

Forecast Errors and Uncertainties in Atmospheric Rivers

DAVID A. LAVERS,^a N. BRUCE INGLEBY,^a ANEESH C. SUBRAMANIAN,^b DAVID S. RICHARDSON,^a
F. MARTIN RALPH,^c JAMES D. DOYLE,^d CAROLYN A. REYNOLDS,^d RYAN D. TORN,^e
MARK J. RODWELL,^a VIJAY TALLAPRAGADA,^f AND FLORIAN PAPPENBERGER^a

^a *European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom*

^b *University of Colorado Boulder, Boulder, Colorado*

^c *Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California*

^d *Naval Research Laboratory, Monterey, California*

^e *Department of Atmospheric and Environmental Science, University at Albany, State University of New York, Albany, New York*

^f *NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland*

(Manuscript received 1 April 2020, in final form 7 May 2020)

ABSTRACT


A key aim of observational campaigns is to sample atmosphere–ocean phenomena to improve understanding of these phenomena, and in turn, numerical weather prediction. In early 2018 and 2019, the Atmospheric River Reconnaissance (AR Recon) campaign released dropsondes and radiosondes into atmospheric rivers (ARs) over the northeast Pacific Ocean to collect unique observations of temperature, winds, and moisture in ARs. These narrow regions of water vapor transport in the atmosphere—like rivers in the sky—can be associated with extreme precipitation and flooding events in the midlatitudes. This study uses the dropsonde observations collected during the AR Recon campaign and the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) to evaluate forecasts of ARs. Results show that ECMWF IFS forecasts 1) were colder than observations by up to 0.6 K throughout the troposphere; 2) have a dry bias in the lower troposphere, which along with weaker winds below 950 hPa, resulted in weaker horizontal water vapor fluxes in the 950–1000-hPa layer; and 3) exhibit an underdispersiveness in the water vapor flux that largely arises from model representativeness errors associated with dropsondes. Four U.S. West Coast radiosonde sites confirm the IFS cold bias throughout winter. These issues are likely to affect the model's hydrological cycle and hence precipitation forecasts.

1. Introduction

Observational campaigns use a range of airborne and surface-based instruments to probe atmosphere–ocean phenomena to improve understanding of these phenomena, and in turn, numerical weather prediction (NWP) models. A campaign, for example, can have a research aircraft to deploy dropsondes to measure atmospheric properties (Ralph et al. 2017), extra radiosondes to better sample weather systems or climate zones (Schäfler et al. 2018; or a research vessel for ocean

measurements (Ralph et al. 2016). The observations taken may then be assimilated into NWP models to first provide a more accurate estimate of the initial state, and second, to compare with the NWP short-range forecasts prior to their assimilation to identify model errors and behavior. In recent years, there has been a range of missions, for example, covering tropical regions (e.g., Doyle et al. 2017), extratropical regions (Geerts et al. 2017; Ralph et al. 2016; Schäfler et al. 2018), polar regions (e.g., Uttal et al. 2002), and cloud processes (Flamant et al. 2018).

In January and February 2018, there was an observational campaign called Atmospheric River Reconnaissance (AR Recon) in which research aircraft released dropsondes into atmospheric rivers (ARs; Ralph et al. 2018) and other dynamically active regions across the eastern North Pacific

 Denotes content that is immediately available upon publication as open access.

Corresponding author: David Lavers, david.lavers@ecmwf.int

DOI: 10.1175/WAF-D-20-0049.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

Ocean, along with radiosondes from sites in California. ARs are important because they are responsible for much of the water vapor flux across the midlatitudes (Ralph et al. 2005, 2017) and as they can be associated with extreme precipitation, flooding, and adverse socioeconomic effects especially in coastal mountainous regions (Ralph et al. 2006; Lavers et al. 2011; Neiman et al. 2011; Ramos et al. 2015). It was the aim of AR Recon in 2018 to provide research measurements and added information to better inform decision-makers and forecasters on AR impacts. In particular, the campaign afforded the opportunity to do diagnostic studies on the capability of NWP systems to model ARs (e.g., Lavers et al. 2018; Stone et al. 2020). Lavers et al. (2018) found that the largest uncertainties in the magnitude of the AR water vapor flux in the ensemble forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) originated from the 850-hPa winds, a standard pressure level typically above the planetary boundary layer (PBL). The specific humidity was also erroneous in the ECMWF IFS forecasts examined, but was subject to less uncertainty. These uncertainties in the water vapor flux can affect the forecasts for high-impact extreme precipitation events driven by ARs and also the location and magnitude of the latent heat release, which has implications for the atmospheric dynamics and predictability (e.g., Berman and Torn 2019). Strong diabatic forcing over the central United States has also been identified as a common precursor six days prior to large forecast busts over Europe (Rodwell et al. 2013).

The AR Recon campaign conducted six intensive observation periods (IOPs) in February and March 2019 in which dropsondes and radiosondes were launched across the northeast Pacific Ocean and California, respectively. These unique dropsonde profiles increase the available data sample on ARs and together with those obtained in 2018 provide an opportunity to further the findings of Lavers et al. (2018). Specifically, one limitation of the previous research was the predominant use of three standard pressure levels (925, 850, and 700 hPa) to calculate the water vapor flux; this was because these are the only levels archived in the ECMWF ensemble system and a consistent assessment was required between the observations and medium-range forecasts. This study builds upon the previous results by considering all available dropsonde pressure levels that were assimilated into the ECMWF IFS, thus allowing for a more complete analysis that includes data at a much higher vertical resolution than considered by Lavers et al. (2018). In so doing, the following questions are addressed. First, what model errors exist in the specific humidity, temperature, and winds during these AR

events? Second, in which layers do the largest water vapor flux and its errors occur? And third, using the 925-, 850-, and 700-hPa surfaces, what is the spread–error relationship of lower-tropospheric water vapor flux in ECMWF IFS forecasts?

2. Data and methods

a. IOPs and U.S. West Coast land-based radiosonde sites

There were six IOPs in AR Recon 2019: 2 February, 11 February, 13 February, 24 February, 26 February, and 1 March 2019 (all centered at 0000 UTC). Figure 1 shows the magnitude of the vertically integrated horizontal water vapor flux (integrated vapor transport, IVT), mean sea level pressure, and dropsonde locations during the IOPs. The dropsondes were deployed by research aircraft (average flight time of 8 h) mostly in regions of intense water vapor flux. These observations were transferred to the World Meteorological Organization Global Telecommunications System (GTS) and ingested into operational NWP systems including the ECMWF IFS. During the six IOPs, 259 dropsondes were released, and together with the 326 dropsondes released in the five IOPs in AR Recon 2018 [Fig. 1 in Lavers et al. (2018)], there were 585 dropsondes available, with 571 that include the 925-, 850-, and 700-hPa surfaces. In the IOPs in 2018, Vaisala RD94 dropsondes were used and their accuracy for pressure, temperature, and relative humidity is 0.4 hPa, 0.2 K, and 2%, respectively (Vaisala 2010). In the 2019 IOPs, the Vaisala RD41 dropsondes were used, with higher accuracy for temperature of 0.1 K (Vaisala 2018). Jensen et al. (2016) indicate the high performance of the humidity sensors, and the wind observations were reported to the nearest 1 m s^{-1} (due to the alphanumeric data reports).

To assess the representativeness of the dropsondes, the 0000 UTC ascents at four U.S. West Coast land-based radiosonde sites were assessed during winter [December–January–February (DJF)] 2017/18 and 2018/19 and summer [June–July–August (JJA)] 2018 and 2019. The sites used were at Salem, Oregon (ID: 72694; 44.9°N, 123.0°W); Medford, Oregon (ID: 72597; 42.4°N, 122.9°W); Oakland, California (ID: 72493; 37.7°N, 122.2°W); and San Diego, California (ID: 72293; 32.8°N, 117.1°W). These four radiosonde sites are shown as magenta dots in Fig. 1.

b. ECMWF ensemble of data assimilations and forecasts

To characterize the initial atmospheric flow uncertainty in the ECMWF IFS, an ensemble of data assimilations (EDA; Isaksen et al. 2010) is employed. During this period, the EDA consisted of one control member

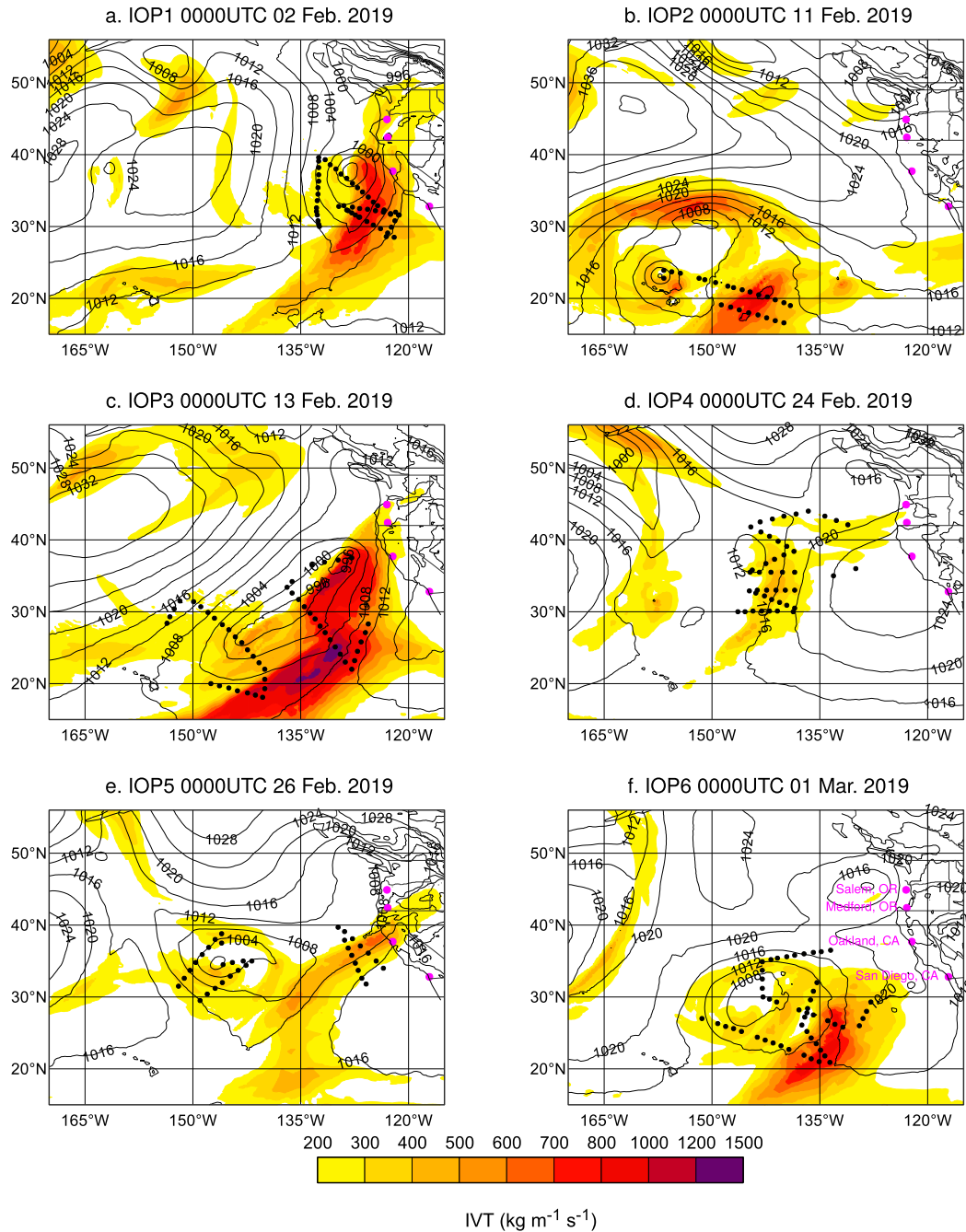


FIG. 1. The six intensive observation periods (IOPs) in AR Recon 2019. (a)–(f) The ECMWF analysis of mean sea level pressure (contours; hPa) and the magnitude of the vertically integrated vapor transport (IVT) (shaded; $\text{kg m}^{-1} \text{s}^{-1}$). Dropsonde locations are given by black dots, and the magenta dots in each panel refer to the radiosonde locations. The names of the radiosonde locations are given in (f).

and 25 perturbed members in which the first-guess (background; 3–15-h lead time) forecasts and observations (including the dropsonde data) were combined (using four-dimensional variational data assimilation) to produce 26 new analyses. In this framework, the variance of the first-guess forecasts indicates the initial flow

uncertainty and the observation errors are derived from the different observation perturbations estimated and applied in each of the 25 perturbed EDA members. The analyses attempt to describe the remaining atmospheric uncertainty after assimilation. Following the data assimilation procedure, the 50 ensemble forecast members

(ENS) that run out to 15 days are produced by a symmetric combination of 6-h forecasts from the EDA analyses, the latest single high-resolution forecast, and singular vectors (Lang et al. 2015). Stochastic perturbations (Leutbecher et al. 2017) are applied within the background forecasts and ENS. The EDA in the 0000 UTC window and ENS data from 0000 UTC were interpolated to the release point of the dropsonde observations.

c. Forecast evaluation and water vapor flux calculation

In data assimilation, it is common practice to assess the model fit to observations using observation-minus-background differences ($O - B$; Desroziers et al. 2005), referred to in this study as departures. Herein, using all dropsonde locations, we calculate $O - B$ in the EDA control member for specific humidity, temperature, and wind speed on all assimilated pressure levels from 1000 to 200 hPa and average these departures in 50-hPa layers. This approach allows for the identification of potential model biases in different layers. Note that each dropsonde report received over the GTS for use in this study typically contained 15–40 levels and thus they have fewer pressure levels than those measured by the dropsondes on their descent to the ocean surface. These pressure levels consist of both standard and significant levels (most of them being significant) and the average $O - B$ departures were evaluated from these levels.

The horizontal pressure-level water vapor flux magnitudes were determined as the product of specific humidity and the horizontal wind speeds and then evaluated in 50-hPa layers. These fluxes were assessed by linearly interpolating specific humidity onto wind levels; similar fluxes were found when interpolating the winds on to specific humidity levels. This approach, without considering the vertical integral of the water vapor flux, guards against higher weighting being given to any particular pressure level.

A reliable ensemble forecasting system should have the property where over many forecasts the mean ensemble variance and error of the ensemble mean are the same (Leutbecher and Palmer 2008). For example, the ensemble variance (EnsVar) can be compared with the variance of the error of the ensemble mean (Error²); the standard deviation of these quantities can also be compared. Rather than discussing errors, however, we discuss the departures of forecasts from observations because each has associated errors and biases. Following Rodwell et al. (2016), we estimate observation uncertainty (ObsUnc²) consistent with the EDA's observation perturbations and calculate the squared bias

between the forecast and observation (Bias²). Writing Depar² for the mean squared departure and DepVar for the variance of the departures (once the bias has been removed), we obtain the following:

$$\text{Depar}^2 = \text{EnsVar} + \text{ObsUnc}^2 + \text{Bias}^2, \quad (1)$$

$$\text{DepVar} = \text{EnsVar} + \text{ObsUnc}^2. \quad (2)$$

This modified spread–error relationship is investigated in water vapor fluxes on the 925-, 850-, and 700-hPa pressure surfaces at 24-h intervals out to 120 h (day 5).

3. Results and discussion

a. Forecast evaluation with dropsonde observations

To illustrate the AR Recon dropsonde data used to calculate the $O - B$ departures, Fig. 2 shows scatterplots of observations versus background (EDA control) forecasts for specific humidity, temperature, wind speed, and water vapor fluxes in the 950–1000-hPa layer. First, as expected given the short forecast range, all variables have a strong linear correlation ranging from 0.86 for water vapor flux (Fig. 2d) to 0.97 for temperature (Fig. 2b). Second, for this particular layer, the mean $O - B$ shows that the observed temperature is 0.23 K warmer (Fig. 2b) and the specific humidity is 0.15 g kg⁻¹ moister (Fig. 2a) than the model, as highlighted by the location of many of the points below the 1:1 lines in Figs. 2a and 2b. Third, the cloud of points for wind speed (Fig. 2c) and water vapor flux (Fig. 2d) have a similar shape implying that the model has a tendency to underestimate low-level wind speeds (>20 m s⁻¹) and subsequently underestimate low-level water vapor fluxes. This likely causes the positive $O - B$ values of 0.41 m s⁻¹ and 6.06 g kg⁻¹ m s⁻¹ for the wind speed and vapor fluxes, respectively (Figs. 2c,d).

The average $O - B$ departures (in the EDA control member) in 50-hPa layers for the specific humidity, temperature, wind speed, and water vapor fluxes are shown in Fig. 3. Figure 3a reveals that the model is drier than observations by ~0.1–0.2 g kg⁻¹ below 850 hPa and moister than observations by ~0.1 g kg⁻¹ between 600 and 850 hPa. The dry departure below 850 hPa would reduce the lower-tropospheric water vapor flux in the model and potentially affect IFS precipitation. The model moist bias at midlevels has less impact on the water vapor fluxes as it is above the low-level jet core; this is confirmed by assessment of the observed and background fluxes in Fig. 3d. For the temperature in Fig. 3b, the model mostly has a cold bias throughout the troposphere when compared with the dropsonde profiles, with the largest departure of about 0.6 K found from 900 to 950 hPa. Over the ocean this layer includes,

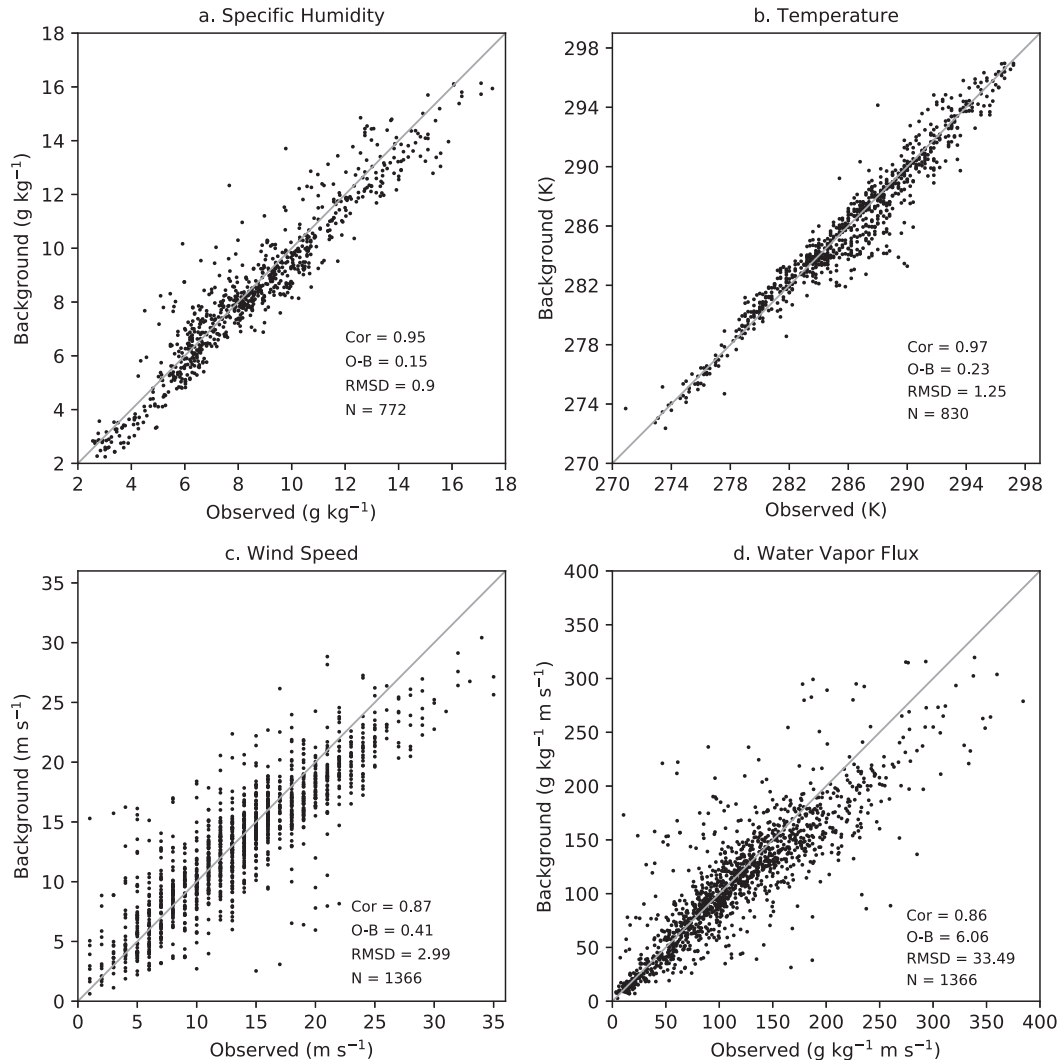


FIG. 2. Scatterplots of the observations vs the background (EDA control) forecasts at the dropsonde locations for (a) specific humidity, (b) temperature, (c) wind speed, and (d) water vapor fluxes in the 950–1000-hPa layer. The linear correlation, mean observed-minus-background ($O - B$) departure, root-mean-square departure (RMSD), number of observations, and the 1:1 lines (in gray) are given. The striped nature of the observed winds in (c) results from the winds being reported to the nearest 1 m s^{-1} .

on average, the PBL top, which is calculated here using the bulk Richardson number (e.g., Lavers et al. 2019). Figure 4 shows scatterplots of the observations versus background for the height and temperature at the PBL top. There is generally good agreement found in the scatterplots and the linear correlation has values of 0.79 and 0.96 for the height and temperature of the PBL top, respectively; the average observed PBL height is 734.4 m (928.0 hPa) and the background PBL height is 743.6 m (926.4 hPa). These findings for the PBL height are in line with those for the midlatitudes in Lavers et al. (2019) and the mean $O - B$ for temperature at the PBL top of 0.44 K corroborates the $O - B$ for

temperature in Fig. 3b. Furthermore, the temperature departures of all layers in Fig. 3b have mean values that are significantly different from zero at the 90% level, as shown by the black error bars that do not cross zero, suggesting that the signal of a cold temperature bias in the model is robust. These model cold biases are similar to previous findings of an approximate 0.5 K bias in the tropics (20°S–20°N) and somewhat smaller biases in the northern midlatitudes at low levels (Ingleby 2017).

For wind speed, there is evidence that the observations are stronger than the background in the 950–1000-hPa layer and in the upper troposphere (Fig. 3c),

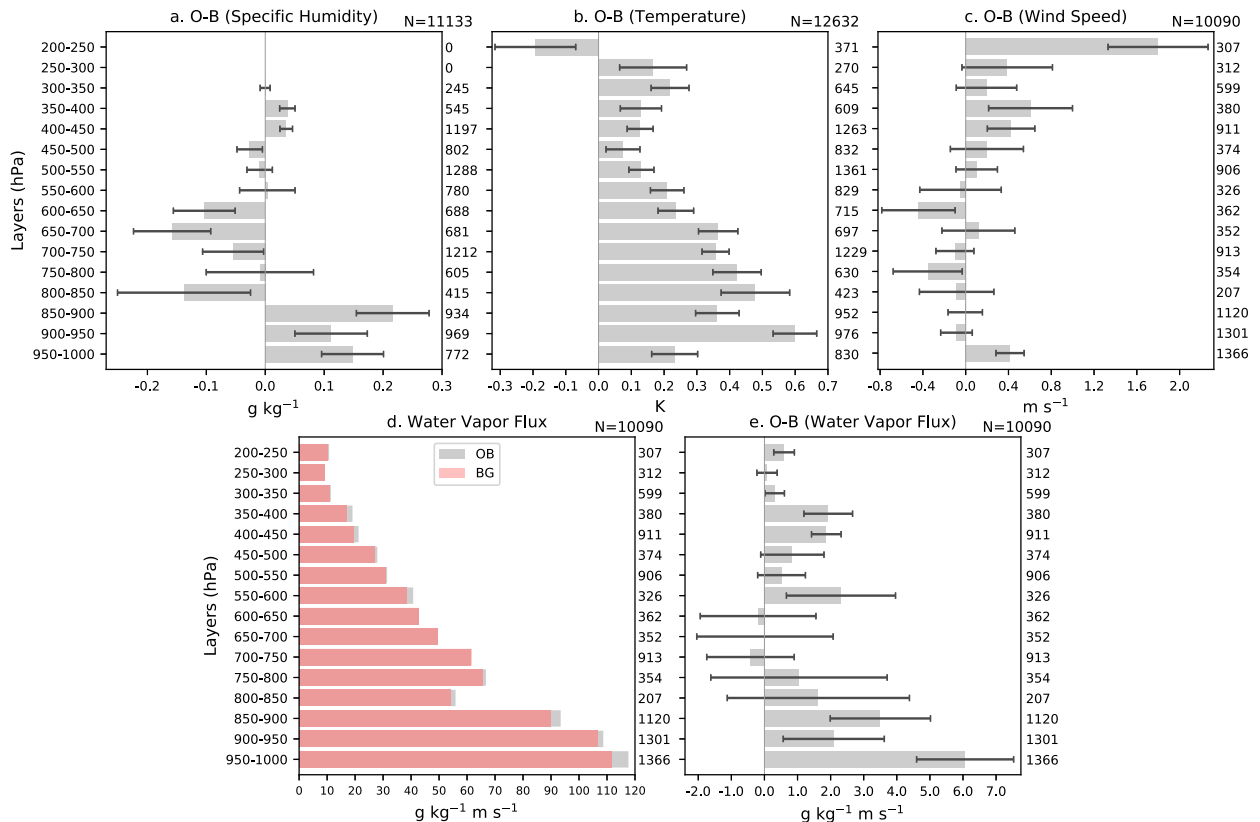


FIG. 3. Shown are $O - B$ departures (in the EDA control member) averaged in 50-hPa layers at the dropsonde locations for (a) specific humidity, (b) temperature, and (c) wind speed. (d) Observed (gray) and background (red) pressure-level water vapor flux magnitude averaged in 50-hPa layers when interpolating specific humidity on to the wind levels and (e) the $O - B$ departures for water vapor flux. The error bars show the 90% confidence interval of the mean. The number of values in each layer is given on the right-hand side of the panel, and the number used across all layers is shown at the top right of each panel.

with the latter perhaps suggesting a model underestimation of the jet stream. In most layers, however, there is little sign of a clear bias. The pressure-level background and observed water vapor fluxes averaged in 50-hPa layers are plotted in Fig. 3d. In the model and observations, the 900–950- and 950–1000-hPa layers have the peak water vapor fluxes, where observed values of 109 and 118 $\text{g kg}^{-1} \text{m s}^{-1}$ are found, respectively. The low-altitude nature of this flux in the PBL may not be adequately captured by the commonly archived 925- and 1000-hPa levels suggesting that forecast studies of ARs using a limited number of pressure levels may not provide accurate flux estimates. With the moisture content of the atmosphere decreasing with height, water vapor fluxes also typically decrease with increasing height, with a notable drop off in water vapor fluxes above 850 hPa in the present study (Fig. 3d). A comparison of the observed (gray bars) and model (red bars) fluxes highlights that the model underestimates water vapor fluxes below 850 hPa, with Fig. 3e showing that the $O - B$ in the 950–1000-hPa layer is

$6.06 \text{ g kg}^{-1} \text{m s}^{-1}$ meaning that the forecast flux is 5.2% below the observed flux. These results indicate that the modeled transport of water vapor through the atmospheric branch of the water cycle within ARs (over the ocean) is insufficient in the IFS forecasts, which has important ramifications for precipitation forecasting and for the correct positioning of moisture for latent heat release (e.g., Reynolds et al. 2019).

To ascertain whether there was dependence between the low-altitude forecast departures (of specific humidity, temperature, winds, and the water vapor flux) and their location relative to the AR, we also assessed the relationship between the root-mean-square departures (RMSD) in the EDA control member and the IVT at the 0000 UTC analysis at the closest model grid point. We use the IVT because it acts as a proxy for AR location, with larger values being mostly situated near the core of the AR. While there is some positive association between RMSD and IVT, there is not a particularly strong correlation evident (not shown). This aspect of the forecast evaluation will be tested further with more

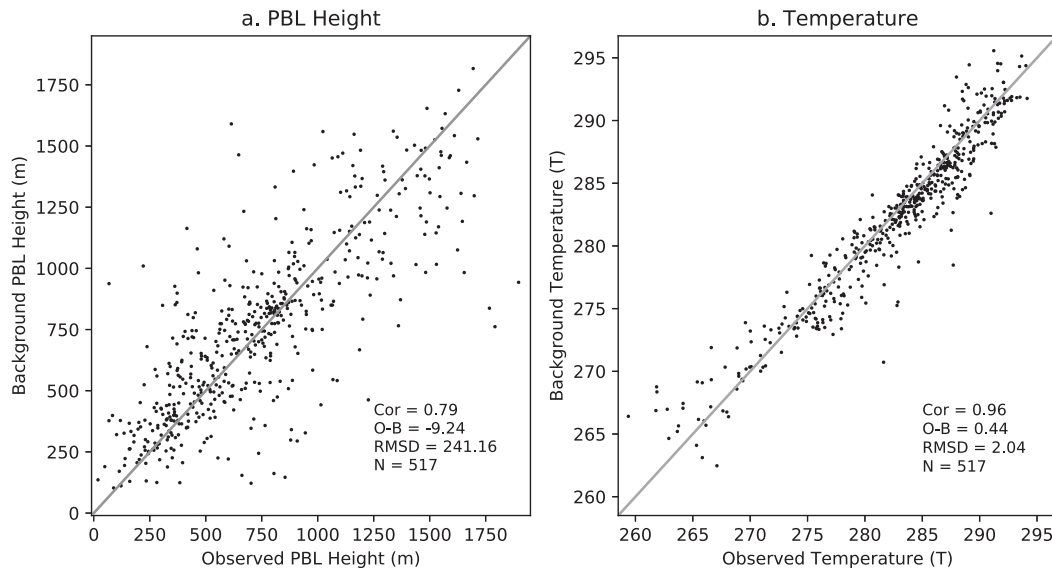


FIG. 4. Scatterplots of the observations vs the background forecasts (in the EDA control member) for (a) PBL height and (b) temperature at the PBL height. The linear correlation, mean $O - B$ departure, root-mean-square departure (RMSD), number of observations, and the 1:1 lines (in gray) are given. The number of dropsondes ($N = 517$) is smaller than the 585 dropsondes available because of the criterion in the Lavers et al. (2019) method to have at least five dropsonde pressure levels at low altitudes (≥ 700 hPa) to ensure that the low levels were well sampled.

dropsonde observations that will become available with future campaigns.

b. Forecast evaluation with U.S. West Coast radiosonde observations

We now consider the $O - B$ departures (in the EDA control member) at the four U.S. West Coast land-based radiosonde sites in Figs. 5 and 6. For the specific humidity during winters (DJF) 2017/18 and 2018/19 (Figs. 5a,d,g,j), there are primarily negative $O - B$ values suggesting the model has an overall wet bias. This model wet bias is somewhat different to the dropsonde results where a model dry bias was found at low altitudes. For specific humidity during summers (JJA) 2018 and 2019 (Figs. 6a,d,g,j), the model bias in the lower troposphere changes sharply with height, with dry biases found below 950 hPa and wet biases above 950 hPa at Salem and San Diego, and a dry bias found between 950 and 800 hPa in Oakland and Medford (with a wet bias in Oakland below 950 hPa). These results illustrate how, over land, biases can vary significantly based on the model's ability to represent the local climate, effects of topography, and land-sea differences at the subgrid scale. For example, the model grid cells over land do not accurately represent subgrid-scale urban areas, which would neglect the urban heat island effect thus potentially leading to temperature biases and moisture content issues. Note,

also that the $O - B$ departures are mostly larger in JJA owing to the higher moisture content present in the summer (cf. Figs. 5a,d,g,j and 6a,d,g,j).

In terms of temperature, positive $O - B$ departures or model cold biases are generally seen in the radiosonde profiles and these departures are broadly similar between DJF (Figs. 5b,e,h,k) and JJA (Figs. 6b,e,h,k) particularly above the PBL. This corroborates the dropsonde results discussed in section 3a and the cold biases presented in Ingleby (2017). In particular, the largest temperature biases throughout the troposphere are found at the southernmost site assessed herein at San Diego, which agrees with the results for the tropics in Ingleby (2017). It is hypothesized that these biases are linked to low-level humidity and cloud problems from several interrelated issues, such as, challenges posed by low vertical resolution satellite data to the data assimilation system and forecast model errors. For the wind speed (Figs. 5c,f,i,l and 6c,f,i,l), the high-altitude winds are generally underestimated by the model, a similar finding to the dropsonde results (Fig. 3c). In the lower troposphere (except Medford; Figs. 5f and 6f) the $O - B$ wind speed departures are negative, and hence different from the dropsondes, meaning that the model winds are mostly too large, which potentially indicates an issue with model roughness and drag at these locations. The different average wind profile at Medford relates to its higher altitude (~ 400 m) and surrounding complex terrain. Comparing DJF (Figs. 5c,f,i,l)

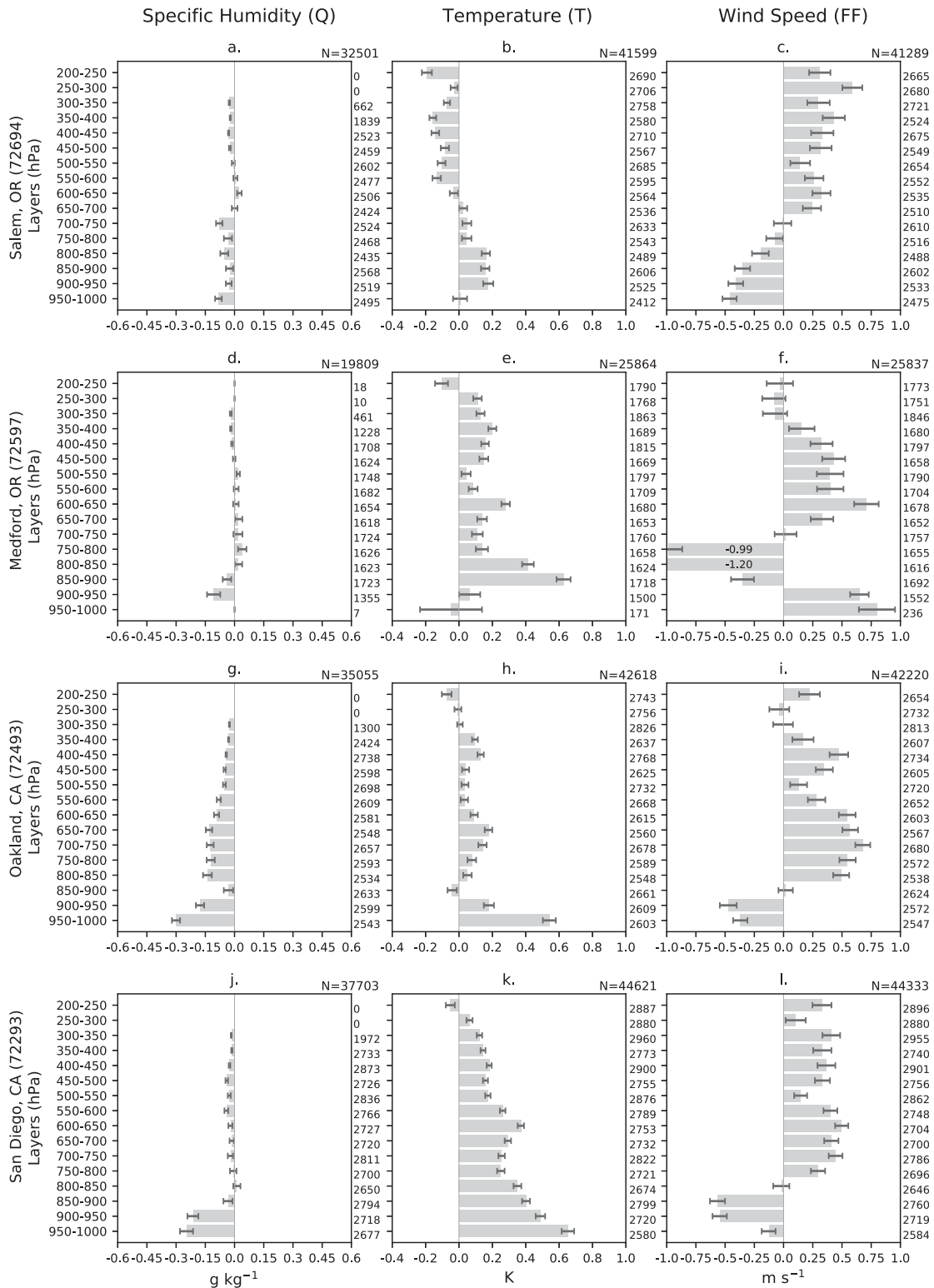


FIG. 5. (a)–(l) $O - B$ departures (in the EDA control member) averaged in 50-hPa layers during DJF 2017/18 and 2018/19 for (left) specific humidity, (center) temperature, and (right) wind speed at the four radiosonde sites. The layers in (d) with less than 100 values are masked out. The bars in (f) that extend beyond the x axis have their mean values given in the bar. Key as in Fig. 3.

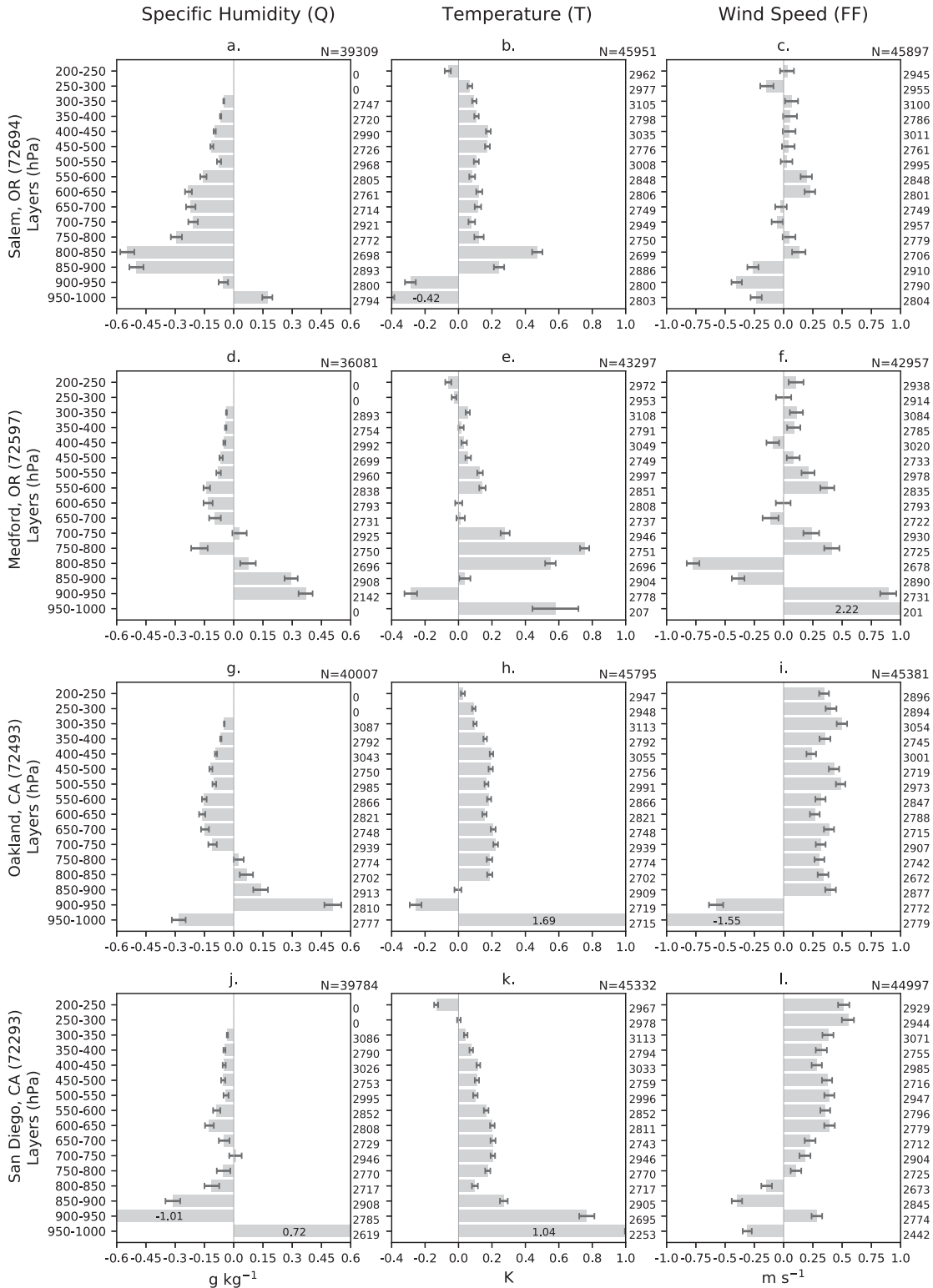


FIG. 6. (a)–(l) $O - B$ departures (in the EDA control member) averaged in 50-hPa layers during JJA 2018 and 2019 for (left) specific humidity, (center) temperature, and (right) wind speed at the four radiosonde sites. The bars in (b), (f), (h), (j), (k), and (l) that extend beyond the x axis have their mean values given in the bar. Key as in Fig. 3.

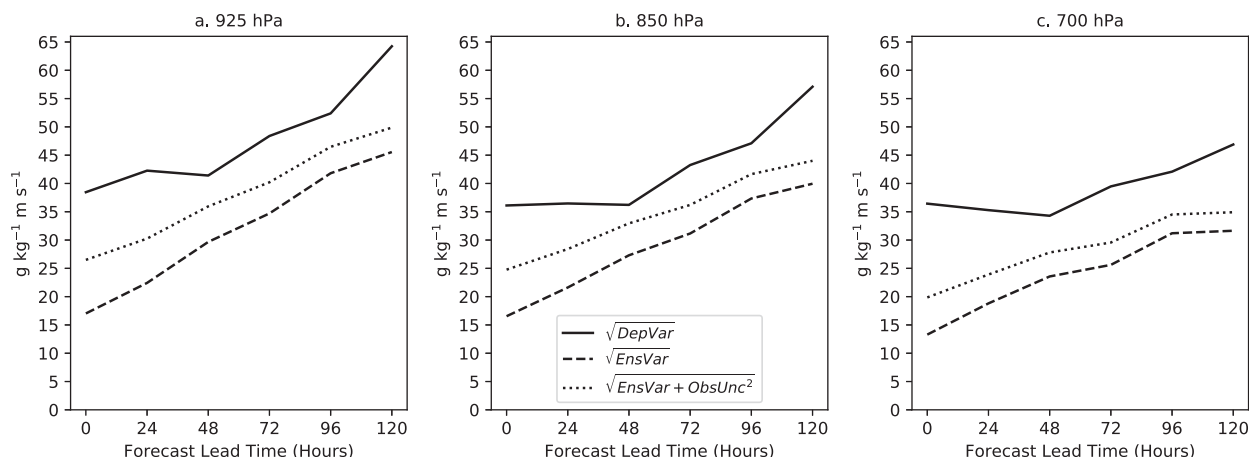


FIG. 7. The modified spread–error relationship in (2) of pressure-level water vapor flux magnitude. The square roots of DepVar (solid), EnsVar (dashed), and EnsVar + ObsUnc² (dotted) are shown at 0–120-h forecast lead times for (a) 925, (b) 850, and (c) 700 hPa.

with the results of JJA (Figs. 6c,f,i,l) shows that the sign of the model wind speed bias at different heights is broadly consistent in both seasons, although the magnitude of the bias may change between summer and winter.

c. Spread–error relationship of water vapor flux at the dropsonde locations

A forecast evaluation of the square root of the modified spread–error relationship in (2) at the dropsonde profiles is now undertaken for water vapor fluxes on the 925-, 850-, and 700-hPa levels and the results are presented in Fig. 7. In general, the standard deviation of the departures $\sqrt{\text{DepVar}}$ (solid lines) increases with lead time with the 925-hPa surface having the largest departures (Fig. 7a). Note that an evaluation of the modified spread–error relationship with a smaller sample size showed that the shape of the error lines depended somewhat on the sample used. Thus, the smallest departures found at $T + 48$ in Fig. 7 are hypothesized to result from the relatively small sample considered and do not indicate that $T + 48$ is less subject to errors. The ensemble standard deviation $\sqrt{\text{EnsVar}}$ (dashed lines) also grows with lead time, but it is not enough to explain the standard deviation of the departures $\sqrt{\text{DepVar}}$. The observation uncertainty is also an important component, since by considering it, the total $\sqrt{\text{EnsVar} + \text{ObsUnc}^2}$ draws closer to the departures (cf. dotted and solid lines in Fig. 7). In this case, the observation uncertainty is thought to largely account for representativeness errors whereby the model resolving processes on a grid does not adequately represent the point observation from the dropsonde. However, this is still not enough to match the standard deviation of the departures, which implies that the forecasts for this quantity are underdispersive,

and this may reflect the need for improved stochastic perturbations to better represent model uncertainty (e.g., of large horizontal gradients in ARs) and ultimately increase the dispersiveness of the ensemble. An alternative approach to address the representativeness issue is through a downscaling of the model fields to the point observation (which would further increase EnsVar). However, this is outside the scope of this study.

4. Conclusions

This investigation has used unique dropsonde observations collected during AR Recon 2018 and 2019 to evaluate the forecasts of ARs in the ECMWF EDA and ENS. First, the $O - B$ departures calculated in the AR Recon IOPs indicate that the EDA control member had a cold bias throughout the troposphere, reaching a maximum value of 0.6 K in the 900–950-hPa layer. This cold bias was also found at four U.S. West Coast radiosonde sites across both winter and summer seasons indicating that the bias is independent of AR occurrence. Furthermore the most noticeable cold bias was found at the southernmost radiosonde site (San Diego), consistent with the cold bias identified in the tropics in previous studies. Second, the EDA control member at the dropsonde profiles is 0.1–0.2 g kg^{-1} drier below 850 hPa and has weaker winds below 950 hPa, which results in an $O - B$ of 6.06 $\text{g kg}^{-1} \text{m s}^{-1}$ for water vapor flux in the 950–1000-hPa layer, or 5.2% of the average observed flux of 118 $\text{g kg}^{-1} \text{m s}^{-1}$. This implies that the forecasted moisture flux within ARs is insufficient, thus potentially affecting the downstream precipitation forecasts and the location of moisture within the model. Note that the largest errors herein are found at a lower

altitude than in Lavers et al. (2018). This results from the use of pressure-level water vapor fluxes compared to the vertical integral of the water vapor transport from three levels in Lavers et al. (2018) which resulted in a higher weight being given to 850 hPa. Third, the evaluation of the modified spread–error relationship revealed the IFS to be underdispersive. This is partly due to representativeness error, whereby the model resolves processes on the model grid scale of 18 km, which is a larger scale than the point observation given by the dropsonde. Even when an attempt is made to account for the representativeness issue through an estimate of the observation uncertainty, we found that the IFS still had too little spread. There is also a possibility that not enough model uncertainty is accounted for in the forecasts.

Further research on this topic could follow multiple directions and here we highlight three possibilities. First, as model improvements are generally implemented annually at NWP centers the future deployment of dropsondes in observational campaigns could be used to assess whether model upgrades are improving the forecasts of ARs and precipitation. Second, data denial experiments (i.e., IFS model forecasts run without these dropsonde observations) could be performed and evaluated to determine the impact of better representing the water vapor flux on precipitation forecasts. Third, a potential limitation of dropsondes is that they are usually available for specific storms during IOPs, which results in a dropsonde sample that is biased toward storms. Herein, we addressed this by using U.S. West Coast radiosonde sites. This shows that the cold bias is more widespread than just within storms and ARs. However, the dry bias at low levels was not present at these land stations which suggests that moisture biases are more dependent on the storms or the surface characteristics. This could be further investigated by using the synoptic radiosonde network in other midlatitude regions that are affected by ARs to ascertain if the forecast departures and uncertainties uncovered herein are found in other locations. This type of diagnostic study would then lead to improved understanding of model behavior and errors.

Acknowledgments. The authors acknowledge financial support from the European Union Horizon 2020 IMPREX project (Grant 641811). We are deeply thankful to the NOAA and U.S. Air Force flight crews for undertaking the missions to provide these dropsonde observations. JDD and CAR acknowledge the support of the Chief of Naval Research through the NRL Base Program, PE 0601153N. ACS acknowledges the support of U.S. Army Corps of Engineers (USACE)-Cooperative Ecosystem Studies Unit (CESU) as part of Forecast Informed Reservoir

Operations (FIRO), Grant W912HZ-15-2-0019 and the California Department of Water Resources Atmospheric River Program, grant 4600010378 TO#15 Am 22. The authors are grateful to the three anonymous reviewers whose comments helped to clarify and improve the paper. Data availability statement: The data used are available through the ECMWF archive (<https://www.ecmwf.int/en/forecasts/datasets/archive-datasets>).

REFERENCES

- Berman, J. D., and R. D. Torn, 2019: The impact of initial condition and warm conveyor belt forecast uncertainty on variability in the downstream waveguide in an ECMWF case study. *Mon. Wea. Rev.*, **147**, 4071–4089, <https://doi.org/10.1175/MWR-D-18-0333.1>.
- Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, <https://doi.org/10.1256/qj.05.108>.
- Doyle, J., and Coauthors, 2017: A view of tropical cyclones from above: The Tropical Cyclone Intensity experiment. *Bull. Amer. Meteor. Soc.*, **98**, 2113–2134, <https://doi.org/10.1175/BAMS-D-16-0055.1>.
- Flamant, C., and Coauthors, 2018: The Dynamics–Aerosol–Chemistry–Cloud Interactions in West Africa field campaign: Overview and research highlights. *Bull. Amer. Meteor. Soc.*, **99**, 83–104, <https://doi.org/10.1175/BAMS-D-16-0256.1>.
- Geerts, B., and Coauthors, 2017: The 2015 Plains Elevated Convection At Night field project. *Bull. Amer. Meteor. Soc.*, **98**, 767–786, <https://doi.org/10.1175/BAMS-D-15-00257.1>.
- Ingleby, B., 2017: An assessment of different radiosonde types 2015/2016. ECMWF Tech. Memo. 807, 71 pp., <https://www.ecmwf.int/en/elibrary/17551-assessment-different-radiosonde-types-2015-2016>.
- Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. ECMWF Tech. Memo. 636, 48 pp., <https://www.ecmwf.int/en/elibrary/10125-ensemble-data-assimilations-ecmwf>.
- Jensen, M. P., D. J. Holdridge, P. Survo, R. Lehtinen, S. Baxter, T. Toto, and K. L. Johnson, 2016: Comparison of Vaisala radiosondes RS41 and RS92 at the ARM Southern Great Plains site. *Atmos. Meas. Tech.*, **9**, 3115–3129, <https://doi.org/10.5194/amt-9-3115-2016>.
- Lang, S. T. K., M. Bonavita, and M. Leutbecher, 2015: On the impact of re-centring initial conditions for ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **141**, 2571–2581, <https://doi.org/10.1002/qj.2543>.
- Lavers, D. A., R. P. Allan, E. F. Wood, G. Villarini, D. J. Brayshaw, and A. J. Wade, 2011: Winter floods in Britain are connected to atmospheric rivers. *Geophys. Res. Lett.*, **38**, L23803, <https://doi.org/10.1029/2011GL049783>.
- , M. J. Rodwell, D. S. Richardson, F. M. Ralph, J. D. Doyle, C. A. Reynolds, V. Tallapragada, and F. Pappenberger, 2018: The gauging and modeling of rivers in the sky. *Geophys. Res. Lett.*, **45**, 7828–7834, <https://doi.org/10.1029/2018GL079019>.
- , A. Beljaars, D. S. Richardson, M. J. Rodwell, and F. Pappenberger, 2019: A forecast evaluation of planetary boundary layer height over the ocean. *J. Geophys. Res. Atmos.*, **124**, 4975–4984, <https://doi.org/10.1029/2019JD030454>.

- Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539, <https://doi.org/10.1016/j.jcp.2007.02.014>.
- , and Coauthors, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, <https://doi.org/10.1002/qj.3094>.
- Neiman, P. J., L. J. Schick, F. M. Ralph, M. Hughes, and G. A. Wick, 2011: Flooding in western Washington: The connection to atmospheric rivers. *J. Hydrometeor.*, **12**, 1337–1358, <https://doi.org/10.1175/2011JHM1358.1>.
- Ralph, F. M., P. J. Neiman, 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, 2016: CalWater field studies designed to quantify the roles of atmospheric rivers and aerosols in modulating U.S. West Coast precipitation in a changing climate. *Bull. Amer. Meteor. Soc.*, **97**, 1209–1228, <https://doi.org/10.1175/BAMS-D-14-00043.1>.
- , and Coauthors, 2017: Dropsonde observations of total integrated water vapor transport within North Pacific atmospheric rivers. *J. Hydrometeor.*, **18**, 2577–2596, <https://doi.org/10.1175/JHM-D-17-0036.1>.
- , M. D. Dettinger, M. M. Cairns, T. J. Galarneau, and J. Eylander, 2018: Defining “atmospheric river”: How the glossary of meteorology helped resolve a debate. *Bull. Amer. Meteor. Soc.*, **99**, 837–839, <https://doi.org/10.1175/BAMS-D-17-0157.1>.
- Ramos, A. M., R. M. Trigo, M. L. R. Liberato, and T. Ricardo, 2015: Daily precipitation extreme events in the Iberian Peninsula and its association with atmospheric rivers. *J. Hydrometeor.*, **16**, 579–597, <https://doi.org/10.1175/JHM-D-14-0103.1>.
- Reynolds, C. A., J. D. Doyle, F. M. Ralph, and R. Demirdjian, 2019: Adjoint sensitivity of North Pacific atmospheric river forecasts. *Mon. Wea. Rev.*, **147**, 1871–1897, <https://doi.org/10.1175/MWR-D-18-0347.1>.
- Rodwell, M. J., and Coauthors, 2013: Characteristics of occasional poor medium-range weather forecasts for Europe. *Bull. Amer. Meteor. Soc.*, **94**, 1393–1405, <https://doi.org/10.1175/BAMS-D-12-00099.1>.
- , S. T. Lang, N. B. Ingleby, N. Bormann, E. Hólm, F. Rabier, D. S. Richardson, and M. Yamaguchi, 2016: Reliability in ensemble data assimilation. *Quart. J. Roy. Meteor. Soc.*, **142**, 443–454, <https://doi.org/10.1002/qj.2663>.
- Schäfler, A., and Coauthors, 2018: The North Atlantic Waveguide and Downstream Impact Experiment. *Bull. Amer. Meteor. Soc.*, **99**, 1607–1637, <https://doi.org/10.1175/BAMS-D-17-0003.1>.
- Stone, R. E., C. A. Reynolds, J. D. Doyle, R. Langland, N. 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>.
- Uttal, T., and Coauthors, 2002: Surface heat budget of the Arctic Ocean. *Bull. Amer. Meteor. Soc.*, **83**, 255–276, [https://doi.org/10.1175/1520-0477\(2002\)083<0255:SHBOTA>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0255:SHBOTA>2.3.CO;2).
- Vaisala, 2010: Vaisala dropsonde RD94. Vaisala Rep., 2 pp., <https://www.vaisala.com/sites/default/files/documents/RD94-Dropsonde-Datasheet-B210936EN-A-LoRes.pdf>.
- , 2018: Vaisala dropsonde RD41. Vaisala Rep., 2 pp., <https://www.vaisala.com/sites/default/files/documents/RD41-Datasheet-B211706EN.pdf>.