aeolus

DISC

Aeolus Data Innovation Science Cluster
DISC

Verification report of the third reprocessing campaign for first FM-A period from September 2018 till June 2019

DISC-Ref.: AED-TN-ECMWF-GEN-080

Issue: V 1.01

Date: 06/06/2023

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and Water Management

Document Change Log

Issue	Date	New pages	Modified pages (after introducing new pages)	Observations	Name
V 1.0	27 Feb. 2023	-	-	Version 1.0 prepared with input from DISC consortium	Saleh Abdalla, Thomas Flament, Isabell.Krisch, Uwe.Marksteiner, N. Masoumzadeh Oliver Reitebuch, Michael Rennie, Dimitri Trapon Fabian Weiler
V 1.01	06 Jun. 2023	-	-	Minor edits mainly in headers (font & page numbering).	Saleh Abdalla

Table of Contents

1 Introduction

1.1 Compliance Statement

The Verification report of the third reprocessing campaign for first FM-A period from September 2018 till June 2019 is fully compliant with the management requirements of the DISC contract.

1.2 Applicable Documents

[AD-1] DLR (2020): DISC Project Management Plan. AED-PMP-DLR-001, V 3.01, 18/09/2020.

1.3 Reference Documents

[RD-1] Abdalla, S., de Kloe, J., Flament, T., Krisch, I., Marksteiner, U., Reitebuch, O., Rennie, M., Weiler, F., Witschas, B. (2020), Verification report of first Reprocessing campaign for FM-B covering the time period 2019-06 to 2019-12, Aeolus DISC Technical Note, AED-TN-ECMWF-GEN-040, v1.

[RD-2] Abdalla, S., Flament, T., Krisch, I., Marksteiner, U., Reitebuch, O., Rennie, M., Trapon, D., Weiler, F. (2021), Verification report of the second repro- cessing campaign for FM-B from 24 June 2019 till 9 October 2020, Aeolus DISC Technical Note, AED-TN-ECMWF-GEN-060, v1.1.

[RD-3] Masoumzadeh, N. et al. (2022), Third Reprocessing campaign for FM-A covering the time period 2018-08 to 2019-06, Aeolus DISC Technical Note, AED-TN-DLR-GEN-061, v2.

[RD-4] Aeolus DISC (2021), The dark side of Aeolus during the IOCV phase. Aeolus DISC Technical Note, AE-TN-DLR-7300-2, issue 1.1, also available as a handover report file AE-TN-ESA-SY-094_Dark_Signal_Analysis_V1_1.pdf. Available from the internal DISC web site: https://csde.esa.int/confluence/display/AEOLUSDISC/dL1B_037?pre-view=%2F143105764%2F143105785%2FAE-TN-ESA-SY-094_Dark_Signal_Analysis_V1_1.pdf

[RD-5] Weiler, F., Kanitz, T., Wernham, D., Rennie, M., Huber, D., Schillinger, M., Saint-Pe, O., Bell, R., Parrinello, T., and Reitebuch, O. (2021), “Characterization of dark current signal measurements of the ACCDs used on board the Aeolus satellite”, Atmos. Meas. Tech., 14, 5153–5177, https://doi.org/10.5194/amt-14-5153-2021.

[RD-6] Weiler, F. (2020), "Detecting hot pixel induced steps in the Aeolus atmospheric signals", Master thesis, Department of Mathematics, University of Innsbruck, 85pp.

[RD-7] Weiler, F. (2021), "Detecting hot pixel induced steps in Aeolus atmospheric signals”, Aeolus DISC Technical Note, AED-TN-L1B-GEN-064, v 1.0.

[RD-8] Weiler, F. (2022), "Reprocessing – Hot Pixel Correction”, Presentation during WM 177 on 12 Jan. 2022, PowerPoint presentation file: E2_WM177_FW_Reprocessing_hot_pixel_correction.pptx

[RD-9] Weiler, F. (2022), "Aeolus 3rd reprocessing FM-A period - DCMZ generation and validation”, Aeolus DISC analysis report in a form of PowerPoint presentation file: FM_A_reprocessing_AUX_DCMZ_validation_FW_25_01_2022.pptx.

[RD-10] Marseille, G.-J., de Kloe, J., Marksteiner, U., Reitebuch, O., Rennie, M. & deHaan, S. (2022), NWP calibration applied to AeolusMie channel winds. Quarterly Journal of the Royal Meteorological Society, 148(743), 1020–1034. Available from: https://doi.org/10.1002/qj.4244

1.4 Acronyms & Abbreviations

An up-to-date list of abbreviations used within DISC and in this document can be found in the Aeolus DISC Wiki: https://csde.esa.int/confluence/display/AEOLUSDISC/Aeolus+DISC+Acronym+List.

2 Background

After the successful first and second Aeolus reprocessing campaigns covering the second Flight Model laser (FM-B) from 24 June 2019 to October 2020 ([RD-1] and [RD-2]), efforts were focused on the third reprocessing campaign covering the first operation period of Flight Model laser (FM-A) from 31 August 2018 to 16 June 2019 (see Figure 1).

This is the third reprocessing campaign ([RD-3]). It was performed based on processor versions deployed in March 2022 (L1bP 7.12, L2aP 3.14.8, L2bP 3.70), used for near-real time (NRT) generation of baseline 14 (B14) data products. The reprocessing covers L1A, L1B, L2A and L2B data products as well as the respective auxiliary files and provides for the first-time low wind bias data products for the FM-A period. In combination with the second reprocessing campaign and the NRT baselines, the third reprocessing campaign ensures a seamless data availability of low wind bias products from the beginning of the mission to this day. The next reprocessing campaign will be performed using the B16 processor version and will cover the complete data period from September 2018 to the release of NRT baseline B16 in spring 2023.

Note that the first usable L2A and L2B products are available from 3 September 2018 onwards since the data collected during the first four days are used for calibration, OWV and NOP.

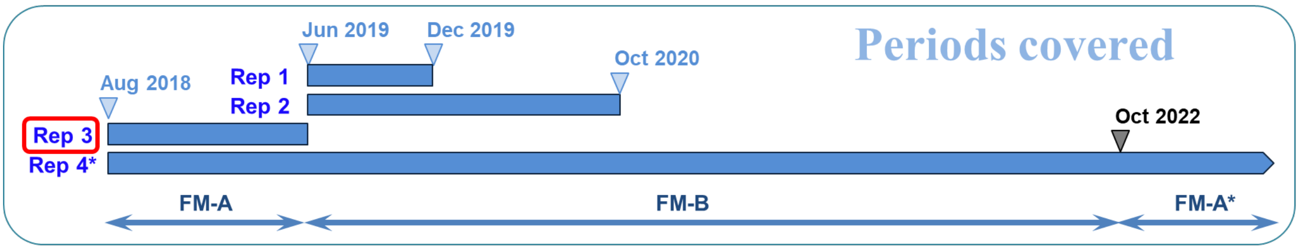


Figure 1: Periods covered by various reprocessing campaigns.

Several aspects of the reprocessing and the verification of L2B products are presented hereafter. The remainder sections of this report are as follows. The reprocessed auxiliary data were validated by comparison against their original counterparts. Section 3 presents the validation of AUX_ZWC data while Section 4 shows the validation of AUX_MRC and AUX_RRC data. Section 5 is dedicated to the efforts towards mitigation of the hot pixel issue. Verification of L2A and L2B products are presented in Sections 6 and 7. Conclusions are listed in Section 8.

3 AUX_ZWC

To verify the content of the reprocessed AUX_ZWC files, comparisons of original and reprocessed AUX_ZWC data were performed for various sub-datasets. Corresponding presentations are available on the EDAFECS-FTP under ***/documentation/documentation_deliverables_DISC/Reprocessing/Reprocessing_3rd_campaign***:

- 20220207 UM RPRO3 Verification AUX_ZWC_1B - AUX_RRC - AUX_MRC - AUX_PAR_1B - EDCF - 1B13.pptx
- 20220310 UM RPRO3 Verification AUX_ZWC_1B - AUX_RRC - AUX_MRC - AUX_PAR_1B - EDCF - 1B14.pptx
- 20211216 UM BW 3rd Reprocessing Campaign ISR MRC RRC PAR1B.pptx
- 20220310 UM BW 3rd Reprocessing Campaign ISR MRC RRC PAR1B.pptx

These reports were created on the basis of a comparison of original NRT data against data reprocessed on the Sandbox. In order to verify the PDGS reprocessed dataset, we rely on a comparison of the content of selected AUX_ZWC files from the PDGS against the Sandbox dataset. The output of both reprocessing activities is supposed to be identical in terms of scientific data content, i.e. mainly ground correction velocities.

In order to more or less representatively cover the full period of the 3rd reprocessing campaign from September 2018 until June 2019, three test weeks have been picked (for the comparison of NRT data against Sandbox B13 and B14 trial runs) while trying to consider the different states of Aeolus during that period, i.e. before and after the switch-off on 2019-01-14 due to a GPS issue.

In addition, the selection of these weeks is supposed to satisfy the demands of most of the verification procedures of the reprocessing. The 1st test week includes the onset of Rayleigh hot pixel [4,1] and provides the possibility to evaluate the improved DCMZ correction in conditions with a low measurement frequency of DUDEs. The 2nd and 3rd test week provide a continuous time period without irregularities (according to https://csde.esa.int/confluence/display/AEOLUSDISC/Aeolus+Activities+History) apart from the onset of Rayleigh hot pixel [2, 1]. **Table 1** provides additional details regarding the three test weeks with their more than 300 AUX_ZWC files in total.

Table 1: Characterization of the three sub-datasets of one week each for comparison and verification.

ZWC data sets	**Baseline** (*PDGS / Sandbox*)	**Software_Ver (L1bP)** (*PDGS / Sandbox*)	**Start time**	**Stop time**	**Number of files** (*PDGS / Sandbox*)
1. NOV 2018	1B14 / 0001	07.12 / 07.12	2018-11-11 01:05.14	2018-11-08 00:39.50	103 / 103
2. MAY 2019	1B14 / 0001	07.12 / 07.12	2019-05-01 00:37.53	2019-05-08 00:38.17	104 / 104
3. MAY 2019	1B14 / 0001	07.12 / 07.12	2019-05-08 00:37.53	2019-05-15 00:38.17	103 / 103

The original AUX_ZWC dataset from the Near-Real-Time (NRT) processing was taken from the directory ***/ADDF/L1B_Calibration/1B02*** on the FTP Server ***aeolus-ds.eo.esa.int***. The AUX_ZWC dataset reprocessed on the Sandbox was taken from ***/data_exchange/reprocessing_nr_3_results_B14_trial_run*** on the EDAFECS commissioning FTP ***Commissioning.aeolus.esa.int*** (User: *aeolus-edafecs*). The dataset reprocessed by PDGS is available from the directory ***/RPRO_3/Operational*** on the reprocessing FTP repository ***aeolus-ops-repo.eo.esa.int*** (User: *aeolus-disc*).

As stated in 20220310 UM RPRO3 Verification AUX_ZWC_1B - AUX_RRC - AUX_MRC - AUX_PAR_1B - EDCF - 1B14.pptx we found that there was an incorrect selection of the AUX_PAR_1B file for the 2nd and 3rd test week on the Sandbox B14 trial run. However, the AUX_PAR_1B files involved for the 3 test weeks:

- AE_OPER_AUX_PAR_1B_**20181031T032914_20181107T042751**_0001 (*Sandbox / all 3 weeks*)
- AE_OPER_AUX_PAR_1B_**20181031T021349_20181107T032915**_0001 (*PDGS 1. week*)
- AE_OPER_AUX_PAR_1B_**20181107**T020537_**20181114**T032839_0001 (*PDGS 1. week*)
- AE_OPER_AUX_PAR_1B_**20190429**T014204_**20190516**T003730_0001 (*PDGS 2. + 3. week*)

only differ with respect to the Energy Drift Correction Factor and selected header information as exemplarily shown in Figure 2. As the Energy Drift Correction Factor should only affect the ISR processing and not the wind measurements, the AUX_PAR_1B files cannot explain the observed 12% difference in the number of valid Rayleigh Ground Correction Velocities (GCVs) as given in Figure 6. In addition, the same AUX_MRC_1B (Figure 5) and AUX_RRC_1B (Figure 6) files were used by the Sandbox and PDGS, at least according to their name. Consequently, it is unclear what causes the discrepancy in the number of Rayleigh GCVs.

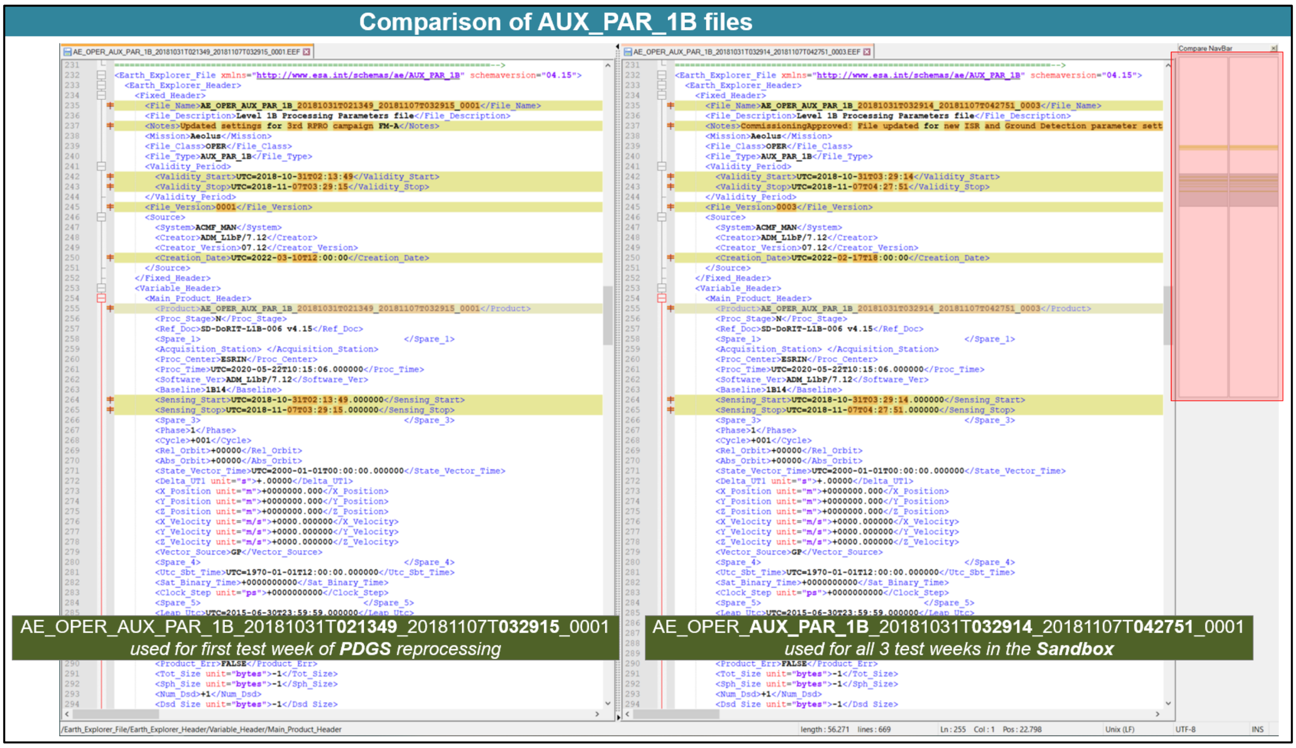


Figure 2: Comparison of the two AUX_PAR_1B files used by PDGS (left) and the Sandbox (right) for the reprocessing of the 1st test week.

The three AUX_ZWC sub-dataset have been analyzed by means of the DLR-HBE 2021v1 tool. The following findings can be noted down:

- The number of AUX_ZWC files agrees for the three test weeks between the Sandbox data set and the 3rd reprocessing data set.
- Different start/stop times of AUX_ZWC files (*also represented in the filenames*) were observed, which are most likely caused by differences in quality control via related parameters in the applied AUX_PAR_1B files and, hence, the validity of Ground Correction Velocities, i.e. the availability/existence of Data Set Records. However, the AUX_PAR_1B and AUX_M/RRC_1B files used for the Sandbox and PDGS reprocessing do not show any differences that could explain a different behavior regarding the quality control of GCVs.
- Exemplarily, for the 1st test week there are 7 AUX_ZWC files with different stop times, i.e. differences in valid GCVs and hence written DSRs.
- Min-Max range of Ground Correction Velocities in the reprocessed dataset after a minimal QC in the HBE tool (*thresholds for Mie and Rayleigh SNR, useful signal and minimum number of ground bins all set to 1*):
  - Mie (Sandbox): min: -7.09 m/s (*week 2*) max: 10.98 m/s (*week 1*)
  - Rayleigh (Sandbox): min: -73.72 m/s (*week 3*) max: 75.55 m/s (*week 1*)
  - Mie (PDGS): min: -6.93 m/s (*week 2*) max: 10.19 m/s (*week 2*)
  - Rayleigh (PDGS): min: -73.72 m/s (*week 3*) max: 32.99 m/s (*week 1*)
- Unlike for the 2nd reprocessing, the ice coverage of the polar regions with its seasonal dependency is assumed to have only a minor impact onto the statistics of the 3rd reprocessing dataset, given the selection of test weeks in autumn (November) and spring (May).
- The major part of the differences in biases (*mean and median in the tables on the right of Figure 5 and Figure 6*) between the three test weeks, namely November versus May, are most likely due to the temporally varying M1 temperature dependence of L1B data. M1 bias correction only takes place for L2B data.
- The bias differences in the Rayleigh channel obtained from a comparison of the individual test weeks of the two datasets, i.e. 1st week Sandbox vs. 1st week PDGS, etc., are presumably caused by the additional valid Rayleigh GCVs in the Sandbox dataset.
- Lower random errors (*std. dev. and MAD in the tables on the right of Figure 5 and Figure 6*) for the 1st test week compared to the 2nd and 3rd test week for both datasets can most likely be explained by the higher signal level (*i.e. better SNR and more valid GCV as can be seen from Figure 3 and Figure 4*) at the start of the Aeolus mission and its subsequent continuous decay over the following months. The high signal level during the beginning of the FM-A period obviously even resulted in many valid GCVs from sea surface for test week #1.
- The random errors (*std. dev. and MAD in the tables on the right of Figure 5 and Figure 6*) for PDGS are smaller than those for the Sandbox dataset, suggesting that the additional GCVs from the latter are of bad quality.

Apart from a so far unsolved difference in valid Rayleigh ground correction velocities between PDGS and Sandbox, the overall quality of the dataset reprocessed by PDGS complies with the expectations and is accepted.

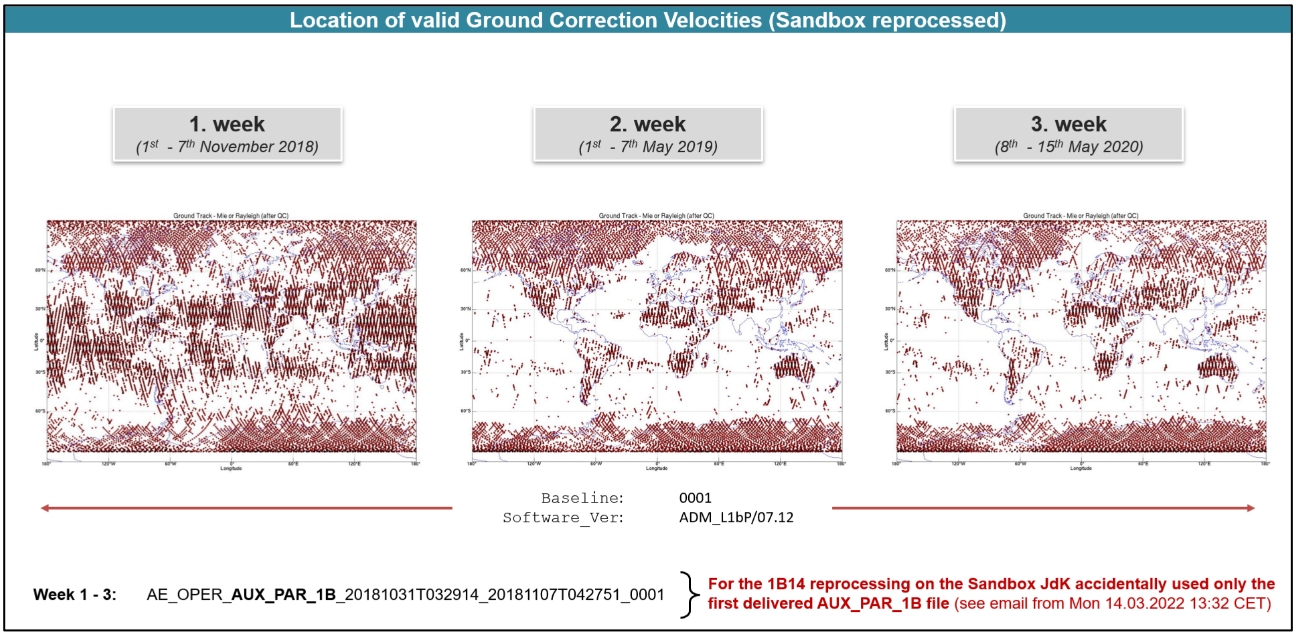


Figure 3: Location of the valid Mie and Rayleigh Ground Correction Velocities contained in the AUX_ZWC files of the 1st, 2nd and 3rd test week of the Sandbox 1B14 reprocessed dataset, along with the applied AUX_PAR_1B file containing the ground detection parameters in the <WVM_Params> section.

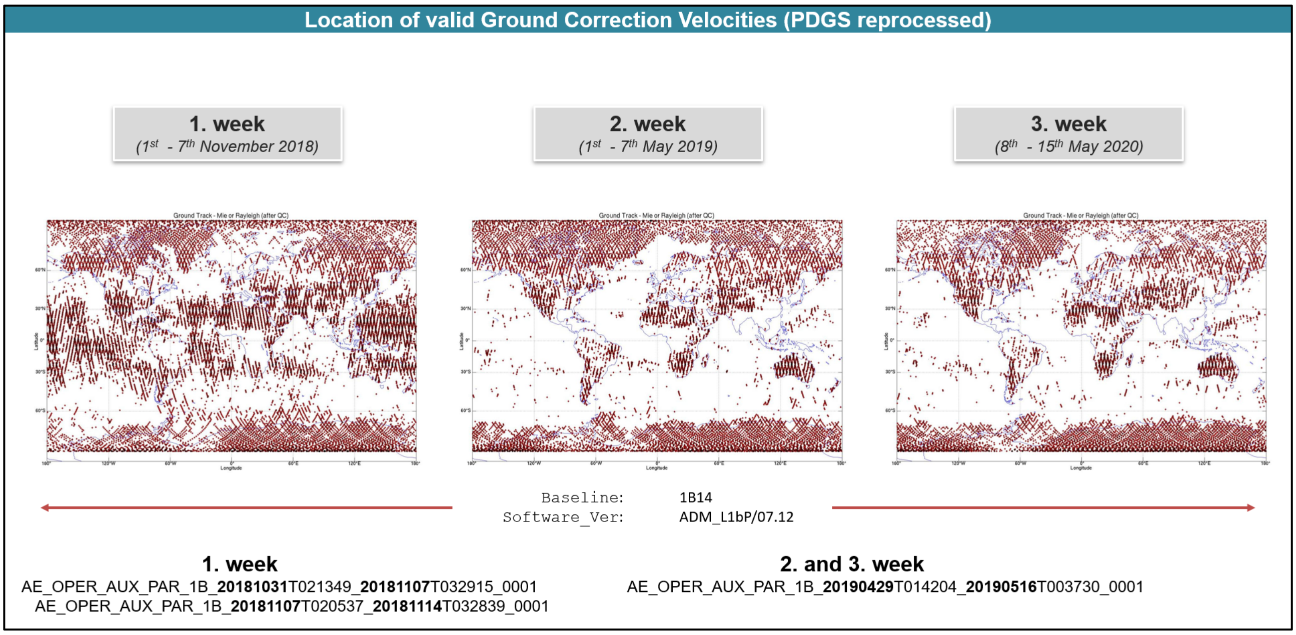


Figure 4: Location of the valid Mie and Rayleigh Ground Correction Velocities contained in the AUX_ZWC files of the 1st, 2nd and 3rd test week of the PDGS 1B14 reprocessed dataset, along with the applied AUX_PAR_1B files containing the ground detection parameters in the <WVM_Params> section.

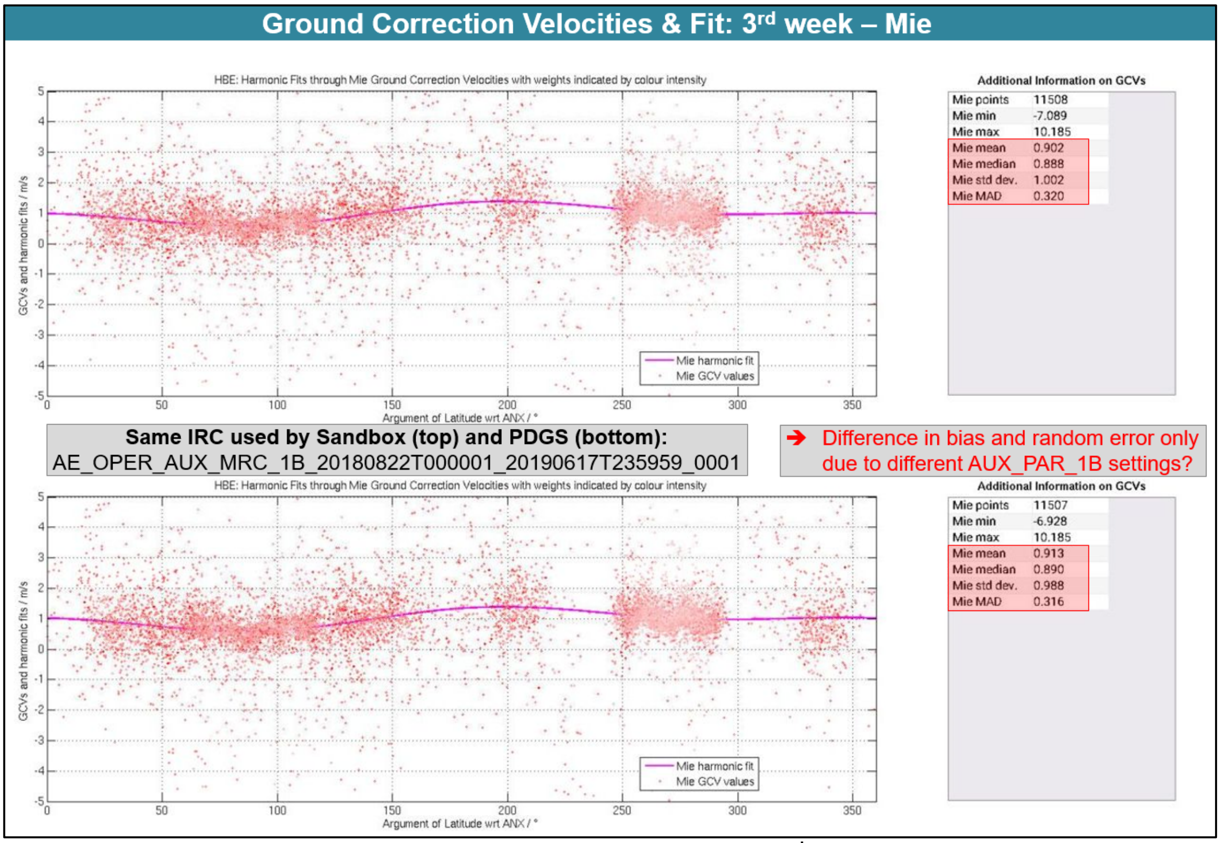


Figure 5: Mie Ground Correction Velocities and harmonic fits of the 3rd test week from the dataset reprocessed by the Sandbox (top) and PDGS (bottom).

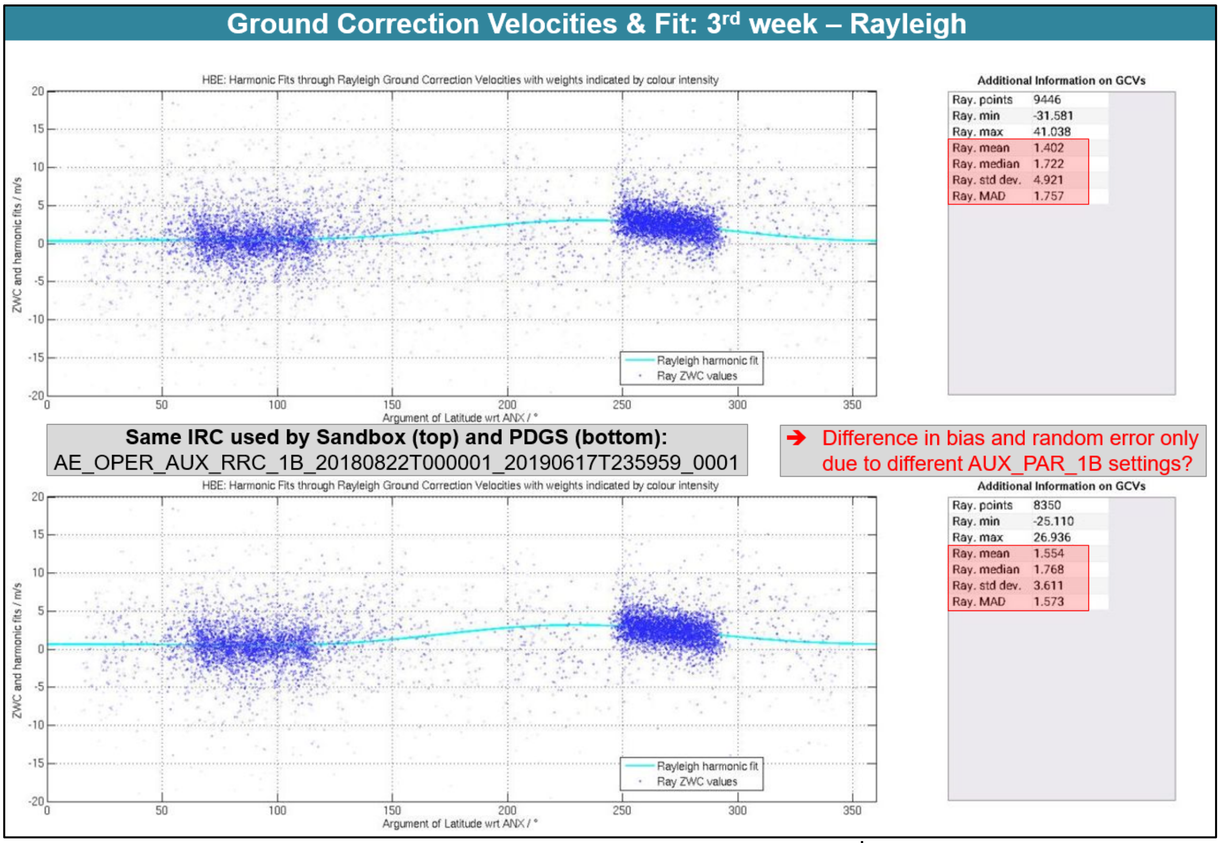


Figure 6: Rayleigh Ground Correction Velocities and harmonic fits of the 3rd test week from the dataset reprocessed by the Sandbox (top) and PDGS (bottom).

4 AUX_MRC and AUX_RRC

To validate the content of the reprocessed AUX_MRC and AUX_RRC files a comparison of the original NRT and the PDGS reprocessed dataset was performed for 13 IRCs from September 2018, November 2018 and May 2019. No checks against the Sandbox results were performed as per default the Sandbox did not provide output for calibration modes (available from Jos de Kloe on request after additional Sandbox runs).

The following IRCs were compared:

- #01, #02, #03, #04 from Aug. 7th, 14th, 20th, 27th
- #09, #10, #11, #12, #13 from Nov. 1st, 8th, 15th, 19th, 26th
- #31, #32, #33, #34, from May 9th, 16th, 23rd, 30th

The original AUX_MRC and AUX_RRC dataset from the Near-Real-Time (NRT) processing was taken from the respective folders under the directories ***/ADDF/L1B_Calibration/1B01***/, ***/ADDF/L1B_Calibration/1B02***/ and ***/ADDF/L1B_Calibration/1B03***/ on the FTP Server ***aeolus-ds.eo.esa.int***. The reprocessed AUX_MRC and AUX_RRC dataset was taken from the directory ***/RPRO_3/Operational*** on the reprocessing FTP repository ***aeolus-ops-repo.eo.esa.int*** (User: *aeolus-disc*).

Table 2: Examples for the version numbering in the filenames of the original NRT, manually reprocessed NRT (September only) and PDGS reprocessed AUX_RRC_1B files (left) as well as start/stop times (bold) of the applied AUX_PAR_1B files (right). From top to bottom: IRC #01, #09, #31, #34.

	AUX_RRC filename	AUX_PAR_1B file used for processing
NRT:	AE_OPER_AUX_RRC_1B_20180907T123614_20180907T125250_0001	AE_OPER_AUX_PAR_1B_**20180712**T000000_**99991231**T235959_0001
NRT (reproc.):	AE_OPER_AUX_RRC_1B_20180907T123614_20180907T125250_0002	AE_OPER_AUX_PAR_1B_20180712T000000_99991231T235959_0003
PDGS RPRO3:	AE_OPER_AUX_RRC_1B_20180907T123614_20180907T125250_0001	AE_OPER_AUX_PAR_1B_**20180822**T000001_**20180919**T115403_0001
:	:	:
NRT:	AE_OPER_AUX_RRC_1B_20181101T185426_20181101T191050_0002	AE_OPER_AUX_PAR_1B_**20180712**T000000_**99991231**T235959_0003
PDGS RPRO3:	AE_OPER_AUX_RRC_1B_20181101T185426_20181101T191050_0002	AE_OPER_AUX_PAR_1B_**20181031**T021349_**20181107**T032915_0001
:	:	:
NRT:	AE_OPER_AUX_RRC_1B_20190509T194053_20190509T195705_0001	AE_OPER_AUX_PAR_1B_**20180712**T000000_**99991231**T235959_0003
PDGS RPRO3:	AE_OPER_AUX_RRC_1B_20190509T194053_20190509T195705_0002	AE_OPER_AUX_PAR_1B_**20190429**T014204_**20190516**T003730_0001
:	:	:
NRT:	AE_OPER_AUX_RRC_1B_20190530T194105_20190530T195717_0001	AE_OPER_AUX_PAR_1B_**20180712**T000000_**99991231**T235959_0004
PDGS RPRO3:	AE_OPER_AUX_RRC_1B_20190530T194105_20190530T195717_0001	AE_OPER_AUX_PAR_1B_**20190529**T232016_**20190606**T003742_0001

For the NRT processing during the 1st FM-A period five different AUX_PAR_1B files were used (three of them mentioned in Table 2), namely version numbers 1 – 5 with the respective inventory dates available from the timeline AeolusPDGS_AUXfileusage_timeline_v4.0.xlsx (https://csde.esa.int/confluence/pages/view-page.action?spaceKey=AEOLUSDISC&title=PDGS+AUX+files). The IRCs from the PDGS reprocessed dataset are obviously based on a correct selection of the provided AUX_PAR_1B files when considering the start/stop times according to the filenames in Table 2.

IRCs #01 – #04 (07./14./20./27.09.2018) also exist in a reprocessed version (Baseline 02) other than from the first two official reprocessing campaigns, hence, the additional entry indicated as “*NRT (reproc.):*” in Table 2.

Figure 7, Figure 8, Figure 9 and Figure 10 are based on the following AUX_RRC_1B and AUX_MRC_1B files taken from the PDGS dataset of the 3rd reprocessing:

01. IRC #001 AE_OPER_AUX_M/RRC_1B_20180907T123614_20180907T125250_0001.EEF
02. IRC #002 AE_OPER_AUX_M/RRC_1B_20180914T111314_20180914T112950_0001.EEF
03. IRC #003 AE_OPER_AUX_M/RRC_1B_20180920T093202_20180920T094838_0001.EEF
04. IRC #004 AE_OPER_AUX_M/RRC_1B_20180927T093950_20180927T095626_0001.EEF
05. IRC #009 AE_OPER_AUX_M/RRC_1B_20181101T185426_20181101T191050_0002.EEF
06. IRC #010 AE_OPER_AUX_M/RRC_1B_20181108T185338_20181108T191002_0002.EEF
07. IRC #011 AE_OPER_AUX_M/RRC_1B_20181115T185226_20181115T190850_0001.EEF
08. IRC #012 AE_OPER_AUX_M/RRC_1B_20181119T194438_20181119T200102_0001.EEF
09. IRC #013 AE_OPER_AUX_M/RRC_1B_20181126T181402_20181126T183026_0001.EEF
10. IRC #031 AE_OPER_AUX_M/RRC_1B_20190509T194053_20190509T195705_0002.EEF
11. IRC #032 AE_OPER_AUX_M/RRC_1B_20190516T194053_20190516T195705_0001.EEF
12. IRC #033 AE_OPER_AUX_M/RRC_1B_20190523T194117_20190523T195729_0001.EEF
13. IRC #034 AE_OPER_AUX_M/RRC_1B_20190530T194105_20190530T195717_0001.EEF

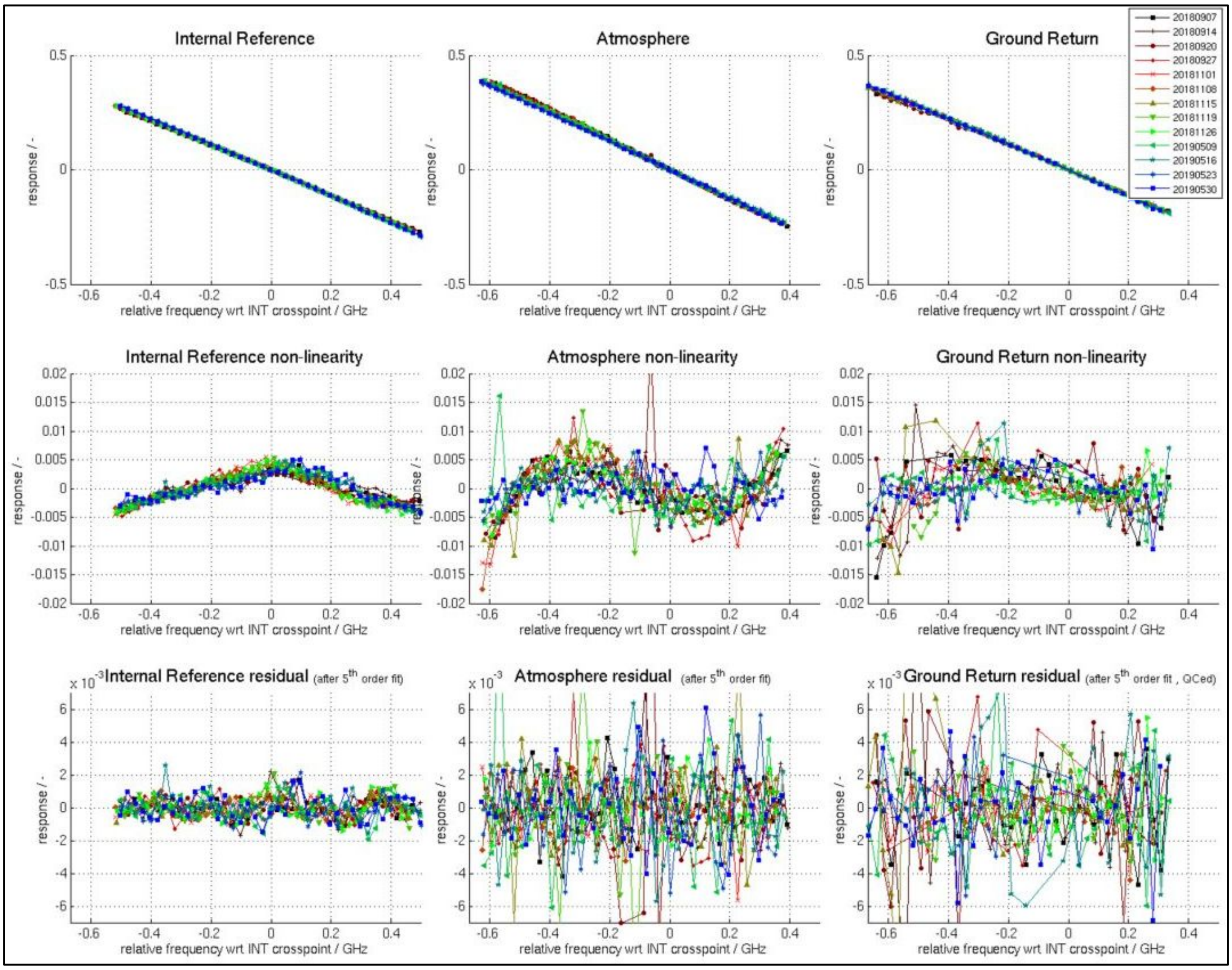


Figure 7: Response curves (top), non-linearities (middle) and residuals (bottom) for the internal reference (left), atmosphere (centre) and ground return (right) of the selected RRCs from the PDGS reprocessed dataset.

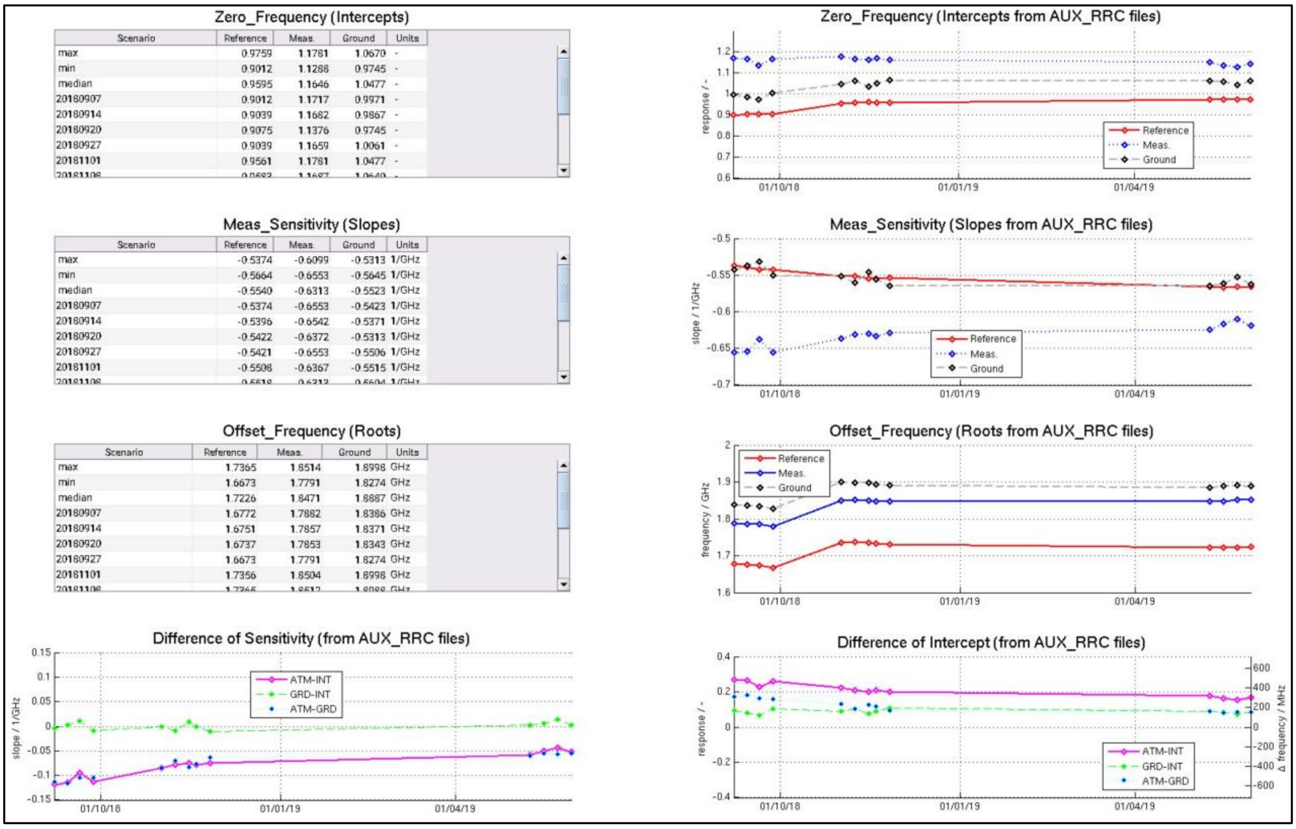


Figure 8: Intercept (Zero_Frequency), slope (Mean_Sensitivity) and offset (roots) for the internal reference, atmosphere and ground return of the selected RRCs from the PDGS reprocessed dataset. Differences of sensitivities and intercepts of the different paths are given at the bottom.

Figure 7 shows the Rayleigh response curves, their non-linearities and residuals for the internal reference, the atmosphere and the ground return of the selected RRCs from the PDGS reprocessed dataset. The results from the reprocessing have an improved quality mainly because of less outliers in the atmospheric and ground return curve compared to NRT dataset.

Figure 8 shows the timelines of the Intercept (Zero_Frequency), slope (Mean_Sensitivity) and offset (roots) for the internal reference, atmosphere and ground return of the selected RRCs from the PDGS reprocessed dataset. Differences of sensitivities and intercepts of the different paths are given at the bottom. Numerous changes in all these characteristics for all three calibration paths are visible from a comparison against the equivalent figure for the NRT dataset. All values are within expectations.

The corresponding figures for the NRT data set are available from the presentation "*20221103 UM RPRO3 Verification PDGS AUX_ZWC_1B AUX_RRC AUX_MRC AUX_PAR_1B.pptx*" on the EDAFECS FTP under /documentation/documentation_deliverables_DISC/Reprocessing/Reprocessing_3rd_campaign/Verification_PDGS_dataset_final.

Figure 9 shows the Mie response curves, their non-linearities and residuals for the internal reference and the ground return of the selected MRCs from the PDGS reprocessed dataset. The results from the reprocessing have an improved quality mainly because of additional valid frequency steps for the ground return compared to NRT dataset. This particularly affects the later MRCs from May 2019, presumably due to improved ground detection and QC applied to the low return signals during this period right before the switch to FM-B.

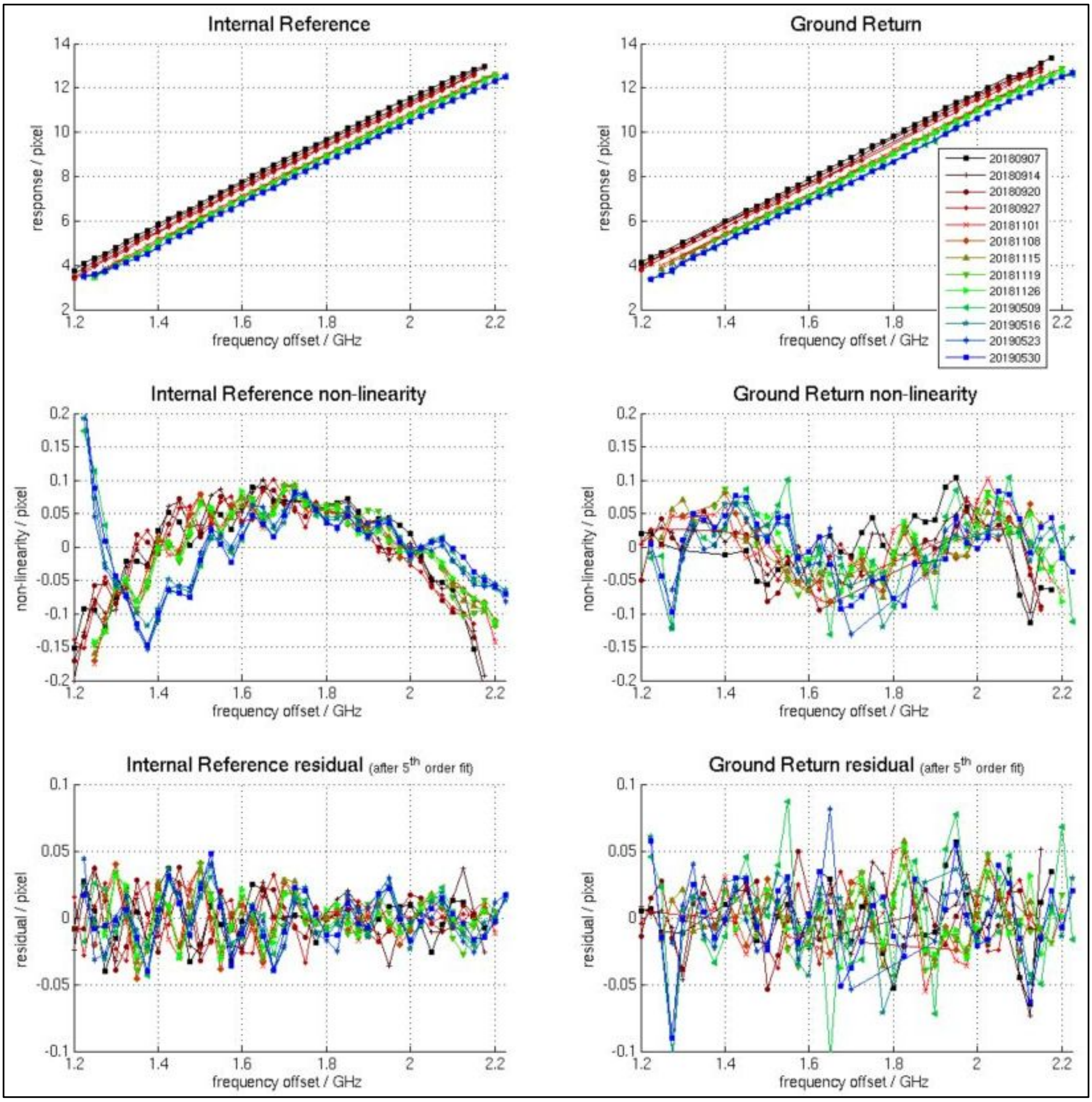


Figure 9: Response curves (top), non-linearities (middle) and residuals (bottom) for the internal reference (left) and ground return (right) of the selected MRCs from the PDGS reprocessed dataset.

Figure 10 shows the timelines of the Intercept (Zero_Frequency) and slope (Mean_Sensitivity) for the internal reference and ground return of the selected MRCs from the PDGS reprocessed dataset. Compared to the NRT dataset the MRCs from the 3rd PDGS reprocessing mainly differ in terms of intercept and slope for the ground return response curves in May, presumably related to the signal decay over the FM-A period together with the benefit from an improved ground detection scheme for the reprocessing. All values are within expectations.

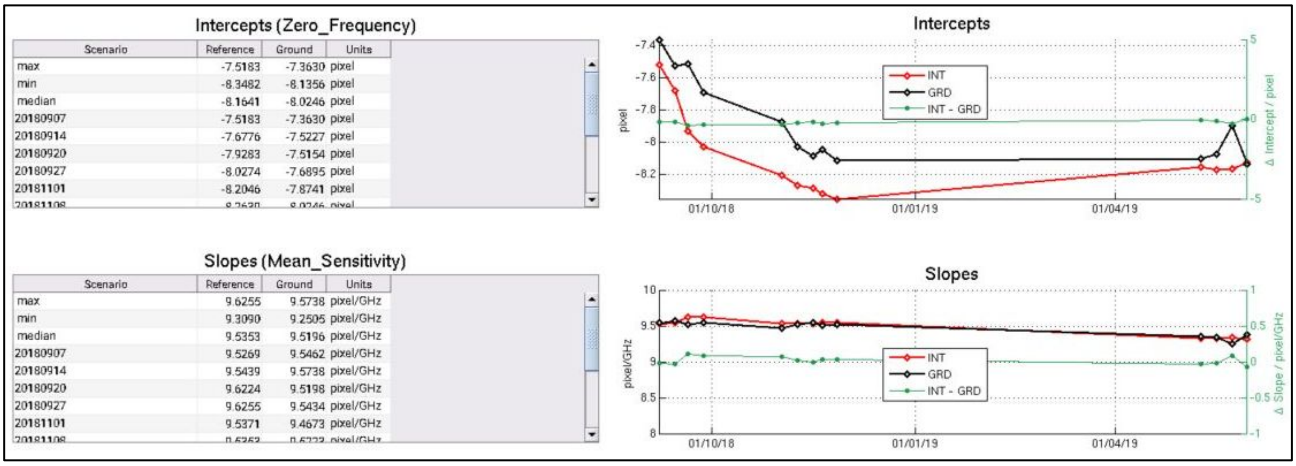


Figure 10: Intercept (Zero_Frequency) and slope (Mean_Sensitivity) for the selected MRCs from the PDGS reprocessed dataset. Differences are indicated as green lines.

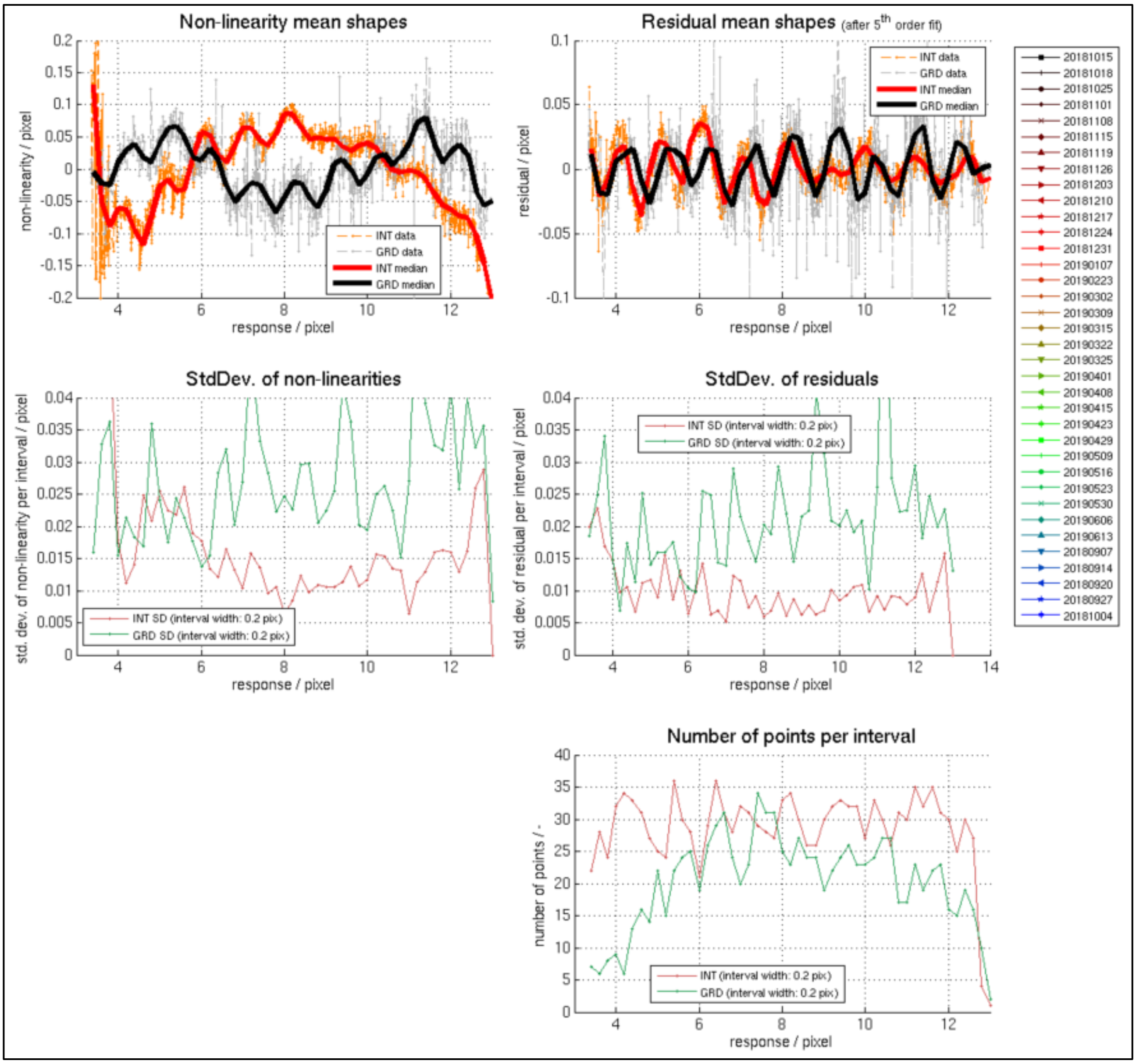


Figure 11: Averaged Mie non-linearity of the internal reference and ground return (atmospheric path) as applied during the 3rd reprocessing (top left). Residuals after a 5th order polynomial fit show the pronounced pixilation effect due to ACCD characteristics. Additional statistics are given in the lower graphs.

Unlike for the FM-A NRT processing, which did not include any Mie non-linearity correction, the 3rd reprocessing campaign made use of a Mie non-linearity curve averaged over all 36 MRC of the FM-A period (#006 – #036 from original NRT data and #001 – #005 manually reprocessed with adapted parameters in the AUX_PAR_1B file). The respective parameters (Num_Sampling_Points_Internal_Reference, Num_Sampling_Points_Atmosphere, Pixel_Positions_Internal_Reference, Fitted_Reference_Pulse_Error_Mie_Response, Pixel_Positions_Atmospheric_Path and Fitted_Measurement_Error_Mie_Response) had been introduced to the AUX_PAR_1B file with Baseline 08. The shape of the non-linearity is shown in Figure 11.

The AUX_MRC and AUX_RRC dataset was analyzed by the means of the DLR Matlab Plot Tool. The following findings can be noted down:

- The reprocessed and the NRT dataset show the same number (within the selected verification period) of IRCs and the same dates.
- The occurrence of different version numberings between NRT and reprocessed AUX_RRC_1B and AUX_MRC_1B is not of concern as its explanation most likely follows the findings from the 2nd reprocessing (email from 2021/06/23 Simone Bucci).
- The AUX_MRC and AUX_RRC files show the expected content and structure (*additional information available in the reprocessed files due to updated processor baseline*). Differences in specific values are explainable by the use of different QC thresholds and the new Mie non-linearity correction in the AUX_PAR_1B file, as well as an improved ground detection scheme.
- Effects of improved quality control are visible for the Rayleigh atmosphere & ground return response curves, namely an improved random error of the residuals of the Rayleigh atmosphere and ground return response curves as well as additional valid frequency steps for the ground return.
- Mie response curves from the reprocessing have an improved quality mainly because of additional valid frequency steps for the ground return compared to NRT dataset.
- The correct usage of the provided AUX_PAR_1B files during the reprocessing can be confirmed.

The overall quality of the reprocessed dataset is as expected, and it is accepted.

5 Mitigation of the hot pixel issue

5.1 Introduction

Here, only a brief a summary of the hot pixel issue is provided. A more detailed description can be found in [RD-4] and [RD-5]. Hot pixels are ACCD memory pixels with enhanced dark current signal levels. Due to the measurement principle of Aeolus and the instrument design, the accuracy of wind and aerosol measurements is very sensitive towards dark current signal changes. To mitigate this issue, regularly performed dark current characterizations, so-called Dark Current in Memory Zone (DCMZ) measurements, are carried out nowadays. These measurements provide a map of the current dark signal state of the ACCD which can be used for the correction. However, this correction has only been activated on 14th of June 2019. Unfortunately, the hot pixel issue was not anticipated before launch and no regular dark current measurements were performed at the beginning of the mission during the FM-A period.

Figure 12 indicates the availability of DCMZ measurements for the FM-A period. At the beginning, DCMZ measurements were carried out only very sporadically. Afterwards, between 2018/11/26 and 2018/12/19 the frequency of DCMZ measurements was increased to one per day. Subsequently, it was increased to two per day until the instrument went to standby on 2019/01/14. From 2019/02/15 onwards the DCMZ frequency was further increased to four per day.

To correct for hot pixel induced steps in periods with a low DCMZ measurement frequency, it is necessary to detect hot pixel induced steps in the backscatter signals measured in wind mode. This approach was already implemented for previous reprocessing campaigns. However, a simplified detection method was used which was not able to detect hot pixels steps for illuminated Rayleigh pixels. Thus, a sophisticated approach was developed and for the first time implemented for this reprocessing campaign [RD-6].

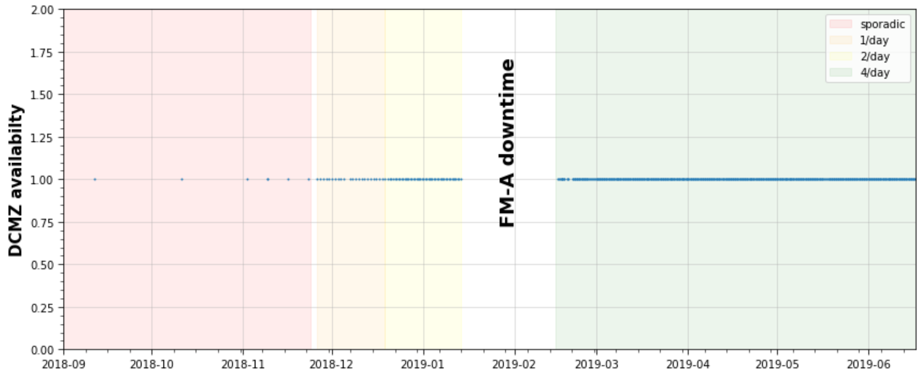


Figure 12: Measurement frequency of DCMZs for the FM-A period from 2018/09/01 to 2019/06/15.

5.2 Methodology

The change point detection algorithm is introduced in detail in [RD-7] and [RD-8]. The algorithm uses backscatter signal intensities (corrected for DCO and SBKG) from ALD_U_N_1A products as input and scans the time series for sudden changes induced by dark current signal changes. It is based on Singular Spectrum Analysis (SSA) which decomposes the time series into structural components such as mean, noise and periodical components. The sequential application of SSA to each time step of the time series allows to detect sudden changes in the mean of the time series. The backscattered signal of the hot pixels is a complex composition of atmospheric measurement and dark current signals. To increase the reliability of detection the time series signal of one additional neighbouring pixel is considered. This leads to the reduction of atmospheric effects on the hot pixel detection and also allows for the detection of subtle hot pixel induced steps.

As a first step, the reprocessing dataset was split into 37 chunks where each chunk holds one week of data. Next, the change point detection was applied for each chunk to every Mie and Rayleigh hot pixel. The algorithm output, i.e. the detected change point indices and amplitudes, was stored for each hot pixel. After each processed chunk the algorithm performance was validated by visually inspecting the step-corrected hot pixel time series. The validation allows to detect problems with the change point detection at an early stage which makes it possible to adapt algorithm parameters and repeat the detection in case of problems.

A summary of the processed hot pixels is provided in Table 3 and Table 4. The third column describes the hot pixel characteristics and the expected impact on the data quality [RD-5]. The last column describes the number of modifications applied to each hot pixel. Note that the dark current rates of Rayleigh hot pixel [0,6] also had to be corrected for increased signals due to atmospheric contamination as a result of issues with the location of the top range gate in DCMZ mode (see section 5.3). Thus, additional modifications besides the correction of hot pixel induced steps were necessary for this hot pixel.

Table 3: List of Mie hot pixels during the reprocessing period. In brackets, the first number refers to the ACCD row (0 to 23) and the second numbers refers to the column (0 to 15). Note that the counting starts from 0. RTS stands for Random Telegraph Signal.

Pixel	Activation	Impact	No. of steps
[15,14]	before IOCV	Low (small RTS amplitudes with low fluctuation rate)	24
[23,2]	before IOCV	Low (very high frequency but low amplitudes)	416
[12,8]	2018-10-21	High (multi-level RTS)	163
[1,14]	2018-10-24	Low (two level RTS with rather low fluctuation rate)	78
[4,12]	2019-01-09	Medium (sporadic mean shifts)	90
[19,1]	2019-03-31	Low (sporadic mean shifts)	12
[9,12]	2019-04-26	High (multi-level RTS, high amplitudes)	30

Table 4: List of Rayleigh hot pixels during the reprocessing period. In brackets, the first number refers to the ACCD row (0 to 23) and the second numbers refers to the column (0 to 15). Note that the counting starts from 0. RTS stands for Random Telegraph Signal.

Pixel	Activation	Impact	No. of steps
[10,1]	2018-09-07	High (multi-level RTS)	206
[4,1]	2018-11-04	Medium (sporadic mean shifts)	51
[14,3]	2018-11-24	Low (sporadic mean shifts, low amplitude)	0
[19,9]	2019-01-27	High (multi-level RTS)	19
[0,6]	2019-02-20	Medium (stable multi-level RTS)	132
[10,15]	2019-03-17	High (multi-level RTS)	53
[2,1]	2019-05-08	High (multi-level RTS, high amplitudes)	22

5.3 Top-range bin issue

On 24th of May 2019 the on-board star tracker settings (SOR-148) and the associated parameters in the ground processing were changed. This also changed the height allocation of the range bins during DCMZ measurements which caused the top Rayleigh bin to be located above the Earth's surface. Thus, dark measurements of the first Rayleigh range bin were polluted with atmospheric signals from 24th of May onwards. To correct for this issue, the first range bin of both channel of the affected DUDE measurements were replaced with the first range bin of the last valid, i.e. non-affected, DCMZ measurement. To detected affected DCMZ measurements, it was checked if the height of the first range bin is below the altitude of the DEM intersection.

5.4 Validation

The outcome of this task was validated for the following test weeks:

- 2018-11-01 – 2018-11-08
- 2019-05-01 – 2019-05-15

As first step, the useful signals of the reprocessed L1B products were checked for remaining hot pixel effects. For this, the signals from the reprocessed dataset were compared with the NRT dataset. It should be mentioned that the L1B processor only contains pixel-wise signals for the Mie channel whereas for the Rayleigh channel only the summed-up signals for filter A and B are contained. This explains why this detailed pixelwise validation step is only possible for Mie hot pixels. Figure 13 compares the signal intensity for Mie pixel [12, 8] obtained from the near-real-time data with the reprocessed data. It demonstrates the successful correction of hot pixel induced steps for this example.

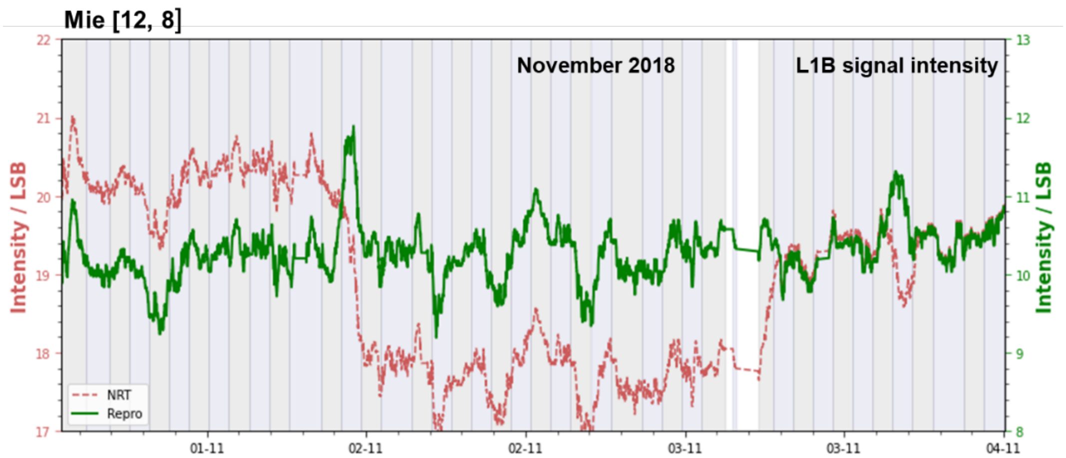


Figure 13: L1B signals of Mie hot pixel [12, 8] for the near-real-time (red) and reprocessed (green) dataset.

As second step, the correction selection of the provided DCMZ file in the reprocessing is checked for selected days of the test weeks. This is done by checking the header of the L1B files which gives information about the used auxiliary files to process the L1B product. This is then examined with the prepared set of AUX_DCMZ files of that day. The detailed results concerning this verification step can be found under [RD-9]. In summary, it can be stated here that both the file selection and scientific results were as expected.

6 Verification of L2A products

6.1 De-noising of SCA backscatter and extinction coefficients: The MLE algorithm concept

Because of the SCA algorithm principle, the aerosol retrievals are highly affected by signal noise. The particle extinction and backscatter coefficients are indeed retrieved independently and the extinction is normalized using the signals in the first range bin. This bin has a low SNR and the noise in this first bin is propagating downwards through the whole extinction profile.

A physical regularization scheme has then been implemented within L2A processor to reduce the noise contamination of SCA optical product. It is called Maximum Likelihood Estimation (MLE). It can mainly be seen as an alternative to the Standard Core Algorithm (SCA) processing of crosstalk corrected signals. It allows retrieval of extinction for particles and extinction to backscatter ratio (i.e. so-called lidar ratio) within predefined physical bounds.

It then assumes that extinction and backscatter are vertically collocated and the lidar ratio retrievals are bounded in expected physical limits, i.e. between 2 sr and 200 sr. A positivity constraint on the extinction and backscatter is also set.

Figure 14 illustrates the improvement of the extinction coefficient for particles retrieved by the MLE scheme compared to the SCA algorithm using the Sandbox reprocessed dataset for 2018. The extinction coefficient in the top bins appears noisy for the SCA but not for the MLE. The MLE then provides denoised extinction for particles with better horizontal homogeneity. Nevertheless, a known positive bias can be seen in lowermost bins close to the ground. Note that these differences are less pronounced with backscatter retrievals.

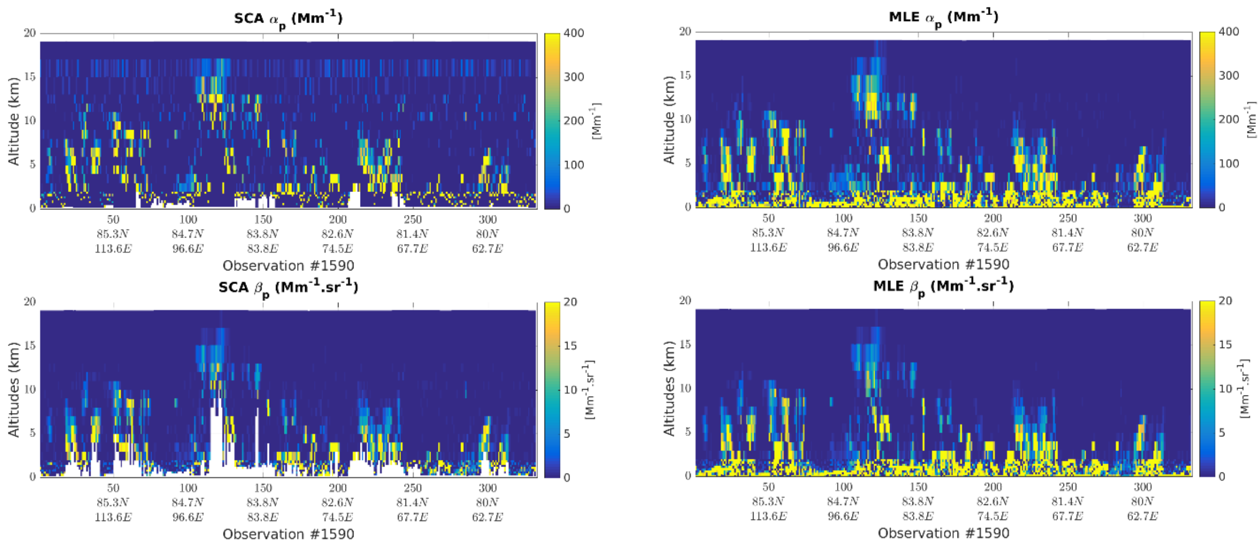


Figure 14: SCA and MLE extinction and backscatter coefficients for particles for a whole orbit in December 2018.

6.2 Overview of impacts on L2A products: Improvements and limitations

The data is homogeneously processed with L2A v3.14.8 which means:

- Final implementation of the physical regularization scheme, i.e. called Maximum Likelihood Estimation (MLE) or Denoising Scheme, to compensate the noise contamination of SCA retrievals. Optical properties of particles are then retrieved from a minimization using the L-BFGS-B open-source algorithm constrained by pre-defined physical bounds.

- The calibration coefficients characterizing the radiometric efficiency of the receivers, i.e. K_{ray} and K_{mie}, are estimated from signal prediction in particle-free regions of the atmosphere (i.e. scattering ratio below 1.164) for mid-altitudes 6 to 16 km. They are calculated per observation using a multiple linear regression based on telescope temperatures oscillations (i.e. information provided by the Accurate Housekeeping Telemetry (AHT) and Thermal Control (TC) processes).

Limitations:

- The L2A product shows limitations for truncated orbits and specific conditions, i.e. top most range bin set below 16km. The radiometric correction indeed includes a selection of bins corresponding to an aerosol-free atmosphere between 6km and 16km. As a result, too few bins are selected when the topmost bin is set in low altitudes, leading to a failure of the regression fit. The K_{ray} and K_{mie} coefficients then exceed a defined validity range.

6.3 Quick analysis of reprocessed dataset

6.3.1 Outliers in SCA backscatter and extinction for L2AP V3.14.8 SANDBOX

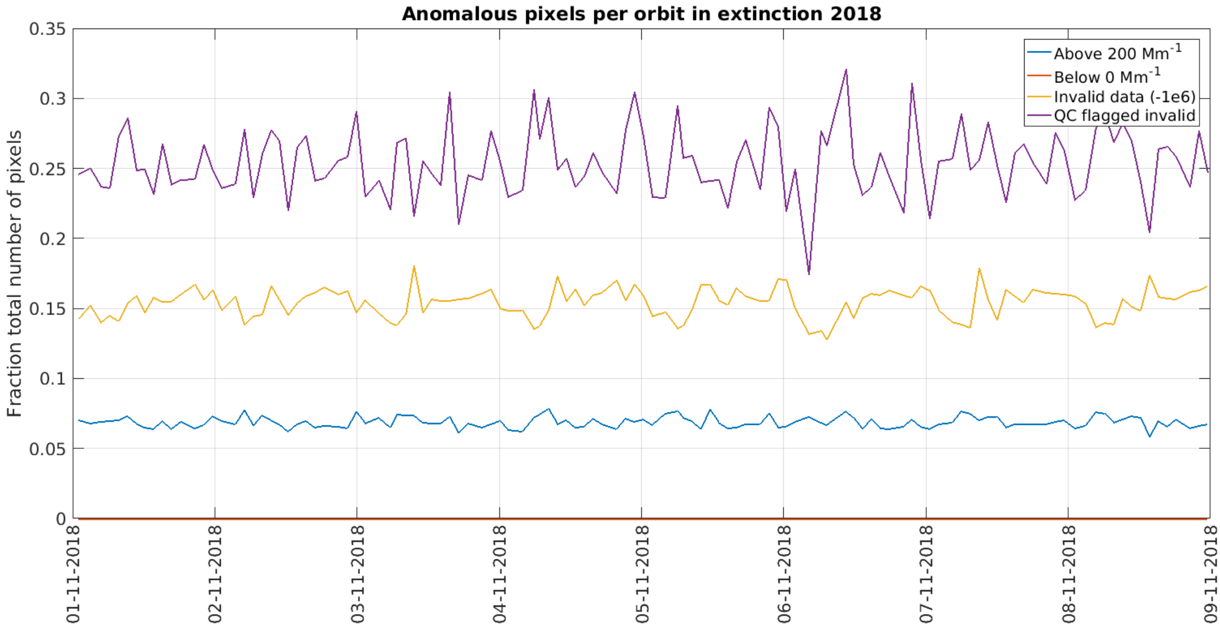
The extinction and backscatter values are sorted in 3 categories to easily characterize the occurrence of anomalous values:

1) values larger than 200 Mm^{-1} for extinction and 200 $Mm^{-1}.sr^{-1}$ for backscatter,
2) values <0,
3) not computed and set to -1.

As regards the extinction coefficients, the number of anomalous values is acceptable for both years. The scores in 2019 look better (i.e. lower fraction of QC flagged invalid and lower values larger than 200 Mm^{-1}). The 2019 dataset shows less negative backscatter except for the orbit #4118 with low top range bin altitude. It can also be noted that fraction of backscatter values larger than 200 $Mm^{-1}.sr^{-1}$ is very low (i.e. less than 2%) compared to the extinction. Illustrations of the SCA scores for trial weeks dataset are illustrated below.

6.3.2 Outliers in MLE backscatter and extinction for L2AP V3.14.8 SANDBOX

Due to the pre-defined physical bounds included in MLE algorithm no negative extinction or backscatter can be seen for both years. Only the top and bottom most (i.e. bin index 1 and 24) are set to -1 for 2018 backscatter whereas the fraction of values set to -1 reaches ~25% in 2019. In terms of backscatter values larger than 200 Mm^{-1} the scores look improved in 2019 compared to 2018 (extinction mean score from ~17% to ~12%, backscatter mean score from ~4% to ~1%). Illustrations of the MLE scores for trial weeks dataset are illustrated below.



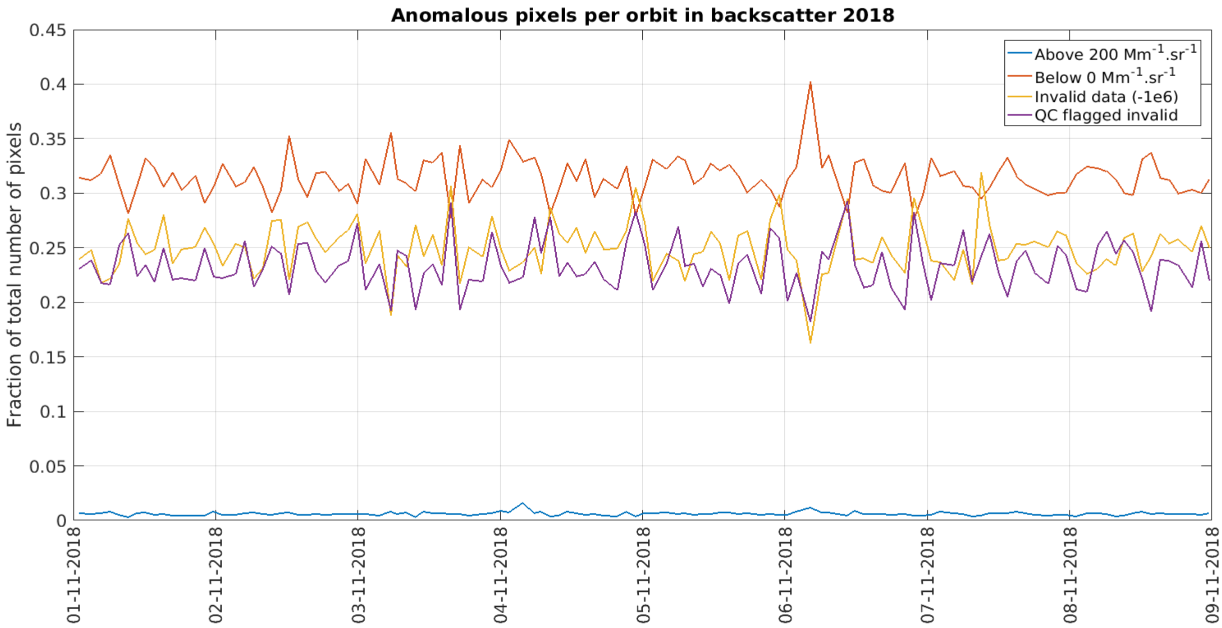


Figure 15: Fraction of SCA anomalous pixels in extinction (top) and backscatter (bottom) for 2018 dataset: outliers with high values (blue), with values <0 (orange) or not computed and set to -1 (yellow).

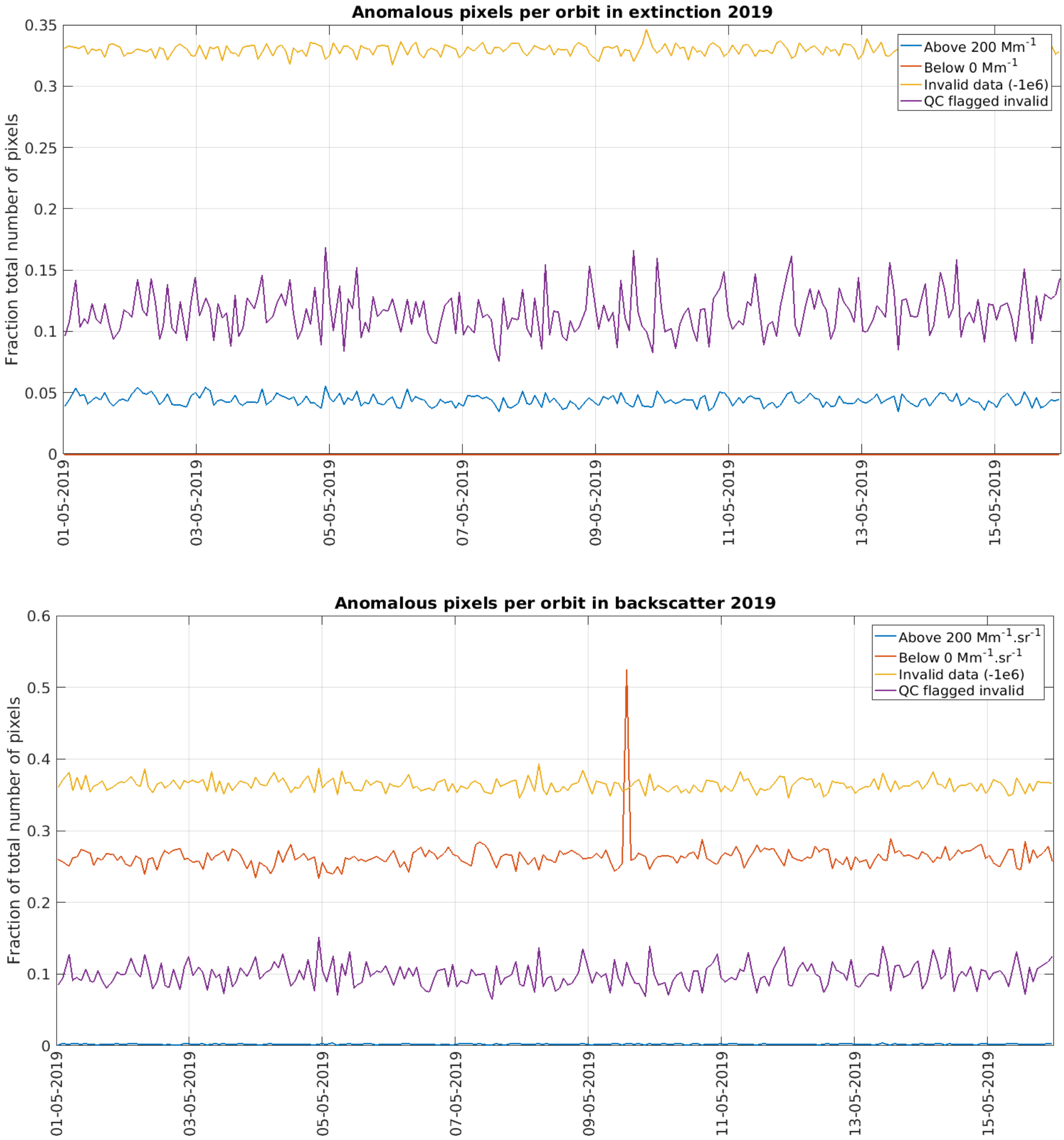
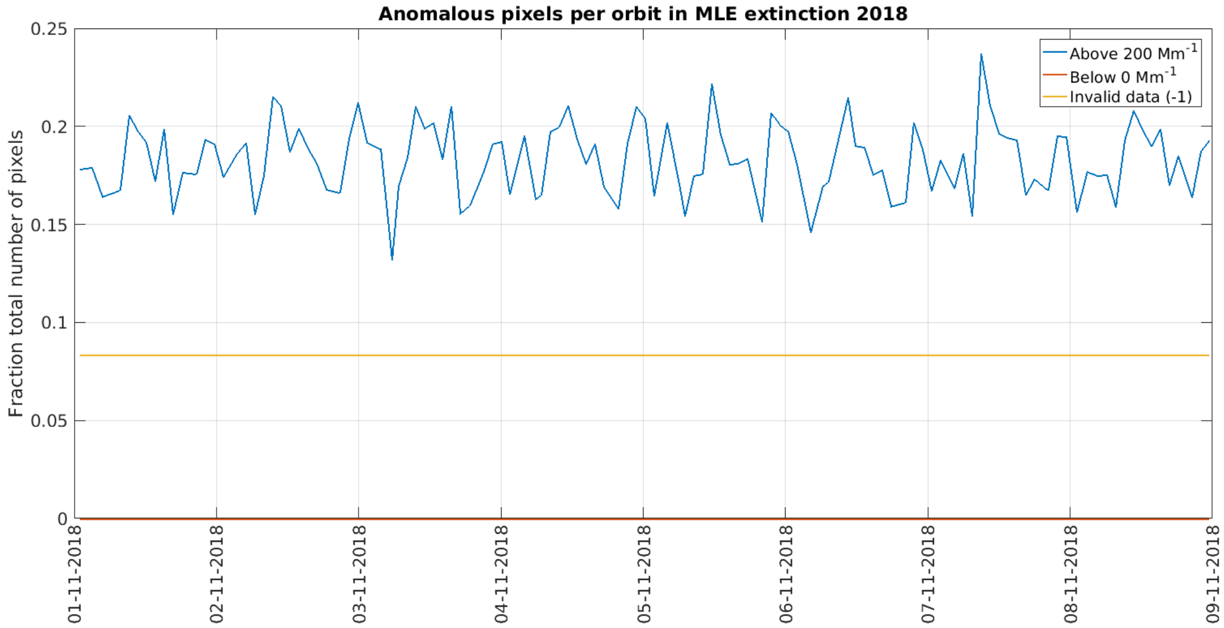


Figure 16: Fraction of SCA anomalous pixels in extinction (top) and backscatter (bottom) for 2019 dataset: outliers with high values (blue), with values <0 (orange) or not computed and set to -1 (yellow).



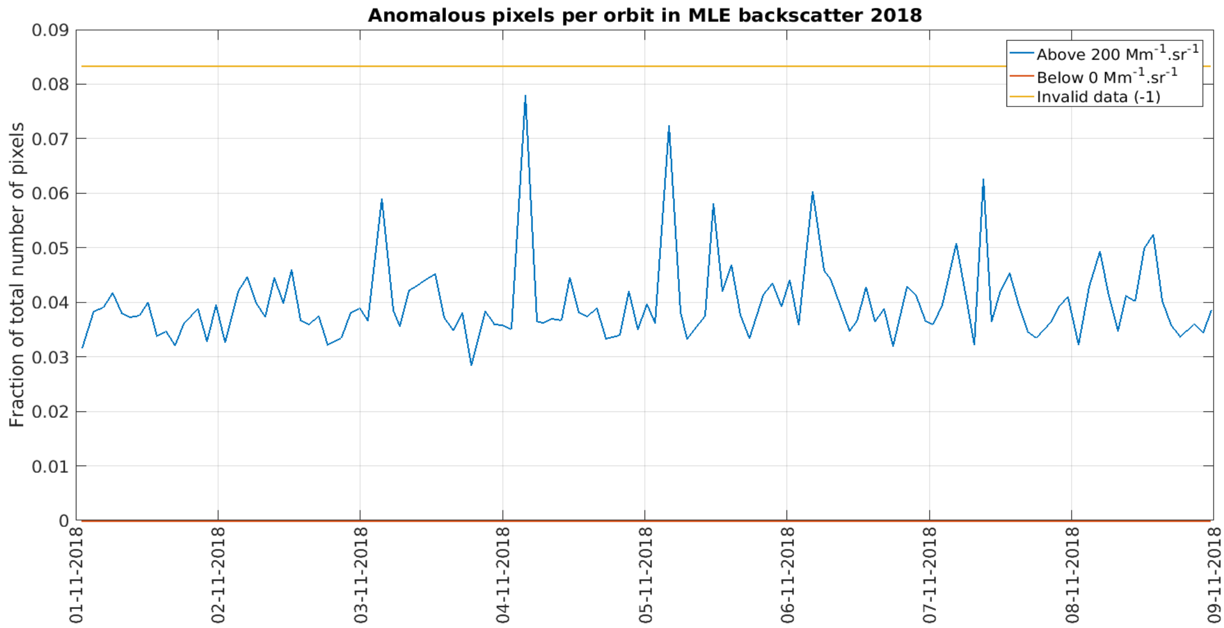
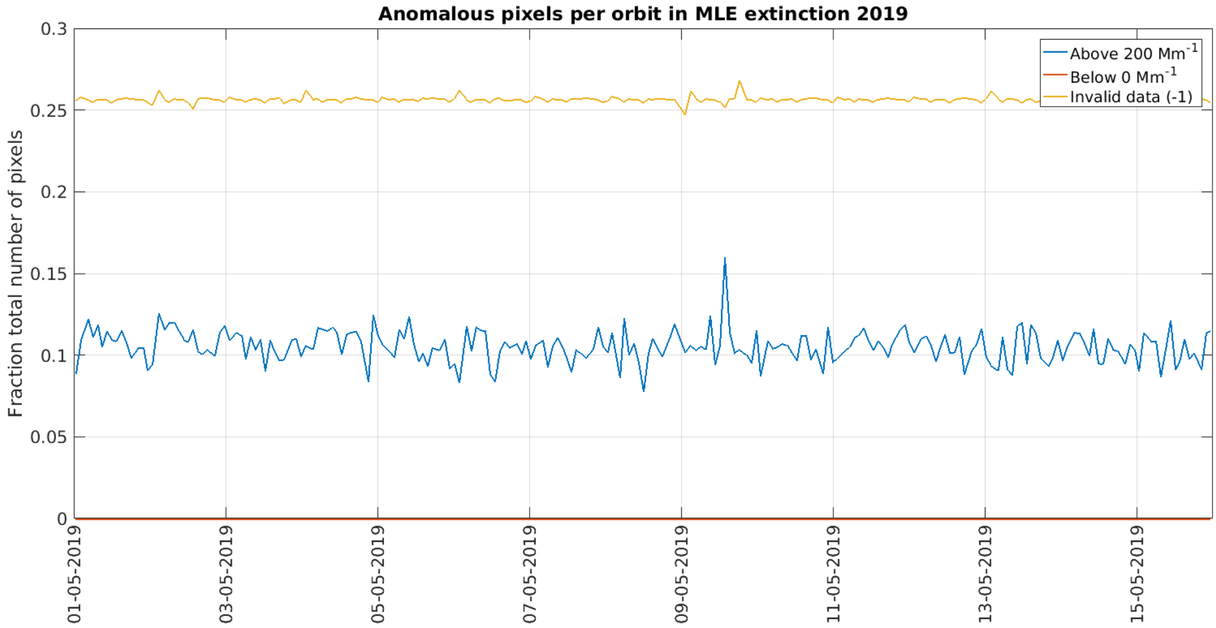


Figure 17: Fraction of MLE anomalous pixels in extinction (top) and backscatter (bottom) for 2018 dataset: outliers with high values (blue), with values <0 (orange) or not computed and set to -1 (yellow).



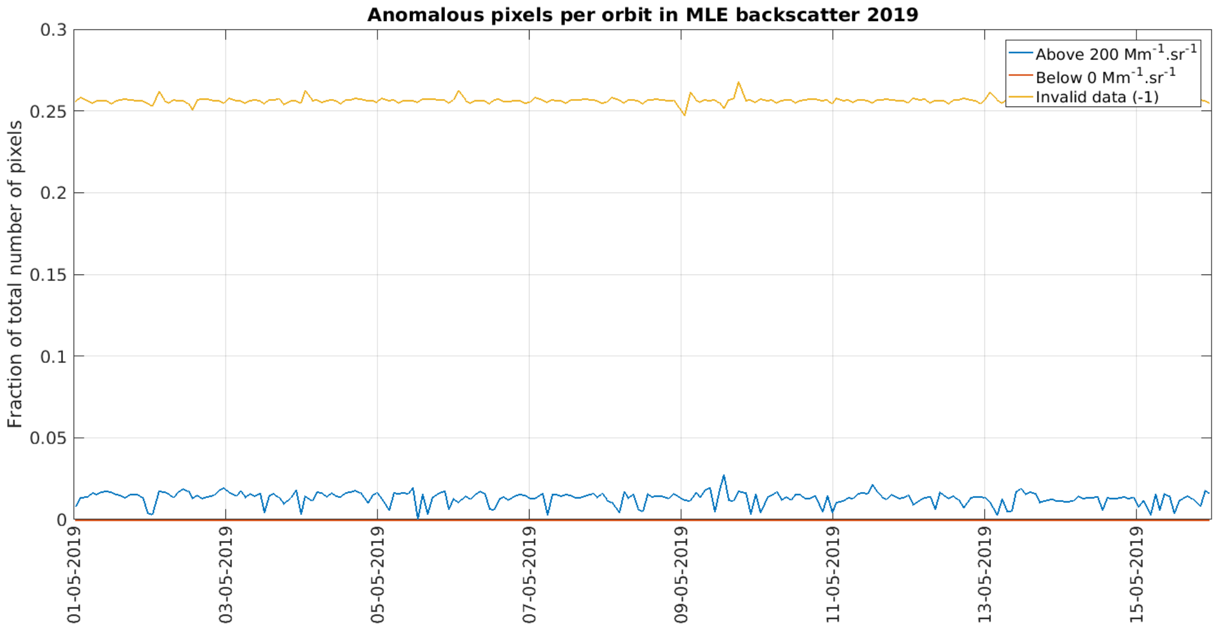


Figure 18: Fraction of MLE anomalous pixels in extinction (top) and backscatter (bottom) for 2019 dataset: outliers with high values (blue), with values <0 (orange) or not computed and set to -1 (yellow).

6.3.3 Outliers in SCA and MLE backscatter and extinction for L2AP V3.14.8 PDGS

The quality of the PDGS dataset has been assessed by measuring the mean absolute differences with SAND-BOX either for calibration coefficients and main L2A products.

Very low differences are then observed for both K_{ray} and K_{mie} (i.e. 0.00001% for 2018 and below 0.15 % for 2019), the highest differences in 2019 corresponding to truncated orbits.

Some differences between both datasets for SCA retrievals are observed but the mean absolute differences per profile rarely exceed 1e-4 m^{-1} for extinction and 1e-5 $m^{-1}.sr^{-1}$ for backscatter. The highest differences correspond to outliers in low SNR regions of the atmosphere.

As regards the MLE scores a significant number of profiles show mean absolute difference per profile above 1e-5 m^{-1} for extinction and 1e-6 $m^{-1}.sr^{-1}$ for backscatter mainly due to a positive bias occurring in low altitudes below 2 km.

6.4 Visual inspection of a few orbits: Orbits with peaks in anomalous pixel fraction

The Level-2A product shows limitations for specific conditions. It includes truncated orbits due to special operations and localized ones with topmost bin set in low altitudes. Examples of both cases are respectively shown below with truncated orbit #2833 (February 17, 2019) and orbit #4118 (May 09, 2019) with topmost bin set in low altitude. In each case the K_{ray} and K_{mie} provided per observation exceed the valid range because of non-optimized regression. Non-physical backscatter can then be observed.

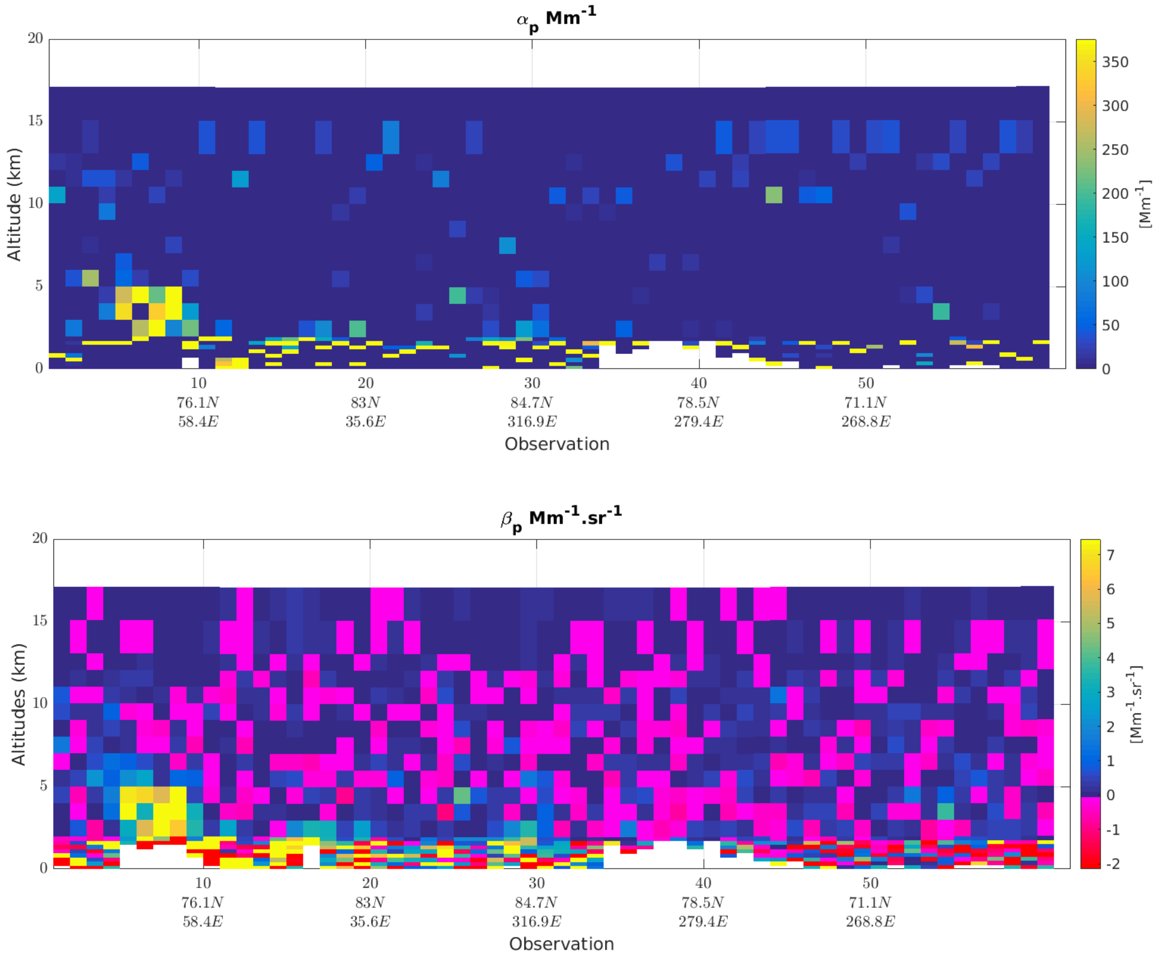


Figure 19: SCA Extinction and backscatter for particles of truncated orbit #2833.

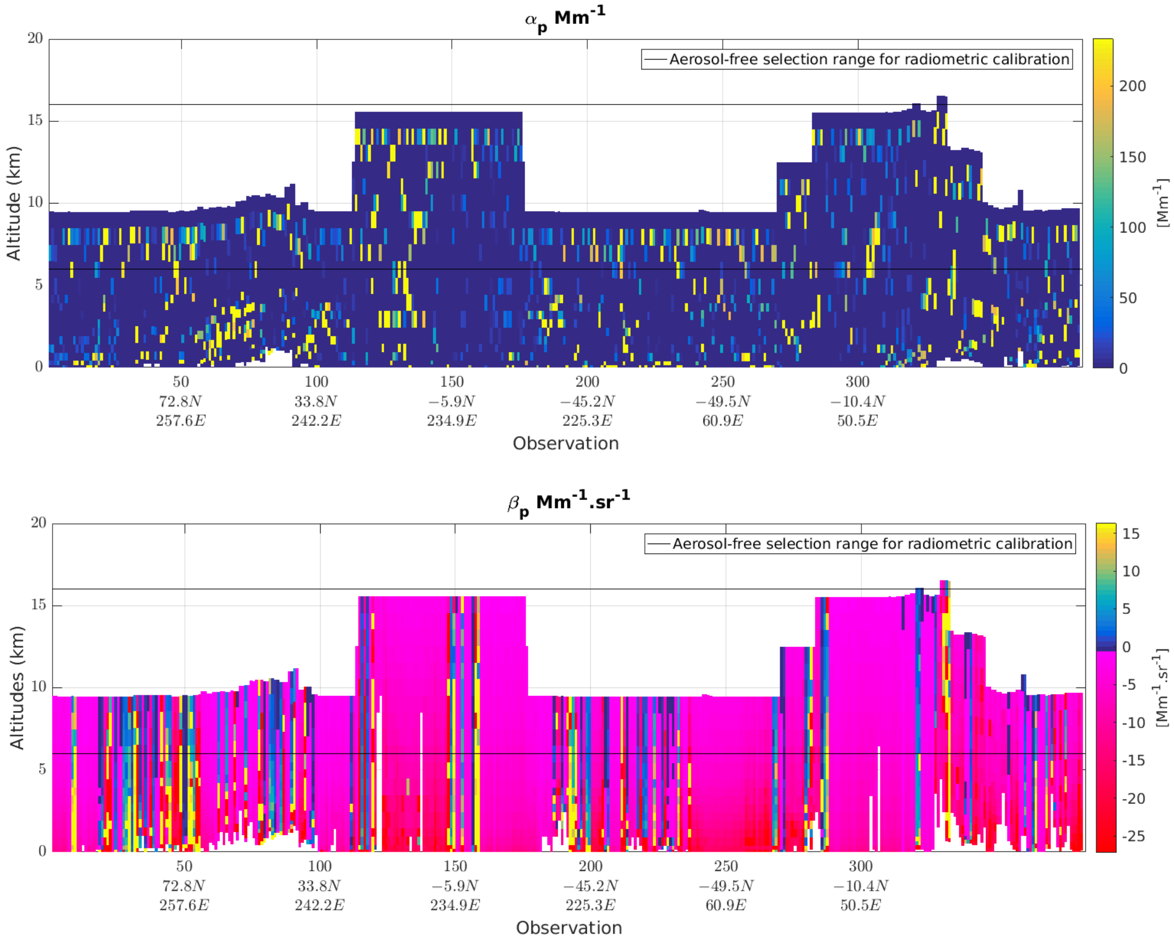


Figure 20: SCA Extinction and backscatter for particles of orbit #4118 with low topmost range bin (i.e. below 16km)

7 Verification of L2B products

7.1 General notes

There are several general notes regarding the data files that contain the Aeolus L2B horizontal line of sight (HLOS) wind speed observations. From the user's point of view, these notes are:

- Data block (.DBL) and the corresponding header (.HDR) files were provided.
- Sensing period (UTC): from 2018-09-03T10:38:38 to 2019-06-16T18:59:04.
- There was a major data gap between 14 January (last data before the gap is at 2019-01-14T01:41:49) and 15 February 2019 (first sensed data after the gap is at 2019-02-15T20:10:05).
- Total number of DBL (and HDR) files is 3643 pairs of files.
- There are files from two different version numbers. Their distribution is shown in Table 5.
- There is no pair of files (DBL and HDR) with duplicate versions.
- There is no data covering the months of July and August (i.e. no proper NH summer or SH-winter).

Table 5: Number of pairs of files for each file version.

Version number	0001	0002	Total
Number of files	3303	340	3634

Verification of Aeolus L2B HLOS wind speed product is done against its counterpart which is computed from the atmospheric fields produced by the Numerical Weather Prediction (NWP) model known as the ECMWF Integrated Forecast System (IFS). This information, which is provided in L2B product, is part of the NRT AUX_MET files, which were produced from the profiles of ECMWF IFS TcO1279 L137 background forecast along Aeolus predicted ground-tracks in NRT. L2B HLOS wind was collocated with the corresponding model fields and the results were compared against the collocations provided as part of L2B. The differences were found to be marginal.

The Aeolus L2B observations are classified into Mie clear, Mie cloudy, Rayleigh clear and Rayleigh cloudy. There are only 790 Mie clear wind observations during the whole period of more than 8 months. This number is an order of magnitude higher than that from the second reprocessing (61 over a longer period of 16 months). Irrespective of the fact the Mie clear data are in rather good agreement with the model (bias of 1 m/s, random error of 5 m/s and a correlation coefficient of 0.95), we believe these are based on noise and this class is not considered any further.

L2B observations were subject to minimal quality control to remove erroneous ones. The data should be valid according to the validity flag in L2B. The HLOS wind error estimate should not exceed the threshold error value for the specific channel where the threshold error value for Mie is 3 and for Rayleigh is 5. Finally, in order to reduce the adverse impact of outliers, any collocated pair with difference between L2B and model HLOS wind values exceeding 5 times the error threshold were removed. In general, the last two criteria were responsible for discarding less than 0.1% of the data.

7.2 L2B Mie cloudy winds

The probability density function (PDF) of Mie cloudy HLOS winds from the whole third reprocessed dataset is shown in Figure 21. PDF's of both Aeolus and ECMWF model (collocated with Mie cloudy observations) are shown. The Mie cloudy PDF shows very good agreement with that of the model. This is a significant improvement compared to the second reprocessing. Mie cloudy PDF shows slightly higher occurrences of extreme values compared to the model. It is difficult to tell which one is better at those regimes. It is expected that any numerical model, including the ECMWF model, reduces the "variability" which may explain the more conservative extreme values of the model. At the same time, the noise in any observing system, including Aeolus, increases variability which may explain the higher extremes of Mie cloudy HLOS winds.

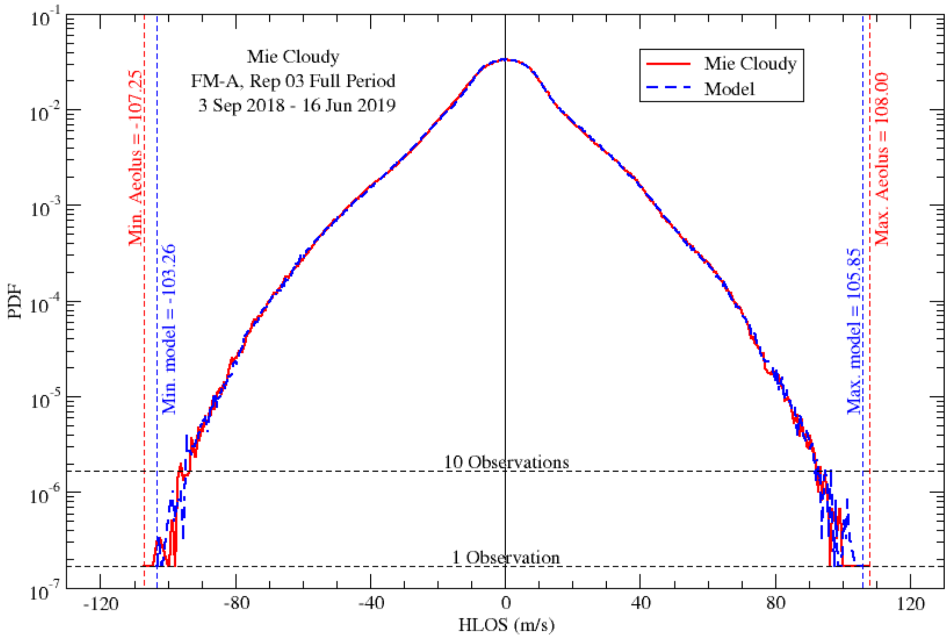


Figure 21: Probability density function (PDF) of Mie cloudy HLOS winds from the whole third reprocessed dataset. The corresponding PDF of ECMWF model collocated with Mie cloudy observations is also shown. The maximum and minimum (maximum of absolute negative values) for both Mie cloudy and the model are also shown.

Figure 22 shows the time series of the number of valid L2B Mie cloudy HLOS wind speed observations, and the bias and scaled (i.e. multiplied by 1.4826) median absolute difference (SMAD) between the reprocessed L2B and ECMWF IFS model. It should be noted that the SMAD, which in theory equals to the standard deviation of the difference (SDD) for normally distributed data, is a robust estimate of random deviation and less sensitive to outliers as in the case of SDD. The curves are smoothed by taking the running means over 30 successive files (orbits).

According to Figure 22, the number of L2B Mie cloudy wind observations that pass quality control in the third reprocessing is slightly lower than that of the second reprocessing. The number of valid observations, however, shows rapid decline during the first two months followed by mild gradual decline by time. The gradual loss in the atmospheric path signal is the main candidate responsible for this reduction. The repeated small drop in the data volume (biweekly dips) is due to moon-blinding flagging. The number of valid observations from ascending and descending orbits are almost equal as expected.

Figure 22 suggests that the global bias, compared to ECMWF model, is around -0.1 m/s (underestimation). However, individual orbits show higher biases. More than 95% of the orbits show biases within -0.45 m/s and +0.25 m/s. This is an improvement with respect to the NRT products and the second reprocessing. The wind biases of ascending and descending orbits are very close over the whole period with a slight tendency to diverge during March 2019.

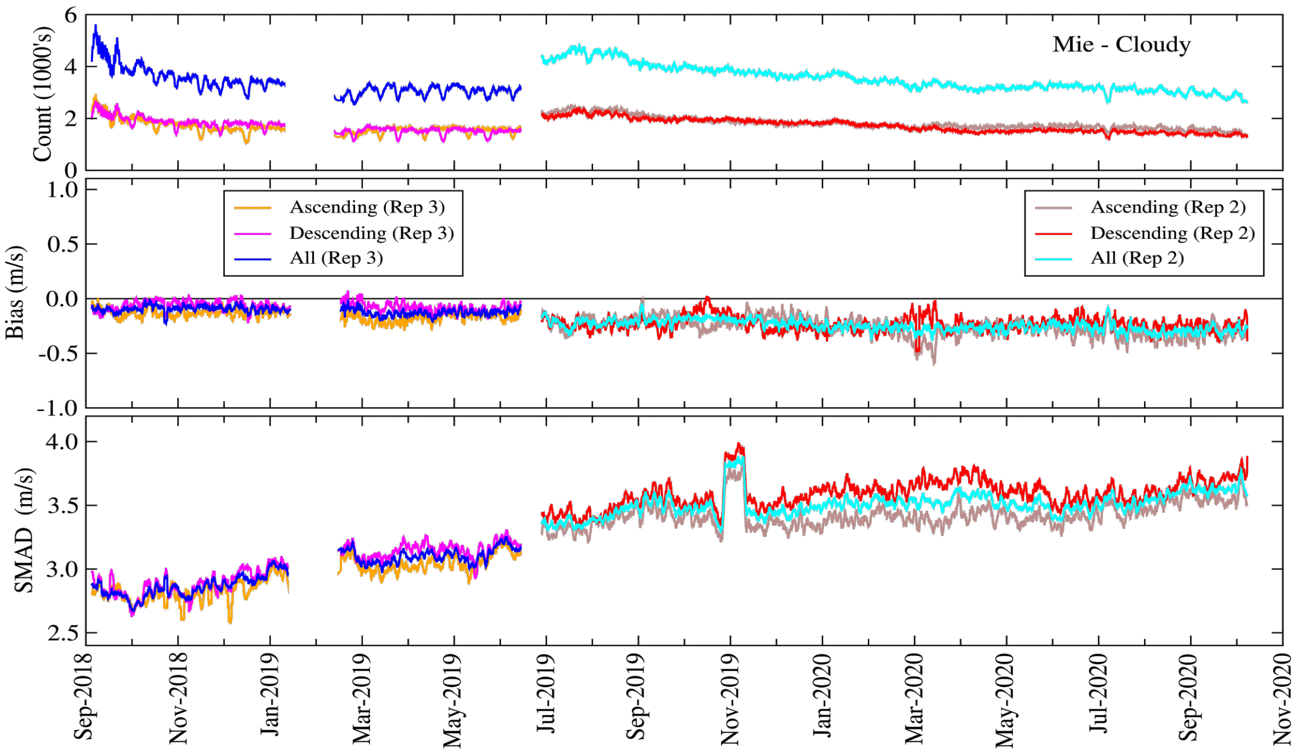


Figure 22: Time series of number of valid L2B Mie cloudy observations, HLOS wind speed bias, and scaled (i.e. multiplied by 1.4826) median absolute difference (SMAD) between L2B Mie cloudy and ECMWF IFS model winds. Running means along 30 successive files (orbits) are shown. Curves for all data as well as for ascending and descending orbits are shown for both the third (Rep 3) and the second (Rep 2) reprocessings.

The random differences between L2B and the model winds is represented by SMAD. The global SMAD values during the earlier stages vary just below 3.0 m/s. Afterwards, the scaled MAD gradually increases to reach values around 3.4 m/s as can be seen in the lower panel of Figure 22. All global SMAD values for the whole of first FM-A period (September 2018-June 2019) are lower than those of from FM-B period (with values more than 2.5 m/s). The standard deviation of the difference (SDD), which is another measure to the random error, follows a similar trend but at higher values (not shown). The gradual reduction in atmospheric path signal during this period is the one to be blamed for the increasing trend of random error.

Note that the random error (represented by the SMAD for example) at the beginning of FM-B period started at the same level as that of end of first FM-A period (lower panel of Figure 22).

As was the case in the second reprocessing (during the FM-B period), the data from ascending orbits show consistently lower SMAD values as can be clearly seen in the lower panel of Figure 22. This feature can be attributed to the greater cloud cover in ascending (18 local solar time) versus descending (06 local solar time) due to increased convection. However, the separation of the ascending and descending SMAD values has been supressed during the first FM-A period. This reduction can be attributed to the improved processing of Mie cloudy wind. Other reasons related to the instrument cannot be discarded.

The comparison between L2B Mie cloudy HLOS wind against its model counterpart is shown in Figure 23 in a form of a density scatter plot for all reprocessed data for this channel/class pair covering the whole globe during the whole period of reprocessing. The agreement between the ~12 million collocations of L2B and model wind is very good with a global bias of -0.10 m/s, SDD of 3.3 m/s and SMAD of 3.0 m/s. The correlation coefficient is 0.979. All Mie-cloudy wind statistics from the third reprocessing are better than their counterpart from the second reprocessing (for FM-B). Some other statistics are provided in the lower right corner of the plot. Note that most of the collocations fall on the symmetric line (the diagonal) as indicated by the orange/red/brown-coloured boxes. The circles and the crosses are very close to each other indicating that L2B and model winds have very close random errors.

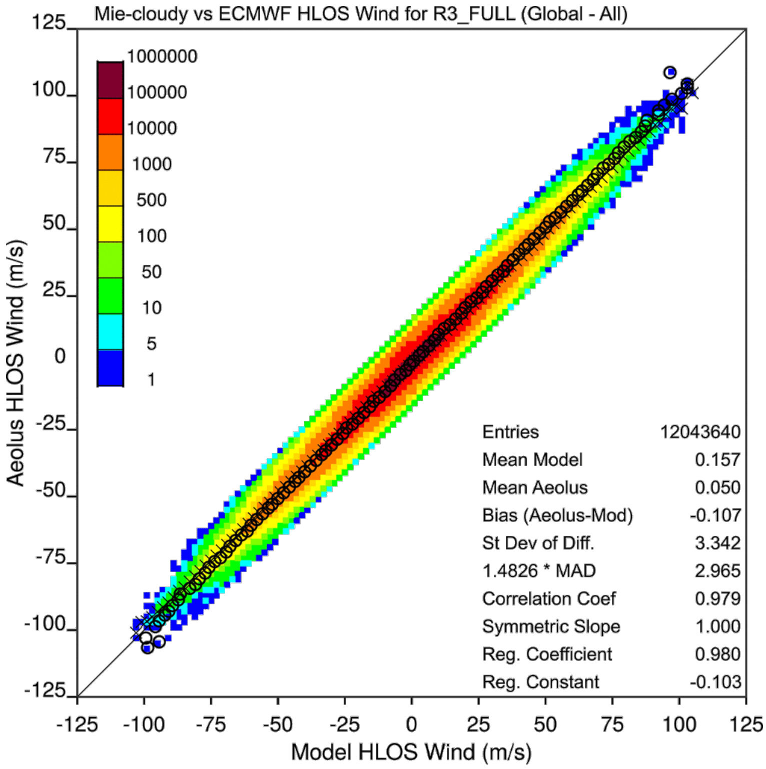


Figure 23: Density scatter plot comparing L2B Mie cloudy to ECMWF model HLOS winds over the whole globe during the whole period of the third reprocessing. Boxes are colour coded based on the number of collocations within the box. The circles are the average model value given L2B value while the crosses are the other way around. Statistics are given in the lower right corner of the plot.

The variation of the bias and the SDD between Mie cloudy HLOS wind and its model counterpart with respect to HLOS wind speed (the mean location of circles and crosses in the scatter plot of Figure 23) with respect to the HLOS wind values is shown in Figure 24. The variation of number of valid observations, bias and SDD (compared to the ECMWF model) as functions of HLOS wind for all third reprocessed dataset are shown in Figure 24. For comparison, the same from the second reprocessing are shown as well. The bias is reduced for the whole range of wind speeds with significant improvements for high wind speeds (faster than ~20 m/s). Thanks to the efforts of a number of the Aeolus DISC consortium members [RD-10], the “wobbly” shape of the bias curve in the second reprocessing has almost disappeared in the third reprocessing. The random error also reduced especially for winds slower than ~40 m/s.

The probability density function (PDF) and the cumulative distribution function (CDF) of Mie cloudy wind bias with respect to the ECMWF model for each individual orbit over the whole first FM-A period are shown in Figure 25. The bias is clearly more on the negative side. The CDF in Figure 25 clearly shows that more than 95% of the orbits showing a bias within -0.45 m/s and +0.25 m/s.

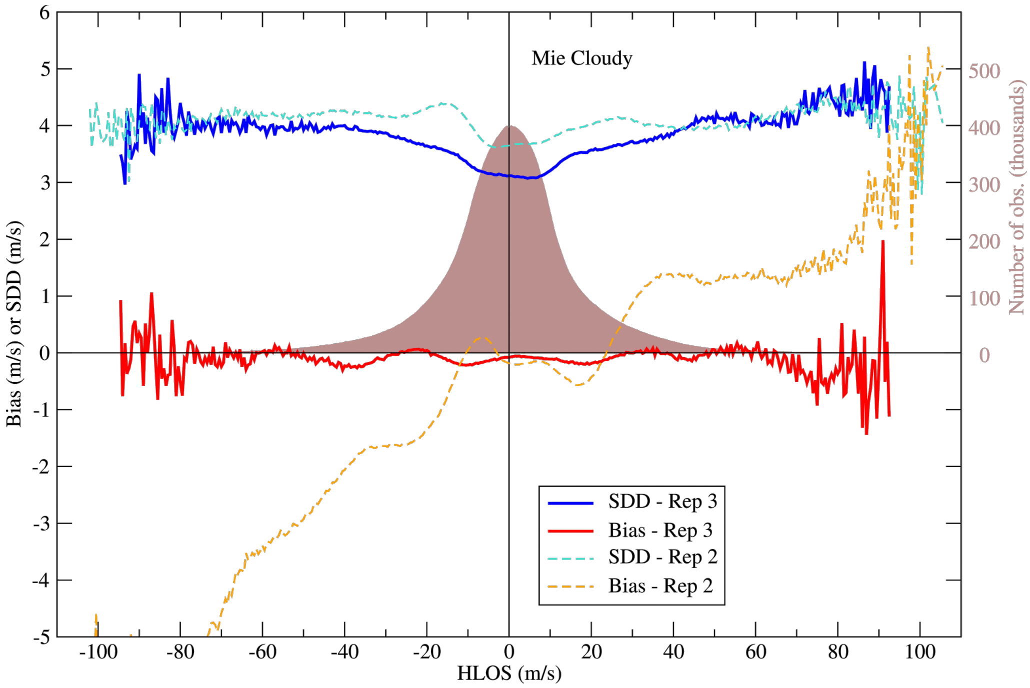


Figure 24: Number of valid Mie cloudy HLOS wind observations, bias and SDD (compared to ECMWF model) as functions of HLOS wind for the whole third reprocessed dataset.

The third reprocessed Mie cloudy dataset was discriminated based on the geographical location being in extratropical Northern Hemisphere (NH, with latitudes northern of 20°N), Southern Hemisphere (SH, with latitudes southern of 20°S) or in the Tropics (for latitudes between 20°N and 20°S). The scatter plots showing the comparison against the ECMWF model in each geographical region are shown in Figure 26. As expected, the Tropical winds are lower than those of the extra tropics. However, the agreement with the model in the Tropics is not as good as in the extra tropics. The differences in statistics between the NH and SH are very small.

Figure 27 shows the geographical distribution of the number of Mie cloudy observations within boxes of 3°X3° in latitudinal and longitudinal directions during the whole third reprocessing period. Note that boxes around the equator are the largest in size (in kilometres) with the size gets smaller moving towards the poles. As one would expect, deserts, including Antarctica, show themselves with small number of cloudy observations. The enhanced number of observations in the Tropics especially over the rainforests, cannot be missed.

Figure 28 and Figure 29 show the geographical distribution of HLOS wind speed bias and SDD, respectively, between Mie cloudy and ECMWF model for the whole third reprocessing dataset. Small negative biases dominate the global map in Figure 28. However, enhanced positive biases can be seen in the monsoon area, Antarctica and several mountain ranges. On the other hand, small SDD values dominate Figure 29 with enhancements over some of the mountain ranges, Intertropical Convergence Zone (ITCZ) and the monsoon area.

The whole third reprocessed L2B Mie cloudy dataset was discriminated based on the orbit whether ascending or descending. The resulting time series for ascending, descending or both (all) are shown in Figure 22 while the scatter plots of the ascending and descending orbits are shown in Figure 30. There is about 5% more data in the ascending orbits (left-hand side panel of Figure 30) compared to the descending orbits (left-hand side panel of Figure 30). The ascending orbit data fit slightly better to the model with SSD, SMAD and correlation coefficient (CC) of 3.31, 2.94 and 0.975, respectively, for the ascending compared to 3.37, 3.01 and 0.974, respectively, for the descending orbits. The differences between statistics for the third reprocessing are smaller than those from the second reprocessing.

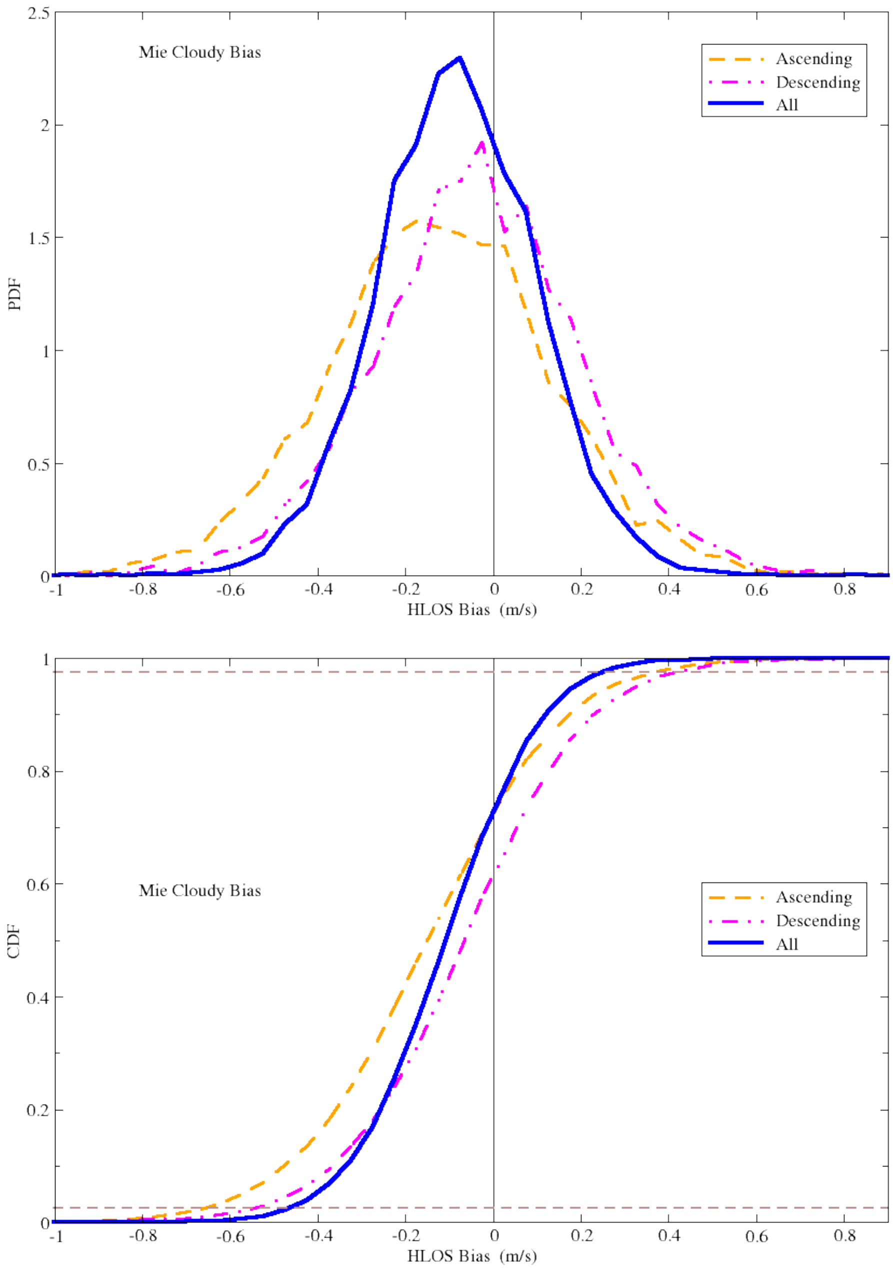


Figure 25: Probability density function (PDF; upper panel) and cumulative distribution function (CDF, lower panel) of Mie cloudy wind bias for the individual orbits over the whole first FM-A period. The horizontal thin dashed lines in the lower panel represent the 2.5% and 97.5% levels.

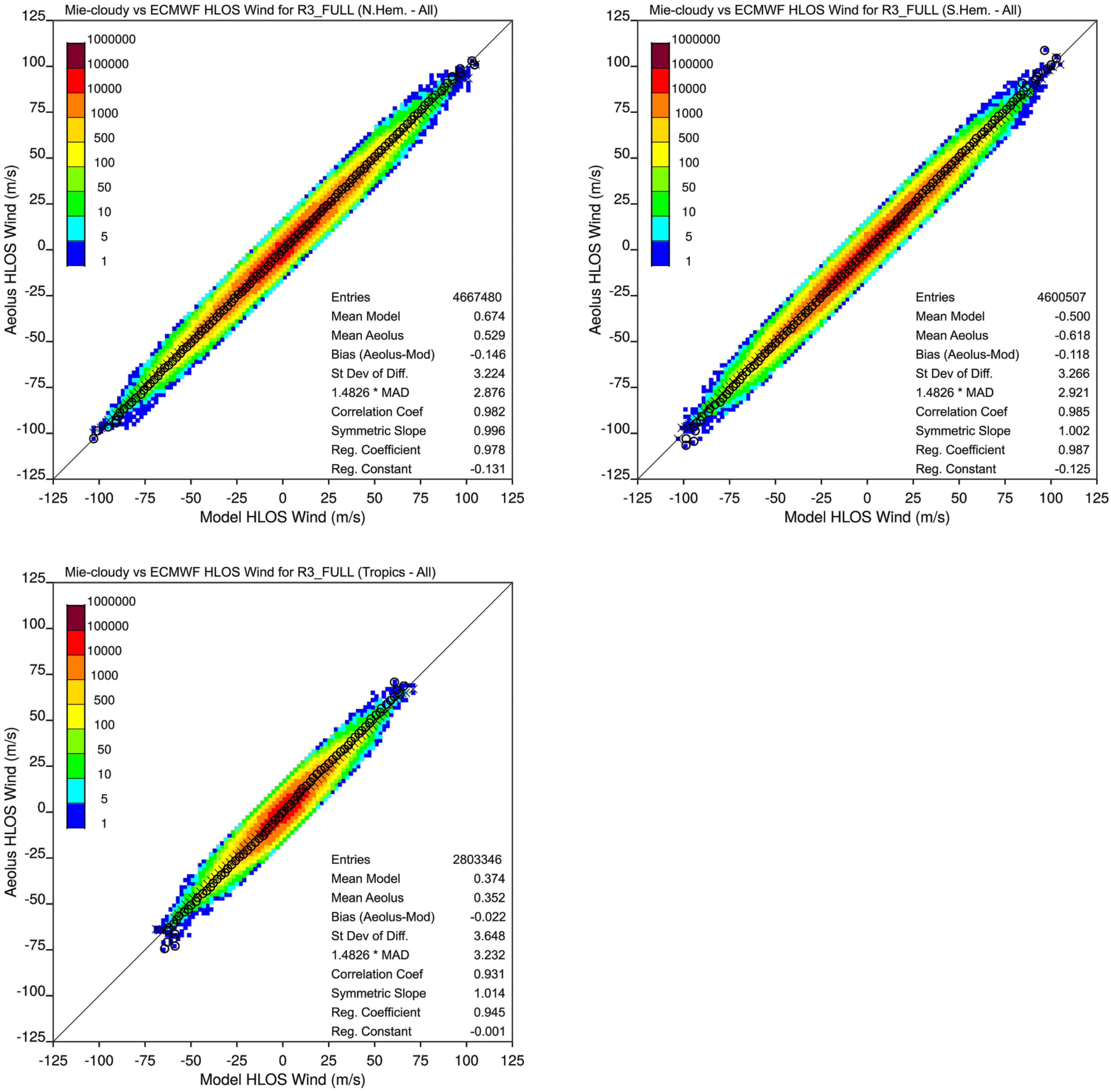


Figure 26: As in Figure 23 but with discrimination between extra tropics (upper) Northern Hemisphere (upper left) and Southern Hemisphere (upper right) and Tropics (lower). Refer to Figure 23 for the meaning of crosses, circles and colour-coding.

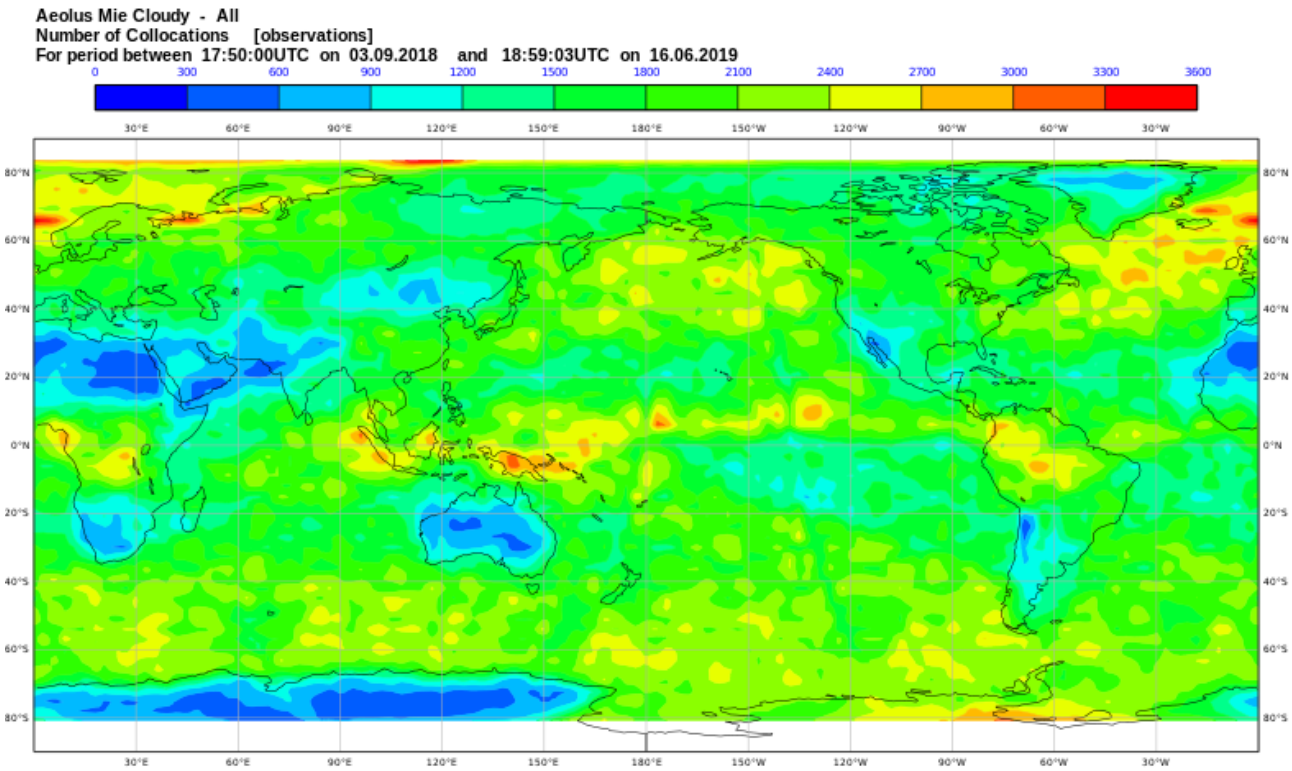


Figure 27: Geographical distribution of number of Mie cloudy L2B observations per 3°X3° latitude-longitude box during the whole third reprocessing.

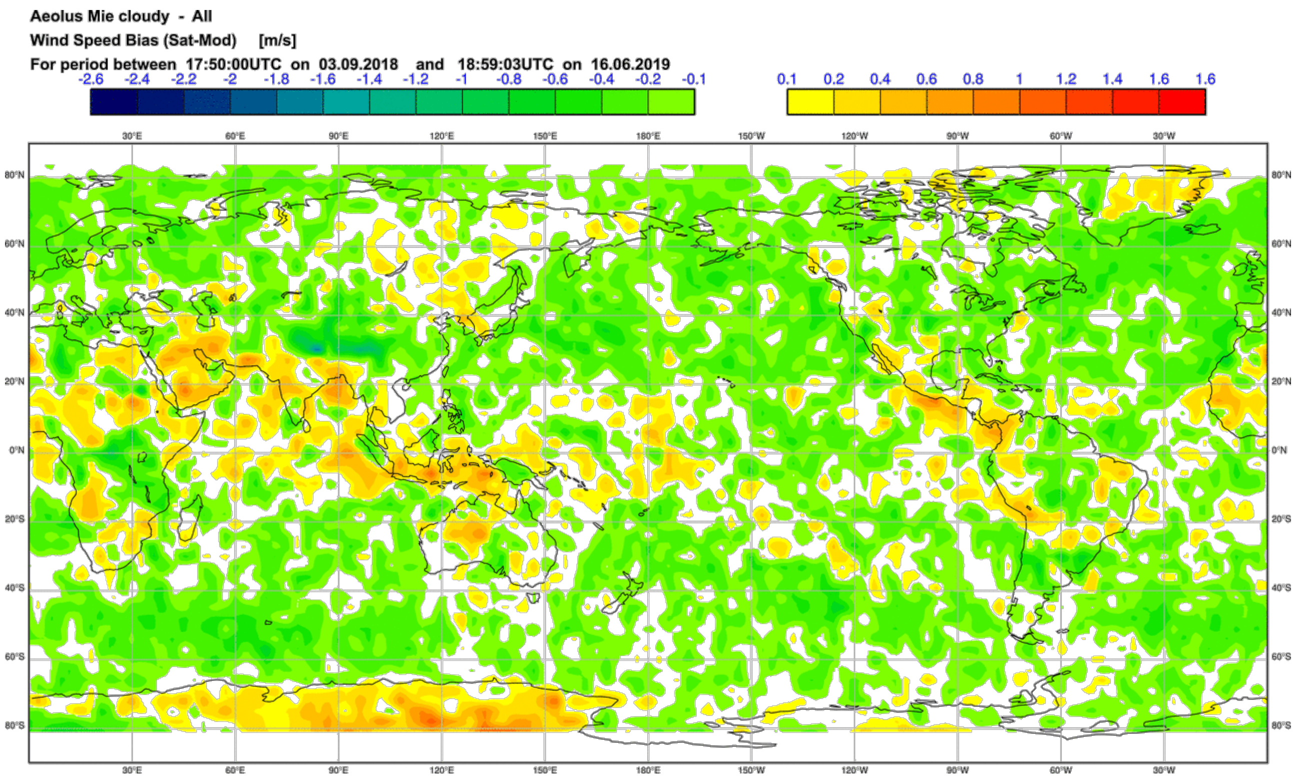


Figure 28: Geographical distribution of HLOS wind speed bias between Mie cloudy and ECMWF model for the whole third reprocessing dataset.

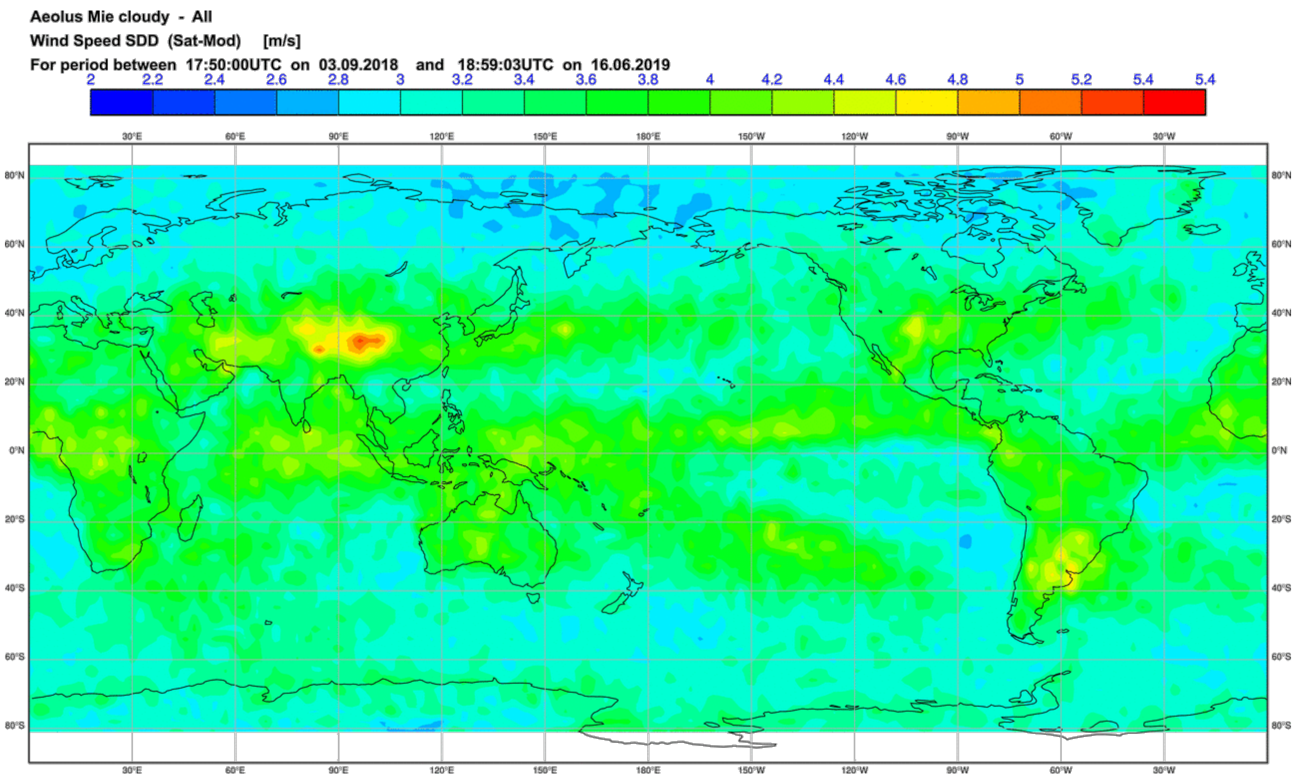


Figure 29: As in Figure 28 but for the Mie cloudy SDD.

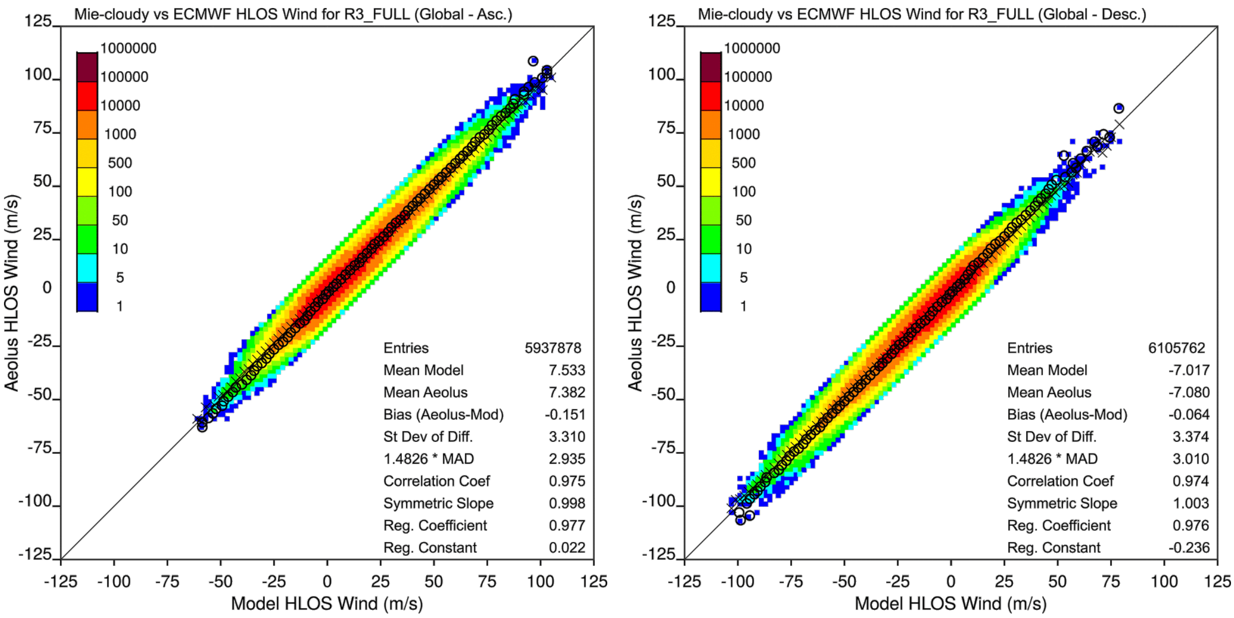


Figure 30: As in Figure 23 but with discrimination between ascending (left) and descending (right) orbits. Refer to Figure 23 for the meaning of crosses, circles and colour-coding.

Profiles of L2B Mie cloudy HLOS winds deviations from ECMWF model (bias, SDD and SMAD) together with the number of collocations at each altitude are shown in Figure 31. There was no Mie cloudy data above ~19-

km altitude (compared to ~27 km for FM-B). This may be attributed to the range-bin settings during FM-A operations.

Mie cloudy wind show slight positive bias compared to ECMWF model (of about 0.1 m/s) between altitudes 3 km and 15 km. The bias at those altitudes is almost altitude independent. The bias gradually increases to -0.5 m/s near the surface and +0.4 m/s at altitude 17 km. SMAD and SDD are smallest (2.5 m/s and 3.0 m/s, respectively) near the surface and increase (almost) linearly with altitude to reach 5.0 m/s (both) at altitude 17 km. The number of observations at both ends of the profile is rather small.

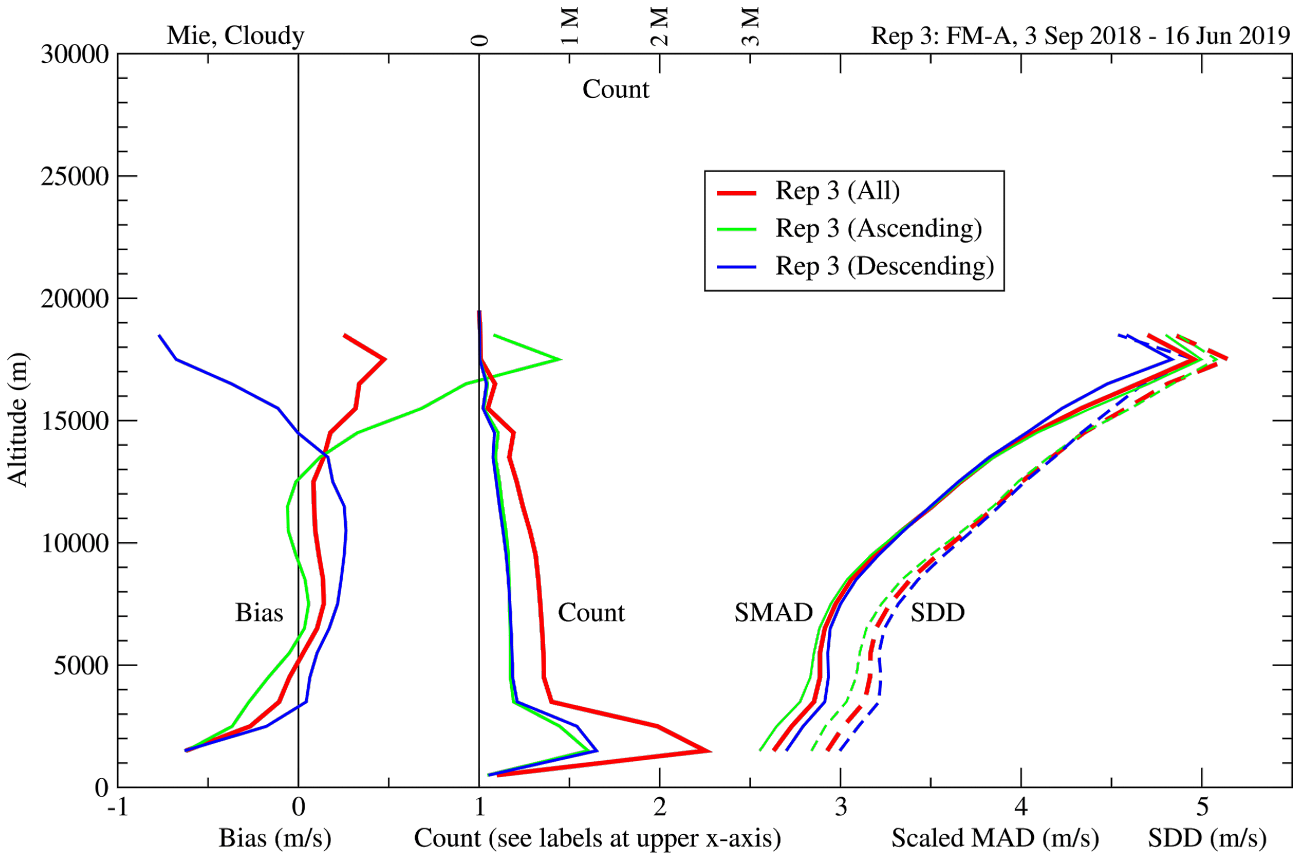


Figure 31: Profiles of L2B Mie cloudy HLOS wind deviations from ECMWF model (bias, SDD and SMAD) together with the number of collocations at each altitude. Profiles show discrimination between ascending and descending orbits in addition to all data.

7.3 L2B Rayleigh clear winds

The probability density function (PDF) of Rayleigh clear HLOS winds from the whole third reprocessed dataset is shown in Figure 32. PDF's of both Aeolus and ECMWF model (collocated with Rayleigh clear observations) are shown. Aeolus Rayleigh clear PDF show higher occurrences of extreme values compared to the model. It is difficult to tell which one is better at those regimes. It is expected that any numerical model, including the ECMWF model, reduces the "variability" which may explain the more conservative extreme values of the model. At the same time, the noise in any observing system, including Aeolus, increases variability which may explain the higher extremes of Rayleigh clear HLOS winds. It was also noted that the deviations between Rayleigh clear wind and the model in the third reprocessing are lower than those in the second reprocessing.

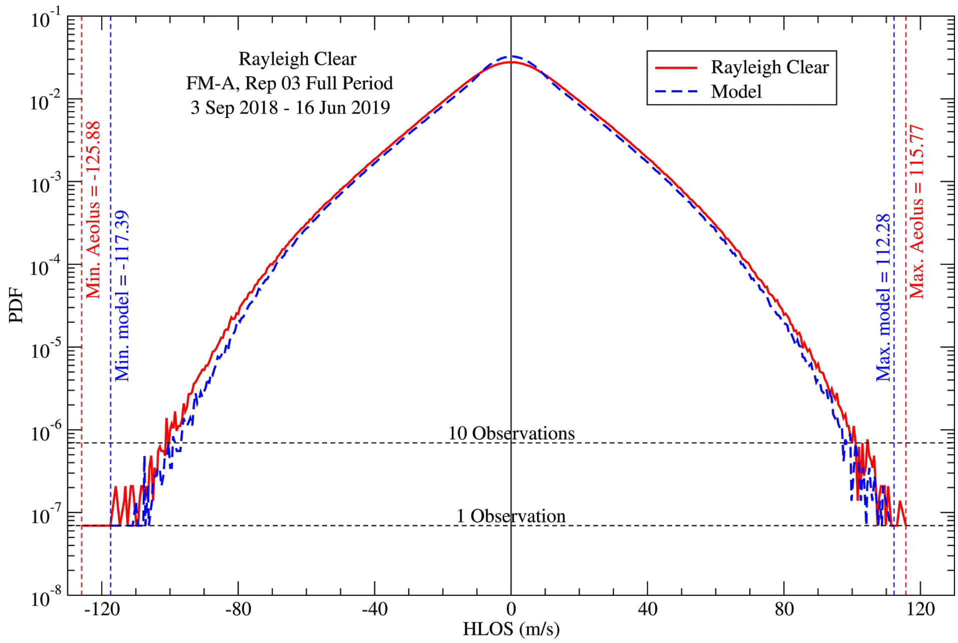


Figure 32: Probability density function (PDF) of Rayleigh clear HLOS winds from the whole third reprocessed dataset. The corresponding PDF of ECMWF model collocated with Rayleigh clear observations is also shown. The maximum and minimum (maximum of absolute negative values) for both Rayleigh clear and the model are also shown.

Figure 33 shows the time series of the number of valid L2B Rayleigh clear HLOS wind speed observations that passed the quality control procedure, and the bias and SMAD (scaled median absolute difference, MAD) between the second and third reprocessed L2B data from one side and the ECMWF IFS model from the other side over the whole period covered by the third reprocessing dataset. Only running means over 30 successive files (orbits) are shown.

Initially, the number of valid observations during each orbit reduced sharply during September 2018. After that, it stabilised until the data gap in January-February 2019. After the gap, the number of observations started slightly higher than that before the gap and started to reduce gradually at a small rate. The gradual loss in the atmospheric path signal is the main candidate responsible for this reduction. The repeated small drop in the data volume (biweekly dips) is due to moon-blinding flagging. The number of valid observations from ascending and descending orbits are almost equal as expected.

On average, the L2B Rayleigh clear wind bias compared to ECMWF model is virtually unbiased (fluctuates around 0.0 m/s). With few exceptions, the bias varies between -0.4 m/s and +0.4 m/s for individual orbits (see bias PDF and CDF curves later) and between -0.1 m/s and +0.1 m/s for the 30-orbit running means as shown in the middle panel of Figure 33. On clear exception with bias higher than 0.1 m/s (with individual orbit bias of about 1 m/s) can be witnessed towards the end of November 2018 as also shown in Figure 34. There was no recorded incident at that time.

There was a separation in the ascending and descending bias curves towards the end of May and start of June 2019 as can be clearly seen in Figure 33 and Figure 34. However, the difference is much lower than those seen the second reprocessing during October 2019 and March 2020 (Figure 33).

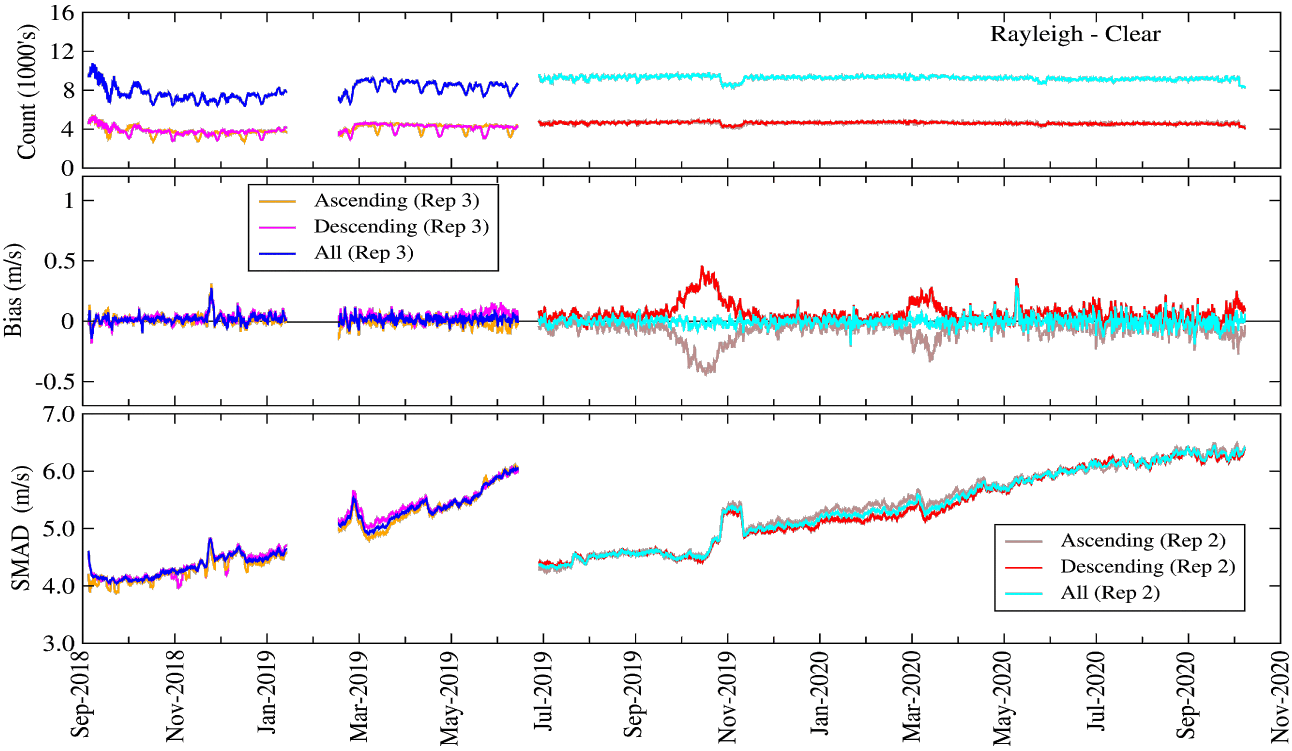
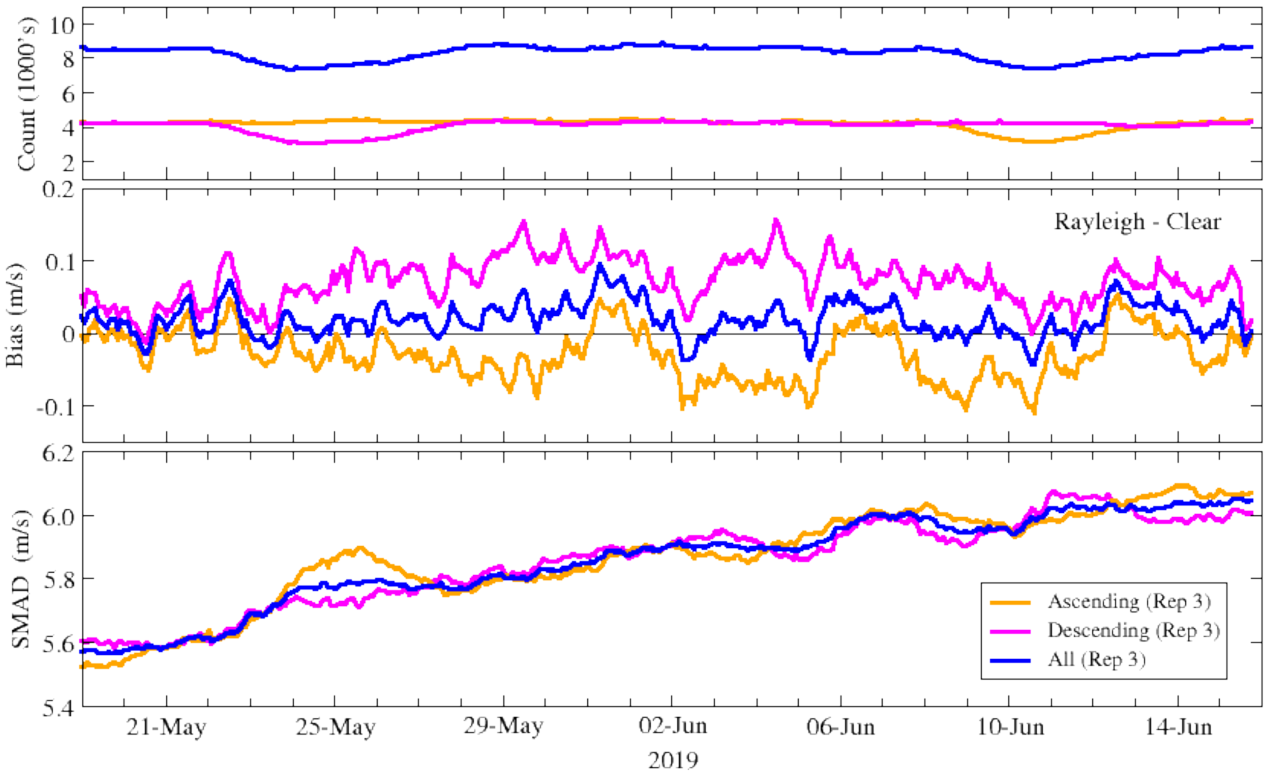


Figure 33: Time series of number of valid L2B Rayleigh clear wind observations, HLOS wind speed bias and SMAD between L2B Rayleigh clear and ECMWF IFS model winds. Running means along 30 successive files (orbits) are shown. Curves for all data as well as for ascending and descending orbits are shown for both the third (Rep 3) and second (Rep 2) reprocessings.

The lower panel of Figure 33 shows that SMAD initially varies at values slightly higher than 4.0 m/s before it started to gradually increase in an almost linear manner until it reached values higher than 6 m/s towards the end of the first FM-A period. The gradual loss of signal is responsible for this degradation. Focus on the periods with jumps in SMAD values are shown in Figure 34. Contrary to the FM-B data (from the second reprocessing), SMAD for the ascending orbits is consistently lower than that of the descending orbits. The difference, however, is not large except for March-April 2019. The SDD curves follow similar pattern (not shown) but at higher values.

Note that the random error (represented by SMAD for example) at the start of FM-B period started at value comparable to the start of the mission (see lower panel of Figure 33). The increasing trend of random error in FM-A is faster than that of FM-B.

The comparison between L2B Rayleigh clear HLOS wind against its model counterpart is shown in Figure 35 in a form of a density scatter plot for all reprocessed data for this channel/class pair covering the whole globe during the whole period of third reprocessing. The agreement between about 29 million collocations of L2B and model wind is rather good without any bias globally and with SDD and SMAD values of 5.9 m/s and 4.8 m/s, respectively. These global values are slightly better than those from the second reprocessing. The correlation coefficient is 0.945 which is slightly worse than that from the second reprocessing. Some other statistics are provided in the lower right corner of Figure 35. Note that most of the collocations fall on the symmetric line (the diagonal) as indicated by the orange/red/brown-coloured boxes. The circles and the crosses are very close to each other for the HLOS wind speed range of -70 m/s to +70 m/s indicating that L2B and model winds have very close random errors for most of the data range. However, this may not be hold for extreme values where the circles and crosses deviate from each other.



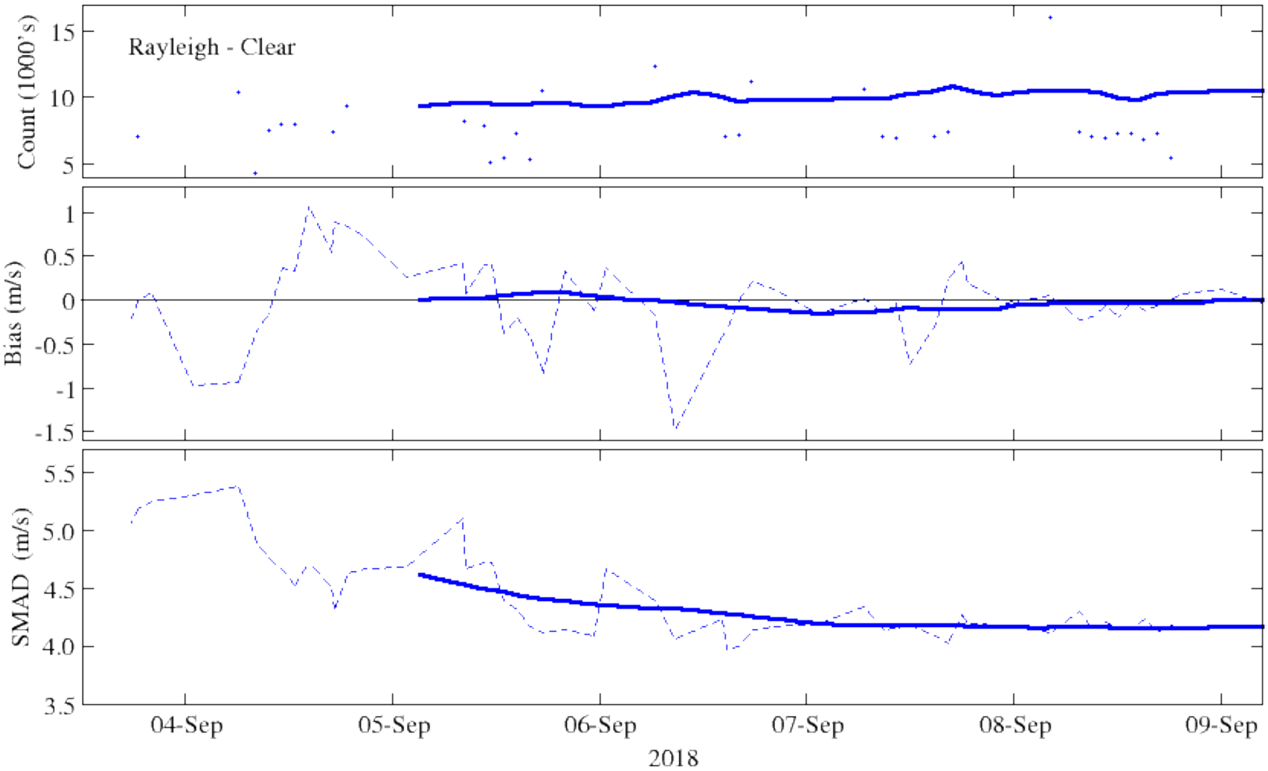
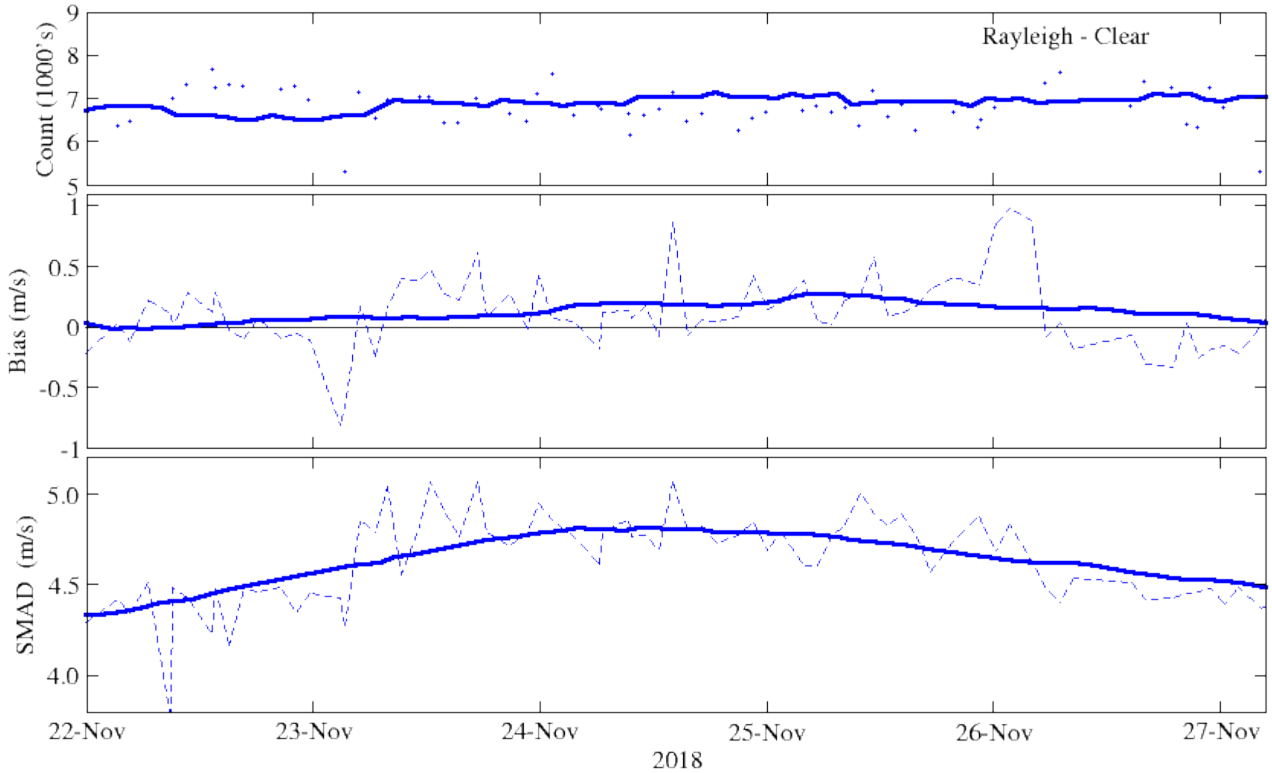


Figure 34: Same as in Figure 33 but focussed on periods with higher-than-normal bias or SMAD during the period of the third (Rep 3) reprocessing. Ascending and descending curves are not shown except for the first panel. Solid lines are for 30-orbit running means and dashed lines are for individual orbits.



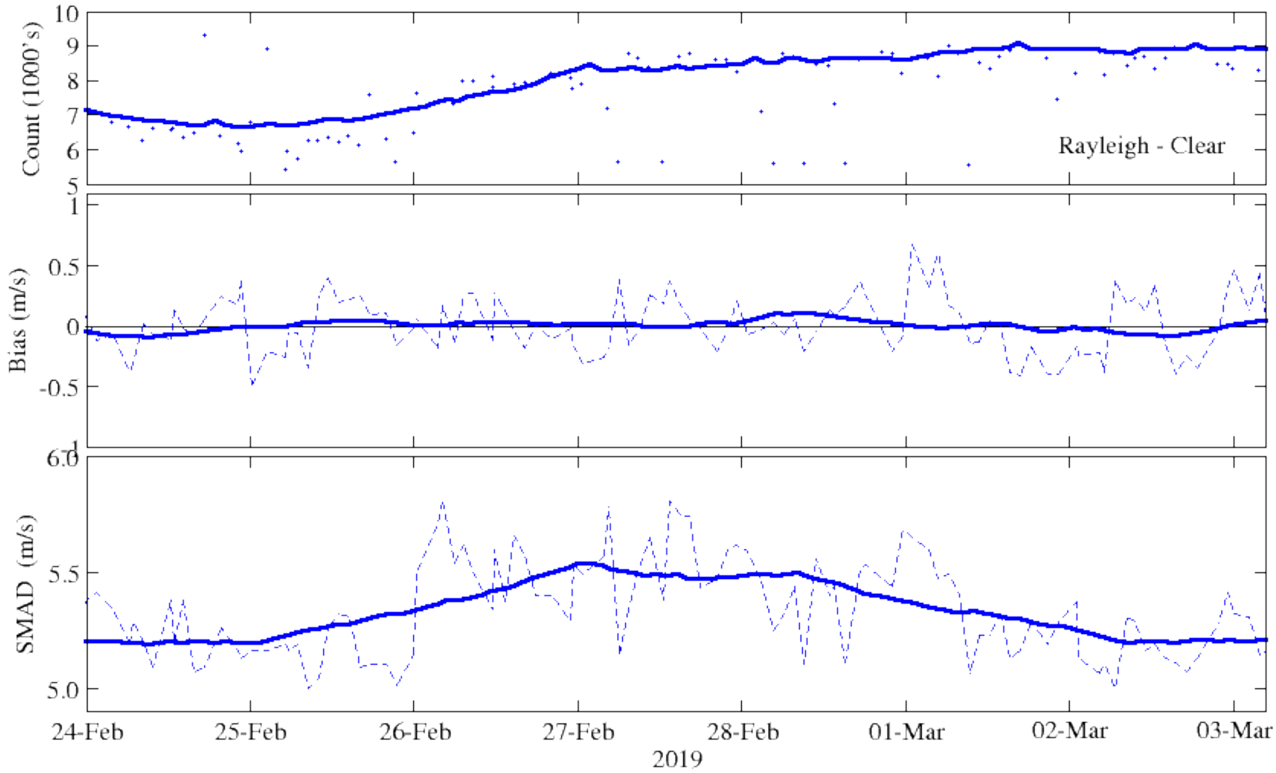


Figure 34: Continued.

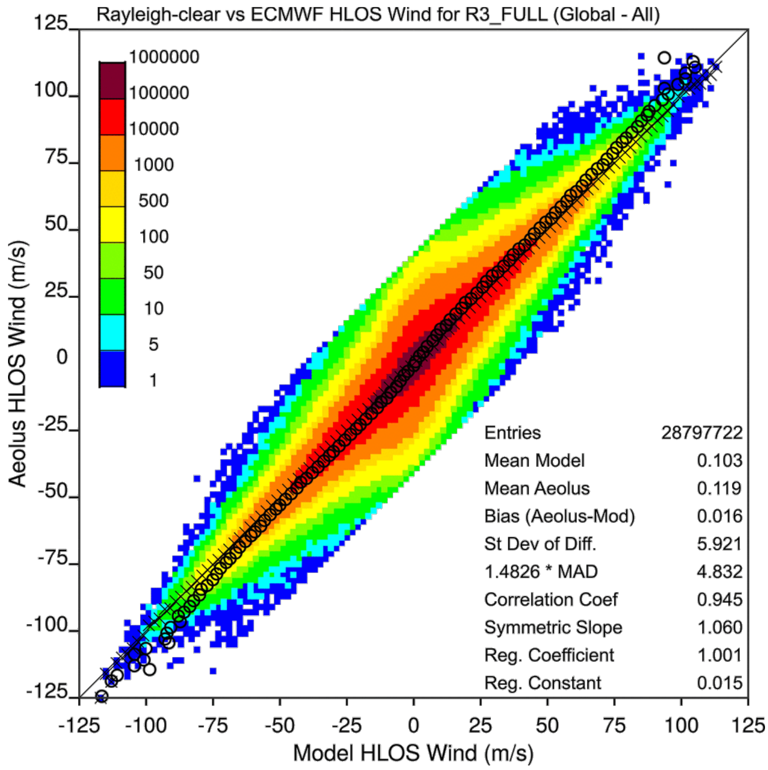


Figure 35: Density scatter plot comparing L2B Rayleigh clear to ECMWF model HLOS winds over the whole globe during the whole period of the third reprocessing. Boxes are colour coded based on the number of collocations within the box. The circles are the average model value given L2B value while the crosses are the other way around. Statistics are given in the lower right corner of the plot.

Figure 36 shows the variation of number of valid observations, bias and SDD (compared to ECMWF model) as functions of HLOS wind for the whole third reprocessed dataset. Both bias and SDD curves show smooth and symmetric behaviour. The bias is rather small (less than 2 m/s) for most of the HLOS winds (-60 m/s to +60 m/s). Rayleigh clear HLOS winds show overestimation with respect to the model (positive bias for positive winds and vice versa for negative winds). Even the bias of 2 m/s suggested by Figure 36 for HLOS values in excess of 10 m/s (in either direction) seems an artifact of the assumption that both Rayleigh clear and model winds have equal errors as can be seen in Figure 37. The actual bias is much smaller. Only extreme wind speeds (more than ~100 m/s) show bias higher than 4 m/s. Note that the number of those cases is rather small. The third reprocessing shows an improvement in biases of high winds compared to the second reprocessing. The SDD is almost constant for a wide range of wind speeds (-80 to 90 m/s). The SDD tends to increase for higher winds. The superiority of the third reprocessing over the second one in terms of SDD is clear for all wind speeds faster than 20 m/s.

The probability density function (PDF) and the cumulative distribution function (CDF) of Rayleigh clear wind bias with respect to the ECMWF model for each individual orbit over the whole first FM-A period are shown in Figure 38. The bias is clearly more on the positive side. The CDF in Figure 38 clearly shows that more than 95% of the orbits showing a bias within -0.4 m/s and +0.4 m/s.

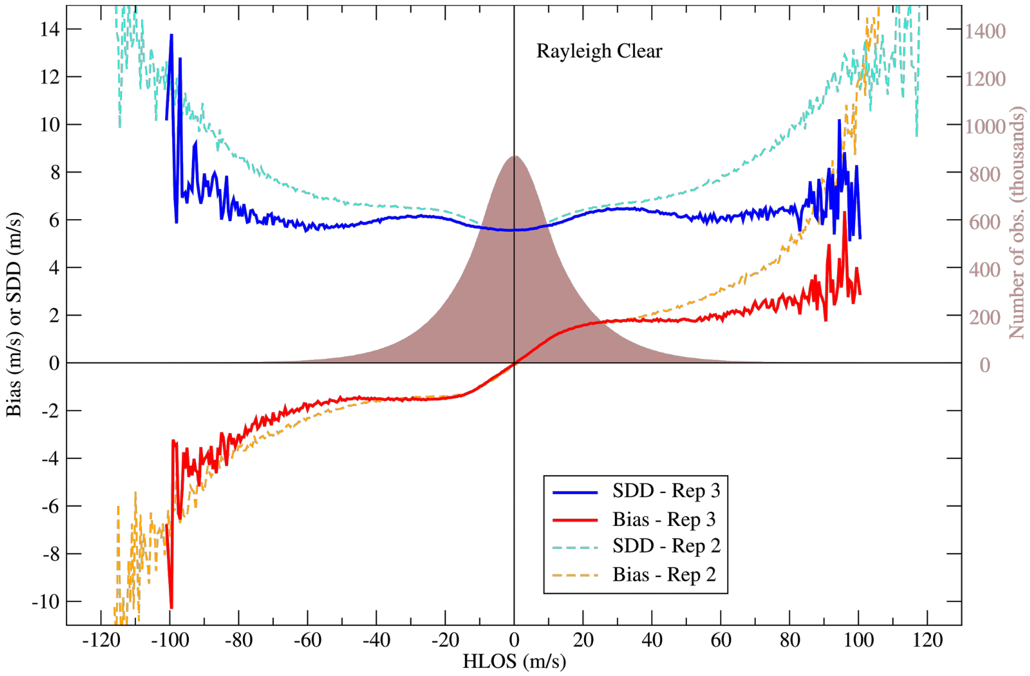


Figure 36: Number of valid Rayleigh clear HLOS wind observations, bias and SDD (compared to ECMWF model) as functions of HLOS wind for the whole third reprocessed dataset.

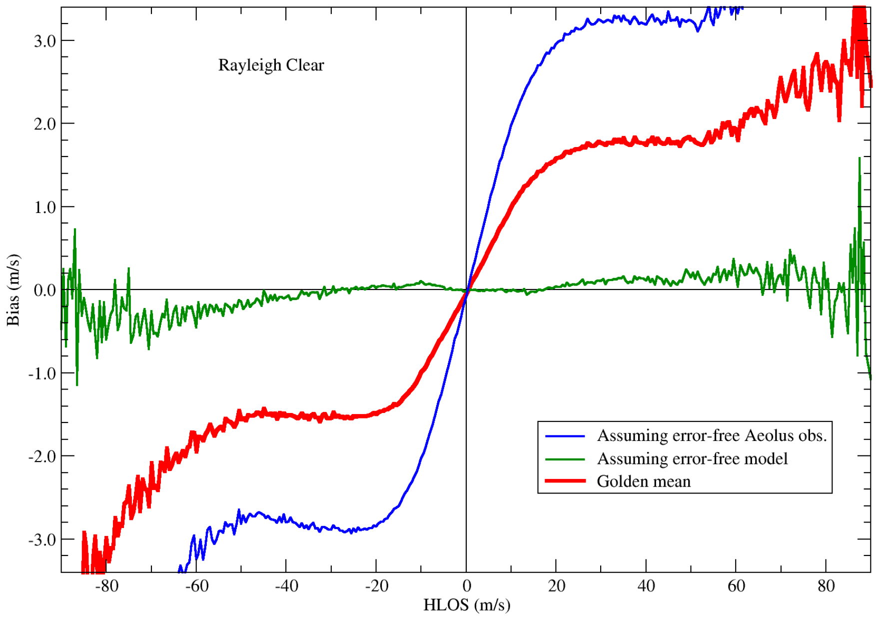


Figure 37: Rayleigh clear HLOS wind bias compared to ECMWF model as functions of HLOS wind for the whole third reprocessed dataset assuming no error in Aeolus data (blue), no error in the model (green) or equal errors in both (red), which is the same bias curve in Figure 36.

The third reprocessed Rayleigh clear dataset was discriminated based on the geographical location being in extratropical Northern Hemisphere (NH; northern of 20°N), Southern Hemisphere (SH; southern of 20°S) or in the Tropics (between latitudes 20°N and 20°S). The scatter plots showing the comparison against the ECMWF model in each geographical region are shown in Figure 39. As expected, the Tropical winds are lower than those of the extra tropics. However, the agreement with the model in the Tropics is not as good as in the extra tropics (compare the correlation coefficients). The distribution of the cloud of boxes in the scatter plots in Figure 39 especially for the Tropics shows skewness at the middle of the cloud. A secondary distribution appears mainly during the seasons with low winds. It is high likely that this is a model issue. This feature was also noticed in the previous reprocessings.

Figure 40 shows the geographical distribution of the number of Rayleigh clear observations within boxes of 3°X3° in latitudinal and longitudinal directions during the whole third reprocessing period. Note that boxes around the equator are the largest with the size gets smaller moving towards the poles. The distribution of the observations is rather uniform over the whole globe with batches of enhanced numbers.

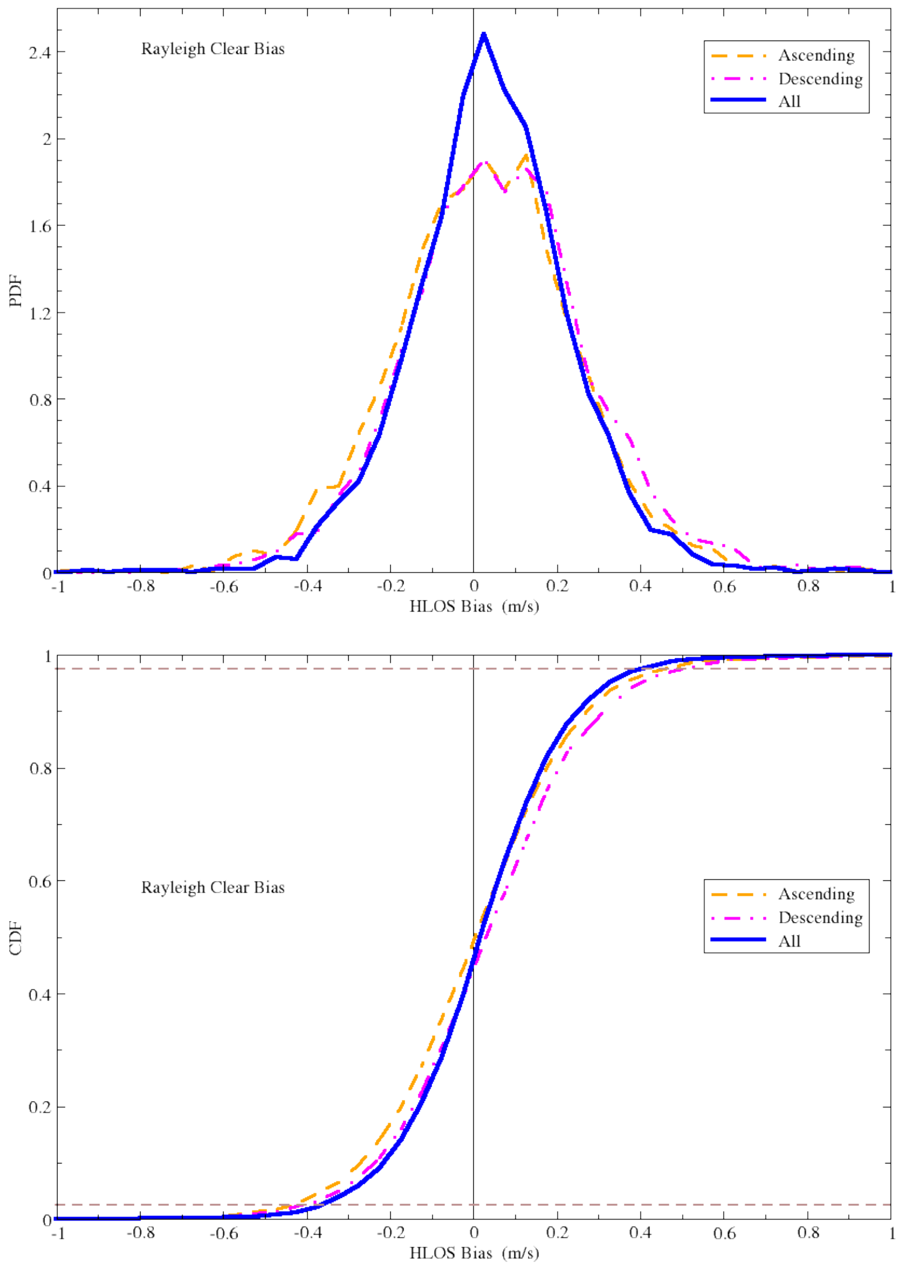


Figure 38: Same as Figure 25 but for Rayleigh clear wind bias.

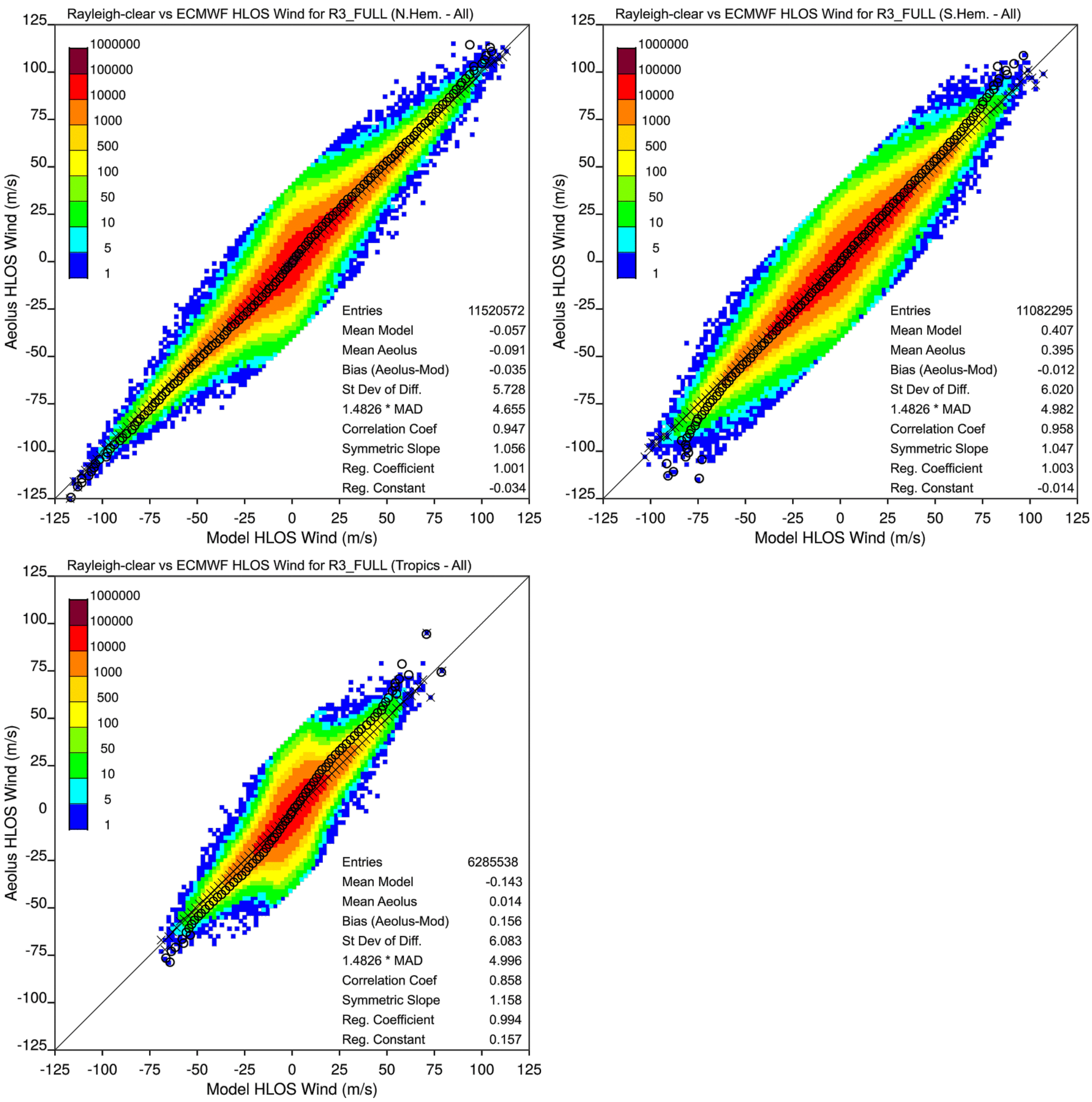


Figure 39: As in Figure 35 but with discrimination between extra tropics (upper) Northern Hemisphere (upper left) and Southern Hemisphere (upper right) and Tropics (lower). Refer to Figure 35 for the meaning of crosses, circles and colour-coding.

Figure 41 and Figure 42 show the geographical distribution of HLOS wind speed bias and SDD, respectively, between Rayleigh clear and ECMWF model for the whole third reprocessing dataset. The bias, as can be seen from Figure 41, is small everywhere with clear distinction between regions with alternating bias signs. The bias is mainly positive in the region bounded by latitudes 20°N and 20°S. The region bounded by latitudes 20°N and 60°N is dominated by negative bias. The same applies to the region bounded by latitudes 20°S and 40°S. Positive bias dominates the regions extending northern of latitude 60°N and the region between latitudes 40°S and 70°S with negative bias dominates the region southern of latitude 70°S.

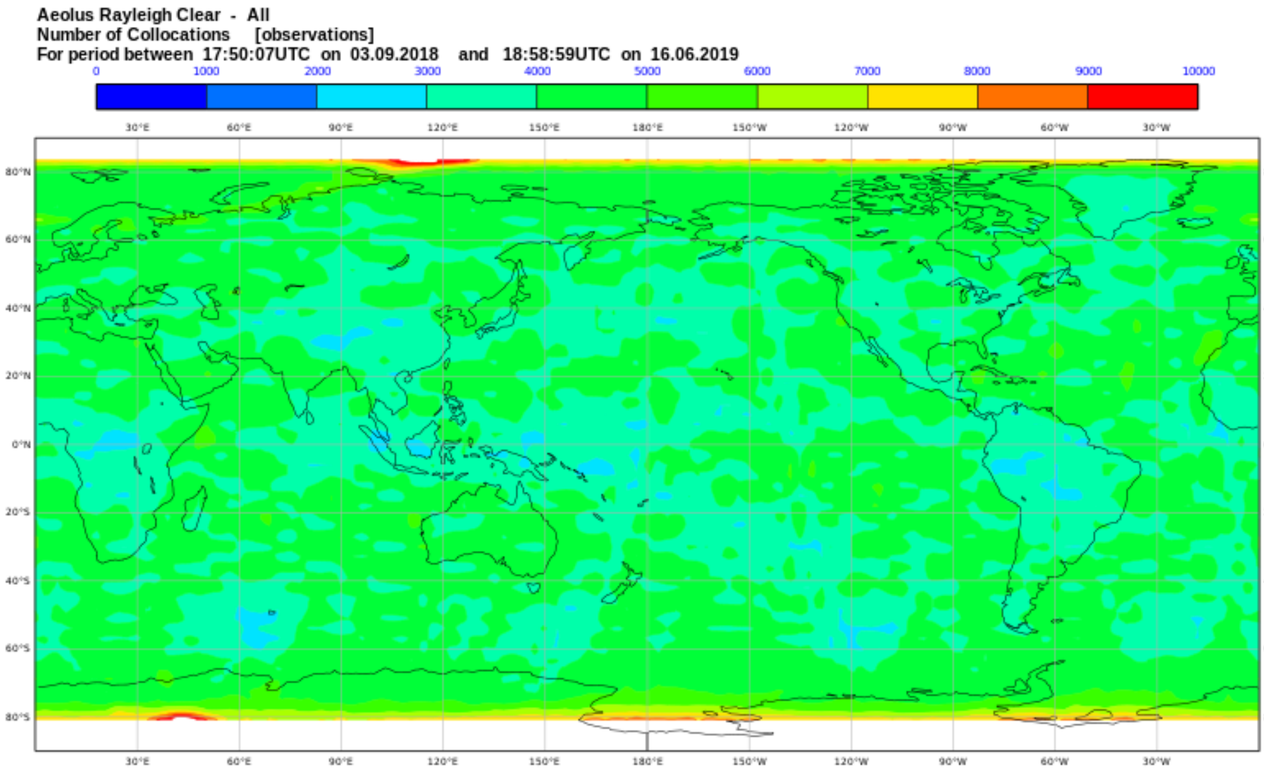


Figure 40: Geographical distribution of number of Rayleigh clear L2B observations per 3°X3° latitude-longitude box during the whole third reprocessing.

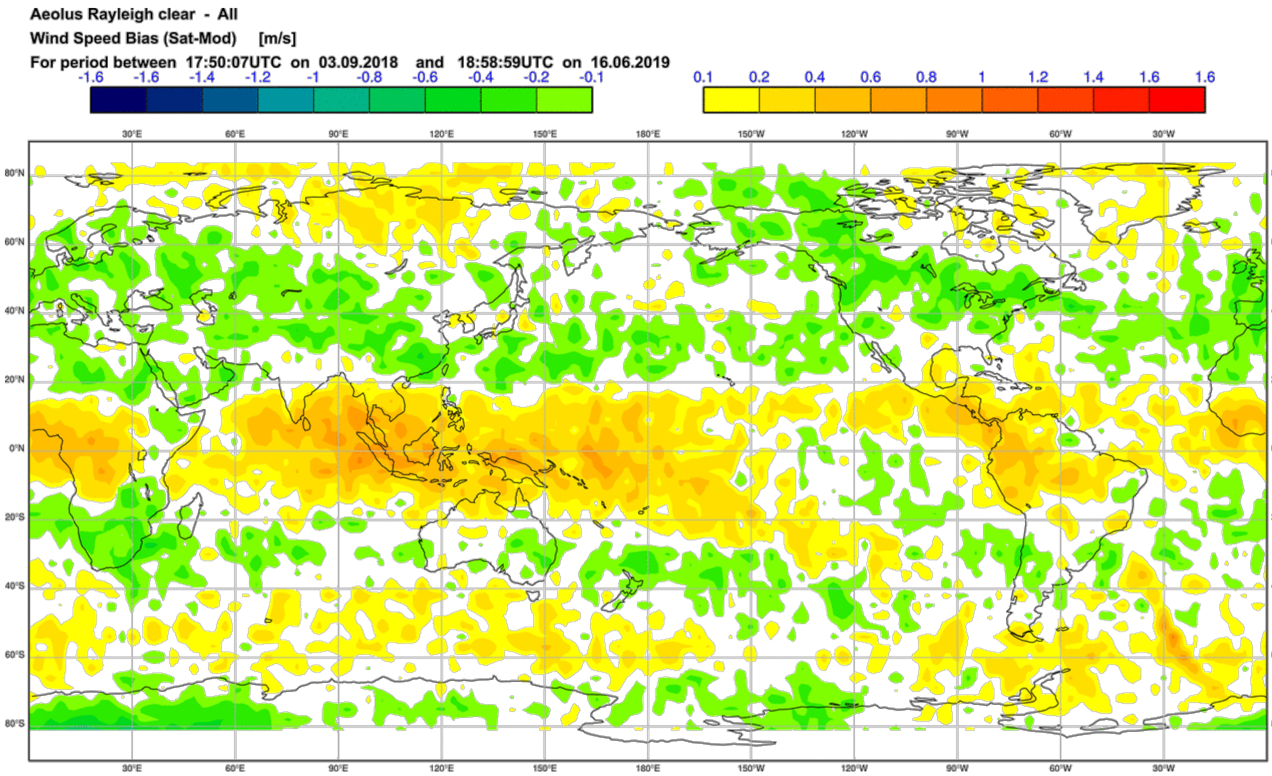


Figure 41: Geographical distribution of HLOS wind speed bias between Rayleigh clear and ECMWF model for the whole third reprocessing dataset.

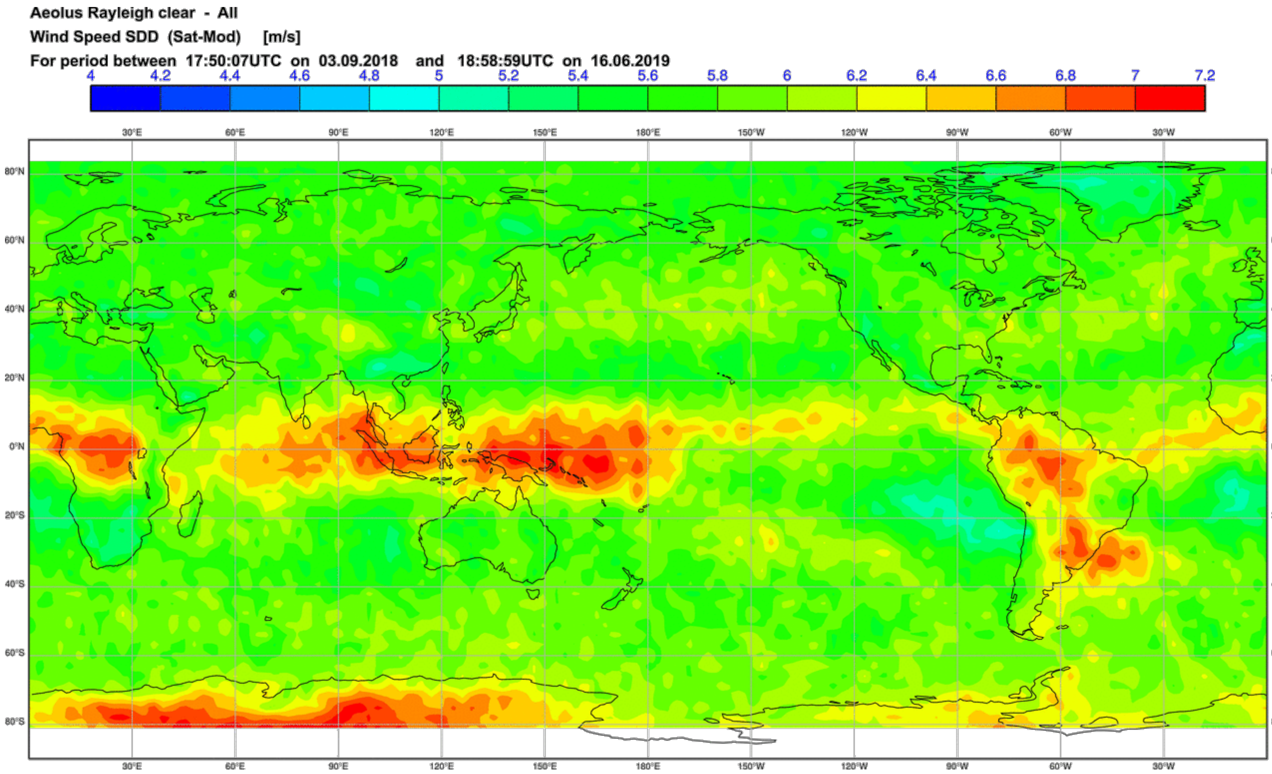


Figure 42: As in Figure 28 but for the Rayleigh clear SDD.

The SDD is rather high with enhanced value over the monsoon region (Tropics), Antarctica and most of South America as can be seen in Figure 42. Smaller SDD values can be seen in the Intertropical Convergence Zone (ITCZ) and over Greenland.

The whole third reprocessed L2B Rayleigh clear dataset was discriminated based on the orbit direction. The resulting time series for ascending, descending or both (all) are shown in Figure 33 above while the scatter plots of the ascending and descending orbits are shown in Figure 43. The number of valid observations during ascending and descending orbits are almost equal. There is almost no difference between the statistics from both orbit directions with descending orbits show marginally better fit to the model.

Profiles of L2B Rayleigh clear HLOS winds deviations from ECMWF model (bias, SDD and SMAD) together with the number of collocations at each altitude are shown in Figure 44. The bias is almost zero up to altitude of 18 km. Beyond that altitude, a linear bias increase can be seen as it goes below -0.5 m/s at altitude 26 km. The higher biases that can be seen at higher altitudes are less significant due to the small volume of valid observations at those altitudes.

The overall SMAD and SDD follow a C-shaped curve. SMAD and SDD show minimum values of 4.0 m/s and 5.0 m/s, respectively, between altitudes 4 km and 12 km. Outside this band, SMAD and SDD increase rapidly and exceed 9 m/s near the surface and above altitude 24 km. The difference between the ascending and descending orbits is very small with a slight advantage in favour of descending orbits.

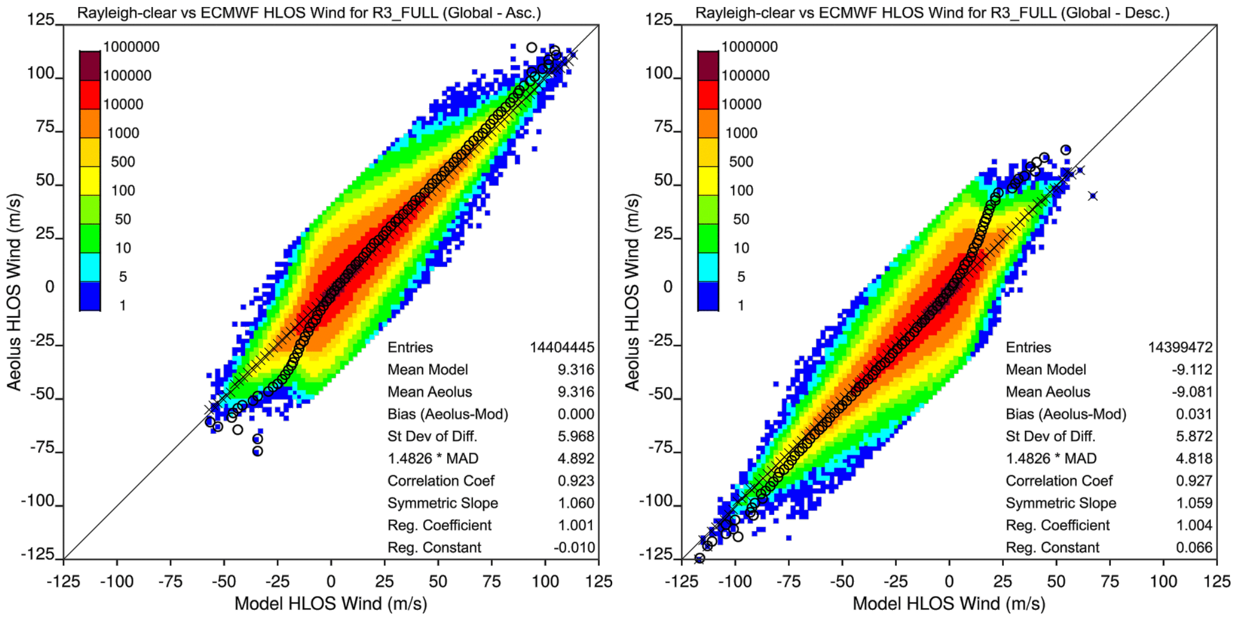


Figure 43: As in Figure 35 but with the discrimination between ascending (left panel) and descending (right panel) orbits. Refer to Figure 35 for the meaning of crosses, circles and colour-coding.

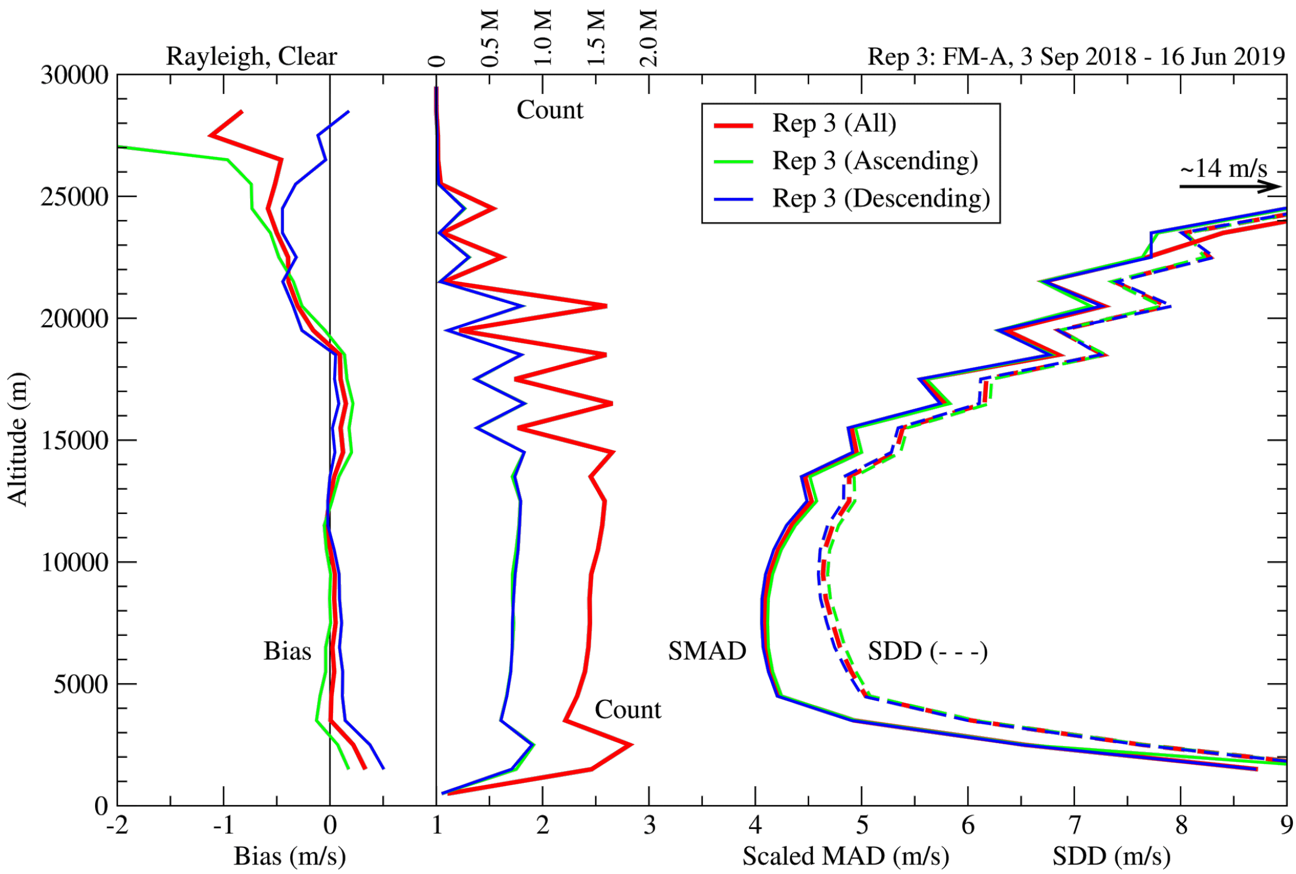


Figure 44: Profiles of L2B Rayleigh clear HLOS wind deviations from ECMWF model (bias, SDD and SMAD) together with the number of collocations at each altitude. Profiles show discrimination between ascending and descending orbits in addition to all data. SDD and SMAD values at the top of the profile, where very little data available, are around 14 m/s.

7.4 L2B Rayleigh cloudy winds

The probability density function (PDF) of Rayleigh cloudy HLOS winds from the whole third reprocessed dataset is shown in Figure 45. PDF's of both Aeolus and ECMWF model (collocated with Rayleigh cloudy observations) are shown. Aeolus Rayleigh cloudy PDF show higher occurrences of extreme values compared to the model. It is difficult to tell which one is better at those regimes as explained earlier.

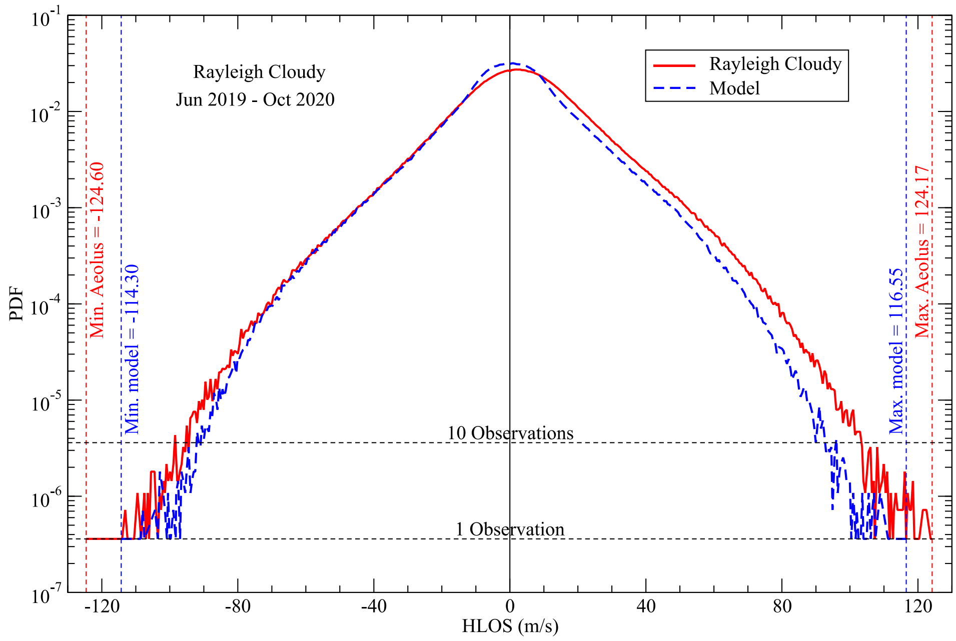


Figure 45: Probability density function (PDF) of Rayleigh cloudy HLOS winds from the whole third reprocessed dataset. The corresponding PDF of ECMWF model collocated with Rayleigh cloudy observations is also shown. The maximum and minimum (maximum of absolute negative values) for both Rayleigh cloudy and the model are also shown.

Figure 46 shows the time series of the number of valid L2B Rayleigh cloudy HLOS wind speed observations, and the bias and SMAD (scaled median absolute difference, MAD) between the reprocessed L2B (both the second and the third reprocessings) and ECMWF IFS model. Only running means over 30 successive files (orbits) are shown.

As it is clearly seen in upper panel of Figure 46, at the start of the mission the number of valid Rayleigh cloudy observations are the highest ever. The data volume continued to decrease gradually till the end of the first FM-A period. The repeated small drop in the data volume (biweekly dips) is due to moon-blinding flagging. The number of valid observations from ascending and descending orbits are almost equal as expected.

As can be seen in the middle panel of Figure 46, the (running) mean bias fluctuates around +1.0 m/s at the start of the mission. The bias then dropped to zero before started to increase as it exceeded +1.0 m/s in January 2019 before the data gap. After the gap, the bias dropped to about +0.5 m/s.

Biases from ascending and descending orbits are very close to each other. However, the bias from ascending orbits is consistently lower than that of descending ones. This is the same behaviour as in the. second reprocessing.

According to the lower panel of Figure 46, SMAD varied initially below 6.0 m/s before it started to increase gradually to exceed 6.0 m/s. There were two short jumps of SMAD. The first happened towards the end of November 2018 (same as seen in Rayleigh clear wind) and the second happened in February 2019 after the data gap.

The lower panel of Figure 46 also suggests that Rayleigh cloudy HLOS wind from descending orbits agrees better with the model compared to the ascending orbits. The SMAD of FM-B started almost at the same level reached at the end of the first FM-A period.

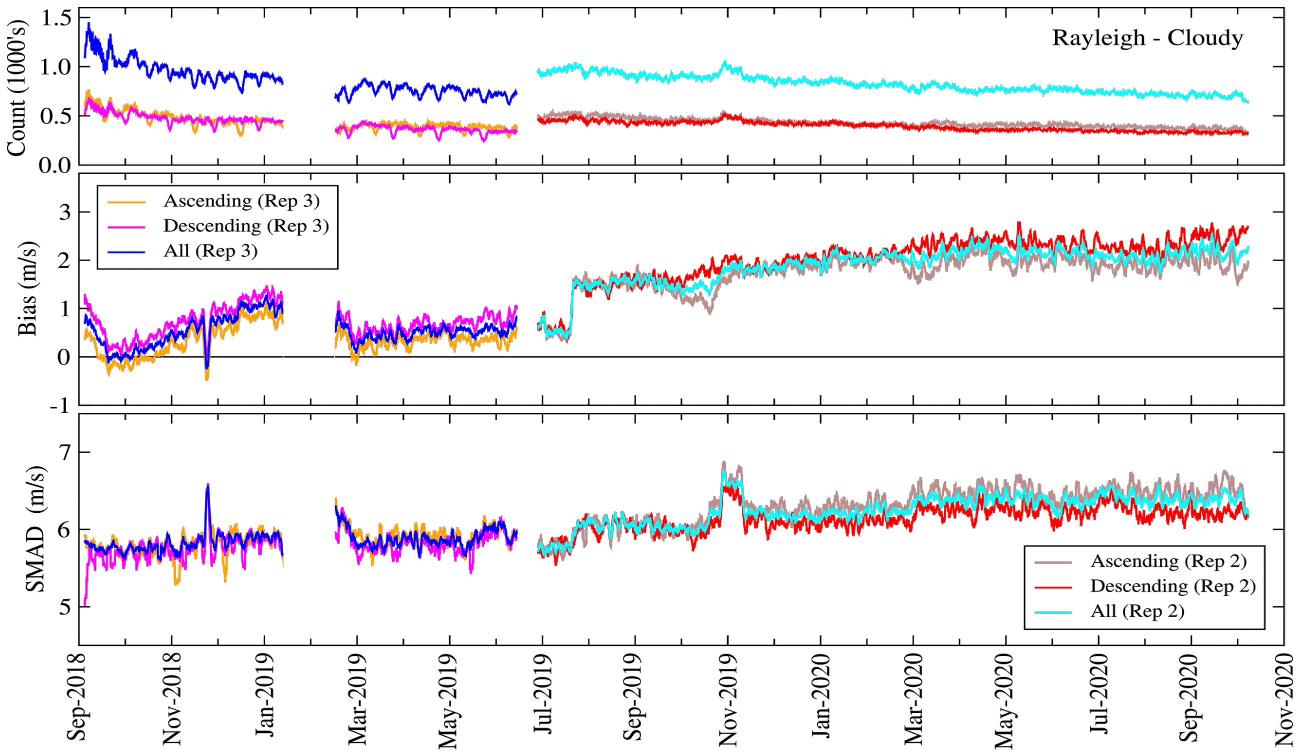


Figure 46: Time series of number of valid L2B Rayleigh cloudy wind observations, HLOS wind speed bias and SMAD between L2B Rayleigh cloudy and ECMWF IFS model winds. Running means along 30 successive files (orbits) are shown. Curves for all data as well as for ascending and descending orbits are shown for both the third (Rep 3) and the second (Rep 2) reprocessings.

The comparison between L2B Rayleigh cloudy HLOS wind against its model counterpart is shown in Figure 47 in a form of a density scatter plot for all reprocessed data for this channel/class pair covering the whole globe during the whole period of the third reprocessing. The agreement between the 3 million collocations of Rayleigh cloudy and model wind is acceptable with a global bias of +0.5 m/s, SDD of 6.6 m/s and SMAD of 5.9 m/s. The correlation coefficient is 0.933. Comparing these values against the ones from the second reprocessing proves that the third reprocessing (or FM-A) is better. Some other statistics are provided in the lower right corner of Figure 47.

Note that most of the collocations in Figure 47 fall on the symmetric line (the diagonal) as indicated by the orange-/red-/brown-coloured boxes. The circles and the crosses deviate slightly from each other at both ends of the distribution. This deviation can be attributed either to nature of Rayleigh cloudy wind being much noisier than the model wind or possibly some model issues at high winds.

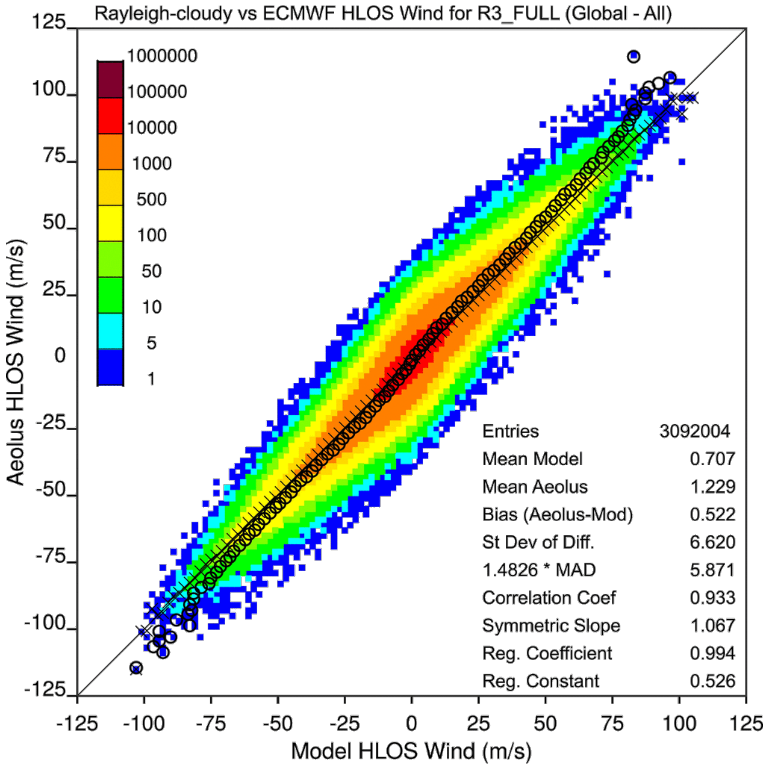


Figure 47: Density scatter plot comparing L2B Rayleigh cloudy to ECMWF model HLOS winds over the whole globe during the whole period of third reprocessing. Boxes are colour coded based on the number of collocations within the box. The circles are the average model value given L2B value while the crosses are the other way around. Statistics are given in the lower right corner of the plot.

Figure 48 shows the variation of number of valid observations, bias and SDD (compared to ECMWF model) as functions of HLOS wind for the whole third reprocessed dataset. The corresponding curves from the second reprocessing (for the FM-B period) are also shown for comparison. The bias distribution for the third reprocessing shows a symmetric shape around zero. Although it is not a perfect symmetry, it is a clear improvement with respect to that of the second reprocessing. Although very close, the bias curve does not pass through the origin. The bias becomes rather high (more than 4 m/s) at extreme values (above 90 m/s). Note that the number of those values is rather small. It is also worthwhile mentioning that the bias of 2 m/s suggested by Figure 48 for HLOS values in excess of 10 m/s (in either direction) seems an artifact of the assumption that both Rayleigh cloudy and model winds have equal errors (as was mentioned in relation to the similar case of Rayleigh clear (see Figure 36 and Figure 37). The actual bias is much smaller.

The distribution of the SDD with respect to the HLOS wind is rather flat at around 7 m/s for most of the range. This value is lower than that from the second reprocessing. However, a great proportion of the dataset, which is in the range of -10 m/s to 10 m/s, has a rather smaller SDD of about 6 m/s.

The probability density function (PDF) and the cumulative distribution function (CDF) of Rayleigh cloudy wind bias with respect to the ECMWF model for each individual orbit over the whole first FM-A period are shown in Figure 49. More than 80% of the orbits show positive bias. The CDF in Figure 49 clearly shows that more than 95% of the orbits have bias within -0.4 m/s and +1.5 m/s.

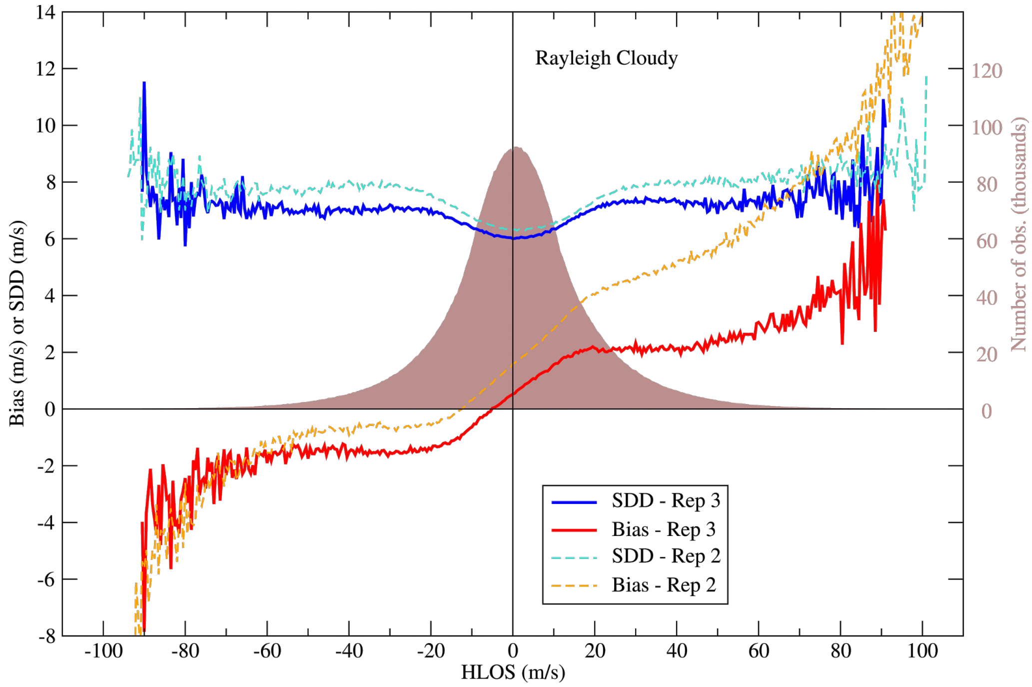


Figure 48: Number of valid Rayleigh cloudy HLOS wind observations, bias and SDD (compared to ECMWF model) as functions of HLOS wind for the whole third reprocessed dataset.

The third reprocessed Rayleigh cloudy dataset was discriminated based on the geographical location being in extratropical Northern Hemisphere (NH; northern of 20°N), Southern Hemisphere (SH; southern of 20°S) or in the Tropics (between latitudes 20°N and 20°S). The scatter plots showing the comparison against the ECMWF model in each geographical region are shown in Figure 50. The Tropical winds are lower than those of the extra tropics. However, the agreement with the model in the Tropics is not as good as in the extra tropics (compare the statistics in the three panels of Figure 50). Furthermore, the distribution of Rayleigh cloudy versus model HLOS winds follows a “stretched” S-shape which is more obvious in the Tropics as can be seen in the lower panel of Figure 50.

Figure 51 shows the geographical distribution of the number of Rayleigh cloudy observations within boxes of 3°X3° in latitudinal and longitudinal directions during the whole third reprocessing period. Note that boxes around the equator are the largest with the size gets smaller moving towards the poles. As one would expect, deserts, including Antarctica, show themselves with small number of cloudy observations. The enhanced number of observations in the Tropics especially over the rainforests, cannot be missed. It is important to note that the number of observations is much lower than those for Rayleigh clear or even Mie cloudy.

Figure 52 and Figure 53 show the geographical distribution of HLOS wind speed bias and SDD, respectively, between Rayleigh cloudy and ECMWF model for the whole third reprocessing dataset. The bias, as can be seen from Figure 52, is positive and rather high everywhere. However, over Antarctica the bias is relatively low. On the other hand, as shown in Figure 53, the SDD is rather high with enhanced values over Antarctica and most of South America and in the tropics especially the ITCZ and the monsoon region.

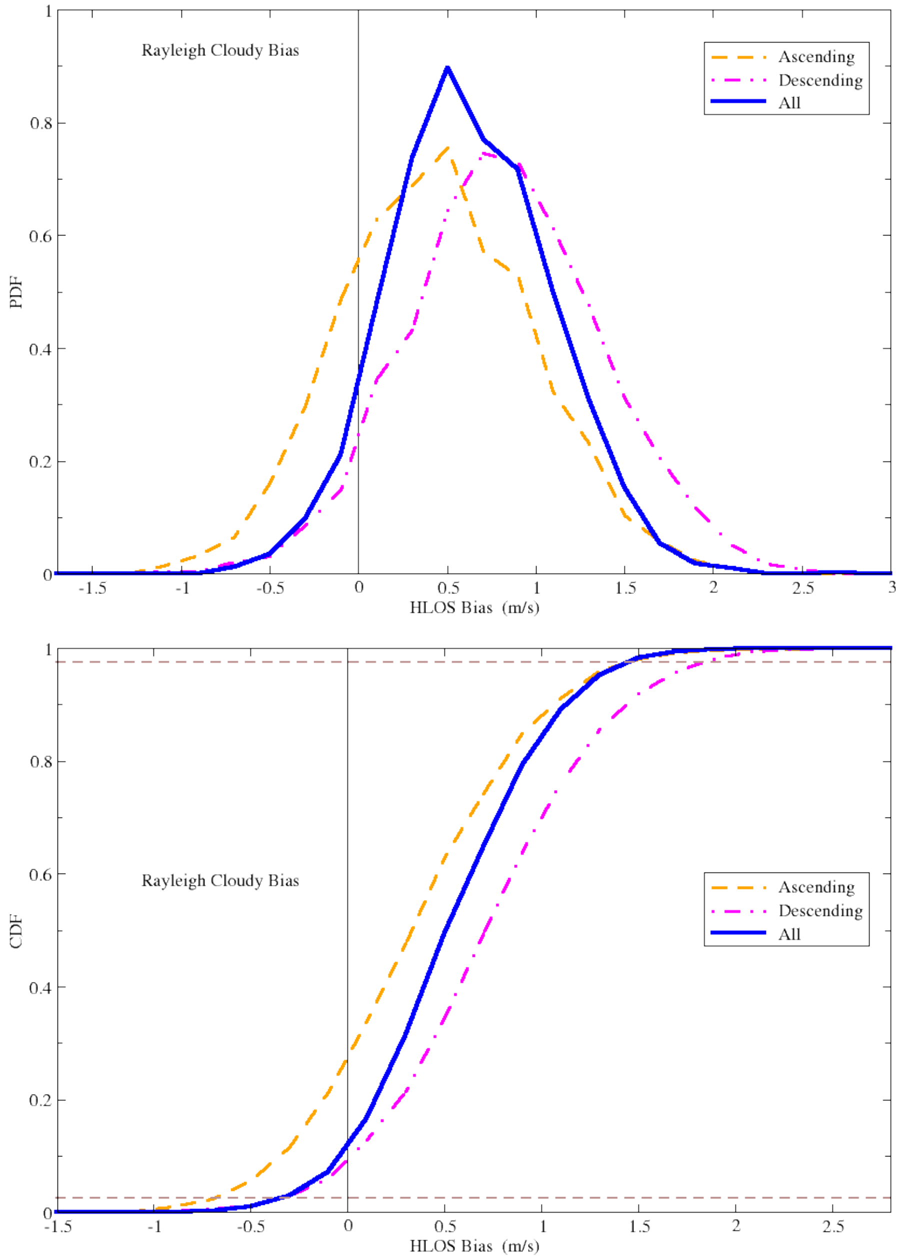


Figure 49: Same as Figure 25 but for Rayleigh cloudy wind bias.

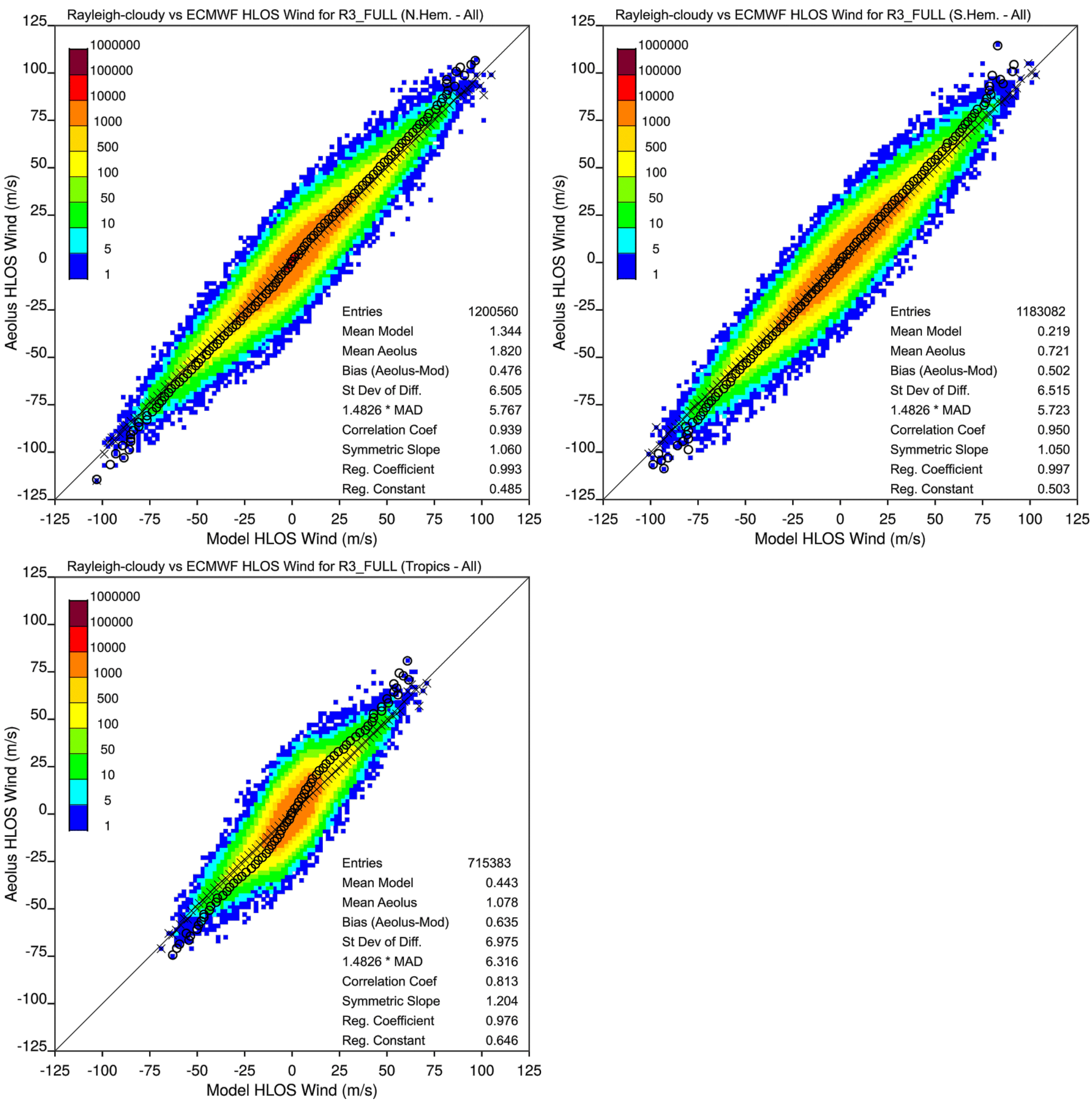


Figure 50: As in Figure 47 but with discrimination between extra tropics (upper) Northern Hemisphere (upper left) and Southern Hemisphere (upper right) and Tropics (lower). Refer to Figure 47 for the meaning of crosses, circles and colour-coding.

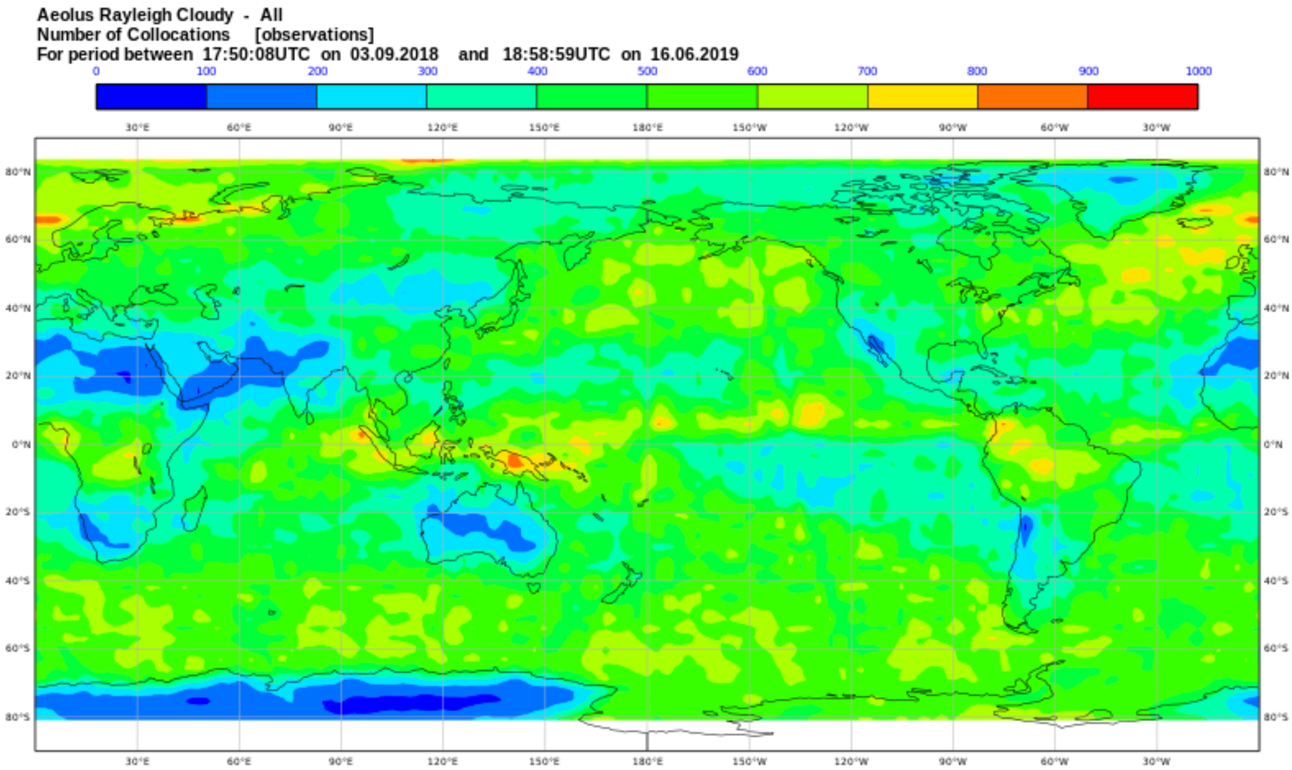


Figure 51: Geographical distribution of number of Rayleigh cloudy L2B observations per 3°X3° latitude-longitude box during the whole third reprocessing.

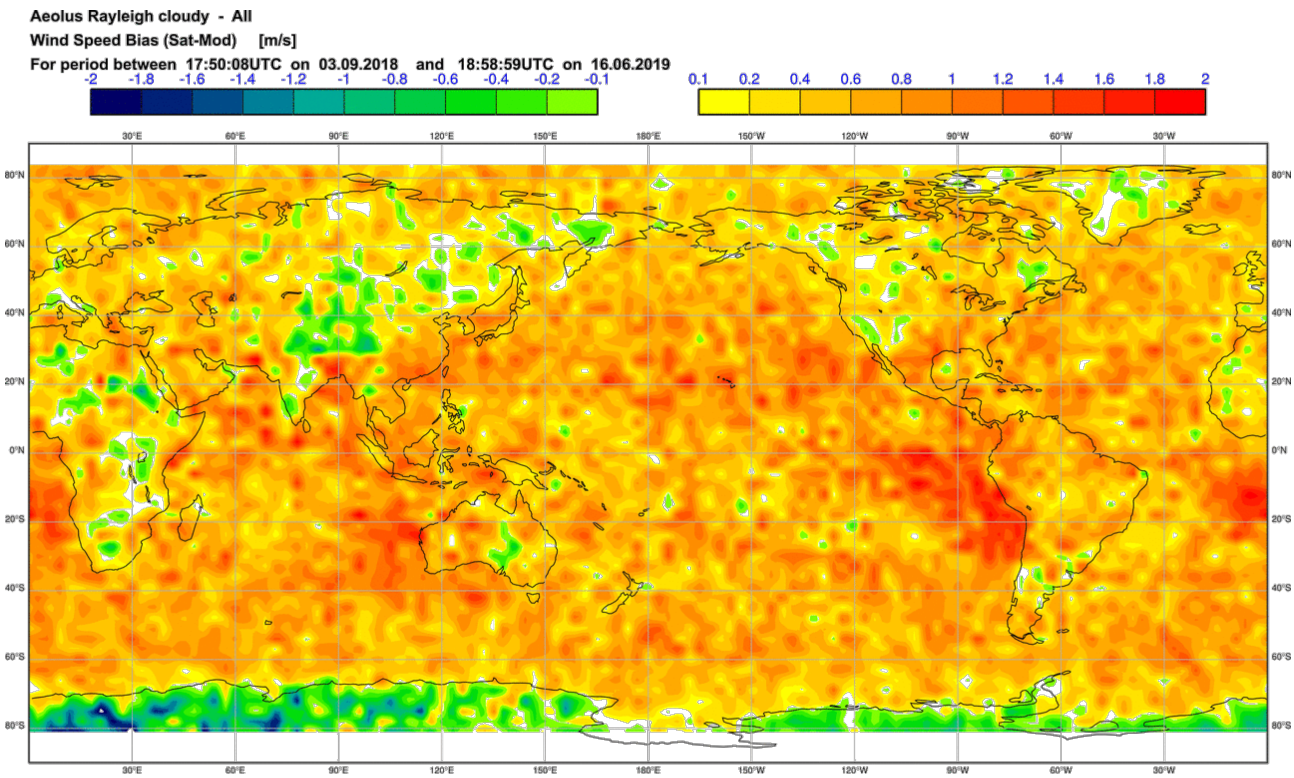


Figure 52: Geographical distribution of HLOS wind speed bias between Rayleigh cloudy and ECMWF model for the whole third reprocessing dataset.

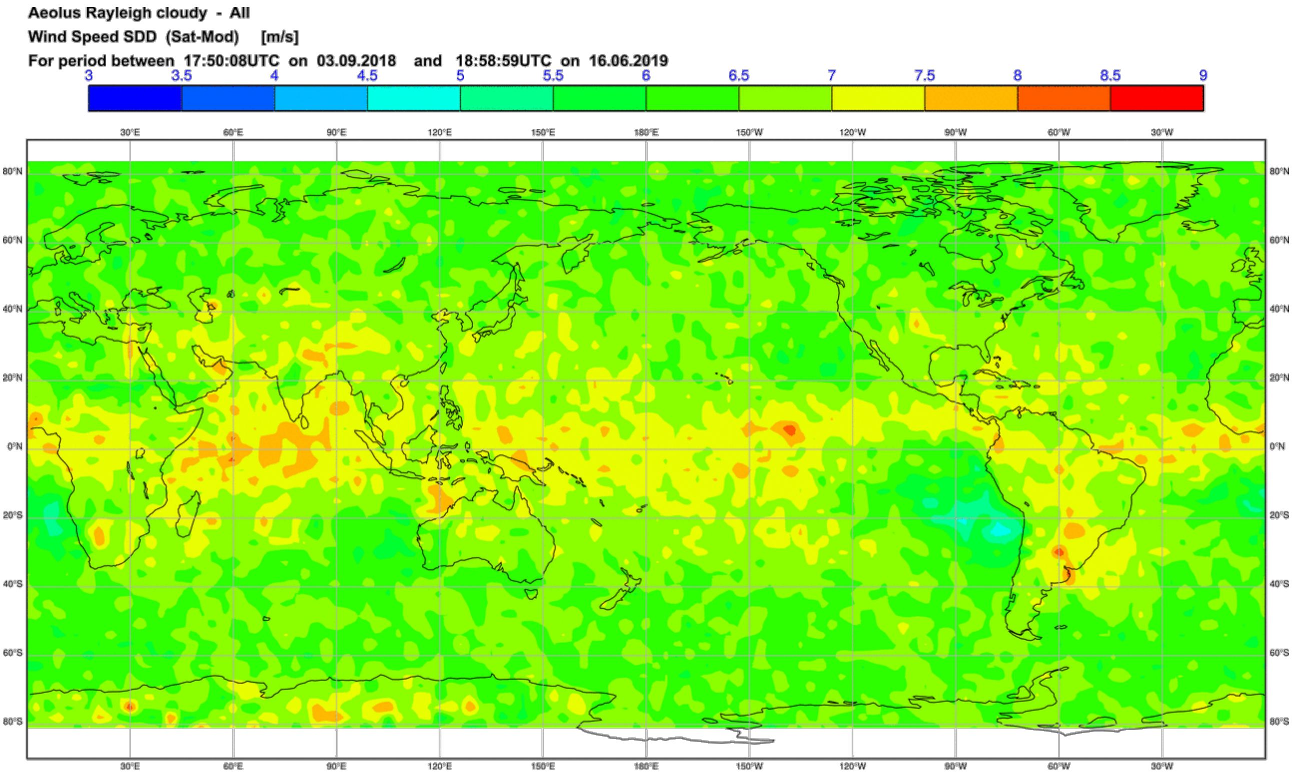


Figure 53: As in Figure 52 but for the Rayleigh cloudy SDD.

The whole reprocessed L2B Rayleigh cloudy data set was discriminated based on the direction of orbit whether ascending or descending. The resulting time series for ascending, descending or both (all) are shown in Figure 46. The scatter plots for ascending and descending orbits are shown in Figure 54. Statistics show that indeed, observations from descending orbits are in better agreement with the model than those from ascending orbits.

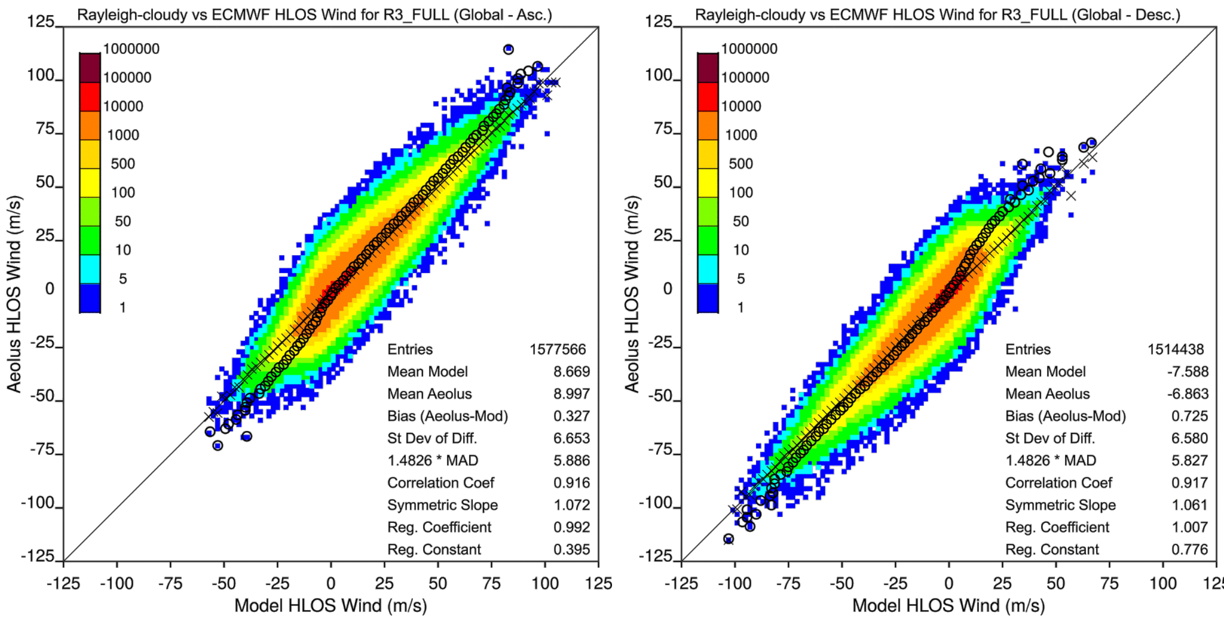


Figure 54: As in Figure 47 but with the discrimination between ascending (left) and descending (right) orbits. Refer to Figure 47 for the meaning of crosses, circles and colour-coding.

Profiles of L2B Rayleigh cloudy HLOS winds deviations from ECMWF model (bias, SDD and SMAD) together with the number of collocations at each altitude are shown in Figure 55. Between altitudes 4 km and 16 km the bias is rather small (~ 0.1 m/s) and independent of the altitude. Below 4 km and above 16 km the bias increases linearly with altitude and reaches 1 m/s near the surface and at the top of the profile. The overall SMAD and SDD profiles in Figure 55 increase linearly from 5.5 m/s and 6.3 m/s near the surface to 6.5 m/s and 7.2 m/s at the top of the profile, for SMAD and SDD respectively.

SMAD and SDD profiles in Figure 55 also suggest that observations from descending orbits are in better agreement with the model. The differences are rather marginal.

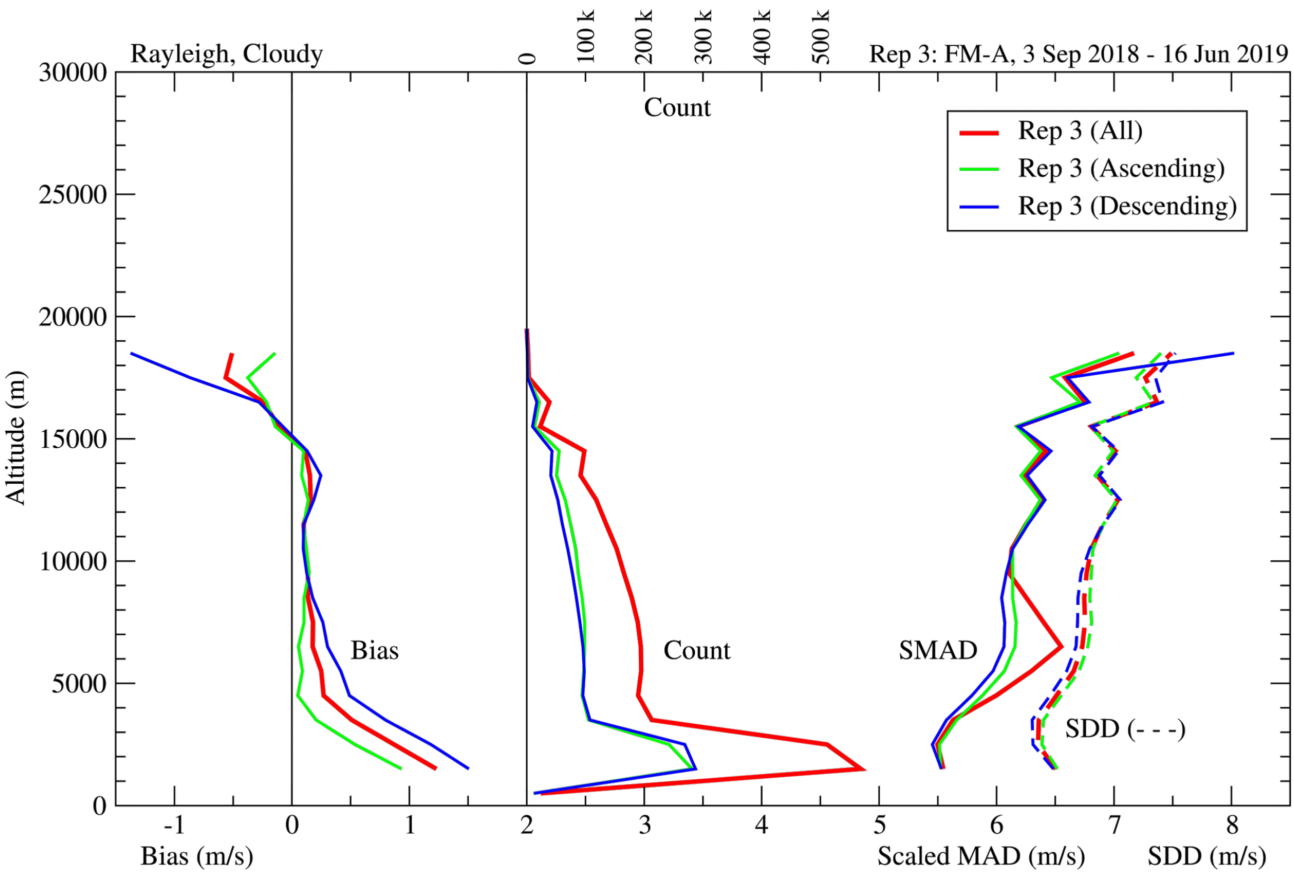


Figure 55: Profiles of L2B Rayleigh cloudy HLOS wind deviations from ECMWF model (bias, SDD and SMAD) together with the number of collocations at each altitude. Profiles show discrimination between ascending and descending orbits in addition to all data.

8 Conclusions

Aeolus data products were reprocessed from August 31, 2018 to June 16, 2019 covering the operation period of the first Flight Model laser (FM-A). This was the third reprocessing campaign performed based on processor versions deployed in March 2022 (L1bP 7.12, L2aP 3.14.8, L2bP 3.70), used for near-real time (NRT) generation of baseline 14 (B14) data products. The reprocessing covers L1A, L1B, L2A and L2B data products as well as the respective auxiliary files and provides for the first-time low wind bias data products for the FM-A period. In combination with the second reprocessing campaign and the NRT baselines, the third reprocessing campaign ensures a seamless data availability of low wind bias products from the beginning of the mission to this day. The next reprocessing campaign will be performed using the B16 processor version and will cover the complete data period from September 2018 to the release of NRT baseline B16 in spring 2023.

Main improvements with respect to the original processing baseline:

- Correction of orbit phase dependent wind bias in L2B product using daily updates of the correlation between ECMWF O-B (observation-background) statistics and temperatures from the telescope primary mirror M1. This correction is applied to Rayleigh and Mie L2B winds and includes also the correction of constant bias drifts. The correction for day d was applied with correlations from day d (which is even more accurate than the previous day, d-1, correlations used for the NRT products), which should slightly improve the bias as well.
- Correction of hot pixel fluctuations based on an algorithm which detects hot pixel induced steps in the backscatter wind signal. No regular dark signal measurements (DCMZ) were performed at the beginning of the FM-A period. However, the hot pixel detection algorithm ensures the correction of hot pixel induced steps also in these periods. Note that in contrast to previous reprocessing campaigns a sophisticated detection algorithm was used which also allows for the correction of illuminated hot pixels located within the two Rayleigh spots.
- Radiometric coefficients K_{ray} and K_{mie} are calculated along the orbit for L2A products using a multiple linear regression based on M1 temperatures. The first estimates are calculated from signal prediction in particle-free regions of the atmosphere for mid-altitudes from 6 to 16 km and then fitted using the regression coefficients to provide both K_{ray}, and K_{mie} per observation for the whole orbit. This feature was not implemented for the NRT FM-A baselines.
- Implementation of a physical regularization scheme, called Maximum Likelihood Estimation (MLE), to improve the processing of crosstalk-corrected signals in the L2A products. The MLE product is an alternative to the SCA product which is less sensitive towards signal noise.
- Availability of the L2A AEL-FM (Feature Mask) and AEL-PRO (Processing) data products. The FM and PRO algorithms are based on EarthCare developments. The FM algorithm provides a high-resolution feature mask and is the prerequisite of the optimal estimation realised in the PRO algorithm. The algorithm provides estimates of cloud and aerosol extinction as well as the lidar ratio. However, due to the unknown data quality these products are flagged invalid.

Further improvements with respect to NRT data products:

- Improvement of the dark signal correction for the uppermost Rayleigh range bin for some DCMZ calibrations of the NRT product from 24th May, 2019 onwards.

- At the beginning of the mission, a large number of instrument tests were performed which lead to degraded data quality. For the reprocessing, the L2B wind products were flagged invalid for such periods, which facilitates the automatic use of L2B products.
- The L2B wind bias correction using M1 temperatures was applied for day d with correlations from the same day d. This is slightly more accurate than using the correlations from the previous day, d-1, as applied during NRT processing for the NRT products. This leads to slight improvement in wind bias.
- In addition to the M1 bias correction, the Rayleigh and Mie winds were corrected for a remaining seasonal orbit-phase-dependent bias. This removes the enhanced opposite bias for ascending and descending orbits during October and March.
- Improved calibration scheme for L2B Mie cloudy winds based on the ECMWF model which leads to an improvement of the systematic and random Mie wind errors.
- The calculation of the geolocation parameters was improved. This removes a longitudinal offset (~8 km at the equator) in the geolocation of all Aeolus products.

The quality of the data products:

- Better quality of signal levels (Rayleigh, Mie) in L1B product was achieved due to better treatment of the hot-pixel dark current levels. This improves L2A and L2B products which are based on L1B products.
- Radiometric calibration of L2A products was performed with consideration of the orbital variations. This leads to lower systematic errors for the backscatter and extinction profiles. Moreover, the accurate radiometric calibration allows the proper monitoring of the signal loss in the atmospheric path over time.
- The Rayleigh (clear) wind is globally unbiased wrt to the ECMWF model. Most of the orbit-phase dependent wind bias observed in NRT products has been removed. The bias varies within ±0.4 m/s for more than 95% of the orbits. This is a considerable improvement wrt the NRT products, which showed M1 induced temporal variations of the Rayleigh bias of several m/s's during this period.
- The wind random differences such as SMAD (scaled median absolute difference = 1.4826*MAD and SDD (the standard deviation of the difference) from the 3rd reprocessing are smaller compared to the second reprocessing. SMAD is less than 5 m/s while SDD is less than 6 m/s. SMAD (and SDD) show a clear increasing trend with time from 4.0 m/s at the beginning of the mission to 6.0 m/s at the end of first FM-B period (June 2019) due to the instrument degradation.
- Rayleigh clear wind is almost unbiased compared to the ECMWF model up to altitudes 20 km. The bias becomes negative (underestimation) above that altitude. SMAD and SDD show minimum values of 4.0 m/s and 5.0 m/s, respectively, between altitudes 4 km and 12 km. Outside this band, SMAD and SDD increase rapidly and exceed 9 m/s near the surface and above altitude 24 km.
- The Mie (cloudy) wind bias wrt the ECMWF model shows global averages around -0.10 m/s (underestimation), with more than 95% of the orbits showing a bias within -0.45 m/s and +0.25 m/s. This is a considerable improvement with respect to the NRT products, which showed temporal variations of the Mie bias of several m/s during this period.
- SMAD (=1.4826*MAD) as a measure of the random error shows values of around 3.0 m/s for Mie cloudy winds at the beginning of the period. Afterwards, the Mie scaled MAD gradually increases to 3.4 m/s.

- Mie cloudy wind show slight positive bias (of about 0.1 m/s) between altitudes 3 km and 15 kn. The bias gradually increases to -0.5 m/s near the surface and +0.4 m/s at altitude 17 km. SMAD and SDD are smallest (2.5 m/s and 3.0 m/s, respectively) near the surface and increase (almost) linearly with altitude to reach 5.0 m/s (both) at altitude 17 km.

Known limitations of the data products:

- Due to initial switch-on of the instrument and related tests, the data products should be only used from 3 September '2018 onwards, for the L2B product data was flagged invalid during periods of degraded data quality (mostly due to dedicated instrument test periods), for L1B and L2A products please check the data exclusion periods and data unavailability periods are available at: https://earth.esa.int/eogateway/instruments/aladin/quality-control-reports
- As the hot-pixel corrections are applied only for the following orbit file, hot pixel fluctuations cannot be corrected within one orbit product, but it should be noted that the L2B processing has a dedicated check for strong biases caused by an uncorrected hot-pixel and will flag winds invalid if they occur.
- There was a slight increase in bias and random error (represented by SMAD and SDD) from 23 to 25 November 2018. Furthermore, there was slight increase in random error from 26 February to 1 March 2019.
- The L2B Mie cloudy winds show a sustained global bias of -0.1 m/s (underestimation). This is a known issue and related to the choice of the Mie estimated error threshold which will be further investigated. The L2B wind from the 3rd reprocessing shows a significant improvement in Mie cloudy bias especially at high winds. The wiggled shape of bias (seen in data from the 1st and 2nd reprocessing campaigns) with respect to the wind speed has almost disappeared in the data from the 3rd reprocessing.
- As the L2B bias correction is based on global O-B statistics, there could be higher regional biases.
- L2B Rayleigh-cloudy winds still show significant bias (up to 1 m/s) and significant higher random error (of around 6 m/s) than Rayleigh-clear winds at the beginning of the period. In contrast to the random error for Rayleigh-clear winds, the Rayleigh-cloudy random wind error did not further increase during the reprocessing period. For future reprocessing campaigns, it is planned to improve Rayleigh-cloudy with a dedicated correction algorithm.
- Due to the unknown quality of the new L2A AEL-FM and AEL-PRO products, these datasets are flagged invalid. However, the data is nevertheless accessible for analysis.