Forecast evaluation of the North Pacific jet stream using AR Recon dropwindsondes

Lavers, David A.; Torn, Ryan D.; Davis, Chris; Richardson, David S.; Ralph, F. Martin; Pappenberger, Florian

DOI:
10.1002/qj.4545

Received: 7 February 2023 | Revised: 14 July 2023 | Accepted: 17 July 2023

DOI: 10.1002/qj.4545

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Forecast evaluation of the North Pacific jet stream using AR Recon dropwindsondes

David A. Lavers[1,2] | **Ryan D. Torn**[3] | **Chris Davis**[4] | **David S. Richardson**[1] | **F. Martin Ralph**[5] | **Florian Pappenberger**[1]

[1]European Centre For Medium-Range Weather Forecasts (ECMWF), Reading, UK

[2]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK

[3]Department of Atmospheric and Environmental Science, University at Albany, State University of New York, Albany, New York, USA

[4]National Center for Atmospheric Research, Boulder, Colorado, USA

[5]Center for Western Weather and Water Extremes (CW3E), Scripps Institution Of Oceanography, University of California, San Diego, La Jolla, California, USA

Correspondence
David A. Lavers, European Centre for Medium-Range Weather Forecasts (ECMWF), Shinfield Park, Reading, RG2 9AX, UK.
Email: david.lavers@ecmwf.int

Funding information
Copernicus Climate Change Service

Abstract

The term jet stream generally refers to a narrow region of intense winds near the top of the midlatitude or subtropical troposphere. It is in the midlatitude jet stream where instabilities and waves may develop into synoptic-scale systems, which in turn makes accurately resolving the structure of the jet stream and associated features critical for atmospheric development, predictability, and impacts, such as extreme precipitation and winds. Using dropwindsonde observations collected during the Atmospheric River Reconnaissance (AR Recon) campaign from 2020 to 2022, this study assesses the North Pacific jet stream structure in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS). Results show that the IFS has a slow-wind bias on the lead times assessed, with the strongest winds ($\geq 50\ m \cdot s^{-1}$) having a bias of up to $-1.88\ m \cdot s^{-1}$ on forecast day 4. Also, the IFS cannot resolve the sharp potential vorticity (PV) gradient across the jet stream and tropopause, and this PV gradient weakens with forecast lead time. Cases with larger wind biases are characterized by higher PV biases and PV biases tend to be larger for cases with a higher horizontal PV gradient. These results suggest that further model-based experiments are needed to identify and address these biases, which could ultimately yield increased forecast accuracy.

KEYWORDS

diagnostics, ECMWF forecasts, jet stream, northern Pacific Ocean, observational campaigns

1 | INTRODUCTION

The term jet stream generally refers to a narrow region of intense winds near the top of the midlatitude or subtropical troposphere. In the midlatitudes, instabilities and waves may develop along the jet stream into synoptic-scale systems, or midlatitude cyclones, which thus makes the jet stream critical for atmospheric development and predictability. Furthermore, their key role in cyclogenesis means they are linked with atmospheric rivers (ARs; e.g., Ralph *et al.*, 2018), warm conveyor belts (WCBs; e.g., Browning, 1986), extreme precipitation and flooding (Ralph *et al.*, 2006; Lavers *et al.*, 2011; Neiman *et al.*, 2011; Corringham *et al.*, 2019), severe winds (e.g., Browning, 2004), cold-air outbreaks (Linkin and Nigam, 2008), and ocean waves (e.g., Cordeira and Bosart, 2010). In the

western United States, for example, they are part responsible, via ARs, for a large proportion of the water supply (Dettinger *et al.*, 2011). The jet stream is also connected with clear-air turbulence which is a hazard for aviation (Koch *et al.*, 2005).

In the midlatitudes, the horizontal temperature gradient between the cold polar and warm subtropical air masses (i.e., the polar-frontal zone) is associated with the presence of the jet stream near the tropopause through the thermal-wind relation. Typically, the jet stream is associated with a strong horizontal potential vorticity (PV) gradient, with lower PV values typical of the troposphere on the warm side and higher values, typical of the stratosphere on the cold side and hence a depressed tropopause. The maximum jet stream wind speed – the jet core – is found in the vicinity of the sharpest PV gradient, which is often characterized by a nearly vertical tropopause. This large horizontal PV gradient can act as a waveguide for Rossby waves (Hoskins and Ambrizzi, 1993; Schwierz *et al.*, 2004; Martius *et al.*, 2010), which are often associated with high-impact weather events (see review by Wirth *et al.*, 2018).

For numerical weather prediction (NWP) models to provide skillful weather forecasts, an accurate estimate of the jet structure is required, as this can impact the propagation of Rossby waves. From linear Rossby wave theory, the dispersion relationship, and hence the phase speed and group velocity of wave packets, depends on both the PV gradient and jet speed (e.g., Rossby, 1945); therefore the structure of the atmosphere near the jet is critical to understanding the evolution of upper-tropospheric midlatitude troughs and ridges. Single-layer, analytical models indicate that a weaker tropopause PV gradient would result in a weaker jet stream and weaker counterpropagation of Rossby waves against the mean flow, which have the potential to partially cancel each other (Harvey *et al.*, 2016) and result in the excessive filamentation of PV and subsequent weakening of Rossby waves (Harvey *et al.*, 2018).

Unfortunately, there are few observations which provide detailed information on the three-dimensional structure of the wind and tropopause structure near upper-tropospheric jets, particularly over the global oceans (e.g., Baker, 2014). This, in turn, affects the NWP initialization and forecasts of the jet stream, and limits the possibility for diagnostic and evaluation studies. Over continental regions, wind and temperature data from commercial aircraft can provide some information about the model's wind speeds, with the model generally characterized by slow biases, particularly at high wind speeds (e.g., Rickard *et al.*, 2001; Cardinali *et al.*, 2004). While aircraft-based data are ubiquitous over land and over certain ocean areas (e.g., flight corridors), most of these data are at a near-constant altitude, making it difficult to obtain the vertical profiles necessary to establish the necessary horizontal and vertical gradients, except near airports. Furthermore, satellite-based observations, including radiance measurements, and atmospheric motion vectors (AMVs) do not have sufficient vertical resolution to provide an accurate estimate of the fine vertical gradients present near the jet, although from 2018 to 2023, Aeolus was an important source of wind profile observations.

As an alternative, dedicated observational campaigns can provide opportunities to undertake model assessments of the jet stream. One recent campaign dedicated to the region around the jet stream was the North Atlantic Waveguide and Downstream Impact Experiment (NAWDEX; Schäfler, 2018). Schäfler *et al.* (2020) used observations taken during NAWDEX to assess the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) and the UK Met Office Unified Model. Like the aircraft-based studies, the results showed that the models had a slow-wind bias in the troposphere (and lower stratosphere) of $-0.41\ \mathrm{m{\cdot}s^{-1}}$ and $-0.15\ \mathrm{m{\cdot}s^{-1}}$ respectively, and large jet stream wind errors of up to $10\ \mathrm{m{\cdot}s^{-1}}$ in individual events, most prominent immediately above the tropopause on the flanks of upper-level ridges. Furthermore, the median vertical shear at and above the tropopause is underestimated by a factor of 1.5 to 5, which may be due to the lower vertical resolution in the model, which can be important near jets where the vertical shear is higher.

Beyond field campaigns, other studies have used model analyses as a proxy for the observed atmospheric state to investigate a model's ability to replicate the PV gradient across the tropopause. Gray *et al.* (2014) found that the PV gradient near the tropopause in the IFS and UK Met Office Unified model decreased with forecast lead time, particularly adjacent to a ridge, and when the model resolution was decreased at longer lead times, indicating that the models could not preserve the sharp PV gradients. Saffin *et al.* (2017) indicates that the weakening of the tropopause PV gradient in the UK Met Office Unified Model is due to the advection, while non-conservative processes, such as from parameterizations, counteract this and sharpen the tropopause.

The observations gathered during the Atmospheric River Reconnaissance (AR Recon; Ralph *et al.*, 2020) campaign provide a unique opportunity to investigate the structure of the North Pacific jet stream. The AR Recon campaign is a research and operations partnership whose main aim is to help better inform decision-makers on water management and flooding in the western United States via the improvement of NWP forecasts. In each winter season, AR Recon uses research aircraft to probe ARs and other dynamically active regions related to ARs,

whereby dropwindsonde observations – of specific humidity, temperature, and winds – are collected and assimilated in real-time into global NWP systems, such as the ECMWF IFS, to improve the initialization of the next forecast. One research aircraft, the National Oceanic and Atmospheric Administration (NOAA) Gulfstream IV-SP (G-IV), is of great interest here because it typically deploys dropwindsondes from an altitude of 150 hPa, which is above the typical jet stream. For a subset of the intensive observing periods (IOPs), the G-IV flight pattern also transected the jet stream axis with dropwindsonde spacing of roughly 100 km, which provides a unique opportunity to investigate the cross-jet structure.

The goal of this study is to utilize the dropwindsonde observations taken by the G-IV in the 2020, 2021, and 2022 AR Recon seasons to evaluate the jet stream and associated PV structure in the ECMWF IFS. In so doing, the following questions are addressed in this article: (1) are there any model biases, for example in terms of wind speed and temperature, in the jet stream region; and (2) how does the jet stream structure evolve in the IFS forecasts? This work is akin to previous studies that have used either sonde transects alone (e.g., Danielsen and Mohnen, 1977; Danielsen *et al.*, 1987; Harvey *et al.*, 2020) or a combination of aircraft data and sondes (e.g., Shapiro, 1974, 1976; Danielsen *et al.*, 1987; Shapiro *et al.*, 1987). The unique aspect of this study is the number of transects with different kinematic and thermodynamic properties (21 over three years), which provides the opportunity to assess whether systematic issues exist within the model, rather than for a single case, which may not be representative. This paper proceeds as follows: Section 2 provides a description of the data and methods, Section 3 provides a model verification both in the raw dropwindsonde data and with respect to PV, while Section 4 provides a summary and discussion of the results.

2 | DATA AND METHODS

2.1 | Dropwindsonde observations

During the 2020, 2021, and 2022 AR Recon seasons, the NOAA G-IV deployed 1,170 troposphere-deep dropwindsondes, with their locations shown in Figure 1a. This illustrates that a broad area of the northern Pacific was sampled, as in 2020 the G-IV was based out of Portland, Oregon, and thus sampled systems closer to the North American continent, while in 2021 and 2022 the G-IV was based in Honolulu, Hawaii, and sampled features further south and west. During all three years, the G-IV deployed Vaisala RD41 dropwindsondes, which have an accuracy for pressure and temperature of 0.4 hPa and 0.1 K respectively (Vaisala, 2018). These dropwindsonde observations – gathered in IOPs and mostly within a six-hour window centred on 0000 UTC – were transferred to the World Meteorological Organization Global Telecommunications System (GTS) and ingested into operational NWP systems including the ECMWF IFS. For the 1,170 dropwindsonde reports received over the GTS, the IFS assimilated a median number of 246 vertical levels of wind data per profile, including both standard and significant levels. This represents fewer pressure levels than the sample collected by the dropwindsondes on their descent to the ocean surface generally due to thinning undertaken within the IFS.

2.2 | ECMWF long-window data assimilation system and forecasts

The long-window data assimilation (LWDA) system of the IFS consists of a 12-hr window where short-range (3–15 hr) or background forecasts are combined with all observations (including the dropwindsonde data) via a four-dimensional variational data assimilation (4D-Var) process. This procedure produces a new LWDA analysis which represents the best estimate of the current atmospheric state at a specific time. It is the short-range forecasts from the LWDA analysis combined with observations in the early-delivery data assimilation cycle that form the initial conditions for the single high-resolution IFS operational forecast (Lean *et al.*, 2019). Herein, these background forecasts and analyses of the LWDA system, retrieved from the ECMWF archive and interpolated within the IFS to the dropwindsonde vertical profiles, are assessed.

For validations beyond the analysis and background forecast, high-resolution IFS forecasts ($0.1° \times 0.1°$ regular grid) valid during the IOPs on model levels were also retrieved from the ECMWF archive. This evaluation considered the forecasts from 0000 UTC on days 2 and 4, where on day 2, the forecasts were available during each IOP every hour (i.e., T + 45, T + 46, to T + 51), and on day 4, they were available during each IOP every three hours (T + 93, T + 96, T + 99). During the 2020–2022 AR Recon campaigns, the operational IFS forecasts had 137 vertical model levels, which in the upper troposphere (200–500 hPa) provided a vertical resolution of about 300 m. In order to create a regular vertical spacing, the model level data were interpolated to 20-hPa resolution pressure levels between 200 hPa and 1,000 hPa. Instantaneous model estimates at the observation location are subsequently calculated by using the (1) nearest-neighbour approach to identify both the nearest horizontal grid point and pressure level and the (2) closest-forecast lead time

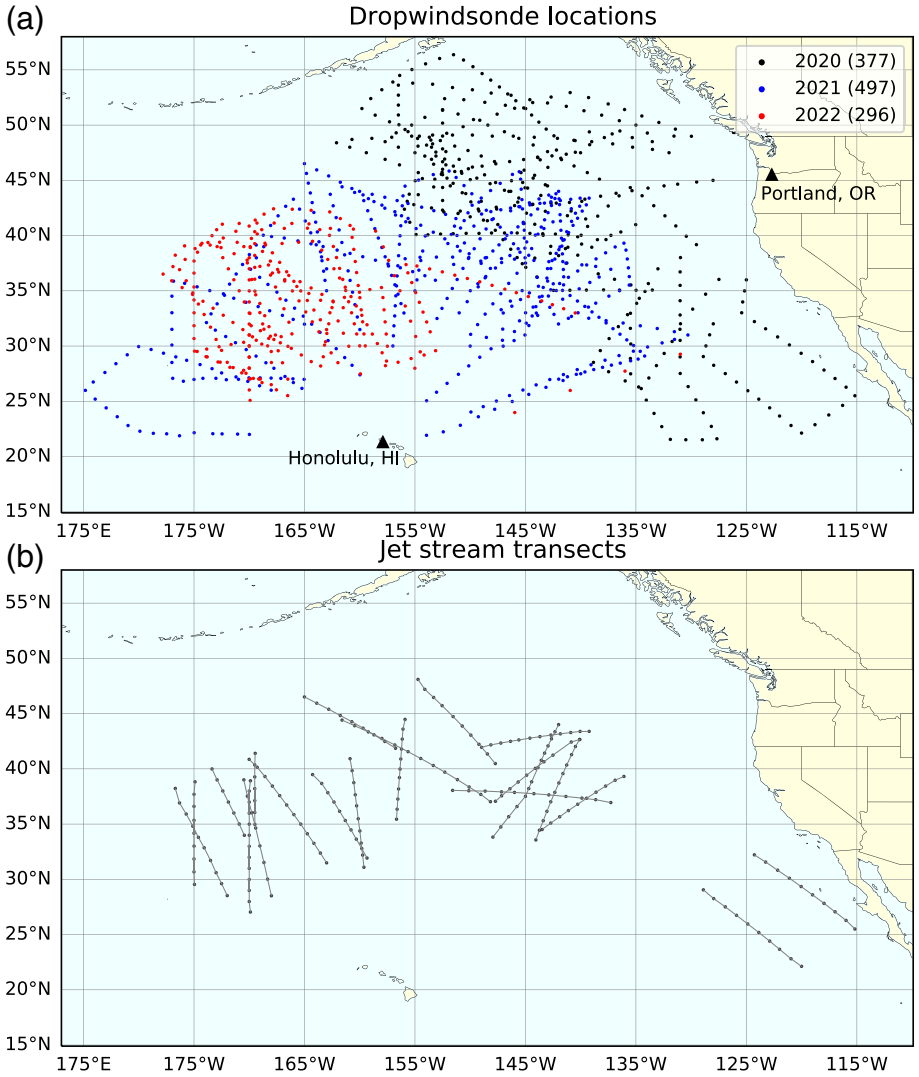

FIGURE 1 (a) The locations of the dropwindsondes deployed by the NOAA G-IV in 2020, 2021, and 2022. The NOAA G-IV was based in Portland, Oregon, in 2020, and in Honolulu, Hawaii, in 2021 and 2022. The number of dropwindsondes available in each year is given in the legend. (b) The 21 jet stream transects (grey lines) and the dropwindsondes along them (grey markers). [Colour figure can be viewed at wileyonlinelibrary.com]

to the dropwindsonde observation time. Note that the dropwindsonde observations and LWDA background and analysis profiles were also interpolated to the same 20-hPa resolution with the nearest-neighbour approach.

2.3 | Forecast evaluation

Historically, a different evaluation convention has been used in the data assimilation and forecast communities, with data assimilation groups calculating the observation-minus-background and the observation-minus-analysis departures (O–B and O–A; e.g., Desroziers *et al.*, 2005) and the forecast community computing the forecast-minus-observation statistic. For consistency and to avoid confusion, herein we use the forecast-minus-observation convention both for the evaluation of the LWDA and high-resolution forecast systems. For the LWDA, the mean and standard deviation – which represent the mean and random errors respectively – of the background-minus-observation (B–O) and analysis-minus-observation (A–O) departures were calculated in 20-hPa layers from the surface (1,000 hPa) to 200 hPa using all assimilated pressure levels from the 1,170 dropwindsonde profiles (261,026 levels for the winds; 263,702 levels for the temperature). This approach allows for the identification of potential model biases and problems in different atmospheric layers for the wind speed and temperature. These statistics were also calculated on a subset of 17 IOPs which were used to investigate the jet stream structure. Furthermore, the relationship between the observed and modelled winds was assessed in the LWDA and high-resolution forecasts using the 20-hPa resolution interpolated dropwindsonde and model data; this used 49,235 pressure levels.

As will be shown below, there appears to be a conditional bias in the wind speed forecasts within the IFS; therefore, the potential source of these biases is investigated in the vicinity of the upper-tropospheric jet stream. Here, the structure of the jet stream and associated

TABLE 1 The 21 jet stream transects used in the composite analysis; these were sampled during 17 intensive observing periods (IOPs).

Date and IOP number	First sonde	Last sonde	Number
15 February 2020 (IOP 8)	21:08	22:31	10
21 February 2020 (IOP 10)	21:07	22:33	10
21 February 2020 (IOP 10)	23:25	00:56	10
27 January 2021 (IOP 7)	21:26	22:38	10
27 January 2021 (IOP 7)	22:38	23:53	12
28 January 2021 (IOP 8)	21:37	22:46	10
22 February 2021 (IOP 15)	21:40	22:50	11
8 March 2021 (IOP 22)	21:00	22:33	14
9 March 2021 (IOP 23)	22:08	23:12	12
11 March 2021 (IOP 25)	22:36	23:36	9
12 March 2021 (IOP 26)	21:11	23:05	13
13 March 2021 (IOP 27)	22:58	00:43	11
3 February 2022 (IOP 8)	21:22	22:40	10
3 February 2022 (IOP 8)	23:22	01:04	13
24 February 2022 (IOP 11)	23:26	00:50	12
25 February 2022 (IOP 12)	21:07	22:49	8
9 March 2022 (IOP 19)	22:08	23:34	9
10 March 2022 (IOP 20)	00:05	00:55	7
11 March 2022 (IOP 21)	21:08	22:34	10
11 March 2022 (IOP 21)	23:40	00:35	7
12 March 2022 (IOP 22)	21:17	22:50	10

Note: The columns refer to the IOP date and number of each season, the time (UTC) of the first and last dropwindsonde in the transects, and the number of dropwindsondes along the transect.

dynamical fields is evaluated by considering the subset of dropwindsondes that represent transects that cross the upper-tropospheric jet at a relatively normal angle relative to the jet axis (no more than 30° normal to jet axis). Table 1 lists the 17 AR Recon IOPs where this occurred and Figure 1b shows the 21 jet stream transects. Each cross-section consists of between seven and 14 individual dropwindsondes, depending on the IOP, with more profiles generally on the warm side of the jet due to aircraft base locations, and limitations on the flight duration. Moreover, the primary purpose of each IOP was to sample the essential atmospheric structures and regions that would yield improvements in US West Coast forecasts (e.g., Cobb *et al.*, 2022), which in many cases did not necessarily mean sampling the jet.

For the analysis and background forecast, the assumption is that the model and observed jet are approximately in the same location; therefore, the validation does not reflect errors in position. By contrast, it is possible that the jet is not in the same geographic location as the observed jet for the day-2 and day-4 forecasts. Consequently, an earth-relative verification may reflect a mismatch of the jet position (i.e., the jet is too far north or south in the forecast), rather than errors in the structure (i.e., a Lagrangian verification). This issue is addressed by shifting the forecast jet position to the analyzed jet position based on the difference in the 'jet centroid' position in the forecast and verifying analysis. Here the jet centroid is defined as the horizontal mass centroid of the 200–300-hPa layer-average wind magnitude in the vicinity of the jet streak. This centroid approach has been used in cyclone tracking and tends to be smoother than comparing a grid point maximum (e.g., Nguyen *et al.*, 2014). Once the centroid is identified in both the forecast and verifying analysis, the difference between the forecast and analysis position is added to each of the observed dropwindsonde latitudes/longitudes, such that the model estimate of the dropwindsondes from

the day-2 and day-4 forecast are extracted in a jet-relative space. The mean absolute differences between the day-2 and day-4 forecast centroid positions and the analysis centroid positions are 82 and 175 km respectively, with a 69-km southward bias in the jet for the day-4 forecast (the only direction and forecast time that contains a statistically significant bias at the 95% significance level; not shown).

The structure of the tropopause is assessed by computing a pseudo-PV based on both observations and the model forecasts at the dropwindsonde locations. PV on a constant pressure surface is given by:

$$\mathrm{PV} = -g\left(\frac{\partial u}{\partial p}\frac{\partial \theta}{\partial y} - \frac{\partial v}{\partial p}\frac{\partial \theta}{\partial x} + \frac{\partial \theta}{\partial p}(f+\zeta)\right), \quad (1)$$

where u is the zonal wind, v is the meridional wind, θ is the potential temperature, f is the Coriolis parameter, and ζ is the vertical component of relative vorticity. For a zonal jet and a sequence of dropwindsondes that are oriented in the meridional direction, it will be possible to compute meridional and vertical gradients in the wind and potential temperature, but not the zonal gradient. Fortunately, the main horizontal gradients near the jet stream are in the cross-jet direction, so this does not substantially limit this analysis, particularly since the same method of calculating PV is employed for both the observations and model fields.

Consequently, for this evaluation, 'PV' is defined as:

$$\mathrm{PV} = -g\left(\frac{\partial u}{\partial p}\frac{\partial \theta}{\partial y} + \frac{\partial \theta}{\partial p}\left(f - \frac{\partial u}{\partial y}\right)\right), \quad (2)$$

where u is instead the wind normal to the dropwindsonde transect and the y direction along the dropwindsonde transect. For both the forecast and observations, all derivatives are computed using either the actual dropwindsonde data, or the model estimate of the dropwindsonde data at the location of the dropwindsondes with centred differences. This approach ensures a consistent comparison in 'PV' between the model and observations and is consistent with the method used by Harvey *et al.* (2020). We note that the dynamic tropopause is defined by the 2-PVU isosurface ($1\ \mathrm{PVU} = 10^{-6}\ \mathrm{K{\cdot}kg^{-1}{\cdot}m^{2}{\cdot}s^{-1}}$).

3 | RESULTS AND DISCUSSION

3.1 | Evaluation of the long-window data assimilation system and the forecast winds using all dropwindsondes

Figure 2a,b shows the mean and standard deviation of the B–O and A–O departures calculated in 20-hPa layers using all assimilated pressure levels from the 1,170 dropwindsonde profiles. First, for the wind speed and air temperature, it is evident that the biases and random errors are all reduced by the data assimilation step, as seen by the red A–O lines being closer to zero than the black B–O lines. Second, the figure shows that the bias and random error for the winds increases with height, which suggests a model slow-wind bias and a poorer fit for the stronger wind speeds of the jet stream (Figure 2a). Third, in Figure 2b, there is a negative B–O bias, or model cold bias, of approximately 0.4 K in the planetary boundary layer (PBL; 940–960 hPa) which is in agreement with previous research (Ingleby, 2017; Lavers *et al.*, 2020); and the relatively large random error of 1 K above the PBL may relate to problems in correctly positioning the moisture there. Figure 2c,d shows the mean and standard deviation of the B–O and A–O departures but now only using those dropwindsondes deployed during the 17 IOPs with the 21 jet stream transects. These results are similar to those using all dropwindsondes suggesting that there is little evidence for different statistics when only considering the IOPs used later for the investigation of the jet stream structure.

Using the 20-hPa resolution interpolated data, the relationship between the observed and model winds in the LWDA and high-resolution forecasts for individual observations is investigated in Figure 3. These scatterplots show that the mean error, or bias, is negative, which means that the model winds, on average, are weaker than the observed winds, implying that there is an overall model slow-wind bias. These biases are also significantly different from zero at the 99% confidence level. When only considering winds at or above $50\ \mathrm{m{\cdot}s^{-1}}$ – defined herein as jet stream winds – the model slow-wind bias is as much as $-1.88\ \mathrm{m{\cdot}s^{-1}}$ on forecast day 4 (Figure 3d), suggesting that the IFS underestimates the strongest jet stream winds. This is furthermore highlighted by the slope of the linear regression lines and the quantile–quantile points for the 95th and 99th percentiles being located below the 1:1 line, results which are most visible in the background and day-2 and day-4 forecasts (Figure 3b–d).

As the lead time decreases, the model fit to the observations improves, as shown by the smaller standard deviation of the departures. For example, for the jet stream winds, the random error reduces from $10.70\ \mathrm{m{\cdot}s^{-1}}$ on forecast day 4 (Figure 3d) to $6.85\ \mathrm{m{\cdot}s^{-1}}$ on forecast day 2 (Figure 3c) to $3.33\ \mathrm{m{\cdot}s^{-1}}$ in the background (Figure 3b). Following the data assimilation procedure, in the analysis the scatter of the points decreases again and the random error reduces further to $2.01\ \mathrm{m{\cdot}s^{-1}}$ (Figure 3a), which is illustrated by the linear regression line almost overlaying the 1:1 line. The slow-wind bias in the jet stream is also partly addressed, as it reduces to $-0.49\ \mathrm{m{\cdot}s^{-1}}$ in the analysis (Figure 3a).

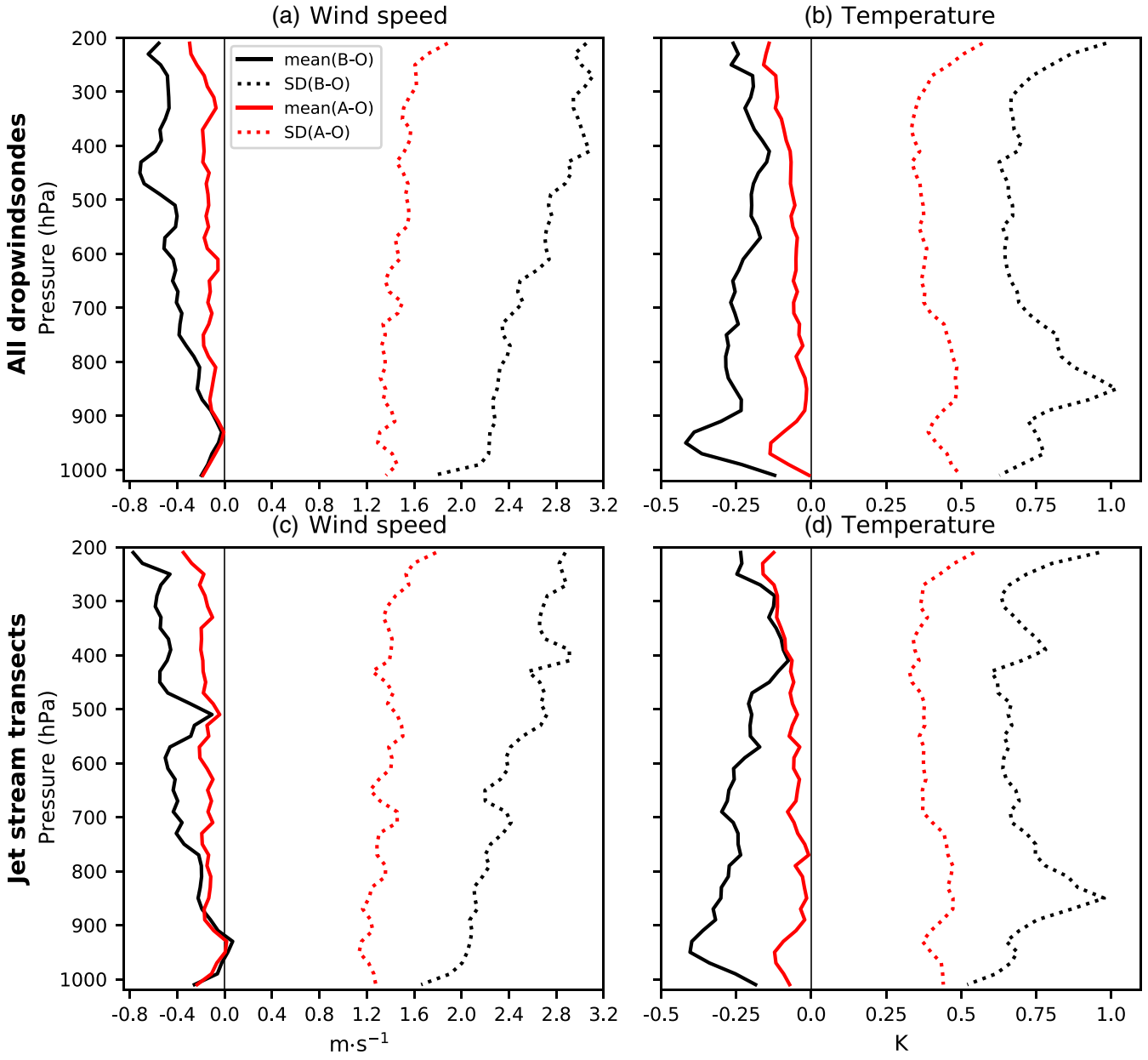

FIGURE 2 The mean (solid lines) and standard deviation (dotted lines) of the background-minus-observation (B–O) and the analysis-minus-observation (A–O) departures (using all available pressure levels in the long-window data assimilation [LWDA]) in 20-hPa layers at all dropwindsonde locations and at only those deployed during the intensive observing periods (IOPs) with the jet stream transects for (a, c) wind speed and (b, d) temperature. The median number of levels available in the layers is 6,148 for the winds and 6,182 for the temperature in the top row; and 2,857 for the winds and 2,855 for the temperature in the bottom row. [Colour figure can be viewed at wileyonlinelibrary.com]

3.2 | Example intensive observing periods on 3 February 2022, 13 March 2021, and 11 March 2022

The above results suggest a very prominent slow-wind bias at wind speeds above 50 $m \cdot s^{-1}$, which is mainly present in the vicinity of jet streams; therefore, the remainder of this analysis focuses on the validation of the jet structure along these dropwindsonde transects. Before evaluating the summary verification statistics, a subset of cases is presented to document the variety of tropopause structures that were present in the observations and how well the model qualitatively captured these features.

On 3 February 2022, the NOAA G-IV aircraft undertook a mission to the northwest of Hawaii to sample at a spacing of roughly 100 km the combination of an AR, WCB, and the jet stream. The sequence of 13 dropwindsondes across the jet stream along 170° W longitude provided

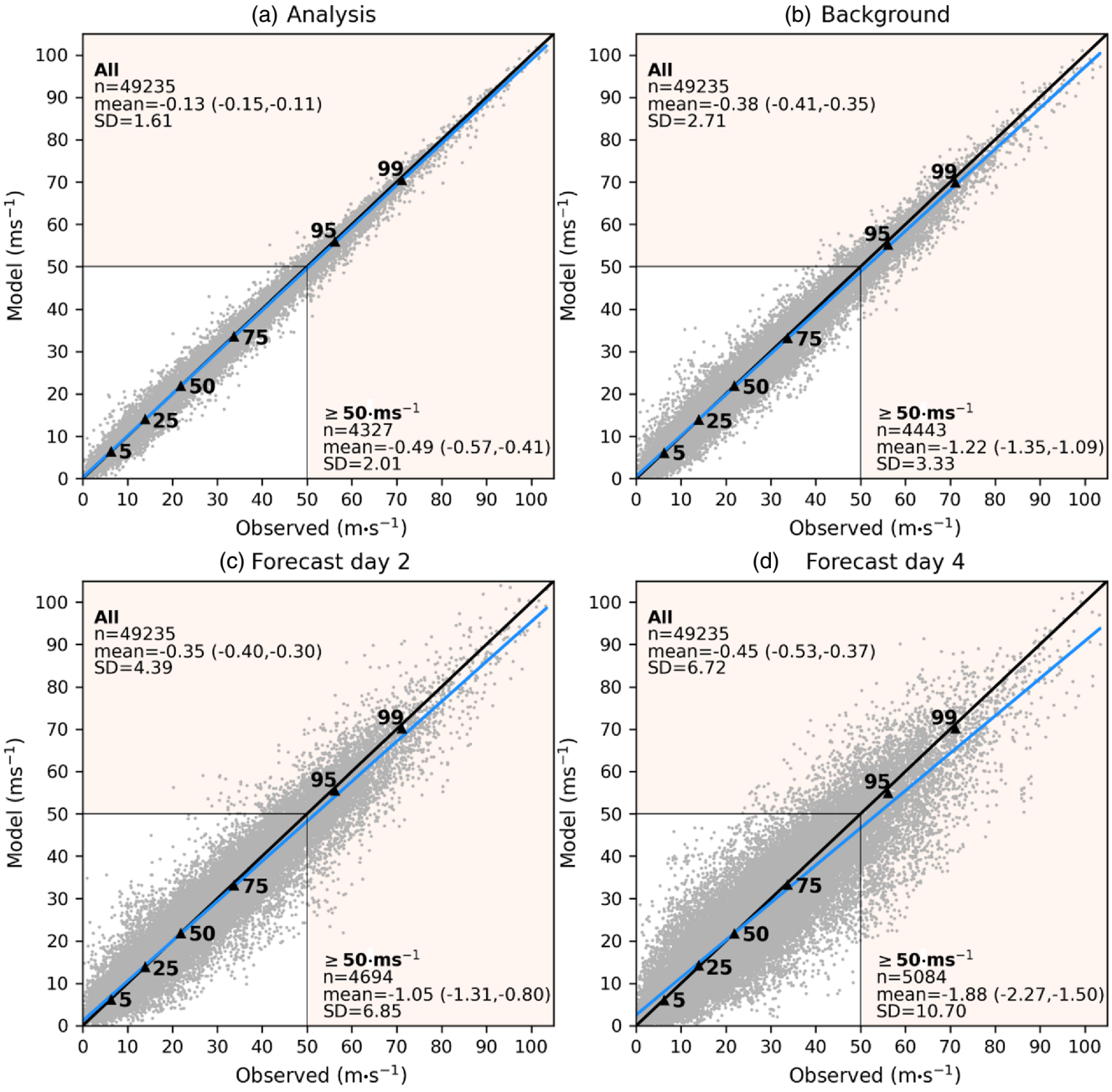

FIGURE 3 Scatterplots of the observed versus model winds for (a) the long-window data assimilation [LWDA] analysis, (b) the LWDA background (3–15 hr) forecasts, (c) forecast day 2, and (d) forecast day 4, in the 20-hPa resolution atmospheric profiles. In each panel, the sample size (n), and the mean and standard deviation of the forecast-minus-observation departures are given for all winds and for those ≥50 m·s^{-1}. The 99% confidence interval of the mean bias is also provided in brackets. Red-shaded regions represent winds ≥50 m·s^{-1} (i.e., jet stream winds) and the 1:1 and linear regression lines are shown in black and blue respectively. Quantile–quantile points are also plotted as black triangles. [Colour figure can be viewed at wileyonlinelibrary.com]

a transect perpendicular to the jet (Figure 4a). At 0006 UTC 3 February 2022, a dropwindsonde measured a maximum wind speed of 102.2 m·s^{-1} at 262.2 hPa at 34.0° N, 169.9° W. The horizontal location of the jet is characterized by a nearly vertical wall of PV, such that the 2-PVU contour is above the dropwindsonde release level to the south of the jet, but is near 400 hPa on the north side (Figure 4b). To the south of the jet, the dropwindsonde cross-section captures a robust upper-level front around 400 hPa, as indicated by the large horizontal potential temperature gradient, colocated with a narrow sloping corridor of PV in excess of 2 PVU. The remaining panels of this figure show the corresponding wind speed, PV, and potential temperature from the IFS at different forecast lead times valid at the time of

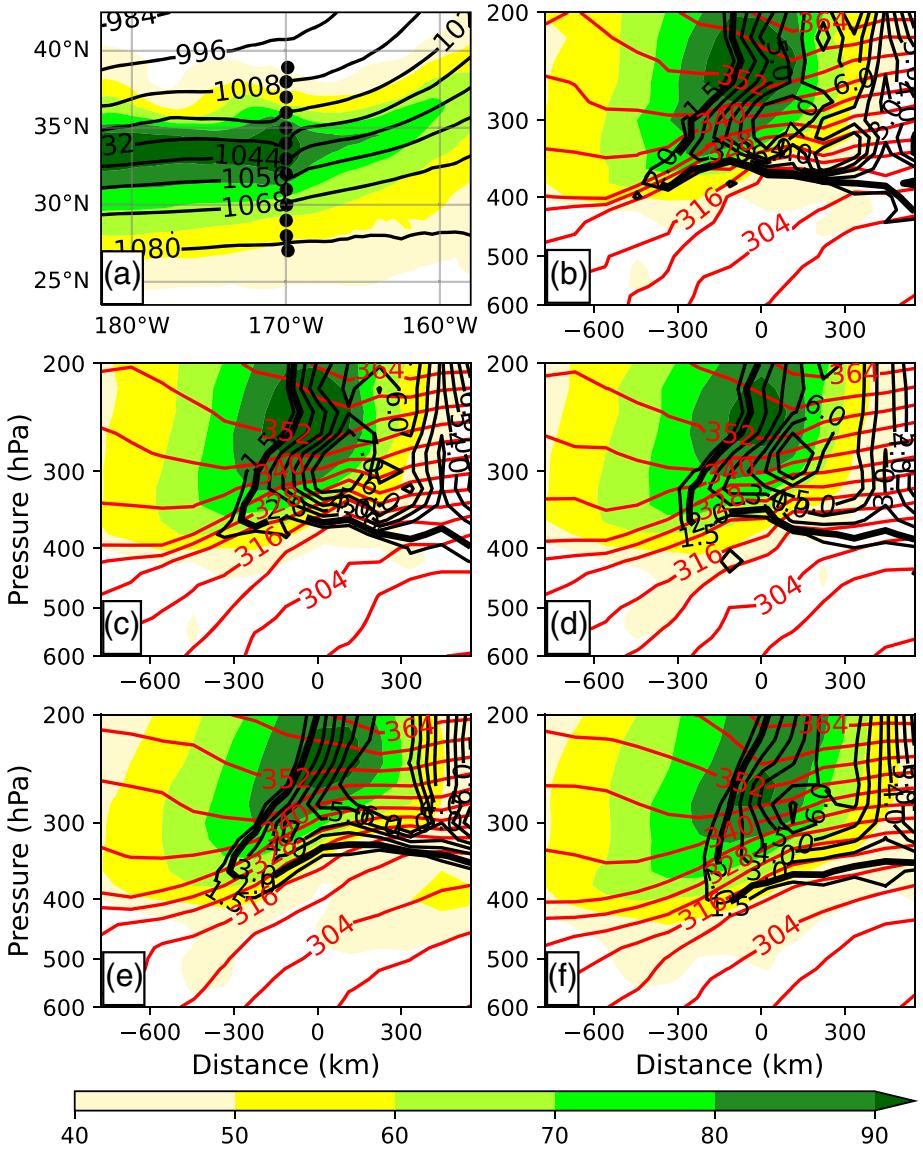

FIGURE 4 An example of a dropwindsonde cross-section on February 3, 2022 (IOP 8). Panel (a) shows the analysis 250-hPa wind speed (shading; units: $m \cdot s^{-1}$) and heights (contours; units: Dm) and the launch location of the dropwindsondes (black dots). Cross-sections of the jet stream in the (b) observations, (c) analysis, (d) background, and on (e) forecast day 2, and (f) forecast day 4. In panels (b–f), the wind speed is shown as shading (units: $m \cdot s^{-1}$), the potential vorticity (PV) is given by black contours (units: PVU) with the 2-PVU contour plotted as a thick line, and the potential temperature is shown by red contours (units: K). The horizontal axis denotes distance relative to the jet maxima, with negative values indicating the warm (i.e., south) side of the jet. [Colour figure can be viewed at wileyonlinelibrary.com]

the dropwindsondes. While the IFS does a reasonable job of capturing both the magnitude and structure of the jet in the analysis and background forecast (Figure 4c,d), the jet magnitude is weaker than observations in the day-2 and day-4 forecasts and the structure is less horizontally extensive (Figure 4e,f). Furthermore, all three forecasts provide a fairly coarse representation of the upper-level front. First, the front only extends 300 km south of the jet core, and second, the PV gradient at the 2-PVU contour is not as sharp, even in the model analysis after the dropwindsonde observations have been assimilated (Figure 4c). There is also evidence that with increasing lead time the sharpness of the upper-level front is weakening or diffusing away (Figure 4f).

Other dropwindsonde transects have a similar ability to capture the structure of the jet at short forecast lead times and distinctive biases that emerge at longer lead times. During 13 March 2021, the G-IV was tasked with sampling the western side of a deep upper-level trough along 150° W to the north of Hawaii. Eleven dropwindsondes provide a transect of the jet streak on the west side of the trough (Figure 5a), though not quite at an angle normal to the jet. This transect depicts a horizontally narrow, but vertically elongated jet exceeding 70 $m \cdot s^{-1}$ at 280 hPa along a nearly vertical wall of PV and hence a nearly vertical tropopause, where the PV is less than 1 PVU and the tropopause is above 200 hPa to the east of the jet, while the PV exceeds 3 PVU as low as 500 hPa within the trough. It is worth pointing out the vertically oriented PV minimum that is located around +300 km in the horizontal. This PV minimum is colocated with the trough axis and is partially an artefact of the method used to calculate the PV in this study. As mentioned in Section 2.3, the dropwindsonde transects do not allow for the calculation of derivatives perpendicular to the cross-section (e.g., $\frac{\partial v}{\partial p}\frac{\partial \theta}{\partial x}$, $\frac{\partial v}{\partial x}$), which are typically small within the jet because the vertical vorticity is

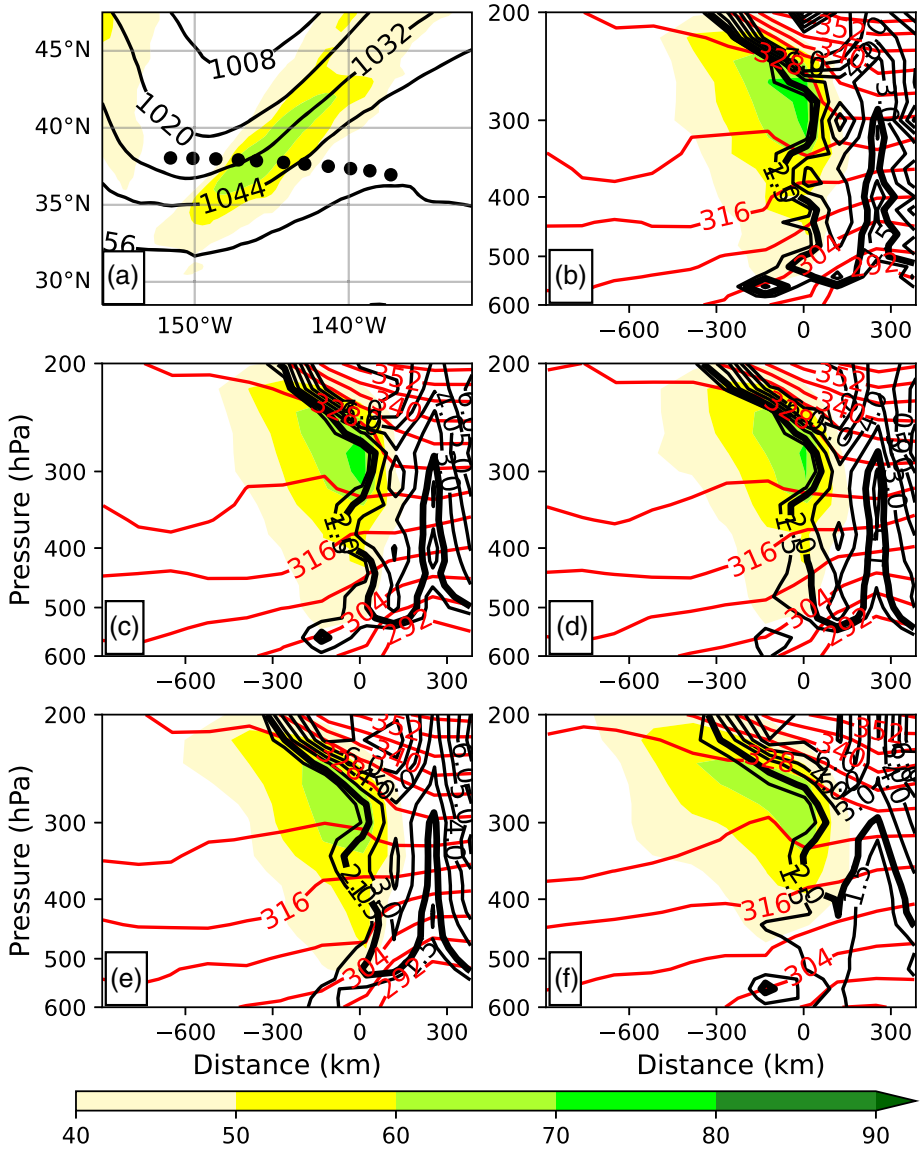

FIGURE 5 As in Figure 4, but for the dropwindsondes at 0000 UTC 13 March 2021 (IOP 27). [Colour figure can be viewed at wileyonlinelibrary.com]

dominated by the wind gradient in the direction of the transect in the vicinity of the jet. By contrast, locations near the trough axis have a greater proportion of curvature vorticity, which requires the perpendicular gradient terms. Estimating these terms from the gridded model data suggests that the $\frac{\partial v}{\partial p}\frac{\partial \theta}{\partial x}$ term does not provide a substantial contribution to PV for this cross-section, while $\frac{\partial v}{\partial x}$ can yield 20%–30% differences in PV at some locations, particularly above and on the cold side of the jet and in locations with curvature (not shown). As a reminder, the PV is computed in the same way in both the model and observations, so this deficiency is present in both and the comparison is still appropriate.

For both the analysis and background forecast, the IFS model is able to replicate the important structures of this jet, including the magnitude of the jet, the elongated structure of the jet, and the depth of the tropopause (Figure 5c,d). In general, the wind speed and PV differences in the analysis and background forecast are less than 4 $m \cdot s^{-1}$ and 0.5 PVU respectively, except at individual points. By contrast, the day-2 forecast had the appropriate jet and PV structure, but the magnitude of the wind is 7 $m \cdot s^{-1}$ slower than the observations in the jet core (Figure 5e). Finally, the day-4 forecast is characterized by significant differences with respect to the dropwindsonde observations, whereby the jet is more horizontally elongated, rather than vertically elongated, weaker than observations, and the tropopause has a less steep orientation, sloping gradually to 400 hPa at 400 km from the jet axis (Figure 5f). Consequently, the wind and PV errors exceed 10 $m \cdot s^{-1}$ and 1.5 PVU respectively, over many points near the tropopause.

The last case described here is from 11 March 2022, which was also characterized by a weaker jet streak at longer lead times. This mission sampled a jet that was zonally oriented near the dateline to the northwest of

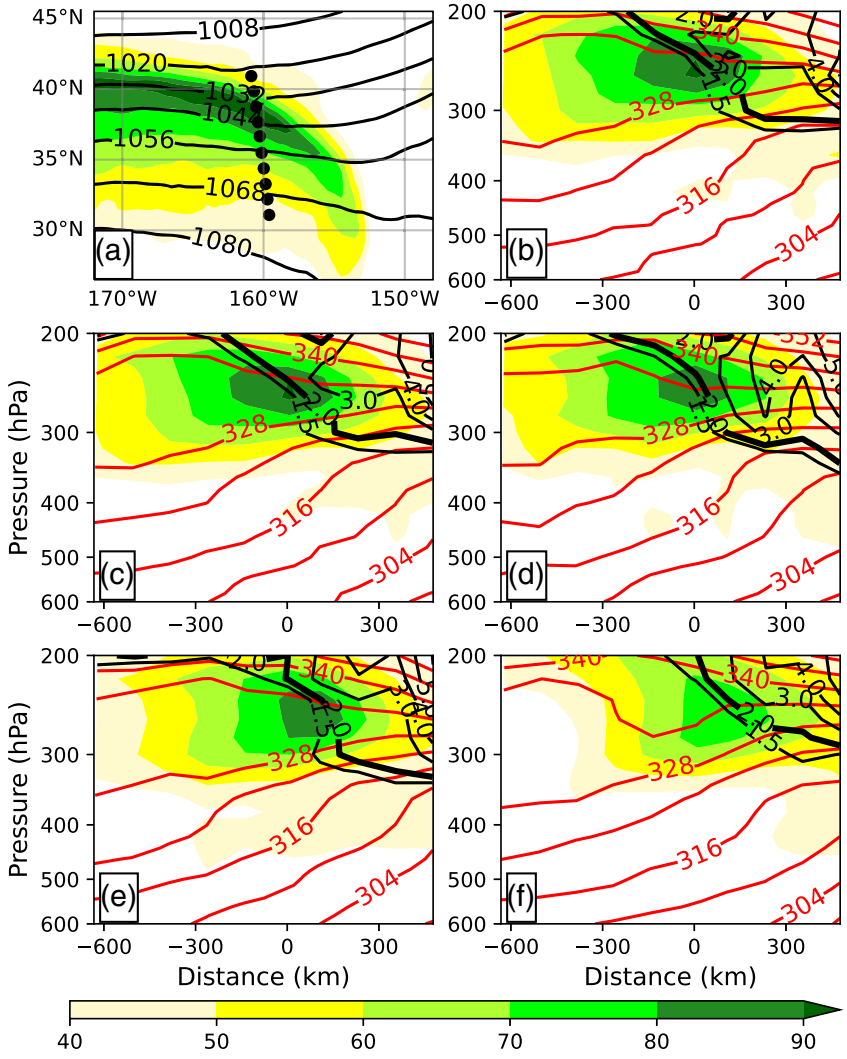

FIGURE 6 As in Figure 4, but for the dropwindsondes at 0000 UTC 11 March 2022 (IOP 21). [Colour figure can be viewed at wileyonlinelibrary.com]

Hawaii and turned anticyclonically along 160° W, where the dropwindsonde cross-section was taken (Figure 6a). Similar to the other two cases, the 90 m·s^{-1} jet was colocated with the steepest point in the tropopause, with the 2 PVU contour extending to 300 hPa on the poleward side of the jet (Figure 6b). Whereas the analysis did a respectable job with both the jet magnitude and location of the tropopause, the background forecast (Figure 6d), day-2 forecast (Figure 6e), and day-4 forecast (Figure 6f) became progressively weaker with increased lead time, such that the jet was 16 m·s^{-1} weaker than observations in the day-4 forecast. Consequently, the day-4 forecast was characterized by a shallower slope to the 2-PVU contour which is used to denote the tropopause.

3.3 | Composite analysis of the jet stream structure

We now use 21 jet stream transects (Table 1) to evaluate the composite errors in the jet stream structure. One of the drawbacks of this approach is the variety of jet stream structures that are present in the dropwindsonde transects, as demonstrated in the previous subsection. Consequently, error composites over all cases are likely to smear out the details that might be present in the errors – even though all cross-sections are horizontally aligned with the jet – due to the variety of jet vertical locations and structures. Therefore, the focus will be on the general trends in the error distribution over all cases. Figure 7 displays the composite mean observation (as line contours) and the composite mean bias (as shading), while Figure 8 provides the mean absolute error of the wind speed, potential temperature, and PV in the LWDA analysis (panels a–c) and background forecasts (panels d–f). The LWDA analysis generally has smaller errors than the LWDA background forecasts, which is due in part to the analysis assimilating the AR Recon dropwindsondes used here. The wind speed bias composite in the background in Figure 7d corroborates the slow bias for the strongest wind speeds found in Section 3.1, as seen by the statistically significant negative bias of up to 1.6 m·s^{-1} between 200 hPa and 300 hPa

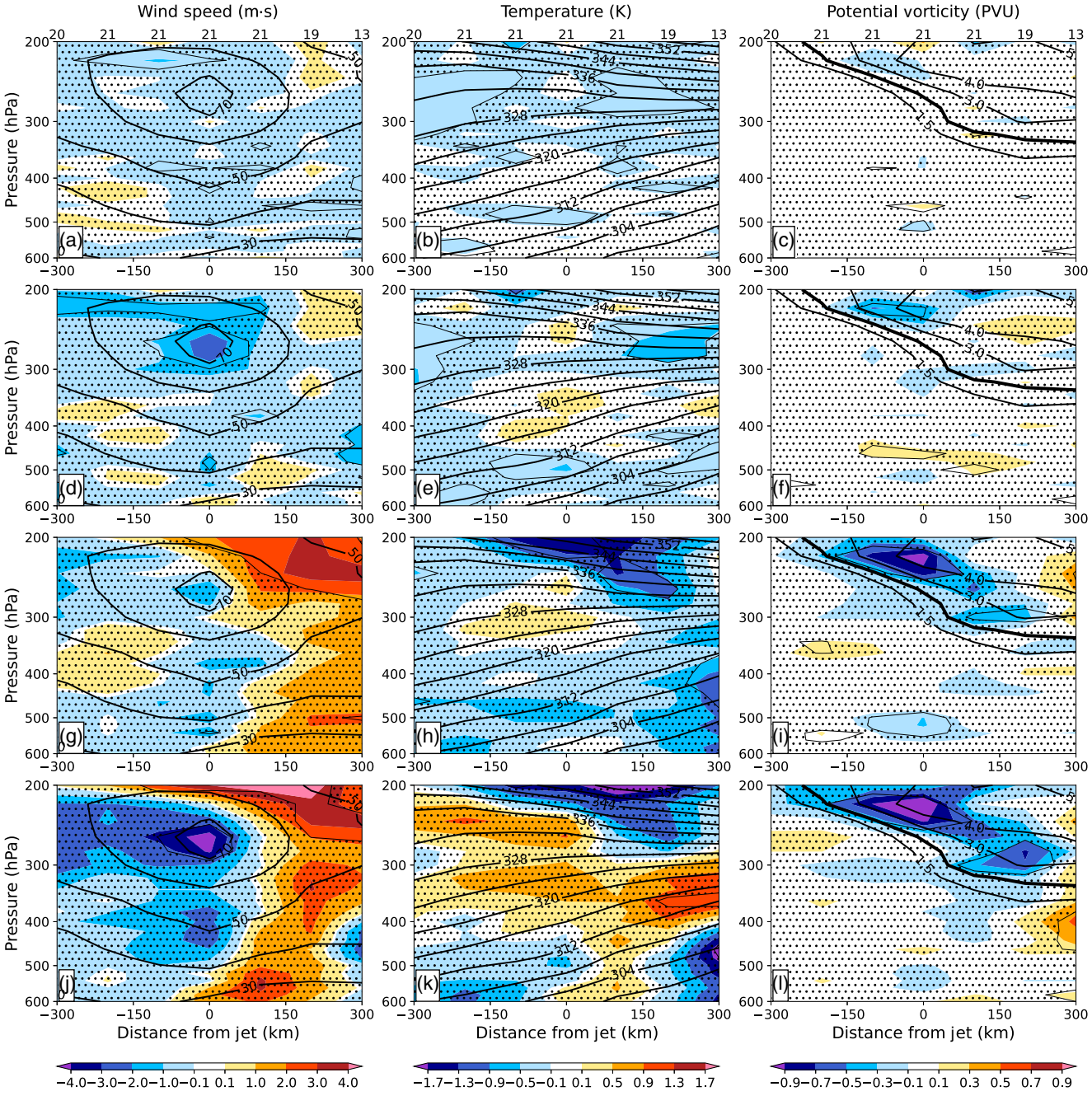

FIGURE 7 Cross-section of the mean bias of the wind speed (left column; units: $m \cdot s^{-1}$), potential temperature (middle column; units: K), and potential vorticity (PV) (right column; units: PVU) in the analysis (top row; a–c), background forecasts (second row; d–f), forecast day 2 (third row; g–i), and forecast day 4 (bottom row; j–l) averaged over all cases (shading). The contours in each figure denote the composite observation field. The horizontal axis denotes distance relative to the jet maxima, with negative values indicating the warm side (generally) to the south of the jet. The numbers along the top of panels (a–c) signify the number of dropwindsonde profiles used in the composite at those locations. Stippled regions indicate where the bias is not statistically significant at the 95% confidence interval based on a Student *t*-test. [Colour figure can be viewed at wileyonlinelibrary.com]

in the jet core. Moreover, there is a significant cold bias 150 km to the north of the jet above 300 hPa (Figure 7e). In terms of PV, the model underestimates this quantity above the jet core (below 200 hPa) on the stratospheric side of the tropopause where the largest horizontal gradient in the observed PV occurs (Figure 7f). For the mean absolute errors, the wind speed errors are generally higher when the observed wind speed is greater than 60 $m \cdot s^{-1}$ (Figure 8a,d), while the temperature and PV have the largest errors above the jet, particularly at 200 hPa (Figure 8b,c,e,f).

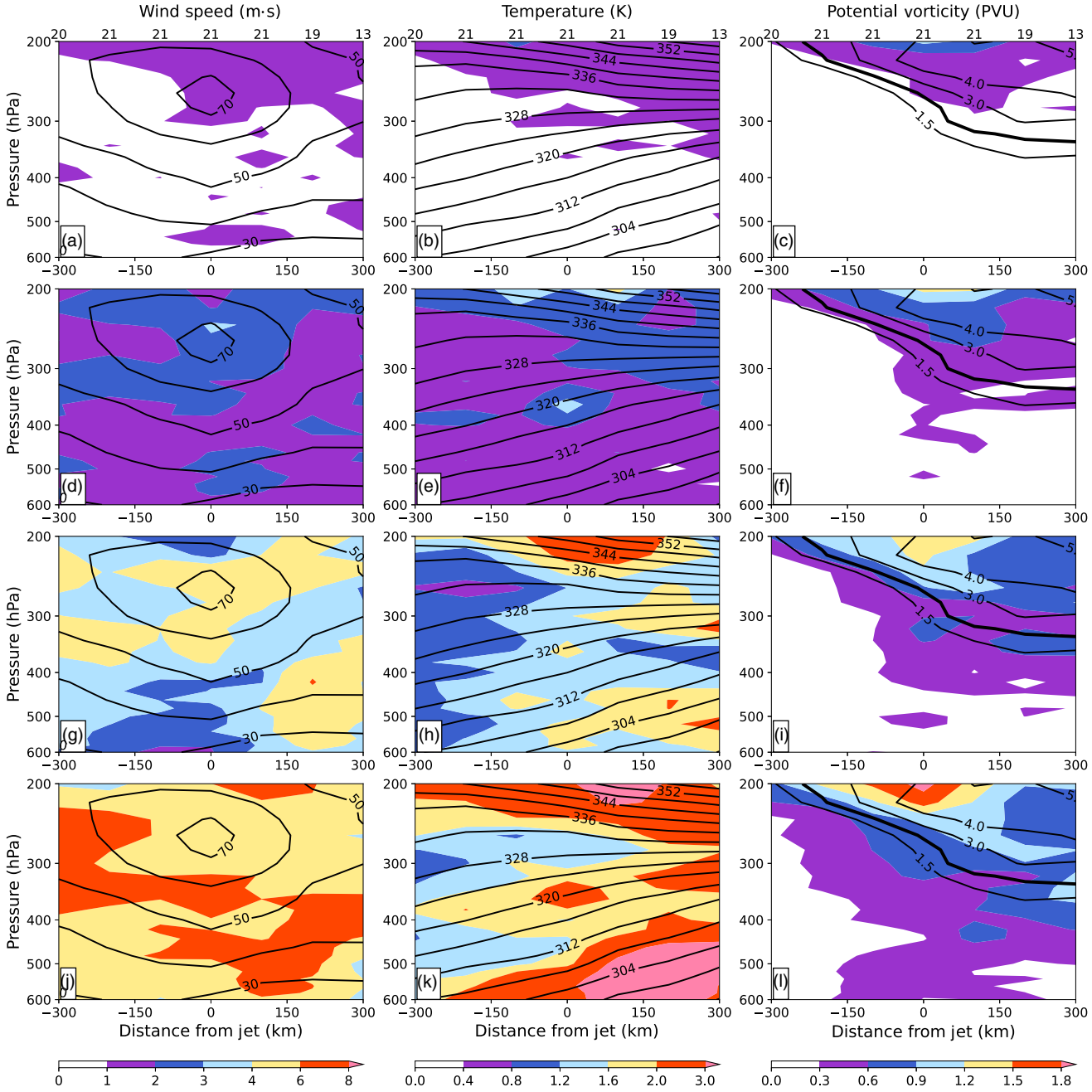

FIGURE 8 Same as Figure 7 but where the shading is the mean absolute error. [Colour figure can be viewed at wileyonlinelibrary.com]

Composite errors for the day-2 and day-4 forecasts exhibit relatively similar patterns, with the magnitudes generally becoming larger with increasing lead time (Figure 7g–l). For the wind speed, there is a weak bias in the jet core that is not statistically significant on day 2, with an extensive region of statistically significant positive wind speed bias on the cold side of the jet within the stratosphere (Figure 7g). For the day-4 forecast, a similar wind speed bias pattern is found, with a positive wind bias on the stratospheric side of the jet, and a more than 4 $m \cdot s^{-1}$ statistically significant slow bias in the jet core (Figure 7j). In terms of potential temperature, cold biases are present above the jet on days 2 and 4 (Figure 7h,k), with the model 1.5 K colder than observations above the jet near 200 hPa. The most striking result is the PV biases, which indicate an increasing negative bias on the stratospheric side of the jet, particularly above the jet (Figure 7i,l). This negative bias region increases from −0.7 PVU in the day-2 forecast to −0.9 PVU in the day-4 forecast and is coincident with the region of negative bias in the background forecast (and analysis) and the highest composite horizontal PV gradient in the cross-section. Furthermore, the day-4 forecast

wind biases (Figure 7j) appear to be qualitatively consistent with the anticyclonic winds that would be obtained by inverting a negative PV anomaly colocated with the negative PV bias. Consequently, the model's PV gradient across the jet stream and tropopause appear to be too weak, or the model is having a difficult time maintaining the sharpness that is present in observations, analyses, and short-range forecasts.

In addition to the biases, the mean absolute errors grow with lead time (cf. Figure 8g–l). Whereas the wind speed error does not show any appreciable relationship with the pattern of the observed winds, temperature errors are maximized in locations where the composite horizontal temperature gradient is the largest (i.e., above the jet around 200 hPa, and on the cold side of the jet between 400 and 600 hPa). Finally, the PV errors are maximized along the tropopause both above and to the cold side of the jet where the horizontal PV gradient is the largest, typically following the region of large PV bias and near 200 hPa, which suggests that the model has increasing errors (with lead time) in the structure of the PV gradients along the tropopause.

While the composite statistics suggest that the IFS has an increasingly difficult time replicating the structure of the jet, it is possible that there are some occasions where the model has especially small or large errors. The relatively small number of transects make it difficult to parse these results based on the structure of the jet, such as with cluster analysis or a self-organizing map. Instead, the individual transects are classified by a number of bulk properties, such as the strength of the jet and the magnitude of the PV gradient.

Before investigating how the bias relates to the structure of the jet, the hypothesis that the magnitude of the jet bias is tied to the PV bias is tested by assessing the mean jet bias versus the mean PV bias in each transect. Given that each transect is not equally distributed around the jet (i.e., there are more profiles on the warm side of the jet, rather than the cold side) and to focus on the jet itself, the wind speed and PV bias for each case is computed only for points within 200 km of the horizontal position of the jet maximum and between 200 and 600 hPa. Figure 9a shows the relationship between the domain-average wind speed and PV biases for each of the transects for the day-4 forecasts. For the background forecast, the bias averaged over each cross-section is generally small (<1 $m \cdot s^{-1}$ and 0.1 PVU for wind speed and PV respectively) and there appears to be no relationship between the domain-average biases (not shown). By contrast, the day-4 forecast biases for individual cases are much larger in magnitude and there is a relationship between the size of the biases (Pearson correlation

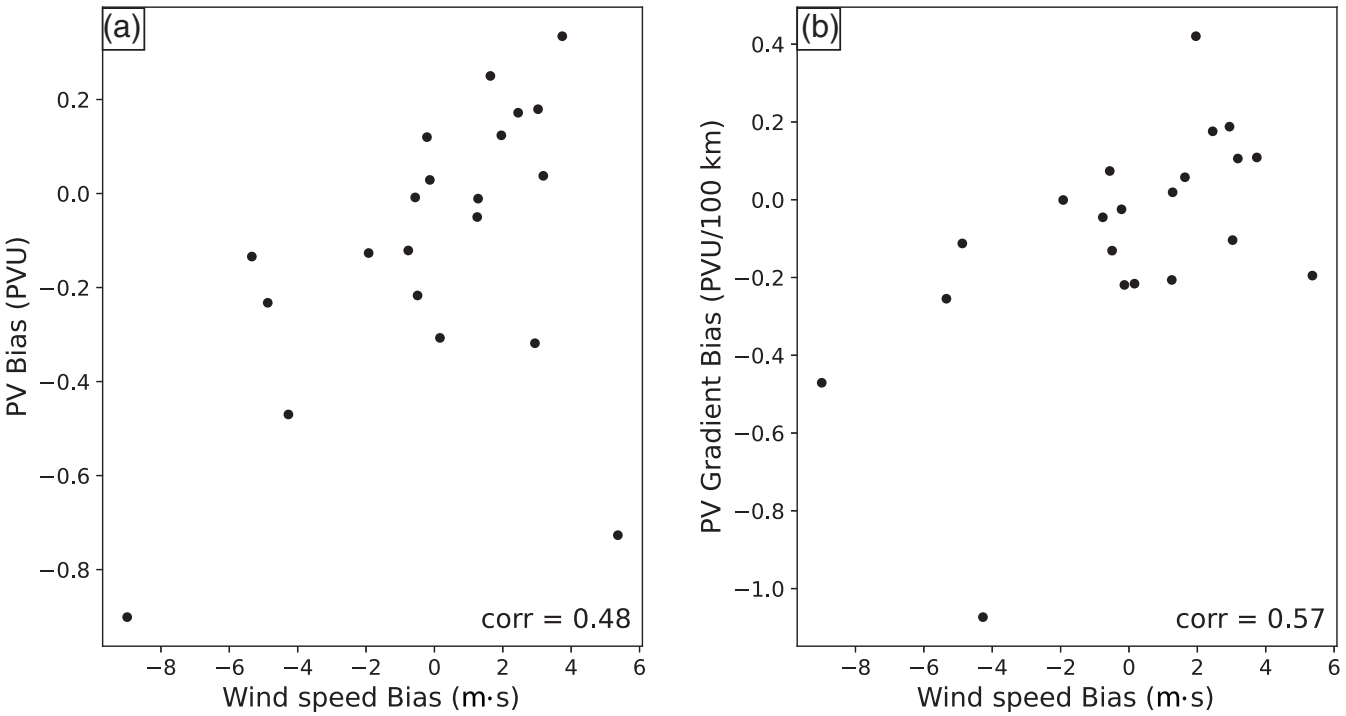

FIGURE 9 Day-4 forecast wind speed bias versus the (a) potential vorticity (PV) bias and (b) bulk PV gradient bias averaged within 200 km of the jet core for each of the dropwindsonde jet transects. The number in the bottom right represents the Pearson correlation coefficient.

coefficient of 0.48), with weak wind cases associated with a negative PV bias, suggesting that the two biases are linked (Figure 9a). The linkage appears to be especially strong for the 9 March 2022 case, which has the largest magnitude wind speed ($-8.8\,m{\cdot}s^{-1}$) and PV bias (−0.9 PVU). For this case, the model jet maximum is above 200 hPa, while the observed jet is at 300 hPa; therefore, the model PV gradient is substantially weaker. Conversely, the 12 March 2021 transect has a − 0.7 PVU PV bias, but a $5.4\,m{\cdot}s^{-1}$ positive wind speed bias. Here, the model's maximum wind speed is weaker than the observed, but the model has a much larger area of higher than $40\,m{\cdot}s^{-1}$ winds relative to the observed transect, which yields a positive bias. The weak jet maximum is consistent with the negative PV bias on the cold side of the jet (not shown). The jet stream itself makes up a fraction of the cross-section; therefore, it is possible that the domain-average wind speed bias does not reflect a bias in the jet stream winds. This possibility is assessed by comparing the domain-average PV bias against the difference in the modelled and observed wind speed percentiles within the pressure and distance criteria given above. This approach helps to isolate the wind speeds associated with the jet, while allowing for different speeds. While the 70th and 80th percentile wind speeds show a higher correlation with the PV bias (Pearson correlation coefficients of 0.61 and 0.62 respectively) relative to the domain-average wind speed bias, the 90th percentile and maximum wind speed biases exhibit slightly lower correlations with the PV bias and the scatterplots are qualitatively similar to Figure 9a. Thus, this result suggests that the relationship between wind speed biases and PV biases is fairly insensitive to the wind speed definition.

Given the large horizontal and vertical gradients in the wind speed and PV, it might be expected that the model may have more difficulty maintaining these gradients for cases with larger observed PV gradients or wind speeds may have larger biases. In order to evaluate that hypothesis, a bulk PV gradient is computed from the dropwindsonde observations for each transect and compared to the mean PV and wind speed bias, calculated using the method described above. There are numerous ways of defining the PV gradient in this context. One method would be to calculate the gradient in the potential temperature along the 2 PVU surface (i.e., the dynamical tropopause). Unfortunately, the 2 PVU surface is above 200 hPa on the warm side of the jet for many of the transects and not all of the dropwindsondes provide reliable data above 200 hPa; therefore, this definition could not be utilized here. Furthermore, it is possible to define the gradient on isentropic surfaces; however, the potential temperature of the wind speed maximum spans 313–356 K, so, it would be difficult to define consistent isentropic surfaces over all transects. Here, the bulk PV gradient is defined as the difference in the 200–300 hPa layer-average PV at +/− 200 km of the wind speed maximum. This layer encompasses the depression of the tropopause on the cold side of the jet for the majority of cases. While this layer is unlikely to be optimal for all cases, it does a fairly good job representing the bulk difference in the tropopause level between the warm and cold side of the jet. The sensitivity of the results to the definition of the PV gradient was tested by calculating the 200–300 hPa PV gradient at +/− 100 km of the wind speed maximum and the PV +/− 40 hPa of the wind speed maximum at +/− 200 km of the jet centre. The results are qualitatively similar and do not substantively change the results presented below (not shown).

Similar to the wind speed–PV biases described above, a relationship between the observed PV gradient and PV biases is minimal in the background forecast (not shown), but is more prominent for the day-4 forecast (Figure 10a). In general, the PV bias becomes more negative with higher horizontal PV gradients (Pearson correlation coefficient of −0.52), suggesting that the model has difficulty maintaining PV as the PV gradient becomes larger. For wind speed (Figure 10b), the day-4 forecast bias has a relatively weak relationship with the PV gradient itself, suggesting that large PV gradients are not necessarily associated with higher wind speed biases.

Finally, theory would suggest a connection between the model's representation of the PV gradient and the observed PV gradient as well as the model's PV gradient and the wind speed bias; therefore, the potential relationship is explored. For this calculation, a bulk PV gradient bias is calculated in the same manner as above for both the model and dropwindsonde data. Whereas the background forecast shows no clear relationship between the bulk PV gradient bias and the bulk PV gradient or wind speed bias (not shown), the model's bulk PV gradient bias exhibits a tendency towards negative values (i.e., the model's PV gradient is weaker than what is observed from dropwindsondes) as the observed PV gradient increases (Pearson correlation coefficient: −0.63; Figure 10c), though this is heavily weighted by the aforementioned 9 March 2022 transect. It is worth pointing out that the bulk PV gradient bias for the day-4 forecast ($-0.1\,PVU100\,km^{-1}$) is an order of magnitude larger than the background forecast's bulk PV gradient bias and is equivalent to 12% of the observed bulk PV gradient averaged over all cases. Furthermore, the cases with a negative day-4 forecast PV gradient bias are characterized by a negative wind speed bias (Pearson correlation coefficient: 0.57; Figure 9b), further suggesting a relationship between the wind speed biases in the vicinity of the jet and PV biases.

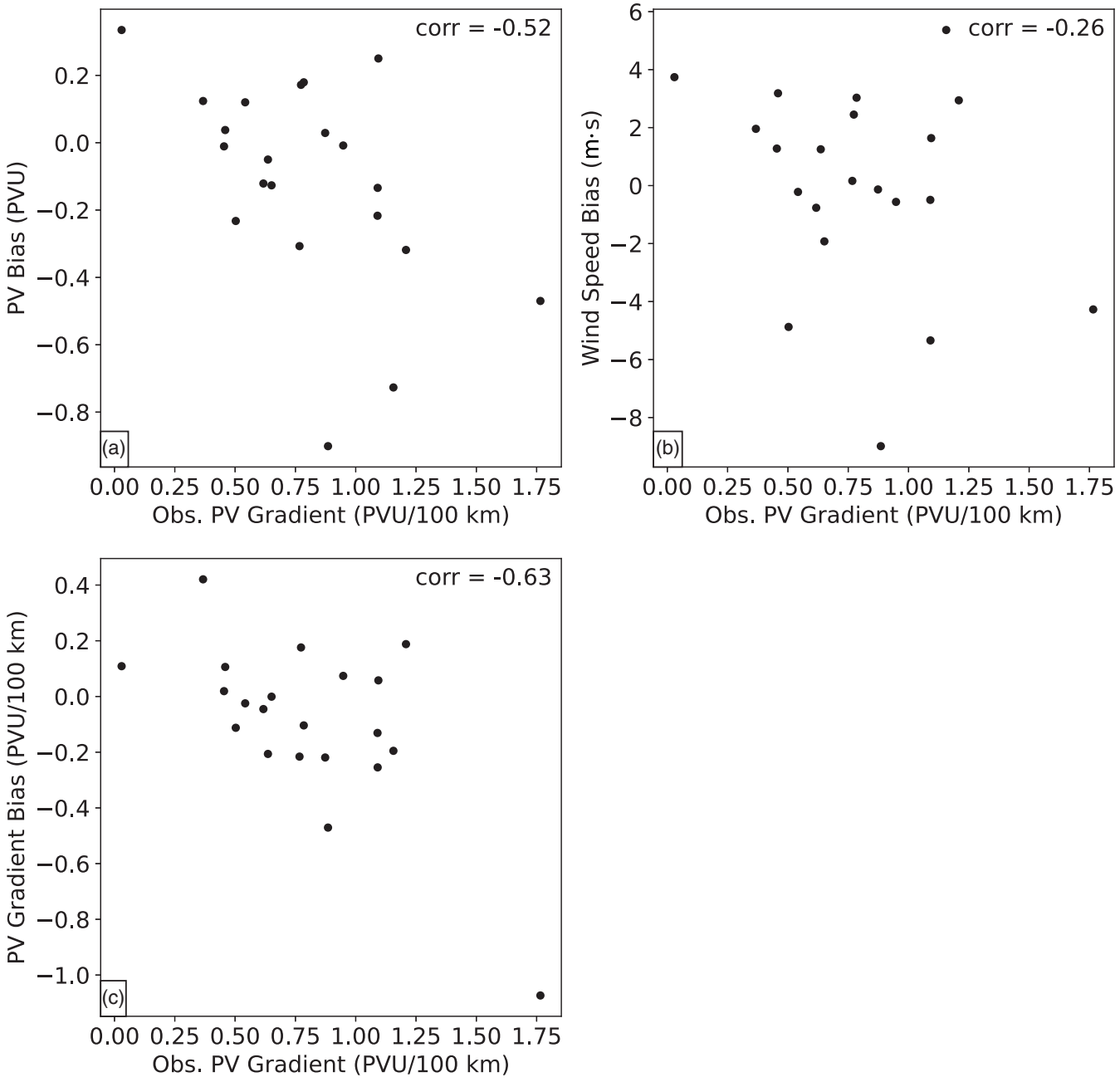

FIGURE 10 The observed horizontal potential vorticity (PV) gradient across the jet versus the biases of day-4 forecast (a) PV, (b) wind speed, and (c) bulk PV gradient averaged within 200 km of the jet core for each of the dropwindsonde jet transects. The number in the top right represents the Pearson correlation coefficient.

4 | SUMMARY AND CONCLUSIONS

Extending the useful predictability range of NWP models in the midlatitudes likely requires accurate predictions of the jet stream and nearby tropopause structure. To that end, this study utilized a unique set of dropwindsonde observation transects collected across the northern Pacific Ocean during three years of AR Recon campaigns in 2020, 2021, and 2022 and compared them to up to day-4 forecasts from the ECMWF IFS. In addition to validating the forecasts against wind and temperature data, this study estimated PV from the dropwindsonde transects and compared it to the model's estimate. These dropwindsondes captured a variety of jet streak and PV structures and hence provided a good dataset for assessing model performance.

IFS forecasts are characterized by slow-wind biases near the jet that increase in magnitude with lead time. For all dropwindsondes from the three years, observations above 50 $m \cdot s^{-1}$ exhibit a greater probability of slow-wind bias than lower wind speeds, with the bias increasing with

time. For the 21 dropwindsonde transects, there was a weak bias within the composite jet core that increased in magnitude from the background to day-4 forecast. In addition, there was a low PV bias on the cold side and above the jet concentrated in the zone of the largest horizontal PV gradient. It is potentially intriguing that the location of the large negative PV bias in the day-4 forecast corresponds to the location where the $\frac{\partial v}{\partial x}$ term is non-trivial in multiple cases. While the model and observed PV are computed with the same method, it is possible that the model may have more vorticity in the $\frac{\partial v}{\partial x}$ term relative to the $\frac{\partial u}{\partial y}$ at longer lead times; however, this cannot be evaluated from the current observation set. Furthermore, cases with a larger horizontal PV gradient tended to have a larger PV and PV gradient bias in the day-4 forecast, while such a pattern did not exist in the background forecast. These results suggest that the IFS has difficulty resolving the sharp PV gradient across the jet stream and tropopause; and this PV gradient further weakens with lead time. The smaller bias in the analysis and background forecast suggest that data assimilation appears to correct some of the model errors, while some combination of model physics and numerics likely reduce the gradient.

These results agree with the observation-based assessment by Schäfler *et al.* (2020) who found winds that were too weak and wind gradients and PV that were too small, relative to lidar data in the Atlantic basin, and to the model-based validation by Gray *et al.* (2014); therefore, this appears to be a characteristic result for the IFS. This error may lead to model issues in handling the interaction between the large-scale atmospheric flow and the development of extratropical cyclones. Given that the PV gradient around the tropopause acts as a waveguide for Rossby wave activity (Martius *et al.*, 2010), this suggests that this model problem could be a source of medium-range forecast errors in the IFS, though this would need to be investigated further. This could be particularly important for midlatitude high-impact weather events, such as ARs and cyclones, which have been found to be highly sensitive to errors in the position of upper tropospheric troughs (e.g., Lamberson *et al.*, 2016; Reynolds *et al.*, 2019).

There are numerous aspects of the current IFS configuration which might be responsible for the PV errors across the tropopause, but these are difficult to discern from the current data. The first possible issue is the vertical resolution of the model levels in the upper troposphere and lower stratosphere (UTLS). With the current IFS setup of 137 model levels, this results in a resolution of about 300 m between 200 hPa and 500 hPa, which may not be sufficient to maintain the sharp wind and temperature gradients that characterize the tropopause region. Numerical diffusion within the IFS may result in the incorrect representation of vortex stripping, which is the process by which sharp PV gradients are generated from an initially smooth PV distribution (Haynes *et al.*, 2001). This stirring acts to remove the intermediate values of PV near the tropopause and dissipates it via small-scale mixing and dissipation, yielding a steep tropopause. Another possibility is overaggressive vertical mixing in the IFS model, which can result in an incorrect representation of the atmospheric energy spectra (e.g., Skamarock *et al.*, 2019). The numerical diffusion meant to help maintain model stability may disrupt this process, particularly in regions of large gradients where this method is more active. Consequently, adding more vertical levels in the vicinity of the tropopause may be beneficial for maintaining this gradient. It is also possible that issues with the model's advection or parameterizations could be contributing to these errors (Saffin *et al.*, 2017).

Along the tropopause, radiative cooling helps to maintain its steepness (Forster and Wirth, 2000; Randel et al. 2007). This is particularly enhanced by the presence of water vapor, whereby more water vapor yields a faster cooling rate (Forster and Wirth, 2000; Ferreira *et al.*, 2016), with cloudtop cooling increasing this even further (Cau *et al.*, 2005). In general, the IFS is known to have a wet bias in the lower stratosphere (Dyroff *et al.*, 2015; Bland *et al.*, 2021), so this problem may be contributing to the cold bias found at about 200 hPa, and hence PV errors, in the transects assessed. Furthermore, as most of the dropwindsondes in this study are near ARs, there are typically relatively higher water vapour concentrations and extensive clouds in the areas sampled, so this model wet bias may be more prominent in these cases.

Large-scale regions of latent heat release can also contribute to the steepening of the tropopause. One place where this could be particularly acute is in the vicinity of WCBs, which are streams of air that transport moist low-PV air from near the PBL to the upper troposphere and hence can raise the tropopause (e.g., Wernli and Davies, 1997; Riemer and Jones, 2010; Grams *et al.*, 2011). Chagnon *et al.* (2013) and Chagnon and Gray (2015) found that a combination of radiative cooling and WCB-related latent heating can produce a dipole of diabatically increased (decreased) PV above (below) the tropopause in troughs related to extratropical cyclones. They emphasized that this near-tropopause effect of diabatic processes may affect Rossby wave propagation. Multiple jet transects were adjacent to ARs that also met the criteria for WCBs (e.g., 15 February 2020, 11–13 March 2021, 24 February 2022), so it may be particularly important to replicate these processes correctly for these cases.

Finally, the region around the jet stream and upper-level fronts is characterized by clear-air turbulence and diffusive mixing (Jaeger and Sprenger, 2007), where these processes can produce PV anomalies and alter the tropopause structure (e.g., Staley, 1960; Shapiro, 1976; Spreitzer *et al.*, 2019). In an NWP model, this process is handled by the turbulent mixing scheme, which in the IFS was developed for use in the PBL. As this mixing scheme is activated in regions of wind shear, this consequently means that this scheme may be overactive in the high-wind-shear UTLS region. In turn, this may be contributing to the weakening or diffusing away of the sharp PV gradients that were found there (e.g., Skamarock *et al.*, 2019) and may be causing a lower-stratospheric moist bias in the IFS (Krüger *et al.*, 2022).

To investigate the issues identified by this work, a variety of modelling experiments and diagnostic studies are being planned to address the source. One possible experiment is to increase the vertical resolution in the UTLS to determine if this will provide a more skillful simulation of the UTLS and jet stream in the IOPs studied herein. Additional experiments could adjust the turbulent mixing coefficients in UTLS region, which may help maintain the gradient. Furthermore, it may be worthwhile to evaluate the water vapour biases on both sides of the jet for these cases and assess how these biases might impact the resulting radiative cooling rate. Finally, future AR Recon missions will continue to provide additional transects of the jet stream, which will yield a larger and more diverse verification dataset.

AUTHOR CONTRIBUTIONS

David Lavers: Conceptualization; formal analysis; investigation; methodology; visualization; writing – original draft; writing – review and editing. **Ryan Torn:** Conceptualization; formal analysis; investigation; methodology; visualization; writing – original draft; writing – review and editing. **Chris Davis:** Conceptualization; writing – review and editing. **David Richardson:** Conceptualization; writing – review and editing. **F. Martin Ralph:** Conceptualization; writing – review and editing. **Florian Pappenberger:** Conceptualization; writing – review and editing.

ACKNOWLEDGEMENTS

The lead author was supported by the Copernicus Climate Change Service, which is implemented by ECMWF on behalf of the European Union. We thank Inna Polichtchouk for discussions on this research and we are deeply grateful to the NOAA flight crews for undertaking the missions to provide these dropwindsonde observations. We also thank the two anonymous reviewers whose comments helped to improve this paper.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data used are available through the ECMWF archive (https://www.ecmwf.int/en/forecasts/access-forecasts/access-archive-datasets).

ORCID

David A. Lavers https://orcid.org/0000-0002-7947-3737
Ryan D. Torn https://orcid.org/0000-0002-4067-6684
Florian Pappenberger https://orcid.org/0000-0003-1766-2898

REFERENCES

Baker, W.E., Atlas, R., Cardinali, C., Clement, A., Emmitt, G.D., Gentry, B.M., Hardesty, R.M., Källén, E., Kavaya, M.J., Langland, R. and Ma, Z. (2014) Lidar-measured wind profiles: the missing link in the global observing system. *Bulletin of the American Meteorological Society*, 95, 543–564. https://doi.org/10.1175/BAMS-D12-00164.1.

Bland, J., Gray, S., Methven, J. and Forbes, R. (2021) Characterising extratropical near-tropopause analysis humidity biases and their radiative effects on temperature forecasts. *Quarterly Journal of the Royal Meteorological Society*, 147(741), 3878–3898. https://doi.org/10.1002/qj.4150.

Browning, K.A. (1986) Conceptual models of precipitation systems. *Wea. Forecasting*, 1, 23–41.

Browning, K.A. (2004) The sting at the end of the tail: damaging winds associated with extratropical cyclones. *Quarterly Journal of the Royal Meteorological Society*, 130, 375–399.

Cardinali, C., Rukhovets, L. and Tenenbaum, J. (2004) Jet stream analysis and forecast errors using GADS aircraft observations in the DAO, ECMWF, and NCEP models. *Monthly Weather Review*, 132, 764–779. https://doi.org/10.1175/1520-0493(2004)132,0764:JSAAFE.2.0.CO;2.

Cau, P., Methven, J. and Hoskins, B. (2005) Representation of dry tropical layers and their origins in ERA-40 data. *Journal of Geophysical Research*, 110, D06110. https://doi.org/10.1029/2004JD004928.

Chagnon, J.M. and Gray, S.L. (2015) A diabatically generated potential vorticity structure near the extratropical tropopause in three simulated extratropical cyclones. *Monthly Weather Review*, 143(6), 2337–2347. https://doi.org/10.1175/MWR-D-14-00092.1.

Chagnon, J.M., Gray, S.L. and Methven, J. (2013) Diabatic processes modifying potential vorticity in a North Atlantic cyclone. *Quarterly Journal of the Royal Meteorological Society*, 139(674), 1270–1282. https://doi.org/10.1002/qj.2037.

Cobb, A., Ralph, F.M., Tallapragada, V., Wilson, A.M., Davis, C.A., Monache, L.D., Doyle, J.D., Pappenberger, F., Reynolds, C.A., Subramanian, A., Black, P.G., Cannon, F., Castellano, C., Cordeira, J.M., Haase, J.S., Hecht, C., Kawzenuk, B., Lavers, D.A., Murphy, M.J., Jr., Parrish, J., Rickert, R., Rutz, J.J., Torn, R., Wu, X. and Zheng, M. (2022) Atmospheric River Reconnaissance 2021: A Review, *Weather and Forecasting* (published online ahead of print 2022). Retrieved Aug 4, 2022, from. https://journals.ametsoc.org/view/journals/wefo/aop/WAF-D-21-0164.1/WAF-D-21-0164.1.xml.

Cordeira, J.M. and Bosart, L.F. (2010) The antecedent large-scale conditions of the "perfect storms" of late October and early November 1991. *Monthly Weather Review*, 138(7), 2546–2569. https://doi.org/10.1175/2010MWR3280.1.

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

Danielsen, E.F., Hipskind, R.S., Gaines, S.E., Sachse, G.W., Gregory, G.L. and Hill, G.F. (1987) Three-dimensional analysis of potential vorticity associated with tropopause folds and observed variations of ozone and carbon monoxide. *Journal of Geophysical Research: Atmospheres*, 92, 2103–2111.

Danielsen, E.F. and Mohnen, V.A. (1977) Project dustorm report: ozone transport, in situ measurements, and meteorological analyses of tropopause folding. *Journal of Geophysical Research*, 82, 5867–5877.

Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2005) Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131, 3385–3396.

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

Dyroff, C., Zahn, A., Christner, E., Forbes, R., Tompkins, A.M. and van Velthoven, P.F.J. (2015) Comparison of ECMWF analysis and forecast humidity data with CARIBIC upper troposphere and lower stratosphere observations. *Quarterly Journal of the Royal Meteorological Society*, 141, 833–844. https://doi.org/10.1002/qj.2400.

Ferreira, A.P., Castanheira, J.M. and Gimeno, L. (2016) Water vapour stratification and dynamical warming behind the sharpness of the Earth's midlatitude tropopause. *Quarterly Journal of the Royal Meteorological Society*, 142(695), 957–970. https://doi.org/10.1002/qj.2697.

Forster, C. and Wirth, V. (2000) Radiative decay of idealized stratospheric filaments in the troposphere. *Journal of Geophysical Research*, 105(D8), 10169–10184. https://doi.org/10.1029/2000JD900052.

Grams, C.M., Wernli, H., Bottcher, M., Campa, J., Corsmeier, U., Jones, S.C., Keller, J.H., Lenz, C.J. and Wiegand, L. (2011) The key role of diabatic processes in modifying the upper-tropospheric wave guide: a North Atlantic case-study. *Quarterly Journal of the Royal Meteorological Society*, 137(661), 2174–2193. https://doi.org/10.1002/qj.891.

Gray, S.L., Dunning, C.M., Methven, J., Masato, G. and Chagnon, J.M. (2014) Systematic model forecast error in Rossby wave structure. *Geophysical Research Letters*, 41, 2979–2987. https://doi.org/10.1002/2014GL059282.

Harvey, B., Methven, J. and Ambaum, M. (2016) Rossby wave propagation on potential vorticity fronts with finite width. *Journal of Fluid Mechanics*, 794, 775–797. https://doi.org/10.1017/jfm.2016.180.

Harvey, B., Methven, J. and Ambaum, M.H.P. (2018) An adiabatic mechanism for the reduction of jet meander amplitude by potential vorticity filamentation. *Journal of the Atmospheric Sciences*, 75, 4091–4106. https://doi.org/10.1175/JAS-D-18-0136.1.

Harvey, B., Methven, J., Sanchez, C. and Schäfler, A. (2020) Diabatic generation of negative potential vorticity and its impact on the North Atlantic jet stream. *Quarterly Journal of the Royal Meteorological Society*, 146, 1477–1497. https://doi.org/10.1002/QJ.3747.

Haynes, P., Scinocca, J. and Greenslade, M. (2001) Formation and maintenance of the extratropical tropopause by baroclinic eddies. *Geophysical Research Letters*, 28(22), 4179–4182. https://doi.org/10.1029/2001GL013485.

Hoskins, B.J. and Ambrizzi, T. (1993) Rossby wave propagation on a realistic longitudinally varying flow. *Journal of the Atmospheric Sciences*, 50, 1661–1671. https://doi.org/10.1175/1520-0469(1993)050,1661:RWPOAR.2.0.CO;2.

Ingleby, B. (2017) An assessment of different radiosonde types 2015/2016. ECMWF *Technical* Memorandum No. 807. https://www.ecmwf.int/en/elibrary/17551-assessment-different-radiosonde-types-2015-2016.

Jaeger, E.B. and Sprenger, M. (2007) A northern hemispheric climatology of indices for clear air turbulence in the tropopause region derived from ERA40 reanalysis data. *Journal of Geophysical Research*, 112, D20106. https://doi.org/10.1029/2006JD008189.

Koch, S.E., Jamison, B.D., Lu, C., Smith, T.L., Tollerud, E.I., Girz, C., Wang, N., Lane, T.P., Shapiro, M.A., Parrish, D.D. and Cooper, O.R. (2005) Turbulence and gravity waves within an upper-level front. *Journal of the Atmospheric Sciences*, 62(11), 3885–3908. https://doi.org/10.1175/JAS3574.1.

Krüger, K., Schäfler, A., Wirth, M., Weissmann, M. and Craig, G.C. (2022) Vertical structure of the lower-stratospheric moist bias in the ERA5 reanalysis and its connection to mixing processes. *Atmospheric Chemistry and Physics*, 22, 15559–15577. https://doi.org/10.5194/acp-22-15559-2022.

Lamberson, W.S., Torn, R.D., Bosart, L.F. and Magnusson, L. (2016) Diagnosis of the source and evolution of medium-range forecast errors for extratropical cyclone Joachim. *Weather and Forecasting*, 31, 1197–1214. https://doi.org/10.1175/WAF-D-16-0026.1.

Lavers, D.A., Allan, R.P., Wood, E.F., Villarini, G., Brayshaw, D.J. and Wade, A.J. (2011) Winter floods in Britain are connected to atmospheric rivers. *Geophysical Research Letters*, 38, L23803. https://doi.org/10.1029/2011GL049783.

Lavers, D.A., Ingleby, N.B., Subramanian, A.C., Richardson, D.S., Ralph, F.M., Doyle, J.D., Reynolds, C.A., Torn, R.D., Rodwell, M.J., Tallapragada, V. and Pappenberger, F. (2020) Forecast errors and uncertainties in atmospheric Rivers. *Weather and Forecasting*, 35, 1447–1458. https://doi.org/10.1175/WAF-D-20-0049.1.

Lean, P., Bonavita, M., Hólm, E., Bormann, N. and McNally, T. (2019) Continuous data assimilation for the IFS. *ECMWF Newsletter*, 158, 21–26. https://www.ecmwf.int/en/newsletter/158/meteorology/continuous-data-assimilation-ifs.

Linkin, M.E. and Nigam, S. (2008) The North Pacific oscillation–West Pacific teleconnection pattern: mature-phase structure and winter impacts. *Journal of Climate*, 21(9), 1979–1997.

Martius, O., Schwierz, C. and Davies, H. (2010) Tropopause-level waveguide. *Journal of the Atmospheric Sciences*, 67, 866–879. https://doi.org/10.1175/2009JAS2995.1.

Neiman, P.J., Schick, L.J., Ralph, F.M., Hughes, M. and Wick, G.A. (2011) Flooding in western Washington: the connection to

atmospheric rivers. *Journal of Hydrometeorology*, 12, 1337–1358. https://doi.org/10.1175/2011JHM1358.1.

Nguyen, L.T., Molinari, J. and Thomas, D. (2014) Evaluation of tropical cyclone center identification methods in numerical models. *Monthly Weather Review*, 142, 4326–4339. https://doi.org/10.1175/MWR-D-14-00044.1.

Ralph, F.M., Cannon, F., Tallapragada, V., Davis, C.A., Doyle, J.D., Pappenberger, F., Subramanian, A., Wilson, A.M., Lavers, D.A., Reynolds, C.A., Haase, J.S., Centurioni, L., Ingleby, B., Rutz, J.J., Cordeira, J.M., Zheng, M., Hecht, C., Kawzenuk, B. and Delle Monache, L. (2020) West coast forecast challenges and development of Atmospheric River reconnaissance. *Bulletin of the American Meteorological Society*, 101, E1357–E1377. https://doi.org/10.1175/BAMS-D-19-0183.1.

Ralph, F.M., Dettinger, M.D., Cairns, M.M., Galarneau, T.J. and Eylander, J. (2018) Defining "atmospheric river": how the glossary of meteorology helped resolve a debate. *Bulletin of the American Meteorological Society*, 99(4), 837–839. https://doi.org/10.1175/BAMS-D-17-0157.1.

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

Randel, W.J., Wu, F. and Forster, P. (2007). The extratropical tropopause inversion layer: global observations with GPS data, and a radiative forcing mechanism. *Journal of Atmospheric Science*, 64(12), 4489–4496. https://doi.org/10.1175/2007JAS2412.1

Reynolds, C.A., Doyle, J.D., Ralph, F.M. and Demirdjian, R. (2019) Adjoint sensitivity of North Pacific Atmospheric River forecasts. *Monthly Weather Review*, 147, 1871–1897.

Rickard, G.J., Lunnon, R.W. and Tenenbaum, J. (2001) The met Office upper air winds: prediction and verification in the context of commercial aviation data. *Meteorological Applications*, 8, 351–360. https://doi.org/10.1017/S1350482701003115.

Riemer, M. and Jones, S.C. (2010) The downstream impact of tropical cyclones on a developing baroclinic wave in idealized scenarios of extratropical transition. *Quarterly Journal of the Royal Meteorological Society*, 136(648), 617–637. https://doi.org/10.1002/qj.605.

Rossby, C.-G. (1945) On the propagation of frequencies and energy in certain types of oceanic and atmospheric waves. *Journal of the Atmospheric Sciences*, 2, 187–204. https://doi.org/10.1175/1520-0469(1945)002,0187:OTPOFA.2.0.CO;2.

Saffin, L., Gray, S., Methven, J. and Williams, K. (2017) Processes maintaining tropopause sharpness in numerical models. *Journal of Geophysical Research – Atmospheres*, 122, 9611–9627. https://doi.org/10.1002/2017JD026879.

Schäfler, A., Craig, G., Wernli, H., Arbogast, P., Doyle, J.D., McTaggart-Cowan, R., Methven, J., Rivière, G., Ament, F., Boettcher, M. and Bramberger, M. (2018) The North Atlantic waveguide and downstream impact experiment. *Bulletin of the American Meteorological Society*, 99, 1607–1637. https://doi.org/10.1175/BAMS-D-17-0003.1.

Schäfler, A., Harvey, B., Methven, J., Doyle, J.D., Rahm, S., Reitebuch, O., Weiler, F. and Witschas, B. (2020) Observation of jet stream winds during NAWDEX and characterization of systematic meteorological analysis errors. *Monthly Weather Review*, 148, 2889–2907. https://doi.org/10.1175/MWR-D-19-0229.1.

Schwierz, C., Dirren, S. and Davies, H. (2004) Forced waves on a zonally aligned jet stream. *Journal of the Atmospheric Sciences*, 61, 73–87. https://doi.org/10.1175/1520-0469(2004)061,0073:FWOAZA.2.0.CO;2.

Shapiro, M.A. (1974) A multiple structured frontal zone-jet stream system as revealed by meteorologically instrumented aircraft. *Monthly Weather Review*, 102, 244–253.

Shapiro, M.A. (1976) The role of turbulent heat flux in the generation of potential vorticity in the vicinity of upper-level jet stream systems. *Monthly Weather Review*, 104, 892–906.

Shapiro, M.A., Hampel, T. and Krueger, A.J. (1987) The Arctic tropopause fold. *Monthly Weather Review*, 115, 444–454.

Skamarock, W.C., Snyder, C., Klemp, J.B. and Park, S. (2019) Vertical resolution requirements in atmospheric simulation. *Monthly Weather Review*, 147, 2641–2656 https://journals.ametsoc.org/view/journals/mwre/147/7/mwr-d-19-0043.1.xml.

Spreitzer, E., Attinger, R., Boettcher, M., Forbes, R., Wernli, H. and Joos, H. (2019) Modification of potential vorticity near the tropopause by nonconservative processes in the ECMWF model. *Journal of the Atmospheric Sciences*, 76(6), 1709–1726. https://doi.org/10.1175/JAS-D-18-0295.1.

Staley, D.O. (1960) Evaluation of potential-vorticity changes near the tropopause and the related vertical motions, vertical advection of vorticity, and transfer of radioactive debris from stratosphere to troposphere. *Journal of the Atmospheric Sciences*, 17, 591–620. https://doi.org/10.1175/1520-0469(1960)017<0591:EOPVCN>2.0.CO;2.

Vaisala. (2018) Vaisala dropwindsonde RD41. Vaisala Report 2pp. [https://www.vaisala.com/sites/default/files/documents/RD41-Datasheet-B211706EN.pdf.].

Wernli, H. and Davies, H.C. (1997) A lagrangian-based analysis of extratropical cyclones. I: the method and some applications. *Quarterly Journal of the Royal Meteorological Society*, 123(538), 467–489. https://doi.org/10.1002/qj.49712353811.

Wirth, V., Riemer, M., Chang, E.K.M. and Martius, O. (2018) Rossby wave packets on the midlatitude waveguide—a review. *Monthly Weather Review*, 146(7), 1965–2001. https://doi.org/10.1175/MWR-D-16-0483.1.

How to cite this article: Lavers, D.A., Torn, R.D., Davis, C., Richardson, D.S., Ralph, F.M. & Pappenberger, F. (2023) Forecast evaluation of the North Pacific jet stream using AR Recon dropwindsondes. *Quarterly Journal of the Royal Meteorological Society*, 1–20. Available from: https://doi.org/10.1002/qj.4545