06.09.2021

Review of the article

"Validation of Aeolus Level 2B wind products using wind profilers, ground-based Doppler wind lidars, and radiosondes in Japan"

submitted by H. Iwai et al.
(AMT)

Scientific significance: Excellent
The paper deals with validation of the Aeolus HLOS winds based on three different reference instruments (Windprofiler, Doppler wind lidars and radiosondes) during different time periods of the mission covering the two different Aeolus laser transmitters (FM-A and FM-B) as well as different data processor versions (2B02 and 2B10). The study is well structured and is an important contribution to the global Aeolus cal/val activities and fits especially well the Aeolus AMT special issue.

Scientific quality: Good
The paper addresses all information that are needed to understand the quality of the used reference measurements data as well as the methods that are used for comparison. Minor issues that could even improve the presented results are addressed below.

Presentation quality: Good
The paper manuscript is clearly structured, all used methodologies are well explained and all figures are clearly visible. The text is well and clearly written. Only minor points are mentioned below that may help to further improve the manuscript.

Review Summary
The paper manuscript "Validation of Aeolus Level 2B wind products using wind profilers, ground-based Doppler wind lidars, and radiosondes in Japan" H. Iwai et al. deals with the validation of Aeolus horizontal line-of-sight (HLOS) winds during different time periods (Oct to Dec 2018, Jun to Dec 2019 and Apr to Oct 2020) based on three different reference data sources coming from windprofilers, two coherent Doppler wind lidars as well as from radiosondes launched from Japan. Hence, these comparisons validate data obtained with the two different Aeolus laser transmitters FM-A and FM-B as well as data based on two different processor versions 2B02 and 2B10. The paper manuscript is clearly structured, all used methodologies are well explained and all figures are clearly visible. The content of the paper is well suited for AMT and especially for the Aeolus special issue. It is recommended to accept the paper manuscript after addressing the minor points that are mentioned below.

Detailed comments

- **General:**
    - The entire analysis uses Aeolus data that is processed with two different processor versions 2B02 and 2B10. Could you somehow address the main differences that were implemented in the processor between version 2B02 and 2B10. This would help the reader to understand why one would expect really differences especially in the systematic error due to the implementation of the M1 mirror temperature correction or the hot pixel correction that was implemented.
    - You compare your determined results with different studies (GPS-RS results with Martin et al.; CDWL results with Witschas et al., WPR results with Guo et al.). But actually, you can use the data of all three instruments as reference. Thus, I am missing a detailed comparison of your own results from the different measurement data of the WPR, the CDWL and the GPS-RS. This should also be available in the summary, maybe with a table that summarizes to retrieved systematic errors based on the respective reference data sets.
    - Sometimes there is a confusion with the "-" sign which sometimes denotes a minus and sometimes denotes a "to". Thus, it is suggested to replace the "-" sign with the word "to" and only use "-" as a minus sign.
    - If possible, I would not spread units over two lines.

- **Fig.1, caption:** bule → blue

- **Line 92:** "two interferometers"…actually, as you also write later in the text, there are three interferometers: one Fizeau interferometer, and two Fabry-Perot interferometers all of them illuminated sequentially.

- **Line 98:** "Fizeau and Fabry Perot as narrowband filters" Somehow this is true, but the bandwidth is still different by a factor of 5 (Fizeau interferometer about 2 GHz, Fabry-Perot interferometers about 11 GHz).

- **Line 99:** "After passing through some optics". You could skip that part of the sentence as it provides no information.

- **Line 106:** "Commissioning phase" which was from x.x to x.x?.

- **Line 110:** "Several different technical, instrumental, and retrieving checks account for this flag." → Could you give a few examples which checks are performed and considered for deriving the validity flag?

- **Line 126:** "There is no significant difference between wind profiler winds and radiosonde winds in the biases and root mean square errors. → What do you mean with "significant"? Is there a difference? If yes, it would be better to quantify here. How do you determine the bias of WPR measurements or rather radiosondes?

- **Line 126:** "294m" → 294 m (space)

- **Line 148:** "Doppler beam swinging (DBS) technique ". Is this a well-known technique? Could you add a reference here?

- **Line 149:** “The Doppler velocity spectra for all range bins of each beam were obtained 10,000 times on average. Since the PRF was 10 kHz, the accumulation time of each beam was 10 s”. If I understand it correctly, the averaging time is one second, or you accumulate 100000 spectra. Or do I misunderstand something here?

- **Line 152:** “Maximum likelihood estimator” Can you give a reference here?

- **Line 153:** “Bias was estimated at 0.02 m s–1 using measurements from a stationary hard target.“ → I guess, this value is only true for single LOS measurements, isn’t it? If yes, this should be clarified here.

- **Line 159:** “As mentioned earlier, we averaged Doppler velocity spectra for all range bins of each beam from 30 min before to 30 min after the passage of Aeolus, and then the vertical profiles of horizontal wind speed and wind direction were acquired by the DBS technique.” → Is it true that you averaged all spectra for 60 minutes and then you determine the mean wind speed? Using such an approach, I would expect the center peak in your power spectrum to be rather broad. Wouldn’t it be better to calculate the wind speed on a e.g. one-minute average, and then calculate the mean over the 60 data points? Furthermore, I do not understand why you calculate and average for your CDWL comparison. Couldn’t you actually just use the profile measured directly during the Aeolus overpass?

- **Line 229:** “Note that this estimation of σ includes the representativeness error due to the spatial and temporal mismatch between Aeolus and reference instruments’ measurements.” → The representativeness error is likely to be different for the different reference instrument measurements. For instance, for the CDWL measurements you could decrease the representativeness error by decreasing the temporal averaging time to only a few minutes. Have you tried if this changes the random error for the comparison?

- **Line 248:** “Furthermore, the range-gate settings of Aeolus were changed on 26 February 2019, which also increased the number of available data points during the baseline 2B10 period.” → What was changed in particular? More range gates in the troposphere, less in the stratosphere? Or just increasing the range bin size? Here it would be helpful to get more details.

- **Line 252:** “slightly positive”. As the bias is 1.6 to 1.8 m/s which is a factor of more than two larger the originally specified, I would skip the word “slightly” here.

- **Line 261:** “Comparison to A2D data”. → You should be careful when comparing to A2D data as they compare LOS winds and not HLOS winds.

- **Line 289:** It is worth mentioning here that only ascending orbits were underflown during the WindVal III campaign.

- **Line 311:** “The main reason for not yet achieving the mission requirement for random errors is probably related to the large representativeness error due to the large sampling volume of the WPR.” → Also the Aeolus laser pulse energy is remarkably less than specified. With the representativeness error you would argue that the actual random error of Aeolus L2B data could meet the requirements. However, this is not true.

- **Line 320:** "…there are relatively many paired data 320 points for comparison (Fig. 6a)." → What does relatively many mean? Can you please quantify. It would also be very helpful to have this information plotted in Fig.6/Fig.7. This would give the possibility to understand how likely the shown bias trend is. For instance, is the negative Mie bias for descending orbits in 10 km altitude real, or just a result of very few data points and thus not reliable.

- **Line328:** "The systematic error was less than that of the baseline 2B02." → Why? Here, you could refer to the differences in 2B02 and 2B10 which I suggested to discuss in the previous part of the manuscript.

- **Line 332:** "However, this result is different from that in the other validation studies conducted during the baseline 2B10 period (Guo et al., 2021)." → What is different? What is the result by Guo et al.? Would be good to write one sentence here such that this information is available without reading Guo et al., 2021.

- **Line 339:** "(descending) orbit, the minimum (maximum) bias is –1.93 (0.54) m s–1 in the altitude range of 5–6 (4–5) km." → this is confusing. Is the bias of -1.93 m/s corresponding to ascending or descending orbits?

- **Fig. 8:** "Monthly averages" → Why do you calculate monthly averages and do not show a time series on a daily basis?

- **Line 358:** "because the Mie return signal does not depend on the laser energy (Martin et al., 2021)." → I would not write it that strong. Indeed, compared to Rayleigh returns, Mie signals are much more depending on the atmospheric backscatter. But of course, if the laser pulse energy is too low, one would not be able to measure at all.

- **Line 369:** "During the baseline 2B02 period, the bias of Rayleigh-clear and WPR HLOS winds slightly increased as the scattering ratio increased (Fig. 10a)." → Would this be better visible when calculating daily means instead of monthly means?

- **Line 391:** "result is similar to that in the comparisons of Aeolus and WPR measurements." → which provides biases of x.x m/s. Would be good to repeat the numbers here.

- **Line 394:** "The values are smaller than the scaled MADs of Rayleigh-clear (Mie-cloudy) versus WPR winds." → Why do you think is this the case? The representativeness is similar, and the respective measurements errors of the WPR or rather the CDWL is corrected, isn't it?

- **Line 423:** "The Rayleigh-clear winds show good coverage and closely follow the shape of the wind profile at altitudes higher than 2 km." → Any ideas or hints what causes outliers as e.g. the one at 8 km for the Rayleigh-clear winds? Still aerosol contamination, as the range gates below provides a valid Mie wind? Would be interesting to see if there is an issue with the cross-talk correction of Mie signals in the Rayleigh channel...

- **Line 435:** "…the clouds were partly existent in the Aeolus observational domain" → This means that clouds are not sufficiently corrected or filtered out in the Rayleigh data product? E.g., the Rayleigh-clear wind in between 3-4 km where also clouds were partly present shows a quite large bias. Is there any way to analyze this altitude in more detail, e.g., by analyzing single Aeolus "measurements" instead of "observations", or is this information not contained in the L2B data product?

- **Line 440:** “There is a possibility regarding horizontal wind gradients in this height region.” → You have actual wind speed and direction available from the radiosonde, right? If yes, it might be good to show if there was indeed a wind shear in this altitude.

- **Line 477:** “Martin et al. (2021) estimated the radiosonde representativeness error, and error sources caused by spatial and temporal displacements need to be considered,” → How do they do that? Based on comparison to measurements or with respect to theoretical assumptions? This would be an interesting side note.