Wind Error Assessment

Criado por Moritz Makowski, última alteração por Aydin Uzun em 01.Agosto 2022

For access to scripts and data locations, as well as a detailed readme for running the framework, see : GitLab

LiDAR Measurements

The scanning Doppler LiDAR is called Leosphere Windcube 200S and provides enhanced measurements of wind speed and wind direction. During the Hamburg campaign it was deployed under the 300 m mast in Hamburg and its exact location is latitude = 53.5193 and longitude = 10.1028. The observation dataset consists of **7170 measurements covering from 27/07/21 to 15/09/21**. In terms of temporal resolution, a measurement is available almost every 10 minutes with some exceptions. The available parameters are **Z** i.e. measurement height above ground level in m, **WSPD** i.e. horizontal wind speed in m/s, **WDIR** i.e. horizontal wind direction in deg, **UZ** i.e. vertical wind speed in m/s and **CNR** i.e. carrier to noise ratio in dB. For further analysis we are interested in **Z, WSPD and WDIR**. In terms of vertical resolution and coverage **Z** ranges from ~25 m to ~2900 m, roughly one sample per 15-25 m, and provides **149 elevation readings**, but please also note that some of these 149 elevation readings have WSPD and WDIR not defined, even though the Z value exists. This usually happens when readings are taken at very high altitudes. In short, this is a really fine resolution compared to what we have in ERA5 model data.

The available ERA5 model data is hourly data so the next step is to **convert these LiDAR measurements to hourly data**. We talked about the temporal resolution of LiDAR measurements. The measurements every 10 minutes do not seem to be taken at sharp hours [e.g. xx:55 xx:05, xx:15, xx:25, xx:35, xx:45 etc., xx is the hour of the day.]. Although it rarely happens, in some cases we have a sharp hour reading. In order to convert this data into hourly data, it was first checked whether there was a measurement in the sharp hour, and if there was a measurement, the representations of that hour in terms of Z,WSPD and WDIR were taken directly, but if there was no measurement in the sharp hour, the average of the two measurements nearest to the sharp hour was taken. The averaging of Z and WSPD arrays is non-trivial, the NaN values have been omitted from the calculation. To take the mean of WDIR arrays, first the the WDIR values are converted to sine and cosine components and the average of them is taken, and then they are converted back to WDIR values.

By doing this and after some reshaping operations I created **3 [for WSPD,WDIR and Z] 3D arrays of size [44x24x149]. 44 is the number of the days [from 01-Aug-2021 to 09-Sep-2021], 24 is the hours of the day and 149 is the available altitude values**. These parameters are stored as an hourly representation of the lidar observations.

ERA5 Model Data

It was decided to **combine hourly ERA5 data on pressure levels with hourly ERA5 data on single levels.**

Let's start with **ERA5 hourly data on pressure levels**. It's **gridded** data on a latitude-longitude grid, which means that after reading the latitude and longitude values, we need to find the latitude and longitude index that corresponds to the LiDAR location given at the beginning. Instead of Z [measurement height above ground level in m] from LiDAR measurements, we have here **37 pressure levels in millibars** related to the vertical resolution. The pressure levels are from 1000 hPa to 1 hPa. The temporal coverage consists of the **44 days of our interest [from 01-Aug 2021 to 09-Sep-2021]**. Because ERA5 provides estimates for a large number of atmospheric, ocean-wave and land-surface quantities, many variables available. The ones we are interested in are and save for later use are "**z**" in m^2 s^-2 i.e. Geopotential i.e. **gravitational potential energy of a unit mass at the locations of 37 pressure levels,** "**u**" in m s^-1 i.e. **U component of wind**, "**v**" in m s^-1 i.e. **V component of wind.**

Second, we need to import the **ERA5 hourly data on single levels**. Again, this is **gridded** data and first we need to find the latitude and longitude index that corresponds to the LiDAR location. The temporal coverage consists of the **44 days of our interest [from 01-Aug 2021 to 09-Sep-2021]**. Now the interest variables are **"u100", "v100", "u10", "v10" and "z"**. u100 is the horizontal speed of air moving towards the east, at a height of 100 meters above the surface of the Earth, in meters per second. v100 is the horizontal speed of air moving towards the north, at a height of 100 meters above the surface of the Earth, in meters per second. u10 and v10 are similarly used to describe the horizontal speed of air moving towards to the respective direction, at a height of 10 meters above the surface of the Earth in meters per second. **Z** is the Geopotential in m^2 s^-2 i.e. **the gravitational potential energy of a unit mass at the surface of the earth, relative to mean sea level.** This parameter does not vary in time.

For the **combination**, given the **ERA5 hourly data on pressure levels** we can **calculate the geopotential heights by dividing the geopotential "z" by the Earth's gravitational acceleration [g** (=9.80665 m s-2)]. Similarly for the **single levels** the **surface Geopotential height** can be calculated by dividing the Geopotential "z" available from the single levels dataset by the Earth's gravitational acceleration. If we **add 10m and 100m to this calculated Geopotential height value**, which is 24.52m, we can **calculate the corresponding Geopotential heights of u10,v10,u100 and v100**. In this way we have **increased the vertical resolution from 37 values to 39 values**. Next, the u and v vectors need to be sorted by their geopotential heights and the data is ready for comparison with the LiDAR. I similarly created **3 [for u,v and Geopotential heights] 3D arrays of size [44x24x39].** 44 is the number of the days, 24 is the hours of the day and 39 the available Geopotential heights. These parameters are stored as an hourly representation of the model data.

Model-Data mismatch (Windspeed and direction)

Next, since the vertical resolutions of the model data and the measured data are different and we want to make a comparison, they need to look similar, i.e. have the same vertical resolution. To fix the temporal resolution, the **range between 06:00 and 18:00** is chosen, since footprints are present at these times. Temporal coverage is the **40-day interval from August 1, 2021 to September 9, 2021**.

First, 13 particle release heights are defined, these are particle release heights from the footprints. Each of these **13 particle release heights** defines a layer and the layer boundaries are the average of the two consecutive layers. The LiDAR measurements and model data (WSPD and WDIR) are **interpolated** such that there are **10 linearly spaced points in each atmospheric layer, giving a total of 130 points.**

For the **LiDAR**, the readings are first **converted into their sine and cosine components** by using the available **raw data in terms of WSPD and WDIR**. Then the **interpolation** is performed **separately for the sine and cosine components** and later the **interpolated WSPD and WDIR values are calculated using these interpolated sine and cosine components**. The **representative of a layer** is a single value and can be easily calculated in terms of LiDAR's WSPD **simply by averaging that layer's interpolated WSPD values**. It's a little bit different for the representative of a layer in terms of WDIR. The **interpolated WDIR values are converted back to sine and cosine components on the unit circle**. **The average of these sine and cosine components on the unit circle is taken and then the angle is calculated as the representative for this laye**r. This operation is **performed for each atmospheric layer, for each hour of the day, and for each day**. This gives us as output the representation of the entire atmosphere during the campaign in terms of **WSPD and WDIR for LiDAR measurements** namely **2 3D arrays of size [40x13x13], 40 is the number of the days, 13 is the hours of the day [from 06:00 to 18:00], 13 is the number of atmospheric layers**. One is for wind speed, the other is for wind direction.

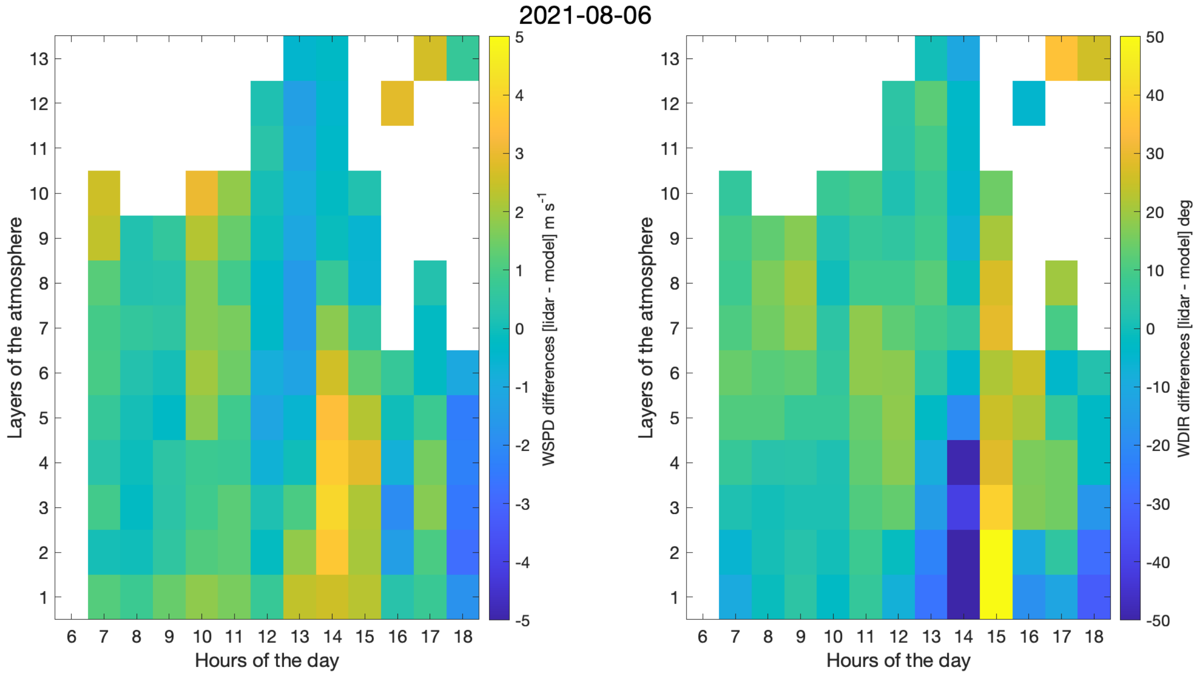
For the **model data**, we have the **u and v components**. The query points are the same as the for LiDAR measurements, namely 10 linearly spaced points in each atmospheric layer, 130 points in total. The **interpolation** is done **separately for u and v components**. Then **each layer is averaged using the interpolated points of u and v**. The **representative of a layer in terms of wind speed and wind direction can be calculated easily by using these average u and v values**. This gives us as output the representation of the whole atmosphere during the campaign in terms of WSPD and WDIR for the ERA5 model, namely **2 3D arrays of size [40x13x13], 40 is the number of the days, 13 is the hours of the day [from 06:00 to 18:00], 13 is the number of atmospheric layers**.

The **mismatches are defined as the distance between the LiDAR representation of the atmosphere and the ERA5 model representation of the atmosphere**, this is easily possible since the vertical and temporal resolutions and units are now the same.

Then I created images with scaled colors for each day to visualize the mismatches between the measurements and the model. The output images can be found at:

```
/Volumes/esm/11-Thesis/03-Scientific-Internship/2021 FP Aydin Uzun/Data/WSPD_WDIR_di
```

Let's investigate this plot representing the mismatches for **06-Aug-2021:**



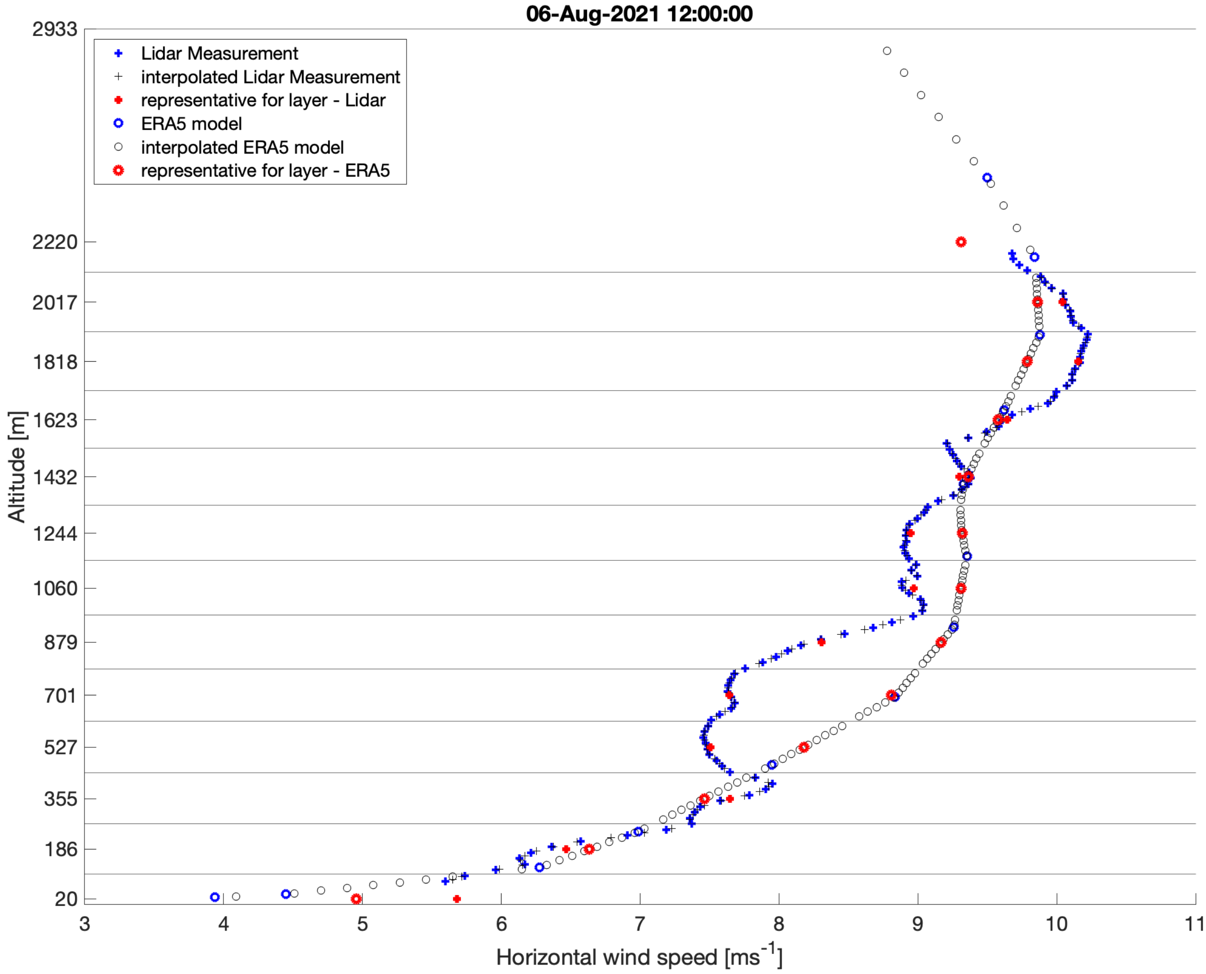
The x-axis shows the hours of the day, the y-axis shows the layers of the atmosphere. It’s a 13 by 13 square. The mismatches in wind speed and wind direction are displayed as an image with scaled colors.

Both the **LiDAR representations and ERA5 model representations [WSPD and WDIR] and the mismatches between them** are saved for later use and can be found at:

```
/Volumes/esm/11-Thesis/03-Scientific-Internship/2021 FP Aydin Uzun/Data/Lidar_ERA5_r
```

Let’s pick an hour from that day and investigate further: **06-Aug-2021 12:00:00**

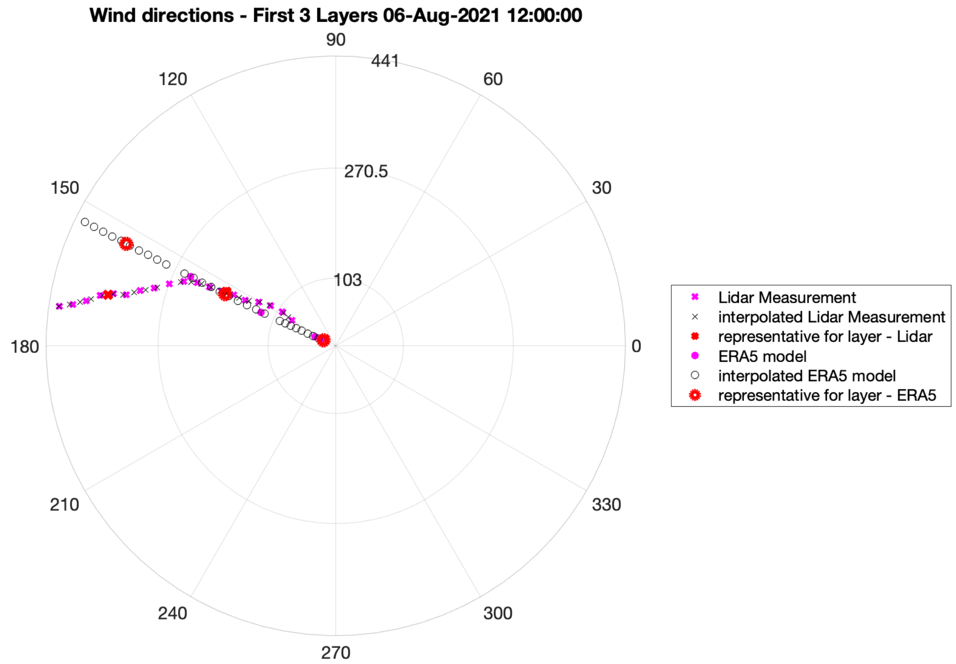
Let's start with the wind speed.

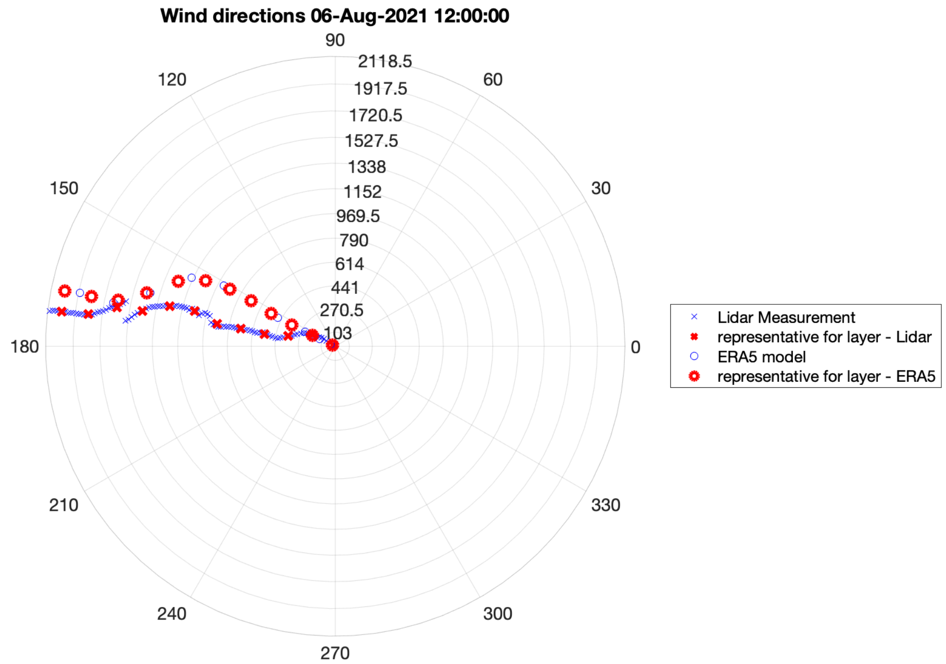


The x-axis is the horizontal wind speed in ms^-1 and y-axis is the altitude in m. The **ticks on the y-axis are the particle release heights**, the **horizontal lines are the boundaries** between the different atmospheric layers. The **circles** here stand for the **model data**, the **'+' signs** here are for the **measurements**. The color **blue** represents the **raw data**, the color **black** represents the **interpolated data**, the color **red** represents the r**epresentatives for each layer** i.e. the red circle in a layer is the mean of all black circles in that layer and the red "+" sign in a layer s the mean of all black '+' signs in that layer. The **distances between the red circles and red '+' signs will be the mismatch** in terms of wind speed. The WSPD plots that visualize the raw data, interpolated data and the representatives for the entire campaign can be found at:

```
/Volumes/esm/11-Thesis/03-Scientific-Internship/2021 FP Aydin Uzun/Data/WSPD_figures
```

Now look at the wind direction plot at the same date time:





The **degree** defines the **wind direction** and the **radius** the **altitude**. The upper plot is an **enlarged version of the bottom plot for the first 3 atmospheric layers**. The **nested circles are the boundaries of the atmospheric layers**. The **circles** here represent the **model data**, the **cross signs** here the **measurement data**. In the **upper** diagram, the color **magenta** stands for the **raw data**, the color **black** for the **interpolated data**, the color **red** for the **representatives**. For the **lower** diagram, the color **blue** stands for the **raw data**, the color **red** for the **representatives.** Again, the distances between the red signs will be the mismatch in terms of wind direction. The WDIR plots that visualize the raw data, interpolated data and the representatives for the entire campaign can be found at:

```
/Volumes/esm/11-Thesis/03-Scientific-Internship/2021 FP Aydin Uzun/Data/WDIR_figures
```

How and why to weight the differences?

We know that the receptors at higher altitudes are not sensitive to surface emissions. So mismatches in the higher atmospheric layers should actually be less important. We need to somehow **weight these mismatches** that we calculated earlier. The idea was to **use the footprint intensities and the vertical scaling factors as weights**.

Pressure scaling factors

Since there are 13 atmospheric layers, there are 13 different scaling factors for each day. The output is the 2D scaling matrix for the entire campaign and the daily scaling factors and can be found at:

```
/Volumes/esm/11-Thesis/03-Scientific-Internship/2021 FP Aydin Uzun/Data/Scaling_fact
```

Combine with footprint intensities to get weighting factors

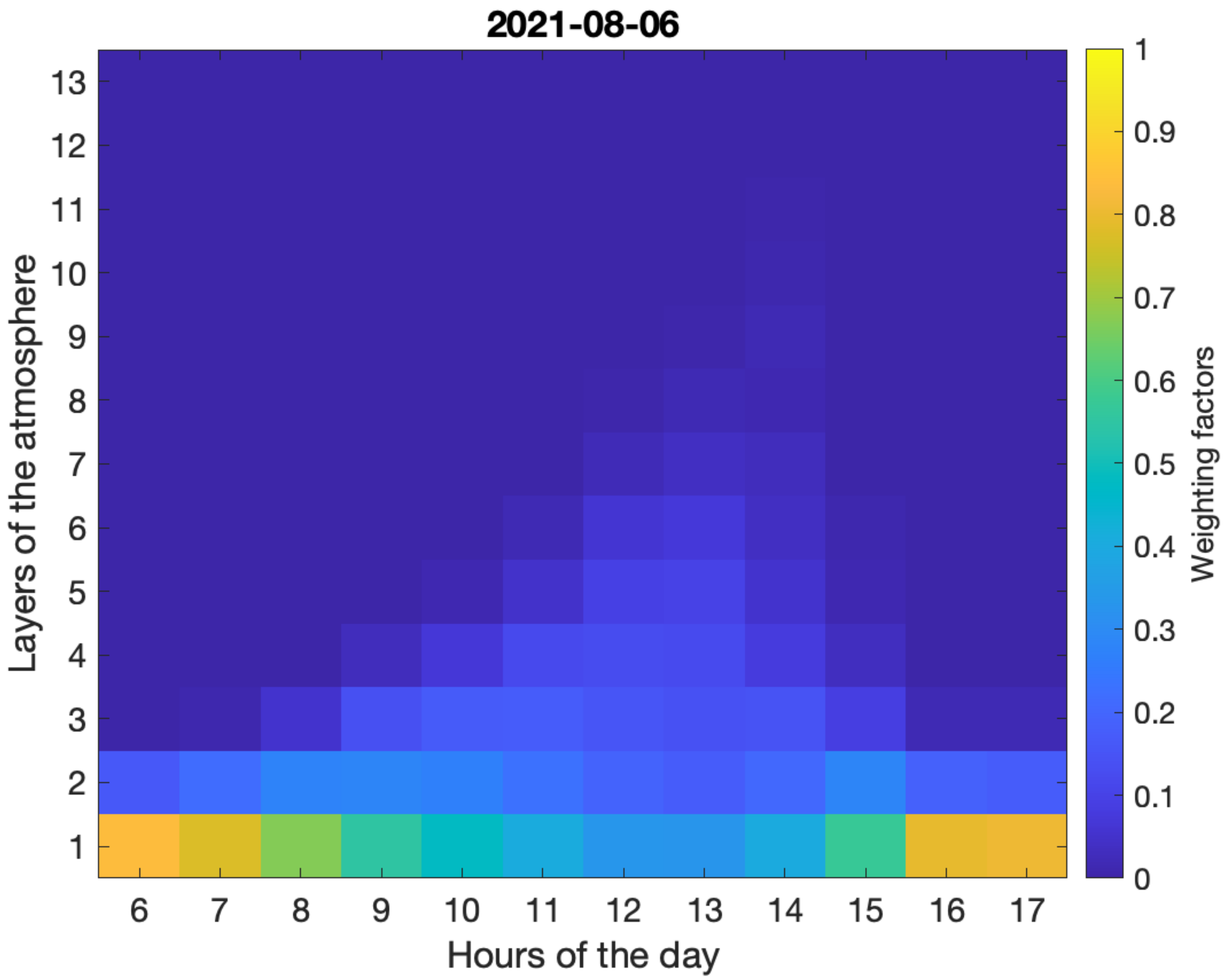
Footprints exist for each **instrument location [mb,mc,md,me], particle release height [13 different values], time series** triple. The **footprint intensities at sharp hours available** are calculated and output as a CSV document, with the first column corresponding to the date and time, the second column corresponding to the instrument's label, and the **third to last column corresponding to the factors for the 13 layers**. These factors are calculated by **multiplying the available scaling factor for the corresponding day by the footprint intensities (basically the sum of all elements of the 2D footprint image) on each layer.** The third column corresponds to the first layer, the fourth column corresponds to the second layer, and so on. By the way, the factors are normalized, so the sum in each row is 1. The output CSV document can be found at:

```
/Volumes/esm/11-Thesis/03-Scientific-Internship/2021 FP Aydin Uzun/Data/HAM_large_fc
```

The derived weighting factors are **averaged across instruments (mb,mc,md,me)** and normalized again and the **weighting factors are displayed as images with scaled colors**, similar to what we did before.

The output images can be found at:

```
Volumes/esm/11-Thesis/03-Scientific-Internship/2021 FP Aydin Uzun/Data/Weighting fac
```



The weighting factors for the selected day (06-Aug-2021) can be found here. Here you can clearly see that the sum of a column in this image is 1. The footprint intensities change depending on the planetary boundary layer.

Weighting the the mismatches

Using the weighting factors previously created and visualised, we can **weight the mismatches between the LiDAR representatives and ERA5 model representatives for both WSPD and WDIR** in order to **find the mean and standard deviation values for the daily mismatch.** This must be done for both WSPD and WDIR. The built-in functions mean and std are used with the weights option and omitting the NaN values.

The output document contains the mean WSPD, WDIR mismatches and standard deviation of WSPD, WDIR mismatches for each day and can be found at:

```
/Volumes/esm/11-Thesis/03-Scientific-Internship/2021 FP Aydin Uzun/Data/Weighted_dai
```

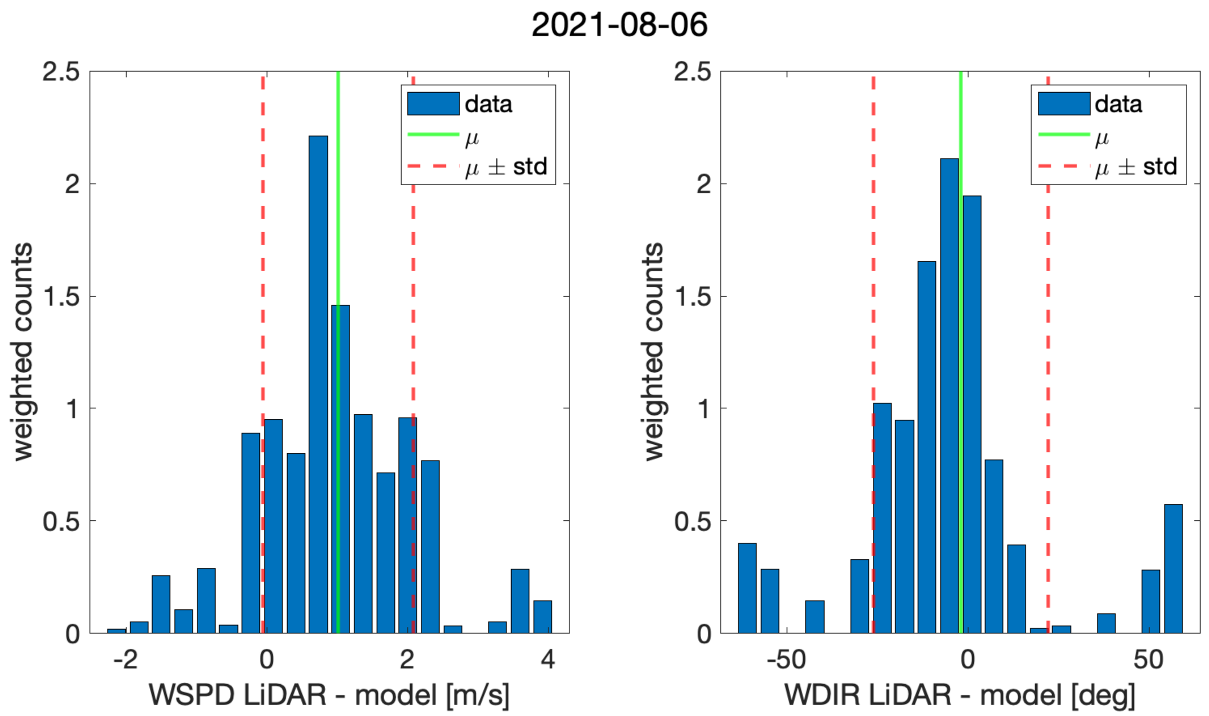
The mean and standard deviation values for some chosen days, which we will use later, can be seen in the table below.

	mean WSPD mismatch [m/s]	std WSPD mismatch [m/s]	mean WDIR mismatch [deg]	std WDIR mismatch [deg]
06-Aug-2021	1.01	1.07	-2	24.11
11-Aug-2021	-0.05	0.91	9.86	20.25

12-Aug-2021	-0.06	0.57	-4.48	19.74
23-Aug-2021	0.7	0.66	5.97	6.47
24-Aug-2021	0.13	0.53	12.72	11.35
31-Aug-2021	0.05	0.7	14.55	9.96
01-Sep-2021	1.07	0.52	-2.52	12.56
03-Sep-2021	1.22	0.51	6.35	8.15
05-Sep-2021	0.3	0.57	12.85	12.9

I also have created **daily weighted histogram plots** using the beneficial function **histwc.m**, a MATLAB function written by Mehmet Suzen on MATLAB Central File Exchange to generate weighted histograms. In addition, the mean and standard deviation of the daily mismatches are highlighted in the plots.

This is how it looks for the day we have chosen:



The weighted histograms for the entire campaign can be found at:

```
/Volumes/esm/11-Thesis/03-Scientific-Internship/2021 FP Aydin Uzun/Data/Weighted his
```

Sem rótulos