- During the mid-FM-B period (1 December 2019 to 31 January 2020), Aeolus L2B near real-time baseline products ‘2B07’ were used. It is revised from ‘2B06’ to 2B07 throughout the paper.
- Tables S1 to S3 are added. Table S1 shows the number of measurements (N) and number of profiles (p) for the validation at the sites, Table S2 for validation over the Canadian Arctic, and Table S3 for validation over the pan-Arctic.
- The following sentence is added in Line 62 for clarification: “We will focus on analyzing random errors instead of systematic errors since, as recommended for operational NWP practice, bias corrected Aeolus data is used in this study (see Sect. 2.1).”
- Belova et al.’s (2021) findings on the systematic and random errors are summarized in line 72: “In related Arctic-based work, Belova et al. (2021) have found consistency between Aeolus winds and a ground-based radar situated in northern Sweden with insignificant biases between the two products (less than 1 ms^{-1}) and slightly increased random errors for Aeolus in the boreal summer, possibly due to sunlight scatter.”
- The passage (line 104) is revised to
  “Prior to 5 March 2019, both Rayleigh and Mie winds were averaged to up to a horizontal resolution of 87 km. Recognizing that Mie scattering in cloudy air yields stronger returns than Rayleigh scattering in clear air, after 5 March 2019, the Mie wind product was provided at a finer horizontal resolution of 12 km.”.
- The data link is added in Line 114: “ECMWF has published the first reprocessed data in fall 2020 (2B10; available at ftp://2018_aeolus_l2b:ecmwf@acquisition.ecmwf.int/), which covers the period between 24 June and 31 December 2019.”
- The threshold values are added in the text (line 125):
  “The thresholds for L2B estimated observation errors during the FM-A period are 4.5 ms^{-1} for the Mie winds and 6.6 to 11 ms^{-1} for the Rayleigh winds, depending on the pressure level, and 5 ms^{-1} for the Mie winds and 8.5 to 12 ms^{-1} for the Rayleigh winds during the early FM-B period. For more details, please refer to Rennie and Isaksen (2020).”
- The following explanation is added in Line 130 to explain the screening threshold of 30 ms^{-1}: “This criterion was obtained from an initial comparison between Aeolus FM-A and ECCC-B (Laroche et al., 2019).”.
- The following passage is added in line 125:
  “The thresholds for L2B estimated observation errors during the FM-A period are 4.5 ms^{-1} for the Mie winds and 6.6 to 11 ms^{-1} for the Rayleigh winds, depending on the pressure level, and 5 ms^{-1} for the Mie winds and 8.5 to 12 ms^{-1} for the Rayleigh winds during the early FM-B period. For more details, please refer to Rennie and Isaksen (2020).”
- The following passage is added in line 143 for clarification:
  “The data used to compare with Aeolus winds in this paper is the assimilated data that is linearly interpolated to Aeolus measurement locations and times. For the linear interpolation between the model’s grid points, the horizontal grid-spacing is 15 km and the vertical grid-spacing varies from approximately 100 m in the PBL to 1 km in the stratosphere (McTaggart-Cowan et al., 2019). The linear interpolation in time is between two consecutive model states, 15 min apart.”
- This passage is added in line 163 for clarification of the process of linear interpolation of ECCC-B:

  “The data used to compare with Aeolus winds in this paper is the assimilated data that is linearly interpolated to Aeolus measurement locations and times. For the linear interpolation between the model’s grid points, the horizontal grid-spacing is 15 km and the vertical grid-spacing varies from approximately 100 m in the PBL to 1 km in the stratosphere (McTaggart-Cowan et al., 2019). The linear interpolation in time is between two consecutive model states, 15 min apart.”

- The raw radiosonde data is measured every 2 s, which results in a profile vertical resolution of 8-15 m. However, the data used from http://weather.uwyo.edu/upperair/sounding.html is the processed radiosonde data provided at standard pressure levels. It has a much coarser resolution than 15 m. The following passage is added in Line 191 for clarification:
  "Vaisala RS92 radiosondes (Mariani et al., 2018) were launched twice daily (45 minutes before synoptic times 00 and 12 Coordinated Universal Time (UTC)). They measure vector wind profiles with a vertical resolution of roughly 15 m depending on ascent speed, up to about 30 km above ground level. The data used (available at http://weather.uwyo.edu/upperair/sounding.html) is the processed radiosonde data provided at mandatory and significant pressure levels (which has a coarser resolution than 15 m). It takes about two hours to reach 30 km altitude (around 10 hPa). The instrumental uncertainty for the wind speed is between 0.4 and 1.0 ms^{-1} and between 0.3 and 0.7 ms^{-1} for the zonal wind component (Dirksen et al., 2014). The error on the zonal wind component due to drift and elapsed time of the ascending balloon is between 0.5 and 1.0 ms^{-1} in the troposphere and UTLS (see Fig.5b in Laroche and Sarrazin, 2013). As a result, the total error for the zonal wind component from these sources of errors is between 0.6 and 1.2 ms^{-1}. Note that the radiosonde data are assimilated in the ECCC and ECMWF systems, which means that the ECCC-B and ERA5 errors are not independent of the radiosonde observation errors. The ECCC Whitehorse site, situated in a wide valley with large lakes, also has radiosondes that operate similarly to the ones at the Iqaluit."
- A few more sentences are added in line 209 on the data quality for in-situ measurements:

  "The uncertainty of the measurements depends on conditions, SNR, and decibel of the return signal. The average vertical wind profile bias to radiosonde is better than 0.27 ms-1.".
- Section 2.5 (line 228) on the data matching process and coincidence criteria is revised for clarification. "For the ground-based validation, the criterion for coincidence of Aeolus overpasses is that the distance from the sites to the measurements be no more than 90 km (horizontal resolution of Rayleigh winds). Using this coincidence criterion, Aeolus overpasses are selected as targets for validation at Iqaluit three times a week at around 21:50, 11:15, and 22:00 UTC, and at Whitehorse twice a week at around 02:25 and 15:30 UTC. The Aeolus measurements are compared to the reanalysis and in-situ measurements that are available in the nearest time. Temporal sampling for each product is as follows: Aeolus overpasses at Iqaluit and Whitehorse are as mentioned above; reanalysis data is provided hourly, on the hour; radiosonde data is from launches at 00 and 12 UTC, with a two-hour time-of-flight to 30 km as mentioned above; Ka-band radar data is provided via 15-minute scans. For example, if Aeolus overpasses selected as a target for validation at the Iqaluit site at 11:15 UTC, since the reanalysis data is sampled hourly, the radiosondes are launched at 00 and 12 UTC, and the Ka-band radar at Iqaluit scans every 15 minutes, the Aeolus HLOS profile would be compared to the reanalysis data and radiosonde measurements at 12 UTC and to the nearest scan by the radar. On the other hand, if the overpass time is 02:25 UTC, the profile would be compared to the ERA5 data at 02 UTC, the radiosonde measurements at 00 UTC, and, again, the nearest scan by the radar.".
- Table 1 is added. It shows the adjusted r-squared and slope of the fitted line for the in-situ comparison.
- The paragraph in line 291 is revised to:
  "Overall, the datasets show strong consistency. ECCC-B and ERA5 are highly mutually consistent (Table 1; with adjusted r-squared greater than 0.97) and therefore show similar consistency with Aeolus (Figs. 3a-b and 4a-b). It can be seen that Aeolus Mie winds are less consistent with ECCC-B, ERA5, and radiosondes at Iqaluit than the corresponding observations at Whitehorse and for the Rayleigh winds. One possible reason for this relates to the fact that the

Mie channel samples winds in the lower atmosphere where winds are harder to assimilate or measure due to topography. Since Iqaluit is situated in tundra valleys with rocky outcrops that can cause increased variability in the wind field while Whitehorse is situated in large valleys with less wind variability due to topography, terrain effects might account for the difference in consistency. In addition, the overall range extent of the HLOS wind samples is between -25 to 25 ms^{-1} at Iqaluit and -45 to 45 ms^{-1} at Whitehouse and r-squared is sensitive to the range of data (note the denominator of the second term in Eq. (5)). Overall, Aeolus data show good agreement with these three datasets with adjusted r-squared greater than 0.8.".

- In order to assess significance, an F-test is performed, and all comparisons are at 99% confidence level, including the comparison between Aeolus and Ka-band radar at Iqaluit. We try to acknowledge more clearly the situation at line 311:

  "Generally, the sampling for these radar measurements is highly limited, which tends to reduce the agreement compared to the other datasets. Nevertheless, the agreement on the variances between Aeolus and the Ka-band radar is at 99% confidence level using F-test. This analysis highlights the importance for programs such as CAWS to continue to provide ground-based radar measurements to ensure independent measurements of the winds for future DWL missions."

- The 99% confidence level on the adjusted r-squared is added on the Fig. 5. The range of the adjusted r-squared for Mie winds is almost overlapping between the seasons. The following paragraph is added in Line 337 for clarification:

  "We also note a slight drop in consistency of the Mie winds for the mid-FM-B period, which took place in winter 2020: for instance, the adjusted r-squared and their 99% confidence intervals, between Mie winds and ECCC-B, are 0.92 ± 0.03 during fall 2018, 0.91 ± 0.01 during summer 2019, and 0.87 ± 0.02 during winter 2020. This decrease in the consistency is almost insignificant."

- The decomposition into different wind-component directions, shown in Figs. 7-8, provides insight into understanding the meteorological conditions that the measurements are sampling, which might be helpful to better understand the dynamical characteristics of this data in both Aeolus and other products. We have slightly modified the text to better explain this analysis (line 375):

  "Furthermore, some ascending and descending HLOS wind measurements cancel in the average owing to simply to the change of the angle of the LOS. To avoid this artefact and to add some insight into the wind features being measured, we also compare the projected HLOS wind vector into its zonal (positive to the east) and meridional (positive to the north) components. The distribution of the zonal-component of the HLOS winds is shown in Fig. 7e and g for Aeolus and ECCC-B HLOS winds. By doing this decomposition, the distributions for ascending and descending measurements are brought into better agreement (Fig. 7f). We also notice that the HLOS winds can provide some information about the vertical variation of the HLOS winds that are projected onto the zonal direction (Figs. 7e and g). For example, for Aeolus the projection of HLOS into the zonal direction for the stratosphere, UTLS, and troposphere are +11.00 ms^{-1}, +4.00 ms^{-1} and +1.00 ms^{-1} respectively for this measurement period and these values (and the standard deviations of their distributions, see the figure legend for values) agree very well with ECCC-B (and ERA5 – not shown). The distributions have mean values that are positive because the winds are mainly westerly over the Arctic in the winter."

- Instead of representing Aeolus, ECCC-B and ERA5 winds separately, Fig. 8 now shows the means and standard deviations of the differences between Aeolus and ECCC-B and ERA5. The means of the differences therefore reflect the remaining bias between the datasets after the dynamic bias correction has been applied. The associated paragraph describing Fig. 8 is also revised. Starting at line 394:
  "We compare the distributions of the differences between the Aeolus wind measurement data and the ECCC-B and ERA5 data during fall 2018, summer 2019, and winter 2020 over the Arctic, as summarized in Fig. 8, which shows the bias and standard deviations of the differences between Aeolus HLOS winds and the ECCC-B HLOS winds, and ERA5 HLOS winds, and their zonal and meridional projections. The measurements are decomposed into Rayleigh (red) and Mie winds (black). They are further decomposed into ascending (indicated with upright triangles) and descending (inverted triangles) measurements. The results, with the bias (the mean values of these differences for the different sampling used) being smaller than 0.7 ms^{-1}, are consistent with our bias correction method. The distributions of the differences in the ascending and descending measurements do not show a significant difference. The discrepancies in the meridional projections of the HLOS winds are smaller because Aeolus picks up mostly the zonal component of the winds due to the direction of the LOS."
- The sentence in line 411 is revised to:

  "Figure 9 shows that Aeolus data consistently has greater standard deviations than ECCC-B during all three periods and for both Rayleigh and Mie winds: its normalized standard deviations are typically within 1.05 to 1.40.".

- We corrected the radial pattern in Figs. 10-11 and Figs. S2-S3 by transforming our data to the EASE (Equal-area scalable earth) grid, described at the NSIDC website (https://nsidc.org/data/ease). The following explanations are added in Line 436:
  "Since the measurement density differs depending on the latitude, the RMSD of the profiles are calculated over nearly equal surface area, using the Equal-Area Scalable Earth (EASE) Grids (Brodzik et al., 2012). Each grid cell is around 104 km^2 which is approximately the square of the along-path resolution of Aeolus Rayleigh winds."
- The estimated errors are decreased in the reprocessed data. We wanted to make a point that the same improvement is not seen in the O-B statistics.
  The following sentences are added for clarification in Line 470:

  "The estimated observational errors have decreased compared to the 2B06 data (Figs. S1 and S2) since the bias due to the M1 mirror temperature dependence is updated on a daily basis and the dark current signals have been removed using improved quality control. However, we do not see the same improvement in the O-B statistics between 2B06 and 2B10 products over the Arctic region."

- The following sentences are added in the discussion section (Line 495):

  "In our analysis of the pan-Arctic region, we found an overall agreement by comparing the distributions of the HLOS winds, ascending and descending HLOS winds, and projections of HLOS winds onto east-west and north-south directions in different atmospheric layers (Fig. 7), and we also compared the distributions of the differences between Aeolus and ECCC-B and ERA5 (Fig. 8). Due to the angle of the HLOS, when comparing the distributions, separating the ascending with descending measurements helps avoid cancelling out part of the HLOS winds and projecting the HLOS winds on to zonal and meridional directions provides some insight on the vertical variation of the HLOS winds."

- Line 506 is revised from
  "We have found some initial evidence that the estimated error product is also a good predictor of RMSD between Aeolus and the reanalysis, which could be useful for constraining future forecasts."
  to
  "We found that the spatial variability of the time-averaged L2B estimated error product is in good agreement with the spatial variability of the RMSD between Aeolus and ECCC-B HLOS winds over the Arctic region. This validates the use of L2B estimated error product as a predictor for the HLOS wind observation errors in data assimilation systems, as proposed by Rennie et al. (2021), to obtain optimal positive impacts on forecasts from assimilating Aeolus winds."
- Some minor changes were introduced to improve readability and correct inconsistencies.