Evaluating wind profiles in a numerical weather prediction model with Doppler lidar

Pyry Pentikäinen[1], Ewan J. O'Connor[2,3], and Pablo Ortiz-Amezcua[4,5]

[1]Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University of Helsinki, Helsinki, Finland
[2]Finnish Meteorological Institute, Helsinki, Finland
[3]Department of Meteorology, University of Reading, United Kingdom
[4]Institute of Geophysics, Faculty of Physics, University of Warsaw, Warsaw, Poland
[5]Andalusian Institute for Earth System Research, Granada, Spain

Correspondence: Pyry Pentikäinen (pyry.pentikainen@helsinki.fi)

Abstract. We use Doppler lidar wind profiles from six locations around the globe to evaluate the wind profile forecasts in the boundary layer generated by the operational global integrated forecast system (IFS) from the European Centre for Medium-range Weather Forecasts (ECMWF). The six locations selected cover a variety of surfaces with different characteristics (rural, marine, mountainous urban, coastal urban).

5   We first validated the Doppler lidar observations at four locations by comparison with collocated radiosonde profiles to ensure that the Doppler lidar observations were of sufficient quality. The two observation types agree well, with the mean absolute error (MAE) in wind speed almost always less than $1 \ \mathrm{ms^{-1}}$. Large deviations in the wind direction were usually seen only for low wind speeds, and is due to the wind direction uncertainty increasing rapidly as the wind speed tends to zero.

Time-height composites of the wind evaluation with one-hour resolution were generated and evaluation of the model winds
10   showed that the IFS model performs best over marine and coastal locations, where the mean absolute wind vector error was usually less than $3 \ \mathrm{ms^{-1}}$ at all heights within the boundary layer. Larger errors were seen in locations where the surface was more complex, especially in the wind direction. For example, in Granada, which is near a high mountain range, the IFS model failed to capture a commonly occurring mountain breeze, which is highly dependent on the sub grid-size terrain features that are not resolved by the model. The uncertainty in the wind forecasts increased with forecast lead time, but no increase in the
15   bias was seen.

At one location, we conditionally performed the wind evaluation based on the presence or absence of a low-level-jet diagnosed from the Doppler lidar observations. The model was able to reproduce the presence of the low-level-jet but the wind speed maximum was about $2 \ \mathrm{ms^{-1}}$ lower than observed. This is attributed to the effective vertical resolution of the model being too coarse to create the strong gradients in wind speed observed.

20   Our results show that Doppler lidar is a suitable instrument for evaluating the boundary layer wind profiles in atmospheric models.

1 Introduction

Weather forecasts affect decision making in several weather sensitive sectors (Ebert et al., 2018). Reliable wind forecasts are important for air and maritime transport (Kojo et al., 2011), the wind energy sector (Roulston et al., 2003; Lew et al., 2011), and for the prediction of other atmospheric phenomena. The development of fog (Gultepe et al., 2007), transport of pollution (Kurita et al., 1985) and the vertical profile of PM_{10} (Sekuła et al., 2021) are all dependent on the winds in the planetary boundary layer. According to the WMO Rolling Review of Requirements Statement of Guidance "wind profiles at all levels outside the main populated areas are a top priority among variables that are not adequately measured by current or planned systems" (Andersson, 2017).

Forecast verification is an important step for understanding both the quality of the forecasts produced, and for the further development of numerical weather prediction (NWP) models. Verification methods are commonly separated into two categories: traditional methods, which consist of various statistics (e.g root-mean-square error, mean absolute error, and standard deviation) calculated by comparing observations to a forecast at specific locations; and spatial methods, which estimate the spatial error in the forecast, (i.e. the forecast has predicted correct features but the location is incorrect (Mass et al., 2002; Casati et al., 2008)), by estimating how much the forecast must be manipulated to match the observations.

In recent decades, spatial verification methods have been favoured over traditional verification methods, as the latter cannot take into account spatial errors in the forecast, and thus may cause the forecast to appear much worse than it is in reality. However, spatial verification methods are difficult to use when the quantity being verified is a vector, as is the case for wind, and are typically used only for uni-dimensional variables (e.g. precipitation). While some spatial verification methods have been adapted for wind forecasts (Skok and Hladnik, 2018), traditional verification methods are still in frequent use (e.g. Pennelly and Reuter, 2017; Olson et al., 2019).

Wind forecast verification has been performed using various instruments, including radiosondes (Houchi et al., 2010; Fovell and Gallagher, 2020), Doppler radar (Salonen et al., 2008; Beck et al., 2014), in-situ observations (Skok and Hladnik, 2018; Fovell and Gallagher, 2020), and from satellite scatterometer observations (Accadia et al., 2007). Olson et al. (2019) used Doppler lidar, sodar, radio-acoustic sounding system and radar wind profiler for wind forecast verification to aid NWP development for wind energy use.

Each instrument has their own benefits and disadvantages. In-situ observations tend to have dense networks giving good horizontal coverage, but the observations are limited to the surface; vertical profiles can be obtained from tall masts, but these are much more sparsely located. Radiosondes are able to capture profiles that encompass the whole troposphere but are usually launched only twice a day and hence have limited data availability. Observations from orbiting satellites also have poor temporal resolution at a single location and the measurement uncertainties tend to be high in relation to ground-based instruments (Accadia et al., 2007). Weather radars can provide radial winds with good spatial and temporal resolution (e.g. Holleman, 2005) but require the presence of hydrometeors to make the measurement.

Doppler lidars provide wind profiles with good temporal and vertical resolution within the planetary boundary layer (Pearson et al., 2009; Päschke et al., 2015; Pichugina et al., 2017; Nijhuis et al., 2018), which can be used to create diurnal composites of

verification metrics and allows for examination of transitions in diurnal wind patterns. Most wind retrievals from an individual scanning Doppler lidar assume that the flow is horizontal and that flow is homogeneous, assumptions that may no longer be valid in complex terrain (Bingöl et al., 2009; Klaas-Witt and Emeis, 2022), and in strongly turbulent situations (Rahlves et al., 2022; Robey and Lundquist, 2022), hence additional checks must be performed to ensure the quality of the wind retrievals. In this paper we evaluate wind forecasts by the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) in six locations with varying surface characteristics against winds retrieved by a Doppler lidar. In section 2 we provide details of the measurement stations, the Doppler lidars, and ECMWF IFS. In section 3 we describe the processing applied to the data and discuss the selected verification metrics, with the evaluation results then presented in section 4.

2 Data

2.1 Stations

Six stations with long-term Doppler lidar wind observations were selected from around the globe (see Fig. 1) for comparison with the operational IFS model winds. These stations have different surface characteristics, ranging from marine, rural, to urban and coastal-urban. Four of the stations were operated by the Atmospheric Radiation Measurement program (ARM, Mather et al., 2016) of the U.S. Department of Energy, one by the Finnish Meteorological Institute (FMI), and one by the Andalusian Institute for Earth System Research (IISTA-CEAMA). The six stations selected were:

SGP: Southern Great Plains (SGP), US, operated by ARM. Located in central USA, near the border of Oklahoma and Kansas. The area is flat and rural, consisting mainly of cattle pasture and wheat fields. The SGP atmospheric observatory is the worlds largest climate research facility with an extensive history of atmospheric measurements.

Darwin: Australia, operated by ARM. Located on the tropical northern coast of Australia, the station is on the southern edge of Darwin International airport within the city of Darwin (population 130 000, 2011). The local building and canopy height is rather shallow and the landscape is mostly flat.

Cape Cod: US, operated by ARM during the Two-Column Aerosol Project. Located in Massachusetts on the northeastern coast, the cape is shaped like a hook, protruding in the Atlantic ocean first to the east and then turning to the north. The station was close to the north eastern shore of the hook, separated by 40 km from the mainland by Cape Cod Bay, and while the station is connected to the mainland, it is mostly surrounded by water.

Graciosa: Azores, Portugal, operated by ARM. The station is located on an airfield on the northern coast of Graciosa Island. The island is part of the Azores island group located in the middle of the Atlantic Ocean, 1600 km west from Portugal. The island is small, having a diameter around 10 km.

Granada: Spain, operated by the Andalusian Institute for Earth System Research (IISTA-CEAMA). Located in the city of Granada (population 232 000, 2018), which is at an altitude of approximately 700 m above sea level (asl), with the Sierra

Table 1. Measurement stations and observation periods. Nominal data coverage refers to the proportion of time during the measurement periods that both Doppler lidar and IFS data are available before filtering for data quality. Altitude refers to altitude above mean sea level

.

Station	Coordinates	Surface characteristics	Period	Nominal data coverage	Altitude: Doppler lidar (model surface)	Approx. local solar time
SGP	36.61°N, 97.49°W	Continental, flat	2012-11-28 – 2020-04-20	88.8%	317 m (311 m)	UTC-7
Darwin	12.43°S, 130.89°E	Coastal, tropical	2012-09-21 – 2014-06-26	89.1%	30 m (14 m)	UTC+9
Graciosa	39.09°N, 28.03°W	Marine	2014-10-21 – 2019-12-31	86.0%	30 m (30 m)	UTC-1
Cape Cod	42.03°N, 70.05°W	Coastal	2012-08-16 – 2013-06-19	99.4%	57 m (2 m)	UTC-5
Granada	37.16°N, 3.61°W	Urban, mountainous	2016-05-03 – 2020–02-11	54.8%	680 m (947 m)	UTC
Kumpula	60.33°N, 25.60°E	Urban, coastal	2018-04-10 – 2020-09-30	88.3%	28 m (10 m)	UTC+2

Nevada, the highest mountain range in continental Spain with peaks above 3500 m asl, nearby to the southeast. Various valleys from the mountain range affect the wind flow in the city, making it spatially very heterogeneous. The Doppler lidar is located on top of a three story building on the western side of the city.

90 **Kumpula:** Finland, operated by the Finnish Meteorological Institute (FMI). Located on the southern coast of Finland within the Helsinki metropolitan area (population 1.5 million). The station is within 5 km of the Baltic Sea to the south and the local coastline contains a small archipelago.

Measurement periods and nominal data coverage for each station are listed in Table 1.

2.2 Doppler lidar

95 The observed wind profiles were obtained from Halo Photonics Streamline and Streamline XR Doppler lidars, which are commercially available heterodyne pulsed systems capable of full-hemispheric scanning (instrument specifications given in Table 2). These instruments have a nominal maximum range of 9 km or more, and were operated at temporal resolutions of 1-5 s. Data from the first 60-90 m in range suffers with contamination from the outgoing pulse and is discarded. Scan schedules for each Doppler lidar comprised sequences of conical scans at constant elevation (Velocity Azimuth display, VAD) interspersed 100 with extended periods of vertical stare and other scan types. For this study, we used the vertical profiles of horizontal winds

4

Table 2. Halo Photonics Streamline and Streamline XR heterodyne Doppler lidar specifications.

Wavelength	1.5 μm
Pulse repetition rate	15 kHz
Nyquist velocity	19.8 m s^{-1}
Sampling frequency	50 MHz
Points per range gate	10
Range resolution	30 m
Pulse duration	0.2 μs
Divergence	33 μrad
Antenna	monostatic optic-fibre coupled

derived from the VAD scans, which were at elevation angles from the horizontal of 60, 70 or 75 degrees, depending on the station.

The standard ARM Doppler lidar scan schedule provides one VAD scan at 60 degrees in elevation from horizontal every 15 minutes, with each VAD scan comprising 8 beams equally-spaced in azimuth and each beam integrating 30 000 pulses. For Granada, the Doppler lidar scan schedule provides one VAD scan at 75 degrees in elevation from horizontal every 10 minutes, with each VAD scan comprising 6 (later 12) beams equally-spaced in azimuth with each beam integrating 30 000 (later 45 000) pulses. For Kumpula, the scan schedule provides one VAD scan at 70 degrees elevation from horizontal every 15 minutes, with each VAD scan comprising 24 beams equally-spaced in azimuth with each beam integrating 45 000 pulses.

For the ARM stations, we used the ARM Doppler lidar wind value-added-product (DLPROFWIND4NEWS, Newsom et al., 2015), and for Granada and Kumpula, the winds were calculated using the method of Päschke et al. (2015).

2.3 The Integrated Forecast System (IFS)

We use the Doppler lidar-retrieved wind profiles to evaluate wind forecasts generated by the European Centre for Medium-Range Weather Forecasts (ECMWF) from their operational Integrated Forecast System (IFS). The IFS is a global numerical weather prediction (NWP) system, which assimilates a wide range of observations and generates a range of forecast products. The forecast products include medium range to seasonal predictions, and both deterministic and ensemble forecasts. In this study we only consider the high resolution deterministic medium-range forecast (referred to as HRES), which currently has a horizontal resolution of approximately 9 km and 137 levels in the vertical. The vertical grid spacing is non-uniform and below 15 km varies from 20 to 300 m with higher resolution closer to the ground. There are about 20 model levels in the lowest 1 km, with the vertical resolution ranging from 20 m to 100 m. Forecasts up to 10 days in length are initialised every 12 hours, and the temporal resolution of the model output is one hour.

The assimilation system ingests wind observations from surface-based instrumentation over land (10 m observations from synoptic stations) and sea (ships and buoys), aircraft, radiosondes, radar wind profilers, weather radars and satellites, with wind speed and direction observations being transformed into wind components (u and v). Satellite observations include ocean scatterometer data, processed using dedicated observation operators and assumed to be equivalent to winds at 10 m above the surface (Hersbach, 2010). Atmospheric motion vectors (AMV) derived from satellite-observed cloud advection are assimilated as single-level wind observations (default operation, Bormann et al., 2003).

It should be noted that the IFS is under constant development, with a new version typically becoming operational every 6–12 months. We do not aim to investigate how changes to the IFS affect the wind forecasts but there have been some potentially significant model updates during the evaluation period; at the end of June 2013, the number of vertical levels was increased from 91 to 137 (from 12 to 20 model levels below 1 km), and, in March 2016, the horizontal grid was changed from a cubic-reduced Gaussian grid to an octahedral-reduced Gaussian grid, resulting in an increase in horizontal resolution from 16 km to 9 km. A full description of the IFS, observations that are assimilated, and the model updates over time can be found from ECMWF documentation: https://www.ecmwf.int/en/publications/ifs-documentation.

For this study we were provided with the model winds for the nearest model grid point to the station from day-ahead forecasts, which have been initialised at 12:00 UTC the previous day and correspond to forecast hours t+12 to t+35.

3 Methodology

3.1 Data processing

Except for very close to surface, the IFS has a coarser vertical resolution than the Doppler lidar. Therefore, prior to the evaluation, the IFS winds were interpolated to match the vertical resolution of the Doppler lidar. To ensure that any differences observed between the data sets were due to the performance of the IFS only, the Doppler lidar data was filtered rather strictly to ensure that only data with low uncertainty was included. First, a Signal-to-Noise-Ratio (SNR) threshold of -20 dB was applied (Manninen et al., 2018). Second, we applied a column-wise threshold filter, which identified the first point with SNR below the -20 dB threshold, and then removed all data above. Note that these two steps could be combined into one filter, but separating the steps gave the opportunity to identify where there was no longer continuity in the wind profile. The effect of the column-wise threshold filter was comparatively small when applied after the SNR filter.

Third, all data points having a wind fit residual (Newsom et al., 2015) greater than $1 \, \mathrm{m \, s^{-1}}$ were discarded. The value of the residual is a good indicator of turbulence, and as the wind retrieval assumes horizontal homogeneity (Päschke et al., 2015), wind retrievals with a high residual value cannot be considered reliable.

In some cases, precipitation can adversely affect the Doppler lidar wind retrievals, for example, when the precipitation intensity is not spatially homogeneous. While most of these cases were captured by the filters presented above, there were a few precipitation periods which were not adequately filtered. These were identified using a simple rain filter, which detected whether there were downdrafts exceeding $2 \, \mathrm{m \, s^{-1}}$ in the lowest 200 m of the Doppler lidar vertical velocity profile. If so, the entire profile was discarded. This is obviously not an accurate method for identifying the presence of precipitation, but was

sufficient for detecting cases where precipitation was likely to compromise the wind retrieval and had not been removed by the
155   other filters.

The order of the filters only matters for the column-wise threshold filter, which must be applied after the SNR filter.

While the filtering of strong turbulence and precipitation is necessary to achieve consistently high data quality, it also removes the impact of these regimes from the results. Therefore, while we are able to evaluate the model in low-turbulent, non-precipitating cases, it is important to remember that we are not able to evaluate the IFS performance in strongly-turbulent
160   or precipitating situations.

Figure 2 displays an example of the impact of applying these filters in terms of a diurnal composite, showing the average fraction of data that would be removed by each filter at Granada. The residual filter and SNR filter each remove the largest fraction of data, with the patterns mostly resembling each other. This is because the increasing uncertainty in radial winds at low SNR would also result in a large residual after attempting the wind retrieval. There are some differences between these two
165   filters; the SNR filter for this instrument and location becomes stronger more rapidly above 1 km than the residual filter, but the residual filter is more effective during the daytime boundary layer at altitudes below 1 km. As the residual filter produces similar results to the SNR filter, with the additional benefit of identifying retrievals suffering from strong turbulence, some wind retrieval methods use a residual filter instead of a SNR threshold (Päschke et al., 2015). Here we use both filters in tandem in order to ensure good data quality. The precipitation filter only removes relatively few profiles, as was intended, but these
170   are the profiles that are expected to have the largest errors. The column-wise threshold filter, which is applied after the SNR filter, mostly removes some data points above the boundary layer, and generally removes an extra 3-6% of data in this region. Figures 2a-d show the fraction of data at the nominal scan temporal resolution removed by each filter.

After filtering, the Doppler lidar wind profiles (resolution 10 to 15 minutes) were averaged to match the 1 hour temporal resolution of the wind profiles from IFS. Averaging was performed separately for the u and v components, from which wind
175   speed and direction were then calculated. Figure 2e shows the resulting impact of all filters in terms of the 1-hour average being compared to IFS.

The fraction of filtered data remaining varies from station to station and is dependent on, among others, the dynamics of the boundary layer, aerosol concentration and the specific Doppler lidar configuration (Frehlich and Kavaya, 1991; Pentikäinen et al., 2020). The dynamics of the boundary layer determine both the altitude to which sufficient aerosol is lofted, but also the
180   variation in the strength of turbulence. The SNR is determined by the aerosol concentration and the Doppler lidar configuration.

3.2   Validation of Doppler lidar winds

The four ARM stations launch radiosondes operationally, which provided an ideal opportunity to validate the Doppler lidar wind retrievals. To compare the winds from radiosonde and Doppler lidar, we performed the filtering of the Doppler lidar as described in section 3.1 and then linearly interpolated the high resolution radiosonde wind profile (about 10 m vertical
185   resolution) to match the vertical resolution of the Doppler lidar wind profile (26 m at these locations). We chose four altitudes for the comparison: 117 m above ground level (agl), 247 m agl, 507 m agl, and 1000 m agl.

Figure 3 shows that, for all locations, there are slightly larger differences between the two instruments at the lowest altitude selected (117 m agl) than above. This was expected as the radiosonde launch may still be impacting the balloon track, as the balloon is still accelerating and swinging, and any difference between Doppler lidar location and radiosonde launch location would have the largest impact near the surface. The distance from Doppler lidar to radiosonde launch location was only 40 m at two stations, Cape Cod and Graciosa, increasing to 120 m at Darwin and 250 m at SGP. The observed winds agree well everywhere except for close to the surface at Darwin. At 117 m agl in Darwin, the radiosonde reports consistently higher wind speeds than the Doppler lidar. This is unlikely to be due to distance only, as SGP still compares well even though the distance between the Doppler lidar and radiosonde launch location is further. At Darwin, radiosonde winds were obtained by radar tracking rather than GPS during the period involved in this study and it is thought that a mismatch between the stated and actual launch location is partly responsible for the discrepancy close to the surface.

Except for 117 m agl in Darwin, the mean absolute difference in wind speed is < 1 m s^{-1}. The mean absolute difference in wind direction was < 10 degrees for every site except Darwin. We attribute this to the large proportion of low wind speeds observed at Darwin; Fig. 3 shows that if low wind speeds < 3 m s^{-1} were ignored, the agreement would also be much better at Darwin. Thus, we are confident in the filtered winds retrieved from Doppler lidar.

Kumpula and Granada lack co-located operational radio soundings so it was not possible to perform the validation of Doppler lidar winds in the same manner. However, the Doppler lidars at these stations were of similar type and configuration, therefore we are also confident in the Doppler lidar winds retrieved from these stations.

3.3 Comparison metrics

Table 3. Evaluation metric definition

Metric	Symbol	Perfect value	Range	Explanation
Wind speed bias	B	0	$-\infty, \infty$	Average wind speed error
Mean absolute wind speed error	MAE_S	0	$0, \infty$	Mean magnitude of wind speed error
Mean absolute wind direction error	MAE_D	0	$0, 180$	Mean magnitude of wind direction error
Mean absolute wind vector error	MAE_V	0	$0, \infty$	Mean magnitude of wind vector error

For model evaluation, we calculated 4 different metrics; wind speed bias B (observation minus model forecast), mean absolute wind speed error MAE_S, mean absolute wind direction error MAE_D, and mean absolute wind vector error MAE_V. The possible ranges for each of the metrics are listed in Table 3. Equations for the metrics are listed in Appendix A. All metrics are calculated with respect to the ground level at the Doppler lidar location. Note that the observation and model surface altitude above mean sea level may not agree in locations with locally varying topography. For the calculation of MAE_D, data with wind speed less than 2 m s^{-1} was ignored, as the wind direction uncertainty increases rapidly as the wind speed tends towards 0 m s^{-1} (e.g. Newsom et al., 2017).

4 Results and discussion

Probability density functions of MAE_V at each station are presented in Fig. 4. The distribution of MAE_V includes all winds remaining below 2500 m after filtering as described in section 3.1. The shape of the distributions follow the Weibull distribution

215 and show how the general performance of the IFS changes with respect to the surface type: being best at Graciosa which is marine, followed by the coastal stations of Cape Cod and Kumpula, and becoming poorer at Darwin and especially over the continental surface at SGP and more mountainous region of Granada. However, Fig. 4 does not give any insight into which regimes may be responsible for the poor performance at certain locations.

4.1 Seasonal variation

220 Two stations, SGP and Graciosa, had sufficiently long observation time series to investigate the seasonal variability of the absolute wind vector error. Figure 5 shows a time series of the 30-day moving averages of the absolute wind vector error over SGP at different altitudes. The mean is consistently higher than the median, which is a result of the absolute wind vector errors following a Weibull distribution, but otherwise the mean and median show a similar pattern.

A seasonal cycle for the absolute wind vector error can be seen, particularly at the lower 3 altitudes selected, with larger

225 errors seen during summer and lower errors seen during winter. However, there is significant variation from year to year. There is no clear signal in absolute wind vector error that could be attributed to the upgrade in the resolution of the IFS.

Fig. 6a, shows the monthly MAE_V from SGP, which clarifies the seasonal cycle suggested by Fig. 5. Again, the winter/summer contrast is obvious in the 3 lowest altitudes selected. At 1 km agl, the monthly MAE_V follows the same pattern from April to September, but displays elevated MAE_V in winter, in contrast to the lower altitudes.

230 The time series of the absolute wind vector error from Graciosa (see Fig. A4) also displays a seasonal cycle, although weaker than at SGP. This cycle can be seen in Fig. 6b, where the monthly MAE_V is largest during winter and lower during summer at all altitudes selected.

4.2 Composites of absolute wind vector error and wind speed bias

The vertical and temporal resolution of Doppler lidar wind profiles enable the generation of diurnal composites of MAE_V and

235 B for each station, as shown in Figs. 7 and 8. These metrics were only calculated for time-height intervals containing at least 50 valid data points. The composites are displayed in approximate local solar time (see Table 1).

The degradation of the 12Z t+12-35 wind forecast over time can be seen in MAE_V at almost all sites, with the improvement seen for the new forecast coming at 00Z (approximate local solar time: Graciosa, 23; Kumpula, 2; Cape Cod, 19; SGP, 17; Darwin, 9; Granada, 0). The forecast degradation with time is also visible in MAE_S and MAE_V (see Figs. A2 and A1), but

240 notably, no degradation in terms of B with respect to the forecast time is seen.

The marine and coastal locations, Graciosa, Kumpula, and Cape Cod, display the lowest values of MAE_V at all altitudes, with values generally $< 3 \mathrm{~m~s^{-1}}$. For these locations, the impact of the surface on the winds is relatively small, and satellite scatterometer observations over the ocean or sea basin have been assimilated.

9

Graciosa shows little diurnal variation in B; for 90% of the composite, $-0.3 > B > 0.2 \mathrm{~m~s}^{-1}$. There is more diurnal influence in B for Kumpula, with negative B extending gradually above the surface during the daytime, indicative of the influence of a growing boundary layer. Kumpula also displays a negative bias of about $-1 \mathrm{~m~s}^{-1}$ close to the surface, which we attribute to the impact of the urban surface not being captured adequately within the model grid box. Cape Cod also shows some diurnal variation; at low altitudes (200-400 m agl), there is a variation in B, with positive values during the night. As seen for Kumpula, there are negative values of B during the day extending from the surface upwards, indicating the influence of a growing boundary layer. Kumpula and Cape Cod are coastal locations, so can experience both shallow marine boundary layers and the deeper boundary layers generated over land.

SGP shows a distinct diurnal pattern where $\mathrm{MAE_V} < 3 \mathrm{~m~s}^{-1}$ within the daytime turbulent boundary layer and can exceed $4 \mathrm{~m~s}^{-1}$ during the night. $\mathrm{MAE_V} > 4 \mathrm{~m~s}^{-1}$ is also seen above the boundary layer. The magnitude of B is very small above and within the upper daytime turbulent boundary layer, even when $\mathrm{MAE_V}$ is large. The lowest 300 m of the daytime boundary layer shows a positive bias reaching $0.7 \mathrm{~m~s}^{-1}$. A large positive bias above $1 \mathrm{~m~s}^{-1}$ is seen in the nocturnal boundary layer, which has a maximum at altitudes around 300 m agl and has values exceeding $1.5 \mathrm{~m~s}^{-1}$. The nocturnal pattern of larger $\mathrm{MAE_V}$ and B suggests the presence of a low-level jet (LLJ), which is explored in more detail in section 4.4.

Darwin exhibits a weak diurnal pattern in $\mathrm{MAE_V}$ with lower values during the morning and the largest values at night. However, there is a distinct feature in B that is present throughout the day; a layer of $B < -1 \mathrm{~m~s}^{-1}$ in the lowest 150 m agl with another layer of $B > 0.5 \mathrm{~m~s}^{-1}$ above this. This bias is likely due to the differences in the local surface experienced by the two measurement types and the surface representation in the model.

For Granada, $\mathrm{MAE_V} > 3 \mathrm{~m~s}^{-1}$ almost everywhere, and can exceed $4 \mathrm{~m~s}^{-1}$ above 1 km. Low values of $\mathrm{MAE_V}$ are only seen in the morning boundary layer, where the wind speeds themselves are also generally low (median wind speed around $2 \mathrm{~m~s}^{-1}$, see Fig. A3). If IFS also produces low wind speeds, then $\mathrm{MAE_V}$ will also be small even if the wind direction is not correct. Hence, these low wind speeds explain the magnitude of B also being relatively low close to the surface. Above the surface B becomes more and more negative with increasing altitude, reaching $-1.5 \mathrm{~m~s}^{-1}$ above 1 km agl. During the afternoon, B is positive close to the surface, $0.5 \mathrm{~m~s}^{-1}$, but still decreases with altitude, with B becoming negative above 1 km. This station also shows the challenges of wind verification in weak wind conditions, where the wind direction is much more difficult to predict (see Figs. A1 and A2).

4.3 Mountain breeze

In Granada, the presence of the Sierra Nevada mountains and generally low wind speeds create conditions suitable for the frequent development of nocturnal mountain breezes (katabatic winds, Ortiz-Amezcua et al., 2022). Above the Doppler lidar station in Granada, the mountain breeze can be observed as a shallow layer of weak easterly and south-easterly winds near the surface with the wind direction matching that of the valleys. There are several valleys leading from the mountains to the city and the surrounding areas, and the resulting topography (see Fig. 1) is too complex to be resolved by the IFS model. The smoothed IFS orography would not be expected to capture mountain breezes arriving from valley directions that are not

resolved. However, it is important to know how large the local model forecast errors may be. Even at low wind speeds, katabatic flows can still impact issues such as the formation of fog (Cuxart et al., 2021) and air quality (Li et al., 2018).

To understand the impact of the mountain breeze on errors in the IFS wind profiles, we separated the evaluation into two classes based on the Doppler lidar wind profiles: cases with a mountain breeze, and cases without. To identify mountain breeze cases in the Doppler lidar winds, the mountain breeze must satisfy 4 conditions: the mountain breeze must be present between 2 hours before sunset and 2 hours after sunrise, although it may persist beyond this time; the wind direction should be between 55° and 145° close to the surface; the mountain breeze should be at least 120 m deep; the mountain breeze should have a duration of at least 1 hour.

Figure 9 presents the diurnal composites of $\mathrm{MAE_D}$ for when the mountain breeze is present and when it is not present. At night, $\mathrm{MAE_D}$ ranges from 60° to 85° near the surface when the mountain breeze is present, and from 45° to 70° when the mountain breeze is not present. The daytime $\mathrm{MAE_D}$ is around 40°. While $\mathrm{MAE_D}$ is larger when the mountain breeze is present, the complex topography and the generally low wind speeds makes forecasting the wind direction in Granada difficult under any circumstances.

IFS forecasts the most common nocturnal wind direction near the surface to be from the south, which matches surface observations at Armilla Airbase, located 4 km south-west of the Doppler lidar station (Ortiz-Amezcua, 2019). However, in the city of Granada, where the Doppler lidar is located, south is the least prevalent wind direction near the surface.

4.4 Low level jets

Low level jets occur frequently at SGP (Song et al., 2005). Reliable LLJ prediction is highly relevant for the wind energy sector in terms of energy production (Gadde and Stevens, 2021) and consequently, profits (Roulston et al., 2003). Additionally, nocturnal precipitation from thunderstorms generated by elevated convection initiated by LLJs is common in the Great Plains region (Gebauer et al., 2018).

To examine the ability of IFS to capture LLJs at SGP, we separated the evaluation into two different classes based on whether an LLJ was present in the Doppler lidar observations. LLJ detection was performed using the method described by Tuononen et al. (2017) , which identifies continuous local wind speed maxima in the vertical wind speed profile being both at least $2 \mathrm{~m~s^{-1}}$ stronger and at least 25% stronger than local wind speed minima above and below. Composites of wind speed bias for LLJ cases and non-LLJ cases are presented in Fig. 10. LLJ cases exhibit a bias of more than $2 \mathrm{~m~s^{-1}}$ up to about 500 m during the night, in contrast to non-LLJ cases, where the bias rarely exceeds $1 \mathrm{~m~s^{-1}}$.

Composites of mean wind speeds for the Doppler lidar and IFS during LLJ cases are shown in Fig. 11. While IFS can clearly produce LLJs at approximately the correct height, the mean wind speed is roughly $2 \mathrm{~m~s^{-1}}$ lower than observed by the Doppler lidar. This is likely due to the effective vertical resolution of IFS smoothing the wind speed profile and reducing both the peak wind speed value and the amount of shear below the jet compared to what is seen in the observations (Houchi et al., 2010).

4.5 Timing errors

Almost all of the largest absolute wind vector errors between IFS and Doppler lidar were due to incorrectly timed frontal passages and storms. While the timing of these fronts and storms may have been out by up to 4 hours, IFS was usually able to produce structures that closely resembled the Doppler lidar observations. Therefore, although the absolute wind vector error could be greater than $20 \mathrm{~m~s^{-1}}$, the IFS may still be considered to be performing rather well. We are limited here to observations from a single Doppler lidar, and we can detect displacement errors only as a temporal error. Given a network of such observations with suitable spatial coverage, it might be possible to evaluate displacement errors in the forecast.

An example of large momentary forecast error due to incorrect timing is shown in Fig. 12. The u and v components of IFS and Doppler lidar winds show similar features, but with IFS displaying an approximately 4-hour delay. Whether this incorrect timing is significant or not is dependent on the intended use of the forecasts (Casati et al., 2008).

5 Conclusions

In this study, we used Doppler lidar observations to evaluate the wind profiles produced by a global weather forecast model. We first validated the wind profiles generated from Doppler lidar observations with co-located radiosonde profiles at four locations, to ensure that the Doppler lidar wind profiles were of sufficient quality after filtering. Radiosonde and Doppler lidar winds were in good agreement, with mean absolute wind speed error almost always less than $1 \mathrm{~m~s^{-1}}$. The only exception was seen at Darwin for the lowest 120 m of the profile, and was attributed to the radiosonde winds having been obtained by radar tracking rather than GPS. Discrepancies in the wind direction occur mainly at low wind speeds ($< 3 \mathrm{~m~s^{-1}}$), which arises from the wind direction uncertainty increasing rapidly with decreasing wind speed, and becoming ambiguous as the wind speed tends to $0 \mathrm{~m~s^{-1}}$.

We then evaluated ECMWF IFS wind profiles in the boundary layer with Doppler lidar winds observations at six stations from around the globe, each having different surface characteristics. IFS performed best at stations largely influenced by the ocean or the sea (Graciosa, Cape Cod and Kumpula), with $\mathrm{MAE_V}$ at all altitudes generally $< 3 \mathrm{~m~s^{-1}}$.

IFS errors increased at stations with more complex surface types and local topography, especially in terms of wind direction. In Granada, a station located at the foot of a tall mountain range, IFS failed to capture a commonly occurring katabatic mountain breeze, the formation and direction of which was dependent on the sub grid-size terrain features that the model was not able to resolve. During the mountain breeze, the mean absolute wind direction error between the Doppler lidar observation and the IFS ranged from $60°$ to $85°$. Unsurprisingly, accurate prediction of winds in locations with complex surface properties require better model horizontal resolution than the global IFS.

The evaluation also showed that IFS was able to reproduce LLJs over SGP with a similar structure observed, but with the mean wind speed in the core of the jet being approximately $2 \mathrm{~m~s^{-1}}$ too low. The wind shear in LLJs below the jet maximum was also too low. This is attributed to the effective vertical resolution of the IFS being too coarse to create the observed vertical gradients in the wind speed.

340　　The uncertainty in the wind forecasts (MAE_V, MAE_S, MAE_D) increased with forecast lead time, but no increase in the bias, B, was seen. Some seasonal dependence was observed in the IFS forecast accuracy, but there was no obvious signal in the time series that could be attributed to be the change in IFS horizontal resolution in March 2016.

　　Almost all of the large absolute wind vector errors were due to incorrectly timed frontal passages and storms, where the IFS produced realistic changes in winds but with a temporal error of up to 4 hours. Evaluating these displacement errors in the
345　forecast would require a network of wind profile observations with suitable spatial coverage.

　　It should be noted that the Doppler lidar observations were filtered to ensure their quality and hence our evaluation may not be fully representative of the IFS performance over all regimes. For example, strongly turbulent periods and precipitating periods have not been evaluated with Doppler lidar observations. For these regimes, observations must be acquired by other instrument types.

350　*Code and data availability.* The Doppler lidar wind profiles and radiosonde profiles for ARM stations are available from the Atmospheric Radiation Measurement (ARM) User Facility portal (ARM, 2013; 2014). Doppler lidar wind profiles from Granada (UGR, 2022) and Kumpula (FMI, 2022) are stored on Zenodo. The ECMWF IFS profiles over each station are available from the ACTRIS-Cloudnet Data portal (http://cloudnet.fmi.fi). The code used to generate the data set from which the figures were generated is available is on Zenodo (Pentikäinen and O'Connor, 2022).

355　**Appendix A: Verification metrics**

Wind speed bias is given by

$$B = \frac{1}{N} \sum_{i=1}^{N} (U_{O_i} - U_{F_i}), \tag{A1}$$

　　where N is the number of data points, U_O is the observed wind speed and U_F is the forecasted wind speed.

　　Mean absolute wind speed error is given by

360　$$MAE_S = \frac{1}{N} \sum_{i=1}^{N} |(U_{O_i} - U_{F_i})|. \tag{A2}$$

　　Mean absolute wind direction error is given by

$$MAE_D = \frac{1}{N} \sum_{i=1}^{N} min(|\delta_{O_i} - \delta_{F_i}|, |360° + \delta_{O_i} - \delta_{F_i}|, |360° + \delta_{F_i} - \delta_{O_i}|), \tag{A3}$$

　　where δ_O is the observed wind direction in degrees and δ_F is the forecasted wind speed in degrees.

Mean absolute wind vector error is given by

$$365 \quad \text{MAE}_V = \frac{1}{N} \sum_{i=1}^{N} |(\mathbf{U_{O_i}} - \mathbf{U_{F_i}})|, \tag{A4}$$

where $\mathbf{U_O}$ is the observed wind vector and $\mathbf{U_F}$ is the forecasted wind vector.

Author contributions. PP developed the software for the post-processing of the Doppler lidar winds the wind profile evaluation. EJO supervised the project and generated the Doppler lidar winds for one of the stations. PP and EJO conceptualised the project and wrote the manuscript. POA generated the Doppler lidar winds for one of the stations and contributed to the discussion on winds.

References

Accadia, C., Zecchetto, S., Lavagnini, A., and Speranza, A.: Comparison of 10-m wind forecasts from a Regional Area Model and QuikSCAT Scatterometer wind observations over the Mediterranean Sea, Mon. Wea. Rev., 135, 1945 – 1960, https://doi.org/10.1175/MWR3370.1, 2007.

Andersson, E.: How to evolve global observing systems, ECMWF Newsletter 153, pp. 37–40, https://doi.org/10.21957/9fxea2, 2017.

Atmospheric Radiation Measurement (ARM) user facility: Balloon-Borne Sounding System (SONDEWNPN). 2013-09-28 to 2022-05-29, ARM Mobile Facility (PVC) Highland Center, Cape Cod MA; AMF1 (M1),Eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal (C1), Southern Great Plains (SGP) Central Facility, Lamont, OK (C1), Tropical Western Pacific (TWP) Central Facility, Darwin, Australia (C3). Compiled by E. Keeler, R. Coulter, J. Kyrouac and D. Holdridge. ARM Data Center [Data set], https://doi.org/10.5439/1021460, 2013.

Atmospheric Radiation Measurement (ARM) user facility: Doppler Lidar Horizontal Wind Profiles (DLPROFWIND4NEWS). 2014-10-21 to 2022-03-20, ARM Mobile Facility (PVC) Highland Center, Cape Cod MA; AMF1 (M1), Eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal (C1), Southern Great Plains (SGP) Central Facility, Lamont, OK (C1), Tropical Western Pacific (TWP) Central Facility, Darwin, Australia (C3). Compiled by T. Shippert, R. Newsom and L. Riihimaki. ARM Data Center [Data set], https://doi.org/10.5439/1178582, 2014.

Beck, J., Nuret, M., and Bousquet, O.: Model wind field forecast verification using multiple-Doppler syntheses from a national radar network, Wea. Forecasting, 29, 331 – 348, https://doi.org/10.1175/WAF-D-13-00068.1, 2014.

Bingöl, F., Mann, J., and Foussekis, D.: Conically scanning lidar error in complex terrain, Meteor. Z., 18, 189–195, https://doi.org/10.1127/0941-2948/2009/0368, 2009.

Bormann, N., Saarinen, S., Kelly, G., and Thépaut, J.-N.: The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data, Mon. Wea. Rev., 131, 706–718, 2003.

Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E. E., Brown, B. G., and Mason, S.: Forecast verification: current status and future directions, Meteor. Appl., 15, 3–18, https://doi.org/10.1002/met.52, 2008.

Cuxart, J., Telisman Prtenjak, M., and Matjacic, B.: Pannonian Basin nocturnal boundary layer and fog formation: role of topography, Atmos., 12, https://doi.org/10.3390/atmos12060712, 2021.

Ebert, E., Brown, B., Göber, M., Haiden, T., Mittermaier, M., Nurmi, P., Wilson, L., Jackson, S., Johnston, P., and Schuster, D.: The WMO challenge to develop and demonstrate the best new user-oriented forecast verification metric, Meteor. Z., 27, 435–440, https://doi.org/10.1127/metz/2018/0892, 2018.

Finnish Meteorological Institute: Doppler lidar wind profiles from Kumpula. Compiled by E. O'Connor. Zenodo [Data set], https://doi.org/10.5281/zenodo.6628968, 2022.

Fovell, R. G. and Gallagher, A.: Boundary layer and surface verification of the High-Resolution Rapid Refresh, Version 3, Wea. Forecasting, 35, 2255 – 2278, https://doi.org/10.1175/WAF-D-20-0101.1, 2020.

Frehlich, R. G. and Kavaya, M. J.: Coherent laser radar performance for general atmospheric refractive turbulence, Appl. Opt., 30, 5325–5352, https://doi.org/10.1364/AO.30.005325, 1991.

Gadde, S. N. and Stevens, R. J. A. M.: Effect of low-level jet height on wind farm performance, Journal of Renewable and Sustainable Energy, 13, 013 305, https://doi.org/10.1063/5.0026232, 2021.

Gebauer, J. G., Shapiro, A., Fedorovich, E., and Klein, P.: Convection initiation caused by heterogeneous low-level jets over the Great Plains, Mon. Wea. Rev., 146, 2615 – 2637, https://doi.org/10.1175/MWR-D-18-0002.1, 2018.

415 Gultepe, I., Tardif, R., Michaelides, S., Cermak, J., Bott, A., Muller, M., Pagowski, M., Hansen, B., Ellrod, G., Jacobs, W., Toth, G., and Cober, S.: Fog research: A review of past achievements and future perspectives., Pure and Applied Geophysics, 164, 1121–1159, 2007.

Hersbach, H.: Comparison of C-Band scatterometer CMOD5.N equivalent neutral winds with ECMWF, J. Atmos. Oceanic Technol., 27, 721–736, https://doi.org/10.1175/2009JTECHO698.1, 2010.

Holleman, I.: Quality Control and Verification of Weather Radar Wind Profiles, J. Atmos. Oceanic Technol., 22, 1541 – 1550, 420 https://doi.org/10.1175/JTECH1781.1, 2005.

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

Klaas-Witt, T. and Emeis, S.: The five main influencing factors for lidar errors in complex terrain, Wind Energ. Sci., 7, 413–431, https://doi.org/10.5194/wes-7-413-2022, 2022.

425 Kojo, H., Leviäkangas, P., Molarius, R., and Tuominen, A.: Extreme weather impacts on transport systems, Tech. rep., VTT Working Papers No. 168, https://publications.vtt.fi/pdf/workingpapers/2011/W168.pdf, accessed: 2022-03-22, 2011.

Kurita, H., Sasaki, K., Muroga, H., Ueda, H., and Wakamatsu, S.: Long-range transport of air pollution under light gradient wind conditions, J. Appl. Meteor. Climatol., 24, 425 – 434, https://doi.org/10.1175/1520-0450(1985)024<0425:LRTOAP>2.0.CO;2, 1985.

Lew, D., Milligan, M., Jordan, G., and Piwko, R.: Value of wind power forecasting, in: Proc. of the 91st American Meteorological Society 430 Annual Meeting, the Second Conference on Weather, Climate, and the New Energy Economy, Washington, DC, https://www.osti.gov/biblio/1011280, 2011.

Li, J., Sun, J., Zhou, M., Cheng, Z., Li, Q., Cao, X., and Zhang, J.: Observational analyses of dramatic developments of a severe air pollution event in the Beijing area, Atmos. Chem. Phys., 18, 3919–3935, https://doi.org/10.5194/acp-18-3919-2018, 2018.

Manninen, A., Marke, T., Tuononen, M., and O'Connor, E.: Atmospheric boundary layer classification with Doppler lidar, J. Geophys. Res. 435 Atmos., 123, 8172–8189, 2018.

Mass, C. F., Ovens, D., Westrick, K., and Colle, B. A.: Does increasing horizontal resolution produce more skillful forecasts?: the results of two years of real-time Numerical Weather Prediction over the Pacific Northwest, Bull. Amer. Meteor. Soc., 83, 407–430, 2002.

Mather, J. H., Turner, D. D., and Ackerman, T. P.: Scientific maturation of the ARM Program, Meteorological Monographs, 57, 4.1–4.19, https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0053.1, 2016.

440 Newsom, R. K., Sivaraman, C., Shippert, T. R., and Riihimaki, L. D.: Doppler Lidar Wind Value-Added Product, Tech. rep., DOE ARM Climate Research Facility, Washington, DC (United States), https://doi.org/10.2172/1238069, 2015.

Newsom, R. K., Brewer, W. A., Wilczak, J. M., Wolfe, D. E., Oncley, S. P., and Lundquist, J. K.: Validating precision estimates in horizontal wind measurements from a Doppler lidar, Atmospheric Measurement Techniques, 10, 1229–1240, https://doi.org/10.5194/amt-10-1229-2017, 2017.

445 Nijhuis, A. C. P. O., Thobois, L. P., Barbaresco, F., Haan, S. D., Dolfi-Bouteyre, A., Kovalev, D., Krasnov, O. A., Vanhoenacker-Janvier, D., Wilson, R., and Yarovoy, A. G.: Wind hazard and turbulence monitoring at airports with lidar, radar and mode-S downlinks: the UFO Project, Bull. Amer. Meteor. Soc., 99, 2275–2294, 2018.

Olson, J., Kenyon, J., Djalalova, I., Bianco, L., Turner, D., Pichugina, Y., Choukulkar, A., Toy, M., Brown, J., Angevine, W., Akish, E., Bao, J.-W., Jimenez, P., Kosovic, B., Lundquist, K., Draxl, C., Lundquist, J., McCaa, J., McCaffrey, K., and Cline, J.: Improving wind energy

450   forecasting through Numerical Weather Prediction model development, Bull. Amer. Meteor. Soc., 100, https://doi.org/10.1175/BAMS-D-18-0040.1, 2019.

Ortiz-Amezcua, P.: Atmospheric profiling based on aerosol and Doppler lidar, Ph.D. thesis, Universidad de Granada, http://hdl.handle.net/10481/57771, 2019.

Ortiz-Amezcua, P., Martínez-Herrera, A., Manninen, A. J., Pentikäinen, P. P., O'Connor, E. J., Guerrero-Rascado, J. L., and Alados-

455   Arboledas, L.: Wind and turbulence statistics in the urban boundary layer over a mountain-valley system in Granada, Spain, Remote. Sens., 14, https://doi.org/10.3390/rs14102321, 2022.

Päschke, E., Leinweber, R., and Lehmann, V.: An assessment of the performance of a 1.5 μm Doppler lidar for operational vertical wind profiling based on a 1-year trial, Atmos. Meas. Tech., 8, 2251–2266, https://doi.org/10.5194/amt-8-2251-2015, 2015.

Pearson, G., Davies, F., and Collier, C.: An analysis of the performance of the UFAM pulsed Doppler lidar for observing the boundary layer,

460   Journal of Atmospheric and Oceanic Technology, 26, 240 – 250, https://doi.org/10.1175/2008JTECHA1128.1, 2009.

Pennelly, C. and Reuter, G.: Verification of the Weather Research and Forecasting Model when forecasting daily surface conditions in Southern Alberta, Atmos.Ocean, 55, 31–41, https://doi.org/10.1080/07055900.2017.1282345, 2017.

Pentikäinen, P., O'Connor, E. J., Manninen, A. J., and Ortiz-Amezcua, P.: Methodology for deriving the telescope focus function and its uncertainty for a heterodyne pulsed Doppler lidar, Atmos. Meas. Tech., 13, 2849–2863, https://doi.org/10.5194/amt-13-2849-2020, 2020.

465 Pentikäinen, P. and O'Connor, E. J.: Processing code for gmd-2022-150. Zenodo [Code], https://doi.org/10.5281/zenodo.7235694, 2022.

Pichugina, Y. L., Banta, R. M., Olson, J. B., Carley, J. R., Marquis, M. C., Brewer, W. A., Wilczak, J. M., Djalalova, I., Bianco, L., James, E. P., Benjamin, S. G., and Cline, J.: Assessment of NWP Forecast Models in Simulating Offshore Winds through the Lower Boundary Layer by Measurements from a Ship-Based Scanning Doppler Lidar, Mon. Wea. Rev., 145, 4277–4301, https://doi.org/10.1175/MWR-D-16-0442.1, 2017.

470 Rahlves, C., Beyrich, F., and Raasch, S.: Scan strategies for wind profiling with Doppler lidar – an large-eddy simulation (LES)-based evaluation, Atmos. Meas. Tech., 15, 2839–2856, https://doi.org/10.5194/amt-15-2839-2022, 2022.

Robey, R. and Lundquist, J. K.: Behavior and mechanisms of Doppler wind lidar error in varying stability regimes, Atmos. Meas. Tech., 15, 4585–4622, https://doi.org/10.5194/amt-15-4585-2022, 2022.

Roulston, M. S., Kaplan, D. T., Hardenberg, J., and Smith, L. A.: Using medium-range weather forecasts to improve the value of wind energy

475   production, Renewable Energy, 28, 585 – 602, 2003.

Salonen, K., Järvinen, H., Järvenoja, S., Niemelä, S., and Eresmaa, R.: Doppler radar radial wind data in NWP model validation, Meteor. Appl., 15, 97–102, https://doi.org/10.1002/met.47, 2008.

Sekuła, P., Bokwa, A., Bartyzel, J., Bochenek, B., Chmura, Ł., Gałkowski, M., and Zimnoch, M.: Measurement report: Effect of wind shear on PM_{10} concentration vertical structure in the urban boundary layer in a complex terrain, Atmos. Chem. Phys., 21, 12 113–12 139,

480   https://doi.org/10.5194/acp-21-12113-2021, 2021.

Skok, G. and Hladnik, V.: Verification of gridded wind forecasts in complex alpine terrain: a new wind verification methodology based on the neighborhood approach, Mon. Wea. Rev., 146, 63 – 75, https://doi.org/10.1175/MWR-D-16-0471.1, 2018.

Song, J., Liao, K., Coulter, R. L., and Lesht, B. M.: Climatology of the low-level jet at the Southern Great Plains atmospheric boundary layer experiments site, J. Appl. Meteor., 44, 1593 – 1606, https://doi.org/10.1175/JAM2294.1, 2005.

485 Tuononen, M., O'Connor, E. J., Sinclair, V. A., and Vakkari, V.: Low-level jets over Utö, Finland, based on Doppler lidar observations, J. Appl. Meteor. Climatol., 56, 2577 – 2594, https://doi.org/10.1175/JAMC-D-16-0411.1, 2017.

University of Granada IISTA-CEAMA: Doppler lidar wind profiles from Granada. Compiled by P. Ortiz-Amezcua and L. Alados-Arboledas. Zenodo [Data set], https://doi.org/10.5281/zenodo.6628923, 2022.

Figure 1. Map displaying the locations of the Doppler lidar observation stations (SGP orange, Cape Cod blue, Graciosa black, Granada cyan, Kumpula green, Darwin magenta), with inset zoomed maps highlighting the most important terrain features (no zoomed map is provided for SGP, which is located in flat plains). For Granada, a topographical map has been selected to highlight the proximity of the mountains and the various valleys leading into the city. The Doppler lidar is located near the western edge of the city, with the eastern edge extending approximately 2 km eastwards to where the mountain valleys begin. The topographical map has been adapted from Granada's Geographical Information Systems website, while other inset maps have been obtained from Open Street Map (© OpenStreetMap contributors 2022. Distributed under the Open Data Commons Open Database License (ODbL) v1.0).
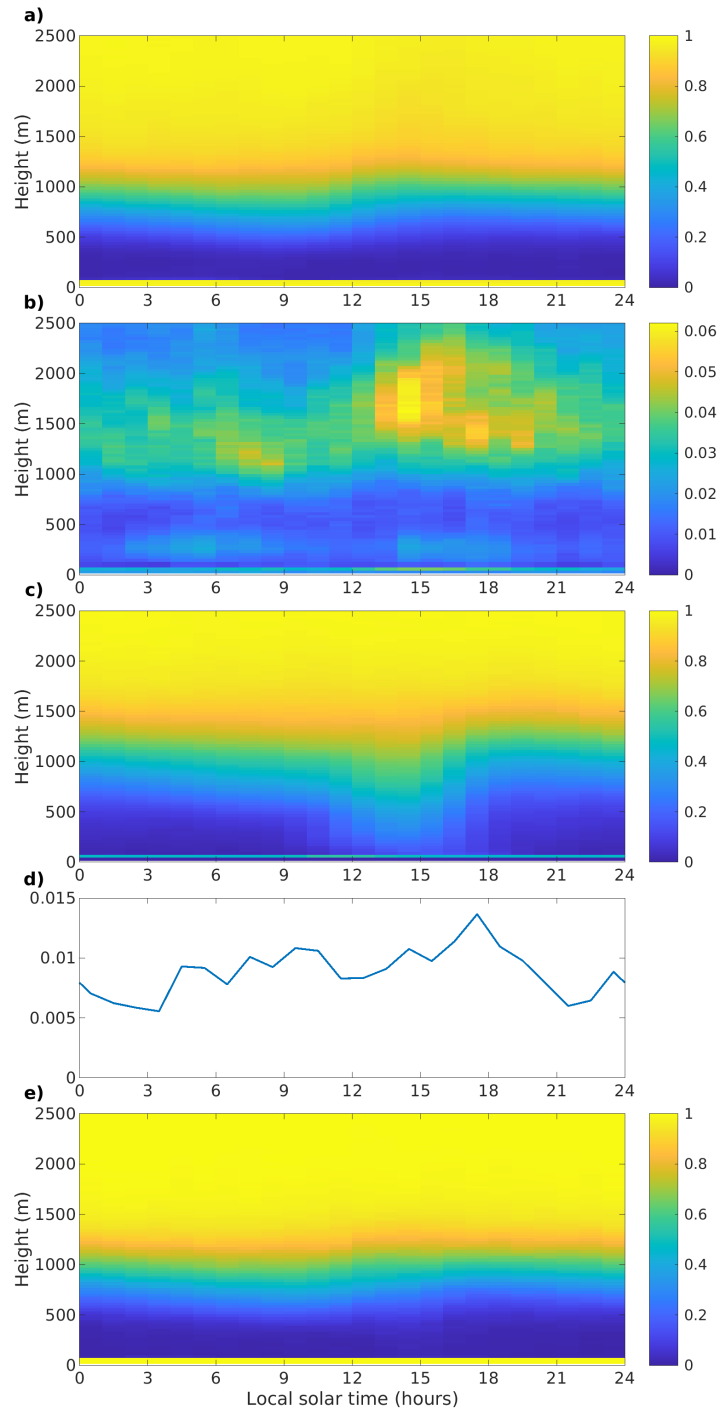
Figure 2. Fraction of Doppler lidar winds filtered at Granada by a) SNR filter, b) column-wise threshold filter, c) residual filter, d) rain filter, e) all filters combined. Note the change in colour scale for b), since this filter is applied after the SNR filter. The rain filter is applied to the whole column. For panels a)-d) the fraction filtered refers to the winds at nominal temporal resolution (10-15 min), and for e) to the 1-hour averaged data.
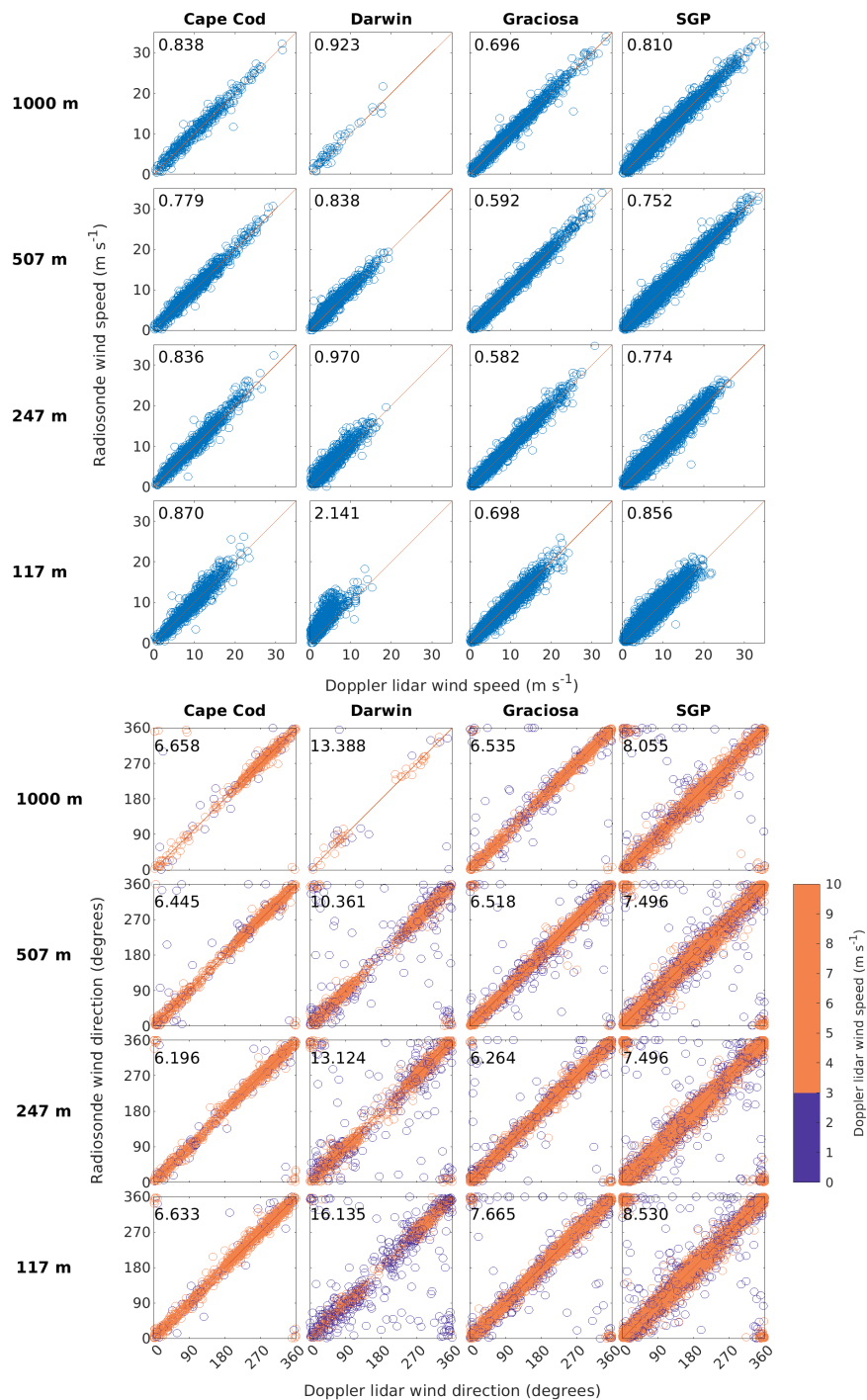
Figure 3. Scatterplots of Doppler lidar versus radiosonde wind speed and direction at four altitudes above ground level and at four locations. The mean absolute difference is also shown in each scatterplot. The wind direction is colour coded in two classes with respect to the wind speed given by the Doppler lidar, denoting whether the wind speed was greater or less than 3 m s^{-1}.
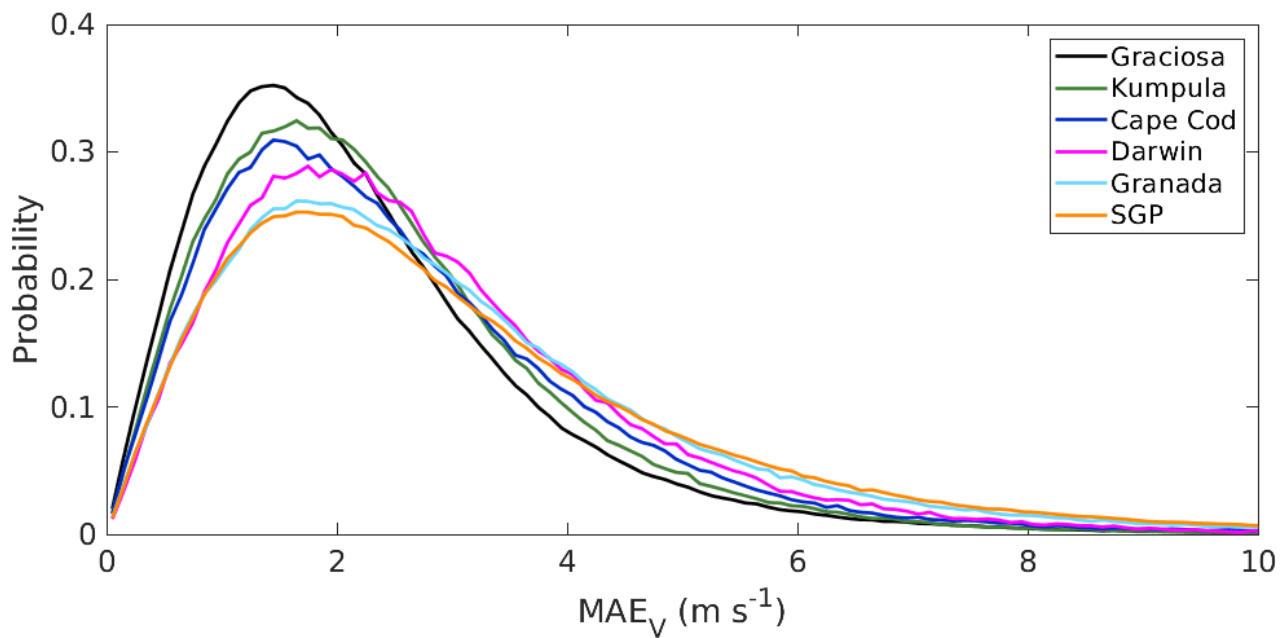
Figure 4. Probability density function of the absolute wind vector error at each station.
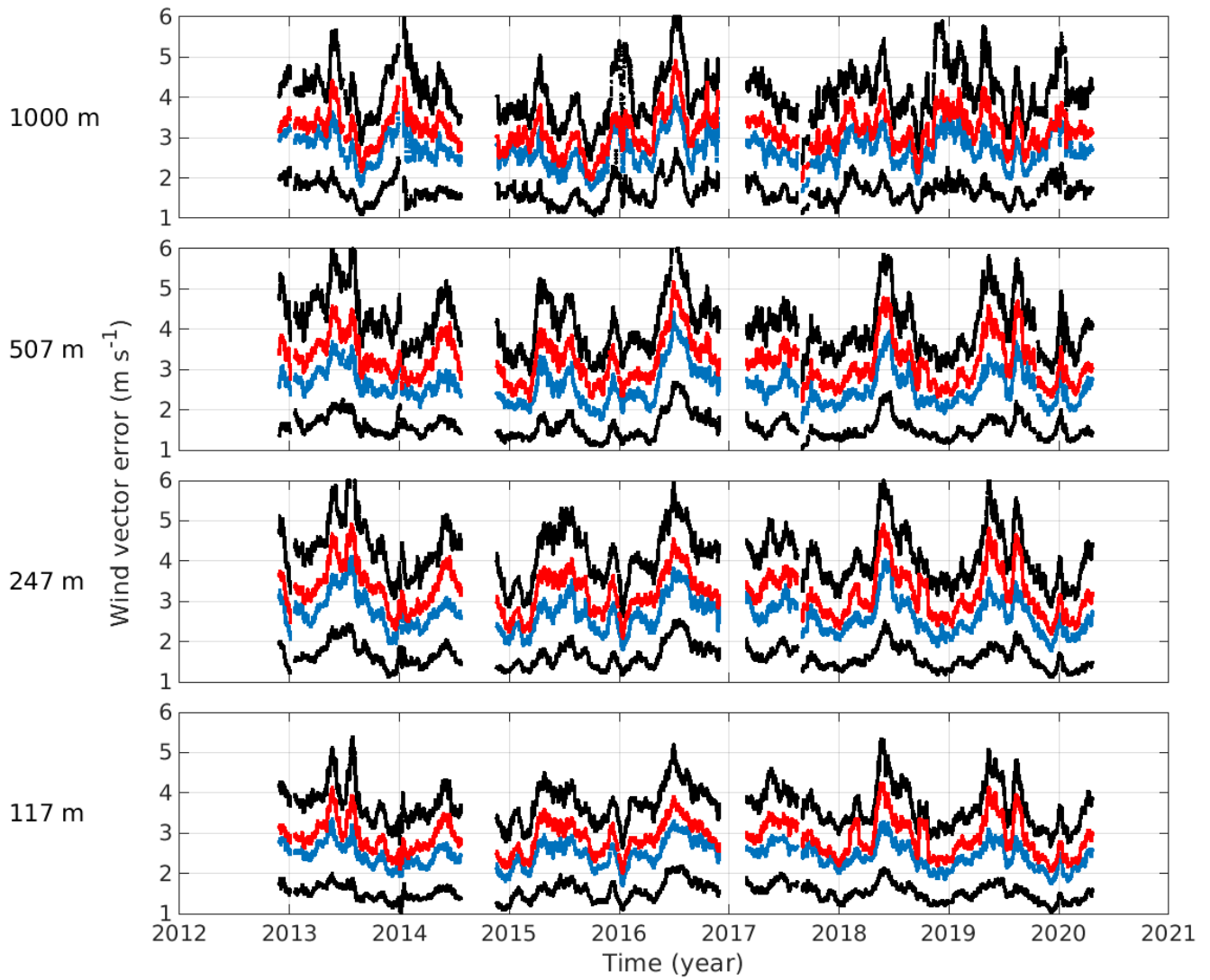
Figure 5. Timeseries of 30-day moving mean (red), median (blue) and 25th and 75th percentiles (black) of wind vector error over SGP at 117 m agl, 247 m agl, 507 m agl, and 1000 m agl.
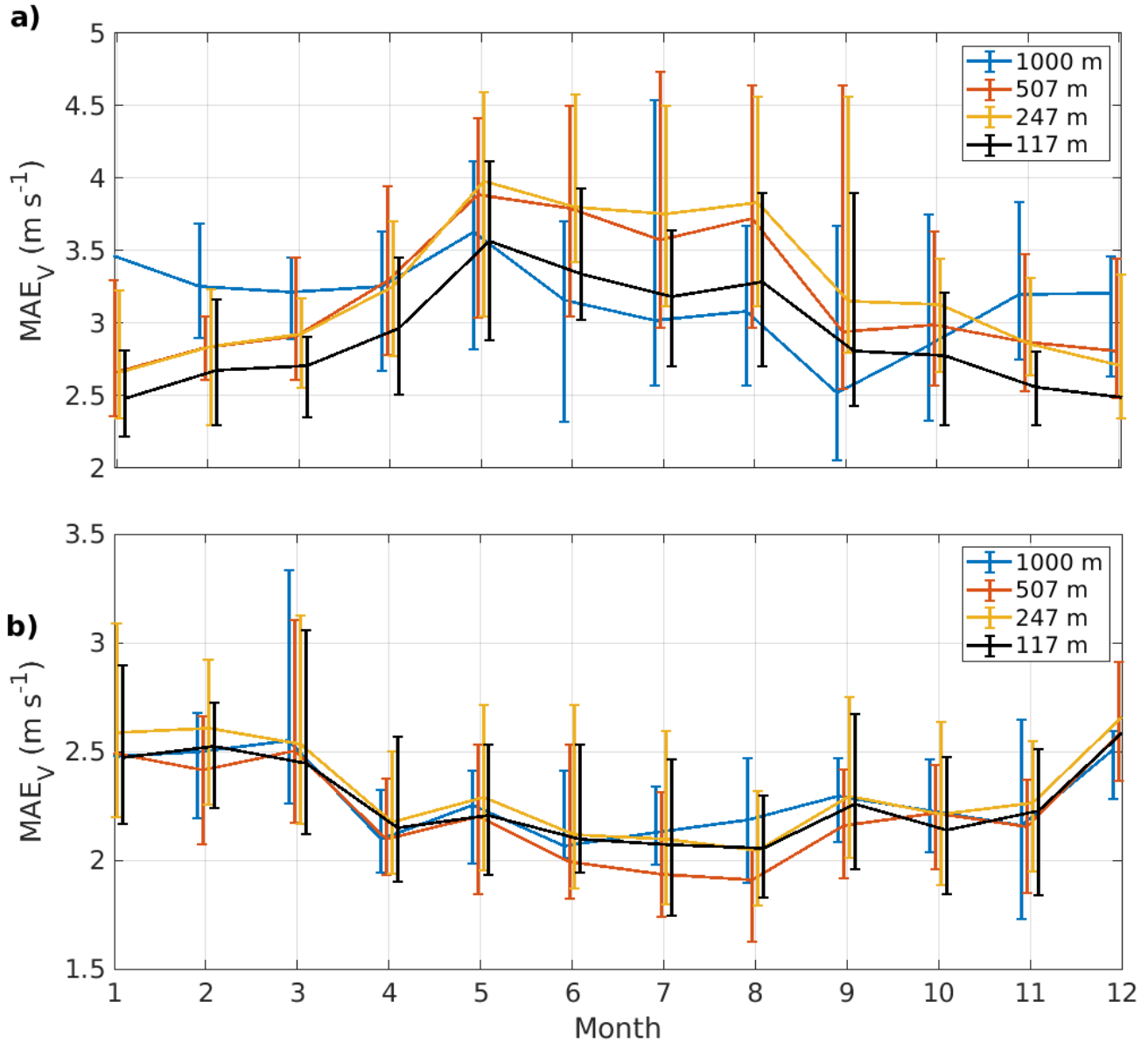
Figure 6. Monthly means of absolute wind vector error at a) SGP and b) Graciosa at 117 m agl, 247 m agl, 507 m agl, and 1000 m agl. The bars show the maximum and minimum monthly average absolute wind vector error across all years.
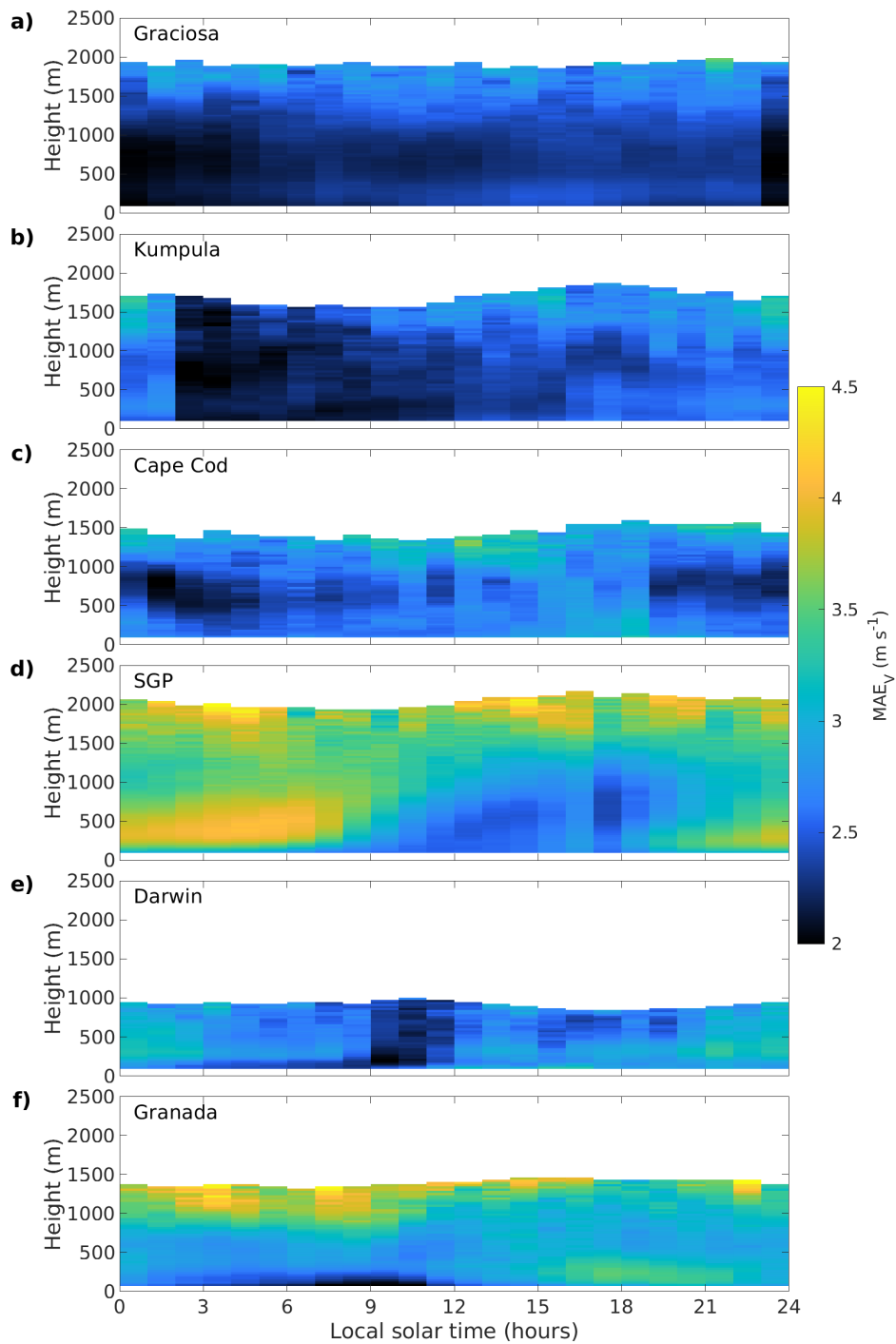
Figure 7. Diurnal composites of mean absolute wind vector error at a) Graciosa, b) Kumpula, c) Cape Cod, d) SGP, e) Darwin, and f) Granada. Time refers to approximate local solar time.
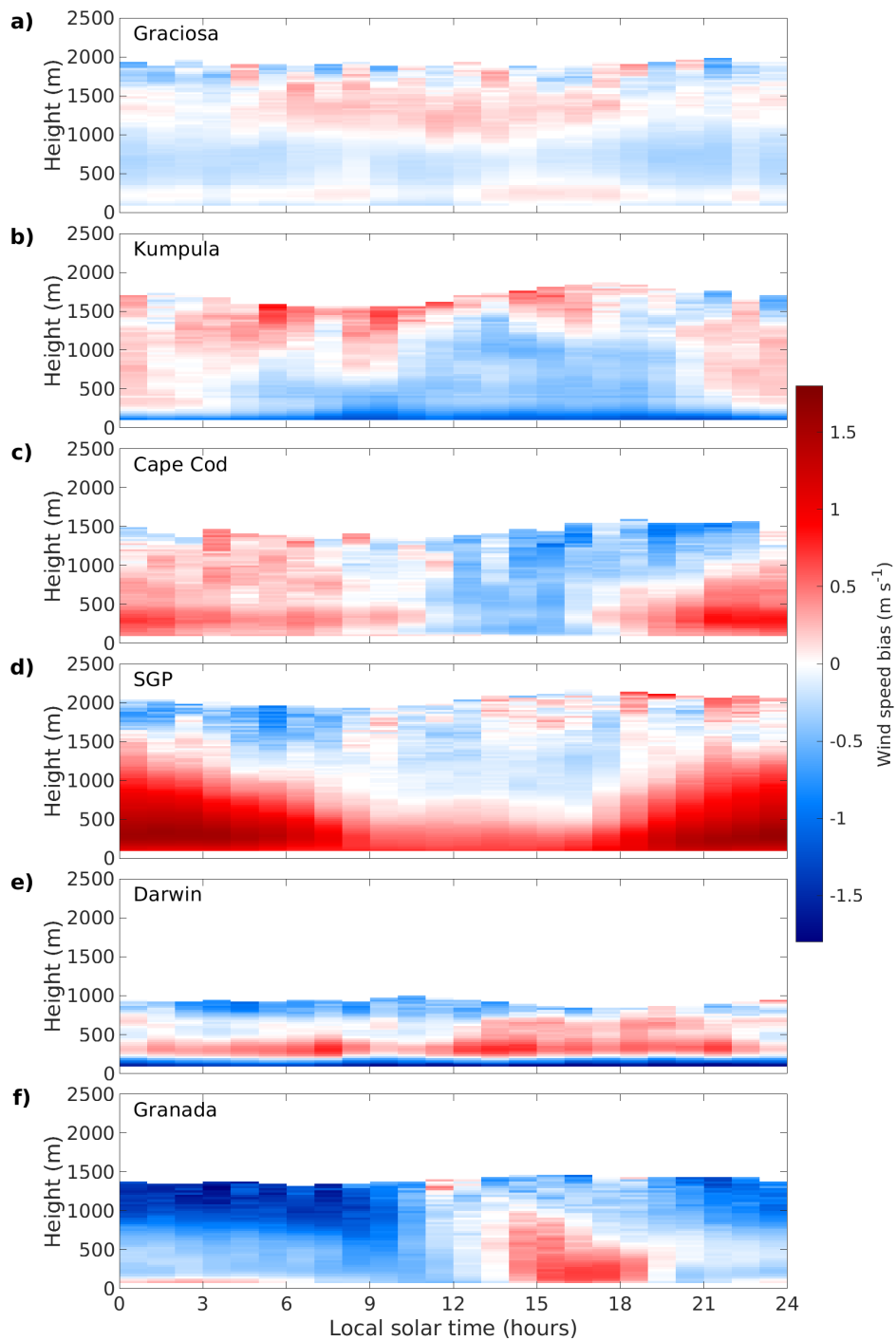
Figure 8. Diurnal composites of wind speed bias at a) Graciosa, b) Kumpula, c) Cape Cod, d) SGP, e) Darwin, and f) Granada. Time refers to approximate local solar time.

26
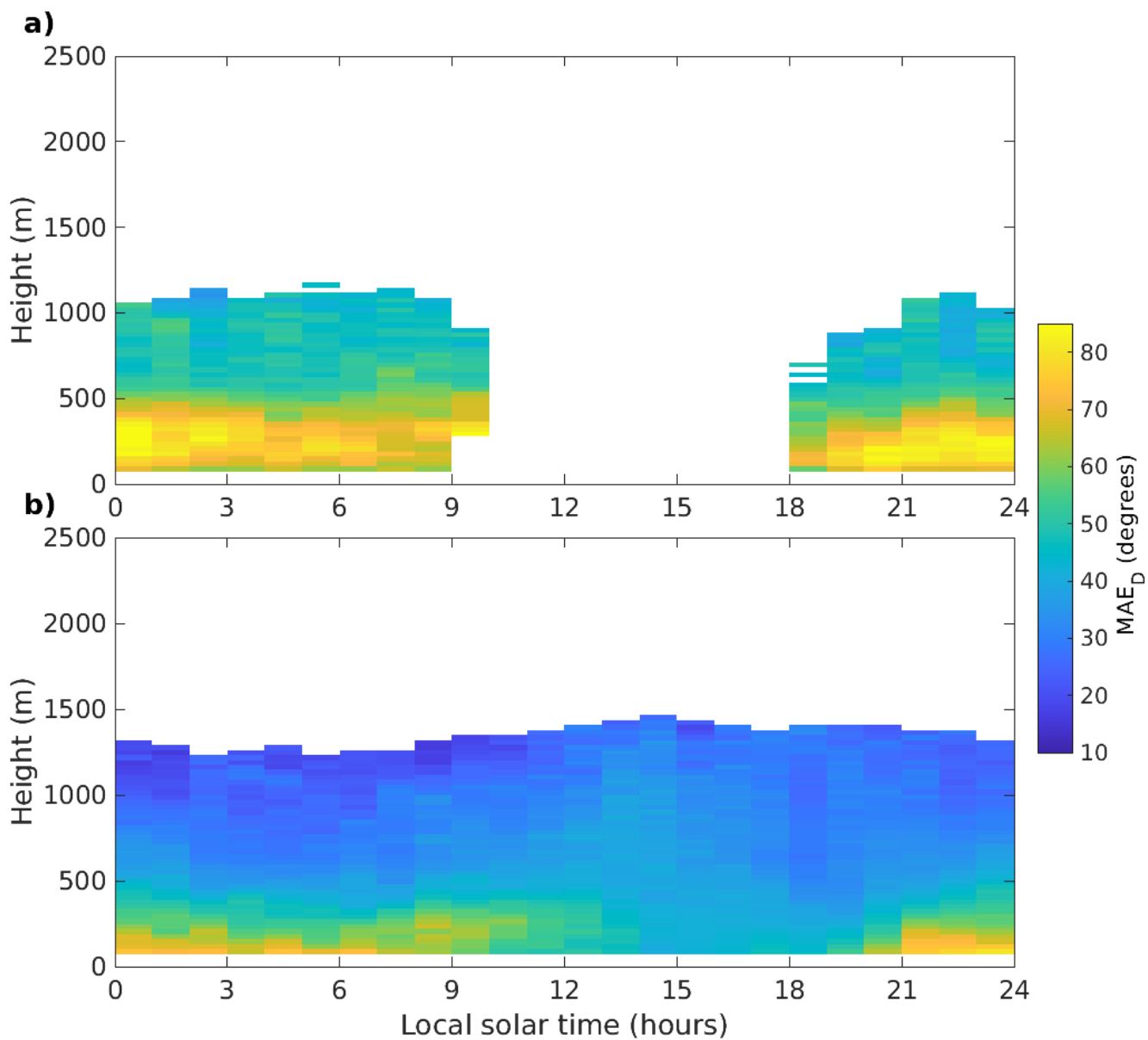
Figure 9. Diurnal composites of mean absolute wind direction error in Granada when a mountain breeze is a) detected, b) not detected, in the Doppler lidar data. Time refers to approximate local solar time.

a)

b)
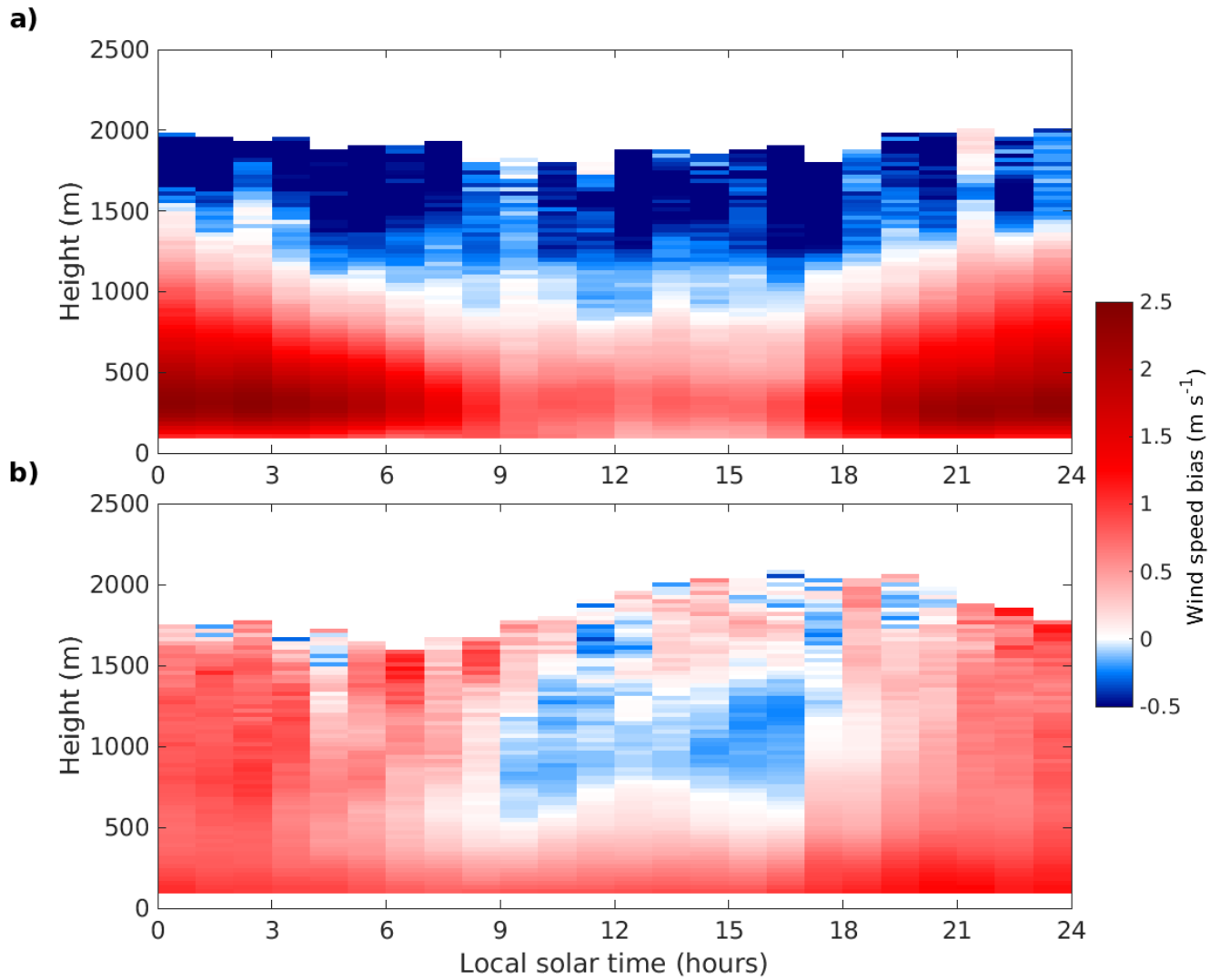
Figure 10. Composites of wind speed bias at SGP for when a) LLJ detected, b) LLJ not detected, in the Doppler lidar data. Time refers to approximate local solar time.
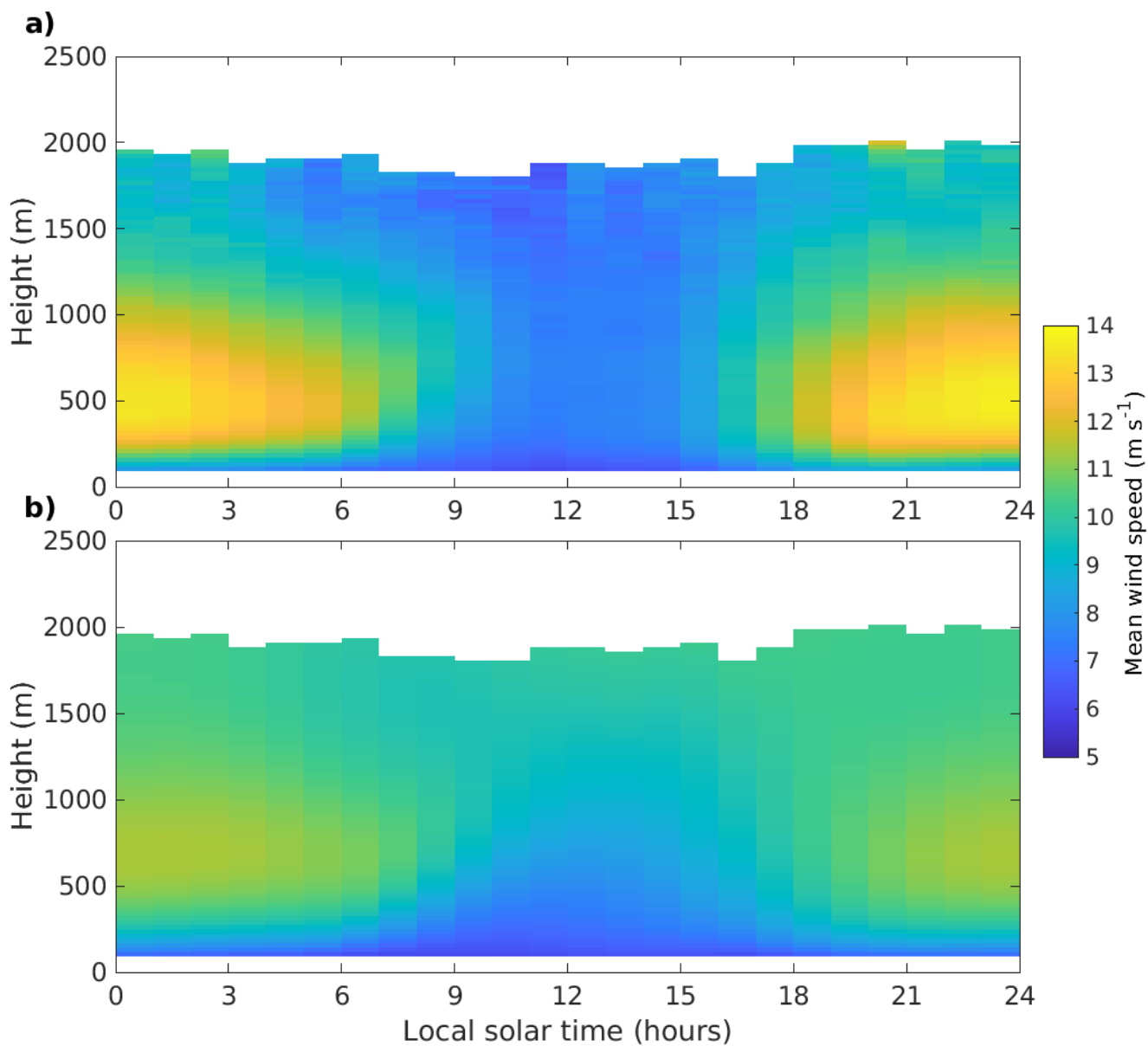
Figure 11. Composites of a) Doppler lidar and b) IFS mean wind speed at SGP when low level jets were detected in the Doppler lidar data. Time refers to approximate local solar time.
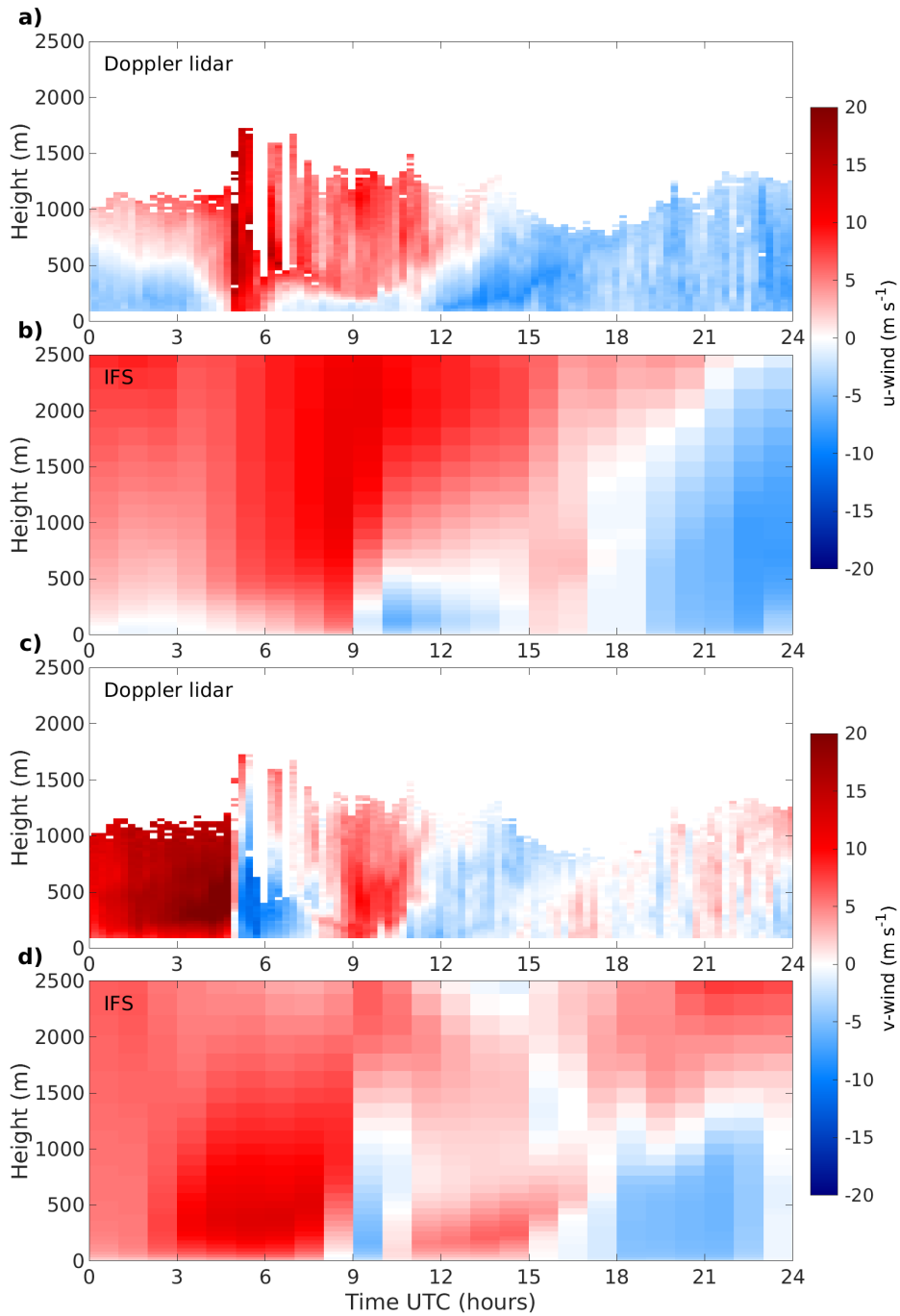
Figure 12. a) Doppler lidar u wind, b) IFS u wind, c) Doppler lidar v wind, d) IFS v wind, at SGP on 7 August 2013.
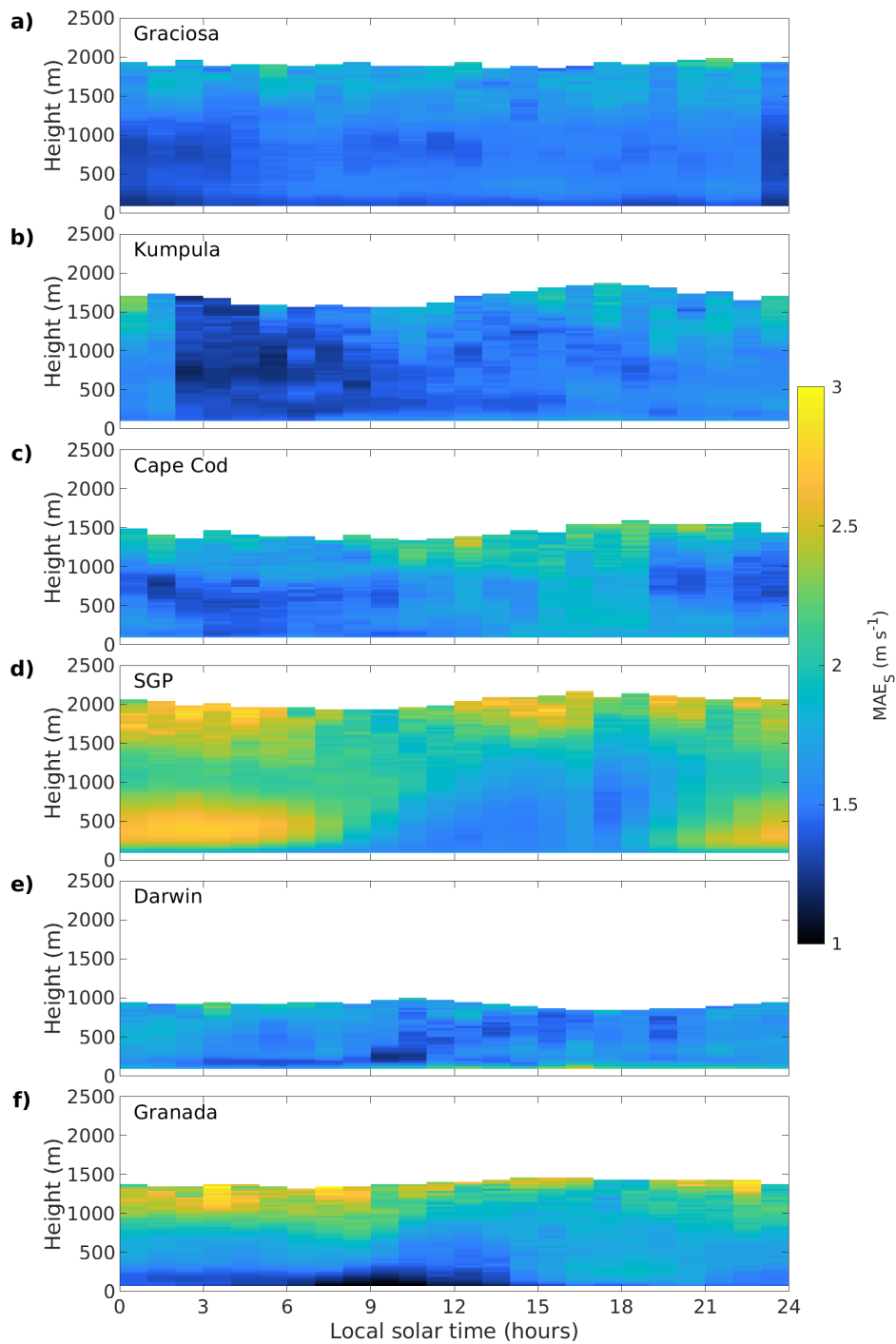
Figure A1. Diurnal composites of mean absolute wind speed error at a) Graciosa, b) Kumpula, c) Cape Cod, d) SGP, e) Darwin, and f) Granada. Time refers to approximate local solar time.
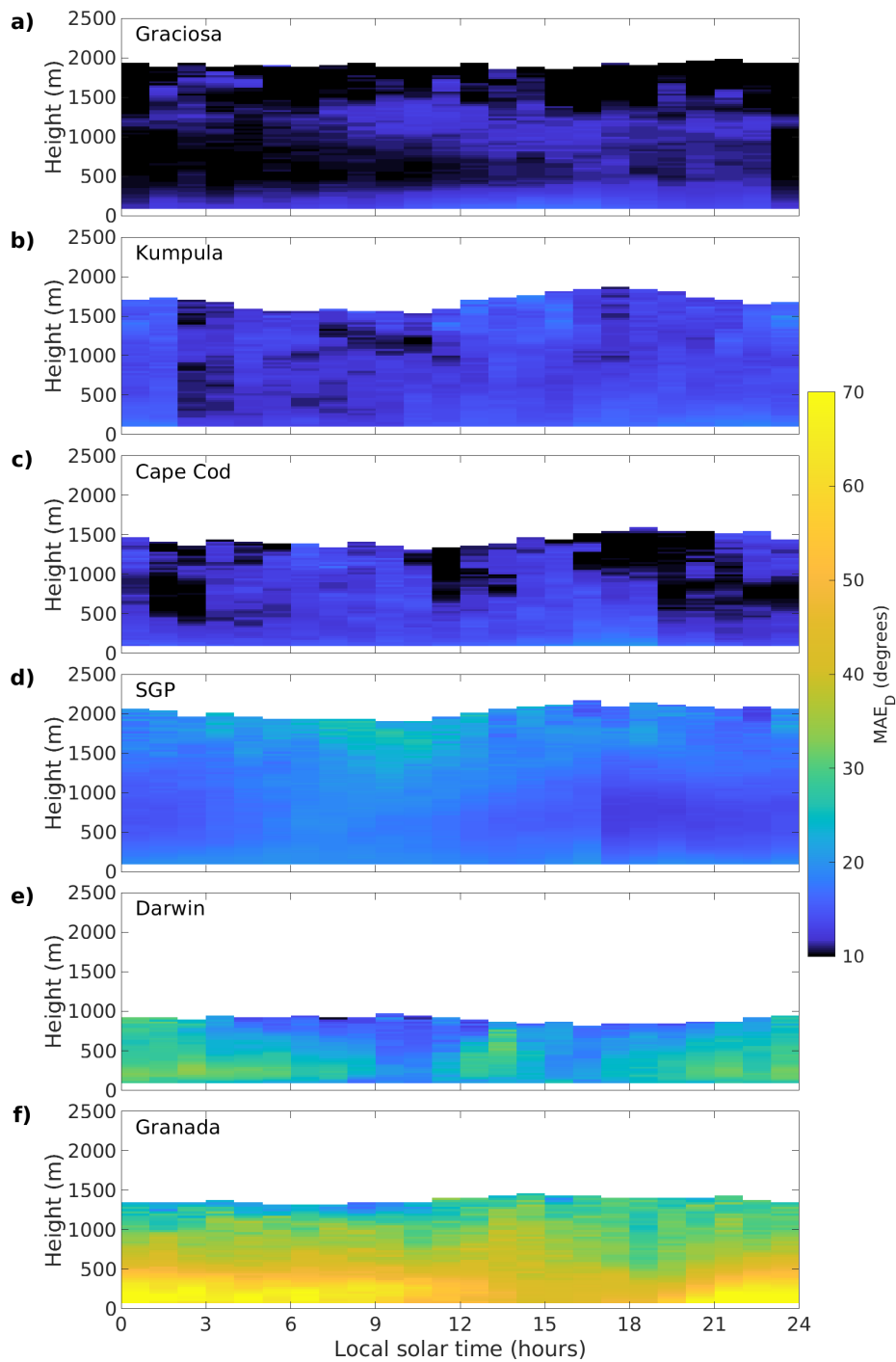
Figure A2. Diurnal composites of mean absolute wind direction error at a) Graciosa, b) Kumpula, c) Cape Cod, d) SGP, e) Darwin, and f) Granada. Time refers to approximate local solar time.
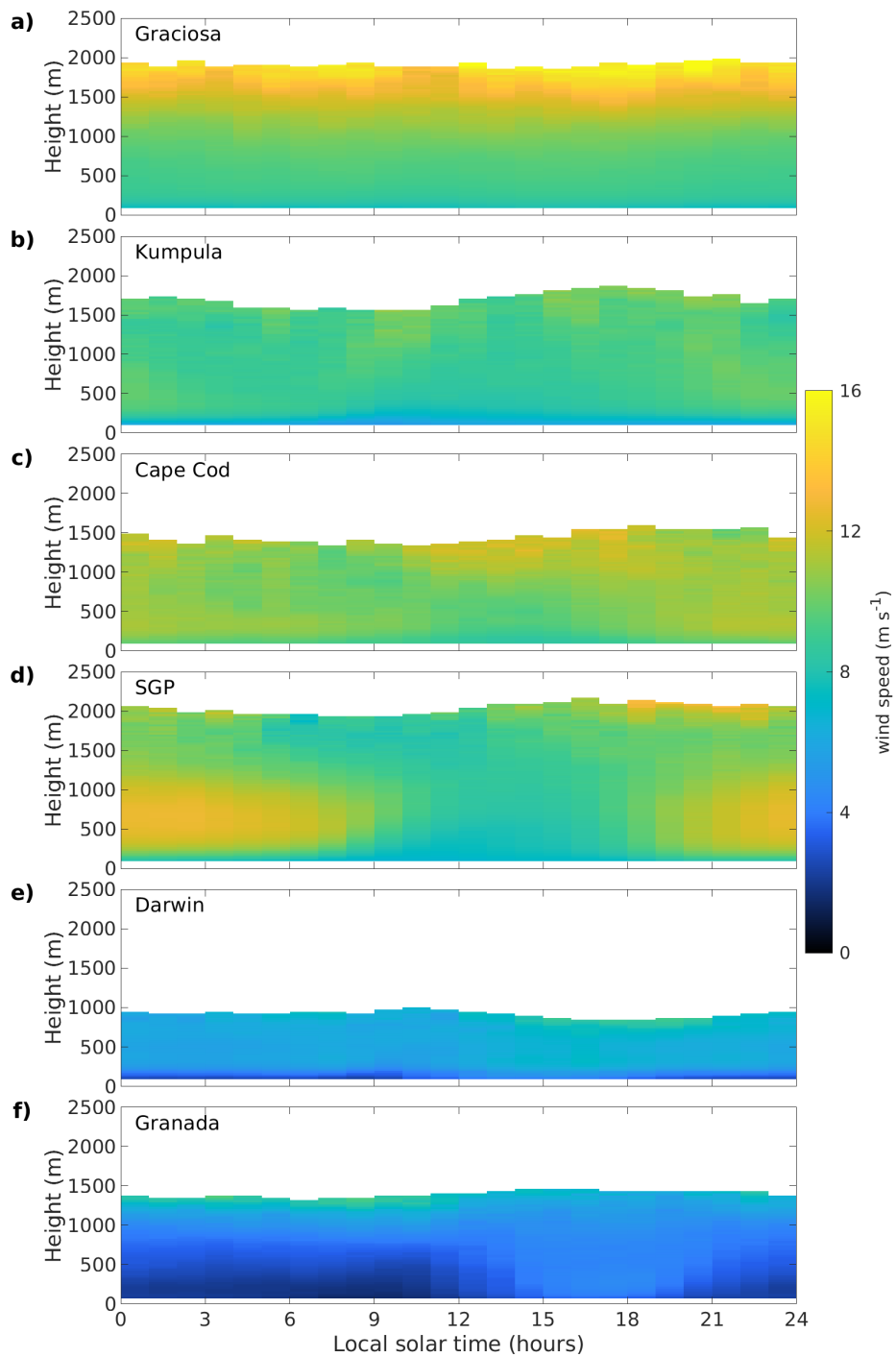
Figure A3. Diurnal composites of mean wind speed measured by the Doppler lidar at a) Graciosa, b) Kumpula, c) Cape Cod, d) SGP, e) Darwin, and f) Granada. Time refers to approximate local solar time.
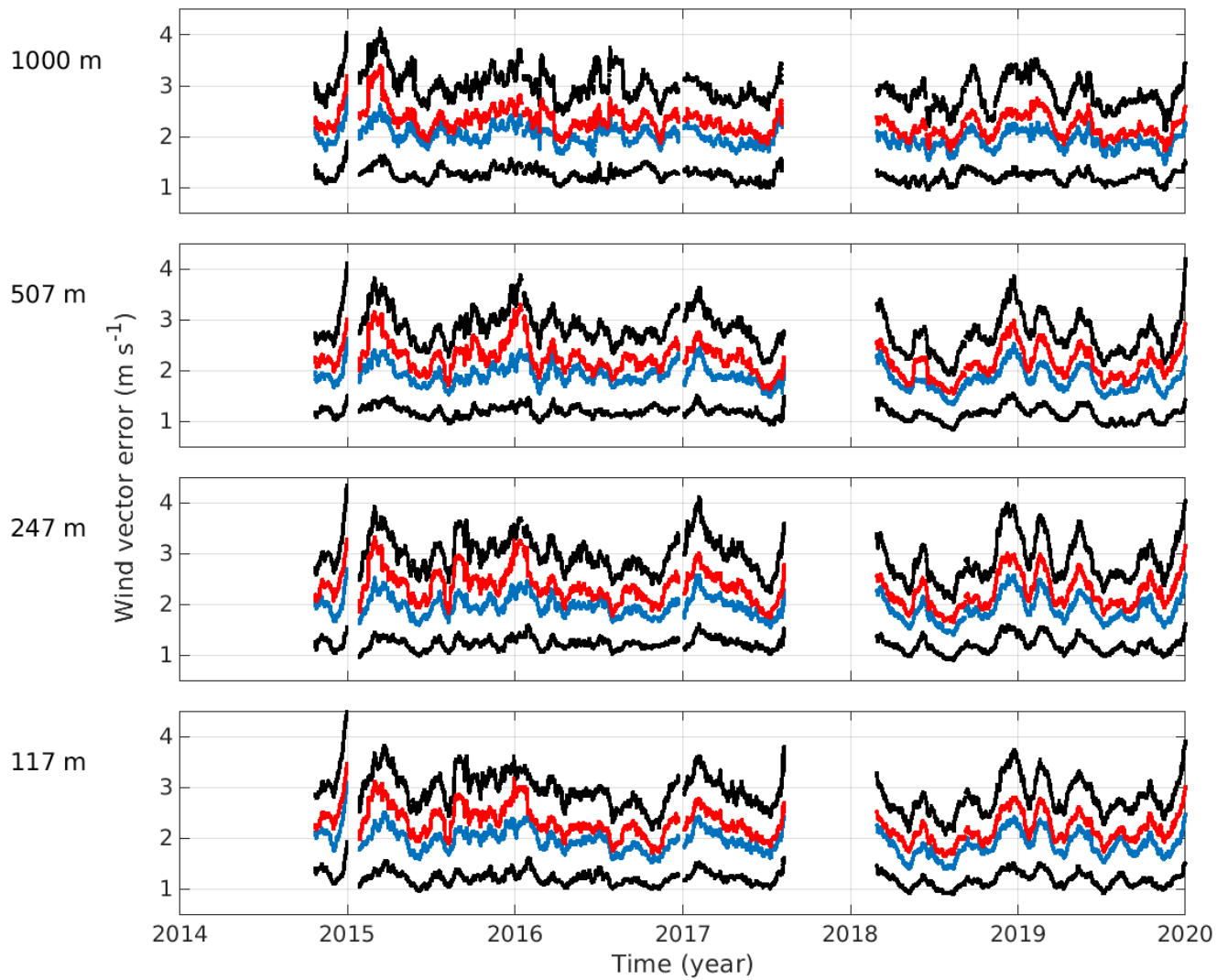
Figure A4. Timeseries of 30-day moving mean (red), median (blue) and 25th and 75th percentiles (black) of wind vector error over Graciosa at 117 m agl, 247 m agl, 507 m agl, and 1000 m agl.