrepBUFR data files.

2) DATA DENIAL EXPERIMENTS

(i) Baseline experiments

The overarching purpose of data denial experiments is to determine the influence of specific observations on forecast skill (i.e., observation impact). Typically, a control run is performed in which all operational observation types are

TABLE 1. Details of the physics parameterization schemes used in the WRF simulations.

Physical process	Scheme(s) used	Reference(s)
Planetary boundary layer scheme	Yonsei University scheme	Hong et al. (2006)
Cloud microphysics	Thompson microphysics scheme	Thompson et al. (2008)
Land surface exchange	Unified Noah land surface model	Chen and Dudhia (2001), Tewari et al. (2004)
Surface layer	Monin–Obukhov similarity scheme	Skamarock et al. (2019)
Radiation	RRTMG shortwave and longwave schemes	Iacono et al. (2008)
Deep convection	Grell–Freitas cumulus scheme (only applied to D01)	Grell and Freitas (2014)

TABLE 2. A summary of the experiments conducted in this study. The letter "Y" denotes the assimilation of AR Recon dropsondes from each IOP and "N" denotes the rejection of dropsondes. Here "S" denotes that the assimilated dropsondes are the high-resolution superobbed version. The "O" denotes the experiment that assimilates dropsondes using the operational version with reduced levels.

Name and IOPs	IOP3 23 Jan	IOP4 24 Jan	IOP5 25 Jan	IOP6 26 Jan	IOP7 27 Jan	IOP8 28 Jan
Control	Y (S)	Y (S)	Y (S)	Y (S)	Y (S)	Y (S)
NoDROP	N	N	N	N	N	N
ManSig	Y (O)	Y (O)	Y (O)	Y (O)	Y (O)	Y (O)
TS2	Y (S)	N	Y (S)	N	Y (S)	N
TS3	Y (S)	N	N	Y (S)	N	N
TSsingle	N	N	N	N	Y (S)	N
SS3	Y (S: 1/3)	Y (S: 1/3)	Y (S: 1/3)	Y (S: 1/3)	Y (S: 1/3)	Y (S: 1/3)
SS5	Y (S: 1/5)	Y (S: 1/5)	Y (S: 1/5)	Y (S: 1/5)	Y (S: 1/5)	Y (S: 1/5)
SS_C130	N	Y (S)	N	N	Y (S)	Y (S)
SS_G4	Y (S)	Y (S)	Y (S)	Y (S)	Y (S)	Y (S)

preprocessed and assimilated. A denial run, which is still a full NWP experiment, is carried out based on the operational observations but with an observation type of interest removed from the data assimilation steps. The forecasts from the denial experiment are then compared with those produced via a control run, with any differences indicating the observation impact of the denied observation type.

In this study, we conducted two baseline experiments to assess the impact of all AR Recon dropsonde data on model analyses and forecasts. The first experiment, referred to as "Control," assimilated all observational data for the West-WRF GSI system [details provided in section 2a(1)]. As for the AR Recon data, we replaced the flight-level and operational version of dropsonde profile data from the operational PrepBUFR files with the superobbed high-resolution dropsonde observations to create a new PrepBUFR file for each IOP (Table 2). The second baseline experiment, referred to as "NoDROP," was identical to "Control," but with the AR Recon data excluded from the data assimilation process. Forecasts are initialized from initial conditions during 0000 UTC 23 January–0000 UTC 28 January generated with both experiments.

(ii) Temporal sampling (TS) experiments

To evaluate the role of aircraft mission frequency, a set of denial experiments denoted by "TS" at the beginning of the acronym are carried out (Table 2, Table S2). Of them, "TS2" assimilates dropsondes collected during IOP3, IOP5, and IOP7, as a representation of flying the targeted system every other day. "TS3" only assimilates dropsondes during IOP3 and IOP6 as a representation of flying the targeted weather system every 3 days. An additional "TSsingle" experiment is also configured that assimilated observations from IOP7 designed to represent a single flight mission at a time just prior to the heaviest precipitation event.

(iii) Spatial sampling (SS) experiments

The second set of denial experiments was configured to investigate the impact of horizontal spatial resolution of the dropsondes in order to explore different deployment scenarios (Table 2). Of these experiments, "SS3" assimilated data from every third

deployed dropsondes (i.e., assimilated 1/3 of the deployed dropsondes). "SS5" assimilated every fifth dropsonde (i.e., only 1/5 of the dropsondes). In addition, the "SS_C130" run assimilated dropsonde data collected from Air Force (AF) C130 aircraft only while "SS_G4" assimilated dropsonde data collected from NOAA G-IV aircraft only.

(iv) The operational experiment

The final experiment explores the potential added benefit of using high-vertical-resolution dropsonde data by assimilating the operational version of dropsonde data. This experiment is referred to as "Oper" and is similar to the "Control" experiment, with the only difference being that the original PrepBUFR data, which includes the operational version of dropsonde data, is assimilated. Figure 2 shows that the number of assimilated direct observations in the Control experiment is significantly higher than that in the ManSig experiment during each assimilation window centered at 0000 UTC.

b. Data validation

In all experiments, meteorological variables are validated against the ERA5 data (Hersbach et al. 2020). We selected the ERA5 data because model forcing data were based on the NCEP GFS products, and thus the ECMWF reanalysis will provide more of an independent validation. Recent studies (e.g., Cobb et al. 2021) have demonstrated that the ERA5 dataset has fewer integrated water vapor transport (IVT) errors when compared to other reanalysis datasets, thus making it a reliable choice for validation. Stage-IV precipitation products at 4-km grid spacing (Du 2011) were employed as a high-resolution validation data for model precipitation from domain 2, following Zheng et al. (2021b).

c. Forecast skill metrics

The Method for Object-Based Diagnostic Evaluation (MODE; Davis et al. 2006, 2009), which is part of the Model Evaluation Tools (MET; Brown et al. 2021), has been employed to validate precipitation and IVT. MODE can identify objects in both the forecast and observation fields and match the observation object with the forecast object based on predefined thresholds. The MODE tool has been applied to object-
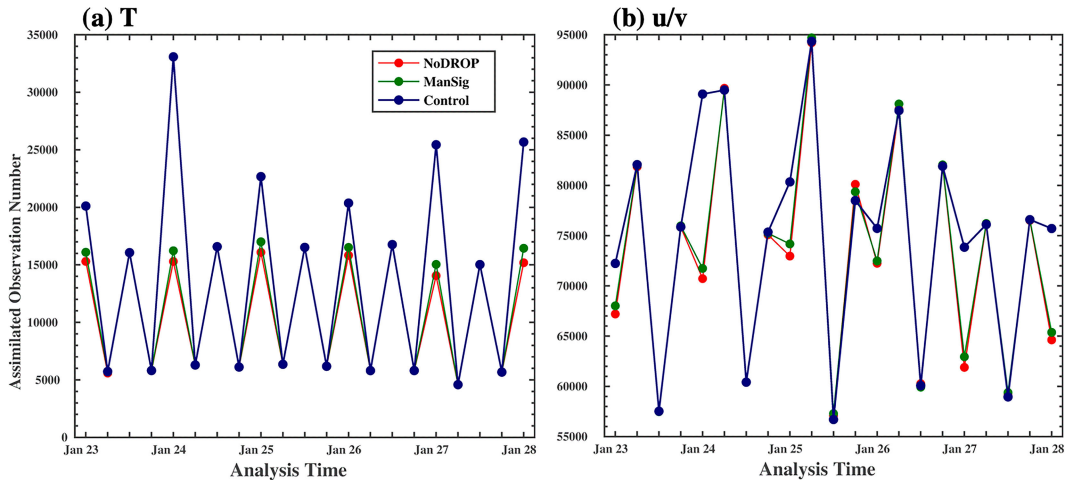
FIG. 2. Numbers of assimilated (a) temperature and (b) horizontal wind observations in the Control, NoDROP, and ManSig experiments during each 6-h assimilation window from 0000 UTC 23 Jan to 0000 UTC 28 Jan.

based validation for ARs in DeHaan et al. (2021), where it was able to correctly identify ARs, and for high-resolution gridded precipitation validation (e.g., Brown et al. 2021). Moreover, MODE provides an intuitive way to interpret the physical meaning of validation results.

MODE outputs attributes (e.g., object size, 90th-percentile values) for each pair of forecast and observation objects to assess the forecast skill. In addition, the total interest value, a summary statistic that computes a weighted average of the shape and intensity attributes of one or more matching forecast-observation paired precipitation areas, represents the overall forecast skill. A value of 1 in the total interest represents a perfect match between a pair of objects (Davis et al. 2009).

3. Synoptic overview

a. A synoptic overview of the sequence IOPs

On 0000 UTC 23 January 2021, an AR (hereafter AR-I), characterized by enhanced integrated water vapor transport (IVT, integrated from 1000 to 300 hPa), is present between two high pressure areas in the Northeast Pacific Ocean (IOP3, Fig. 3a). At 300 hPa, an ~80 m s^{-1} upper-level jet lies just west to the core of AR-I (Fig. 4a). South of 30°N, there is another enhanced IVT region transporting water vapor westward from Hawaii to the date line (Fig. 3a). This feature is associated with an inverted surface pressure trough near 165°W between 20° and 25°N in the downstream of a 500-hPa trough near the date line (Fig. 4a). The G-IV aircraft samples the upper-level trough and northwestern peripheral regions of the tropical moisture plume (Figs. 3a and 4a). AR-I propagates eastward by ~20° longitude (Fig. 3b) and dissipates after making landfall at 1200 UTC 24 January (not shown). The associated upper-level jet streak also propagates eastward, and two Air Force (AF) C-130 aircraft sample this jet streak and AR-I on 24 January (IOP4, Figs. 3b and 4b). Meanwhile, the previous two highs merge into one well-defined high centered around 42°N, 152°W (Fig. 3b), with northward moisture

advection from tropics on its southwestern periphery (Fig. 3b). The surface inverted pressure trough continues to develop and interacted with the upper-level trough between 170°W and the date line (Figs. 3b and 4b) through potential vorticity advection. A G-IV flight samples the strengthening low pressure trough and core of the moisture advection (Figs. 4b and 3b). The moisture advection exhibits further development at 0000 UTC 25 January (IOP5) with two IVT maxima formed: one centered at 30°N, 167°W, the other centered to its northeast near 45°N (Fig. 3c). A low pressure region forms along the western edge of the moisture advection (Fig. 3c), along with a closed upper level low (Fig. 4c), both of which are sampled by the G-IV.

One day later (26 January and IOP6, Fig. 3d), a high-latitude low pressure system develops with the center near 55°N, 145°W, helping to form a new AR (hereafter AR-II) with maximum IVT near 43°N, 140°W over the enhanced pressure gradient on the southwestern flank of the low. The enhanced IVT in the AR is detached from the northern center of the meridional moisture advection on the previous day and is also associated with a zonal jet stream along 148°W (Figs. 3d and 4d). One C-130 aircraft samples AR-II and the southern flank of the jet stream. The meridional moisture advection along the western side of the high pressure ridge is advected further westward and merges with the enhanced upstream IVT. AR-II propagates southeastward the following day and makes landfall over Northern California and southern Oregon at 0000 UTC 27 January (IOP7, Fig. 3e). Two upper-level jet streaks are located at the western side and the southern base of the closed low offshore the Pacific Northwest (Fig. 4e). One Air Force C-130 and the G-IV sample AR-II and its trailing area, respectively (Fig. 3e). AR-II moves southward and makes landfall on 28 January (IOP8) in Central California with decreased IVT, while the upstream moisture advection wraps around the high pressure system and a 500-hPa deep trough forms offshore (Figs. 3f and 4f). An Air Force C-130 continues to sample AR-II and the G-IV samples the leading part of the upper-level jet streak (Figs. 3f and 4f). AR-II continues to impact Southern California the following day (0000 UTC 29 January 2021) but
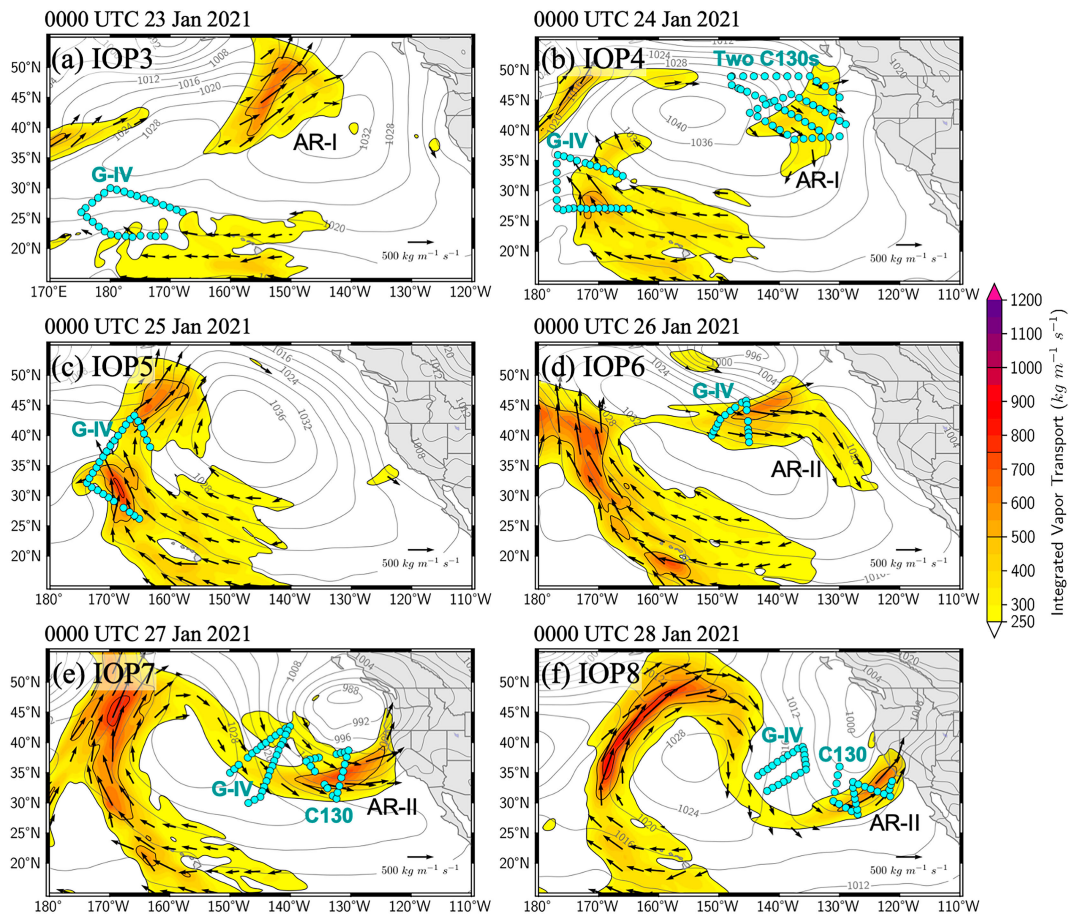
FIG. 3. An overview of the IVT vectors (black arrows) and amplitude (shading, kg m^{-1} s^{-1}) and MSLP (gray contours, hPa) in the ERA5 data, and dropsonde distributions (filled cyan markers). Analysis is valid from (a) 0000 UTC 23 Jan (IOP3) to (f) 0000 UTC 28 Jan 2021, with a time interval of 24 h.

weakened (not shown). After 1200 UTC 29 January, AR-II dissipates and moves southward out of California.

b. A summary of the impacts in California

AR-II made landfall from 0000 UTC 27 January to 0000 UTC 29 January, classified as a long-duration AR (e.g., Zhou et al. 2018). Light to moderate precipitation was distributed along the Northern California coast into the central valleys. Moderate to heavy precipitation was observed over the Central California coast during 1200 UTC 26 January–1200 UTC 27 January (Fig. 5a), along with snowfall in the Northern California mountains and the Sierra Nevada (hereafter Sierra).

The maximum precipitation occurred from 1200 UTC 27 January to 1200 UTC 28 January (Fig. 5b). Moderate to locally heavy precipitation was observed along the Central California coast as well as inland valleys, with the heaviest precipitation ~280–390 mm falling in San Luis Obispo County. Daily records were set in Merced (36.1 and 35.8 mm), Modesto (63.2 and 23.9 mm), Paso Robles (35.3 and 74.7 mm), and Stockton (34.8 and 36.3 mm) on both 27 and 28 January, respectively, and in Fresno (45.2 mm), Hanford (37.3 mm), Santa Barbara (56.9 mm), and Santa Maria (58.4 mm) on 28 January (Fig. 5b).

Moderate to heavy snow (~1–3 ft; 30.5–91.5cm) was observed in the Northern California mountains and Sierra.

As the AR moved southward, the precipitation band also shifted southward with the maximum precipitation (211 mm) over southern Santa Barbara County (Fig. 5c). Heavy snowfall was observed over the southern Sierra (WPC 2021). This impactful precipitation event has been identified as one of the U.S. 2021 Billion-Dollar Weather and Climate Disasters, as documented in the NOAA's "Priorities for Weather Research Report" (NOAA Science Advisory Board 2021). Notably, it stands out as the sole billion-dollar disaster event in the contiguous United States during January 2021. The investigation of this event in our study will improve the sampling strategy for similar events and deepen our understanding of the predictability of such high-impact events in NWP models.

4. Results

a. Temporal sampling

The impact of mission frequency on model analyses and precipitation forecasts is assessed by analyzing the outputs of
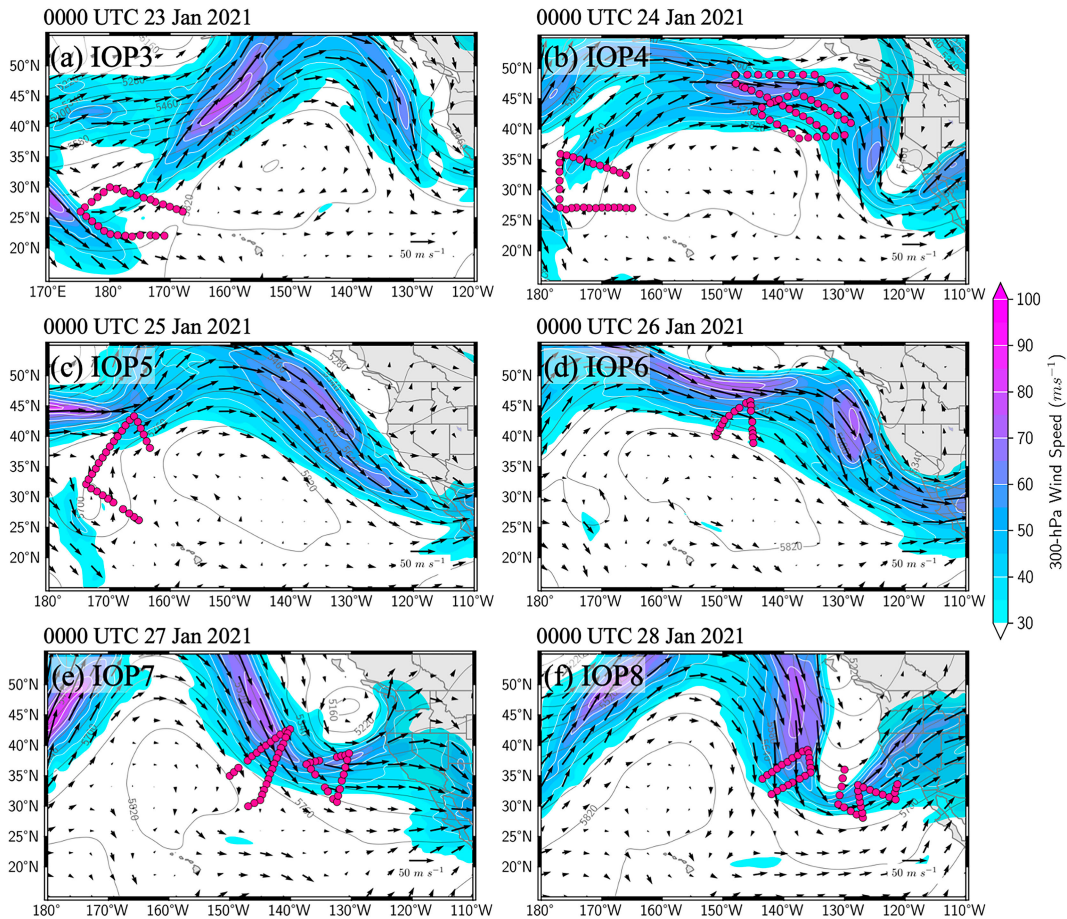
FIG. 4. An overview of upper-level systems, including geopotential height at 500 hPa (contours, m), 300-hPa wind speed (shading, m s^{-1}) and wind vectors (black arrows, m s^{-1}). Analyses are valid at the same time as in Fig. 3. Deep pink markers indicate AR Recon dropsonde locations.

Control, TS, and NoDROP experiments (Table 2 and Table S2). In total, the Control experiment that assimilates AR Recon dropsonde data from all IOPs has increased the temperature and humidity observations within the model domain (Fig. 1a) by 60.6% and 119.1%, respectively, relative to the NoDROP experiment (Table 3). About 14% more wind observations are assimilated in the Control experiment than in the NoDROP experiment, despite there being a large volume of AMV observations. However, it is worth pointing out that dropsonde data samples all-weather conditions while AMVs are often sparse below thick clouds (Velden et al. 2005; Santek et al. 2019). The TS2 experiment, which includes dropsonde observations every other day, also assimilates substantially higher amounts of temperature (24.8% more) and humidity (47.5% more) observations than the NoDROP experiments. The numbers of assimilated in situ observations in TS3 and TSsingle are comparable, which are slightly higher than that in the NoDROP experiment. The statistics presented in Table 3 indicate that incorporating dropsonde observations can significantly increase the available humidity and temperature data in regional modeling systems.

1) IMPACT ON INITIAL CONDITIONS

The impact of mission frequency on the initial conditions of the model has been analyzed for a representative IOP (i.e., IOP5 on 25 January, Fig. 6). This IOP represents the first 0000 UTC analysis when the TS2 and TS3 differ in the assimilation of AR Recon dropsonde data. Large differences between the model analysis and ERA5 are found within and around the 250-IVT-unit contour that indicates the moisture advection extending poleward from Hawaii. For all experiments, the maximum differences are around the two IVT maxima (>500 kg m^{-1} s^{-1}) over the moisture advection. For instance, the Control analysis shifts the northern moisture advection core to the east and underestimates the IVT maxima near 160°W, 45°N relative to ERA5 (Fig. 6a). Meanwhile, the Control analysis shifts the IVT center more west-southwestward near the southern leg of the G-IV flight path between 25° and 32°N (Fig. 6a).

Control and TS2 experiments both assimilate dropsondes from IOP5 and IOP3, so differences between them can be attributed to the assimilation of dropsondes from IOP4 (24 January). The largest difference (~240 IVT units) is near the southern leg of the G-IV flight for IOP4 (Fig. 6b), demonstrating that
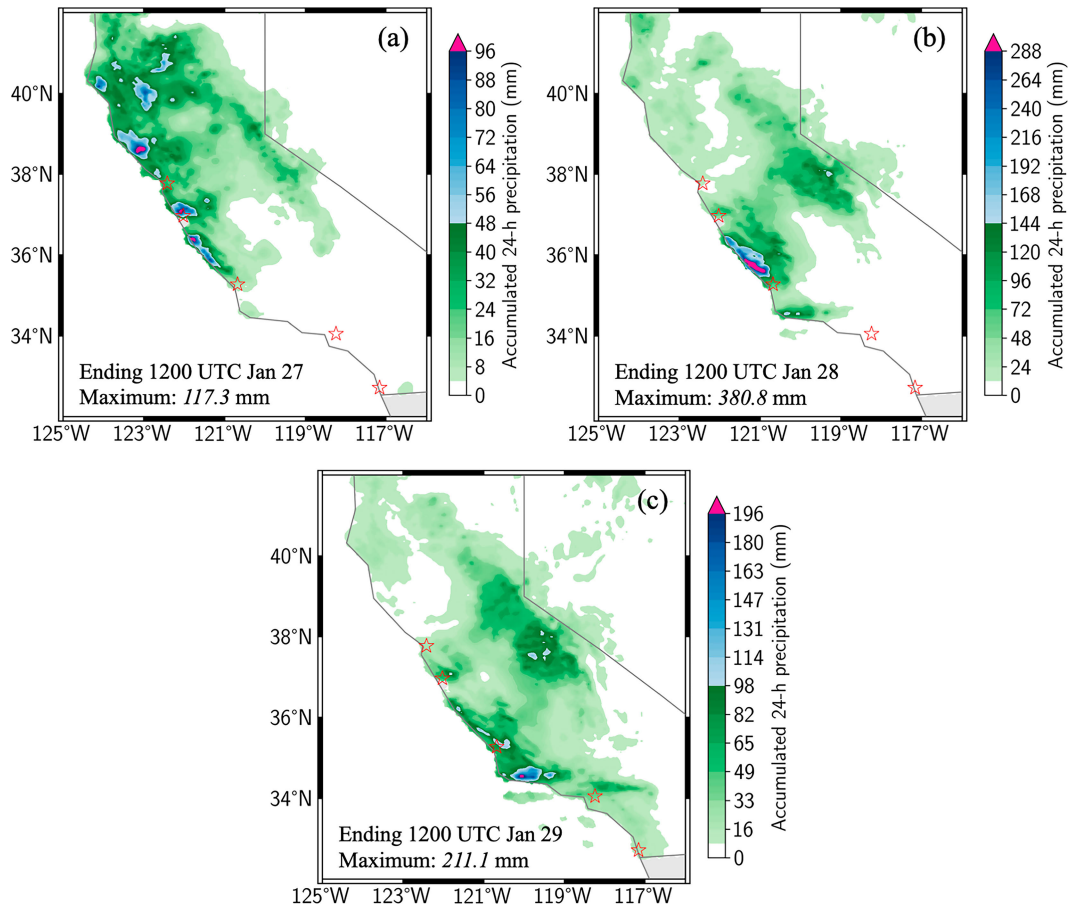
FIG. 5. Stage-IV accumulated 24-h precipitation (mm) ending at (a) 1200 UTC 27 Jan, (c) 1200 UTC 28 Jan, and (c) 1200 UTC 29 Jan 2021. Red stars on each panel from north to south denote the following five cities: San Francisco, Santa Cruz, San Luis Obispo, Los Angeles, and San Diego, respectively.

dropsondes from the previous day can significantly impact the analysis field with a cycling system. Differences in the TS2 and TS3 analyses (Fig. 6d) are completely due to the assimilation of dropsondes deployed from the G-IV flight during IOP5 in the TS2 experiment, which enhances the IVT maxima and shifts the moisture advection core between 25° and 32°N more westward. The amplitude of model differences along the G-IV flight path is comparable to the difference between model analyses and the reanalysis, suggesting that the increment provided by dropsondes

TABLE 3. Counts of assimilated temperature (T), humidity (Q), and wind observation (U, V) in the Control, TS2, TS3, TSsingle, and NoDROP experiments at 0000 UTC accumulated from 23 to 28 Jan. The number in parentheses denotes the percentage of increase compared to the NoDROP.

	T	Q	U, V
NoDROP	91 756 (0.0%)	42 280 (0.0%)	409 654 (0.0%)
TSsingle	103 137 (12.4%)	52 584 (24.3%)	421 646 (2.9%)
TS3	101 108 (10.2%)	50 289 (18.9%)	419 209 (2.3%)
TS2	114 531 (24.8%)	62 356 (47.5%)	433 466 (5.8%)
Control	147 350 (60.6%)	92 654 (119.1%)	466 966 (14.0%)

alone can be comparable to the initial condition errors in the targeted weather system (Figs. 6c,d). This finding generally is valid in other meteorological fields, such as the mean sea level pressure (MSLP) and 500-hPa geopotential height (not shown).

Since TS3 only assimilates dropsonde data from IOP3 (23 January), the differences between TS3 and NoDROP (Fig. 6f) represent the impact of assimilating IOP3 dropsondes 48 h previously and how it cycles through the system. Overall, the differences are noisier and more widespread than in Figs. 6b and 6d with the maximized values downstream of the low pressure area and the IVT core from 20° to 30°N. There are also signs of spatial shifts, including a southward shift of the IVT north of the low pressure. This further demonstrates that assimilated dropsonde data can continue to influence the analysis for the key features in a cycling system even two days later, aligning with Weissmann et al. (2011).

Differences in two baseline experiments—Control and NoDROP—represent the impact from dropsondes in the current assimilation window and from all the previous IOPs. Therefore, the difference fields (Fig. 6h) are observed around the IOP5 flight path, downstream of each previous IOP flight path, and in some areas far away from any IOP flight path
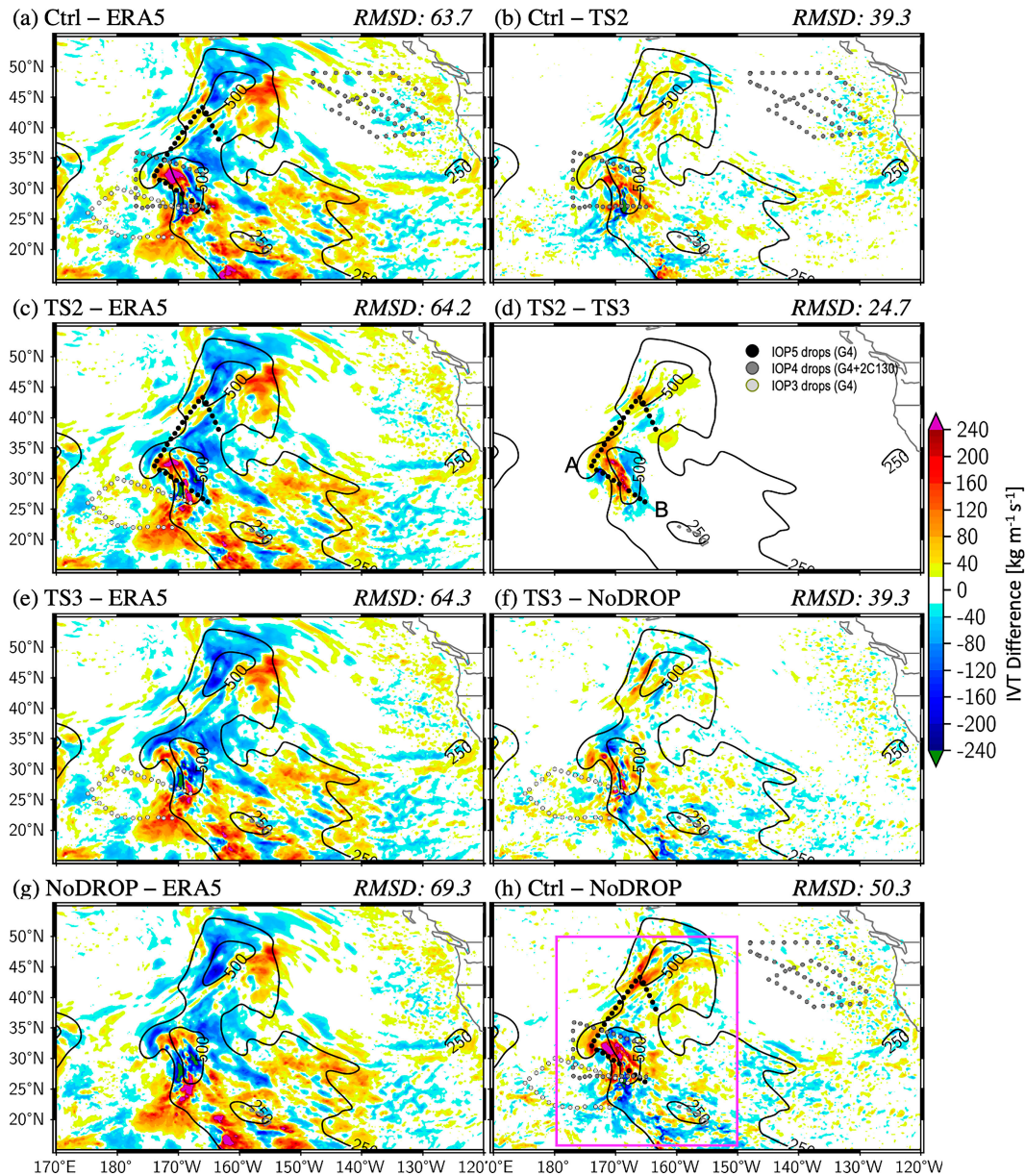
FIG. 6. Data impact on the initial condition of IVT at 0000 UTC 25 Jan 2021 (i.e., IOP5). (a),(c),(e),(g) IVT differences (shading, kg m^{-1} s^{-1}) between each experiment and the ERA5 data, arranged in descending order of AR Recon mission frequency, with (a) representing the highest frequency and (g) representing zero missions. (b),(d),(f),(h) Differences (shading, kg m^{-1} s^{-1}) between two experiments. Black contours are the analyzed IVT amplitude in ERA5 starting from 250 kg m^{-1} s^{-1}. The filled circles in (b), (d), (f), and (h) are the locations of additional dropsondes in the experiment as the minuend. The number in the top right of each pane is the root-mean-square difference (RMSD) of IVT amplitude (kg m^{-1} s^{-1}) based on the shaded difference field over a subset region [magenta box in (h)] spanning from the date line to 150°W longitude and from 15° to 50°N latitude.

(due to numerical noise in the cycling). The differences are maximized around two moisture advection IVT maxima (>500 units), suggesting the importance of dropsondes data in analyzing fields near the moisture advection cores and the sharpest IVT gradients.

The root-mean-square difference (RMSD) between the Control analysis and the ERA5 reanalysis over a subset domain

(15°–50°N, 170°W–180°) is 63.7 IVT units (Fig. 6a), ~8.1% less than that between the NoDROP analysis and the ERA5 data (69.3 IVT units, Fig. 6g). The assimilation of dropsonde data from the three IOPs results in an RMSD of 50.3 IVT units between the Control and NoDROP experiments (Fig. 6h), This value represents ~73% of the RMSD between the NoDROP and ERA5 data, demonstrating the effectiveness of AR Recon

dropsonde data in influencing model analysis within a cycled modeling system.

In the two experiments that assimilated dropsondes from IOP5, IVT differences from two experiments that did not assimilate IOP5 dropsondes reach 50% of the full amplitude along the southern leg of the G-IV flight path (Figs. 6d,h). A cross-sectional analysis along the southern flight path (Fig. 7a) from the northwest to southeast waypoints reveals the impact of dropsondes on dynamics. Significant pressure-level horizontal vapor flux amplitudes indicate locations of enhanced vapor fluxes (Fig. 7a).

Maximized vapor flux amplitudes appear near 975 hPa between 169° and 170°W, reaching 190 g m (kg s)$^{-1}$, and near waypoint A (Fig. 7a). The first maximum tilts upward on both sides, creating maxima between 700 and 800 hPa from 170° to 170.5°W and between 500 and 900 hPa near 168°W. These are associated with the moisture advection core and moisture wrapped around to the northwest of the low pressure system (Fig. 7a). Minima are observed near 900 hPa from 171° to 172.5°W with the minimum value of ∼10 units, near 750 hPa around 169°W, and between 300 and 550 hPa from 171° to 173.9°W, associated with subsiding cold, dry air from higher latitudes. Compared to ERA5, vapor flux maxima could be underestimated by more than 50%, e.g., from 168.5° to 170°W below 850-hPa level in the NoDROP run, near 168°W around 700 hPa, and near waypoint A (Fig. 7g). However, vapor flux minima are overestimated in the NoDROP run, e.g., near 900 hPa from 171° to 172.5°W.

Differences between the Control and NoDROP analyses (Fig. 7h) are opposite to those between the NoDROP analysis and the ERA5 data (Fig. 7g), suggesting that the assimilation of all dropsondes can significantly correct the model initial analyses. Here, we assume ERA5 as the ground truth. This assumption is based on the overall better validation results of ERA5 compared to other reanalysis data when verified with in situ observations. However, it should be noted that the true atmosphere state, particularly over oceans, is unknown. For example, the overestimation in the NoDROP experiment near 172°W between 900 and 1000 hPa has been significantly reduced in the Control (Figs. 7a,g) and TS2 (Fig. 7c) experiments, which assimilates data from IOP5. Underestimation of vapor flux amplitudes from 168° to 170°W below the 800-hPa pressure level has generally been improved in the Control (Figs. 7a,g,h). The moisture and wind speed cross-sectional analyses indicate that the observed improvement is mainly attributable to reducing the underestimation error of wind speed in the southerly direction within the 800–925-hPa layer. Substantial differences between the Control and NoDROP experiments can be attributed to the direct assimilation of the dropsondes for the analysis window (Fig. 7d), such as along 172°W and near 168°–169°W. Nevertheless, assimilating dropsondes from previous IOPs (e.g., IOP3) yielded additional benefits, such as near the minimum vapor flux area around 169°W between 650 and 800 hPa.

The same cross section in Fig. 7 is produced for other variables: the specific humidity (Fig. S3), horizontal wind speed (Fig. S4), and temperature (Fig. S5). Large discrepancies exist between the NoDROP analysis and the ERA5 data, especially in dynamical features like the dry, cold-air intrusion

(inversion) along 172°W. Experiments without dropsondes produce warmer and moister conditions compared to the ERA5 data (Figs. S3 and S5). Overall, dropsondes are critical for an accurate representation of fine-scale vertical structures (i.e., inversion layers) and strong gradients in wind, moisture, and vapor flux.

2) IMPACT ON FORECAST SKILL

(i) Forecasts initialized during IOP5

To better understand the impacts of mission frequency on forecast skill, we examine the impact on short-term predictions for the same IOP (i.e., IOP5) that we previously analyzed in detail for differences between model analyses. The 12-h forecast of IVT initialized at 0000 UTC 25 January is employed as a short-term prediction example (Fig. 8). At 12 h, the core of the northern portion of the moisture advection in the ERA5 data has moved eastward by approximately 6° longitude near 47°N since the initial time (Fig. 8a), while the southern core has strengthened but not propagated much since the initial time (Fig. S6a).

The investigated region is focused on the northern moisture advection (Fig. 8),[1] which is downstream of the assimilated dropsondes from G-IV missions. In the NoDROP forecast, the core of the northern portion of the tropical moisture advection and its northern quadrant are underestimated by ∼200 kg m^{-1} s^{-1} (Fig. 8g). The Control forecast improves the underestimation within the northern core of the moisture advection between 46° and 52°N in the NoDROP by up to ∼160 kg m^{-1} s^{-1} due to the assimilation of dropsondes from IOP5 (Fig. 8d) and IOP3 (Fig. 8f). The Control run predicts weaker IVT at the southwest edge of the moisture advection between 40° and 45°N than the NoDROP run does, which seems to be associated with the assimilation of data from IOPs 4 and 5. The average RMSD between the model analysis and the ERA5 data for the two experiments (Control and TS2), which assimilate dropsondes from IOP5, is 57.45 IVT units (Figs. 8a,c). Notably, this value reflects a 6.9% reduction compared to the average RMSD observed in the two experiments (TS3 and NoDROP) that do not assimilate IOP5 data (Figs. 8e,g). Results for the southern moisture advection region are mixed (see Fig. S6 and its description), likely due to the interaction between observation impacts and lateral boundary condition errors (Torn et al. 2006).

Results for the 72-h IVT forecasts show that dropsondes from all IOPs overall improve the underestimation along the coast of Central California and in the north of the inland IVT maximum (Fig. S7). The skill of landfalling IVT in the Control and TS2 experiments is higher than in the TS3 and NoDROP experiments (Fig. S7). Meanwhile, the overall underestimation of IVT along the coast and inland in model runs (Fig. S7) results in the underforecast of the precipitation in the coastal areas and over Sierra (Figs. 9a,c,e,g). Assimilating dropsondes from all IOPs improves this underforecast of the heaviest precipitation area along the central CA coast, such as from north of San Luis Obispo to Santa Cruz (Figs. 9h,g) and removes a significant underestimation (>240 mm) over San Luis Obispo

---

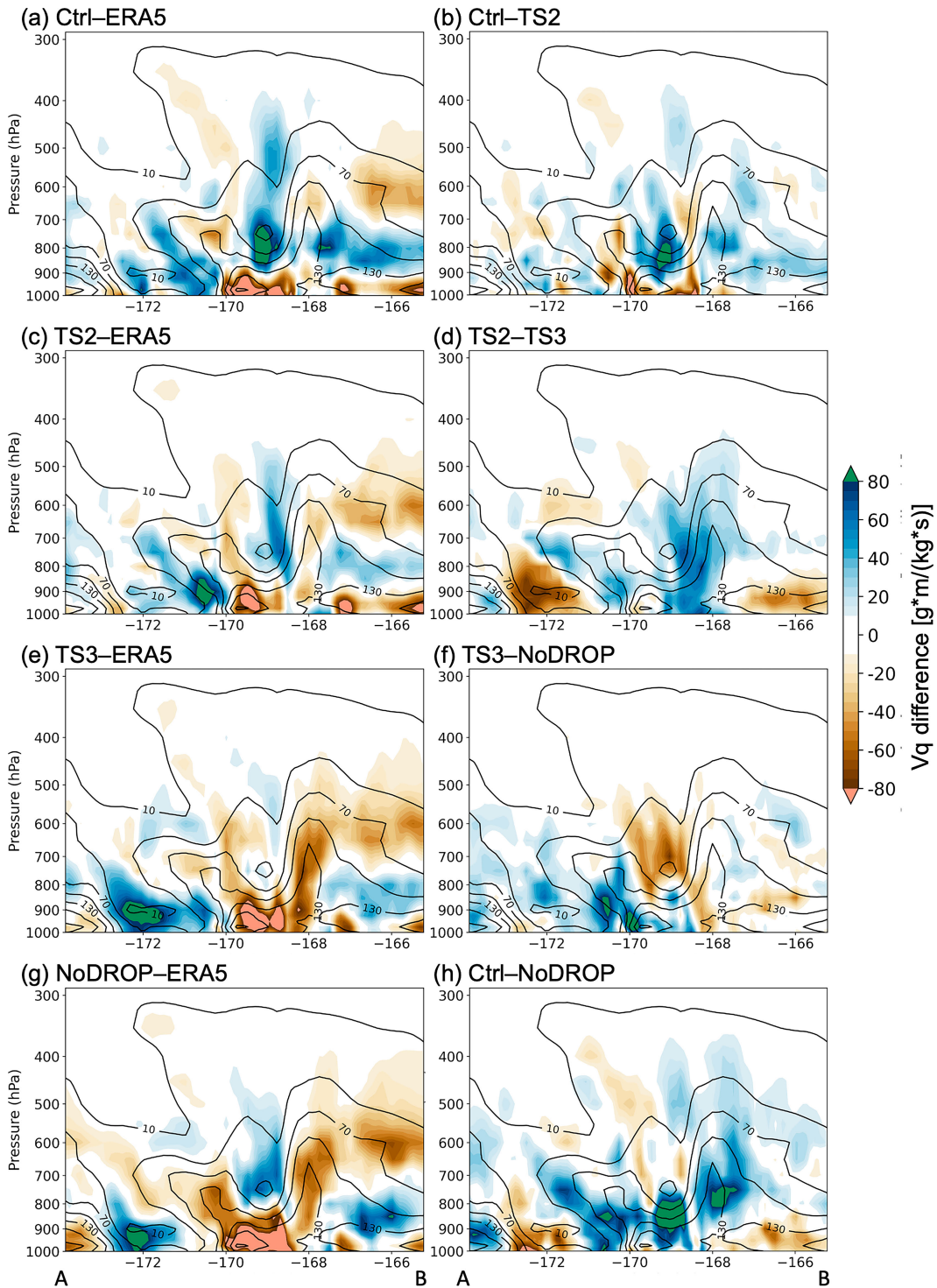[1] This region is also less affected by the lateral boundary condition errors.

Fig. 7. Vertical cross section for amplitude of the horizontal vapor flux [g m (kg s)$^{-1}$] along the flight path from A (32.12°N, 173.87°W) to B (26.17°N, 165.23°W), which are around the two G-IV waypoints labeled in Fig. 6d. (a),(c),(e),(g) The shading indicates the differences in vapor flux amplitudes between model analyses and the ERA5 data. (b),(d),(f),(h) The shading indicates differences between two experiments. Black contours for each panel represent ERA5 vapor flux. The analysis is valid at 0000 UTC 25 Jan 2021.
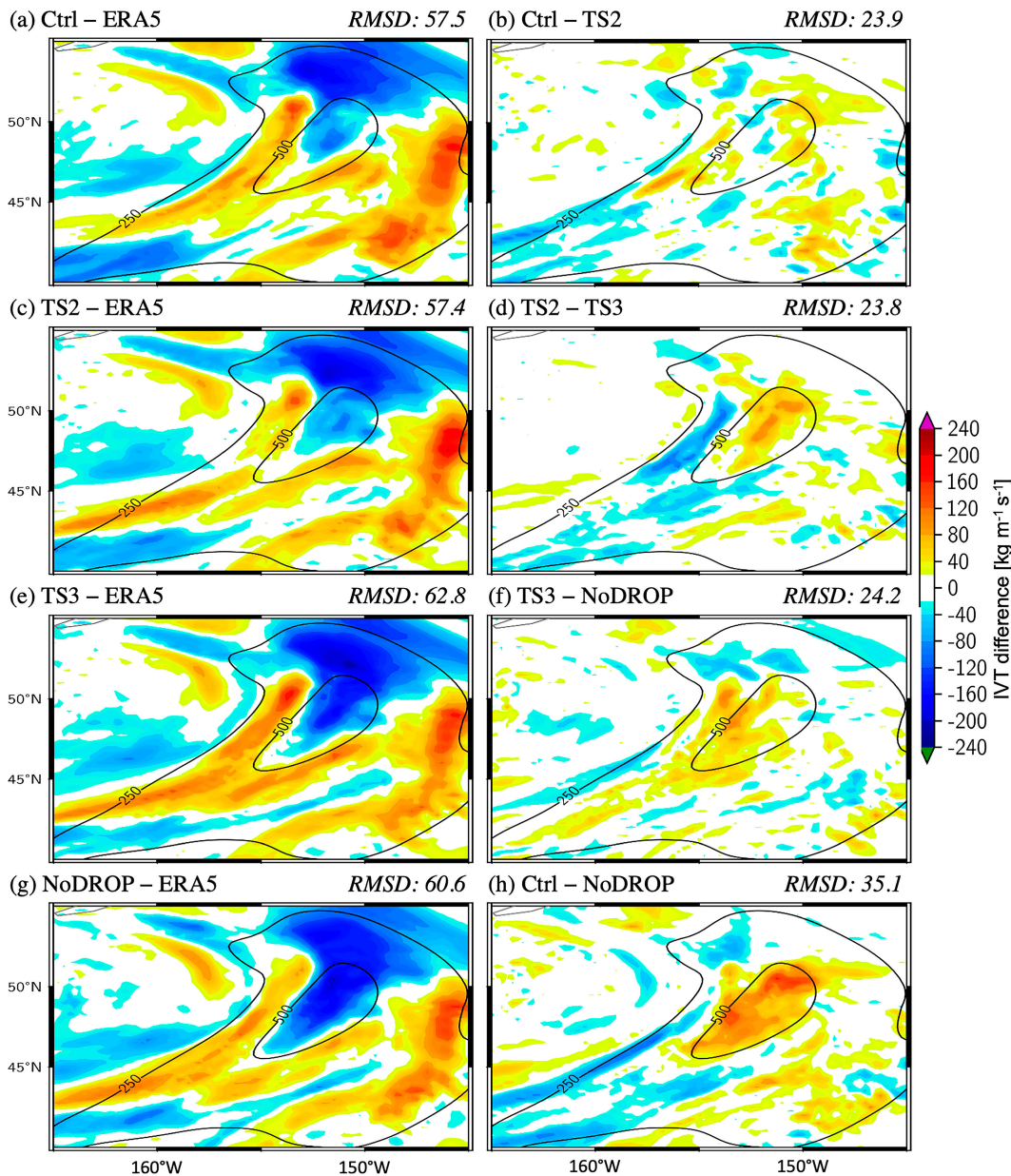
FIG. 8. As in Fig. 6, but for the 12-h IVT forecast valid at 1200 UTC 25 Jan 2021. The RMSD is calculated over the plotting domain.

County in the NoDROP run (Fig. 9g). The improved precipitation along the coast and over Sierra in the Control run is also mainly attributable to the contribution of dropsondes from IOPs 4 and 5 (Figs. 9b,d), which is consistent with improved IVT forecasts (Figs. S7b,d). The Control run reduces the RMSD for precipitation over the investigated domain in the NoDROP run by 13.6% (Figs. 9a,g) and increases the spatial correlation by 8.4%.

(ii) MET-MODE-based precipitation validation for all cycles

The main goal of AR Recon is to improve the forecast skill of precipitation that is associated with high-impact landfalling

ARs. Here we are focused on validation for record-high precipitation along the Central California coast during 1200 UTC 27–28 January. This region was also the domain of interest during operational flight planning for the sequence of the IOPs (Cobb et al. 2024). The MET-MODE tool, as described in section 2c, is used to assess the skill for precipitation and IVT.

The observed values within the heaviest precipitation object (76 mm or 3 in.), created after convolving the raw precipitation fields, show that the precipitation maxima are parallel to the Central California coast (Fig. 10a). The best match between the observed object and a forecast object is found in
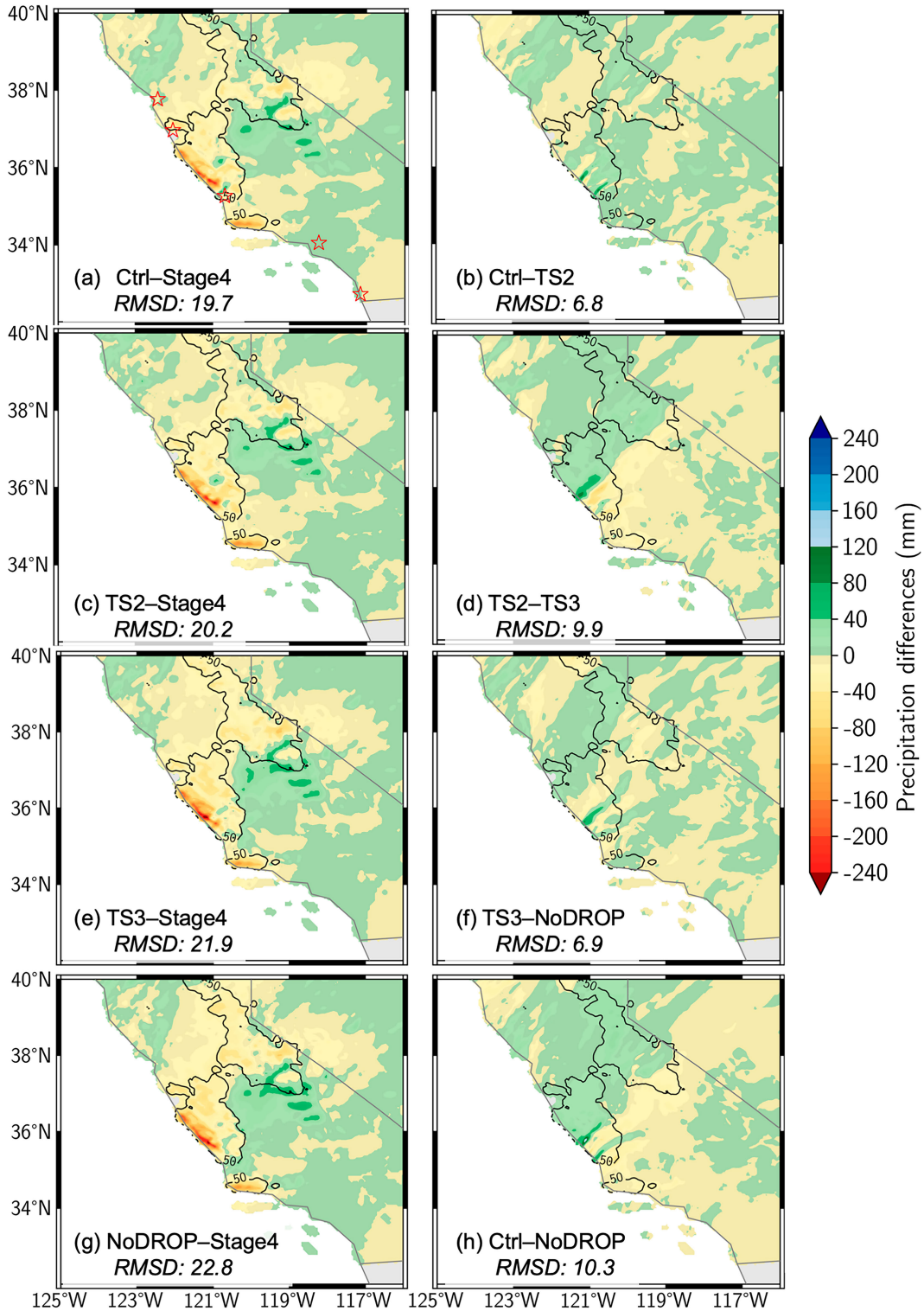
FIG. 9. As in Fig. 8, but for the 24-h accumulated precipitation (mm) from 1200 UTC 27 Jan to 1200 UTC 28 Jan. The initialization time is at 0000 UTC 25 Jan 2021. The validation data are based on the Stage-IV precipitation data. The black contour outlines the 50-mm precipitation in Stage-IV. Red stars in (a) from north to south denote the following five cities: San Francisco, Santa Cruz, San Luis Obispo, Los Angeles, and San Diego, respectively.
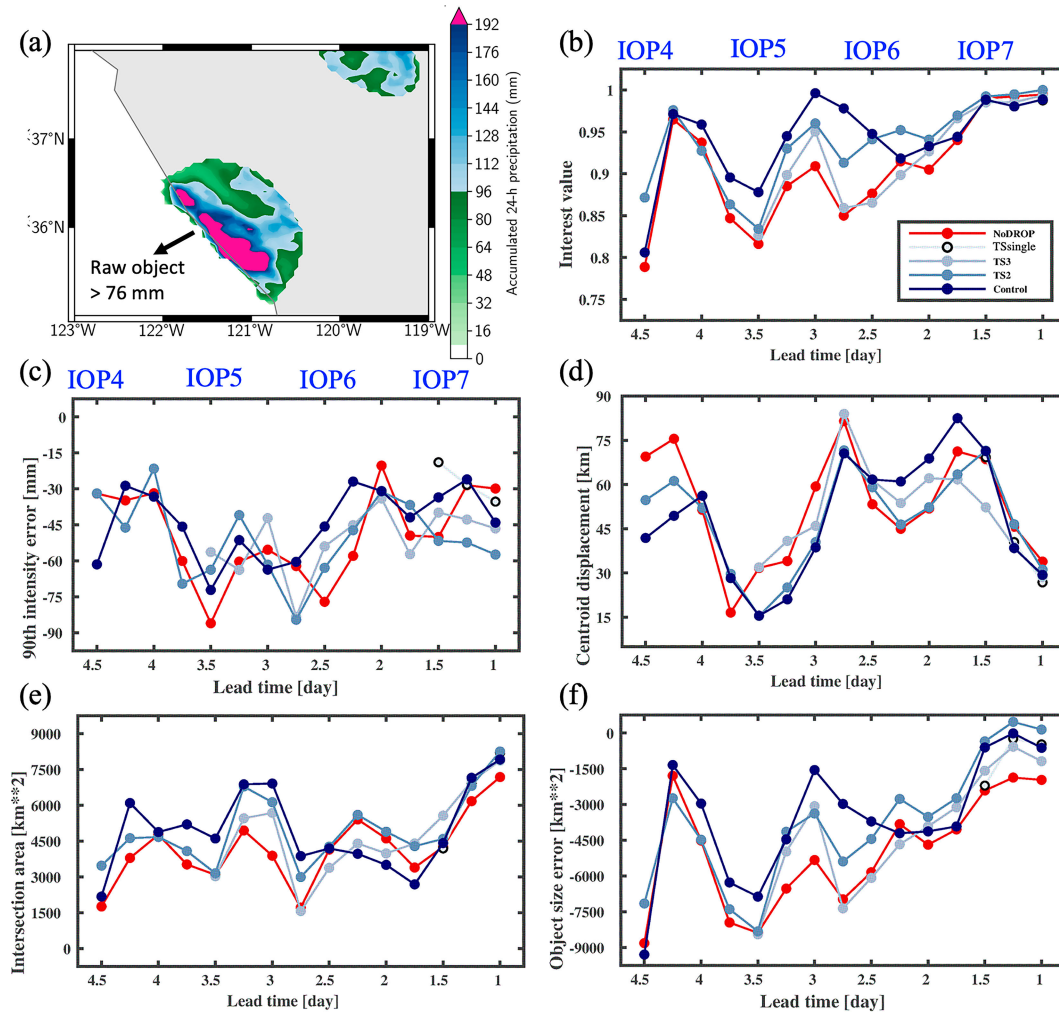
FIG. 10. (a) The MET-MODE object including raw values for accumulated 24-h precipitation greater than 76 mm using Stage-IV data from 1200 UTC 27 Jan to 1200 UTC 28 Jan. (b) The interest value as a comprehensive metric for validating the observed coastal object in (a) based on different experiments for the 24-h precipitation time window ending from a lead time of day 4.5 (IOP4, at 0000 UTC 24 Jan) to day 1 (12 h after IOP7 or at 1200 UTC 27 Jan) with a time interval of 6 h. (c) As in (b), but for the 90th percentile of the precipitation amount within the object (mm). (d) As in (b), but for the object centroid displacement (km). (e) As in (b), but for the intersection area between the observed and model forecasted objects (km^2). (f) As in (b), but for the object size errors (km^2) for the object validation. The blue text above (b) and (c) denotes the IOPs at the forecast lead time of days 4.5, 3.5, 2.5, and 1.5.

the Control run among the TS and baseline experiments from day 2.5 to day 4.5 lead times (Fig. 10b). This period includes a crucial forecast lead time (i.e., 3 days) for water management. The lowest interest value is generally seen in the NoDROP run from day 1 to day 4.5 lead times (Fig. 10b). The contrast between the Control and NoDROP experiments demonstrates an overall positive impact of assimilating dropsondes, particularly for longer than 24-h lead time. Note that positive impacts associated with the assimilation of dropsondes are also apparent in forecasts initialized at 0600, 1200, and 1800 UTC. Compared with the baseline experiments, the TS2 experiment has a slightly lower interest value from day 2.5 to day 4 while the TS3 experiment has a significantly lower interest value (Fig. 10b).

To compare the skill of different physical metrics that contributed to the total interest, we present the validation results for the key attributes of the precipitation (Figs. 10c–f). The 90th-percentile intensity errors do not show notable differences among experiments. However, the NoDROP and TS3 experiments tend to have the lowest values more frequently compared to the Control and TS2 experiments, particularly during the 0000 UTC cycles (Fig. 10c). Validation for the TSsingle run, which differs from the NoDROP run starting at day 1.5 (IOP7), shows that the direct assimilation of a large number of dropsondes during the concurrent time window (after the AR has made landfall) can improve the intensity of the precipitation during the short range. Validation results for the centroid displacement from different experiments are mixed (Fig. 10d).
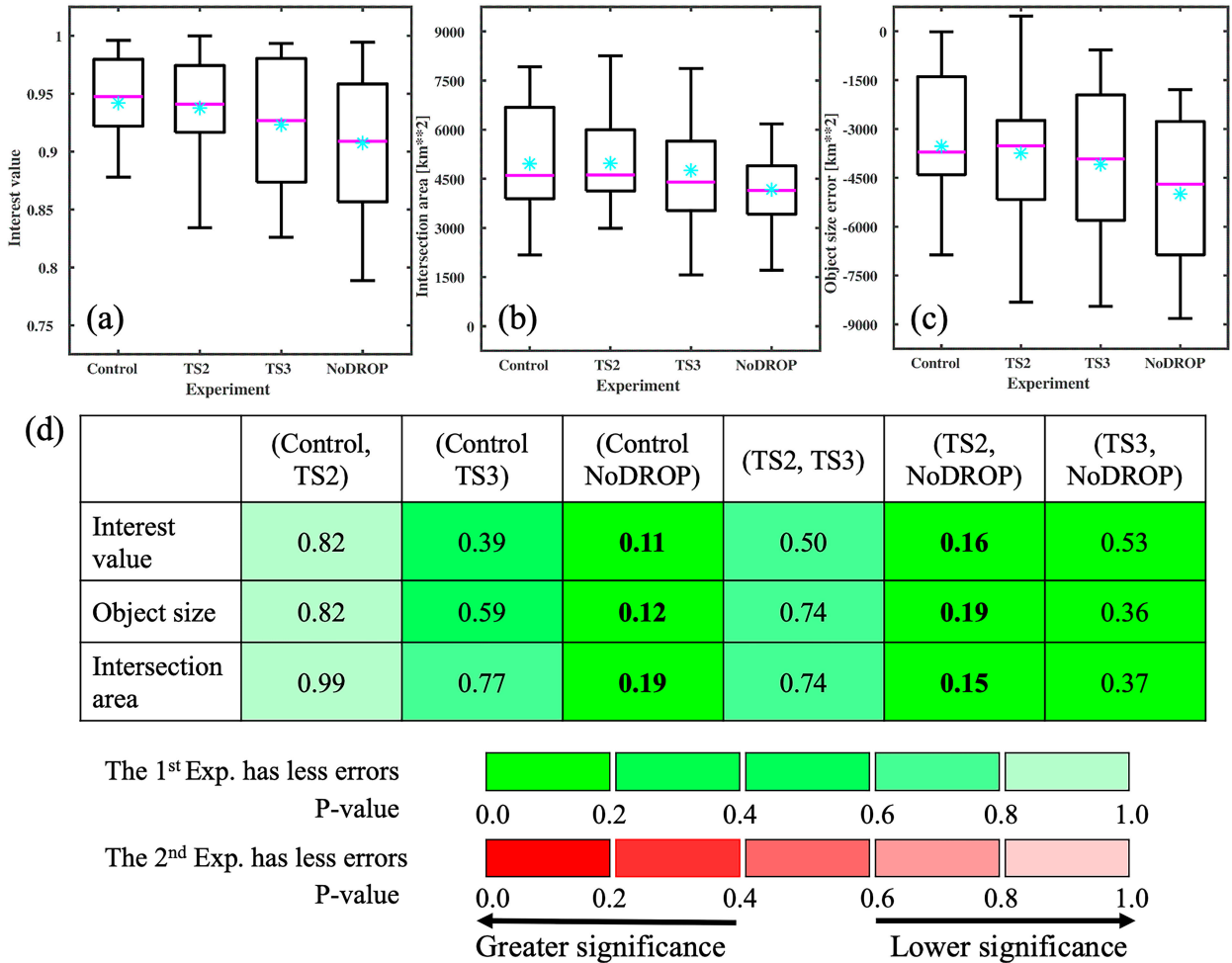
FIG. 11. Boxplot of (a) the interest value, (b) the intersection area, and (c) the object size error for the coastal object validation in Fig. 10. The boxplots are calculated by combining all 19 lead times together with the nonmatched forecast object excluded in the corresponding lead time. The bottom and the top of each box represents the 25th percentile and the 75th percentile, respectively. The magenta line in the middle of the box is the median. The cyan asterisk is the mean value of each experiment. The magenta horizontal line is the median of each data. (d) The p value, representing the degree of significance for the mean value differences between two experiments. The green shading in (d) correspond to that the first experiment in the parentheses has less errors for the three metrics while the red shades show the second experiment has less errors. Bold values in the chart of (d) show two experiments are significantly different at the 80% confidence levels.

Forecasting the spatial coverage of the heaviest precipitation is crucial for effective water management and risk mitigation, especially in areas with wildfire burn scars or antecedent saturated soil moisture. Object size errors and intersection areas represent the inaccuracies in the coverage of heaviest precipitation across models and the overlapping areas between predicted and observed objects (Figs. 10e,f). The intersection area of paired objects for the Control and TS2 experiments is higher than that for the TS3 and NoDROP experiments from day 2.75 through day 4.25 and in the short range (e.g., day 1). The Control and TS2 experiments exhibit less object size errors compared to the TS3 and NoDROP experiments from day 1 through day 4.25. The TSsingle experiment also has less errors in the short range compared to the TS3 and NoDROP experiments (e.g., day 1–1.25, Fig. 10f).

Notably, the TS3 and NoDROP experiments frequently exhibit the largest object size errors, indicating that reducing mission frequency or completely removing all of the dropsonde data in this case would degrade the forecast skill for heavy precipitation coverage. For instance, in the Control run, the object size error at a lead time of day 2.75 is about -3000 km^2, whereas the error is amplified to nearly -7000 km^2 in the TS3 and NoDROP runs, yielding a substantial missed precipitating area, approximately 60% of the observed heavy precipitation areas (11 472 km^2, Table S3).

To quantify the distribution of the overall skill in each experiment and provide the significance levels (a two-sample Student's t test) for the model differences, we present the boxplots for all lead times and differences in mean values between two model runs in Fig. 11. Skill in terms of mean interest

value, intersection area, and object size decreases as the number of IOP used decreases (Figs. 11a–c). The Control experiment has higher average skill in all three metrics than the TS2 experiment, but they are overall comparable as indicated by the high *p* values (Fig. 11d). The Control experiment also has higher skill than the TS3 experiment, and lower *p* values between them indicate larger differences between the Control and TS3 experiments than between the Control and TS2 experiments. The TS2 run has higher skill than the TS3 run, but the difference is not significant.[2] Both the Control and TS2 experiments have significantly higher skills than the NoDROP experiment, specifically for the object size. Even with significantly reduced mission frequency, the TS3 run has higher skill than the NoDROP run, indicating that experiments with as few as one full mission still have higher skill than those without dropsondes.

The MODE tool was also used to identify the IVT object and validate the forecasted object with the ERA5 object. IVT validation results show that the improved skill of precipitation, particularly the coverage of the heaviest precipitation, is associated with the improved forecast of the axis angle of the AR object along the coast and offshore and the 90th percentile of the IVT object (Fig. S8).

b. Spatial sampling

1) IMPACT ON INITIAL CONDITIONS

The analysis with an assimilation window centered at 0000 UTC 23 January (Fig. 3 and Table 2, IOP3) is employed to illustrate the impact of dropsonde spatial resolution. Since this cycle is the first to assimilate dropsonde data during the investigation period, it provides a clean comparison of the observation impact among the experiments. There are large differences between model analyses and ERA5 along the southern leg of the G-IV flight track and to the north of it (Figs. 12a,c,e,g). This region is downstream of the moisture advection west of Hawaii (Fig. 3a), with an associated upper-level trough (Fig. 4a). Impacts of the full-horizontal-resolution dropsonde data are maximized (~100 IVT units) along the southern leg, where horizontal gradients in IVT are large, while the differences along the northern leg are ~40–80 IVT units (Fig. 12h). These impacts exhibit alternating positive and negative patterns at a horizontal scale of hundreds of kilometers (Figs. 12b,h). Impacts of reduced-resolution dropsondes, such as in the SS5 and SS3 experiments, are spatially broad and quasi-Gaussian (Figs. 12d,f), which is a reflection of the background error covariance with sparse observations.

A west–east cross section that connects two waypoints on the southern leg of the G-IV flight shows that model analyses exhibit stronger vapor transport between 800 and 925 hPa characterized by large vertical vapor flux gradients (Figs. S9a,c,e,g). Model analyses tend to underestimate the vapor flux amplitude near the maximum points at various levels. The assimilation of full-horizontal-resolution dropsondes overall reduces the discrepancy between the NoDROP analysis and the ERA5 data, especially in regions with sharp horizontal and vertical gradients

(Fig. S9h). The impacts of reduced horizontal resolutions on the vertical structure are more homogeneous due to the sparsity of observations, whereas the inclusion of all of the dropsondes significantly improves error correction and captures intricate structure in the results for IOP3 (Figs. S9b,d,f,h).

2) IMPACT ON FORECAST SKILL

The impact of the observation sampling on forecast skill is first illustrated in IOP3. During operational flight planning, ensemble sensitivity (Ancell and Hakim 2007; Torn and Hakim 2008; Chang et al. 2013; Zheng et al. 2013; Hill et al. 2020) and adjoint sensitivity (Doyle et al. 2014; Reynolds et al. 2019; Doyle et al. 2019) tools were applied to inform the design of flight tracks (Cobb et al. 2024). For this IOP, one forecast metric used in the ensemble sensitivity is an MSLP metric over the north portion of the Kona low (Otkin and Martin 2004) valid at 0000 UTC 24 January (Fig. 3b). Therefore, we verified the forecasts valid at 0000 UTC 24 January in the domain of 25°–40°N, 165°W–180°, which is in the north portion of the Kona low.

The 24-h model forecasts exhibit positive error in the MSLP (Figs. 13a,c,e,g), indicating a weaker Kona low. This is qualitatively associated with the underestimation of the IVT amplitude in the northwestern side (near date line) of the TME. Meanwhile, the southwestern edge of the verification box is slightly underestimated, suggesting a stronger pressure gradient over 25°–30°N, 170°W–180° (e.g., Fig. 13g). The positive error in MSLP is larger in the NoDROP and SS5 experiments than in the SS3 and Control experiments (Fig. 13), suggesting that assimilating higher-horizontal-resolution data improves the skill of the Kona low forecast.

To compare the skill in different SS experiments for the precipitation forecasts, we verify the same MODE object for the heaviest coastal precipitation used in Fig. 10a from lead times of day 4.5 to day 1 (Fig. 14). Among the Control, SS3, and SS5 experiments, the Control run that assimilates full-horizontal-resolution data shows the highest skill in the interest value, the intersection, and the object size for lead times > 1.5 days (Figs. 14b,e,f). The skill in the Control experiment is significantly higher than that in the SS5 experiment for the intersection area (Fig. 15d). In contrast, the SS5 experiment, which increases the dropsonde spacing to 5 times, shows the lowest skill based on the interest value, the intersection and object size metrics specifically for lead times > 2 days. When considering all lead times, there is an overall decreasing trend in skill for interest value, intersection area, and object size as the dropsonde spacing increases (Figs. 15a–c). It is noteworthy that the SS3 experiment shows comparable skill with Control for the mean value of the intersection area (Fig. 15b).

In addition to the sensitivity experiments for dropsonde horizontal spacing, we conducted the SS_G4 and SS_C130 experiments to compare the impacts of dropsondes deployed from NOAA G-IV aircraft and the AF C-130 aircraft, respectively. The G-IV aircraft is designed for high-altitude and long-range travel, with a maximum altitude of ~13 500 m (~150 hPa), while the C-130 aircraft is designed for short takeoff and landing operations and typically flies at lower

---

[2] Note that the sample size for the box plot is limited to 15–17, making the assessment of statistical significance challenging.
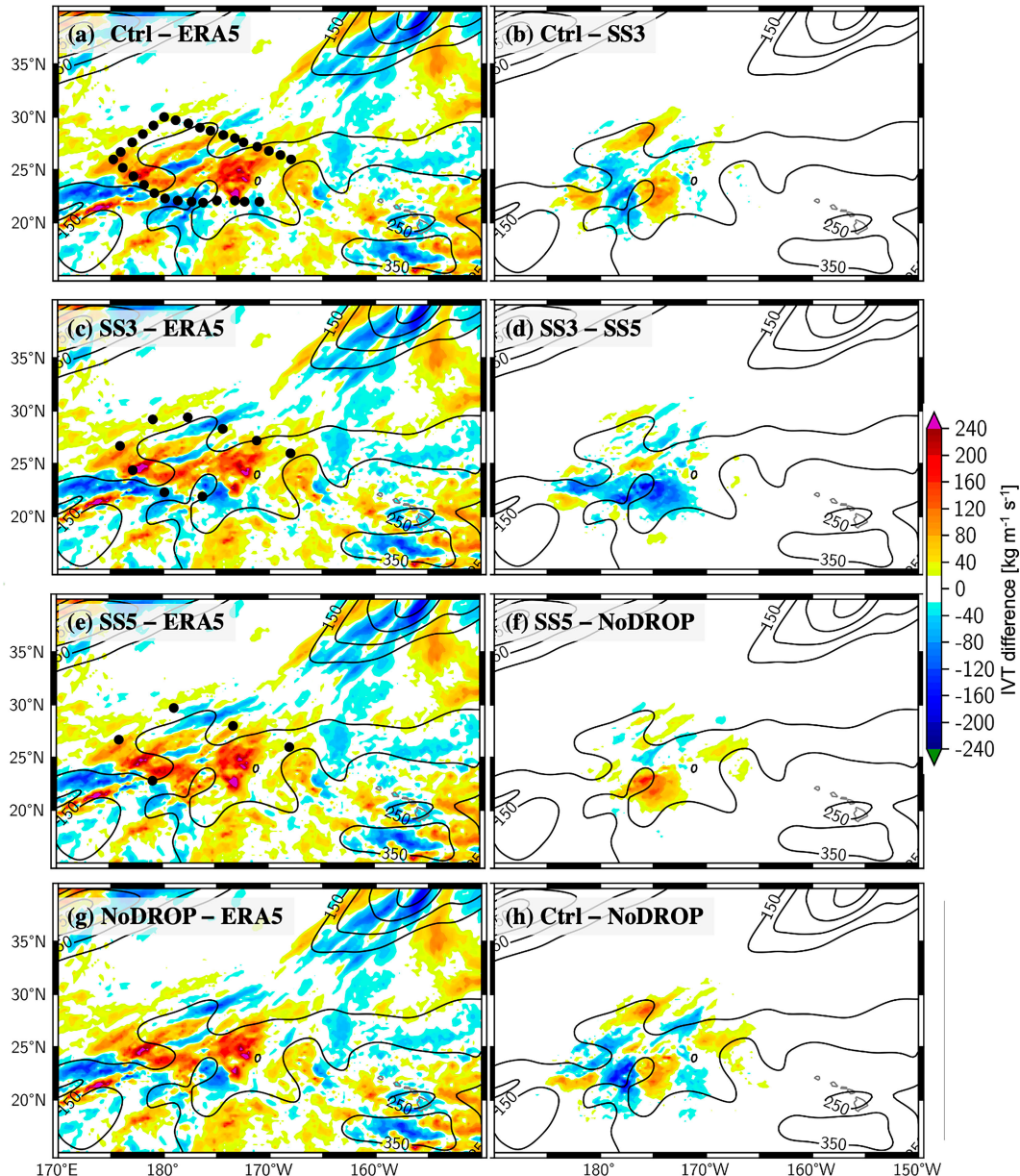
FIG. 12. As in Fig. 6, but for the SS experiments for the analysis time of 0000 UTC 23 Jan. (a),(c),(e),(g) Arranged in descending order based on the horizontal resolution of AR Recon dropsondes, with (a) representing the inclusion of full dropsonde spatial resolution and (g) representing zero dropsondes. (b) The IVT differences (shading, kg m^{-1} s^{-1}) between Control and SS3. (d),(f),(h) As in (b),but for the differences between SS3 and SS5, SS5 and NoDROP, and Control and NoDROP, respectively. Black contours are the analyzed IVT amplitude (kg m^{-1} s^{-1}) in ERA5 starting from 150 IVT units with an increase of 100 units. Filled black circles in (a), (c), and (e) indicate the locations of dropsondes assimilated during the analysis window centered at 0000 UTC 23 Jan for the Control, SS3, and SS5 experiments, respectively.

altitudes of ~8000 m (~325 hPa). The skill of SS_G4 is overall higher than that of SS_C130 in terms of interest value, object size, and intersection area particularly during days 1–4 lead times (Figs. 14b,e,f). However, it is noteworthy that this comparison is unfair to the C-130 because the C-130 aircraft were deployed for IOPs 4, 7, and 8 only, while G-IV aircraft were sent out during all six IOPs. By the time of the start of

the heaviest precipitation, the SS_C130 run had only assimilated C-130 dropsonde data in IOP4 and IOP7, with the former sampling a decaying AR that made landfall in the Pacific Northwest instead of California. Nevertheless, comparing these two experiments shows that the G-IV aircraft, which flies higher, covers longer distance, and followed the moisture advection as a precursor of the landfalling AR, and the AR itself, is critical to
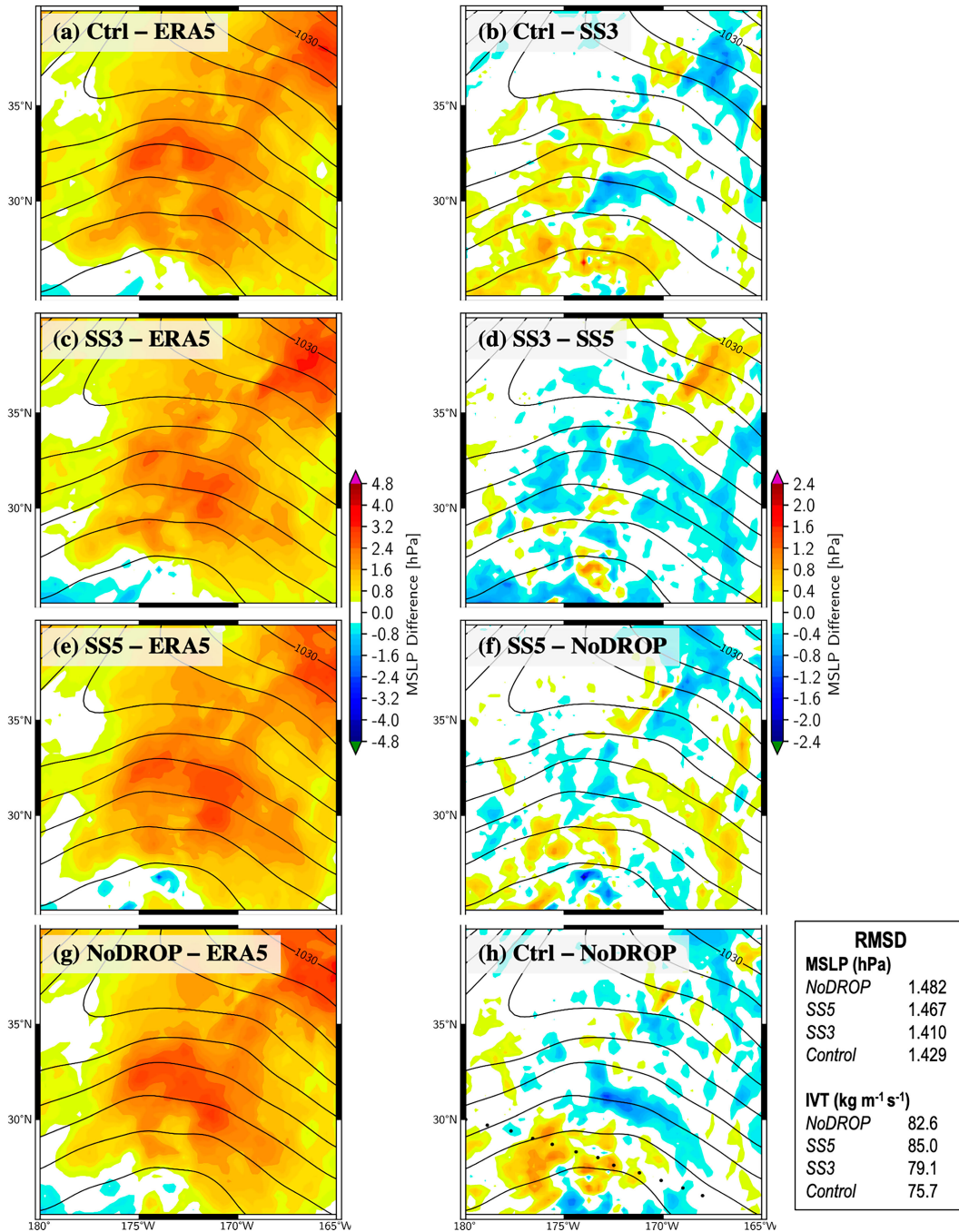
FIG. 13. As in Fig. 12, but for the differences in MSLP in the forecast valid at 0000 UTC 24 Jan. The forecasts are initialized at 0000 UTC 23 Jan. The text box on the bottom right is a summary of the RMSD between each experiment and the ERA5 data for MSLP and IVT based on the domain of 25°–40°N, 165°W–180°.

improving skill in predicting the heaviest precipitation in California. In fact, the average skill of SS_G4 is comparable to that of Control (Fig. 15) in the three metrics. However, the skill of SS_G4 has more outliers and is not as stable as in the Control run, indicating that having data from both aircraft benefits the forecast.

c. Comparison of high-vertical-resolution and reduced-vertical-level profiles

As of this writing, many operational models, such as the NCEP GFS, only assimilate dropsonde data at mandatory and significance levels (Zheng et al. 2021b). The ManSig experiment
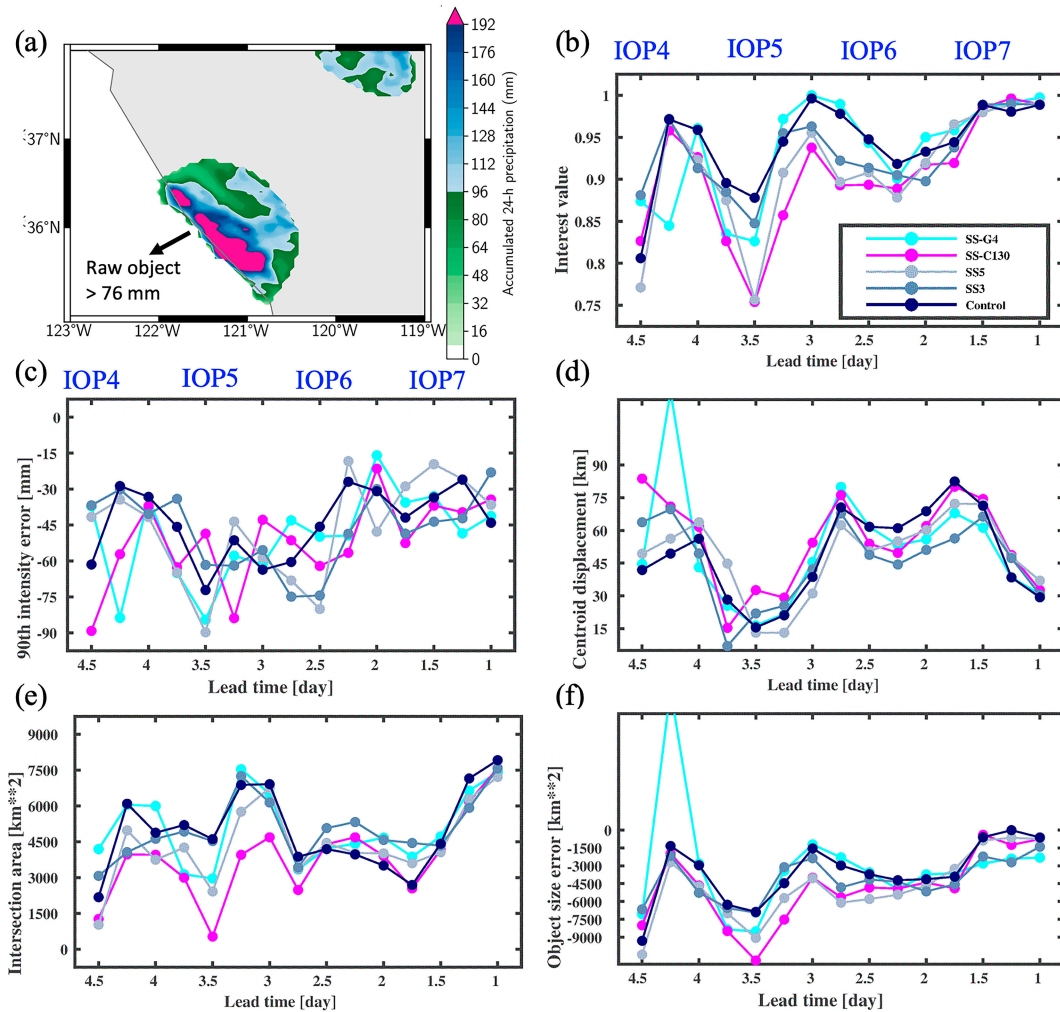
FIG. 14. As in Fig. 10, but for the Control, SS3, SS5, SS_C130, and SS_G4 experiments.

is conducted to compare the skill of assimilating reduced-level dropsonde data with that of assimilating the higher-resolution dropsonde data (i.e., the superobbed dropsonde data in the Control run). Thus, comparisons of the Control and ManSig experiments test the impact of using high-vertical-resolution data (Control) versus lower-vertical-resolution data (ManSig).

The same cross section for IOP3 (Fig. S9) is made to compare the initial conditions of the ManSig and Control runs (Fig. 16). The differences in horizontal vapor flux amplitude between the ManSig analysis and the ERA5 data resemble other experiments (e.g., Fig. S9a) with underestimation in the maxima and overestimation in the strong vertical gradients (Fig. 16a). The differences between the Control and ManSig analyses are most pronounced in regions with strong vertical gradients, such as between −182°W (i.e., 178°E) and −179°W, and along −175° and 172.5°W (Fig. 16b). To disentangle the impacts on wind and moisture, the same cross section for wind speed and specific humidity is also shown in Figs. 16c,d and 16e,f. The ManSig run underestimates the wind speed below the upper-level jet roughly between 300 and 450 hPa on

the western half of the cross section and between 250 and 400 hPa on the eastern half (Fig. 16c). Differences between the Control and ManSig experiments show a layer of positive values under the upper-level jet (Fig. 16d), demonstrating that the higher-vertical-resolution observations help to better represent the downward extension of the upper-level jet.

The discrepancy in specific humidity between the ManSig experiment and the ERA5 data is primarily an overestimation of the moisture in the dry intrusion regions, such as between −182° and −175°W from 600 to 900 hPa and along −173°W from 600 to 700 hPa (Fig. 16e). Impacts of the vertical resolution are most pronounced in regions with sharp vertical moisture gradients (Fig. 16f).

The largest differences in IVT between the Control and ManSig analyses are along the southern leg of the G-IV flight (Fig. S10c). A secondary region with significant differences is along the northern leg. Note that the discrepancy between the Control and ERA5 data is larger than that between the ManSig and ERA5 data near the southwestern waypoint of the flight path (Figs. S10a,b). Results for the IWV differences
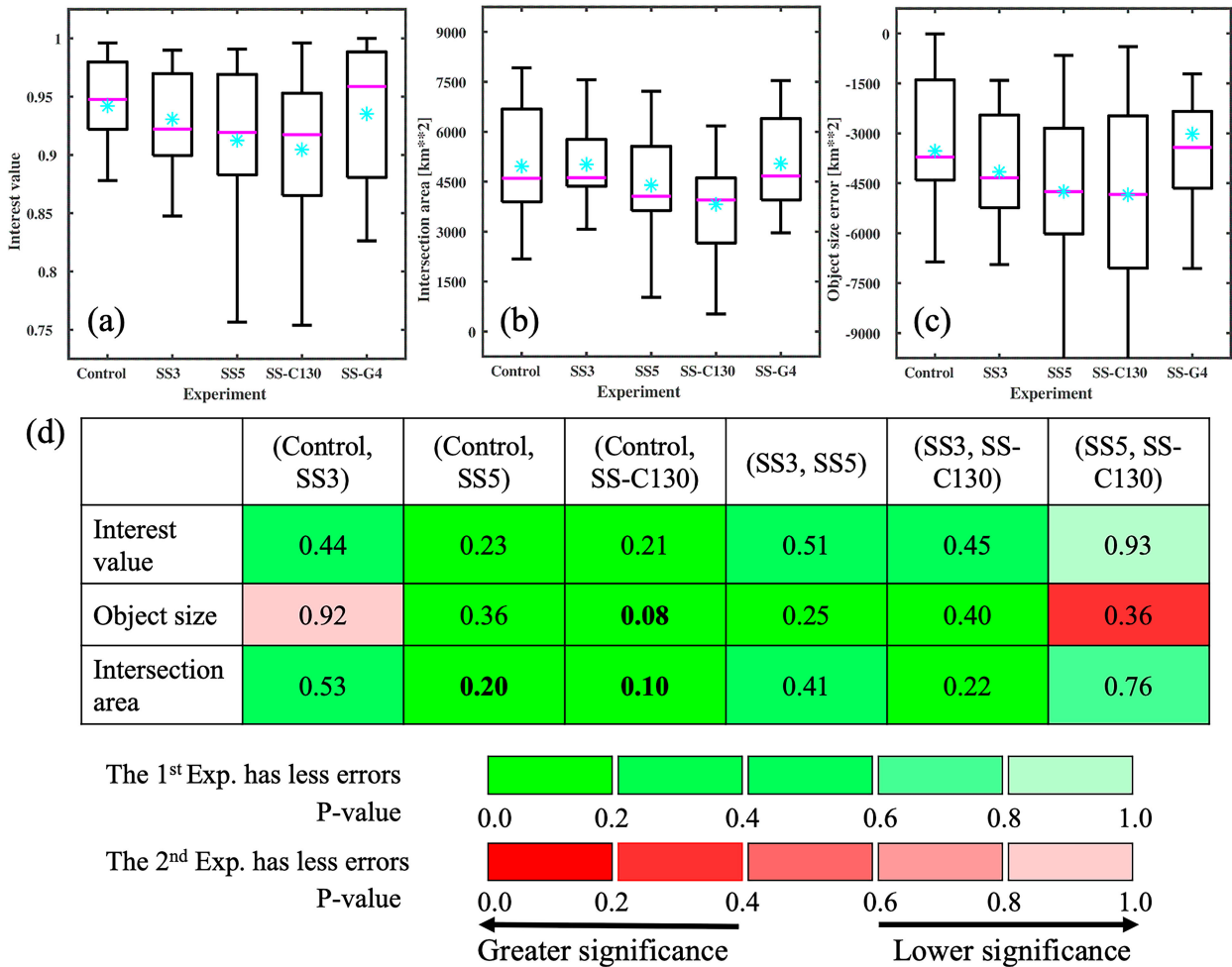
FIG. 15. As in Fig. 11, but for the SS experiments.

(Fig. S10d) show that the wind components dominate the IVT difference on the western side of the flight path between ~175°E and 180°, while the moisture components dominate the eastern side between 180° and 170°W from 20° to 28°N, where the elongated IWV plume is maximized.

Overall, the skill in the Control and ManSig experiments for the investigated period are comparable, and none of the metrics show statistically significant differences between the two experiments (Fig. 17, Fig. S11). Both experiments exhibit significant improvement over NoDROP, underscoring once again the importance of assimilating AR Recon dropsondes in accurately forecasting the precipitation object size and intersection area. The Control experiment show slightly higher skill in precipitation intensity than the ManSig experiment for the shorter lead times (e.g., days 1–2.5, Fig. 17c, Fig. S11c).

5. Discussion and conclusions

In this paper, we employ data denial experiments to explore the impacts of AR Recon mission frequency and dropsonde spatial resolution on regional model analyses and forecasts of a

2021 high-impact AR event in California. The event took place during a 6-day intensive observing period (IOP) of 2021 AR Recon, triggering heavy precipitation in both coastal California and the Sierra Nevada.

Experiments are conducted to represent scenarios of various flight missions and spatial resolutions during parallel week-long cycled simulations with the West-WRF and 4DEnVar system. Overall, the results from this case study indicate that:

- Dropsondes improved the representation of ARs in the model analyses when using ERA5 as the ground truth, especially near sharp horizontal and vertical gradients, such as the dry intrusion and the inversion layer. The benefits in model analyses are transferred to positive impacts on the forecast skill of ARs and QPF, particularly for lead times >1 day. This finding is consistent with the results from Reynolds et al. (2019), in which they found that the optimal perturbations of moisture for ARs occur in regions where the humidity gradient is large, acting to fill the drier regions, rather than in regions where the amount of humidity is greatest.
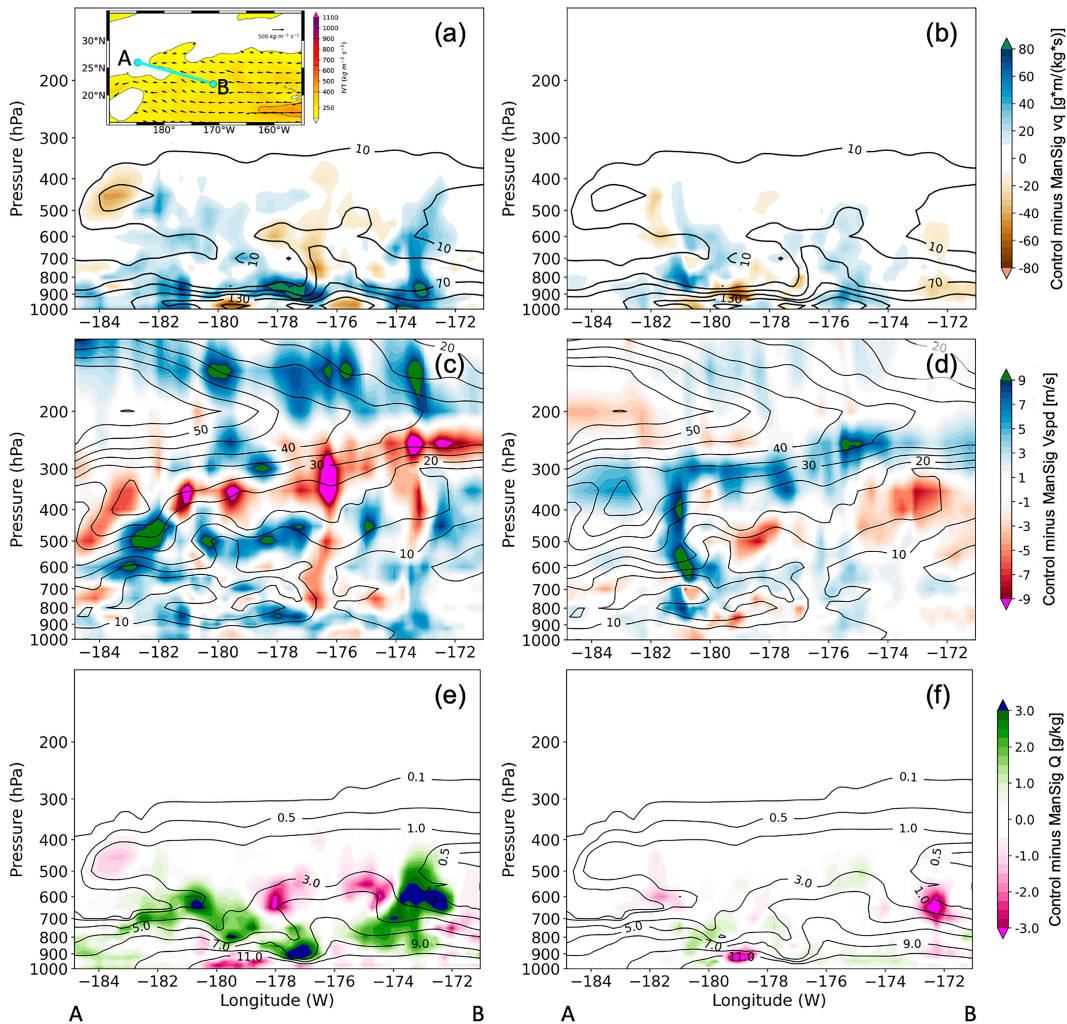
FIG. 16. (left) Difference between the ManSig analysis and the ERA5 data (shaded) and (right) difference between the Control and ManSig analyses. (a),(b) Vapor flux amplitude [kg m (kg s)$^{-1}$]; (c),(d) wind speed (m s^{-1}); and (e),(f) specific humidity (g kg^{-1}). The cross section is from A (25.99°N, 184.83°W) to B (22.03°N, 171.05°W). The analysis is valid at 0000 UTC 23 Jan 2021.

- Reduced mission frequency resulted in degraded skill in forecasting the heaviest precipitation coverage. QPF skill was significantly higher for the daily and every other day mission scenarios than for the one-mission-every-three-days and no-flight scenarios.
- Reduced dropsonde horizontal spatial resolution overall degraded forecast skill. The scenario with the dropsonde spacing increased to 5 times the original exhibits the worst skill. Increasing the dropsonde horizontal spatial resolution reduces phase errors that could exist near the sharp horizontal gradients.
- The inclusion of two types of aircraft (G-IV and C-130s), sampling different regions, is an effective strategy to enable the benefits of missions on consecutive way. G-IV samples weather features further upstream and at higher altitudes, allowing more time for observations to improve future forecasts via better background forecasts.

- Assimilating superobbed high-resolution dropsondes and only assimilating data at mandatory and significant levels show similar average skill. However, the former shows slightly higher skill in the precipitation intensity than the latter.

This study suggests some promising guidance for flight planning during future operational AR Recon missions. Results indicate that flights on consecutive days benefit the forecast more than single flights, consistent with Stone et al. (2020) and Zheng et al. (2021b). Moreover, results suggest that future missions should maintain current dropsonde spacing, particularly along the strong horizontal gradients of wind, moisture, and IVT in ARs, consistent with the adjoint results of Reynolds et al. (2019) and Doyle et al. (2019). Last, in addition to the AF C-130 aircraft that are often used to sample the AR, the NOAA G-IV brings the capability of sampling the full depth of the troposphere and the jet stream.
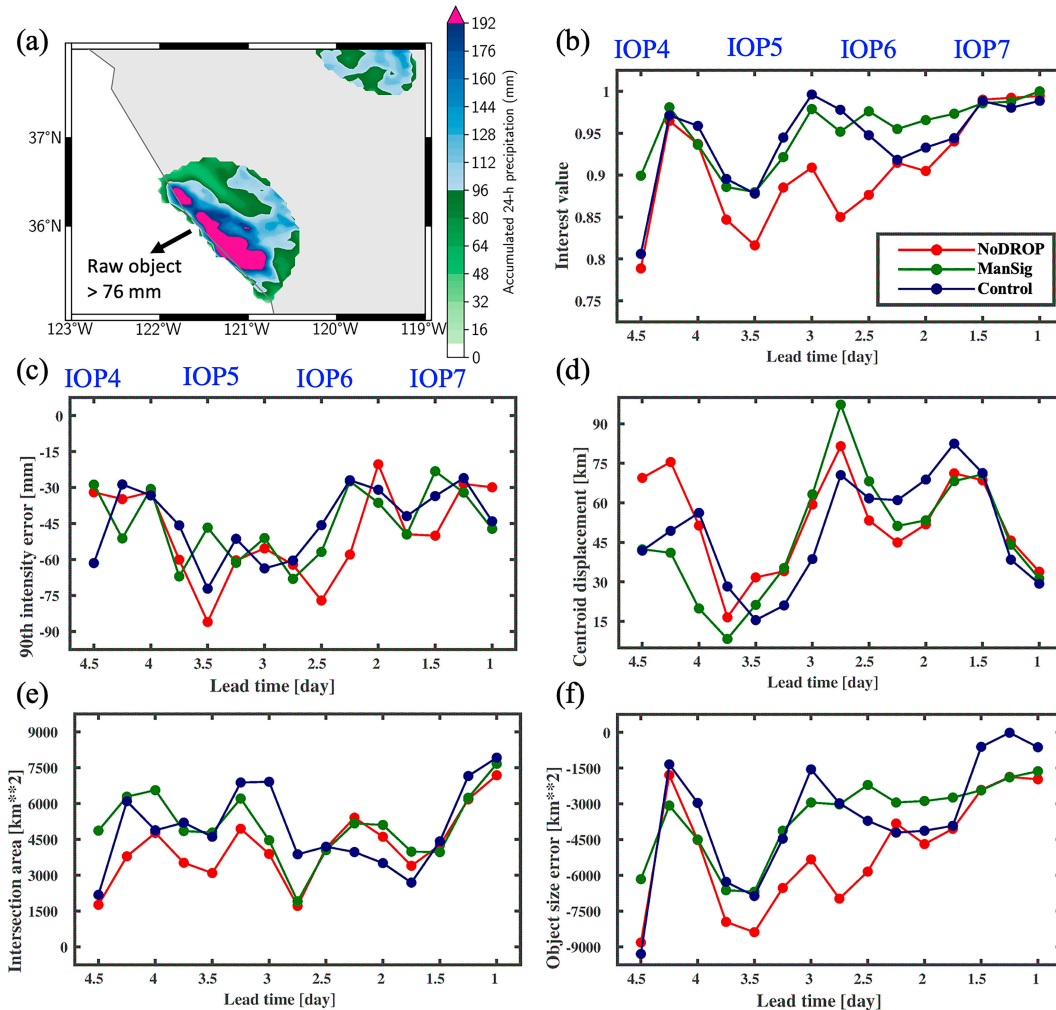
FIG. 17. As in Fig. 10, but for ManSig. Control and NoDROP results are included for comparison.

A caveat of this study is that, except for the Control run, all other experiments only reduced the resolution and/or mission frequency and did not increase them due to the lack of flights at higher frequency (e.g., every 12 h) and denser dropsondes (e.g., ~45 km). Therefore, higher-observation-resolution experiments are recommended if future flights increase the frequency and dropsonde spatial resolution. It is also noteworthy that this is a case study, and the results might change with more cases, specifically cases with different flow regimes (Majumdar et al. 2010). Future investigation will focus on the impacts on an operational global modeling system, such as the NCEP GFS. In addition, future work will also investigate other long-sequence flights during different weather regimes, such as the zonal, fast-moving atmospheric flow during a high-AR-activity period from 6 to 18 January 2023.

Users' page (https://dtcenter.org/community-code/model-evaluation-tools-met).

REFERENCES

Aberson, S. D., 2008: Large forecast degradations due to synoptic surveillance during the 2004 and 2005 hurricane seasons. *Mon. Wea. Rev.*, **136**, 3138–3150, https://doi.org/10.1175/2007MWR2192.1.

——, 2010: 10 years of hurricane synoptic surveillance (1997–2006). *Mon. Wea. Rev.*, **138**, 1536–1549, https://doi.org/10.1175/2009MWR3090.1.

Ancell, B., and G. J. Hakim, 2007: Comparing adjoint- and ensemble-sensitivity analysis with applications to observation targeting. *Mon. Wea. Rev.*, **135**, 4117–4134, https://doi.org/10.1175/2007MWR1904.1.

Bergot, T., 2001: Influence of the assimilation scheme on the efficiency of adaptive observations. *Quart. J. Roy. Meteor. Soc.*, **127**, 635–660, https://doi.org/10.1002/qj.49712757219.

Brown, B., and Coauthors, 2021: The Model Evaluation Tools (MET): More than a decade of community-supported forecast verification. *Bull. Amer. Meteor. Soc.*, **102**, E782–E807, https://doi.org/10.1175/BAMS-D-19-0093.1.

Burpee, R. W., J. L. Franklin, S. J. Lord, R. E. Tuleya, and S. D. Aberson, 1996: The impact of omega dropwindsondes on operational hurricane track forecast models. *Bull. Amer. Meteor. Soc.*, **77**, 925–934, https://doi.org/10.1175/1520-0477(1996)077<0925:TIOODO>2.0.CO;2.

Chang, E. K. M., M. Zheng, and K. Raeder, 2013: Medium-range ensemble sensitivity analysis of two extreme Pacific extratropical cyclones. *Mon. Wea. Rev.*, **141**, 211–231, https://doi.org/10.1175/MWR-D-11-00304.1.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Cobb, A., L. Delle Monache, F. Cannon, and F. M. Ralph, 2021: Representation of dropsonde-observed atmospheric river conditions in reanalyses. *Geophys. Res. Lett.*, **48**, e2021GL093357, https://doi.org/10.1029/2021GL093357.

——, and Coauthors, 2024: Atmospheric river reconnaissance 2021: A review. *Wea. Forecasting*, https://doi.org/10.1175/WAF-D-21-0164.1, in press.

Davis, C. A., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, https://doi.org/10.1175/MWR3145.1.

——, ——, ——, and J. Halley-Gotway, 2009: The Method for Object-based Diagnostic Evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC spring program. *Wea. Forecasting*, **24**, 1252–1267, https://doi.org/10.1175/2009WAF2222241.1.

DeHaan, L. L., A. C. Martin, R. R. Weihs, L. Delle Monache, and F. M. Ralph, 2021: Object-based verification of atmospheric river predictions in the northeast Pacific. *Wea. Forecasting*, **36**, 1575–1587, https://doi.org/10.1175/WAF-D-20-0236.1.

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

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

Doyle, J. D., C. Amerault, C. A. Reynolds, and P. A. Reinecke, 2014: Initial condition sensitivity and predictability of a severe extratropical cyclone using a moist adjoint. *Mon. Wea. Rev.*, **142**, 320–342, https://doi.org/10.1175/MWR-D-13-00201.1.

——, C. A. Reynolds, and C. Amerault, 2019: Adjoint sensitivity analysis of high-impact extratropical cyclones. *Mon. Wea. Rev.*, **147**, 4511–4532, https://doi.org/10.1175/MWR-D-19-0055.1.

Du, J., 2011: NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data, version 1.0. UCAR/NCAR–Earth Observing Laboratory, accessed 1 June 2023, https://doi.org/10.5065/D6PG1QDD.

Feng, J., and X. Wang, 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). *Mon. Wea. Rev.*, **147**, 3069–3089, https://doi.org/10.1175/MWR-D-18-0305.1.

Grell, G. A., and S. R. Freitas, 2014: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos. Chem. Phys.*, **14**, 5233–5250, https://doi.org/10.5194/acp-14-5233-2014.

Guan, B., N. P. Molotch, D. E. Waliser, E. J. Fetzer, and P. J. Neiman, 2010: Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements. *Geophys. Res. Lett.*, **37**, L20401, https://doi.org/10.1029/2010GL044696.

Haase, J. S., M. J. Murphy, B. Cao, F. M. Ralph, M. Zheng, and L. Delle Monache, 2021: Multi-GNSS airborne radio occultation observations as a complement to dropsondes in atmospheric river reconnaissance. *J. Geophys. Res. Atmos.*, **126**, e2021JD034865, https://doi.org/10.1029/2021JD034865.

Hamill, T. M., F. Yang, C. Cardinali, and S. J. Majumdar, 2013: Impact of targeted winter storm reconnaissance dropwindsonde data on midlatitude numerical weather predictions. *Mon. Wea. Rev.*, **141**, 2058–2065, https://doi.org/10.1175/MWR-D-12-00309.1.

Healy, S. B., 2011: Refractivity coefficients used in the assimilation of GPS radio occultation measurements. *J. Geophys. Res.*, **116**, D01106, https://doi.org/10.1029/2010JD014013.

Henn, B., K. N. Musselman, L. Lestak, F. M. Ralph, and N. P. Molotch, 2020: Extreme runoff generation from atmospheric river driven snowmelt during the 2017 Oroville Dam spillways incident. *Geophys. Res. Lett.*, **47**, e2020GL088189, https://doi.org/10.1029/2020GL088189.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hill, A. J., C. C. Weiss, and B. C. Ancell, 2020: Factors influencing ensemble sensitivity–based targeted observing predictions at convection-allowing resolutions. *Mon. Wea. Rev.*, **148**, 4497–4517, https://doi.org/10.1175/MWR-D-20-0015.1.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Keclik, A. M., C. Evans, P. J. Roebber, and G. S. Romine, 2017: The influence of assimilated upstream, preconvective dropsonde observations on ensemble forecasts of convection initiation

during the Mesoscale Predictability Experiment. *Mon. Wea. Rev.*, **145**, 4747–4770, https://doi.org/10.1175/MWR-D-17-0159.1.

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

Kren, A. C., L. Cucurull, and H. Wang, 2020: Addressing the sensitivity of forecast impact to flight path design for targeted observations of extratropical winter storms: A demonstration in an OSSE framework. *Meteor. Appl.*, **27**, e1942, https://doi.org/10.1002/met.1942.

Langland, R. H., and Coauthors, 1999: The North Pacific Experiment (NORPEX-98): Targeted observations for improved North American weather forecasts. *Bull. Amer. Meteor. Soc.*, **80**, 1363–1384, https://doi.org/10.1175/1520-0477(1999)080<1363:TNPENT>2.0.CO;2.

Lord, S. J., X. Wu, V. Tallapragada, and F. M. Ralph, 2023a: The impact of dropsonde data on the performance of the NCEP Global Forecast System during the 2020 atmospheric rivers observing campaign. Part I: Precipitation. *Wea. Forecasting*, **38**, 17–45, https://doi.org/10.1175/WAF-D-22-0036.1.

——, ——, ——, and ——, 2023b: The impact of dropsonde data on the performance of the NCEP global forecast system during the 2020 atmospheric rivers observing campaign. Part II: Dynamic variables and humidity. *Wea. Forecasting*, **38**, 721–752, https://doi.org/10.1175/WAF-D-22-0072.1.

Majumdar, S. J., 2016: A review of targeted observations. *Bull. Amer. Meteor. Soc.*, **97**, 2287–2303, https://doi.org/10.1175/BAMS-D-14-00259.1.

——, C. H. Bishop, R. Buizza, and R. Gelaro, 2002a: A comparison of ensemble-transform Kalman-filter targeting guidance with ECMWF and NRL total-energy singular-vector guidance. *Quart. J. Roy. Meteor. Soc.*, **128**, 2527–2549, https://doi.org/10.1256/qj.01.214.

——, ——, B. J. Etherton, and Z. Toth, 2002b: Adaptive sampling with the ensemble transform Kalman filter. Part II: Field program implementation. *Mon. Wea. Rev.*, **130**, 1356–1369, https://doi.org/10.1175/1520-0493(2002)130<1356:ASWTET>2.0.CO;2.

——, K. J. Sellwood, D. Hodyss, Z. Toth, and Y. Song, 2010: Characteristics of target areas selected by the ensemble transform Kalman filter for medium-range forecasts of high-impact winter weather. *Mon. Wea. Rev.*, **138**, 2803–2824, https://doi.org/10.1175/2010MWR3106.1.

——, M. J. Brennan, and K. Howard, 2013: The impact of dropwindsonde and supplemental rawinsonde observations on track forecasts for Hurricane Irene (2011). *Wea. Forecasting*, **28**, 1385–1403, https://doi.org/10.1175/WAF-D-13-00018.1.

Martin, A., F. M. Ralph, R. Demirdjian, L. DeHaan, R. Weihs, J. Helly, D. Reynolds, and S. Iacobellis, 2018: Evaluation of atmospheric river predictions by the WRF Model using aircraft and regional mesonet observations of orographic precipitation and its forcing. *J. Hydrometeor.*, **19**, 1097–1113, https://doi.org/10.1175/JHM-D-17-0098.1.

Masutani, M., and Coauthors, 2013: Observing System Simulation Experiments; justifying new Arctic observation capabilities. NCEP Office Note 473, 20 pp., https://repository.library.noaa.gov/view/noaa/6965.

McMurdie, L. A., and Coauthors, 2022: Chasing snowstorms: The Investigation of Microphysics and Precipitation for Atlantic Coast-Threatening Snowstorms (IMPACTS) campaign. *Bull. Amer. Meteor. Soc.*, **103**, E1243–E1269, https://doi.org/10.1175/BAMS-D-20-0246.1.

NOAA Science Advisory Board, 2021: A report on priorities for weather research. NOAA Science Advisory Board Rep., 119 pp., https://sab.noaa.gov/wp-content/uploads/2021/12/PWR-Report_Final_12-9-21.pdf.

OFCM, 2019: National winter season operations plan. Office of the Federal Coordinator for Meteorology Doc. FCM-P13-2019, 84 pp., https://www.icams-portal.gov/resources/ofcm/nwsop/2019_nwsop.pdf.

——, 2022: National winter season operations plan. Office of the Federal Coordinator for Meteorology Doc. FCM-P13-2022, 124 pp., https://www.icams-portal.gov/resources/ofcm/nwsop/2022_nwsop.pdf.

Otkin, J. A., and J. E. Martin, 2004: A synoptic climatology of the subtropical kona storm. *Mon. Wea. Rev.*, **132**, 1502–1517, https://doi.org/10.1175/1520-0493(2004)132<1502:ASCOTS>2.0.CO;2.

Payne, A. E., and Coauthors, 2020: Responses and impacts of atmospheric rivers to climate change. *Nat. Rev. Earth Environ.*, **1**, 143–157, https://doi.org/10.1038/s43017-020-0030-5.

Pu, Z., X. Li, C. S. Velden, S. D. Aberson, and W. T. Liu, 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. *Wea. Forecasting*, **23**, 62–79, https://doi.org/10.1175/2007WAF2007006.1.

Ralph, F. M., and M. D. Dettinger, 2011: Storms, floods, and the science of atmospheric rivers. *Eos, Trans. Amer. Geophys. Union*, **92**, 265–266, https://doi.org/10.1029/2011EO320001.

——, P. J. Neiman, and G. A. Wick, 2004: Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the winter of 1997/98. *Mon. Wea. Rev.*, **132**, 1721–1745, https://doi.org/10.1175/1520-0493(2004)132<1721:SACAOO>2.0.CO;2.

——, ——, and R. Rotunno, 2005: Dropsonde observations in low-level jets over the northeastern Pacific Ocean from CALJET-1998 and PACJET-2001: Mean vertical-profile and atmospheric-river characteristics. *Mon. Wea. Rev.*, **133**, 889–910, https://doi.org/10.1175/MWR2896.1.

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

——, and Coauthors, 2020: West Coast forecast challenges and development of atmospheric river reconnaissance. *Bull. Amer. Meteor. Soc.*, **101**, E1357–E1377, https://doi.org/10.1175/BAMS-D-19-0183.1.

——, and Coauthors, 2021: Radiosonde data collected during California storms. UC San Diego Library Digital Collections, accessed 22 February 2024, https://library.ucsd.edu/dc/object/bb60495334.

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

——, and Coauthors, 2023: Impacts of northeastern Pacific buoy surface pressure observations. *Mon. Wea. Rev.*, **151**, 211–226, https://doi.org/10.1175/MWR-D-22-0124.1.

Romine, G. S., C. S. Schwartz, R. D. Torn, and M. L. Weisman, 2016: Impact of assimilating dropsonde observations from MPEX on ensemble forecasts of severe weather events. *Mon. Wea. Rev.*, **144**, 3799–3823, https://doi.org/10.1175/MWR-D-15-0407.1.

Rutz, J. J., W. J. Steenburgh, and F. M. Ralph, 2014: Climatological characteristics of atmospheric rivers and their inland

penetration over the western United States. *Mon. Wea. Rev.*, **142**, 905–921, https://doi.org/10.1175/MWR-D-13-00168.1.

Santek, D., and Coauthors, 2019: 2018 Atmospheric Motion Vector (AMV) intercomparison study. *Remote Sens.*, **11**, 2240, https://doi.org/10.3390/rs11192240.

Schindler, M., M. Weissmann, A. Schäfler, and G. Radnoti, 2020: The impact of dropsonde and extra radiosonde observations during NAWDEX in autumn 2016. *Mon. Wea. Rev.*, **148**, 809–824, https://doi.org/10.1175/MWR-D-19-0126.1.

Skamarock, W. C., and Coauthors, 2019: A description of the Advanced Research WRF Model version 4. NCAR Tech. Note NCAR/TN-556+STR 145 pp., https://doi.org/10.5065/1dfh-6p97.

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

Sun, W., Z. Liu, C. A. Davis, F. M. Ralph, L. Delle Monache, and M. Zheng, 2022: Impacts of dropsonde and satellite observations on the forecasts of two atmospheric-river-related heavy rainfall events. *Atmos. Res.*, **278**, 106327, https://doi.org/10.1016/j.atmosres.2022.106327.

Szunyogh, I., Z. Toth, R. E. Morss, S. J. Majumdar, B. J. Etherton, and C. H. Bishop, 2000: The effect of targeted dropsonde observations during the 1999 winter storm reconnaissance program. *Mon. Wea. Rev.*, **128**, 3520–3537, https://doi.org/10.1175/1520-0493(2000)128<3520:TEOTDO>2.0.CO;2.

Tewari, M., and Coauthors, 2004: Implementation and verification of the unified Noah land surface model in the WRF model. *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 14.2a, https://ams.confex.com/ams/84Annual/techprogram/paper_69061.htm.

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

Torn, R. D., and G. J. Hakim, 2008: Ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663–677, https://doi.org/10.1175/2007MWR2132.1.

——, ——, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490–2502, https://doi.org/10.1175/MWR3187.1.

van Leeuwen, P. J., 2015: Representation errors and retrievals in linear and nonlinear data assimilation. *Quart. J. Roy. Meteor. Soc.*, **141**, 1612–1623, https://doi.org/10.1002/qj.2464.

Velden, C., and Coauthors, 2005: Recent innovations in deriving tropospheric winds from meteorological satellites. *Bull. Amer. Meteor. Soc.*, **86**, 205–224, https://doi.org/10.1175/BAMS-86-2-205.

Waliser, D., and B. Guan, 2017: Extreme winds and precipitation during landfall of atmospheric rivers. *Nat. Geosci.*, **10**, 179–183, https://doi.org/10.1038/ngeo2894.

Wang, X., and T. Lei, 2014: GSI-based four-dimensional ensemble–variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP global forecast system. *Mon. Wea. Rev.*, **142**, 3303–3325, https://doi.org/10.1175/MWR-D-13-00303.1.

Weissmann, M. L., and Coauthors, 2011: The influence of assimilating dropsonde data on typhoon track and midlatitude forecasts. *Mon. Wea. Rev.*, **139**, 908–920, https://doi.org/10.1175/2010MWR3377.1.

Weisman, M. L., and Coauthors, 2015: The Mesoscale Predictability Experiment (MPEX). *Bull. Amer. Meteor. Soc.*, **96**, 2127–2149, https://doi.org/10.1175/BAMS-D-13-00281.1.

Wick, G., and Coauthors, 2020: NOAA's Sensing Hazards with Operational Unmanned Technology (SHOUT) experiment: Observations and forecast impacts. *Bull. Amer. Meteor. Soc.*, **101**, E968–E987, https://doi.org/10.1175/BAMS-D-18-0257.1.

WPC, 2021: Major West Coast winter storm—2021: Storm summaries—January 27–29, 2021, NOAA/NWS/WPC, accessed 1 November 2023, https://www.wpc.ncep.noaa.gov/storm_summaries/2021/storm1/storm1_archive.shtml.

Zhang, Z., and F. M. Ralph, 2021: The influence of antecedent atmospheric river conditions on extratropical cyclogenesis. *Mon. Wea. Rev.*, **149**, 1337–1357, https://doi.org/10.1175/MWR-D-20-0212.1.

——, ——, and M. Zheng, 2019: The relationship between extratropical cyclone strength and atmospheric river intensity and position. *Geophys. Res. Lett.*, **46**, 1814–1823, https://doi.org/10.1029/2018GL079071.

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

——, and Coauthors, 2021a: Data gaps within atmospheric rivers over the northeastern Pacific. *Bull. Amer. Meteor. Soc.*, **102**, E492–E524, https://doi.org/10.1175/BAMS-D-19-0287.1.

——, and Coauthors, 2021b: Improved forecast skill through the assimilation of dropsonde observations from the atmospheric river reconnaissance program. *J. Geophys. Res. Atmos.*, **126**, e2021JD034967, https://doi.org/10.1029/2021JD034967.

Zhou, Y., H. Kim, and B. Guan, 2018: Life cycle of atmospheric rivers: Identification and climatological characteristics. *J. Geophys. Res. Atmos.*, **123**, 12 715–12 725, https://doi.org/10.1029/2018JD029180.

Zhu, Y., and R. E. Newell, 1998: A proposed algorithm for moisture fluxes from atmospheric rivers. *Mon. Wea. Rev.*, **126**, 725–735, https://doi.org/10.1175/1520-0493(1998)126<0725:APAFMF>2.0.CO;2.