Observation of jet stream winds during NAWDEX and characterization of systematic meteorological analysis errors

Andreas Schäfler
Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Ben Harvey
National Centre for Atmospheric Science, University of Reading, Reading, United Kingdom

John Methven
Department of Meteorology, University of Reading, Reading, United Kingdom

James D. Doyle
Naval Research Laboratory, Monterey, California

Stephan Rahm
Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Oliver Reitebuch
Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Fabian Weiler
Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Benjamin Witschas
Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Contact:
Dr. Andreas Schäfler
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)
Institut für Physik der Atmosphäre
Oberpfaffenhofen, 82234 Weßling
andreas.schaefler@dlr.de

Early Online Release: This preliminary version has been accepted for publication in *Monthly Weather Review*, may be fully cited, and has been assigned DOI 10.1175/MWR-D-19-0229.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

Abstract

Observations across the North Atlantic jet stream with high vertical resolution are used to explore the structure of the jet stream, including the sharpness of vertical wind shear changes across the tropopause and the wind speed. Data was obtained during the North Atlantic Waveguide and Downstream impact EXperiment (NAWDEX) by an airborne Doppler wind lidar, dropsondes and a ground-based Stratosphere-Troposphere radar. During the campaign small wind speed biases throughout the troposphere and lower stratosphere of only -0.41 m s^{-1} and -0.15 m s^{-1} are found respectively in the ECMWF and UK Met Office analyses and short-term forecasts. However, this study finds large and spatially coherent wind errors up to ±10 m s^{-1} for individual cases, with the strongest errors occurring above the tropopause in upper-level ridges.

ECMWF and Met Office analyses indicate similar spatial structures in wind errors, even though their forecast models and data assimilation schemes differ greatly. The assimilation of operational observational data brings the analyses closer to the independent verifying observations but it cannot fully compensate the forecast error. Models tend to underestimate the peak jet stream wind, the vertical wind shear (by a factor of 2-5) and the abruptness of the change in wind shear across the tropopause, which is a major contribution to the meridional potential vorticity gradient. The differences are large enough to influence forecasts of Rossby wave disturbances to the jet stream with an anticipated effect on weather forecast skill even on large scales.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

1. Introduction

The existence and behavior of the North Atlantic jet stream is central to the weather experienced across Europe in all seasons. Weather systems having major impacts on surface conditions, such as mid-latitude cyclones, the fronts embedded within them and mesoscale convective systems, are all influenced strongly by interaction with the jet stream. Their structure and evolution is affected by the location of strong vertical wind shear, as well as wave and vortex disturbances at tropopause level that develop as the jet stream meanders and contorts. Meandering jet streams coincide with strong gradients of potential vorticity (PV) along the isentropic surfaces intersecting the tropopause. These gradients serve as a waveguide for propagating Rossby waves (Hoskins and Ambrizzi, 1993; Schwierz et al. 2004; Martius et al. 2010). Disturbances to the waveguide at the entrance (western) end of the storm track can have a major effect on surface weather thousands of kilometers downstream through the propagation of disturbance energy in the form of Rossby wave packets (see recent review by Wirth and Riemer, 2018). Therefore, a detailed representation of the jet stream structure is important not only locally in forecasting upper-tropospheric winds, but also has far-reaching consequences for predicting surface weather system development.

Accurate prediction of Rossby waves is sensitive to the representation of the jet stream structure and associated PV gradient, even though their wavelength exceeds the width of the strongest PV gradient regions by several orders of magnitude. This introduces a resolution dependence to jet stream prediction. It has been demonstrated that global numerical weather prediction (NWP) models fail to maintain sufficiently sharp PV gradients at the tropopause and Rossby wave amplitude decreases with lead time (Gray et al. 2014; Saffin et al. 2017). If the PV gradient is too smooth in a model then advection of disturbances by the jet stream and counter-

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

propagation of Rossby waves against the zonal flow are both expected to be too weak. Harvey et al. (2016) showed analytically that although these effects on Rossby wave phase speed cancel to first order, in more accurate estimates phase speed must always decrease (slower eastward). Harvey et al. (2018) used wave activity theory to show that when the PV gradient is too smooth in a model, then Rossby wave amplitude is also predicted to decay. The lead-time dependence of the PV gradient forecast error, both in horizontal gradient along an isentropic surface (Gray et al. 2014) and vertical gradient (Saffin et al. 2017), indicates that the NWP models struggle to represent the tropopause, an issue that is expected to be even more prominent in climate prediction models due to their lower spatial resolution. Davies and Didone (2013) showed how forecast errors of PV propagate and amplify along the jet stream waveguide and Baumgart et al. (2018) have quantified the extent to which different dynamical mechanisms contribute to the growth of PV forecast error from uncertainty in the initial conditions.

In this study we examine high resolution observations of the jet stream (detailed in Section 2) and compare them with the representation of jet-stream winds in meteorological analyses and short-term forecasts. It is an open question to what extent they are able to represent the observed wind speed distribution, especially the strength of the vertical wind shear on either side of the tropopause, which is of crucial importance for an accurate representation of the meridional PV gradient and Rossby wave evolution.

In the 1990s and early 2000s, several studies that used in situ observed winds onboard commercial airliners to validate NWP winds reported on significant wind speed biases in meteorological analyses (Tenenbaum 1991, 1996; Rickard et al. 2001; Cardinali et al. 2003). Multi-case averaging revealed wind speed biases increasing with observed wind speeds and reaching values of up to 5-10 % (Rickard et al. 2001). Cardinali et al. (2003) found that jet streak

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

winds are too weak by 2 to 5 % in data-dense regions over the US and by 5 to 9 % in data-sparse regions over Canada. The continuous increase of vertical and horizontal resolution in NWP models, the continuous increase in quality, amount and resolution of aircraft and satellite observations and their improved application has led to a substantially improved representation of winds in NWP analyses. As depicted by Petersen (2016), Northern Hemispheric wind errors decreased by about 40% for 24-h forecasts between 1984 and 2004. Houchi et al. (2010) compared winds in different climate regions using high vertical-resolution radiosondes from 85 stations and ECMWF short-term forecasts in the year 2006. They found qualitative agreement of observed and modelled wind distributions at all levels. However, they note a substantial underestimation of vertical wind shear and its variability associated with small scale vertical wind gradients that are not well represented by ECMWF short-term forecasts, particularly due to the limited vertical resolution of the model. Based on multi-month analysis differences between ECMWF and the National Centers for Environmental Prediction (NCEP), Baker et al. (2014) estimate an uncertainty of winds at 300 hPa in the order of 2-3 m s^{-1} over the northern North Atlantic. More recently, Belmonte Rivas and Stoffelen (2019) compared surface winds represented by ERA5 with Advanced Scatterometer (ASCAT) observations, and found systematic circulation errors, in the sense that surface winds are too cyclonic across ocean basins in the re-analysis and meridional winds are too weak in mid-latitudes. These surface wind errors were attributed to underestimation in directional wind turning (the Ekman spiral) across the boundary layer of the ECMWF model. Therefore, it can be anticipated that errors at tropopause level will not have the same characteristics as surface wind errors.

In this study we compare operational meteorological analyses and short-term forecasts of two global NWP centers, the ECMWF and the United Kingdom Met Office, with a unique set of

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

wind profile observations across the tropopause that was obtained during the North Atlantic Waveguide and Downstream impact EXperiment (NAWDEX). NAWDEX was conducted in autumn 2016 with the aim to examine the structure of the jet stream, the impact of diabatic processes on the jet stream disturbances and their influence on high-impact weather downstream (Schäfler et al. 2018). For the first time, an established Doppler wind lidar payload onboard the research aircraft DLR Falcon performed dedicated observations of the jet stream winds providing both high vertical and horizontal resolution, which is not available from other observational sources. Additionally, the wind lidar data set is supplemented by dropsonde and ground-based wind profiler observations to provide a wider coverage and to investigate the observational reliability of the wind lidar.

In Section 2 we provide an overview of the observation and model data and the methods applied to validate analyses and short-term forecasts of ECMWF and Met Office. In Section 3, a case study is presented with coordinated wind lidar and dropsonde observations of a jet stream near Iceland on 23 September 2016. Section 4 contains a statistical evaluation of the horizontal wind and vertical wind shear representation during the NAWDEX field phase based on the wind lidar data set and wind profiler observations. Discussion of the results and conclusions are given in Section 5. The implications of the findings are presented in Section 6.

2. Data and methods

a. Airborne observations: Doppler Wind Lidar and Dropsondes

During NAWDEX, wind observations onboard the DLR Falcon were obtained by two Doppler wind lidar systems; the ALADIN Airborne Demonstrator (A2D, Reitebuch et al. 2009; Lux et al. 2018, Marksteiner et al. 2018) and the 2-µm Doppler wind lidar system (Weissmann et al. 2005, Witschas et al. 2017). In this study we rely on observations of the horizontal wind

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

vector measured by the 2-µm Doppler wind lidar (in the following abbreviated as DWL). Additionally, we use wind observations measured by in situ sensors in the nose-boom of the aircraft and by dropsondes that were released during coordinated flights with the High Altitude and Long Range Research Aircraft (HALO; Schäfler et al. 2018).

The coherent and heterodyne detection DWL measures range resolved profiles of the horizontal wind vector beneath the aircraft through detection of frequency shifts between emitted and retrieved laser signals. The DWL uses a wavelength of 2022.54 nm in an atmospheric window with low absorption of water vapor enabling wind measurements up to the maximum flight altitude of ~12 km, depending on aerosol column beneath. The DWL transmits short laser pulses with a length of 400-500 ns, a repetition rate of 500 Hz and an energy of 1-2 mJ to the atmosphere beneath the aircraft. The signal is partly scattered back to the aircraft by aerosols and cloud particles where it is received by a telescope and analyzed for frequency shift Δf which is proportional to the wind speed v_{LOS} in the line of sight (LOS) according to $\Delta f = (2f_0 \cdot v_{LOS})/c$, where f_0 is the laser frequency, c is the speed of light and $\lambda_0 = c/f_0 = 2022.54$ nm is the laser wavelength. To be able to derive a horizontal wind vector from LOS measurements, the DWL uses a double-wedge scanner to measure LOS winds at different pointing directions. A conical step-and-stare scan pattern (Velocity Azimuth Display (VAD)-technique) around the vertical axes with an off-nadir angle of 20° provides 21 LOS observations per one scanner revolution. A mean wind vector in the measurement volume can be derived by combining these 21 LOS velocities at different viewing direction. A wind profile is derived every 42s, i.e. the time that is required for one complete scanner revolution with 21 LOS observations including an averaging of 1s per LOS position and the scanner movement. Wind vectors are derived at a vertical

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

resolution of 100 m. For a more detailed instrument description of the DWL and the algorithms for the wind retrieval the interested reader is referred to Witschas et al. (2017).

During NAWDEX, the DLR Falcon successfully observed approaching cyclones and evolving jet streams surrounding Iceland. Eight flights were performed with the DWL between 17 September and 9 October 2016 (see Fig. 1a and overview in Schäfler et al. 2018) corresponding to a total measurement time of 22:55 h and a total distance of ~17,000 km. In a total of 1922 measurement profiles between 0 km and 12 km altitude, 77541 horizontal wind measurements were obtained which corresponds to a total data availability of about 33.8 % resulting from low concentration of the required aerosol or cloud scatterers in the frequently sampled clean and dry tropospheric and lower stratospheric air at high latitudes. However, the NAWDEX data set provides a maximum in data availability where the average wind shows a maximum, between 8 km and 10 km altitude (Fig. 1b). The maximum data availability of 80 % at 9.4 km altitude corresponds to ~18:20 h of observations and a flight distance of 13,500 km. The mean profile separation, i.e. the horizontal resolution, which depends on the speed of the aircraft and the time for one scanner revolution (~42 s) is approximately 8.6 km. The distribution of all observations shows that winds up to 91 m s^{-1} were sampled which represents the highest wind speeds that have been observed by the DWL since its first airborne deployment in 2001.

To assess the accuracy (systematic error) and precision (random error) of the DWL during the campaign, typically comparisons with independent observation types are conducted. During three DLR Falcon research flights (RF02, RF03 and RF04) on 17, 21 and 23 September, coordinated flights with HALO provide 15 dropsondes that are used for a comparison with DWL winds. Dropsondes are small instrument carriers consisting of temperature, pressure and humidity sensors as well as a GPS receiver that transmit their data to the Airborne Vertical

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

Atmospheric Profiling Systems (AVAPS; UCAR/NCAR 1993; Hock and Franklin 1999) onboard the aircraft that consists of a data acquisition and processing unit. AVAPS is a well-established dropsonde system to provide high quality and high resolution profile data from the flight altitude down to the ground (e.g., Wang et al. 2015). During NAWDEX the Vaisala dropsonde version RD94 was used (Vaisala 2017) and the data was quality-controlled using the automatic post-processing Earth Observing Laboratory (EOL) Atmospheric Sounding Processing Environment (ASPEN, https://www.eol.ucar.edu/software/aspen) software. Wind speed accuracy is in the order of 0.2-0.3 m s^{-1} (Holger Vömel 2019, personal communication).

The dropsonde wind observations were vertically interpolated to the DWL vertical resolution of 100 m and after accounting for the drift of the dropsonde, the spatially closest DWL observation was used for comparison. Figure 1c shows a scatter plot for 529 pairs of wind observations from the DWL and dropsondes ranging between 4 m s^{-1} and 55 m s^{-1}. Although, the mean horizontal distance between sets of the compared observations is 10.8 km and maximum distances up to 29 km are reached, no dependence on the distance difference between both observations is discernible. The good agreement is reflected by a high correlation coefficient of 0.99. A linear fit reveals a slope value of 0.99 and an intercept of -0.004 m s^{-1}. The mean bias is 0.05 m s^{-1} and the standard deviation is 1.87 m s^{-1}. A more restrictive selection of data points, with a maximum horizontal distance between dropsonde and DWL of 10 km leads to a reduced number of 245 observations for the comparison and a reduced standard deviation of 1.50 m s^{-1}. These results are in agreement with earlier findings that are summarized in Table 1 following Witschas et al. (2020). Slight differences between the different campaigns may arise from different weather situations and related wind variability and aerosol loads resulting in different signal-to-noise ratios, differences in the retrieval algorithms and quality-control thresholds, or

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

differences in the spatial-temporal collocation. Nevertheless, these results demonstrate the high accuracy and precision of the DWL.

b. Wind profiler data at South Uist

In addition to the airborne observations described above, the stratospheric-tropospheric wind profiler (STP) located on the island of South Uist in the Outer Hebrides, Scotland (Winston, 2004; location indicated in Fig. 1a) provides an overview of the wind conditions during the extended NAWDEX campaign period (10 September – 20 October 2016). The ATRAD STP installed at the site has an operating frequency centered at 64 MHz and is able to provide wind measurements up to an altitude of 20 km with a vertical resolution of 500 m. It runs continuously providing data to European meteorological services through the EUMETNET E-PROFILE Program (http://eumetnet.eu/activities/observations-programme/current-activities/e-profile/). Very high frequency (VHF) radio waves are generated by a 12x12 antenna array. The directional beams are partially scattered off irregularities in the atmospheric refractive index, and the LOS winds are derived from the Doppler-shifted return frequency. Horizontal wind components are constructed from a cyclic sequence of 5 vertical and near-vertical beam pointing directions known as Doppler Beam Swinging. The dwell time for each direction is 1 minute, giving a maximum temporal frequency of 5 minutes, however to reduce measurement errors the data transmitted on the global telecommunication system (GTS) via the E-PROFILE network is averaged over 30 minute periods, and it is this data that is utilized here (data is available for download from the Met Office, 2008). Typical measurement areas at ~10 km altitude are 5x5 km. The STP data was assimilated at ECMWF and Met Office.

The accuracy of the current configuration of the South Uist wind profiler has not been assessed systematically against independent high resolution observations, however, a number of

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

similar STP systems from the same manufacturer located in Australia have recently been evaluated against collocated radiosonde observations by Dolman and Reid (2018). They find the line of best fit between the individual wind components measured by the two techniques to be in the range 0.93-0.97. Earlier STP systems have been systematically evaluated by Dibbern et al. (2001) who found typical mean wind speed biases relative to radiosonde measurements of order 0.09 m s^{-1} with a standard deviation of 1.5 m s^{-1}.

c. Modelled winds

For the comparison, we use ECMWF operational analysis and short-term forecast fields from the atmospheric high resolution model (HRES, IFS cycle 41r2) with spectral truncation TCo1280 (Malardel et al. 2016). The data was retrieved from ECWMF's Meteorological Archival and Retrieval System (MARS) and interpolated to a 0.125°x0.125° longitude-latitude grid (~14 km). The IFS is a hydrostatic atmospheric model that uses a hybrid-pressure vertical coordinate with 137 levels that transition from terrain-following surfaces into pressure surfaces with increasing altitude (Simmons and Burridge 1981). To compare with wind observations, first the pressure at each level is calculated by using the surface pressure before the geopotential height can be derived from integrating the hydrostatic equation using pressure and temperature profiles. Details on the vertical discretization and altitude calculation can be found in the IFS documentation in Part III: Dynamics and Numerical procedures (available at https://www.ecmwf.int/en/forecasts/documentation-and-support). We use 6-h analysis fields (0000, 0600, 1200 1800 UTC) in combination with hourly forecasts initialized from 0000 and 1200 UTC for the intermediate time steps (e.g., Schäfler et al. 2010) as higher temporal frequency reduces the error in interpolating model data to observation points. For example, this strategy is used by many authors for air mass trajectory calculations, despite the differences

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

between analyses and short-range forecasts, because the reduced interpolation error has been shown to reduce net trajectory error (e.g., Stohl et al. 2001).

The NAWDEX wind observations are also compared with operational analyses and forecasts from the UK Met Office using the Met Office Unified Model (MetUM). The MetUM is a non-hydrostatic fully compressible model with deep atmosphere dynamics. The model version in use in 2016 was the GA6.1/GL6.1 science configuration (Walters et al. 2017) operating with a horizontal N768 grid (approx. 17 km grid-spacing in mid latitudes), with 70 vertical levels on a terrain-following hybrid-height Charney–Phillips grid. Since this model is formulated in hybrid-height coordinates, no vertical integration is required to derive altitude values. To compare with the observations, the wind components are output on model levels and simply interpolated in the horizontal and vertical to the coordinates of the observations using linear interpolation in space and time. Forecasts are initialized from analyses at 6-h intervals (0000, 0600, 1200 and 1800 UTC) with data output at 1-h intervals.

Please note that the DWL profile data is an independent data set meaning that it was not assimilated by the IFS or MetUM data assimilation systems. In contrast, all dropsondes released during NAWDEX (Schäfler et al. 2018) and the STP data were distributed on the GTS and assimilated in the ECMWF (Schindler et al. 2020) and the Met Office prediction systems.

Figure 2 shows the distribution of IFS and MetUM model levels between ground and 15 km altitude in comparison with the vertically constant resolution of 100 m for the DWL and 500 m for the STP at South Uist. In the region 8 - 14 km where the jet stream is typically observed, the IFS provides 19 vertical levels with a mean vertical distance of ~300 m ranging from 290 m to 310 m. The MetUM provides 11 levels at a mean vertical separation of ~550 m ranging from 460 m to 630 m in this region. As we are interested in the model capability to capture the

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

observed sharp gradients at the tropopause, we perform the comparisons at the vertical resolution of the DWL and by linearly interpolating the model data in the vertical to the observation location. Likewise, the 1-hourly model data is bi-linearly interpolated in the horizontal to the profile location and linearly in time to the observation time (Schäfler et al. 2010). Please note that for the dropsondes, the model data was interpolated to the location along the fall trajectory of each dropsonde (tracked by GPS). In case of the wind profiler we used data at a 6-hourly time resolution and only compare profiles at the time of the analysis to avoid an influence of short-term forecast error.

3. Case Study

a. Synoptic overview

First, a case study on NAWDEX Intensive Observation Period (IOP) 3 on 23 September 2016 is presented that comprises HALO (RF 03), DLR Falcon (RF 04) and the FAAM Bae 146 (RF 01) flights that observed ascending air masses within cyclone Vladiana (Schäfler et al. 2018). In this paper the focus is on the flight of the DLR Falcon southeast of Iceland between 0710 UTC and 1017 UTC (Fig. 3) that was coordinated with HALO between 0800 UTC and 0900 UTC. After the joint leg, the DLR Falcon returned to Keflavik and HALO turned southwestward to observe a strong warm conveyor belt (WCB) related to cyclone Vladiana (Oertel et al. 2019). At 0900 UTC the center of cyclone Vladiana (V) was located south of Iceland and a second low to the west (Fig. 3a). The occluded frontal system related to Vladiana is visible in the increased relative humidity at 700 hPa north and west of the cyclone center and in the clouds along the cold and warm fronts in the eastern and south-eastern sector of the cyclone. In the upper-level outflow of the WCB, which can be seen from the approaching high-level clouds (Fig. 4), a weak ridge has formed with its axis from northwestern Scotland towards

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

Iceland (Fig. 3b). On their coordinated leg, the DLR Falcon and HALO entered a region of increased jet stream winds along the northeast flank of the ridge (Fig. 3b). Increased jet stream winds follow the 2 PVU contour on the 320 K isentropic surface (compare Figs. 3a and 3b) and a second wind speed maximum occurred along the western flank of the ridge. On the coordinated leg dropsonde observations were made by the HALO aircraft (see colored dots in Fig 3b). The aircraft were separated by only 50-km horizontal distance along the coordinated flight leg. Additionally, the flight was located relatively close to the wind profiler in South Uist, Scotland (Fig. 3b) that was observing the jet stream while it moved over the station.

b. Observations and model evaluation

Figure 5a shows DWL wind speed observations along the entire 2340 km long flight between 0710 UTC and 1017 UTC (see track in Fig. 3a). After take-off at Keflavik, the Falcon initially loitered near Iceland between 0710 UTC and 0800 UTC to wait for the HALO aircraft to join the coordinated flight leg between 0800 UTC and 0900 UTC towards the southeast and after that returned along the same track to Iceland. In the first part of the flight leg, the data coverage in clean and dry air is low and restricted to a band extending from 1000 m to about 1500 m beneath the aircraft and to the lowest ~2 km above the ocean. In the upper band, the signal intensity is high near the aircraft whereas an increased load of sea salt aerosol and low-level clouds increases the atmospheric return near the surface (c.f. low level clouds northeast of the WCB-induced cirrus in Fig. 4). The data coverage improves and the observed wind speeds increase up to a maximum of 58 m s^{-1} when both aircraft approached the upper-level cirrus clouds at about 0825 UTC and entered the region of the jet stream. The return along the same flight track causes the symmetry in the wind field in Fig. 5a. The following discussion concentrates on the coordinated part and the return flight with increased upper-level winds

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

between 5 km and 12 km altitude (grey box in Fig. 5a). The DWL observations in this subset and the complementary in situ and dropsonde observations (Fig. 5b) depict the jet stream. Dropsonde winds above and below the DWL observations confirm that, despite the limited data coverage, the DWL captured the entire vertical extent of the jet stream. Maximum wind speeds follow the dynamical tropopause with increased static stability above, as visible from the large vertical gradient of potential temperature. In the following we use the term tropopause as a synonym for the dynamical tropopause, where PV equals 2 PVU. North of cyclone Vladiana, a colder Arctic air mass was advected beneath the ascending warm air and formed a tropopause fold structure along the transect that was also intersected on the return flight. The ascending warm air mass with elevated tropopause altitude can be characterized by two separate regions. The first part with tropopause altitudes of about 9 km (~0812-0826 UTC and 0948-1000 UTC) features low data coverage in the tropospheric air mass indicating a lack of cirrus clouds, while the second region with the tropopause located at about 10 km altitude (~0826-0948 UTC) is characterized by increased returns from the DWL due to the cirrus clouds.

Figures 5c and 5d show differences of horizontal wind speed between ECMWF IFS and Met Office MetUM forecasts (using +8h, +9h and +10h forecasts for the IFS and +2h, +3h and +4h for the MetUM) and DWL observations, respectively. The IFS shows coherent areas of increased negative wind speed differences above and below the tropopause corresponding to underestimated winds with peak values of up to -17 m s^{-1}. The MetUM wind speed differences are slightly weaker and feature positive and negative regions that range between -10.5 m s^{-1} and 9.5 m s^{-1}. Please note that the depicted error structures are mirrored on the return flight towards Iceland. The consistency of the wind speed differences derived from the three measurement types; DWL, in situ and dropsondes, underlines the reproducibility and representativeness of the

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

measurements. The dropsonde profiles suggest that largest differences occurred near the tropopause. The IFS and MetUM wind speed differences differ substantially, although, it can be noted that the most negative differences in the MetUM tend to occur at approximately the same location as in the IFS. Interestingly, the IFS and MetUM tropopause altitude is different as can be seen from the PV distribution in Fig. 6. The tropopause fold and leading edge of the tropospheric air mass appear earlier along the section in the MetUM which corresponds to a northwestward shift. Similarly, the second increase in tropopause altitude, i.e. the region of low PV values that was approached at about 0820 UTC in the MetUM (Fig. 6a) and is located further northwest along the flight track than in the IFS (Fig. 6b). Towards the southeast of the flight section MetUM overestimates the jet stream wind (Fig. 5d), most likely caused by a different representation between the models of the dynamics associated with the WCB outflow of Vladiana, that is suggested by the higher diagnosed tropopause in the MetUM compared to the IFS in this region. Although this indicates the importance of a correct representation of the tropopause altitude, a vertical shift would be expected to show up as a vertical dipole-like structure in the wind speed differences, while this is not the structure found.

To investigate the representation of winds near the tropopause in more detail, observed and modelled wind profiles at the location of the six dropsondes are examined (Fig. 7). The close correspondence of DWL measurements (dots) and dropsonde winds (colour lines) for these six profiles, is consistent with the general statistical comparison shown in Fig. 1c. The maximum wind speed was observed by the DWL at the location of the easternmost dropsonde with 57.5 m s^{-1} at 10.1 km altitude. Unfortunately, the associated dropsonde was launched at a lower altitude of 8.6 km (after HALO descended to a lower flight level) and therefore did not capture this wind maximum (Fig. 5b). A qualitative comparison of the observations (Fig. 7a) and the IFS profiles

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

interpolated to the observation points (Fig. 7b) shows that the altitude of the wind maxima coincides well, while both the strength of the wind maximum and the vertical gradients are underestimated resulting in increased negative wind speed differences in the jet stream above 9 km (Fig. 7c). The observations exhibit a step-like change in vertical wind shear at ~10 km altitude, which is not represented in the IFS. The MetUM forecasts (Fig. 7e) show a more realistic representation of the peak wind speeds. However, the strong vertical gradients are underestimated especially above the wind maximum where the observed step-like change in wind speed with height is not represented correctly which results in increased wind speed differences (Fig. 7f).

To account for the variability in tropopause altitude along the flight and the height of the wind maximum that differs between the dropsonde locations, wind speeds are displayed with respect to their vertical distance to the tropopause identified by 2 PVU (Figs. 7g-l). Using the tropopause as a reference is an established approach to investigate tropopause sharpness and related trace chemical gradients (e.g., Birner 2006, Pan et al. 2004). In tropopause-relative coordinates, the observed wind profiles transecting the jet stream (sondes 2 to 6) collapse on each other showing that the observed peak wind speed and abrupt change in vertical wind shear is approximately co-located with the dynamic tropopause defined in terms of simulated PV. However, there are differences using the tropopause of the IFS (Fig. 7g) and the MetUM (Fig. 7j). For example, the maximum wind in DWL observations at the easternmost dropsonde profile (dots in Fig. 7g) is situated less than 300 m above the IFS tropopause, while the MetUM tropopause is only 100 m above this DWL wind maximum (Fig. 7j). These displacements are less than the model level spacing in the IFS and MetUM and therefore better correspondence cannot be expected. Although the tropopause location has some inherent uncertainty, difference

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

features from multiple profiles are more coherent in the tropopause-relative framework. The distributions of modelled wind speeds (Figs. 7h and k) and respective differences (Figs. 7i and l) emphasize the finding that the IFS underestimates the wind maxima and tropopause sharpness and that the MetUM performs better in terms of wind speeds and gradients in this particular case. Note also that the observations are compared with longer lead time forecasts for the IFS than for the MetUM (due to the operational forecast frequency). Nevertheless, this analysis shows that the wind speed differences are influenced by diverse uncertainties related to the representation of the peak winds, the strength of vertical wind shear on the stratospheric and tropospheric sides of the tropopause and uncertainty in tropopause altitude.

Figure 7 shows that the vertical gradient of wind speed is under-represented on both sides of the tropopause over a considerable distance (more than a km), which spans several model levels in both the IFS and MetUM. To further investigate the structure of vertical wind shear, Figure 8a shows the magnitude of the vertical shear in the vector wind, calculated at points along the cross section, as derived from the DWL and dropsonde observations. Thin, but horizontally extended, layers of high vertical wind shear are observed along the tropopause and also ~1 km above it. Although each layer is too thin to be resolved in the NWP data (Fig. 8 b and c), both models indicate increased vertical shear above the tropopause. The important question for Rossby wave propagation is whether the vertical wind shear above and below the tropopause is too weak in the models on average, since this would imply a weaker PV gradient.

For a quantitative comparison, Fig. 9 shows horizontal averages of wind speeds and vertical shear in a tropopause-relative framework for this flight. Figure 9a and 9b reiterate the finding of increased wind errors above the tropopause in the IFS compared to MetUM (see also from Fig. 5c and d). Vertical wind shear is higher on the stratospheric side of the tropopause in both

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

models (Fig. 9 c and d), however, clearly underestimated compared to the observations. The higher spread in the observed vertical shear is dominated by the small-scale layers (Fig. 8a) that cannot be represented at the current model resolution. The maximum observed vertical shear by the DWL with a 100-m vertical resolution is 0.23 s^{-1}, which certainly is a local extreme. For this case study, the median observed vertical shear is 0.031 s^{-1} above and 0.013 s^{-1} below the tropopause. Corresponding median values are 0.018 s^{-1}/0.010 s^{-1} for the IFS and 0.021 s^{-1}/0.013 s^{-1} for the MetUM which indicates a significant underestimation of shear, especially above the tropopause, in this case.

4. Statistical assessment of wind speed differences

Section 3 focused on the structure of the observed wind speeds and vertical shear for one case study and gave an indication of significant uncertainties in the representation of jet stream winds in global NWP models, especially at the level of the mid-latitude tropopause. To investigate whether these uncertainties were systematically occurring features during NAWDEX, the following section addresses campaign statistics based on the entire DWL data set and the wind profiler data at South Uist (location in Fig.1).

a. Wind lidar data set

Frequency distributions for all DWL wind speed observations from NAWDEX in tropopause-relative coordinates make use of the IFS definition of the tropopause in Fig. 10a and the MetUM tropopause in Fig. 10b. Both wind distribution and mean and median wind curves look similar. Small differences between both can be explained by slightly variable tropopause altitudes as discussed in section 3b. The highest average winds peak around the tropopause with a maximum median (mean) wind speed of ~41 m s^{-1} (~38 m s^{-1}) which is found in the 500 meters

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

below the tropopause. Above and below the tropopause, winds quickly decline. The altitude range from 1 km above to 2 km below the tropopause provides slightly weaker maxima in the frequency distributions indicating broader distributions and thus more variability in the winds. The highest data coverage from the DWL is found around the tropopause, which is a result from the chosen flight altitude. Some increased frequencies above the tropopause appear at high wind speeds and are related to situations where the tropopause altitude rapidly decreases in the stratospheric air, i.e. on the cyclonic shear side of the jet stream, for example at ~0810 UTC in Fig. 5b. In such situations high wind speeds are attributed to low tropopause altitudes.

The median (mean) wind speed difference of -0.41 m s^{-1} (-0.68 m s^{-1}) for the IFS and -0.15 m s^{-1} (-0.28 m s^{-1}) for the MetUM derived from the 77541 modelled and observed wind speeds is small. Frequency distributions of the differences for 1 km altitude bins relative to the tropopause provide information on the vertical distribution of biases in the IFS (Fig. 10c) and MetUM (Fig. 10d). Generally, the median (mean) differences are small at all altitudes ranging between -1.54 m s^{-1} (-1.72 m s^{-1}) and 0.38 m s^{-1} (0.30 m s^{-1}) in the IFS, and -0.9 m s^{-1} (-1.0 m s^{-1}) and 0.36 m s^{-1} (0.22 m s^{-1}) in the MetUM. Please note that most of the wind speed differences are found to be statistically significant based on the 95% confidence interval that was calculated from 1000 bootstrap samples. Interestingly, the highest variability in the differences is visible in the altitude bin directly above the tropopause in both models indicating increased uncertainty in the representation of the winds at this location. This is particularly striking when viewing individual frequency curves for each range bin (Fig. 11). The differences in the first kilometer above the tropopause provide a significantly broader distribution (standard deviation of 3.98 m s^{-1} for the IFS and 3.82 m s^{-1} for the MetUM) compared to the mean curve (standard deviation of 3.23 m s^{-1} for the IFS and 3.17 m s^{-1} for the MetUM).

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

Figure 10e, f show the magnitude of vertical shear for the DWL data set. The vertical distribution of median and mean vertical shear using IFS and MetUM is remarkably similar around the tropopause. Observed median (mean) values in the troposphere range from 0.01 s^{-1} (0.013 s^{-1}) to 0.016 s^{-1} (0.02 s^{-1}) with values decreasing with height towards the tropopause. Above the tropopause vertical shear values jump up to values of 0.021 s^{-1} (0.023 s^{-1}) before they again decrease to ~0.014 s^{-1} (0.017 s^{-1}). The increased difference between mean and median levels relates to the skewed distributions at all altitudes. The vertical shear difference to the DWL observations of the IFS (Fig. 10 g) and the MetUM (Fig. 10 h) show an underestimation at all levels with the smallest errors in the 2 km below the tropopause. This is in agreement with the case study presented in Fig. 9. Expressed as a ratio of observed and modelled vertical shear, the factor of underestimation ranges between 1.3 and 5 for the median in both models. The underestimation is lower (factor 1.5 to 2) in the upper troposphere where observed vertical shear is small and directly above the tropopause where the simulated vertical shear shows a maximum (c.f. Fig. 10 e, f).

One could ask to what extent this result is reproducible in a different year or season. Therefore, we repeated the statistical comparison for the WindVAL-I campaign that was conducted from Iceland in the period 11 to 29 May 2015 and that used the same DWL instrument to measure horizontal wind speed (Reitebuch et al. 2017; Marksteiner et al. 2018). Fig. A1a shows again increased data coverage around the tropopause. Although the mean winds are smaller than during NAWDEX and almost constant with altitude for this campaign (Fig. A1a), again the largest variability in the wind speed differences occurs in the altitude bin directly above the tropopause (Fig. A1b). Vertical wind shear (Fig. A1c) also shows a comparable distribution with weakest differences in the upper troposphere. As during NAWDEX, the vertical

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

shear in the IFS (Fig. A1d) is too weak at all altitudes with underestimation ratios ranging between 2 and 3.5 being higher in the lower troposphere.

b. Ground-based wind profiler data set

To investigate the representativeness of the DWL comparison with NWP data, the ECMWF and Met Office analysis data are additionally compared with STP wind profiles at South Uist providing a continuous time series in the NAWDEX observation area. During the NAWDEX period the wind situation above South Uist is characterized by large variability (Fig. 12a). Especially in the first half of the period, repeated passages of strong wind events accompanied by increased tropopause variability are noticeable. The tropopause location in MetUM and IFS are located at similar altitudes with a mean difference of approximately 100 m. Jet stream observations are related to IOP 1 (tropical cyclone Ian) on 17 September, IOP 2 (cyclone Ursula) on 22 September, IOP 3 (Vladiana) from 23 to 25 September and IOP 4 (tropical storm Karl) from 27 to 29 September. Increased winds on 3 and 7 October can be related to IOP 6 (the Stalactite Cyclone) and IOP 8, respectively. In the second half of the time series, upper-level wind speeds, as well as the variability of the tropopause, become lower as a block established over Europe (Schäfler et al. 2018).

Figure 12b shows 6-h forecasts from the Met Office which correspond to the background forecasts in the data assimilation process. In the one-month period, two obvious situations appear that feature increased wind speed differences. First, frontal passages, which can be identified from tilted isentropes, most often feature overestimated wind speeds in the lower troposphere. Second, situations with strong upper-level winds, elevated tropopause altitudes and sharp vertical gradients in winds and static stability predominantly feature underestimated wind speeds in the first 2 km above the tropopause. Figure 12c shows the Met Office analysis profiles compared

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

with the STP observations. Obviously the data assimilation of the STP observations reduces the errors in the background field. However, negative analysis differences remain in situations of increased errors in the 6h forecast, e.g. on 12, 17 and 24-25 September. The comparison of ECMWF analysis profiles with the STP observations (Fig. 12c) reveals very similar errors, even in situations of large tropopause variability, which is remarkable as both forecasting systems use different data assimilation schemes and models. Consistent with the DWL observations, the diagnosed wind speed errors show increased uncertainty of the winds above the tropopause with a tendency of an underestimation, especially above tropopause ridges.

5. Conclusions

A unique set of comprehensive airborne and ground-based wind profile observations was used to characterize the structure of the jet stream and to evaluate the representation of winds across the tropopause in the two state-of-the-art global operational NWP forecasting systems of the ECMWF and the Met Office. The study covers the high latitude North Atlantic Ocean where the availability of conventional data sources for winds are sparse. The NAWDEX period was characterized by high wave activity and variable predictability (Schäfler et al. 2018).

The independent (not assimilated) DWL data set features 1922 wind profiles at high horizontal (8.6 km profile spacing) and vertical resolution (100 m) during 8 flights. Comparison of DWL wind profiles with dropsondes demonstrates the low measurement error, which is needed to quantify meteorological analysis errors. Although NWP models are characterized by lower horizontal and vertical resolution, compared to the DWL data, the average representation of the winds is remarkably good. Statistical assessment using the DWL data set provided median (mean) biases of -0.41 m s^{-1} (-0.68 m s^{-1}) for the IFS and -0.15 m s^{-1} (-0.28 m s^{-1}) for the MetUM. The comparison with temporally continuous lidar profiles requires a temporal interpolation from

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

NWP analysis and forecast data, so it is likely that forecast errors may have affected the differences with NWP data. The longer forecast intervals that were used for the ECMWF data (forecasts initialized at 0000 and 1200 UTC) compared to the MetUM (initialized at 0000, 0600, 1200 and 1800 UTC) may have caused slightly higher average negative wind speed differences in the IFS. NWP profiles were found to be smoother and less detailed for the IFS compared to the MetUM. Diagnosed average biases are smaller at all altitudes relative to the early 2000s that were characterized by biases in the order of 5-10 % (Tenenbaum 1991, 1996; Rickard et al. 2001; Cardinali et al. 2003). This study corroborates that recent advances in NWP connected to improved data assimilation methods, improved data quality and availability, as well as increased model resolution and better formulation, have led to a significant improvement of the wind analysis quality in the mid-latitudes. However, Horányi et al. (2015) have shown that already small scale systematic observational wind errors in the order of 1 m s^{-1} are able to significantly deteriorate forecast quality after 24 h.

This study also shows that wind errors still reach values exceeding ±10 m s^{-1} (i.e. about 3σ of the difference distributions) for individual cases and that error structures are of large extent and spatially correlated (up to ~500 km in the horizontal and 1-2 km in the vertical) in the analyses and short-range forecasts of ECMWF and Met Office. DWL measurement errors are found to be smaller than the errors in NWP data and typically uncorrelated. Forecast and analysis error structures are most prominent immediately above the tropopause on the flanks of upper-level ridges where strongest vertical wind-shear occurs (e.g., Fig. 5). The same wind error structures are found in the comparison of modelled profiles with the STP radar profiler data over a 6-week period (Fig. 12). The spatial structure of near-tropopause errors is similar in ECMWF and Met Office short-range forecasts and analyses, even though the forecast models and data

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

assimilation schemes differ greatly. Moreover, increased wind uncertainty directly above the tropopause could be confirmed for the WindVAL-I campaign in 2015.

The different observation types, used in this study, have very different sampling characteristics. The DWL observations represent samples from 8.6 km line segments, the STP profiler measurements represent a volume of size 5 km x 5 km x 500 m (at 10 km) averaged over 30 minutes, while the dropsondes are effectively point measurements along the sonde trajectory. These are compared with winds from NWP models represented on a grid with an approximate horizontal spacing of 15 km and vertical level spacing of 300 m in the IFS, 17 km and 550 m in the MetUM (see Fig. 2). Therefore, such a validation of NWP data will inevitably be affected by a representation (sampling) error (e.g. Janjić et al. 2017). For this reason, data assimilation uses an assigned observation error that is a combination of instrument and representation error. Weissmann et al. (2005) estimate the representation error to range between 1.5 m s^{-1} for a point measurement in a 40 km grid box and 0.15 m s^{-1} for a line measurement through that box. They argue that typical assigned observation errors of 2-3 m/s may be too high. To account for the difference in the representation of the data, the observations could be averaged before comparing. However, this study aimed at investigating how far the models deviate from "nature" as observed by the DWL and STP. The large horizontal and vertical scales of the correlated wind error structures (several hundred km horizontally and 1-2 km vertically) can be represented on the grids used by the NWP models. Furthermore, error features persisted for extended periods of time (hours to several days) in the time-series of the STP (Fig. 12). The magnitude of the errors (up to 10 m s^{-1}) and the systematic occurrence at the flank of and above ridges indicates that these structures cannot be explained by representation and measurement error alone.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

The analysis of vertical wind shear revealed that observed values rapidly increase above the tropopause and that median vertical shear is underestimated in both models at all altitudes by a factor of 1.5 to 5. This is line with Houchi et al. (2010) who found an underestimation by a factor of 2.5 to 3 for vertical shear of the zonal and meridional wind and illustrate that most of the missing vertical shear can be explained by the lower vertical resolution of the model profiles. By vertically averaging winds they estimate an effective vertical resolution for wind shear of 1.7 km for the IFS version in 2006 with 91 model levels. Furthermore, the missing small-scale variability of vertical wind shear that was demonstrated along the DWL cross section (Fig. 8) is in line with their findings.

6. Implications of the findings

Underestimation of vertical shear by models has implications locally for the nature and intensity of turbulence and the parametrization of subgrid-scale processes (Houchi et al. 2010). For example, by changing the bulk Richardson number used in parametrization. In addition, the under-estimation of the change in vertical shear across the tropopause that has been discovered here has a non-local, large-scale consequence: the dynamics of Rossby wave propagation depend on the meridional gradient in the PV distribution which is dominated by the change in vertical shear. Direct calculation of Ertel PV and its gradient across the jet stream from observations requires measurements of horizontal wind and temperature with high resolution in both the vertical and horizontal. This is very difficult to achieve, although Harvey et al. (2020) present an example from a high density dropsonde section crossing the jet stream in NAWDEX IOP4. However, the meridional gradient in quasi-geostrophic PV, q, across a zonal flow, u (see Hoskins and James, 2014) can be estimated using the DWL wind data (without coincident high resolution temperature profile data):

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

$$\frac{\partial q}{\partial y} = \beta - \frac{\partial^2 u}{\partial y^2} - \frac{1}{\rho_R}\frac{\partial}{\partial z}\left(\rho_R \frac{f^2}{N^2}\frac{\partial u}{\partial z}\right) \approx \beta - \frac{2(u_e - u_J)}{L^2} - \frac{f^2}{\Delta z}\left(\frac{\Lambda_s}{N_s^2} - \frac{\Lambda_t}{N_t^2}\right)$$

where ρ_R(z) is a reference density profile (assumed to vary less quickly with z than u(z) to derive the right side approximation), f is Coriolis parameter, β is its meridional gradient, N_t and N_s are the Brunt-Vaisala frequencies for troposphere and stratosphere and Λ_t and Λ_s are the respective vertical wind shears separated by a specified distance Δz across the tropopause zone. The horizontal curvature term is estimated by centred difference over cross-jet scale, L, where u_J represents the jet core speed and u_e is the environmental wind speed at distance L from the core. At 62 N, f = 1.3 x 10^{-4} s^{-1} and β = 1.1 x 10^{-11} m^{-1} s^{-1}. Using numbers from the observed cross-section Fig. 5b, it is estimated that the meridional wind curvature term is approximately 8-12β (using L=600 km, u_J =50 m s^{-1} and u_e = 30 m s^{-1}) and the vertical wind curvature term is as much as 2000-2500β (using Δz of 100 m, N_s = 2 x 10^{-2} s^{-1}, N_t = 10^{-2} s^{-1}, Λ_s = -3 x 10^{-2} s^{-1} , Λ_s = 10^{-2} s^{-1}) illustrating how dominant the change in vertical wind shear is in the estimate of meridional PV gradient in the regions where errors are observed. If the same change in vertical shear in the model is spread over 1 km (compare profiles in observations and analyses in Fig. 7) then this term would be 10 times smaller in the model (although still dominant).

Background forecasts (+6h) for the atmospheric column above the STP profiler at South Uist showed similar wind error structures above the tropopause with higher amplitude than seen in the analyses. This indicates that data assimilation reduces the background forecast model error but cannot eliminate it. Future work is needed to evaluate whether assimilated wind profiles tend to improve near-tropopause wind fields through sharpening the gradients. Pilch Kedzierski et al. (2016) found that static stability increments tend to strengthen the tropopause gradients. Schindler et al. (2020) demonstrate an overall positive impact of additional wind information from NAWDEX radiosonde and dropsonde observations on the mid-tropospheric flow.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

Additional research is needed to quantify errors of other quantities across the tropopause and how these uncertainties relate to our findings. Pilch Kedzierski et al. (2016) indicate an excessively diffuse tropopause in terms of temperature gradients as verified by radio-occultation observations. Another important quantity is water vapor providing a tropopause-based step change in concentration. The resulting sharp peak in longwave radiative cooling at the tropopause is able to strengthen the positive Ertel PV anomaly above, and negative PV anomaly below, the tropopause (Chagnon et al. 2013, Spreitzer et al. 2019) thus increasing tropopause sharpness (Ferreira et al. 2015). Saffin et al. (2017) used the MetUM with PV tracers that diabatic processes, including longwave cooling, microphysics and the turbulent mixing parametrization all act to increase the tropopause PV contrast while the non-conservative numerical effects associated with the dynamical core of the model compete, acting to reduce the PV contrast. In forecasts, the PV anomalies associated with these tendencies saturate in about 24 hours indicating that the model has found its own climatological balance of processes at the tropopause. However, the true balance affecting tropopause structure in the atmosphere, where numerical effects are absent and the tropopause is typically much sharper, is not known. Furthermore, the NAWDEX observations show that a major increase in model vertical resolution near the tropopause (by at least a factor of 3) would be required to resolve the abrupt change in both vertical wind shear and static stability there, indicating scope to increase forecast skill through better representation of the tropopause and its influence on the propagation of Rossby waves.

In August 2018 the European Space Agency (ESA) Aeolus satellite mission was launched, carrying the first wind lidar in space. It is expected to contribute significantly to improved representation of the winds in global analyses and forecasts (e.g., Stoffelen et al. 2005; ESA

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

2008; Reitebuch 2012). It will be interesting to evaluate to what extent a large number of observations from Aeolus in oceanic regions with hitherto sparse wind data coverage will impact winds in the mid-latitudes and more specifically at the tropopause.

Acknowledgments

The DLR Falcon contribution to NAWDEX received funding from DLR, the Naval Research Laboratory (NRL), the European Space Agency (ESA) within the WindVal-II project (Contract 4000114053/15/NL/FF/gp) and the European Facility for Airborne Research (EUFAR; project NAWDEX-Influence). The authors thank the German Science Foundation (DFG) for supporting the HALO contribution to the NAWDEX campaign within the priority program SPP1294 HALO. The authors are grateful for the HALO and Falcon pilots who did a fantastic job to coordinate both aircraft for several coordinated flight legs which allowed us comparing the different data sets. Additionally, we thank the Met Office and the University of Manchester, especially Prof. Geraint Vaughan, for operating the ST Wind Profilers during NAWDEX, including the one on South Uist. We thank Dr. Florian Ewald for providing the SEVIRI satellite image for this publication. In addition we thank ECMWF for providing data access in the framework of the Support Tool for HALO Missions (SPDEHALO) project. B. Harvey is funded through the National Centre for Atmospheric Science national capability programme. J. Doyle acknowledges support from the Chief of Naval Research through the NRL Base Program, PE 61153N. The authors thank Dr. Sonja Gisinger for her valuable comments on the manuscript.

Appendix

In 2015, the WindVAL-I campaign was conducted from Iceland using the same set of instruments on-board the Falcon. Unlike NAWDEX, this campaign focused rather on the preparation of the Aeolus calibration and validation in various wind and cloud scenes than on

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

specifically observing jet stream situation (Reitebuch et al. 2017). Figure A1 shows all 141906 DWL wind observations in tropopause-relative coordinates that were measured from 14 research flights in the surrounding of Iceland.

References

Baker, W. E., and Coauthors, 2014: Lidar-Measured Wind Profiles: The Missing Link in the Global Observing System. *Bull. Amer. Meteor. Soc.*, **95 (4)**, 543–564, doi:10.1175/BAMS-D-12-00164.1.

Baumgart, M., M. Riemer, V. Wirth, F. Teubler, and S.T. Lang, 2018: Potential Vorticity Dynamics of Forecast Errors: A Quantitative Case Study. *Mon. Wea. Rev.*, **146**, 1405–1425, https://doi.org/10.1175/MWR-D-17-0196.1

Belmonte Rivas, M. and A. Stoffelen, 2019: Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT, *Ocean Science*, **15 (3)**, 831--852, doi:10.5194/os-15-831-2019

Birner, T., 2006: Fine-scale structure of the extratropical tropopause region. *J. Geophys. Res.: Atmospheres*, **111 (D4)**, doi:10.1029/2005JD006301.

Cardinali, C., L. Rukhovets, and J. Tenenbaum, 2004: Jet Stream Analysis and Forecast Errors Using GADS Aircraft Observations in the DAO, ECMWF, and NCEP Models. *Mon. Wea. Rev.*, **132 (3)**, 764–779, doi:10.1175/1520-0493(2004)132<0764:JSAAFE>2.0.CO;2.

Chagnon, J., S. Gray, and J. Methven, 2013: Diabatic processes modifying potential vorticity in a North Atlantic cyclone. *Quart. J. Roy. Meteor. Soc.*, **139**(674), 1270–1282. doi:10.1002/qj.2037

Chouza, F., O. Reitebuch, A. Benedetti, and B. Weinzierl, 2016: Saharan dust long-range transport across the Atlantic studied by an airborne Doppler wind lidar and the MACC model. *Atmos. Chem. Phys.*, **16 (18)**, 11 581–11 600, doi:10.5194/acp-16-11581-2016.

Davies, H.C. and M. Didone, 2013: Diagnosis and Dynamics of Forecast Error Growth. *Mon. Wea. Rev.*, **141**, 2483–2501, https://doi.org/10.1175/MWR-D-12-00242.1

Dibbern, J., W. Monna, J. Nash, and G. Peters, 2001: COST Action 76 - final report. Development of VHF/UHF wind profilers and vertical sounders for use in European observing systems. *European Commission*.

Dolman, B. K., I. M. Reid, and C. Tingwell, 2018: Stratospheric tropospheric wind profiling radars in the Australian network. *Earth, Planets and Space*, **70**: 170, doi:10.1186/s40623-018-0944-z

ESA, 2008: ADM-Aeolus Science Report. *SP-1311* ed., URL: http://esamultimedia.esa.int/docs/EarthObservation/SP-1311ADM-Aeolus_Final.pdf

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

Ferreira, A. P., J. M. Castanheira, and L. Gimeno, 2015: Water vapour stratification and dynamical warming behind the sharpness of the Earth's midlatitude tropopause. *Quart. J. Roy. Meteor. Soc.*, **142**(695), 957–970. doi:10.1002/qj.2697

Gray, S., C. Dunning, J. Methven, G. Masato, and J. Chagnon, 2014: Systematic model forecast error in Rossby wave structure. *Geophys. Res. Lett.*, **41 (8)**, 2979–2987, doi:10.1002/2014GL059282

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

Harvey, B., J. Methven, and M.H. Ambaum, 2018: An Adiabatic Mechanism for the Reduction of Jet Meander Amplitude by Potential Vorticity Filamentation. *J. Atmos. Sci.*, **75**, 4091–4106, https://doi.org/10.1175/JAS-D-18-0136.1

Harvey, B., J. Methven, C. Sanchez, A. Schäfler, 2020: Diabatic generation of negative potential vorticity in the North Atlantic and its impact on the jet stream. *Q. J. R. Meteorol. Soc.*, 1-22, doi:10.1002/qj.3747

Hock, T. F., and J. L. Franklin, 1999: The NCAR GPS Dropwindsonde. *Bull. Amer. Meteor. Soc.*, **80 (3)**, 407–420, doi:10.1175/1520-0477(1999)080*h*0407:TNGD*i*2.0. CO;2.

Horányi, A., C. Cardinali, M. Rennie, and L. Isaksen, 2015: The assimilation of horizontal line-of-sight wind information into the ECMWF data assimilation and forecasting system. Part II: The impact of degraded wind observations. *Quart. J. Roy. Meteor. Soc.*, **141,** 1233-1243, doi:10.1002/qj.2551

Hoskins, B.J. and T. Ambrizzi, 1993: Rossby Wave Propagation on a Realistic Longitudinally Varying Flow. *J. Atmos. Sci.*, **50**, 1661–1671, https://doi.org/10.1175/1520-0469(1993)050<1661:RWPOAR>2.0.CO;2

Hoskins, B., and I. N. James, 2014: Rotation in the atmosphere. *Fluid Dynamics of the Midlatitude Atmosphere*, B. J. Hoskins and I. N. James, Eds., John Wiley and Sons, 125–148, https://doi.org/10.1002/9781118526002.ch8.

Houchi, K., A. Stoffelen, G. J. Marseille, and J. De Kloe, 2010: Comparison of wind and wind shear climatologies derived from high-resolution radiosondes and the ECMWF model. *J. Geophys. Res.: Atmospheres*, **115 (D22)**, doi:10.1029/2009JD013196.

Janjić, T., N. Bormann, M. Bocquet, J. A. Carton, S. E. Cohn, S. L. Dance, S. N. Losa, N. K. Nichols, R. Potthast, J. A. Waller, P. Weston, 2017: On the representation error in data assimilation, *Q. J. R. Meteorol. Soc.*, **144**: 1257– 1278. https://doi.org/10.1002/qj.3130

Lux, O., C. Lemmerz, F. Weiler, U. Marksteiner, B. Witschas, S. Rahm, A. Schäfler, and O. Reitebuch, 2018: Airborne wind lidar observations over the North Atlantic in 2016 for the pre-launch validation of the satellite mission Aeolus. *Atmos. Meas. Tech.*, **11 (6)**, 3297–13 3322, doi:10.5194/amt-11-3297-2018.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

Malardel, S., N. Wedi, W. Deconinck, M. Diamantakis, C. Kühnlein, G. Mozdzynski, M. Hamrud, and P. Smolarkiewicz, 2016: A new grid for the IFS. ECMWF Newsletter, 146, 23–28.

Marksteiner, U., C. Lemmerz, C., O. Lux, S. Rahm, A. Schäfler, B. Witschas, and O. Reitebuch, 2018: Calibrations and Wind Observations of an Airborne Direct-Detection Wind LiDAR Supporting ESA's Aeolus Mission. *Remote Sens.*, **10(12),** 2056; doi:10.3390/rs10122056

Martius, O., C. Schwierz, and H. Davies, 2010: Tropopause-level waveguide. *J. Atmos. Sci.*, **67 (3)**, 866–879, doi:10.1175/2009JAS2995.1.

Met Office, 2008: Wind Profiler Observations, Part of the Met Office MetDB System. NCAS British Atmospheric Data Centre, https://catalogue.ceda.ac.uk/uuid/9e22544a66ba7aa902ae431b1ed609d6

Oertel, A., M. Boettcher, H. Joos, M. Sprenger, H. Konow, M. Hagen, and H. Wernli, 2019: Convective activity in an extratropical cyclone and its warm conveyor belt - a case study combining observations and a convection-permitting model simulation. *Quart. J. Roy. Meteor Soc.*,**145**, 1406-1426 doi: 10.1002/qj.3500.

Pan, L. L., W. J. Randel, B. L. Gary, M. J. Mahoney, and E. J. Hintsa, 2004: Definitions and sharpness of the extratropical tropopause: A trace gas perspective. *J. Geophys. Res.: Atmospheres*, **109 (D23)**, doi:10.1029/2004JD004982.

Petersen, R. A., 2016: On the Impact and Benefits of AMDAR Observations in Operational Forecasting—Part I: A Review of the Impact of Automated Aircraft Wind and Temperature Reports. *Bull. Amer. Meteor. Soc.*, **97 (4)**, 585–602, doi:10.1175/BAMS-D-14-00055.1

Pilch Kedzierski, R., L. Neef, and K. Matthes, 2016: Tropopause sharpening by data assimilation. *Geophys. Res. Lett.*, **43 (15)**, 8298–8305, doi:10.1002/2016GL069936.

Reitebuch, O., C. Lemmerz, E. Nagel, U. Paffrath, Y. Durand, M. Endemann, F. Fabre, and M. Chaloupy, 2009: The Airborne Demonstrator for the Direct-Detection Doppler Wind Lidar ALADIN on ADM-Aeolus. Part I: Instrument Design and Comparison to Satellite Instrument. *J. Atmos. Oceanic Technol.*, **26**, 2501–2515, doi:10.1175/2009JTECHA1309.1.

Reitebuch, O., 2012: The space-borne wind lidar mission ADM-Aeolus. in Schumann U. (Ed.): Atmospheric Physics – Background, Methods, Trends. Springer Series on Research Topics in Aerospace. ISBN 978-3-642-30182-7, p. 815-827.

Reitebuch, O., C. Lemmerz, O. Lux, U. Marksteiner, B. Witschas, R. R. Neely, 2017: WindVal– Joint DLR-ESA-NASA Wind Validation for Aeolus; Final Report Contract No. 4000114053/15/NL/FF/gp, European Space Agency (ESA), Noordwijk, The Netherlands.

Rickard, G. J., R. W. Lunnon, and J. Tenenbaum, 2001: The Met Office upper air winds: Prediction and verification in the context of commercial aviation data. *Meteor. Appl.*, **8 (3)**, 351–360, doi:10.1017/S1350482701003115.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

Saffin, L., S. Gray, J. Methven, and K. Williams, 2017: Processes Maintaining Tropopause Sharpness in Numerical Models. *J. Geophy. Res.: Atmospheres*, **122 (18)**, 9611– 9627, doi:10.1002/2017JD026879.

Schäfler, A., A. Dörnbrack, C. Kiemle, S. Rahm, and M. Wirth, 2010: Tropospheric Water Vapor Transport as Determined from Airborne Lidar Measurements. *J. Atmos. Oceanic Technol.*, **27 (12)**, 2017–2030, doi:10.1175/2010JTECHA1418.1.

Schäfler, A., and Coauthors, 2018: The North Atlantic Waveguide and Downstream Impact Experiment. *Bull. Amer. Meteor. Soc.*, **99**, 1607-1637, doi:10.1175/20 BAMS-D-17-0003.1.

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

Schwierz, C., S. Dirren, and H. Davies, 2004: Forced waves on a zonally aligned jet stream. *J. Atmos. Sci.*, **61 (1)**, 73–87, doi:10.1175/1520-0469(2004)061*h*0073: FWOAZA*i*2.0.CO;2.

Simmons, A. J., and D. M. Burridge, 1981: An Energy and Angular-Momentum Conserving Vertical Finite-Difference Scheme and Hybrid Vertical Coordinates. *Mon. Wea. Rev.*, **109 (4)**, 758–766, doi:10.1175/1520-0493(1981)109*h*0758:AEAAMC*i*2.0.CO;2.

Spreitzer, E., R. Attinger, M. Boettcher, R. Forbes, H. Wernli, and H. Joos, 2019: Modification of Potential Vorticity near the Tropopause by Nonconservative Processes in the ECMWF Model. *J. Atmos. Sci.*, **76**, 1709-1726, https://doi.org/10.1175/JAS-D-18-0295.1

Stoffelen, A., and Coauthors, 2005: The atmospheric dynamics mission for global wind field measurement. *Bull. Amer. Meteor. Soc.*, **86 (1)**, 73–88, doi:10.1175/BAMS-86-1-73.

Stohl, A., L. Haimberger, M.P. Scheele, and H. Wernli (2001): An intercomparison of results from three trajectory models. *Meteorol. Applications*, **8**, 127-135.

Tenenbaum, J., 1991: Jet Stream Winds: Comparisons of Analyses with Independent Aircraft Data over Southwest Asia. *Wea. Forecasting*, **6 (3)**, 320–336, doi:10.1175/1520-0434(1991) 006<0320:JSWCOA>2.0.CO;2.

Tenenbaum, J., 1996: Jet Stream Winds: Comparisons of Aircraft Observations with Analyses. *Wea. Forecasting*, **11 (2)**, 188–197, doi:10.1175/1520-0434(1996)011<0188:JSWCOA> 2.0.CO;2.

UCAR/NCAR - Earth Observing Laboratory. (1993). NCAR Airborne Vertical Atmospheric Profiling System (AVAPS). UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.5065/D66W9848 Retrieved March 22, 2017/D66W9848

Vaisala, 2017: *Drosponde RD94*. URL: https://www.vaisala.com/sites/default/files/documents/RD94-Datasheet-B210936EN-B.pdf.

Walters, D., M. Brooks, I. Boutle, T. Melvin, R. Stratton, S. Vosper, H. Wells, K. Williams, N. Wood, T. Allen, and A. Bushell, 2017: The Met Office unified model global atmosphere

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

6.0/6.1 and JULES global land 6.0/6.1 configurations. *Geoscientific Model Development*, **10(4)**, 1487-1520.

Wang, J. J., and Coauthors, 2015: A Long-Term, High-Quality, High-Vertical-Resolution GPS Dropsonde Dataset for Hurricane and Other Studies. *Bull. Amer. Meteor. Soc.*, **96 (6)**, 961–973, doi:10.1175/BAMS-D-13-00203.1.

Weissmann, M., R. Busen, A. Dörnbrack, S. Rahm, and O. Reitebuch, 2005: Targeted observations with an airborne wind lidar. *J. Atmos. Oceanic Technol.*, **22 (11)**, 1706–1719, doi:10.1175/JTECH1801.1.

Winston H., 2004: Vaisala wind profilers support: modernization of upper-air data collection network in the UK. *Vaisala News,* **164**, 24–25

Witschas, B., S. Rahm, A. Dörnbrack, J. Wagner, and M. Rapp, 2017: Airborne Wind Lidar Measurements of Vertical and Horizontal Winds for the Investigation of Orographically Induced Gravity Waves. . *J. Atmos. Oceanic Technol.*, **34 (6)**, 1371–1386, doi:10.23 1175/JTECH-D-17-0021.1.

Witschas, B., C. Lemmerz, A. Geiß, O. Lux, U. Marksteiner, S. Rahm, O. Reitebuch, and F. Weiler, 2020: First validation of Aeolus wind observations by airborne Doppler Wind Lidar measurements. *Atmos. Meas. Tech. Discuss.*, https://doi.org/10.5194/amt-2019-432, in review.

Tables

Table 1: Overview of research campaigns with quantitative comparisons of dropsonde and DWL wind speeds following Witschas et al. (2020).

Campaign	Year	Bias / [m s^{-1}]	Standard deviation/ [m s^{-1}]	Number of observations	Reference
NAWDEX	2016	0.05	1.87	529	
WindVal	2015	-0.03	1.46	938	Reitebuch et al. (2017)
SALTRACE	2013	0.08	0.92	1329	Chouza et al. (2016)
A-TREC	2003	1.2	0.00	740	Weissmann et al. (2005)

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

Figure Caption List

- FIG. 1. (a) Location of DWL wind observations during DLR Falcon flights RF02 to RF09. Black dot marks wind profiler at South Uist, Scotland. (b) Horizontal wind speed vs. altitude for all DWL observations (grey dots). Average winds (thick black line), 25/75 % percentile (thin black lines) and data availability (green line) for each 100 m range gate. (c) Comparison of collocated DWL and dropsonde wind speeds color-coded by horizontal distance between the observations. Red line shows the linear regression line.
- FIG. 2. Vertical distribution of observed and modelled wind data for the DWL (dark blue), the wind profiler at South Uist, Scotland (light blue), the ECMWF IFS (orange) and the Met Office MetUM model (yellow). Please note that IFS model level altitudes vary with surface pressure and temperature profile. The model level distribution is obtained by averaging altitudes for all analysis times (0000, 0600, 1200, 1800 UTC) over South Uist for the period 10 Sep to 19 Oct 2016.
- FIG. 3. ECMWF IFS operational forecast for 23 Sep 2016, 0900 UTC (+09 h): (a) Relative humidity at 700 hPa (color shading), 2 PVU at 320 K (thick black contour) and mean sea level pressure (thin grey contours, in hPa). Purple V indicates the position of cyclone Vladiana. (b) Horizontal wind speed (color shading) and geopotential height (black contours, in dm) at 300 hPa. (a) and (b) are superimposed by flight tracks of the DLR Falcon (0710–1020 UTC, red line) and HALO (0736–1636 UTC, grey line) and (b) shows the coordinated leg between 0800 and 0900 UTC (white line). Colored dots mark the position of six dropsondes released from HALO. Purple triangle shows location of South Uist wind profiler.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

- FIG. 4. Meteosat SEVIRI satellite image at 0830 UTC, 23 Sep 2016 superimposed by flight track of HALO (white) and DLR Falcon (red and orange for the coordinated flight leg between 0800 and 0900 UTC). The satellite image matches with the mid-point in time of the coordinated leg when the aircraft reached the outflow of cyclone Vladiana.
- FIG. 5: (a, b) DWL (colored areas), dropsonde (colored observations along arrows) and in situ (colored line contour on top of DWL observations) wind observations and the respective differences to short-range forecast fields of (c) the ECMWF IFS and (d) the Met Office MetUM on 23 Oct 2016. (a) shows observations along the complete flight while (b, c, d) show a subsection indicated by the dark grey box in (a). (b, c, d) are superimposed by potential temperature (black contours) and dynamical tropopause (2 PVU, thick black contour) from IFS (b, c) and MetUM (d). Colored dots at the top of each dropsonde agree with dropsonde marks in Fig. 3.
- FIG. 6. As in Fig. 5 (b, c, d) but with PV (colored) as represented in the ECMWF IFS (a) and Met Office MetUM (b).
- FIG. 7. Observed and modelled wind speeds for dropsonde (lines) and DWL profiles (dots): (a, g) observations, (b, h) IFS, (c, i) differences to IFS, (d, j) observations, (e, k) MetUM and (f, l) differences to MetUM. Distributions with respect to altitude (a-f) and in tropopause relative altitudes (g-l) using the respective dynamical tropopause of IFS (g-i) and MetUM (j-l). Lidar profiles are closest to the dropsondes at the release time and color coding represents color coding as shown in Fig. 3 and 5.
- FIG. 8. Magnitude of the vertical shear in vector wind for (a) DWL (colored areas) and dropsonde (colored observation along arrows, see also Figs. 3 and 5), (b) the ECMWF

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

IFS and (c) the Met Office MetUM (subset region is indicated in Fig. 5a) on 23 Oct 2016. Thick black contour marks the dynamical tropopause of the IFS (a, b) and (c) MetUM.

- FIG. 9. Distributions of wind speed (a, b) and magnitude of vertical wind shear (c, d) in tropopause-relative coordinates for the subset of the research flight on 23 Sep 2016 shown in Fig. 8. Box-whisker plots for distributions of the DWL observations (blue), the IFS (orange) and the MetUM (red). Mean values are shown by the white lines on the box-whiskers and the colored dots. Black diamond markers on the right hand axes indicate statistical significant difference of the medians at the 95% confidence interval using a Wilcoxon Rank-Sum test.
- FIG. 10. Histograms of (a, b) DWL wind speed (color shading) and (e, f) DWL wind shear magnitude in 1 km altitude bins relative to the (a, e) IFS and (b, f) MetUM dynamical tropopause. Histograms of differences between analysis/short-term forecasts of ECMWF IFS and DWL and Met Office MetUM and DWL wind speeds (c, d) and wind shear magnitude (g, h). Black (grey) solid line shows median (mean) value of the DWL observations (a,b and e,f) and the differences (c,d and g,h)in each altitude bin. Black (grey) dashed line in a,b and e,f show median (mean) values from the NWP forecast in each altitude bin. Red line indicates the data availability in each altitude bin. Black diamonds markers indicate altitude bins with median differences that are statistically significant using the 95 % confidence intervals calculated from 1000 bootstrapping samples.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

- FIG. 11. Histogram of the differences between modelled and observed wind speeds for (a) IFS and (b) MetUM for all altitude bins (dark grey lines) shown in Fig. 10. The distribution for all observations is shown as blue line and the bin representing the first kilometer above the tropopause by the orange line.
- FIG. 12. Time series of (a) STP wind speeds (in m s^{-1}) at a 6 hourly time resolution measured at South Uist Scotland and (b, c, d) the differences of modelled and observed winds (in m s^{-1}). (b) uses +06 h MetUM forecasts, (c) MetUM operational analyses and (d) IFS operational analyses winds. All panels are superimposed by potential temperature (thin contours) and the dynamical tropopause (2 PVU contour) of ECMWF (a, d) and Met Office (b, c). The dashed line in (a) represents the Met Office dynamical tropopause.
- FIG. A1. (a, b) as in Fig. 10a,c and (c, d) as in Fig. 10e,g but for the WindVAL-I campaign conducted from Iceland in May 2015.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

Figures

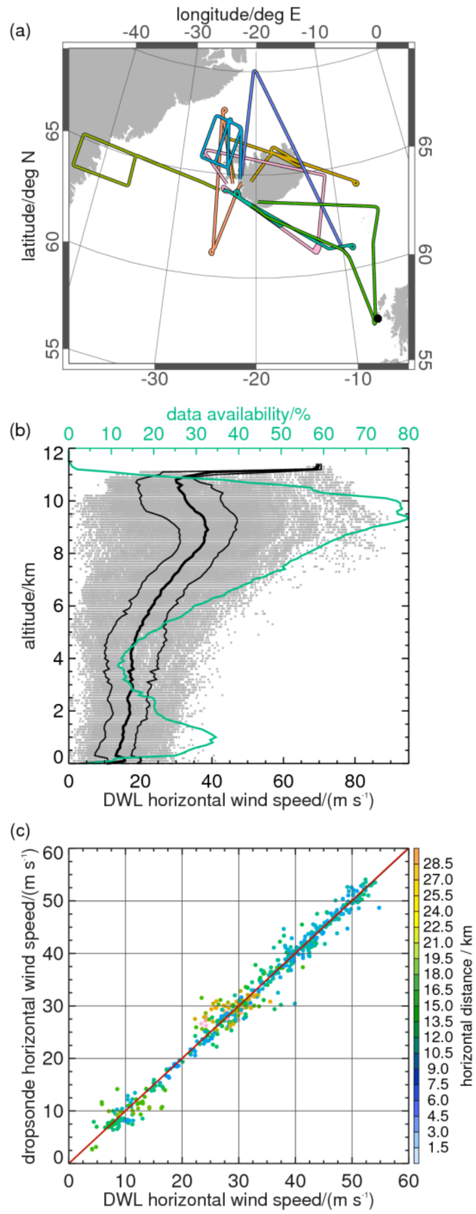

FIG. 1. (a) Location of DWL wind observations during DLR Falcon flights RF02 to RF09. Black dot marks wind profiler at South Uist, Scotland. (b) Horizontal wind speed vs. altitude for all DWL observations (grey dots). Average winds (thick black line), 25/75 % percentile (thin black lines) and data availability (green line) for each 100 m range gate. (c) Comparison of collocated DWL and dropsonde wind speeds color-coded by horizontal distance between the observations. Red line shows the linear regression line.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

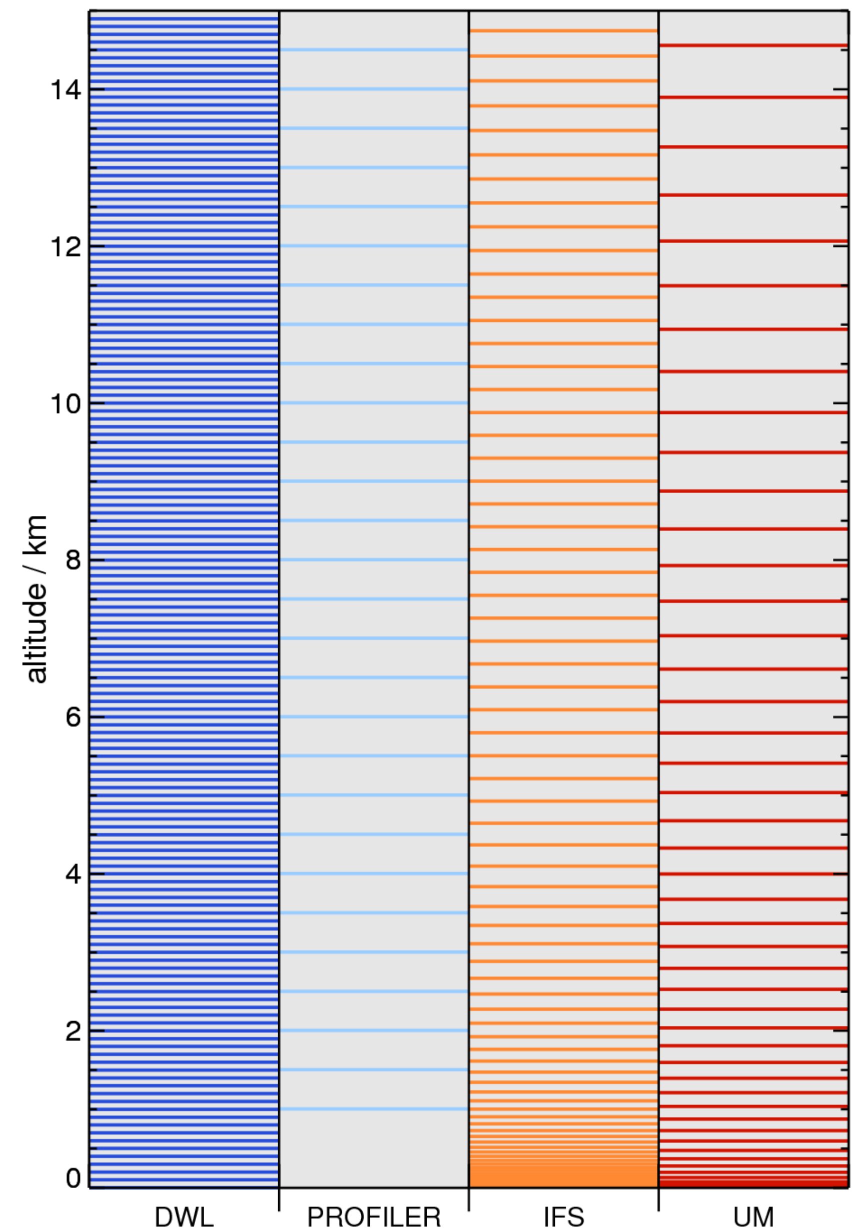

FIG. 2. Vertical distribution of observed and modelled wind data for the DWL (dark blue), the wind profiler at South Uist, Scotland (light blue), the ECMWF IFS (orange) and the Met Office MetUM model (yellow). Please note that IFS model level altitudes vary with surface pressure and temperature profile. The model level distribution is obtained by averaging altitudes for all analysis times (0000, 0600, 1200, 1800 UTC) over South Uist for the period 10 Sep to 19 Oct 2016.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

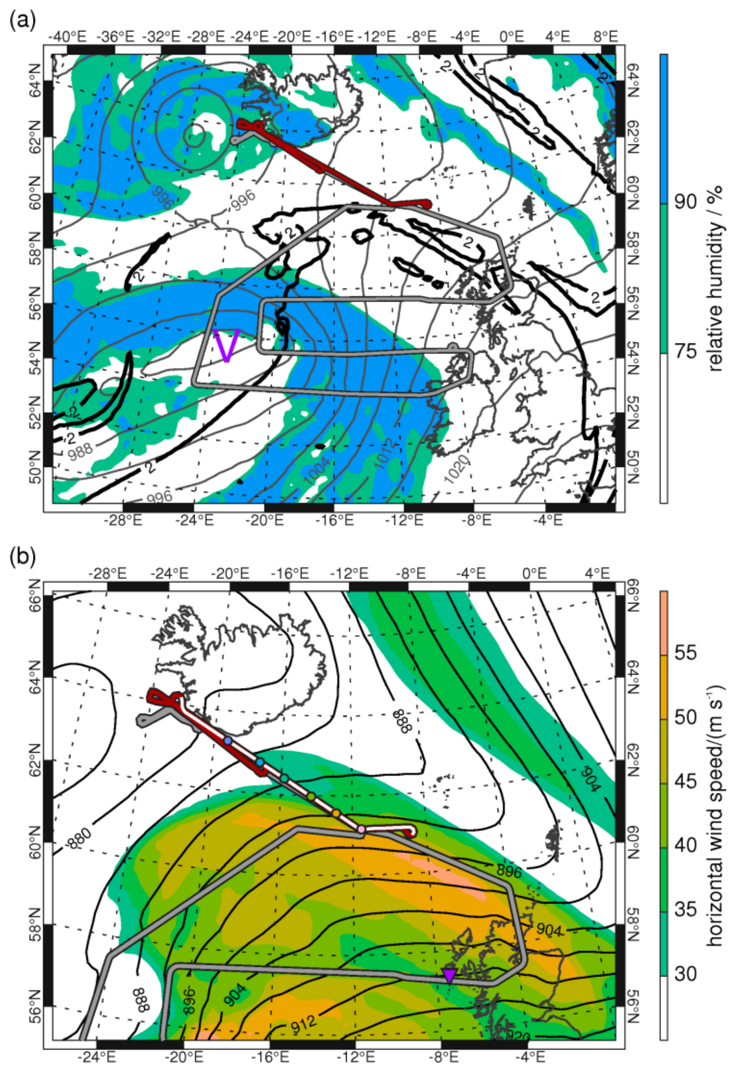

FIG. 3. ECMWF IFS operational forecast for 23 Sep 2016, 0900 UTC (+09 h): (a) Relative humidity at 700 hPa (color shading), 2 PVU at 320 K (thick black contour) and mean sea level pressure (thin grey contours, in hPa). Purple V indicates the position of cyclone Vladiana. (b) Horizontal wind speed (color shading) and geopotential height (black contours, in dm) at 300 hPa. (a) and (b) are superimposed by flight tracks of the DLR Falcon (0710–1020 UTC, red line) and HALO (0736–1636 UTC, grey line) and (b) shows the coordinated leg between 0800 and 0900 UTC (white line). Colored dots mark the position of six dropsondes released from HALO. Purple triangle shows location of South Uist wind profiler.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

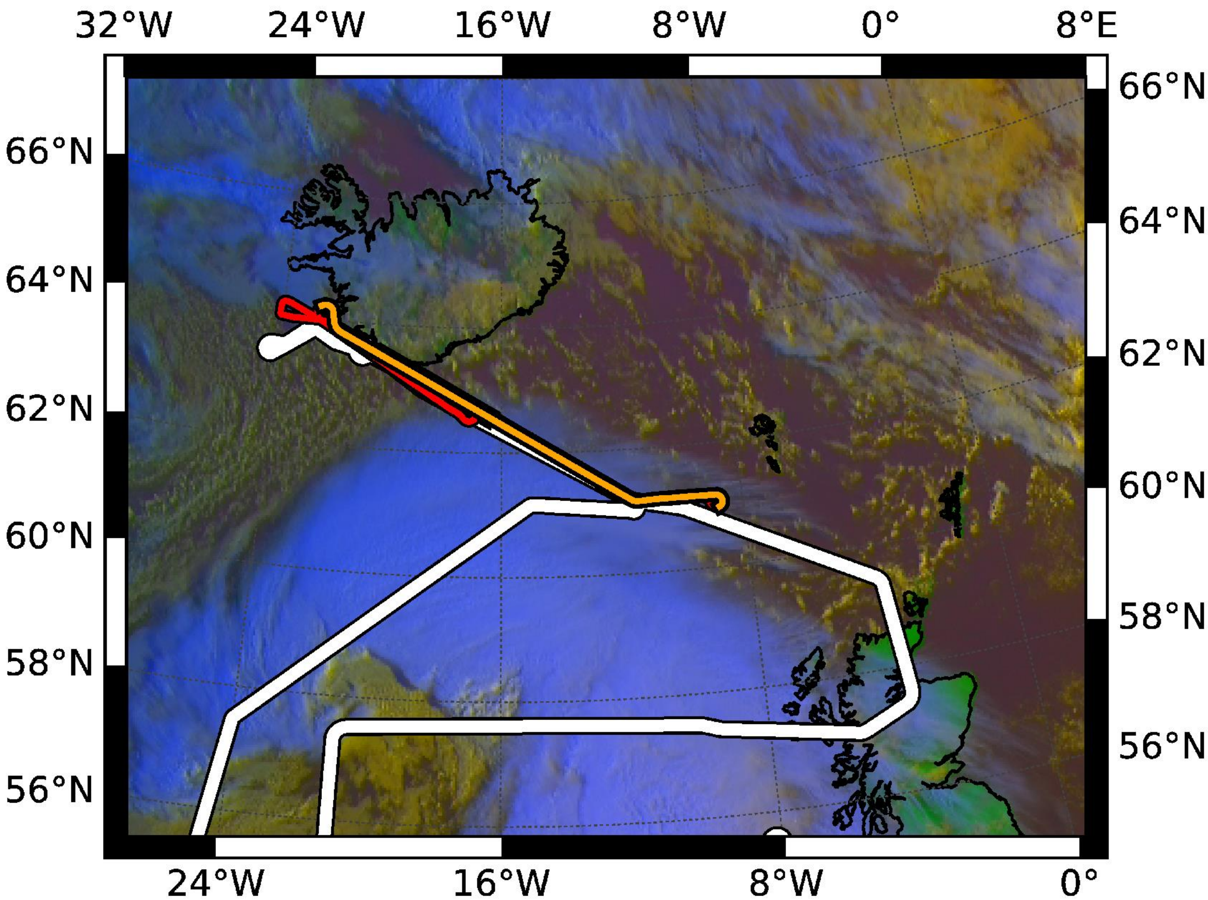

FIG. 4. Meteosat SEVIRI satellite image at 0830 UTC, 23 Sep 2016 superimposed by flight track of HALO (white) and DLR Falcon (red and orange for the coordinated flight leg between 0800 and 0900 UTC). The satellite image matches with the mid-point in time of the coordinated leg when the aircraft reached the outflow of cyclone Vladiana.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

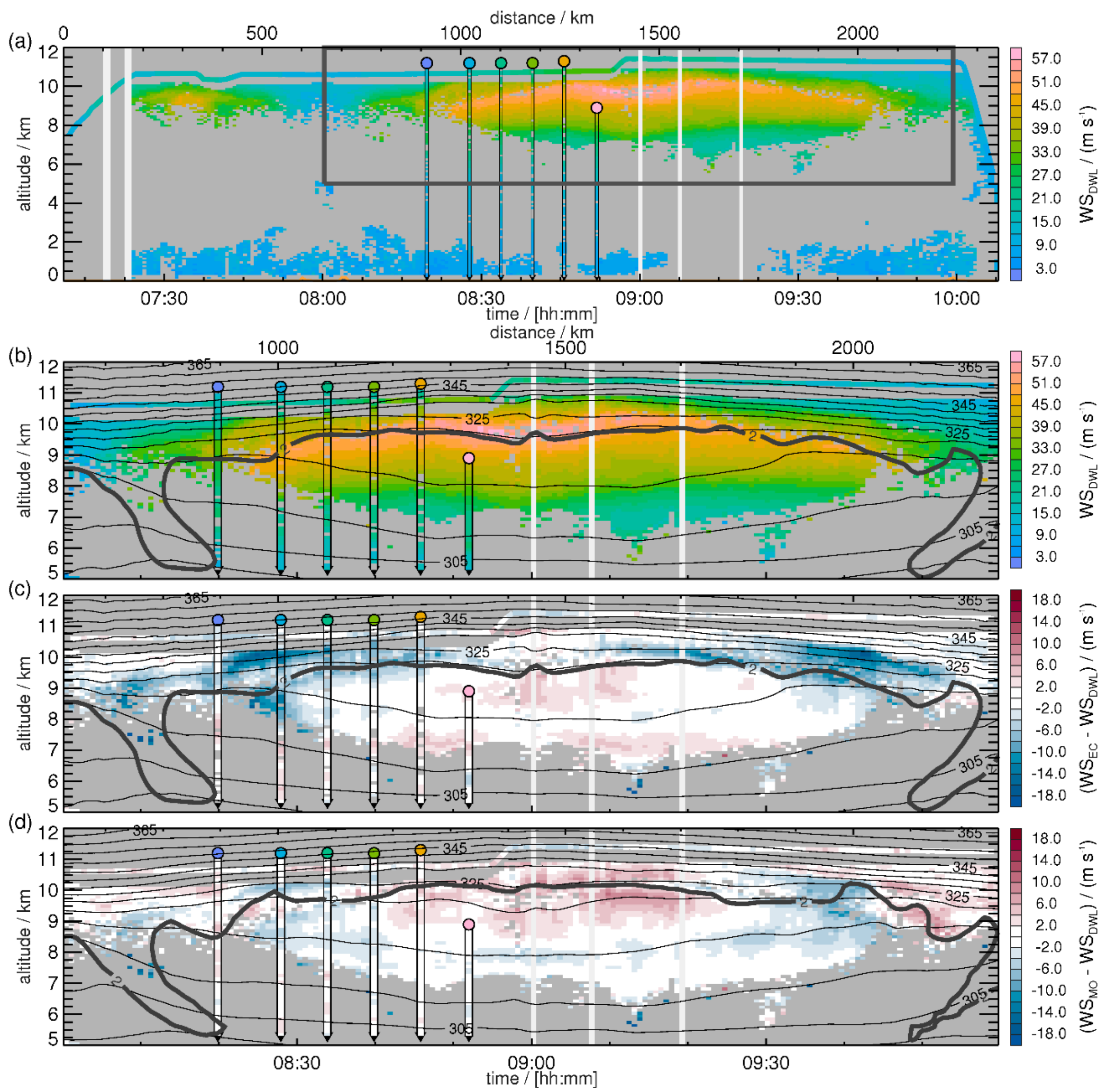

FIG. 5. (a, b) DWL (colored areas), dropsonde (colored observations along arrows) and in situ (colored line contour on top of DWL observations) wind observations and the respective differences to short-range forecast fields of (c) the ECMWF IFS and (d) the Met Office MetUM on 23 Oct 2016. (a) shows observations along the complete flight while (b, c, d) show a subsection indicated by the dark grey box in (a). (b, c, d) are superimposed by potential temperature (black contours) and dynamical tropopause (2 PVU, thick black contour) from IFS (b, c) and MetUM (d). Colored dots at the top of each dropsonde agree with dropsonde marks in Fig. 3.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

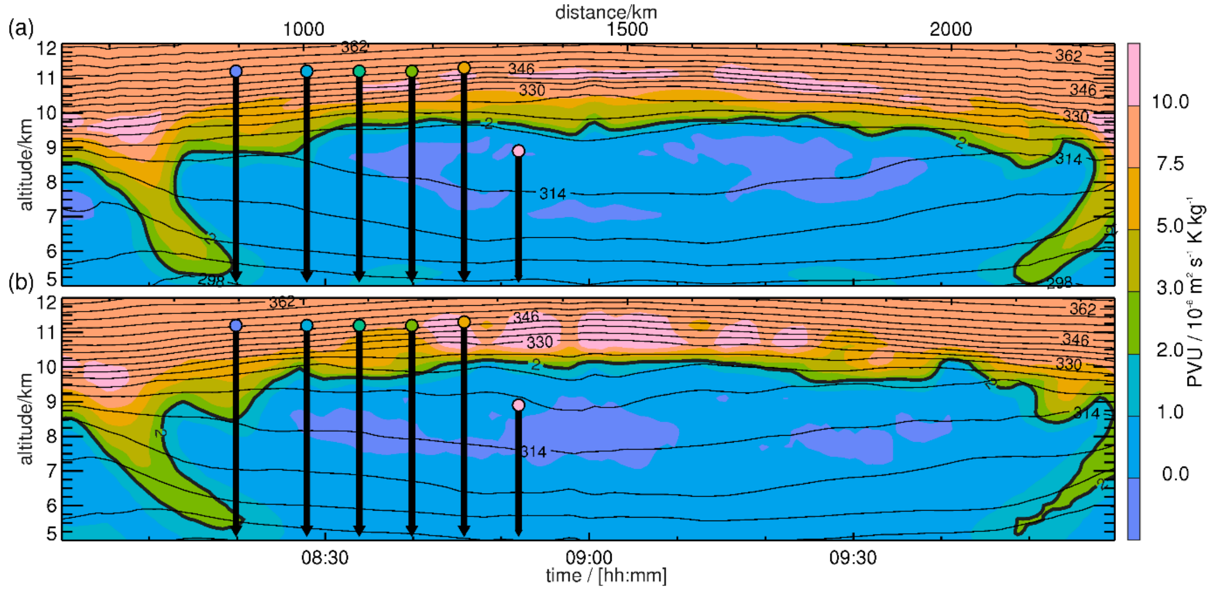

FIG. 6. As in Fig. 5 (b, c, d) but with PV (colored) as represented in the ECMWF IFS (a) and Met Office MetUM (b).

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

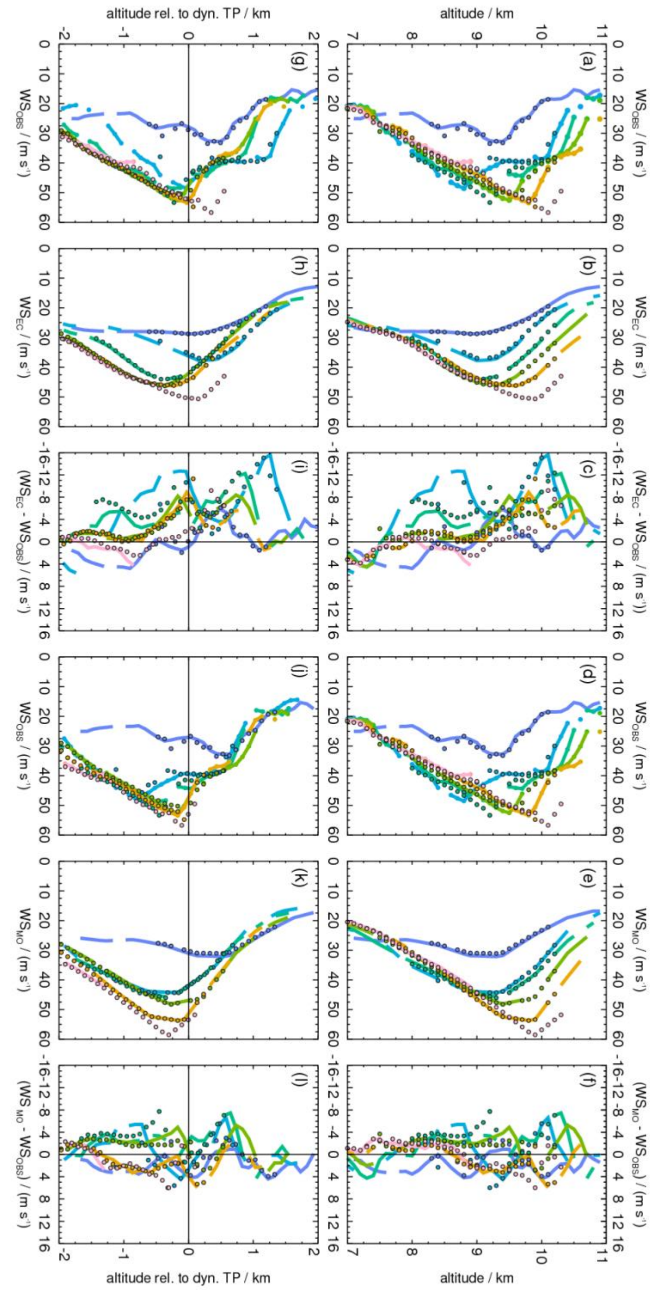

FIG. 7. Observed and modelled wind speeds for dropsonde (lines) and DWL profiles (dots): (a, g) observations, (b, h) IFS, (c, i) differences to IFS, (d, j) observations, (e, k) MetUM and (f, l) differences to MetUM. Distributions with respect to altitude (a-f) and in tropopause relative altitudes (g-l) using the respective dynamical tropopause of IFS (g-i) and MetUM (j-l). Lidar profiles are closest to the dropsondes at the release time and color coding represents color coding as shown in Fig. 3 and 5.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

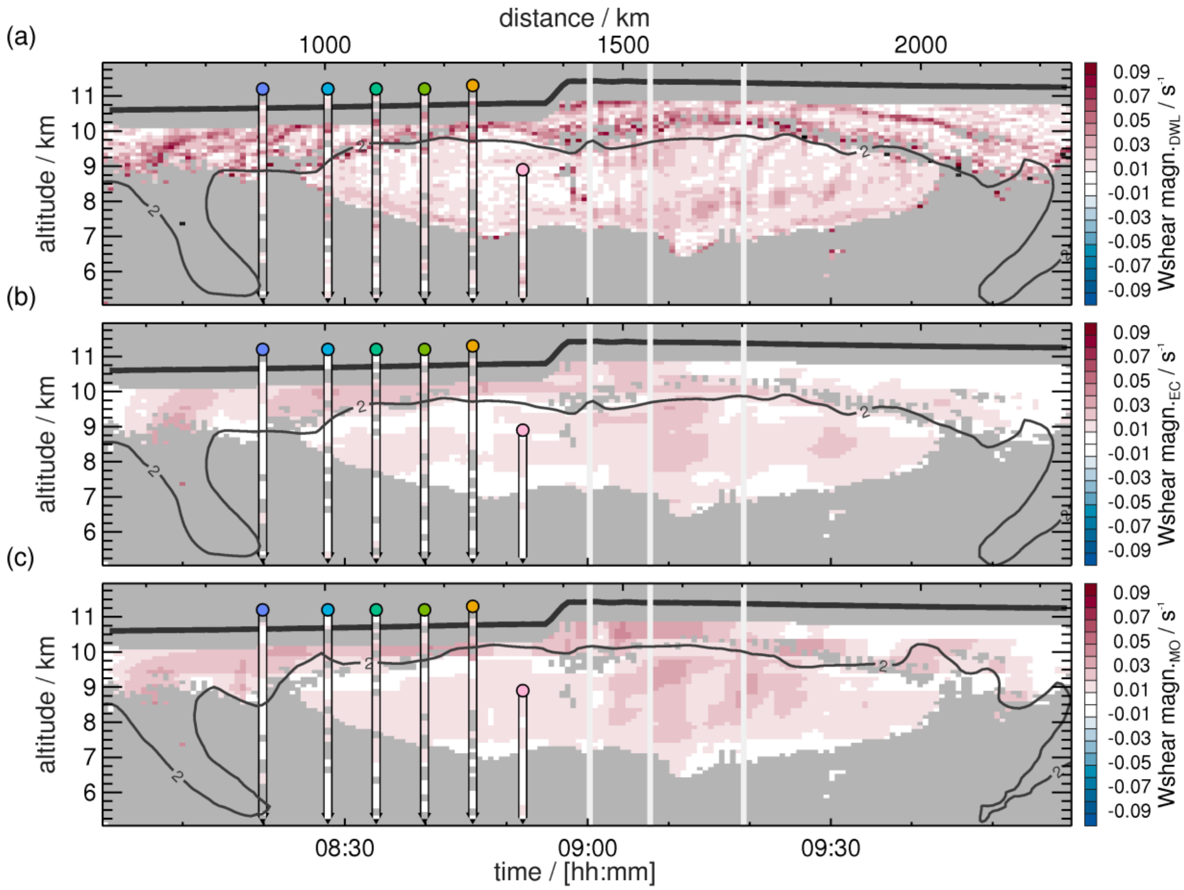

FIG. 8. Magnitude of the vertical shear in vector wind for (a) DWL (colored areas) and dropsonde (colored observation along arrows, see also Figs. 3 and 5), (b) the ECMWF IFS and (c) the Met Office MetUM (subset region is indicated in Fig. 5a) on 23 Oct 2016. Thick black contour marks the dynamical tropopause of the IFS (a, b) and (c) MetUM.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

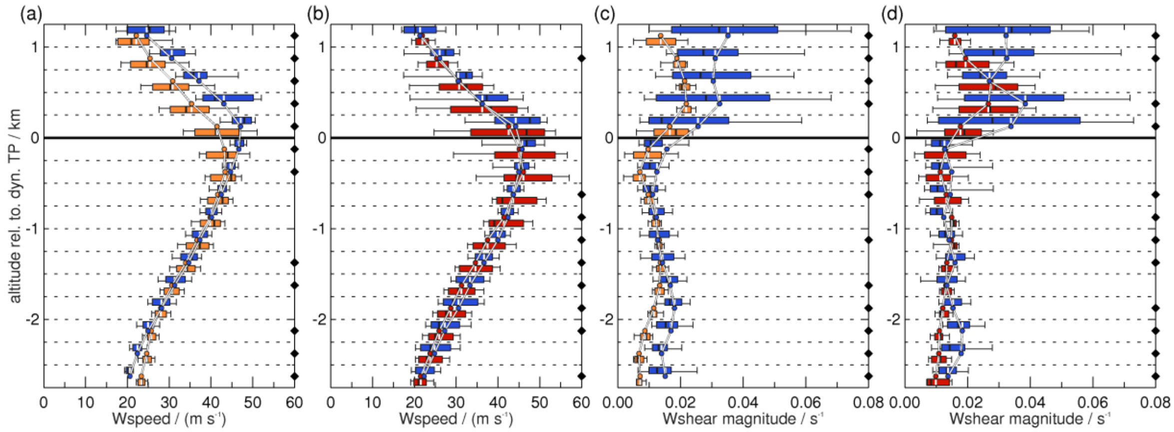

FIG. 9. Distributions of wind speed (a, b) and magnitude of vertical wind shear (c, d) in tropopause-relative coordinates for the subset of the research flight on 23 Sep 2016 shown in Fig. 8. Box-whisker plots for distributions of the DWL observations (blue), the IFS (orange) and the MetUM (red). Mean values are shown by the white lines on the box-whiskers and the colored dots. Black diamond markers on the right hand axes indicate statistical significant difference of the medians at the 95% confidence interval using a Wilcoxon Rank-Sum test.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

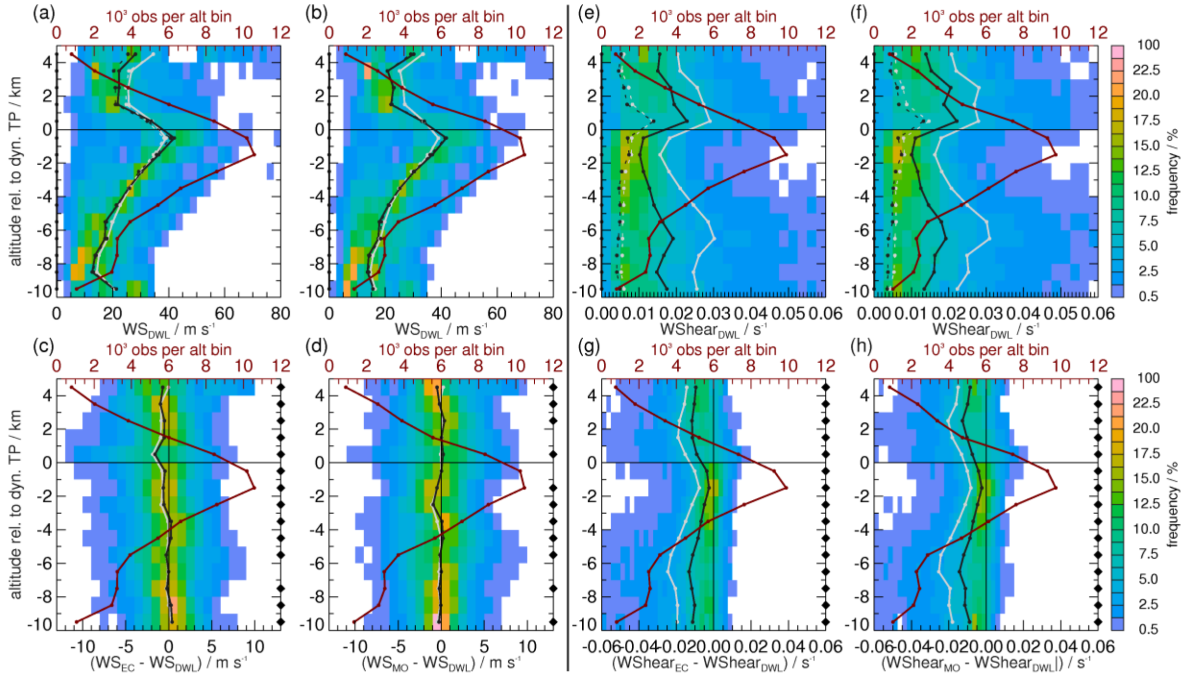

FIG. 10. Histograms of (a, b) DWL wind speed (color shading) and (e, f) DWL vertical wind shear magnitude in 1 km altitude bins relative to the (a, e) IFS and (b, f) MetUM dynamical tropopause. Histograms of differences between analysis/short-term forecasts of ECMWF IFS and DWL and Met Office MetUM and DWL wind speeds (c, d) and vertical wind shear magnitude (g, h). Black (grey) solid line shows median (mean) value of the DWL observations (a,b and e,f) and the differences (c,d and g,h)in each altitude bin. Black (grey) dashed line in a,b and e,f show median (mean) values from the NWP forecast in each altitude bin. Red line indicates the data availability in each altitude bin. Black diamonds markers indicate altitude bins with median differences that are statistically significant using the 95 % confidence intervals calculated from 1000 bootstrapping samples.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

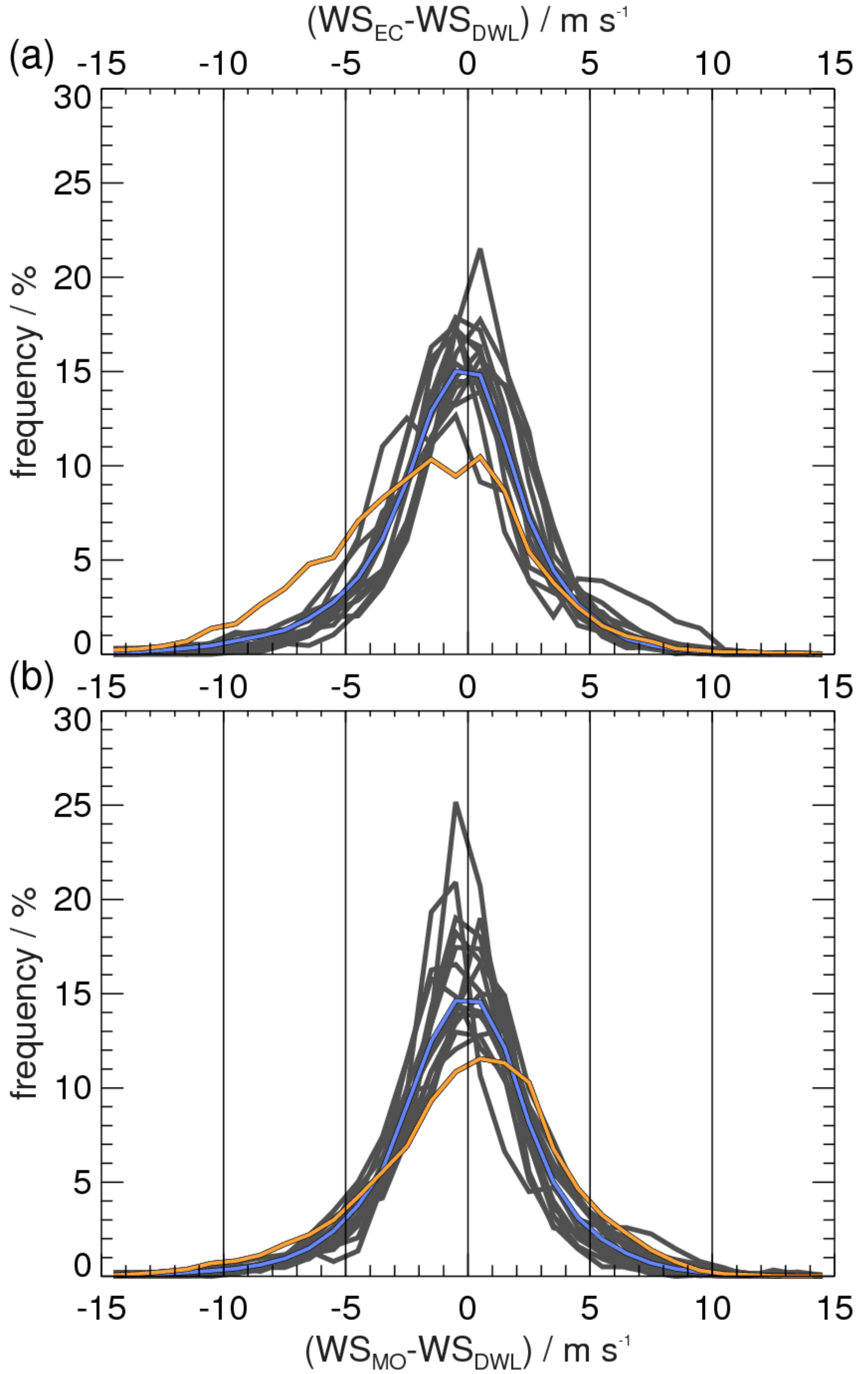

FIG. 11. Histogram of the differences between modelled and observed wind speeds for (a) IFS and (b) MetUM for all altitude bins (dark grey lines) shown in Fig. 10. The distribution for all observations is shown as blue line and the bin representing the first kilometer above the tropopause by the orange line.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

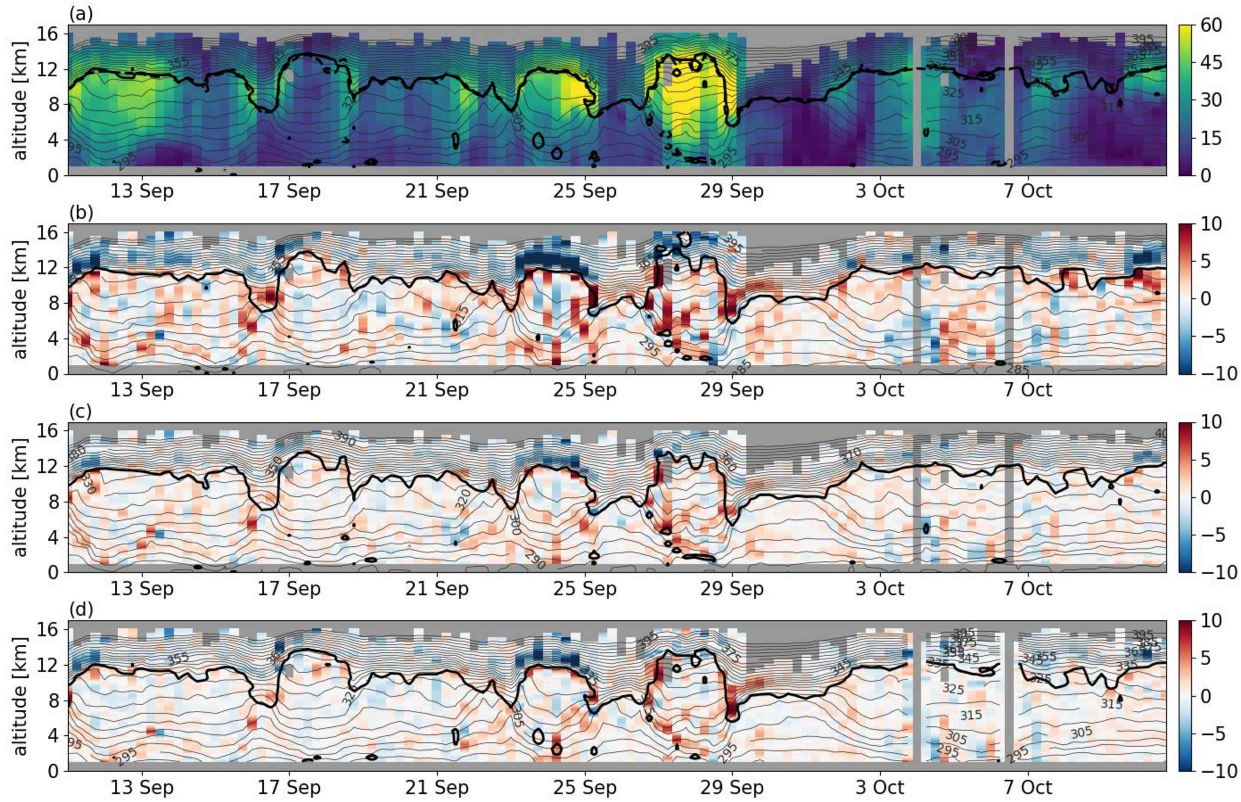

FIG. 12. Time series of (a) STP wind speeds (in m s^{-1}) at a 6 hourly time resolution measured at South Uist Scotland and (b, c, d) the differences of modelled and observed winds (in m s^{-1}). (b) uses +06 h MetUM forecasts, (c) MetUM operational analyses and (d) IFS operational analyses winds. All panels are superimposed by potential temperature (thin contours) and the dynamical tropopause (2 PVU contour) of ECMWF (a, d) and Met Office (b, c). The dashed line in (a) represents the Met Office dynamical tropopause.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.

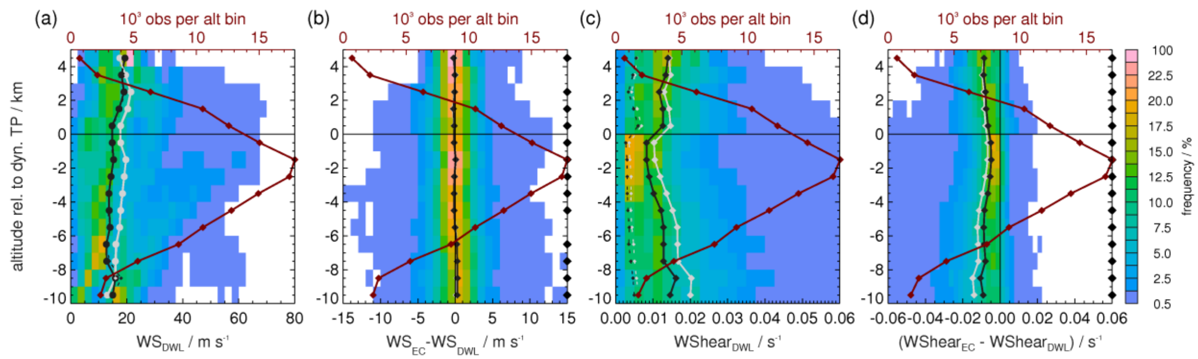

FIG. A1. (a, b) as in Fig. 10a,c and (c, d) as in Fig. 10e,g but for the WindVAL-I campaign conducted from Iceland in May 2015.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-19-0229.1.