Aeolus Data Innovation Science Cluster
DISC

Verification report of the second reprocessing campaign for FM-B from 24 June 2019 till 9 October 2020

DISC-Ref.: AED-TN-ECMWF-GEN-060

Issue: V 1.1

Date: 10/12/2021


DLR
cnrs
DoRIT
serco
s[&]t
ABB
Royal Netherlands Meteorological Institute
Ministry of Infrastructure and Water Management
ECMWF

Document Change Log

Issue	Date	New pages	Modified pages (after introducing new pages)	Observations	Name
V 1.0	29 Nov. 2021	-	-	Version 1.0 prepared with input from DISC consortium	Saleh Abdalla, Isabell.Krisch, Uwe.Marksteiner, Oliver Reitebuch, Michael Rennie, Fabian Weiler
V 1.1	10 Dec. 2021	-	-	Version 1.1: Section for verification of L2A product was added. Maps of number of observations were added. Minor corrections/updates were introduced as appropriate.	Saleh Abdalla, Thomas Flament, Isabell.Krisch, Uwe.Marksteiner, Oliver Reitebuch, Michael Rennie, Dimitri Trapon Fabian Weiler

Table of Contents

1 Introduction

1.1 Compliance Statement

The Verification report of the second reprocessing campaign for FM-B from 24 June 2019 till 9 October 2020 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] AE-TNDLR-7300-2, issue 1.1, the dark side of Aeolus during the IOCV phase

[RD-2] https://www.aeolus.esa.int/confluence/display/CALVAL/Calval+Black+List

[RD-3] L1B_E2_FW_WM084_Reprocessing_dcmz_validation.pptx, DISC analysis report

[RD-4] L1B_E2_NW_FW_WM108_Reprocessing_dcmz_aux_validation.pptx, DISC analysis report

[RD-5] DLR and Witschas, B. (2019). Analysis of in-orbit Rayleigh spectrometer transmission vurves. AE.HR.DLR. WP7300. SPACE.RSP, page 53.

[RD-6] DLR and Witschas, B. (2020). Analysis of Rayleigh spectrometer transmission vurves based on instrument spectral registration (ISR) measurements and investigation of instrumental drift. AED-TN-DLR-L1B-03, page 67.

[RD-7] AE-TN-ECMWF-GEN-040, issue 1.0, Verification report of first reprocessing for FM-B covering the time period 2019-06 to 2019-12

[RD-8] Weiler, F. et al. (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-9] https://www.aeolus.esa.int/confluence/display/CALVAL/Calval+Black+List

[RD-10] L1B_E2_NM_FW_WM_125_2nd_Reprocessing_dcmz_aux_validation.pptx

[RD-11] L1B_E2_NM_FW_WM_125_2nd_Reprocessing_dcmz_aux_validation.pptx

[RD-12] L1B_NM_FW_WM_148_2nd_Reprocessing_dcmz_PDGS_validation.pptx

[RD-13] Marseille, G.-J., de Kloe, J., Marksteiner, U., Reitebuch, O. and Rennie, M. (2021). NWP calibration applied to Aeolus Mie channel winds, submitted to Q J R Meteorol Soc.

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 Aeolus reprocessing campaign covering the early operation period of the second Flight Model laser (FM-B) 24 June to 31 December 2019, efforts were focused on the second reprocessing campaign covering a longer period of FM-B from 24 June 2019 till 9 October 2020. Although, the reprocessing starts with 24 June 2019, when Aeolus started operating in FM-B, the first usable L2A and L2B products are available from 28 June 2019 onwards since the data collected during the first four days are used for calibration, OWV and NOP.

The processor versions deployed operationally in NRT in October 2020 (L1bP 7.09.1, L2aP 3.11.1, L2bP 3.40.2) to generate baseline 11 (B11) data products were used for the second reprocessing. The reprocessing provided level 1 and 2 (L1A, L1B, L2A and L2B) data products as well as all corresponding auxiliary files. Thus, a homogenous high quality and low-biased data product is ensured for the period of the second reprocessing in addition to the period covered by the NRT processing using B11 which extends from 24 June 2019 (start of the second reprocessing) till 26 May 2021 (the date when B11 was replaced by B12 in the NRT processing). The only difference between the reprocessed data and the NRT production is the details of "telescope primary mirror, M1, temperature-based bias correction" (or simply "M1 bias correction"; see below).

The second reprocessing introduced the following improvements with respect to the original processing and/or first reprocessing baseline:

- correction of wind bias in L2B product using daily updates of the correlation between ECMWF IFS O-B (observation-background) statistics and temperatures from the telescope primary mirror M1. This eliminates large wind biases for the NRT product produced between 1 January (the day just after the first reprocessing stops) and 20 April 2020 (the date on which M1 bias correction was activated in NRT processing). This correction is applied to Rayleigh and Mie L2B winds. It implicitly corrects for constant drifts in bias). Unlike the NRT M1 bias correction which uses data from the previous day to correct the biases of the day, the reprocessing considers the data of the day for that leading to even lower biases. This is one of the advantages of the reprocessing campaigns.
- Multiple linear regression based on M1 temperature is used to calibrate radiometric coefficients K_{ray} and K_{mie} along the orbit. The radiometric coefficients are needed for L2A products. Furthermore, the parameter attenuated backscatter is provided for the whole period.
- correction of hot pixels per orbit in time periods between dark signal measurements (DUDE) for L1B product, which is an improvement compared to updating the hot-pixel correction only 4 times per day after DUDE for the NRT data products; improvement of the hot-pixel correction for the uppermost Rayleigh range bin for some DUDE calibrations of the NRT product up to 21 October 2019.
- L2B wind products were flagged invalid for periods of known instrument tests or degraded data quality, which eases the automatic use of L2B products.

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 L2B products are presented in Section 7. Suggestions for future work are given in Section 9. Finally, conclusions are listed in Section 8.

3 AUX_ZWC

In order to validate the content of the reprocessed AUX_ZWC files, comparisons of original and reprocessed AUX_ZWC data were performed for sub-dataset various sub-datasets. According presentations are available on the EDAFECS-FTP under

/documentation/documentation_deliverables_DISC/Reprocessing/Reprocessing_2nd_campaign for:

1. the preparation of the verification against a dataset reprocessed via the Sandbox
    - 20210126 UM RPRO2 AUXZWC1B Validation Sandbox Summary.pptx
2. the verification of selected PDGS test periods
    - 20210624 UM - RPRO2 Validation of AUX_ZWC v2.pptx
3. the verification of the final PDGS dataset
    - 20210624 UM - RPRO2 Validation of final AUX_ZWC dataset v2.pptx
    - 20211115 UM - Validation of RPRO2 AUX_MRC and AUX_RRC files.pptx

In order to more or less representatively cover the full period of the 2nd reprocessing campaign from June 2019 until October 2020, five test weeks have randomly been picked while aiming at a preferably even distribution of these over the whole period. Table 1 provides additional details regarding the five test weeks with their in total more than 500 AUX_ZWC files. No checks of Sandbox results against the final dataset from PDGS were performed.

The discrepancy marked by a red star * in Table 1 can be explained according to an email by Simone Bucci (DOS) from 2021/06/23: "*The missing AUX_ZWC was due to an error in delivery package preparation. I confirm that the data has been correctly generated, please find it in the OPS REPO […]*"

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

ZWC data sets	**Baseline** (*original / reprocessed*)	**Software_ Ver (L1bP)** (*original / reprocessed*)	**Start time** (*original / reprocessed*)	**Stop time**	**Number of files** (*original / reprocessed*)	**# of 0002 Files** (*original / reprocessed*)
1. AUG 19	1B05 / 1B11	07.06 / 07.09	2019-08-03 23:59.11	2019-08-11 01:29.47	105 / 105	none
2. DEC 19	1B07 / 1B11	07.07 / 07.09	2019-12-17 23:07.35	2019-12-25 00:37.59	105 / 105	none
3. MAR 20	1B07 / 1B11	07.07 / 07.09	2020-03-22 22:42.11	2020-03-30 00:11.59	105 / 105	none
4. JUN 20	1B09 / 1B11	07.08 / 07.09	2020-06-21 22:41.35	2020-06-29 00:11.35	**104 / 103***	none
5. SEP 20	1B10 / 1B11	07.08 / 07.09	2020-09-13 22:41.35	2020-09-21 00:11.35	105 / 105	none

The original AUX_ZWC dataset from the Near-Real-Time (NRT) processing was taken from the directories ***/ADDF/L1B_Calibration/1B05***, ***/ADDF/L1B_Calibration/1B07***, ***/ADDF/L1B_Calibration/1B09*** and ***/ADDF/L1B_Calibration/1B10*** on the FTP Server ***aeolus-ds.eo.esa.int***. The reprocessed AUX_ZWC dataset was taken from the directory ***/RPRO_2/Operational*** on the reprocessing FTP repository ***aeolus-ops-repo.eo.esa.int*** (User: *aeolus-disc*).

No files with version number 0002 in the filename have been generated during the reprocessing despite identical file name in Near-Rreal-Time dataset. In an email from 2021/06/23 Simone Bucci (DOS) explained: *„In the Technical Note we provided to DISC, related to the 2nd Reprocessing Strategy on Reference Platform ESA-EOPG-EOEP-TN-37 and also reported as Reference Doc by the DISC TN on Reprocessing, on Page 9, it's explained: The NRT and Reprocessed data will be published with different baselines and it'll be through the combination validity time / baseline that the data will be identified. This is important to be specified due to the file versioning that with the reprocessing performed in Reference will remain to "0001" and not increased … The difference between the 1st and 2nd Campaign is the execution environment. There is an increment in versioning if a version 1 of the file just generated is found. In an empty platform like the Reference, the version is always 0001 because the reprocessed data are the first processed. We'll have an exception for the 5 weeks data processed the past week, that data will have a version 2 due their presence with version 0001.*"

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

- The correct usage of AUX_MRC/RRC files before and after the Rayleigh Cover Temperature (RCT) change and the Laser Frequency Adjustment (LFA) on 2020-07-22 can be confirmed.
- The number of AUX_ZWC files agrees for the five test weeks between the NRT data set and the 2nd 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. The latter, in turn, affect the composition of the filename. In comparison to the original data set (*e.g. AE_OPER_AUX_PAR_1B_20190624T000000_99991231T235959_0005.EEF ingested with Baseline 09*) the reprocessed data set has undergone only very loose quality control (*e.g. AE_OPER_AUX_PAR_1B_20190627T000000_20190809T102324_0002.EEF as delivered to DOS for the 2nd reprocessing campaign*). For the 2nd reprocessing campaign the ground detection parameters have been adapted such that significantly more valid Ground Correction Velocities are provided within the AUX:ZWC files (as can be seen in Figure 1 vs. Figure 2 and top and vs. bottom in Figure 3 and Figure 4, respectively), and the user him/herself is responsible for a meaningful quality control with respect to the purpose of his/her application. The relevant parameters for the ground detection in the Wind Velocity Mode (WVM) section of the AUX_PAR_1B files differ accordingly (*the Mie and Rayleigh Useful Signal Thresholds and the Rayleigh Ground Detection Signal Derivative*). The AUX_PAR_1B files used for processing of the five test weeks of the original NRT data were obtained from the directory
  /AUX_files_ingested/AeolusPDGS_AUXfileusage_from_ADDF_2021_02_04.tar.gz
  on the EDAFECS ftp server. The most recent AUX file list is provided on the Confluence Wiki under
  https://csde.esa.int/confluence/pages/viewpage.action?spaceKey=AEOLUSDISC&title=PDGS+AUX+files
  as **Aeolus PDGS AUXfileusage_timeline_v3.8.xlsx**.

- Differences in biases (*mean and median as in the tables on the right of* Figure 3 and Figure 4) between of original and reprocessed Mie & Rayleigh channel datasets for the five test weeks are assumed to be due to the usage of different IRCs.
- The variability in the number of valid Ground Correction Velocities among the 5 test weeks is strongly related to the ice coverage of the polar regions, i.e. depending on the season. An additional influence by the signal decay on the instruments atmospheric path seems to be secondary.

The overall quality of the reprocessed dataset seems more or less as expected and is accepted.

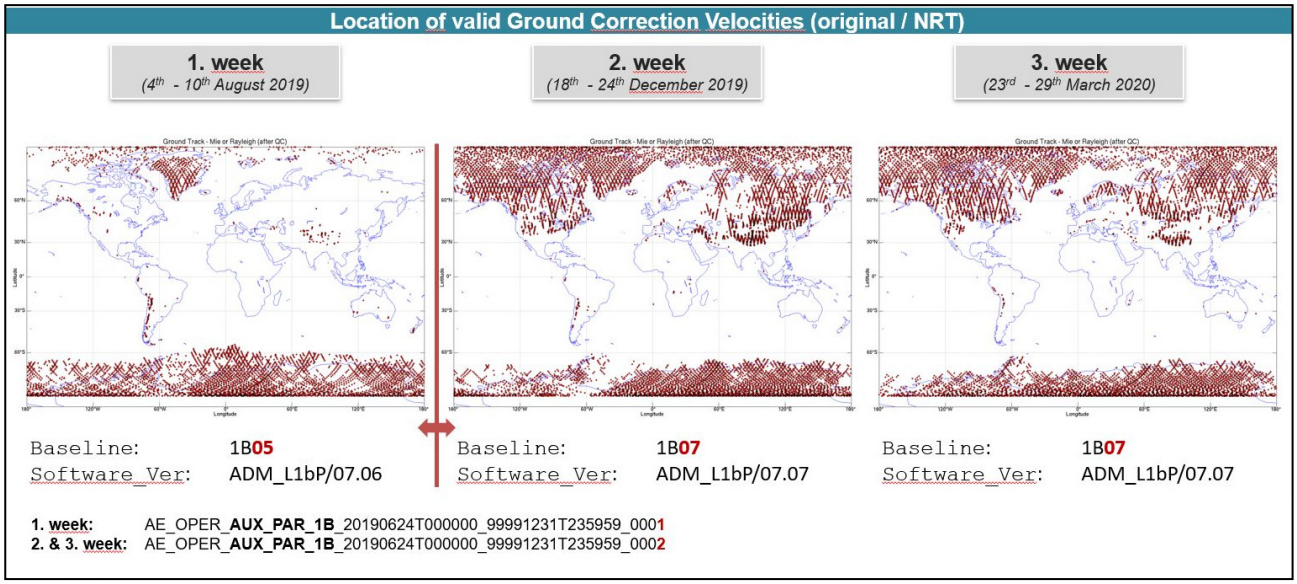


Figure 1: Location of the valid Ground Correction Velocities contained in the AUX_ZWC files of the 1st, 2nd and 3rd week of the original NRT data set, along with the applied AUX_PAR_1B files containing the ground detection parameters in the <WVM_Params> section.

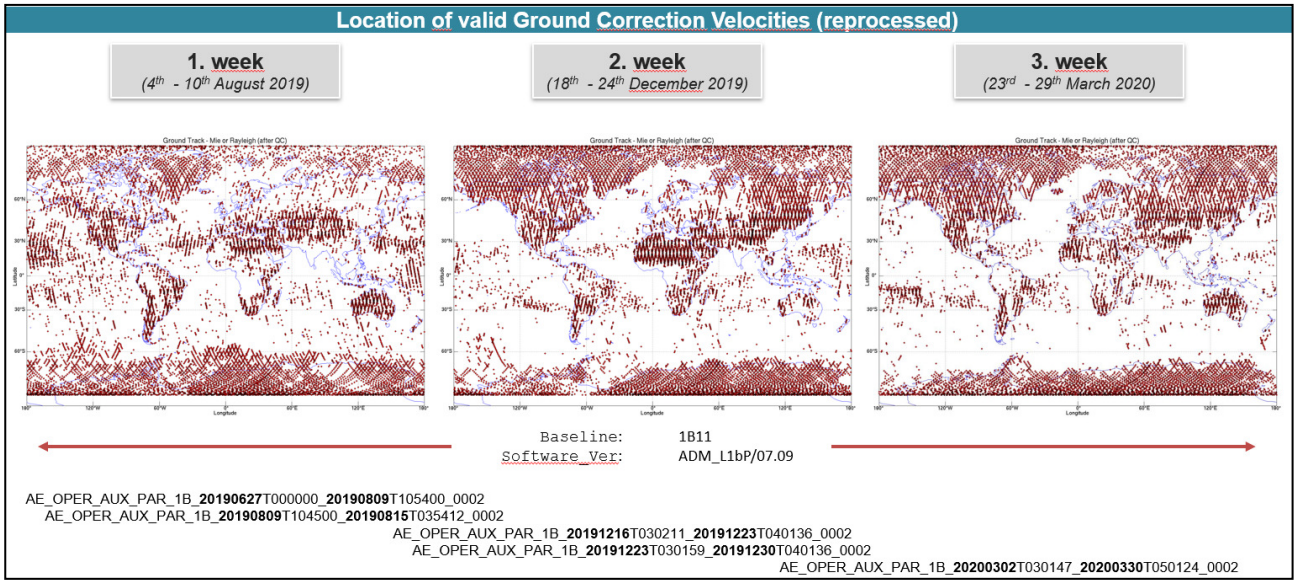


Figure 2: Location of the valid Ground Correction Velocities contained in the AUX_ZWC files of the 1st, 2nd and 3rd week of the reprocessed data set, along with the applied AUX_PAR_1B files containing the ground detection parameters in the <WVM_Params> section.

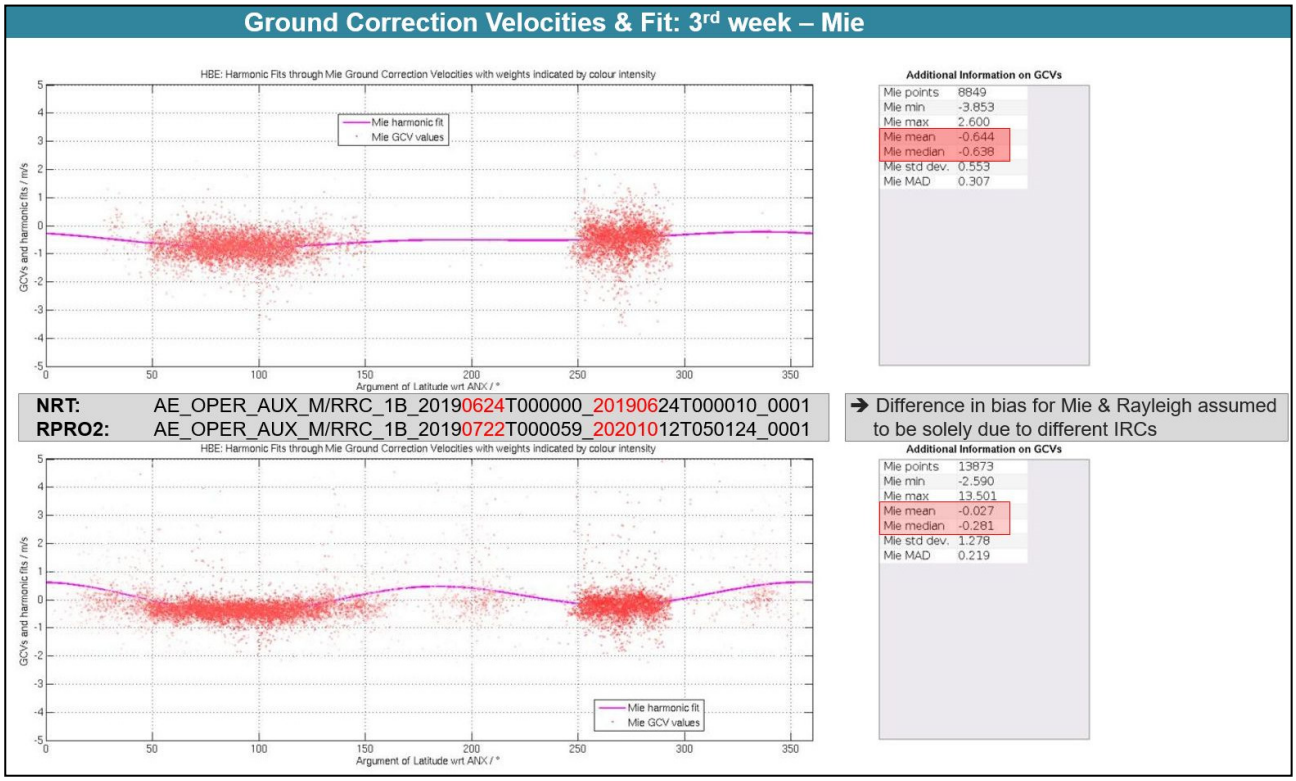


Figure 3: Mie Ground Correction Velocities and harmonic fits of the 3rd test week for the original NRT data set (top) and the data set of the 2nd reprocessing campaign (bottom).

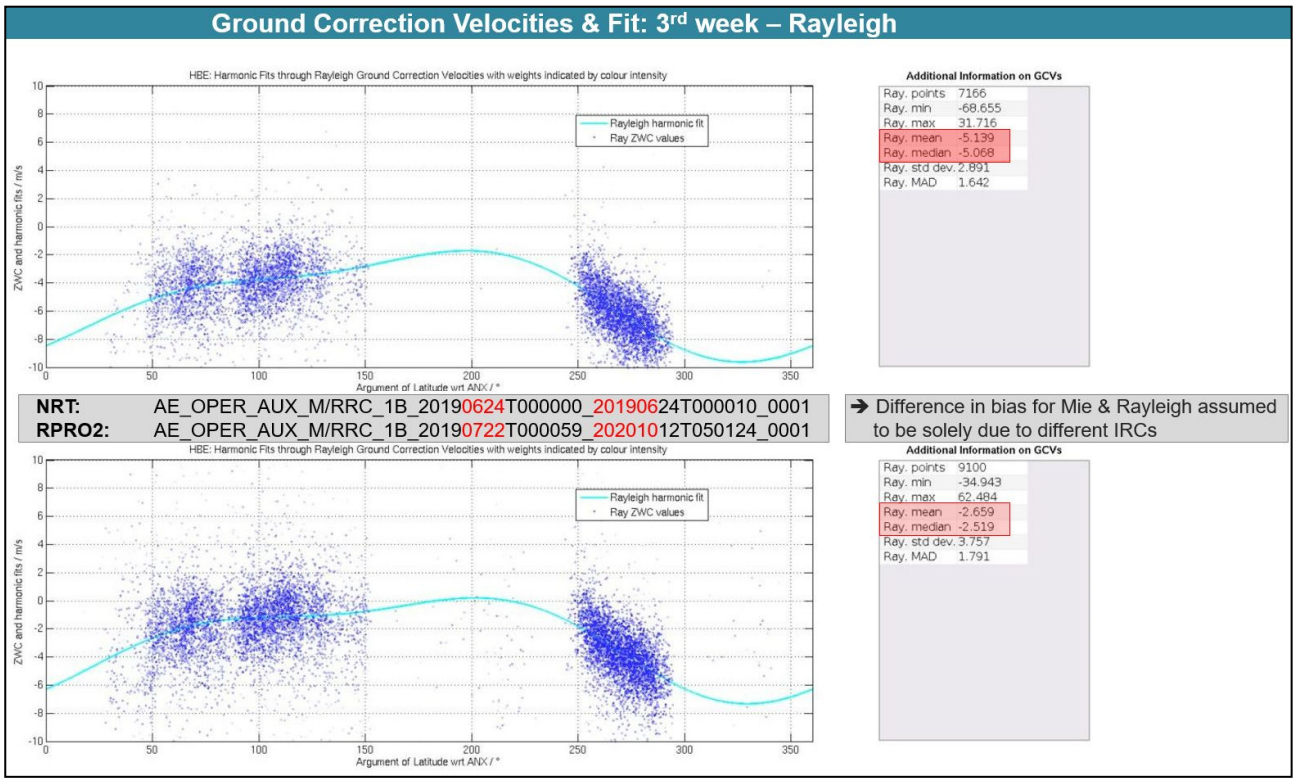


Figure 4: Rayleigh Ground Correction Velocities and harmonic fits of the 3rd test week for the original NRT data set (top) and the data set of the 2nd reprocessing campaign (bottom).

4 AUX_MRC and AUX_RRC

In order to validate the content of the reprocessed AUX_MRC and AUX_RRC files (*25 files each starting with IRC #37 from 2019-07-04 and ending with IRC #61 from 2020-02-24*), comparisons of the original and reprocessed dataset were performed. No checks against the Sandbox results were performed. An according presentation describing the verification of the final PDGS dataset is available on the EDAFECS-FTP under ***/documentation/documentation_deliverables_DISC/Reprocessing/Reprocessing_2nd_campaign/20210928 UM BW - RPRO2 Validation of final AUX_ISR AUX_MRC AUX_RRC datasets.pptx***.

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/1B05/ ;
/ADDF/L1B_Calibration/1B06/ ; and
/ADDF/L1B_Calibration/1B07/
on the FTP Server ***aeolus-ds.eo.esa.int***.

The reprocessed AUX_MRC and AUX_RRC dataset was taken from the directory
/RPRO_2/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 and reprocessed AUX_RRCs as well as start/stop times of the applied AUX_PAR_1B files.

	AUX_RRC filename	Used AUX_PAR_1B file
Original:	AE_OPER_AUX_RRC_1B_20190704T194047_20190704T195711_0001	AE_OPER_AUX_PAR_1B_**20180712**T000000_99991231T235959_0005
Reprocessed:	AE_OPER_AUX_RRC_1B_20190704T194047_20190704T195711_0001	AE_OPER_AUX_PAR_1B_**20190627**T000000_20190809T105400_0002
:	:	:
Original:	AE_OPER_AUX_RRC_1B_20190808T194059_20190808T195711_0001	AE_OPER_AUX_PAR_1B_**20190624**T000000_99991231T235959_0001
Reprocessed:	AE_OPER_AUX_RRC_1B_20190808T194059_20190808T195711_0002	AE_OPER_AUX_PAR_1B_**20190627**T000000_20190809T105400_0002
:	:	:
Original:	AE_OPER_AUX_RRC_1B_20190923T181559_20190923T183211_0001	AE_OPER_AUX_PAR_1B_**20190624**T000000_99991231T235959_0001
Reprocessed:	AE_OPER_AUX_RRC_1B_20190923T181559_20190923T183211_0001	AE_OPER_AUX_PAR_1B_**20190819**T030159_20191104T040112_0002
:	:	:
Original:	AE_OPER_AUX_RRC_1B_20200224T181459_20200224T183123_0001	AE_OPER_AUX_PAR_1B_**20190624**T000000_99991231T235959_0002
Reprocessed:	AE_OPER_AUX_RRC_1B_20200224T181459_20200224T183123_0001	AE_OPER_AUX_PAR_1B_**20200224**T030135_20200302T040112_0001

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 reprocessing and the NRT dataset show the same number of IRCs and the same dates.
- Concerning the identical version numbering of near-real time and reprocessed AUX_RRC and AUX_MRC files the same explanation holds as provided in the AUX_ZWC chapter above (email from 2021/06/23 Simone Bucci).
- One AUX_R/MRC file (*IRC #42*) of the reprocessed data set shows the version number 0002. The reason is unclear. (*A similar issue was found in the dataset of the 1st reprocessing.*)

- The AUX_MRC and AUX_RRC files show the expected content/structure (*additional information in reprocessed dataset compared to the NRT dataset due to updated processor baseline*), in particular the only new IRC from February 24th, 2020. Slight differences in specific values are explainable by the use of different QC thresholds in the AUX_PAR_1B file.
- The correct usage of the provided AUX_PAR_1B files during the reprocessing can be confirmed.
- Based on adapted AUX_PAR_1B parameters, the reprocessed dataset again (*as for the 1st reprocessing*) contains correctly calculated intercepts and slopes for the first four RRCs and MRCs #37 - #40 (*see Figure 5 and Figure 6*).
- Effects of improved quality control are visible for the Rayleigh atmosphere & ground return response curves.
- An improved random error of the residuals of the Rayleigh atmosphere and ground return response curves after subtraction of a polynomial fit of 5th order can be observed (*see Figure 5 and Figure 6*).
- Effects of improved quality control are visible for Mie ground return response curves (*in the non-linearity*).
- Small differences in slopes and intercepts of both, Mie internal reference and atmosphere are visible and are most likely caused by differences in the MieCore parameters and/or quality control parameters in the AUX_PAR_1B file.

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

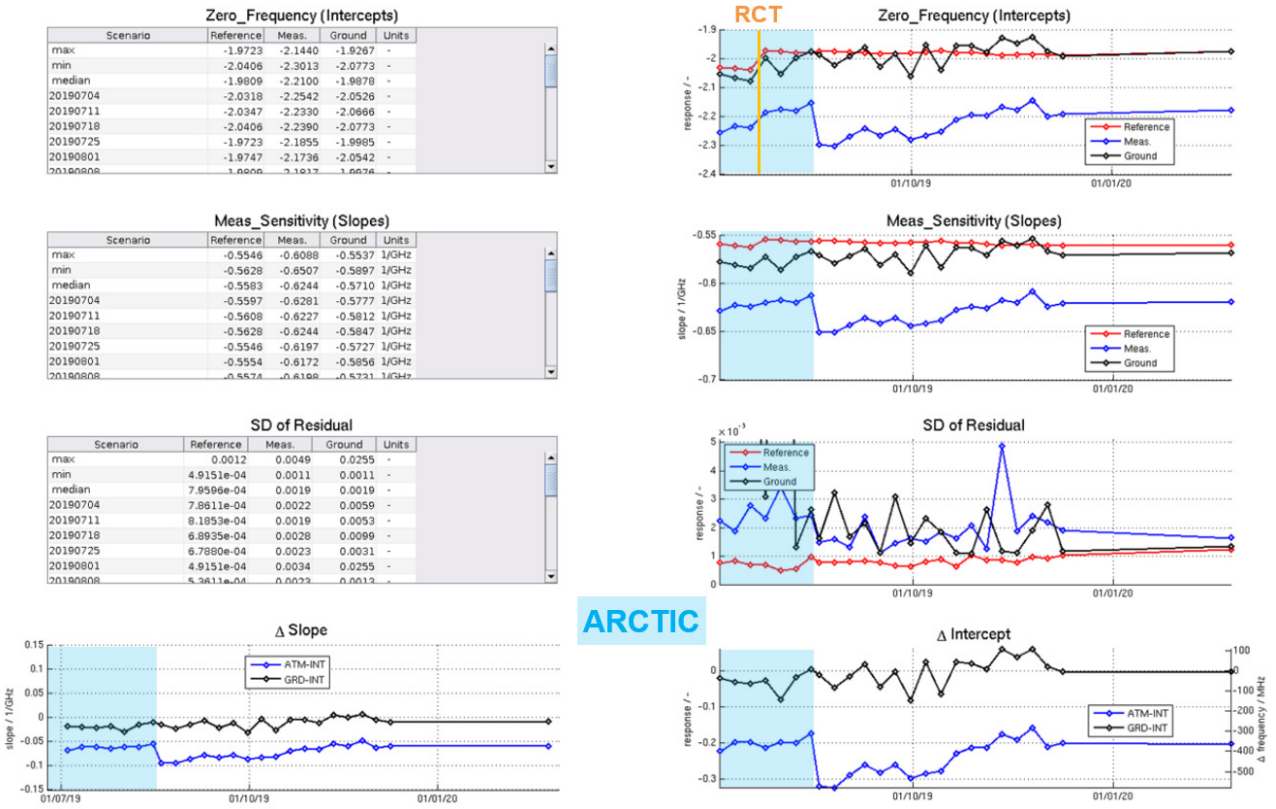


Figure 5: Rayleigh response calibration intercepts, slopes and residuals of the original dataset.

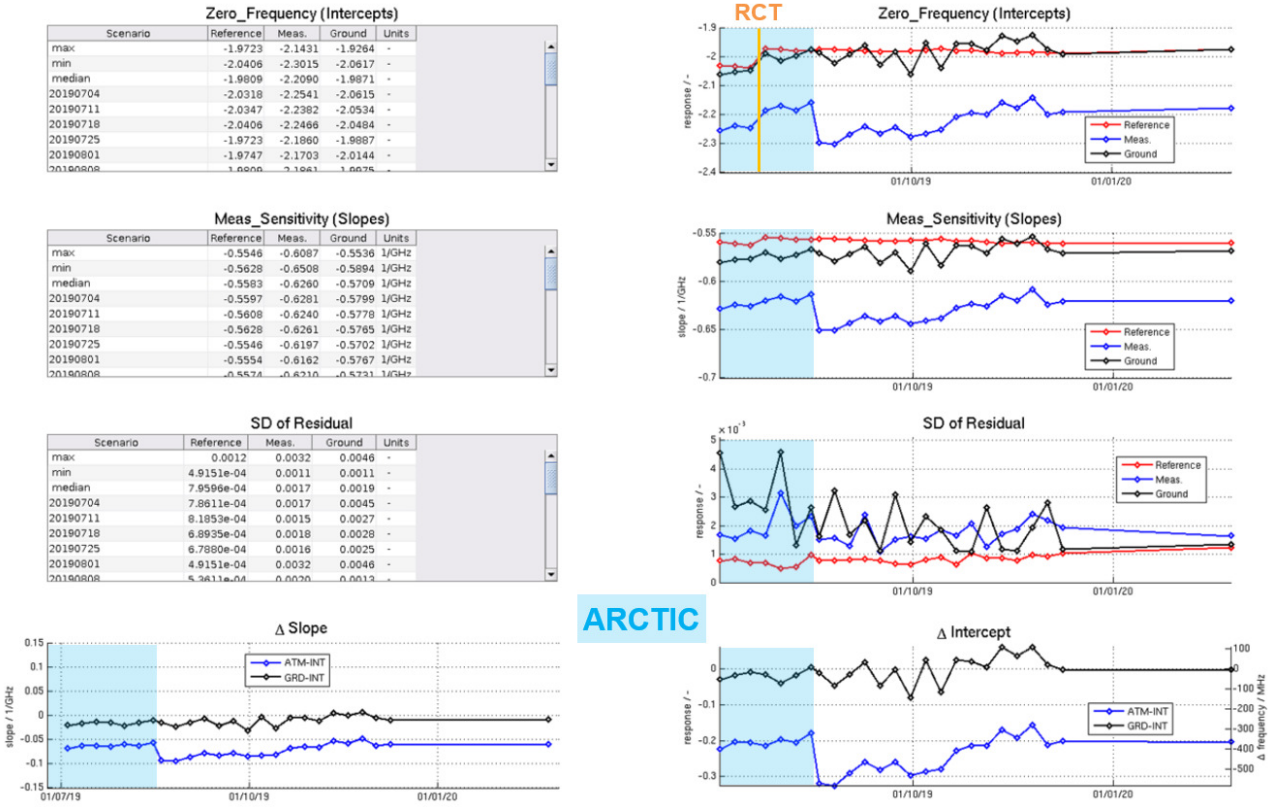


Figure 6: Rayleigh response calibration intercepts, slopes and residuals of the reprocessed dataset.

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 can be found in [RD-1], [RD-7] and [RD-8]. 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. These measurements provide a map of the current dark signal state of the ACCD which can be used for the correction. However, further analysis showed that some hot pixels show step-like dark current signal transitions. The transitions occur at random and unpredictable times and manifest as steps in the atmospheric return signals. Due to the update frequency of 6 hours for DCMZs, the applied dark current signal correction is imperfect. In case a hot pixel changes its level, the wind and aerosol products are biased until the next DCMZ measurement is performed and used for the correction. For the reprocessing there is the possibility to detect hot pixel induced steps in the atmospheric return signals, i.e. in ALD_U_N_1A wind signals. With this information artificial DCMZ can be created to correct for remaining hot pixel induced signal steps.

5.2 Hot pixel characterization

A complete history of hot pixel appearances can be found in the blacklisted data section of the Confluence Wiki [RD-9]. At the start of the reprocessing period in July 2019, 15 hot pixels (7 in the Mie channel and 8 in the Rayleigh channel) were already present and 22 additional pixels (13 Mie and 9 Rayleigh) got activated during the reprocessing period. Table 3 and Table 4 list hot pixels which were present during the reprocessing period. The column “Detectable” indicates if the correction method described in the following sections could be applied to the hot pixel. This is necessary because for some hot pixels (e.g. Rayleigh pixels which are covered by one of the two Rayleigh filter spots i.e. Rayleigh column positions 1, 2, 9 and 10), the atmospheric signal contribution is usually too dominant to detect dark signal induced steps on top of the large atmospheric signal variation. Furthermore, hot pixels with dark current signal fluctuations less than 0.5 LSB at measurement level are flagged as non-detectable as well. The last column describes the characteristics of the hot pixel and the expected impact on the quality of the winds [RD-8].

5.3 Correction approach and generation of DCMZ files

The same approach as for the first reprocessing was also used for this campaign. The approach is described in detail in [RD-7]. It is based on analysing ALD_U_N_1A data which holds atmospheric return signals of both channels for each ACCD pixel. The atmospheric return signals are analysed in combination with dark current signal information provided by the DCMZ measurements. In case the dark current signal correction value significantly changed between two DCMZ measurements, the atmospheric signal previous to the dark current signal change is analysed. In most cases the hot pixel induced steps in the atmospheric signals are detected by visual inspection. In some cases, it was possible to apply a CUSUM based approach to get the index of the step automatically.

Table 3: List of Mie hot pixels during the reprocessing period. Red shaded pixels got active during the reprocessing period. The first column describes the position of the pixel in the memory zone of the ACCD. The both entries indicate the row (0 to 23) and the column position (0 to 15). RTS stands for "Random Telegraph Signal".

Pixel	Activation	Detectable	Impact
[15, 14]	before IOCV	x	Low (small RTS amplitudes with low fluctuation rate)
[23, 2]	before IOCV	x	Low (sporadic mean shifts)
[12, 8]	2018-10-21	✓	High (multi-level RTS)
[1, 14]	2018-10-24	✓	Low (two level RTS with low rather low fluctuation rate)
[4, 12]	2019-01-09	x	Low (sporadic mean shifts)
[19, 1]	2019-03-31	✓	High (multi-level RTS)
[9, 12]	2019-04-26	✓	High (multi-level RTS)
[3, 10]	2019-07-19	✓	Medium (two-level RTS with rather low amplitudes)
[8, 12]	2019-08-08	✓	High (multi-level RTS)
[4, 10]	2019-10-03	✓	Low (sporadic mean shifts)
[3, 2]	2019-10-03	✓	Low (sporadic mean shifts)
[1, 9]	2020-02-28	✓	High (multi-level RTS)
[14, 5]	2020-03-14	✓	Medium (rather stable Multi-level RTS)
[2, 2]	2020-05-10	✓	High (multi-level RTS)
[12, 6]	2020-06-28	✓	Low (sporadic mean shifts)
[6, 1]	2020-07-17	✓	High (Two-level RTS with high fluctuation rate)
[23, 12]	2020-07-24	✓	Low (sporadic mean shifts)
[7, 6]	2020-08-17	✓	High (multi-level RTS)
[17, 1]	2020-09-13	✓	High (multi-level RTS)
[7, 12]	2020-09-17	✓	Low (sporadic mean shifts)

The detected hot pixel step can be corrected by updating the dark current signal correction of the affected pixel in the DCMZ file. For this, depending on the location of the hot pixel step, it may be necessary to create an artificial DCMZ measurement at the right position or to modify an existing one. In the latter case, nothing of the DUDE file is changed except the dark current rate value of the involved pixel. In case, a DUDE file has to be created, a template based on a DCMZ file processing using L1bP 7.09.1 is used where besides the dark current rates only certain fields of the header are adapted. This is needed to make sure that the L1B processor selects the correct DCMZ file. For the update of the dark current rates the content of the DUDE file closest prior to the generated file is copied and only the dark current rate of the involved pixel is changed. Note that the dark current signal correction in the L1B processor works at orbit level and not at observation level. Thus, there are still hot pixel related effects which cannot be fully corrected.

For the complete period of 2nd reprocessing campaign, the AUX_DCMZ1B files had to regenerated on the Sandbox using L1bP 7.09.1 as a result of a major file format change. For the first time DCMZ correction for the solar background range gate (range gate #25) has been introduced. The dark current correction value for each solar range gate pixel are reported in a new parameter in the AUX_DCMZ _1B files.

Table 4: List of Rayleigh hot pixels during the reprocessing period.

Pixel	Activation	Detectable	Impact
[10, 1]	2018-09-07	x	High (multi-level RTS)
[4, 1]	2018-11-04	x	Low (sporadic mean shifts)
[14, 3]	2018-11-24	x	Low (sporadic mean shifts, low amplitude)
[19, 9]	2019-01-27	x	High (multi-level RTS)
[0, 6]	2019-02-20	✓	Medium (stable multi-level RTS)
[10, 15]	2019-03-17	✓	High (multi-level RTS)
[2, 1]	2019-05-08	x	High (multi-level RTS)
[10, 7]	2019-06-15	x	Medium (Two-level RTS with low fluctuation rate)
[19, 1]	2019-08-01	x	High (multi-level RTS)
[19, 15]	2019-08-17	✓	Low (sporadic mean shifts)
[23, 3]	2019-08-29	✓	Medium (multi-level RTS, rather low amplitude)
[23, 5]	2019-12-21	✓	High (multi-level RTS)
[7, 9]	2020-01-28	✓	Low (sporadic mean shifts)
[6, 14]	2020-04-18	✓	Medium (stable multi-level RTS)
[13, 1]	2020-05-07	x	Medium (Two-level RTS with low amplitudes)
[23, 9]	2020-07-06	✓	Medium (stable multi-level RTS)
[2, 12]	2020-08-22	✓	High (multi-level RTS)

For the 1st reprocessing campaign covering the time period between 1 July 2019 to 31 December 2019, the DCMZ package including the created files was already prepared using L1bP 7.08 and validated [RD-7]. However, due to the different DCMZ file formats between L1bP 7.08 and 7.09, it was not possible to use these files for the second reprocessing campaign due to the missing solar background information. To overcome this issue, the closest L1bP 7.09 DCMZ file is selected for each created L1bP 7.08 DCMZ file. Then, information about the atmospheric dark current signals, i.e. range gates 1 to 24, from the old file was copied to the selected 7.09 DCMZ file. In addition, various file parameters such as the file name or the validity period were aligned. This made sure that each file of the old format also contains the solar background correction values and can be used in the reprocessing.

In preparation of the creation of DCMZ files for the period starting at 1 January 2020, the nominally measured DCMZ files of this period also had to be reprocessed on the Sandbox using L1bP 7.09.1. This step was necessary to make sure the created DCMZ files, which are a modified copy of the measured files, are in the correct file format and contain the solar background information.

For the complete reprocessing, 680 additional DCMZ files were created and 263 of the nominally measured DCMZ were modified. Note that the file version of the modified measured files was incremented by one to make sure that those files are preferred over the nominally measured files during the reprocessing. For the sake of brevity, a complete list of all provided DCMZ files is not given in this document.

5.4 Validation

The outcome of this task was validated for the following test weeks ([RD-10], [RD-11] and [RD-12]):

- 4-10 August 2019
- 18-24 December 2019

- 23-29 March 2020
- 22-28 June 2020
- 14-20 September 2020

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 pixel-wise validation step is only possible for Mie hot pixels.

Figure 7 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. Another example is depicted in Figure 8 for Mie hot pixel [17, 1] on 15 September 2020.

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-10], [RD-11], or [RD-12]. In summary, it can be stated here that both the file selection and scientific results were as expected.

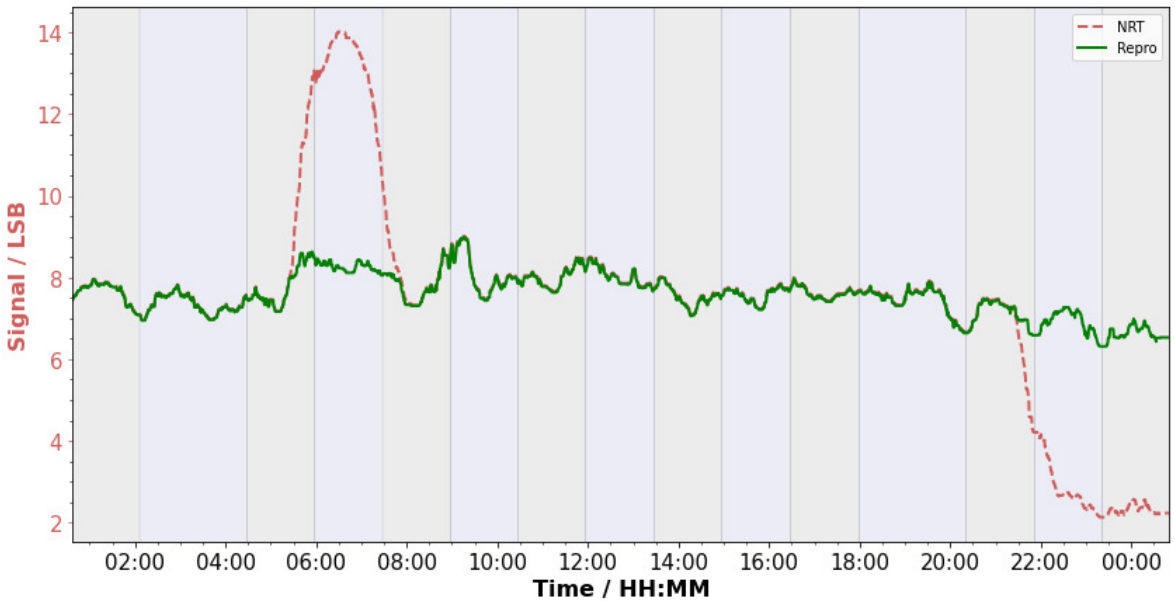


Figure 7: Median filtered (window size of 400 observations) L1B signals of Mie hot pixel [12, 8] on 25 March 2020 for the near-real-time (NRT) and reprocessed (repro) dataset.

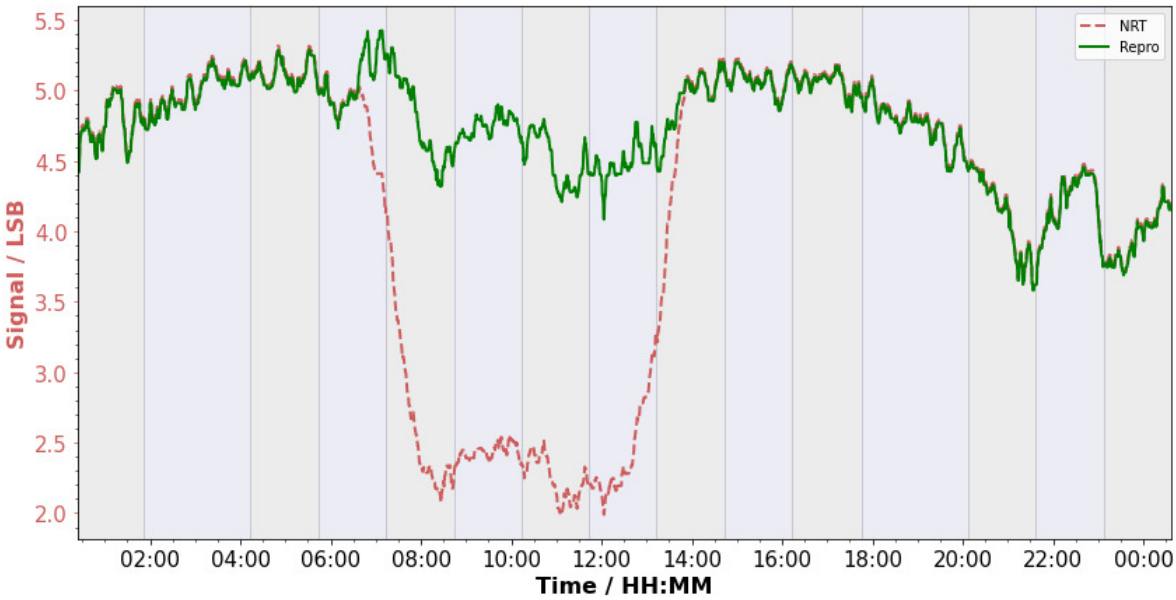


Figure 8: Median filtered (window size of 400 observations) L1B signals of Mie hot pixel [17, 1] on 15 September 2020 for the near-real-time (NRT) and reprocessed (repro) dataset.

6 Verification of L2A products

6.1 Overview of impacts on L2A products

6.1.1 Improvements and limitations

The data is homogeneously processed with L2A v3.11.1 which means:

- A new radiometric correction of the calibration coefficients K_{ray} and K_{mie} (i.e. coefficients characterizing the radiometric efficiency of the receiver) has been implemented into L2A processor V3.11.1 taking into account the variations along the orbit. The coefficients estimated from signal prediction in particle-free regions of the atmosphere for mid-altitudes 6 to 16 km are now 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) which is computed to fit the K_{ray} and K_{mie}.

Limitations:

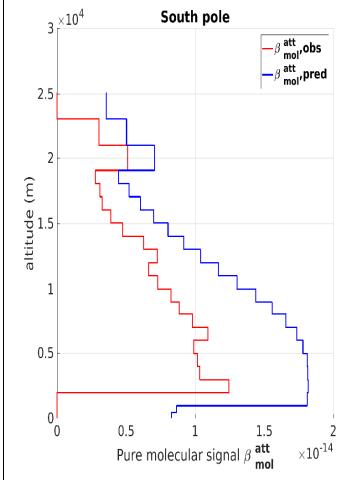
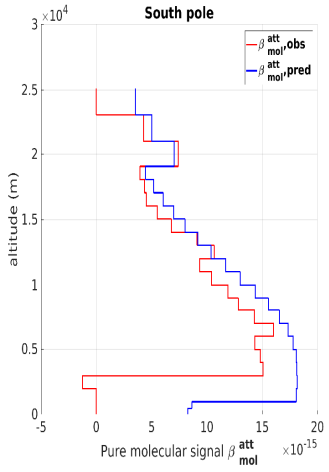
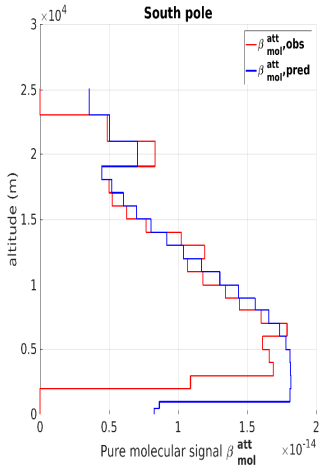
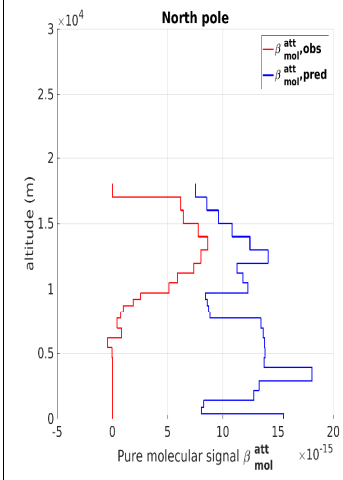
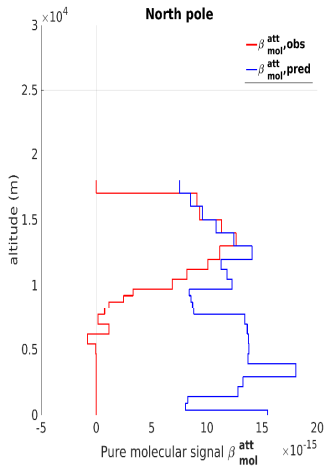
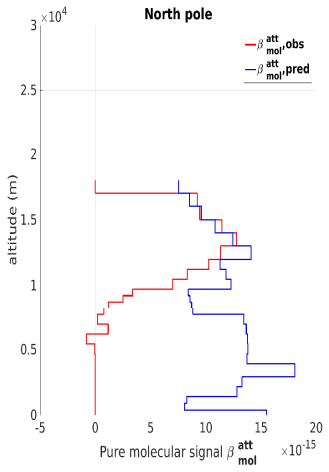
- During two weeks between 28/10/2019 and 10/11/2019, ALADIN used range bins settings designed for the validation of Atmospheric Motion Vectors (AMV). The L2A product shows limitations for the corresponding period: the Mie Core Algorithm (MCA) can be used, but the Standard Correct Algorithm (SCA) cannot be fully exploited, i.e. the AMV range bin settings not being properly handled for mid-altitudes where several Rayleigh bins match one single Mie bin. As a result, SCA products are only partially valid.

6.1.2 Example

The example below shows the improvement of the radiometric correction of K_{ray} and K_{mie} calibration coefficients. Within the first calibration corresponding to L2Ap V3.09 the K_{ray} and K_{mie} are taken from nadir pointing data (IRC). The second scheme implemented in L2Ap V3.10 calibration performed a signal prediction in aerosol-free regions of the atmosphere between 6 to 16km. An orbit averaged value is then applied on the whole orbit for each coefficient. The third version implemented in L2Ap V3.11.1 takes into account the in-orbit oscillations by fitting the coefficients using a multiple linear regression based on telescope temperatures information. Both K_{ray} and K_{mie} are then provided at observation scale, i.e. for each vertical profile.

As shown in Table 5 the SCA is overestimating the observed pure molecular signal in both clear sky and cloudy regions. Nevertheless, the implementation of new calibration schemes within L2Ap V3.10 and L2Ap V3.11.1 allowed to reduce these signal overestimation especially in mid altitudes.

Table 5: Two example profiles showing the improvement in calibration from L2Ap V3.09 to 3.10 and 3.11.1.

	V3.09: Calibration from IRC (nadir pointing)	V3.10: Calibration from signal prediction in clear sky	V3.11.1: Calibration from signal prediction in clear sky and regression based on telescope temperatures
1 profile in clear sky conditions			
1 profile in cloudy conditions			

6.2 Quick analysis of reprocessed dataset

6.2.1 K_{ray} and K_{mie} series

6.2.1.1 K_{ray} and K_{mie} from PDGS dataset

The K_{ray} and K_{mie} parameters, i.e. characterizing the radiometric performance of the two channels, deviate in valid range but significant oscillations can be observed for three distinct periods. In early month of FM-B period the orbit per orbit deviation appears to be higher as can be seen in Figure 9 and Figure 10.(e.g. red boxes). Higher variability can also be seen for the eclipse season (e.g. green boxes). Finally, significant oscillations are observable during AMV campaign in September to November 2019 (e.g. (black boxes).

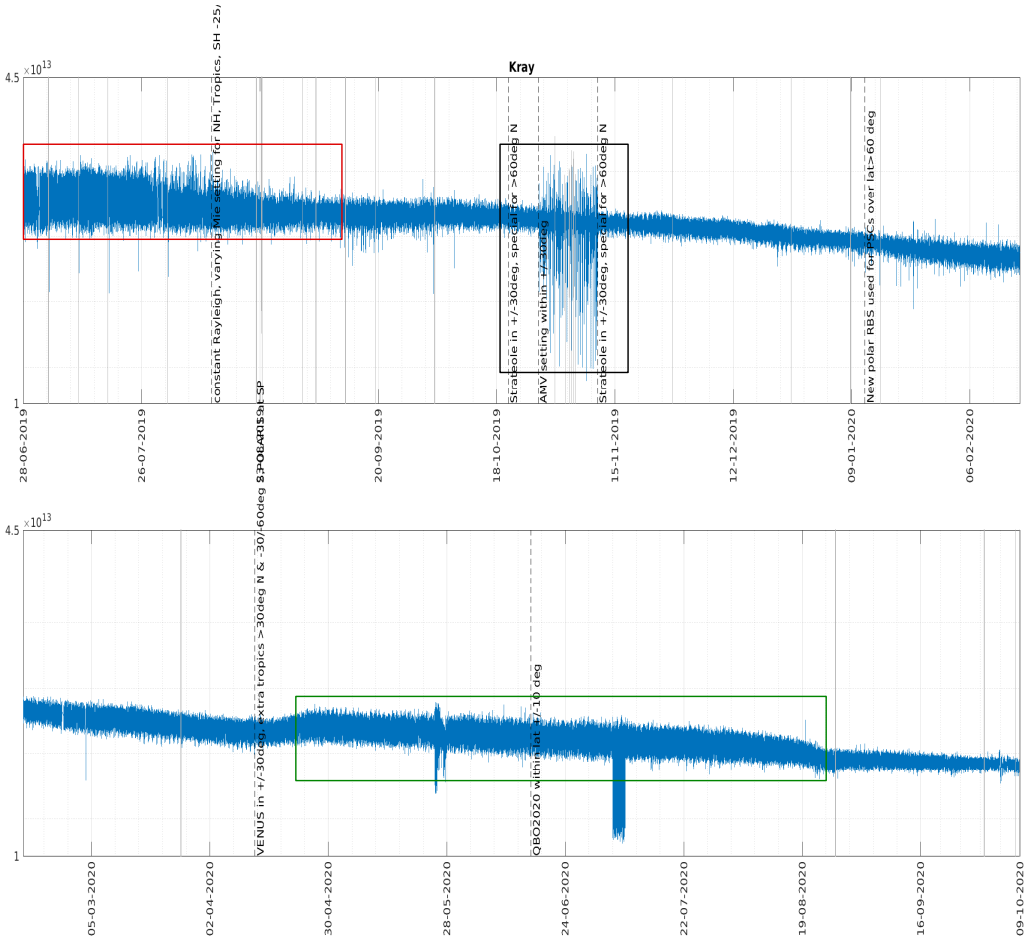


Figure 9: Time series of K_{ray} from reprocessed data

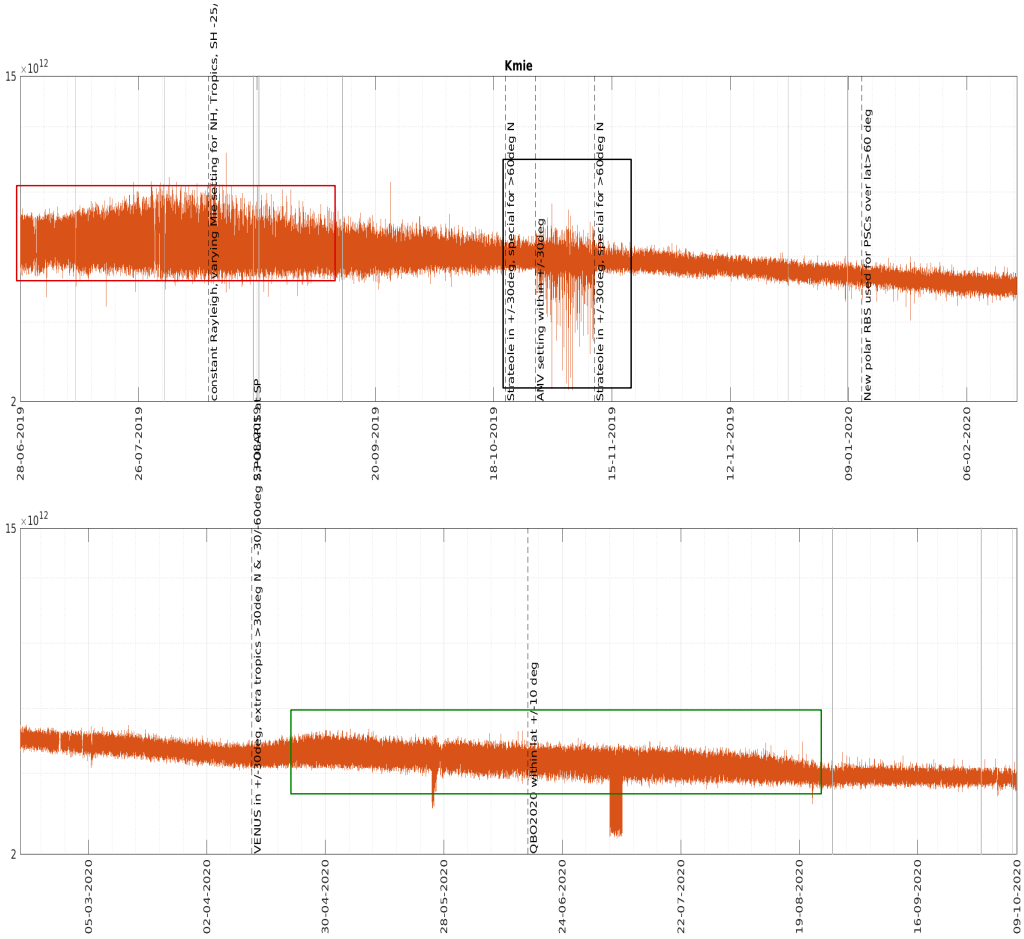


Figure 10: Time series of K_{mie} from reprocessed data

6.2.1.2 K_{ray} and K_{mie} intercomparisons to Sandbox reprocessing

In order to validate the PDGS second reprocessing of L2A data a check with the SANDBOX dataset has been performed. Small differences occurred for both coefficients, usually << 1% and the highest differences can be seen in Early FM-B period as shown in Figure 11 and Figure 12

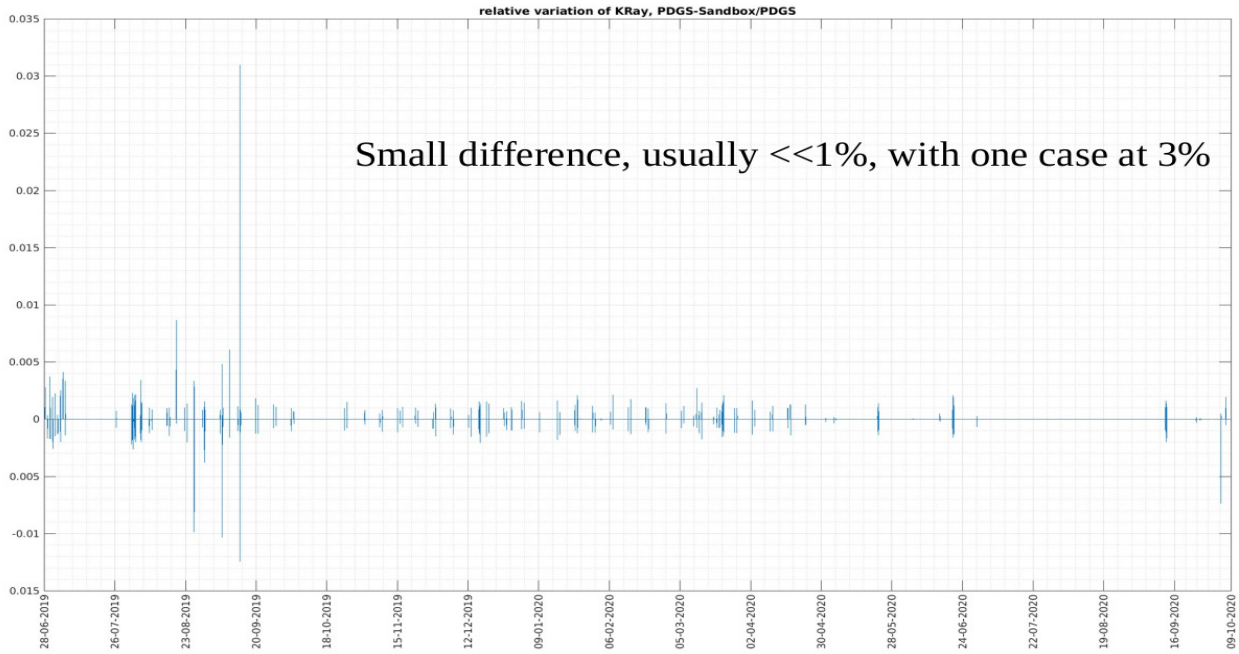


Figure 11: K_{ray} intercomparison between PDGS and SANDBOX data

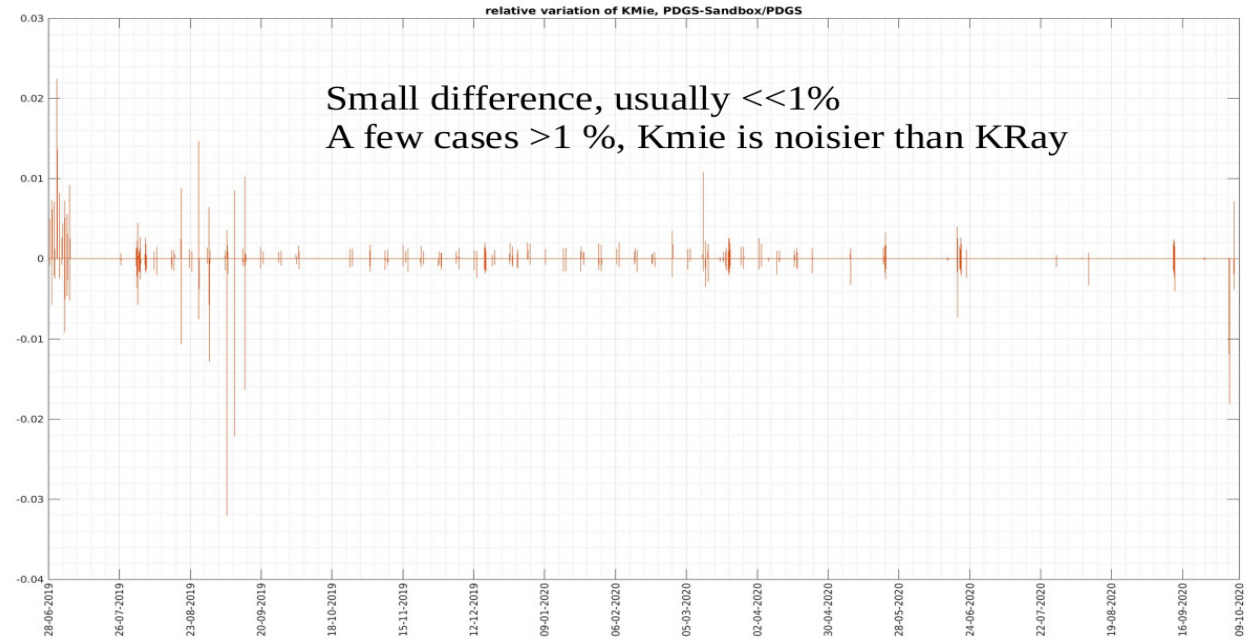


Figure 12: K_{mie} intercomparison between PDGS and SANDBOX data

6.2.2 Outliers in SCA backscatter and extinction

The backscatter values are sorted in 3 categories to easily characterize the occurrence of anomalous values. Fraction of anomalous pixels in backscatter: Backscatters with values larger than $200 \cdot 10^{-6}$ $sr^{-1} \cdot m^{-1}$, with values ≤ 0 or not computed and set to -1 are shown in Figure 13. Abrupt changes (e.g. during the first two weeks of November 2019) are caused by changes in the Range Bin Settings. Thin or thick bins change the influence of noise on the L2A retrieval.

During the first two weeks of November, a special setting was used for the validation of AMV winds and the L2A could not apply its SCA on this data, making the amount of "-1" rise. In the product, a large fraction of extinction values is set to 0 by the algorithm as can be seen from Figure 14. Changes of RBS are then less visible but we observe the same pattern as for the backscatter.

A check of the mid-bin extinction has also been made as illustrated in Figure 15. The extinction being normalized using the signals in the first top bin with low SNR, the noise is propagating downwards and leads to over/underestimation of the extinction coefficient. The mid-bin vertical scale was then designed to overcome the oscillating propagation of errors by averaging two consecutive bins. The fraction of anomalous pixels in mid-bin extinction is significantly reduced compared to extinction product (i.e. from ~0.7% to ~0.3%). An impact of the July 2020 M1 temperatures test can be seen (e.g. red box).

A higher variability of the backscatter can be seen in early FM-B period.

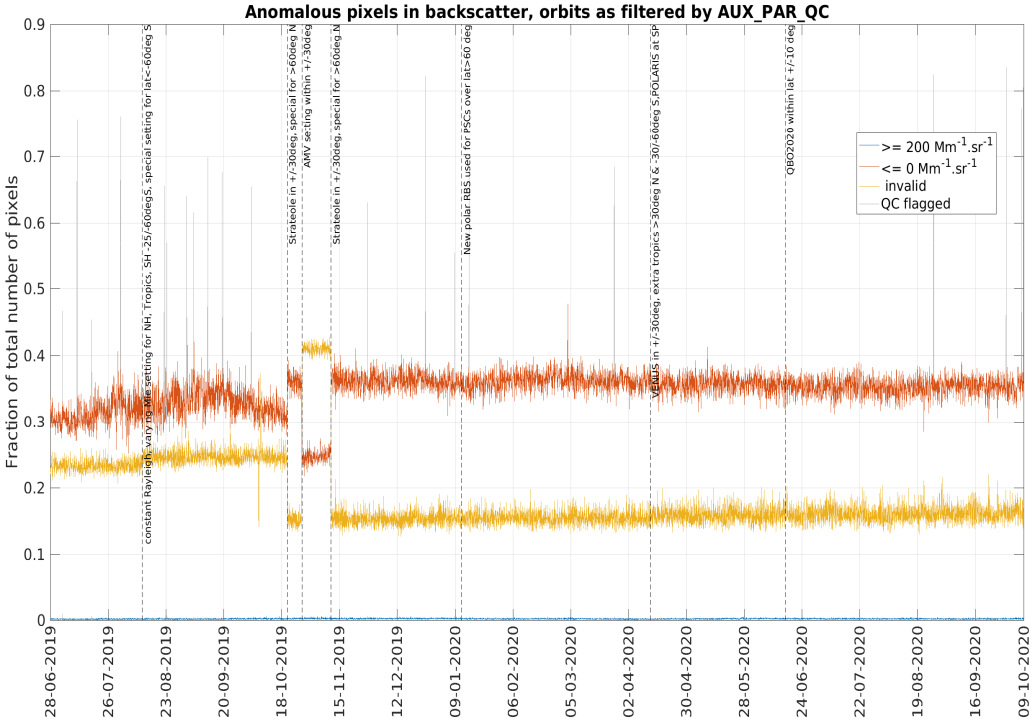


Figure 13: Fraction of anomalous pixels in backscatter: Backscatters with values larger than $200 \cdot 10^{-6}$ $sr^{-1} \cdot m^{-1}$ (blue), with values ≤ 0 (orange) or not computed and set to -1 (yellow). Note that the blue line is very close to zero.

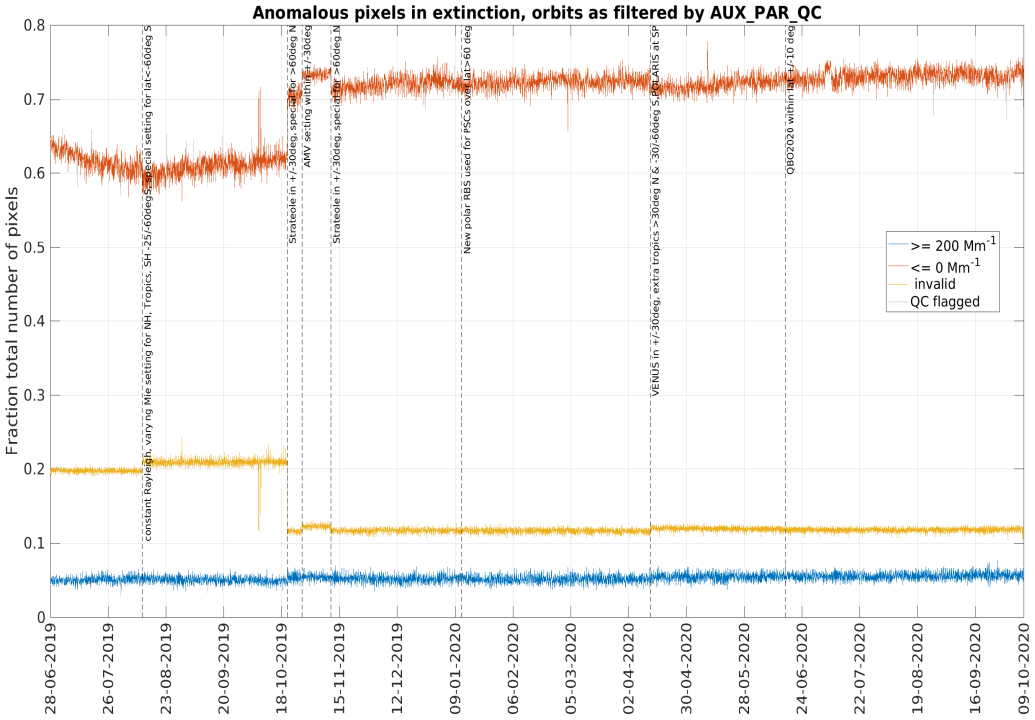


Figure 14: Fraction of anomalous pixels in extinction: extinction with values larger than $200 \cdot 10^{-6}$ m^{-1} (blue), with values ≤ 0 (orange) or not computed and set to -1 (yellow).

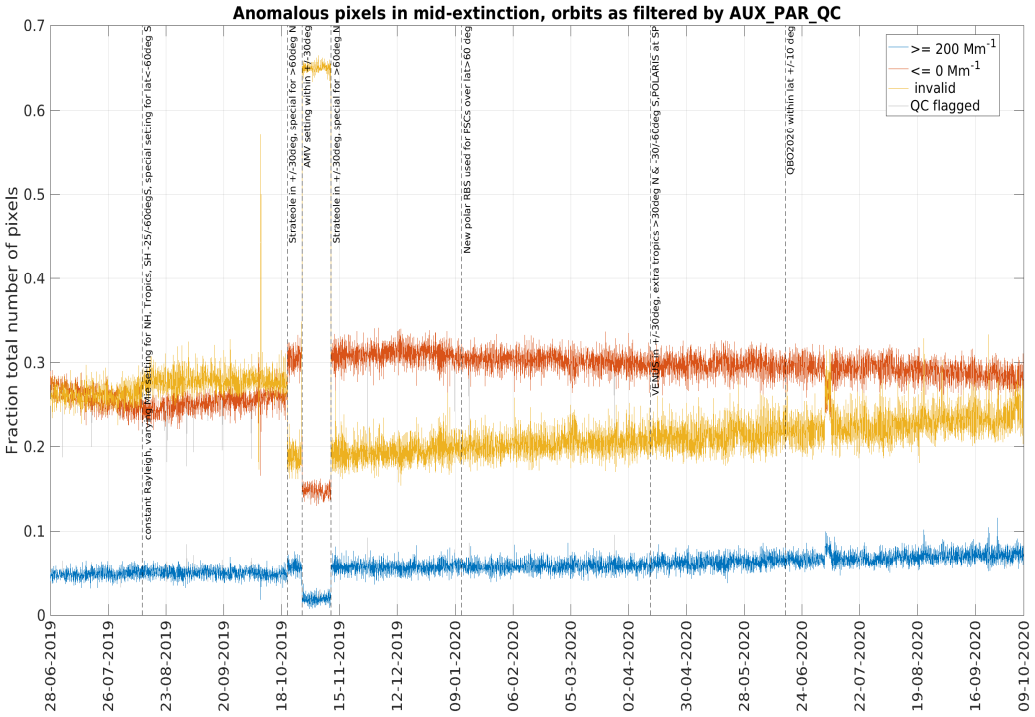


Figure 15: Fraction of anomalous pixels in mid-extinction: mid-bin extinction with values larger than $200 \cdot 10^{-6}$ m^{-1} (blue), with values ≤ 0 (orange) or not computed and set to -1 (yellow).

6.2.3 Outliers in MCA extinction

In the MCA extinction, very few pixels exceed $200 \cdot 10^{-6}$ $sr^{-1} \cdot m^{-1}$ and there is no negative values in output. Changes of RBS within AMV campaign period are visible. A higher variability and a continuous decrease of anomalous pixels can be seen for early FM-B period.

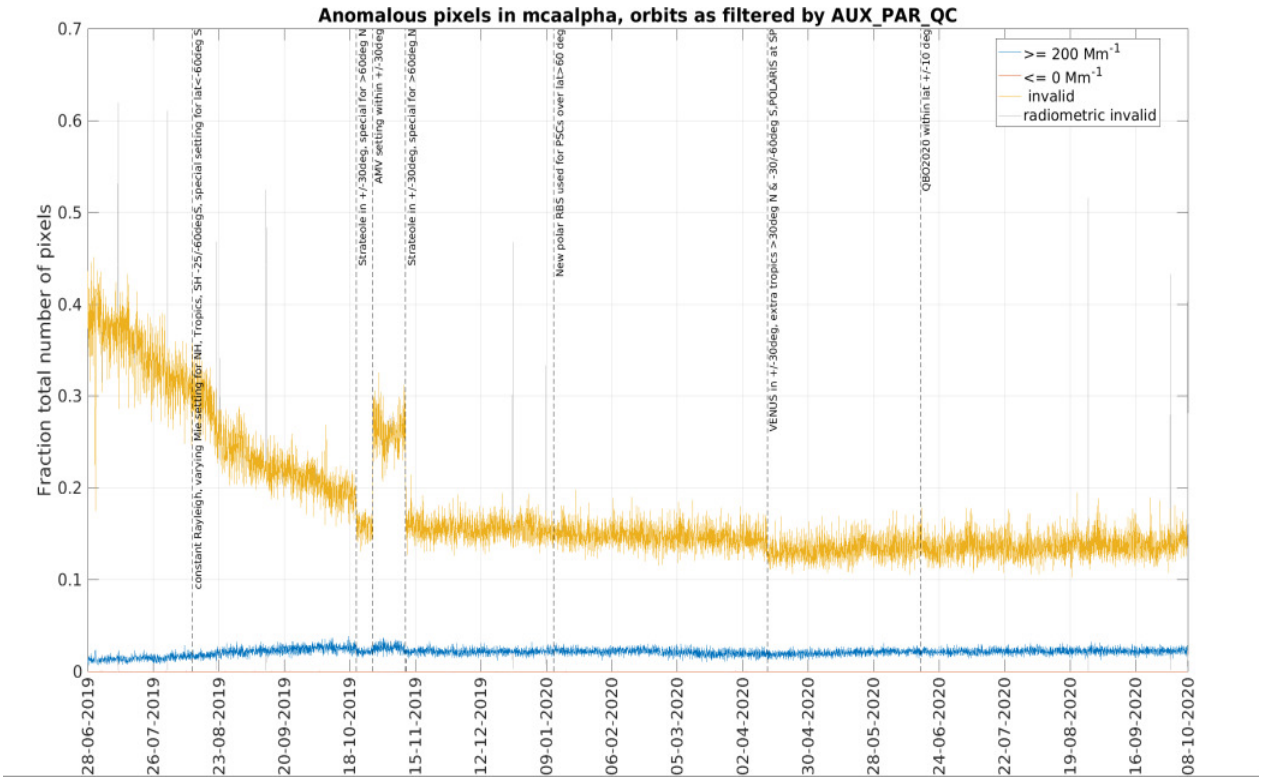
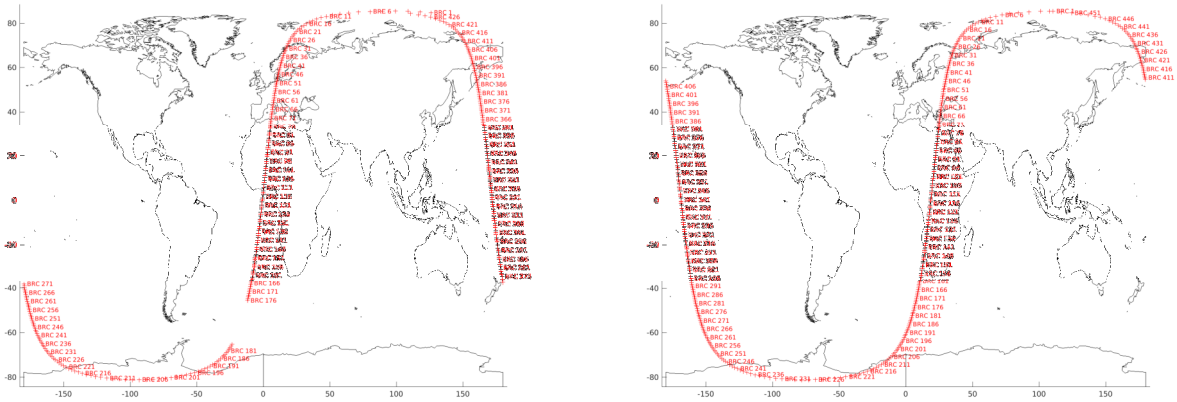


Figure 16: Fraction of anomalous pixels in extinction: extinction with values larger than $200 \cdot 10^{-6}$ m^{-1} (blue), with values ≤ 0 (orange) or not computed and set to -1 (yellow).

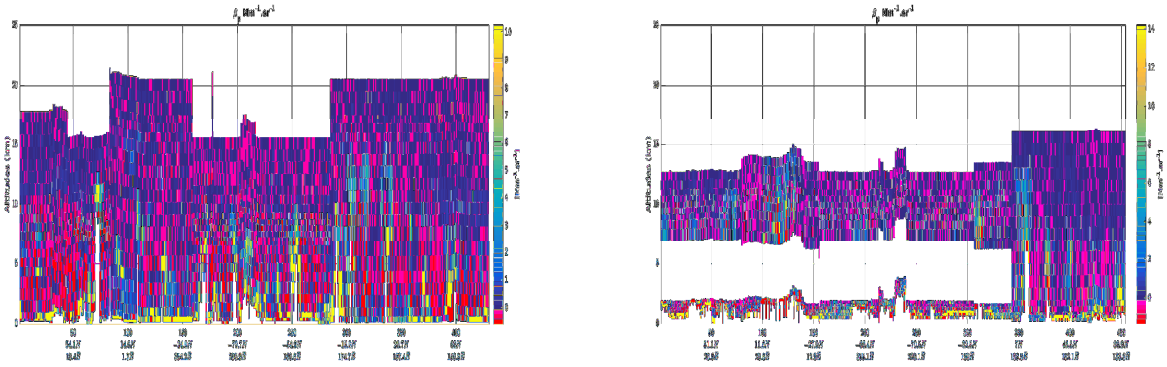
6.3 Visual inspection of a few orbits

6.3.1 Orbits 6507 and 6522 (7 Oct and 8 Oct 2019) with peaks in anomalous pixel fraction

Around the beginning of October, we can spot two peaks in anomalous pixels as shown in Figure 13 and Figure 14. The peaks correspond to two orbits starting on 7 Oct 2019 at 5:31 (orbit #6507) and 8 October 2019 at 4:14 (orbit #6522). As they are happening with ~24-hour interval, they sample the same area. It only happens for these two days. Closer look at backscatter and extinction product reveals partially valid data for orbit 6522 in mid-altitudes due to AMV settings.

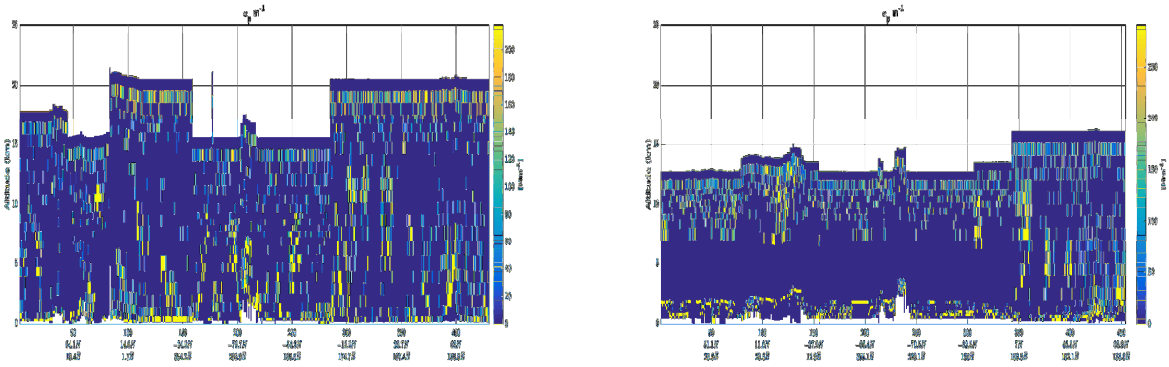


Geolocation of orbits 6507 (left) and 6522 (right).



Backscatter of orbits 6507 (left) and 6522 (right)

Figure 17: Geolocation and backscatter of orbits 6507 and 6522.



Extinction of orbits 6507 (left) and 6522 (right)

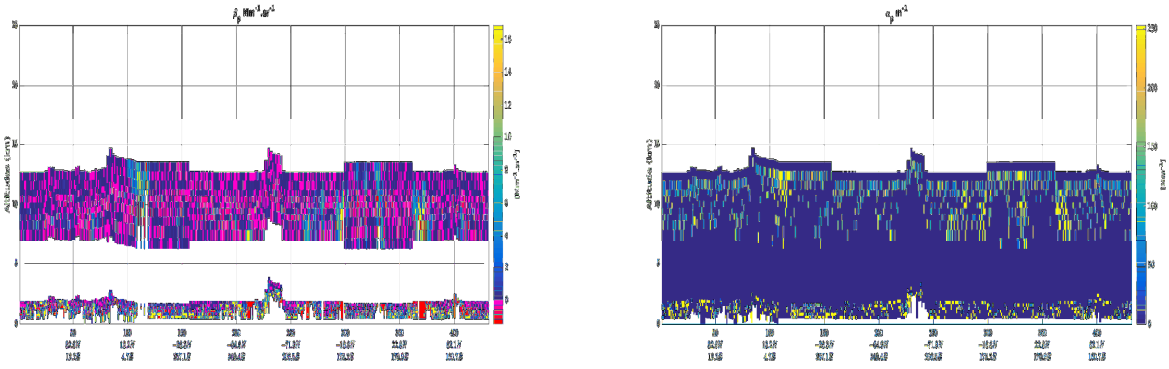
Figure 18: Extinction of orbits 6507 and 6522.

6.3.2 Orbit 7046 with AMV range bin settings

Orbit 7046 was acquired on 10 November 2019, starting at 5:18 UTC. During the first two weeks of November, the range bin setting was set to facilitate the validation of Atmospheric Motion Vectors from geostationary satellites. This RBS cannot be processed by the current L2A because of the specific matching between Mie and Rayleigh (It cannot process two Rayleigh bins in one Mie bin).

This is visible as a clear band in the backscatter in

Figure 19. The extinction is supposed to be 0 within this band and the backscatter product has been set to not a number (NaN) values.



Backscatter (left) and extinction (right) retrieved from orbit 7046

Figure 19: backscatter and extinction retrieved from orbit 7046.

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 2019-06-28T15:22:23 to 2020-1009T23:46:22.
- Total number of DBL (and HDR) files is 6953 pairs of files.
- There is only pair of files (DBL and HDR) with duplicate versions. Both versions are identical except for the version number and the processing time.
- There are files from two different version numbers Their distribution is shown in Table 6.

Table 6: Number of pairs of files for each file version.

Version number	0001	0002	Total
Number of files	6405	548	6953

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 61 Mie clear wind observations during the whole period of more than 15 months, and since we believe these are based on noise only, this class is not further discussed.

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 second reprocessed dataset is shown in Figure 20. PDF's of both Aeolus and ECMWF model (collocated with Mie cloudy observations) are shown. The Mie cloudy PDF shape shows some irregularities especially in the range of HLOS values of -80 to 60 m/s. Furthermore, Mie cloudy PDF shows 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.

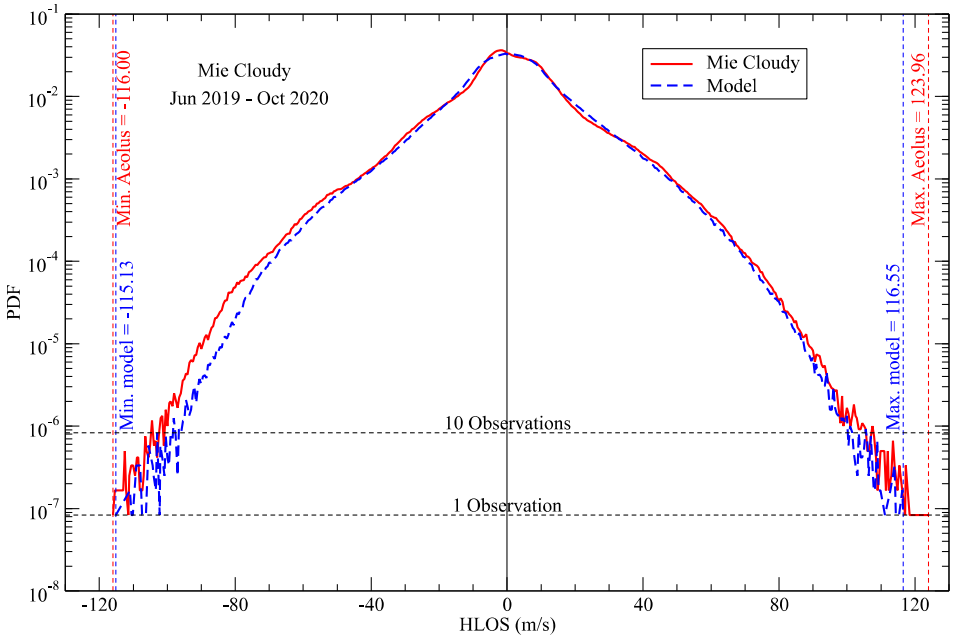


Figure 20: Probability density function (PDF) of Mie cloudy HLOS winds from the whole second reprocessed dataset. The corresponding PDF of ECMWF model collocated with Mie cloudy observations is also shown. The maximum and minimum (maximum of absolute negative values) for both Mie cloudy and the model are also shown.

Figure 21 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 21, the number of L2B Mie cloudy wind observations that pass quality control in the second reprocessing is slightly higher than that of the first reprocessing. The number of valid observations, however, shows mild gradual decline by time as it was the case for the first reprocessing. The gradual loss in the atmospheric path signal during this period is the main candidate responsible for this reduction. This reduction is not so clear in the NRT data due to changes in processing settings with time.

Figure 21 suggests that the bias, compared to ECMWF model, is small (bias of most individual orbits, not shown, varies between -0.4 m/s and +0.4 m/s) with an average of about -0.25 m/s. This is slightly higher than the first reprocessing by few centimetres per second (mainly due to the enhanced bias that started towards the end of November 2019). During the periods between early October and the mid-November 2019 and between mid-February and mid-March 2020, the wind biases of ascending and descending orbits diverge.

The random differences between L2B and the model winds is represented by SMAD. The SMAD values during this period varies around 3.5 m/s as can be seen in Figure 21. This is relatively low value. The two-week period 2019/10/28 – 2019/11/10, when a special range-bin setting (250-500 m bins globally) was used for the

purpose of comparison to Atmospheric Motion Vectors (AMVs), witnessed enhancements in SMAD values. The SMAD (and SDD, not shown) values show mild gradual increase with time for both processings. Again, the gradual reduction in atmospheric path signal during this period is the one to be blamed for that.

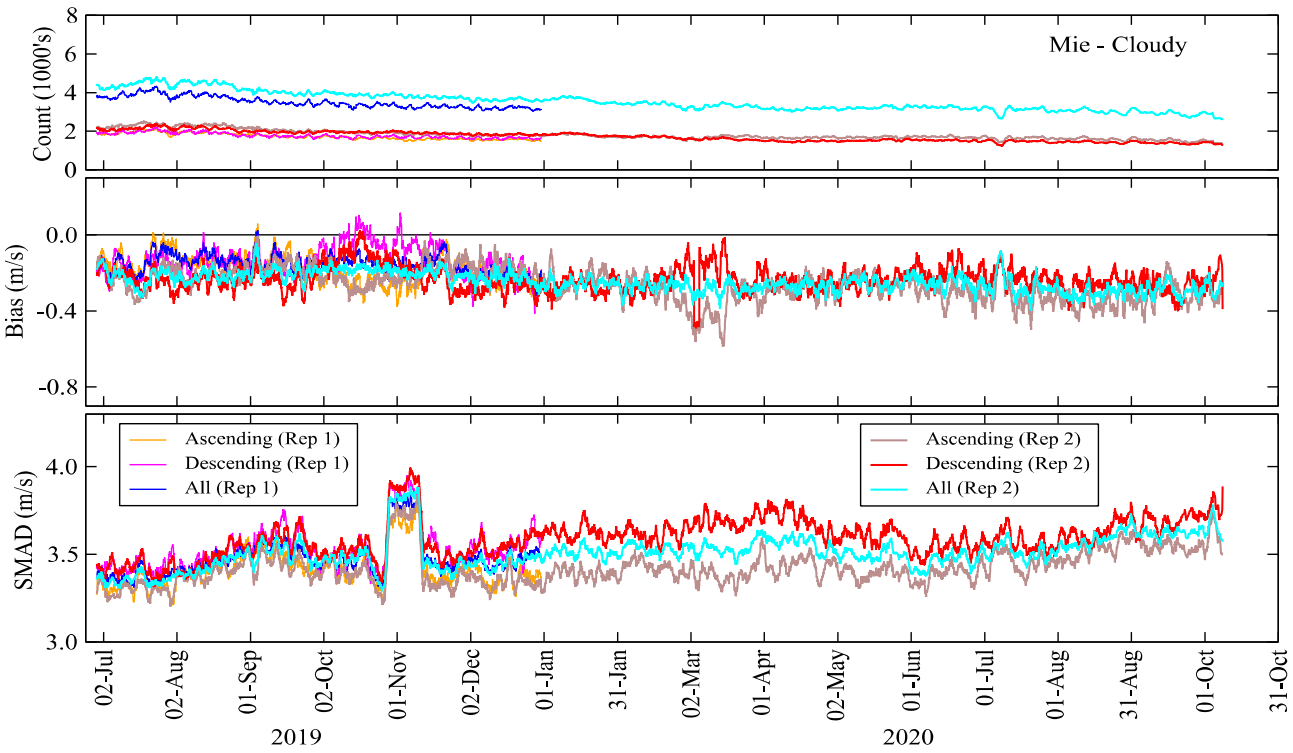


Figure 21: Time series of number of valid L2B Mie cloudy observations, HLOS wind speed bias, and scaled (i.e. multiplied by 1.4826) median absolute difference (SMAD) between L2B Mie cloudy and ECMWF IFS model winds. Running means along 30 successive files (orbits) are shown. Curves for all data as well as for ascending and descending orbits are shown for the both first (Rep 1) and second (Rep 2) reprocessings. Note that the first reprocessing curves end at the end of 2019.

One distinct feature of the SMAD curves in the lower panel of Figure 21, is the separation of the ascending and descending SMAD values. The data from ascending orbits show consistently lower SMAD values. During the period from December 2019 and May 2020, the difference exceeds 10%. The greater cloud cover in ascending (18 local solar time) versus descending (06 local solar time) due to increased convection may be responsible for this feature.

The comparison between L2B Mie cloudy HLOS wind against its model counterpart is shown in Figure 22 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 ~25 million collocations of L2B and model wind is very good with a global bias of -0.24 m, SDD of 3.9 m/s and SMAD of 3.5 m/s. The correlation coefficient is 0.973. 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. Mie cloudy winds are slightly lower than the model for HLOS wind speeds below -70 m/s and slightly higher for HLOS winds above 75 m/s.

One of the features of the scatter plot in Figure 22 is the variable bias (location of circles and crosses with respect to the diagonal) with respect to the HLOS wind values. This can be exemplified in Figure 23 which shows the variation of number of valid observations, bias and SDD (compared to ECMWF model) as functions

of HLOS wind for all second reprocessed dataset. The "wobbly" shape of the bias curve cannot be missed. In fact, the shape of the bias variability with respect to HLOS wind changes with time. Figure 24 shows the change of Mie cloudy HLOS wind bias compared to ECMWF model as function of HLOS wind for 5 different weeks (4-10 August 2019, 18-24 December 2019, 23-29 March 2020, 22-28 June 2020 and 14-20 September 2020) of the second reprocessed dataset. The locations of peaks and troughs of the weekly bias curves shift with time. This also can be inferred from the monthly scatter plots shown in Figure 25. This Mie cloudy non-linearity and the migration of the peaks with time was studied and a solution was found by members of the Aeolus DISC consortium [RD-13]. The solution will be implemented in the forthcoming reprocessings to alleviate this undesirable feature.

The SDD of Mie cloudy observations with respect to the model in Figure 23 shows some irregular behaviour. For example, there is a change in the SDD values at around -15 m/s with values lower that that have higher SDD values. This may be a consequence of the "wobbly" shape of the bias curve.

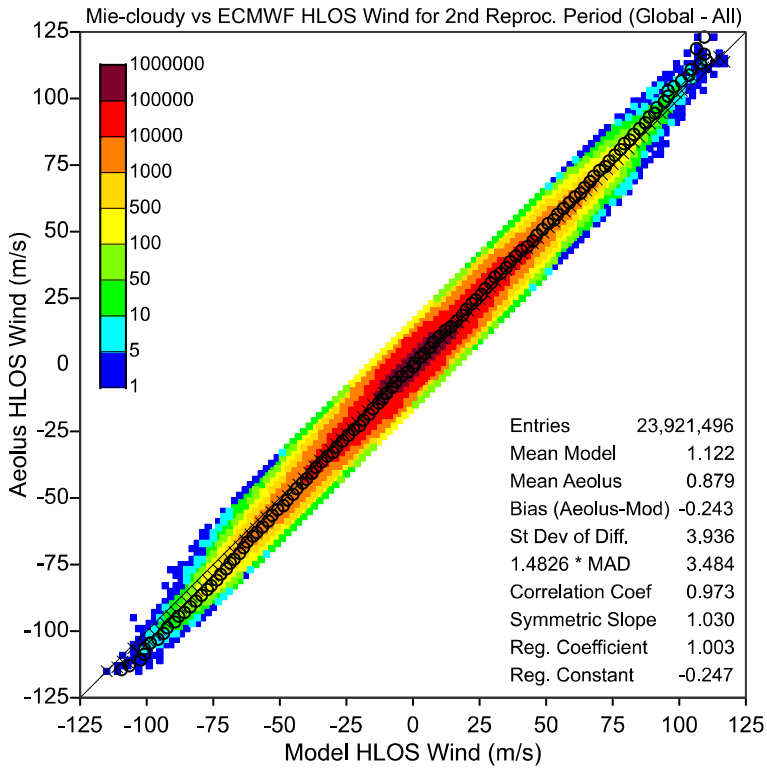


Figure 22: Density scatter plot comparing L2B Mie cloudy to ECMWF model HLOS winds over the whole globe during the whole period of the second reprocessing. Boxes are colour coded based on the number of collocations within the box. The circles are the average model value given L2B value while the crosses are the other way around. Statistics are given in the lower right corner of the plot.

The second reprocessed Mie cloudy dataset was discriminated based on the geographical location being in extratropical Northern Hemisphere (NH), Southern Hemisphere (SH) or in the Tropics. 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 difference in statistics between the NH and SH is very small.

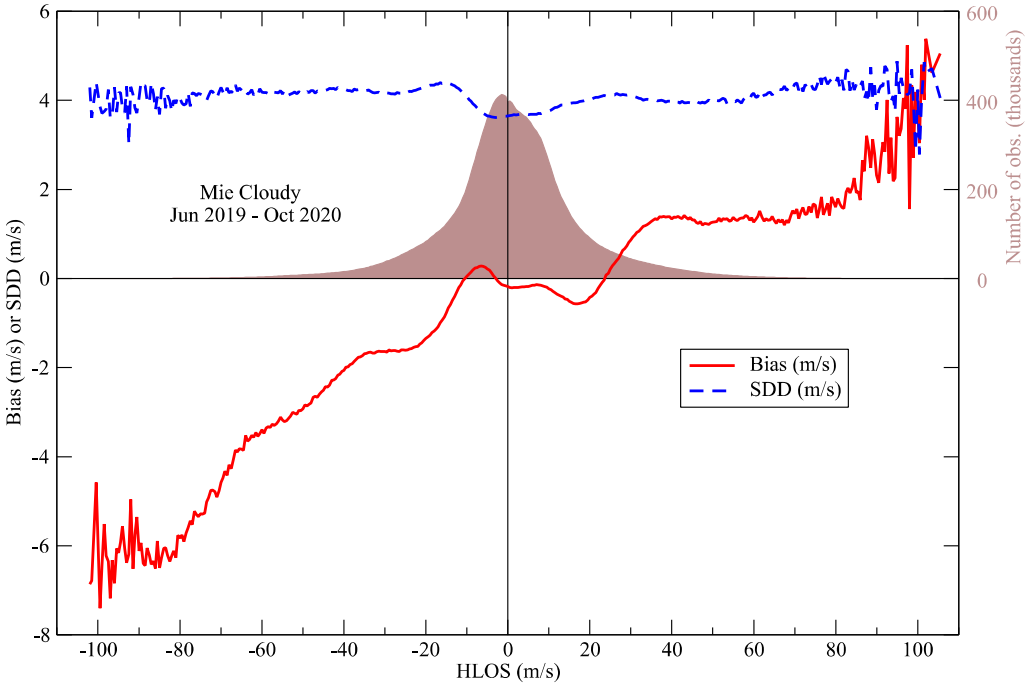


Figure 23: Number of valid Mie cloudy HLOS wind observations, bias and SDD (compared to ECMWF model) as functions of HLOS wind for the whole second reprocessed dataset.

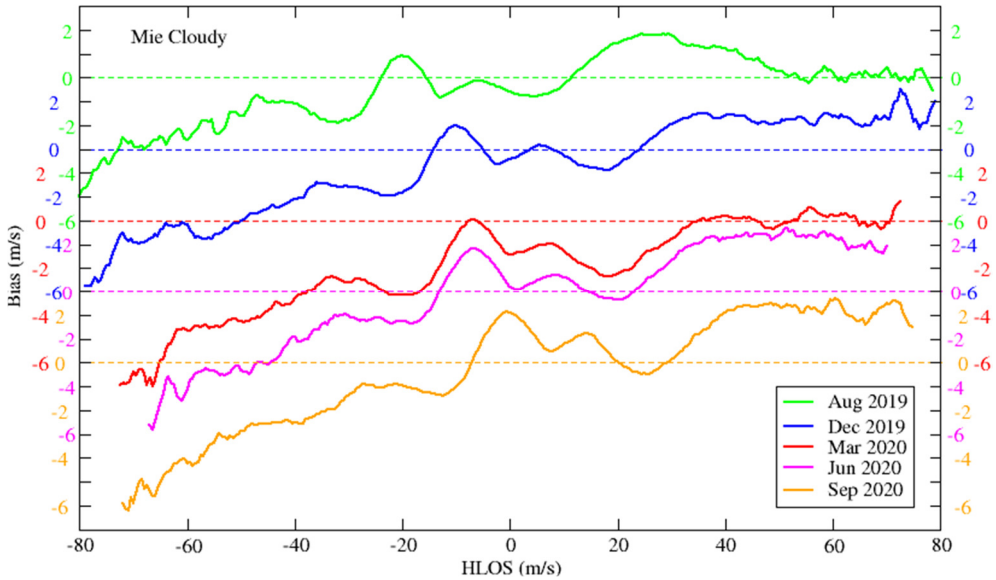


Figure 24: Change of Mie cloudy HLOS wind bias compared to ECMWF model as function of HLOS wind for 5 different weeks (4-10 August 2019, 18-24 December 2019, 23-29 March 2020, 22-28 June 2020 and 14-20 September 2020) of the second reprocessed dataset. The x-axis (HLOS wind) for each week is shifted with labels colour coded as the curve colour of the corresponding week.

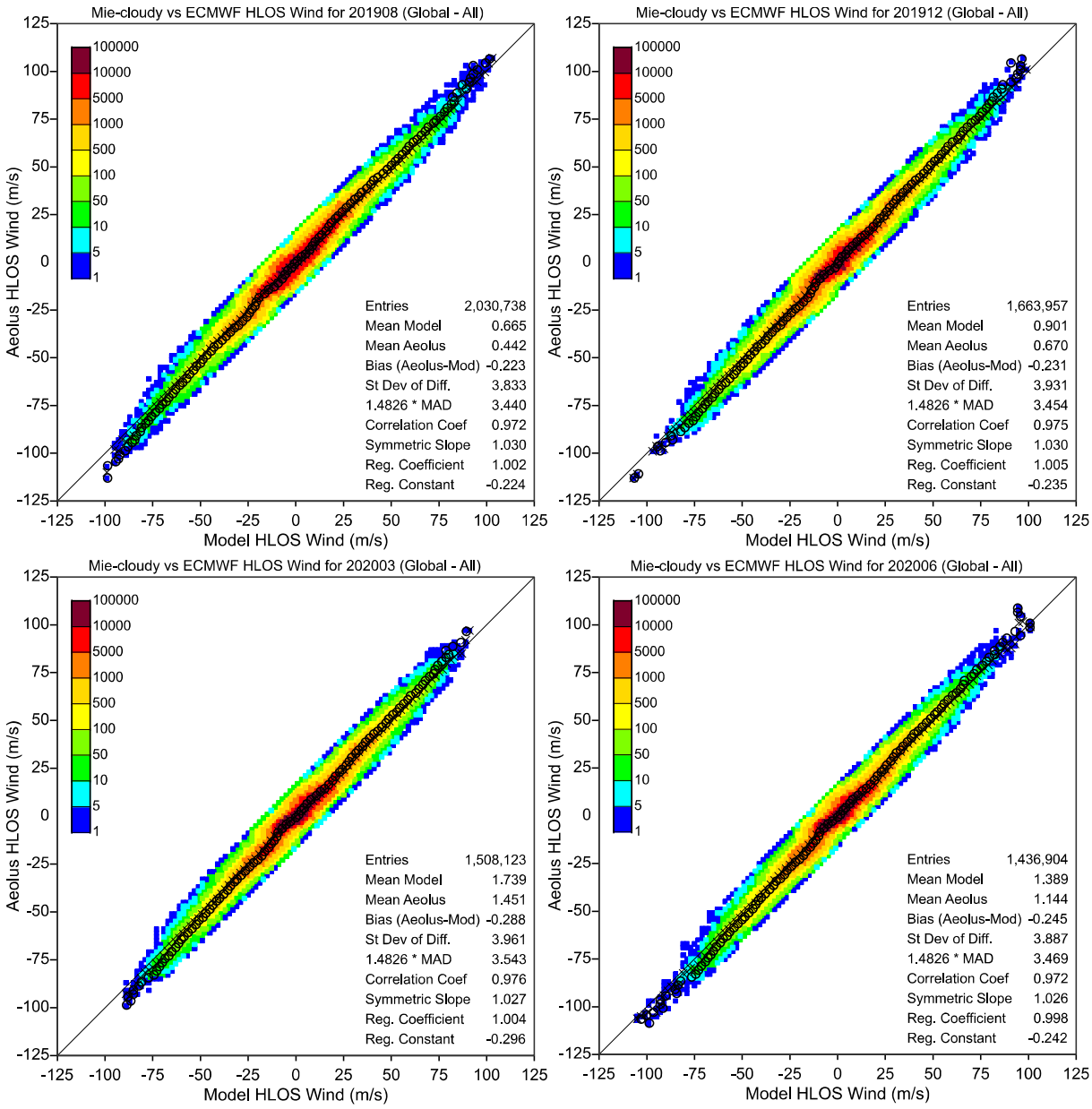


Figure 25: As in Figure 22 but each scatter plot represents the monthly data of 4 selected months: August 2019 (upper left), December 2019 (upper right), March 2020 (lower left) and June 2020 (lower right). Refer to Figure 22 for the meaning of crosses, circles and colour-coding.

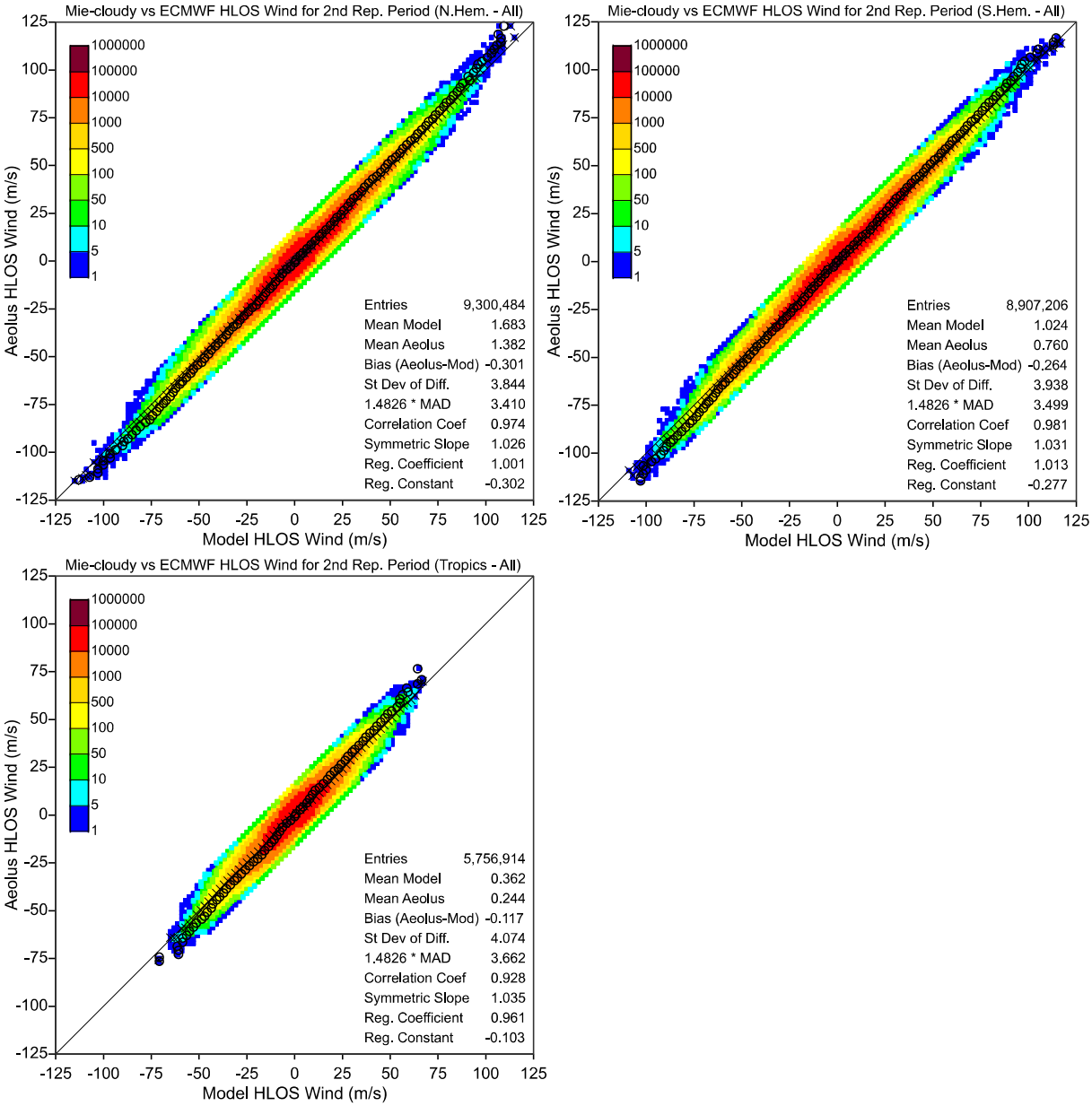


Figure 26: As in Figure 22 but with discrimination between extra tropics (upper) Northern Hemisphere (upper left) and Southern Hemisphere (upper right) and Tropics (lower). Refer to Figure 22 for the meaning of crosses, circles and colour-coding.

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 second 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 amount 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 second reprocessing dataset. Small negative biases dominate the global map in Figure 28. However, enhanced positive biases can be seen on Antarctica and several mountain ranges. On the other hand, small SDD values dominate Figure 29 with enhancements some of the mountain ranges, Intertropical Convergence Zone (ITCZ) and the monsoon area.

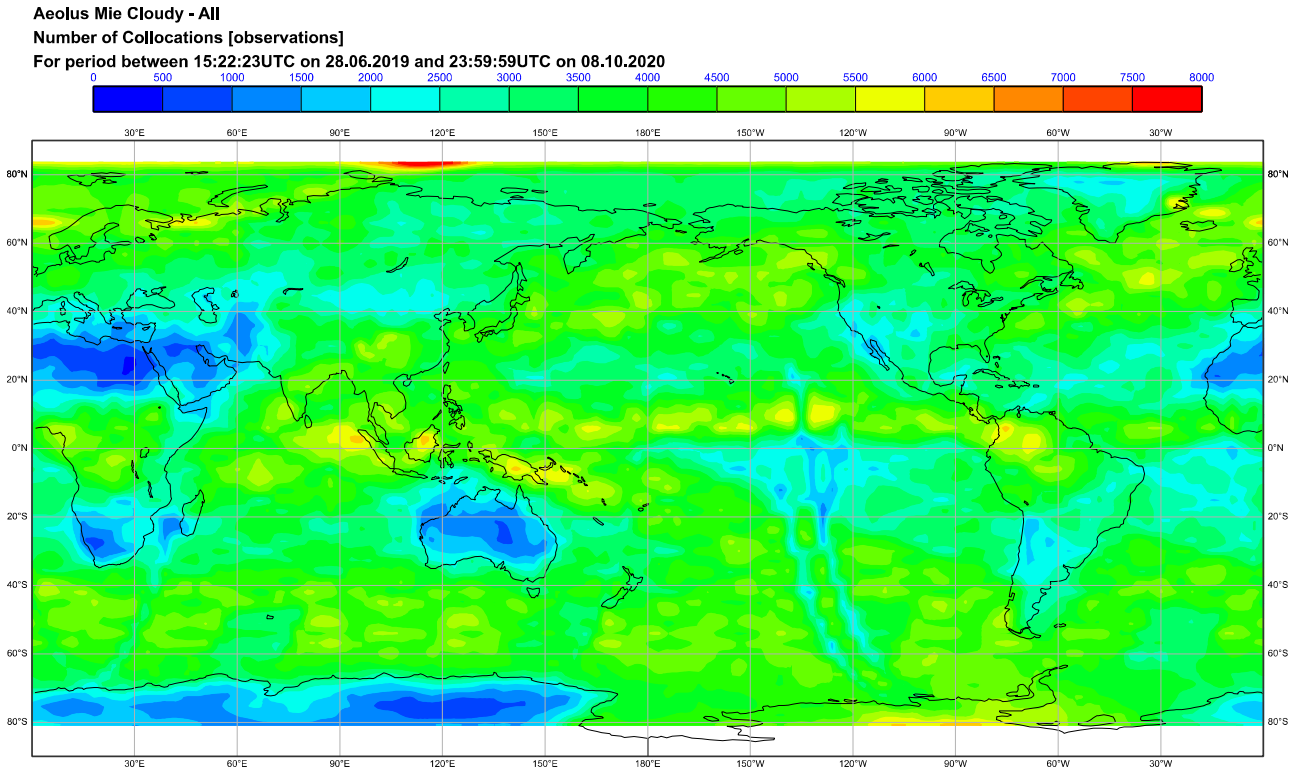


Figure 27: Geographical distribution of number of Mie cloudy L2B observations per 3°X3° latitude-longitude box during the whole second reprocessing.

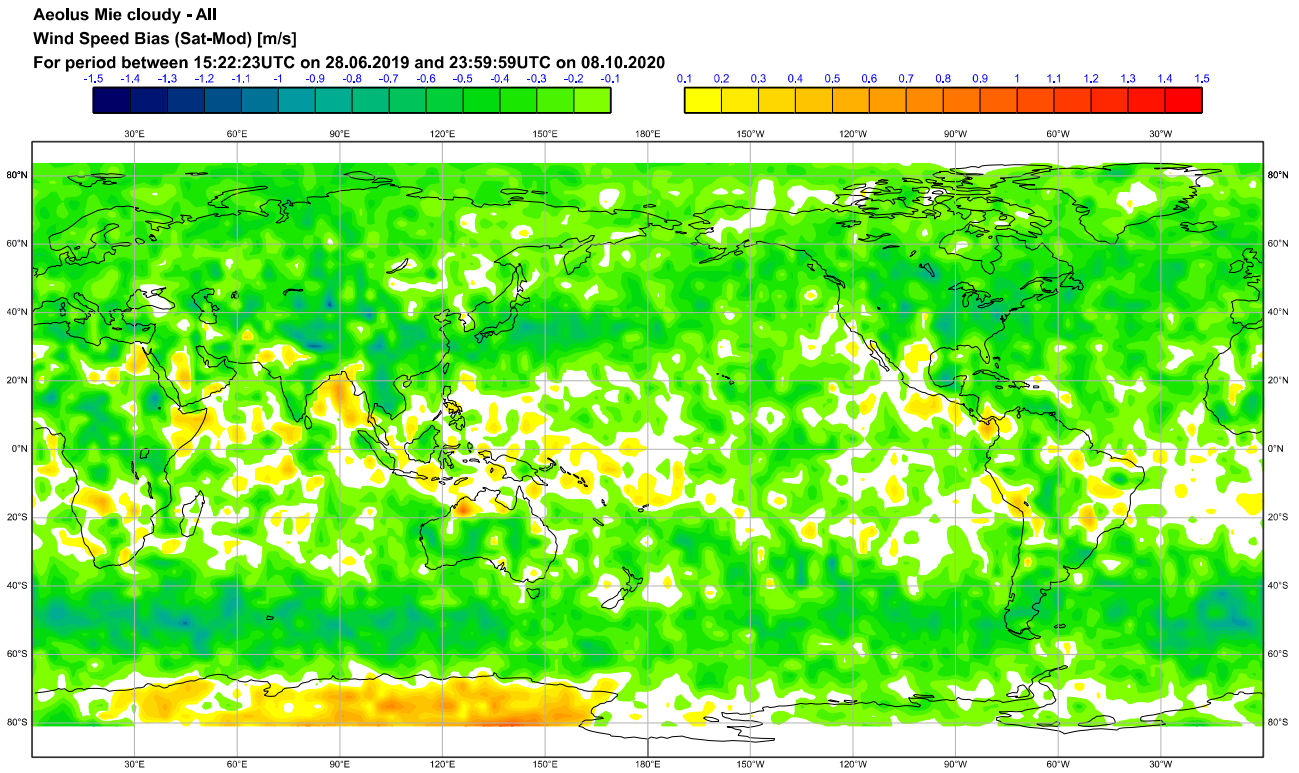


Figure 28: Geographical distribution of HLOS wind speed bias between Mie cloudy and ECMWF model for the whole second reprocessing dataset.

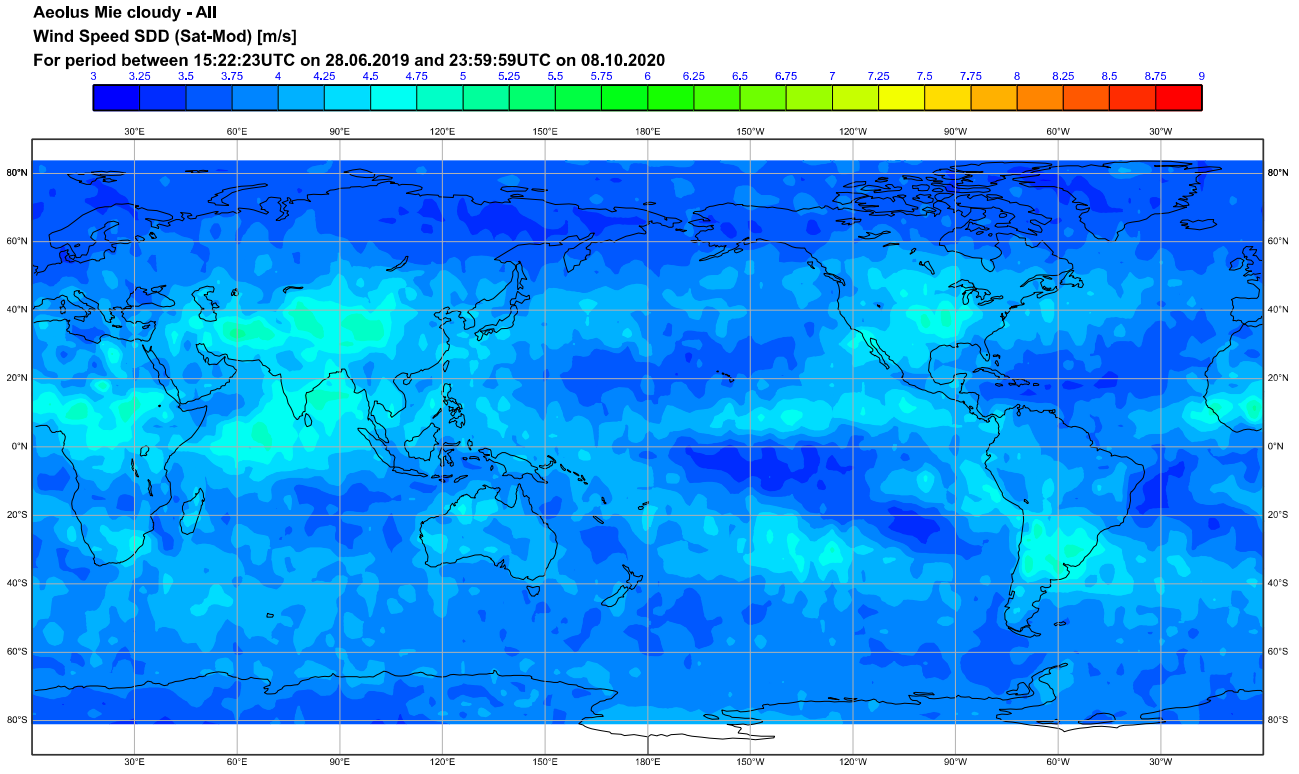


Figure 29: As in Figure 28 but for the Mie cloudy SDD.

The whole second 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 21 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 3.8, 3.4 and 0.971, respectively, for the ascending compared to 4.0, 3.6 and 0.966, respectively, for the descending orbits.

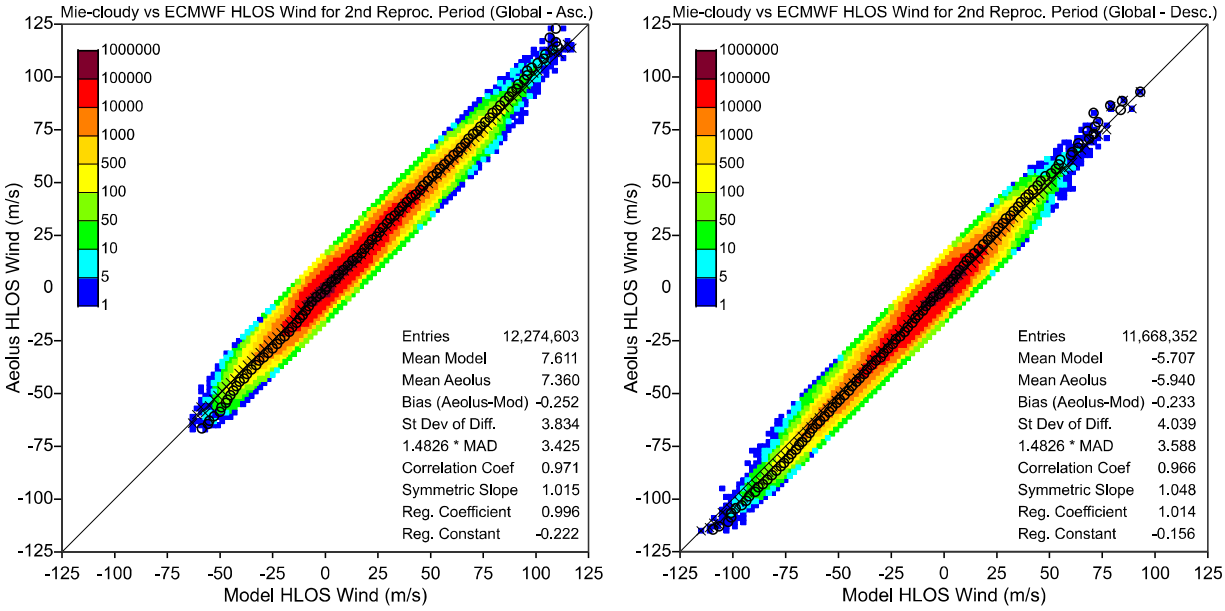


Figure 30: As in Figure 22 but with discrimination between ascending (left) and descending (right) orbits. Refer to Figure 22 for the meaning of crosses, circles and colour-coding.

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. The overall bias is rather small (about -0.2 m/s) and almost independent of the altitude between altitudes 3 and 23 km. However, for the lowest part of the profile the bias is as high as -0.8 m/s and for the upper part the bias as high as -1.6 m/s. The number of observations at both ends of the profile is rather small.

Between altitude 7 and 15 km, the biases from ascending and descending orbits are anti-correlated. A possible interpretation is that the vertical motion combined with the fact that convection may be very different between the morning and afternoon overpasses. This also provide a logical explanation why it is only seen for the Mie channel, which observes cloud tops, and not for the Rayleigh channel. For altitudes above 15 km, where not much data exists, and below 7 km the correlation is positive.

The overall SMAD and SDD increase linearly with altitude from 3.2 and 3.5 m/s, respectively, towards the bottom of the profile to peak at around 4.0 and 4.3 m/s, respectively, at 16-km altitude. At higher altitudes, both SMAD and SDD decrease linearly until reaching local minima of 3.5 and 3.8 m/s, respectively, at 21-km altitude. Beyond that, SMAD and SSD increase linearly to exceed 5 m/s at the top of the profile. Data from ascending orbits are in better agreement (lower SMAD and SDD values) with the model compared to data from descending orbits.

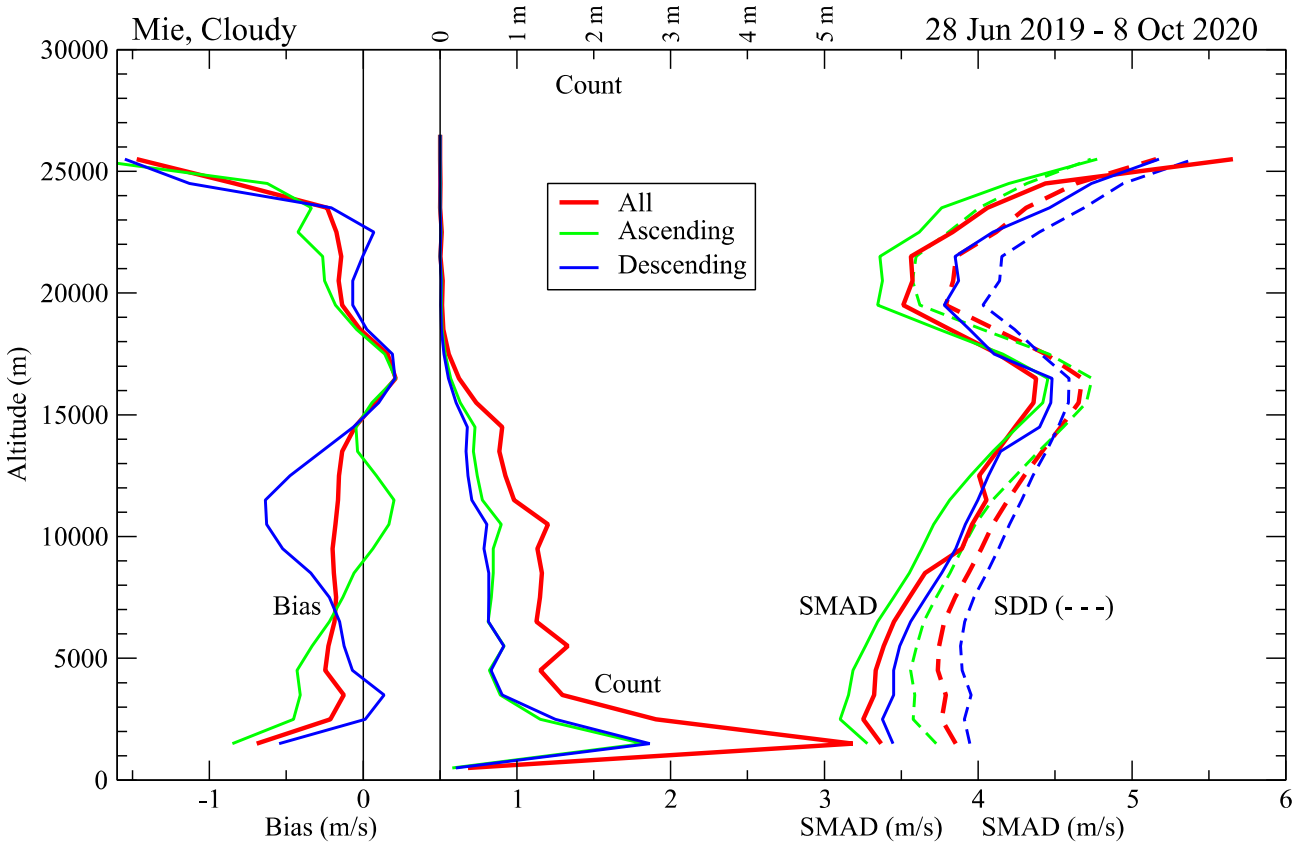


Figure 31: Profiles of L2B Mie cloudy HLOS wind deviations from ECMWF model (bias, SDD and SMAD) together with the number of collocations at each altitude. Profiles show discrimination between ascending and descending orbits in addition to all data.

7.3 L2B Rayleigh clear winds

The probability density function (PDF) of Rayleigh clear HLOS winds from the whole second 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.

Figure 33 shows the time series of the number of valid L2B Rayleigh clear HLOS wind speed observations, and the bias and SMAD (scaled median absolute difference, MAD) between both first and second reprocessed L2B and ECMWF IFS model. Only running means over 30 successive files (orbits) are shown.

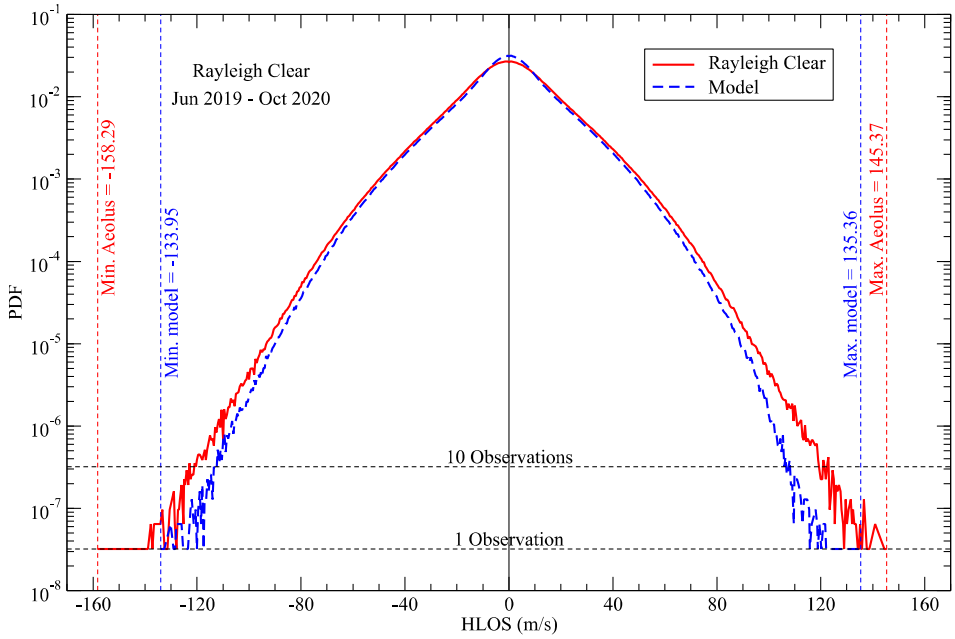


Figure 32: Probability density function (PDF) of Rayleigh clear HLOS winds from the whole second reprocessed dataset. The corresponding PDF of ECMWF model collocated with Rayleigh clear observations is also shown. The maximum and minimum (maximum of absolute negative values) for both Rayleigh clear and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

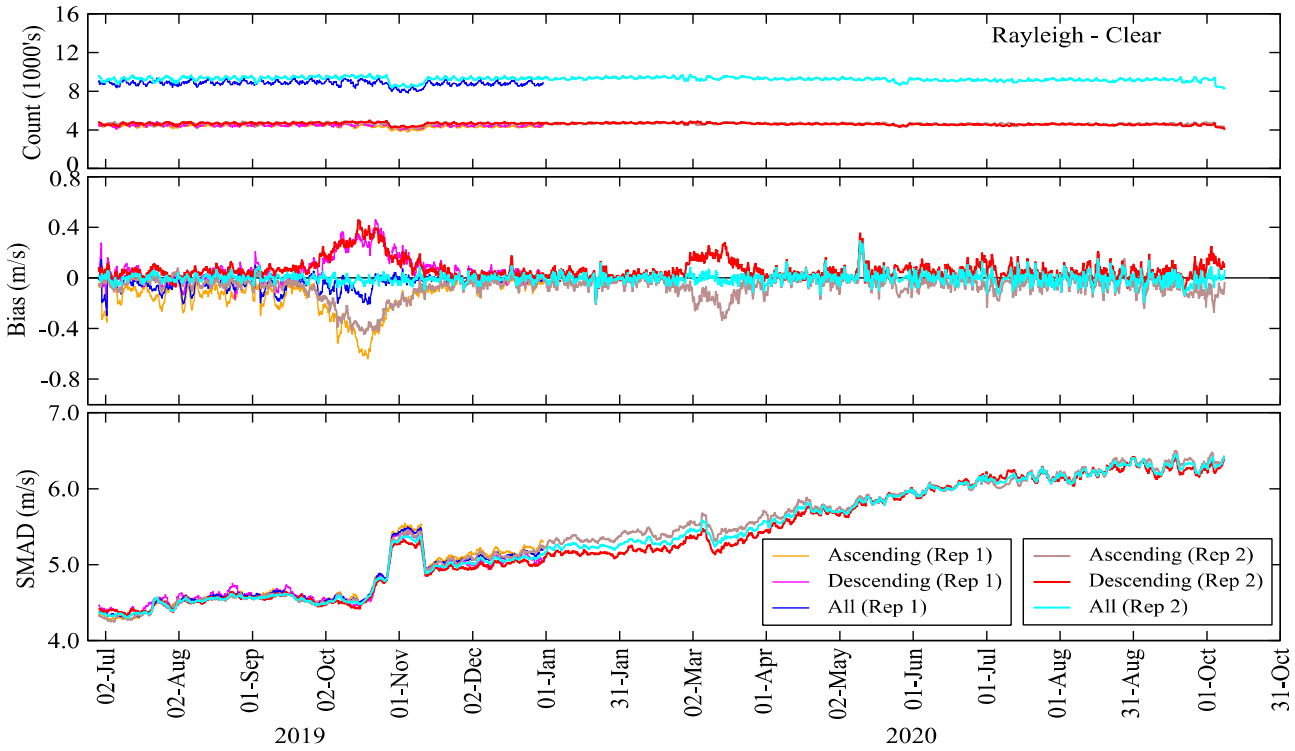


Figure 33: Time series of number of valid L2B Rayleigh clear wind observations, HLOS wind speed bias and SMAD between L2B Rayleigh clear and ECMWF IFS model winds. Running means along 30 successive files (orbits) are shown. Curves for all data as well as for ascending and descending orbits are shown for the both first (Rep 1) and second (Rep 2) reprocessings. Note that the first reprocessing curves end at the end of 2019.

Apart from a very small dip in the number of valid L2B Rayleigh clear wind observations during the two-week period (28 October – 10 November 2019) of special range-bin setting (intended for AMV comparison), another week towards the end of May 2020 and the last few days of the period, the data volume was stable over the whole period covered by the second reprocessing dataset as can be seen in the upper panel of Figure 33. The number of valid observations from ascending and descending orbits are almost equal as expected.

The number of valid Rayleigh clear wind observations in the second reprocessing is slightly higher than that of the first reprocessing as can be seen in the upper panel of Figure 33.

On average, the L2B Rayleigh clear wind bias compared to ECMWF model fluctuates around 0.0 m/s. With few exceptions, the bias varies between -0.9 m/s and +0.9 m/s for individual orbits (not shown) and between -0.2 m/s and +0.2 m/s for the 30-orbit running means as shown in the middle panel of Figure 33. One can recognise three exceptions with rather high biases: from late September to mid-November 2019, from end of February to end of March 2020 and from the end of September 2020 till the end of the second reprocessing period (9 October 2020). The last period looks as if is a repetition of the first period but in 2020.

The bias with respect to the model from ascending and descending orbits are almost equal except for the three exceptions above when the biases from ascending and descending orbits oppose each other. The highest difference is during the first period (late September to mid-November 2019) when bias of ascending orbits is lower than the model by as high as 0.4 m/s while the bias from the descending orbits is higher than the model by as high as 0.4 m/s (the values are based on the running means). This makes the maximum difference between the descending and ascending winds to be about 0.8 m/s during that period.

In general, the middle panel of Figure 33 suggests that the Rayleigh clear HLOS wind bias in the second reprocessing is slightly smaller than that of the first reprocessing especially for the ascending orbits.

The lower panel of Figure 33 shows that SMAD varies around 4.4 m/s initially before it started to jump towards mid July 2019. Possible reason for this is the laser frequency adjustment (LFA) and Rayleigh cover temperature (RCT). A calibration file for the Rayleigh winds (RBC file) derived from an ISR (Instrument Spectral Registration) was applied on 4 July 2019. By modifying the LFA and RCT, there seems to be a need for a new AUX_RBC file (Rayleigh-Brillouin Correction). SMAD values are then became stable till second half of October 2019 when a linear increase started due to the drop off in atmospheric path signal levels, which the Rayleigh channel winds are very sensitive to. This trend was disrupted by the SMAD jumps during the special range-bin setting (for AMV comparison) period when SMAD values approached 5.5 m/s. The linear increasing trend continued afterwards till the end of the reprocessing period. However, there was a noticeable drop in SMAD at the beginning of March 2020 and another drop, but less pronounced, at mid-April 2020. Furthermore, the SMAD appears as if is stabilised from the end of August 2020 till the end of the period. There is a possibility that this stabilisation is due to a seasonal cycle as there is a slight drop of SMAD in September 2019.

Furthermore, the lower panel of Figure 33 shows that SMAD for the descending orbits is consistently lower that that of the ascending orbits. The difference, however, is not large except for the period from December 2019 and April 2020 (and the period of the special range-bin setting for AMV comparison).

The comparison between L2B Rayleigh clear HLOS wind against its model counterpart is shown in Figure 34 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 second reprocessing. The agreement between more than 62 million collocations of L2B and model wind is rather good without any bias globally and with SDD and SMAD values of 6.1 m/s and 5.2 m/s, respectively. The correlation coefficient is 0.948. 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 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.

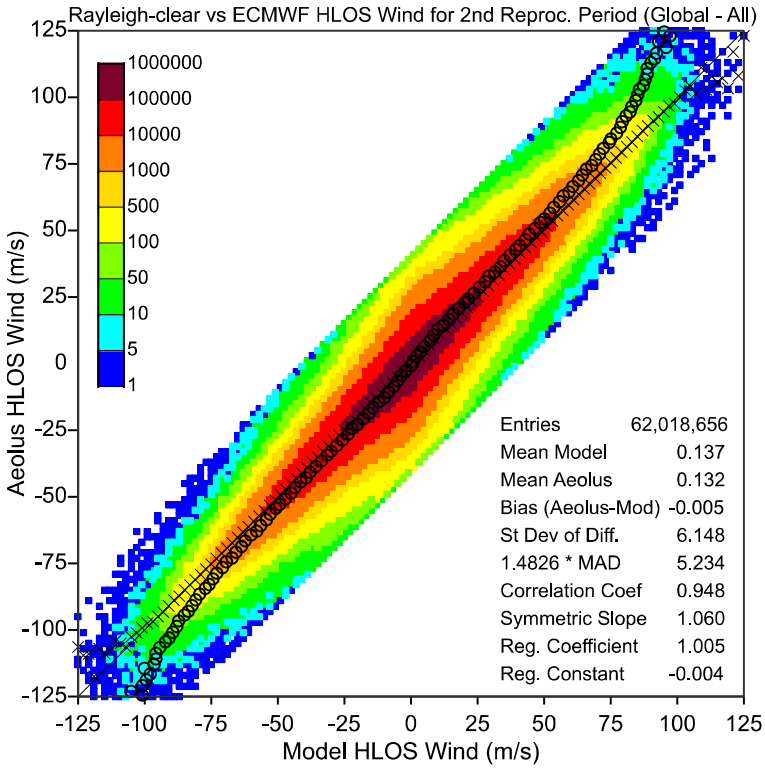


Figure 34: Density scatter plot comparing L2B Rayleigh clear to ECMWF model HLOS winds over the whole globe during the whole period of the second reprocessing. Boxes are colour coded based on the number of collocations within the box. The circles are the average model value given L2B value while the crosses are the other way around. Statistics are given in the lower right corner of the plot.

Figure 35 shows the variation of number of valid observations, bias and SDD (compared to ECMWF model) as functions of HLOS wind for the whole second reprocessed dataset. Both bias and SDD curves show smooth and symmetric behaviour. The bias is rather small (about 1.5 m/s) for most of the HLOS winds (-60 m/s to +60 m/s). Positive HLOS winds show overestimation with respect to the model and vice versa for the negative HLOS values. The bias becomes rather high (about 10 m/s) at extreme values (above 110 m/s). Note that the number of those values is rather small. The SDD increases slowly with the increase of HLOS. For most of the HLOS values, the SDD is between 5 and 6 m/s.

The evolution of the agreement between Rayleigh clear and ECMWF model is represented by four selected monthly scatter plots in Figure 36. The increase in SDD and SMAD values as time progressing can be easily seen. The degradation is more pronounced at the extreme values.

The second reprocessed Rayleigh clear dataset was discriminated based on the geographical location being in extratropical Northern Hemisphere (NH), Southern Hemisphere (SH) or in the Tropics. The scatter plots showing the comparison against the ECMWF model in each geographical region are shown in Figure 37. 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).

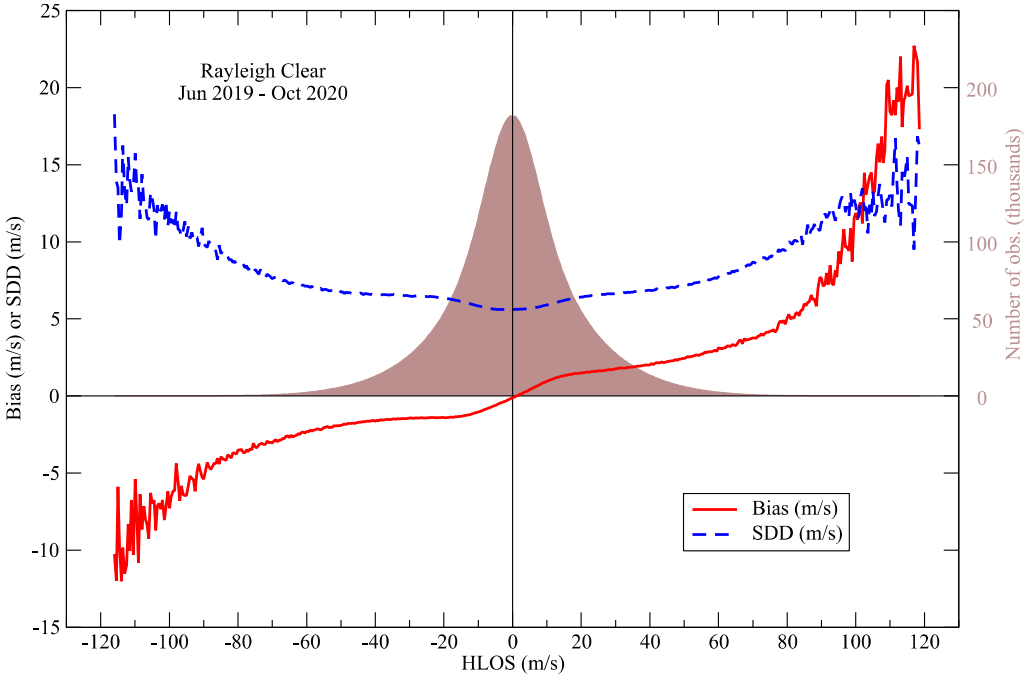


Figure 35: Number of valid Rayleigh clear HLOS wind observations, bias and SDD (compared to ECMWF model) as functions of HLOS wind for the whole second reprocessed dataset.

The distribution of the cloud of boxes in the scatter plots in Figure 37 especially for the Tropics shows skewness at the middle of the cloud. Monthly scatter plots for each region were plotted and selected plots are shown in Figure 38 for the Tropics and in Figure 39 for the extra tropics. The existence of a secondary vertical cloud of boxes centred around the model HLOS value of zero is clearly seen in the Tropics especially during the months of December 2019 and March 2020. Same observation can be made in the SH extra tropics during December 2019 and in the NH extra tropics during August 2020, which are the summer period in each hemisphere. Therefore, this secondary distribution appears mainly during the seasons with low winds. It is high likely that this is a model issue.

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 second 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 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 second reprocessing dataset. The bias, as can be seen from Figure 41, is small everywhere with clear distinction between Tropics, Greenland and Antarctica all with positive bias on one side and the extra tropics with negative bias. On the other hand, as shown in Figure 42, the SDD is rather high with enhanced value over Antarctica, the mountain ranges and the monsoon region. Smaller SDD values can be seen in the southeast Pacific.

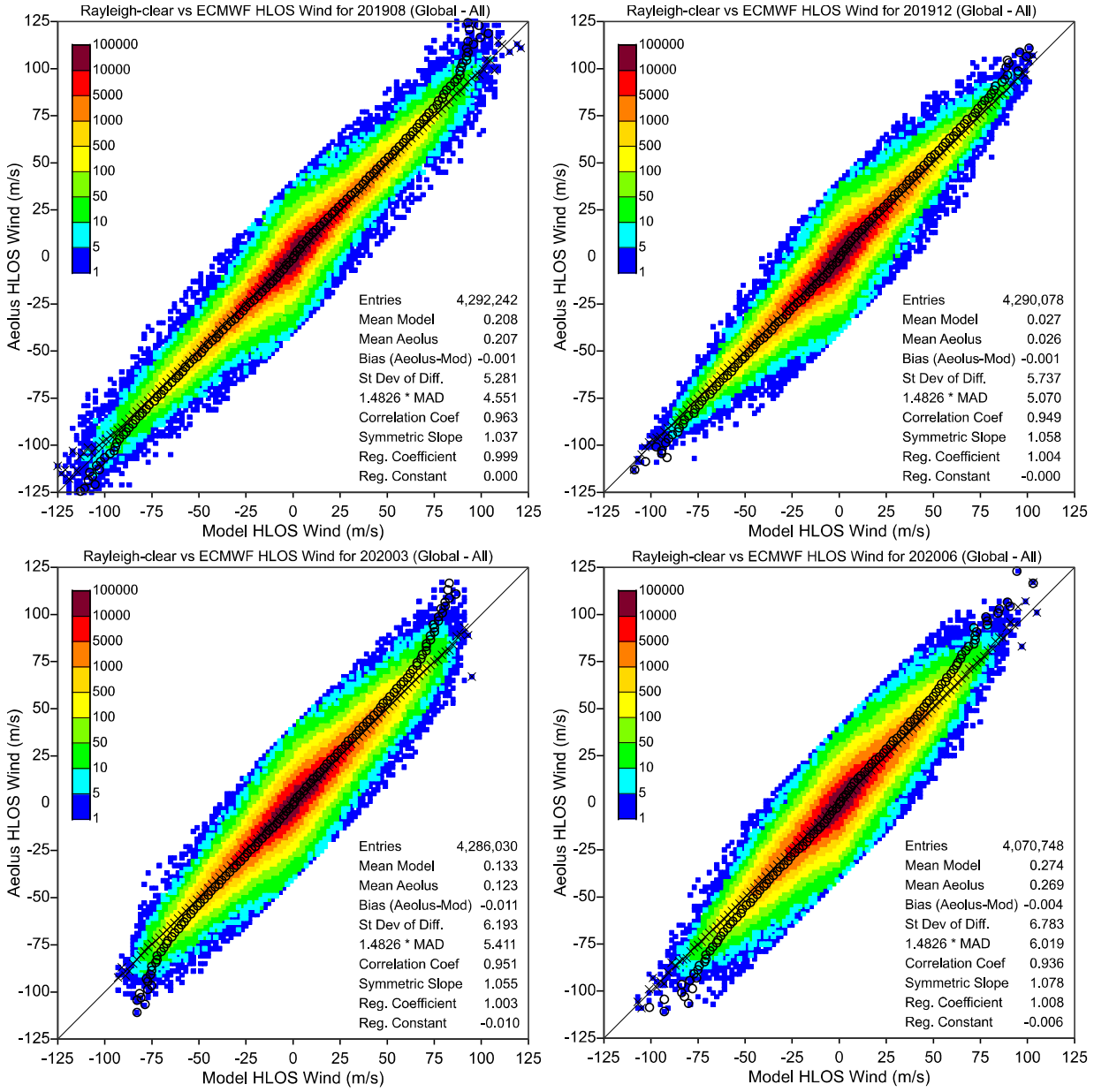


Figure 36: As in Figure 34 but each scatter plot represents the monthly data of 4 selected months: August 2019 (upper left), December 2019 (upper right), March 2020 (lower left) and June 2020 (lower right). Refer to Figure 34 for the meaning of crosses, circles and colour-coding.

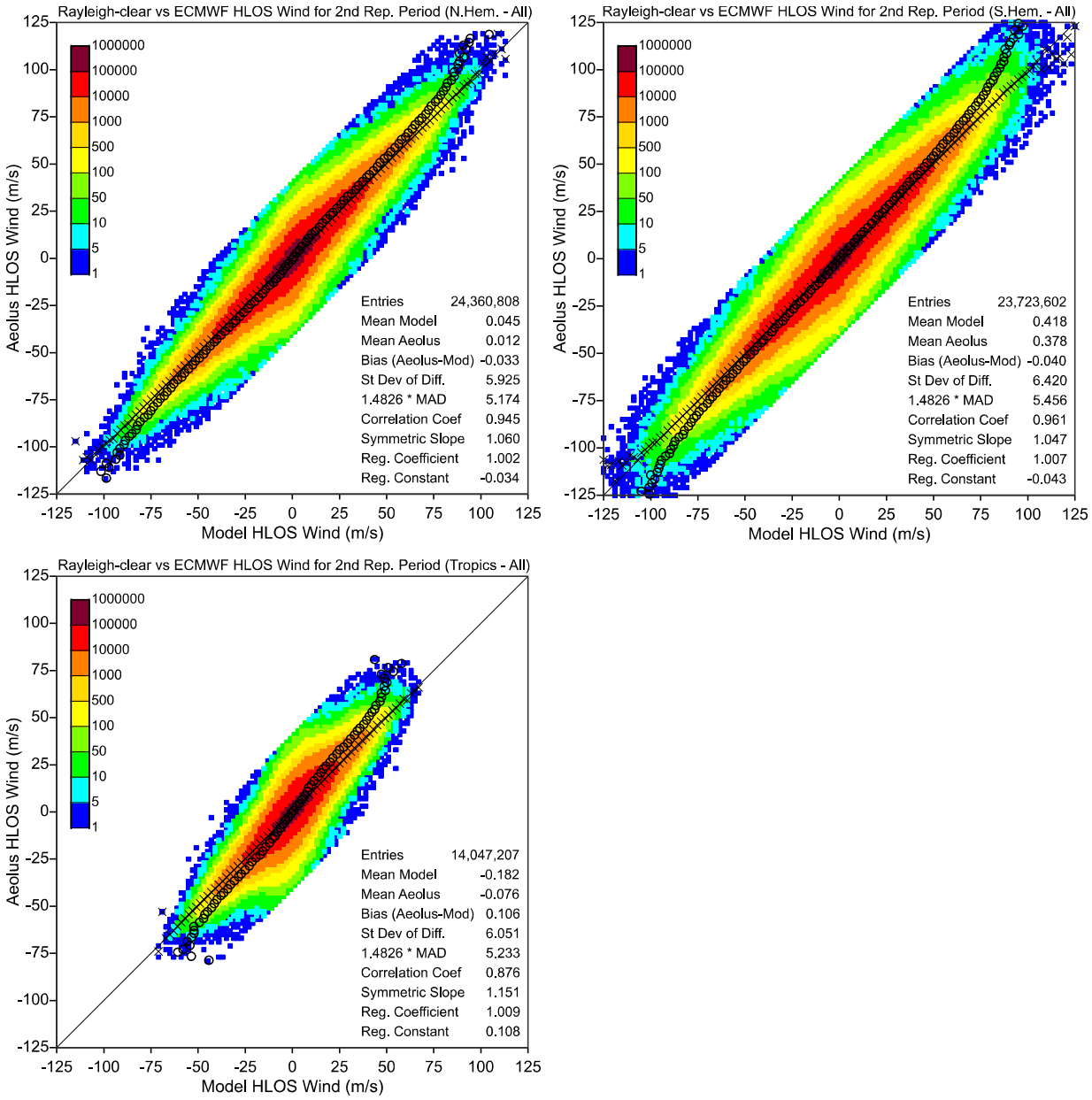


Figure 37: As in Figure 34 but with discrimination between extra tropics (upper) Northern Hemisphere (upper left) and Southern Hemisphere (upper right) and Tropics (lower). Refer to Figure 34 for the meaning of crosses, circles and colour-coding.

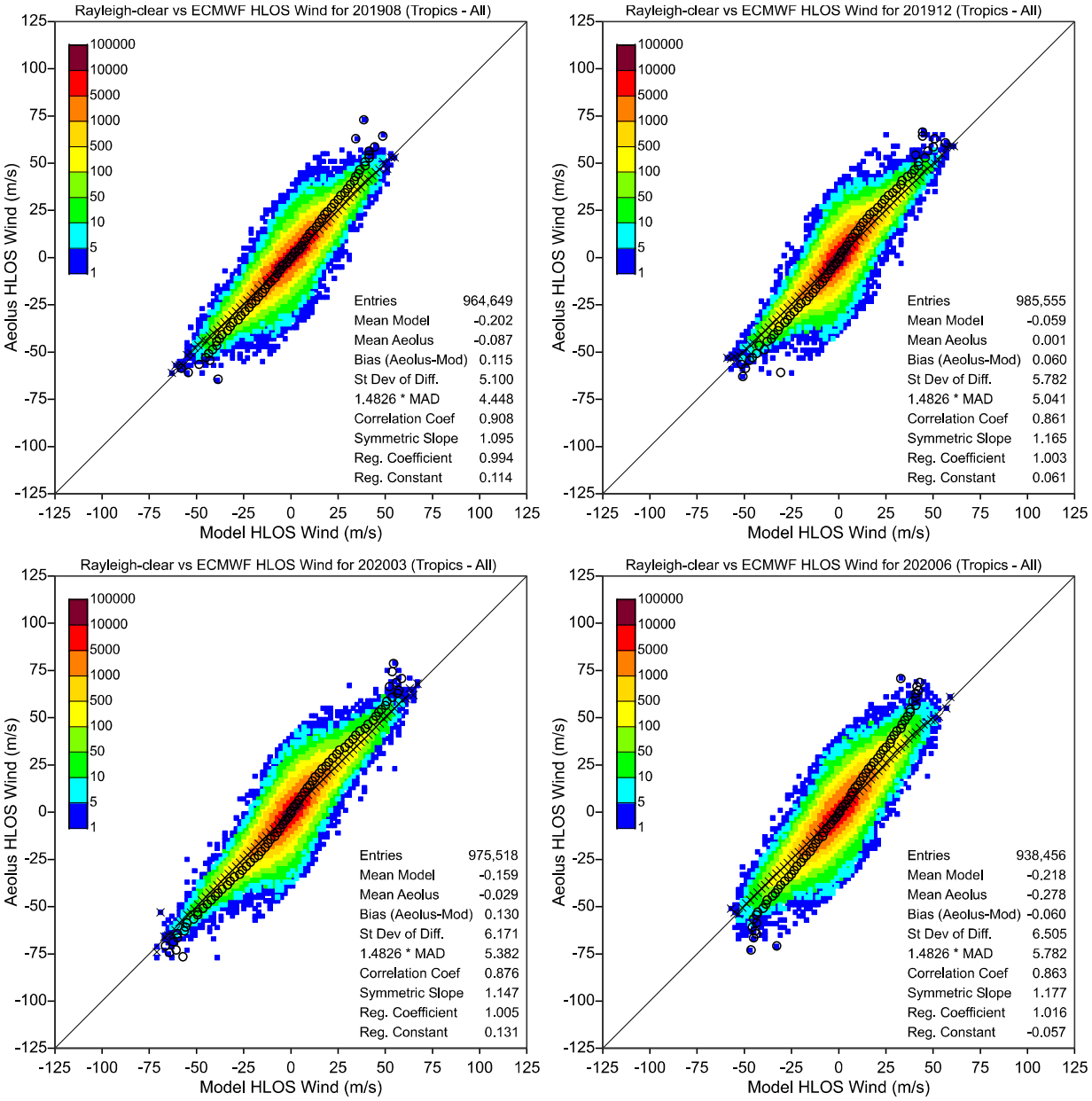


Figure 38: Monthly scatter plots as in Figure 36 but only for the Tropics (lower). Refer to Figure 34 for the meaning of crosses, circles and colour-coding.

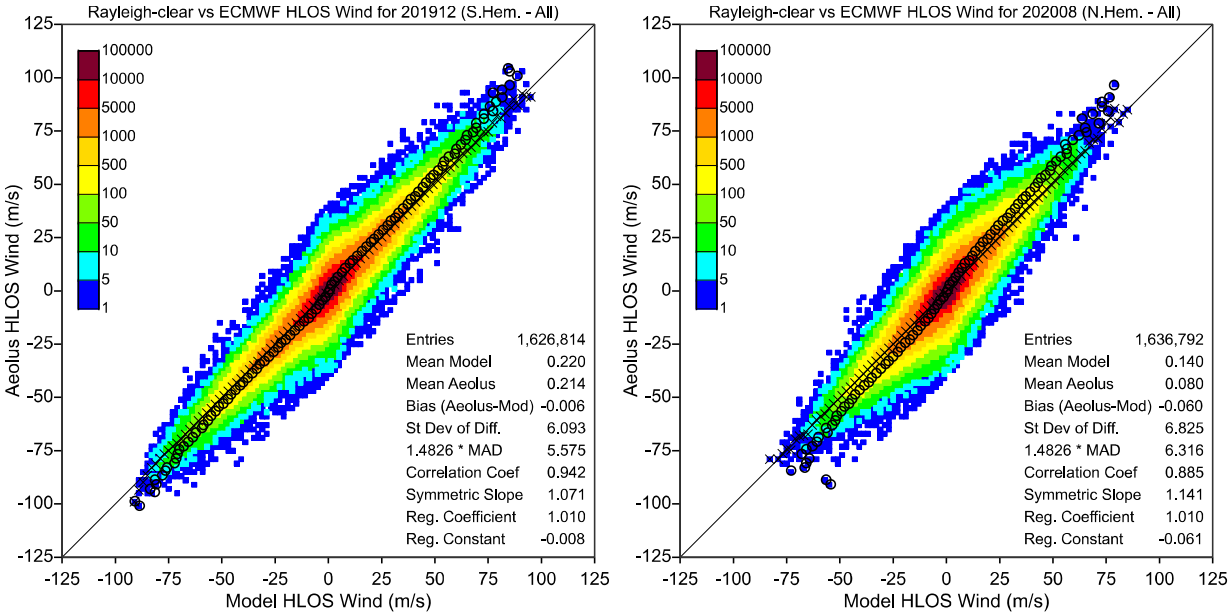


Figure 39: Selected monthly scatter plots of Rayleigh-clear observations for extra tropics: December 2019 for Southern Hemisphere (left) and August 2020 for the Northern Hemisphere (right). Refer to Figure 34 for the meaning of crosses, circles and colour-coding.

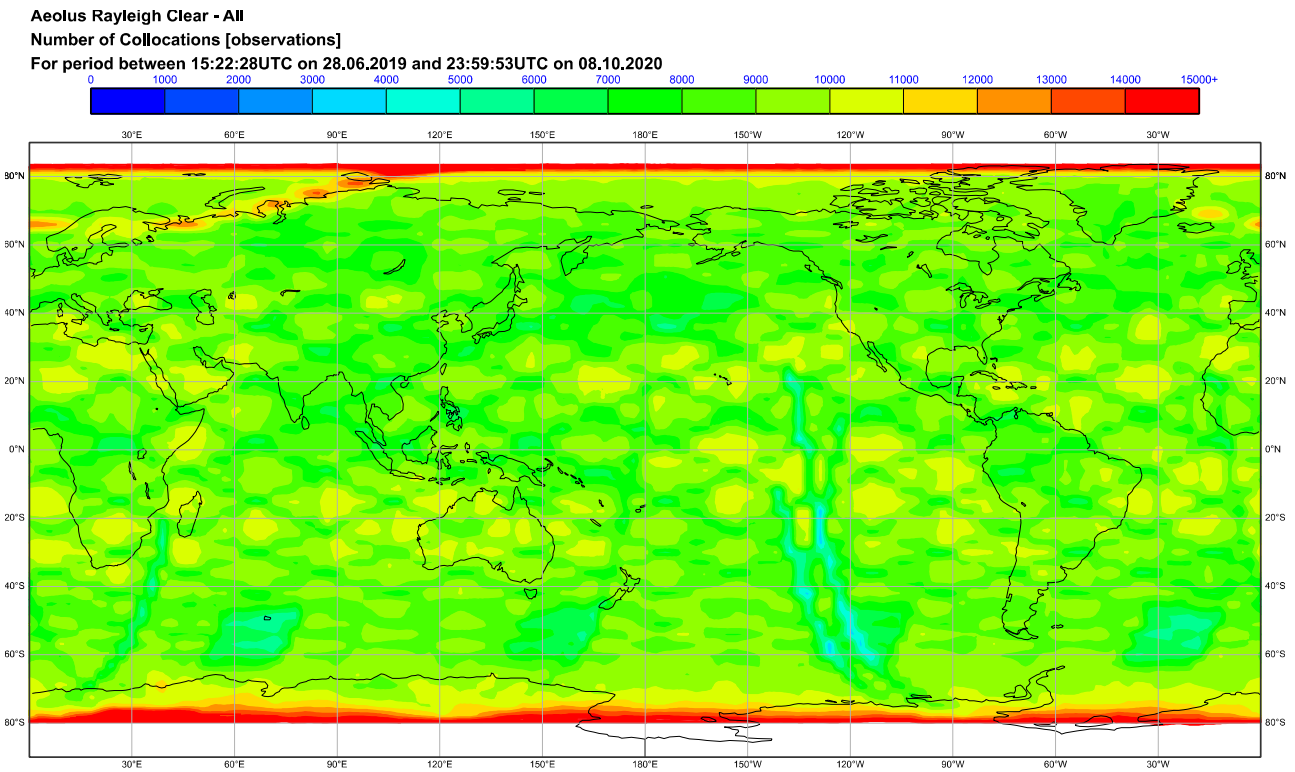


Figure 40: Geographical distribution of number of Rayleigh clear L2B observations per 3°X3° latitude-longitude box during the whole second reprocessing.

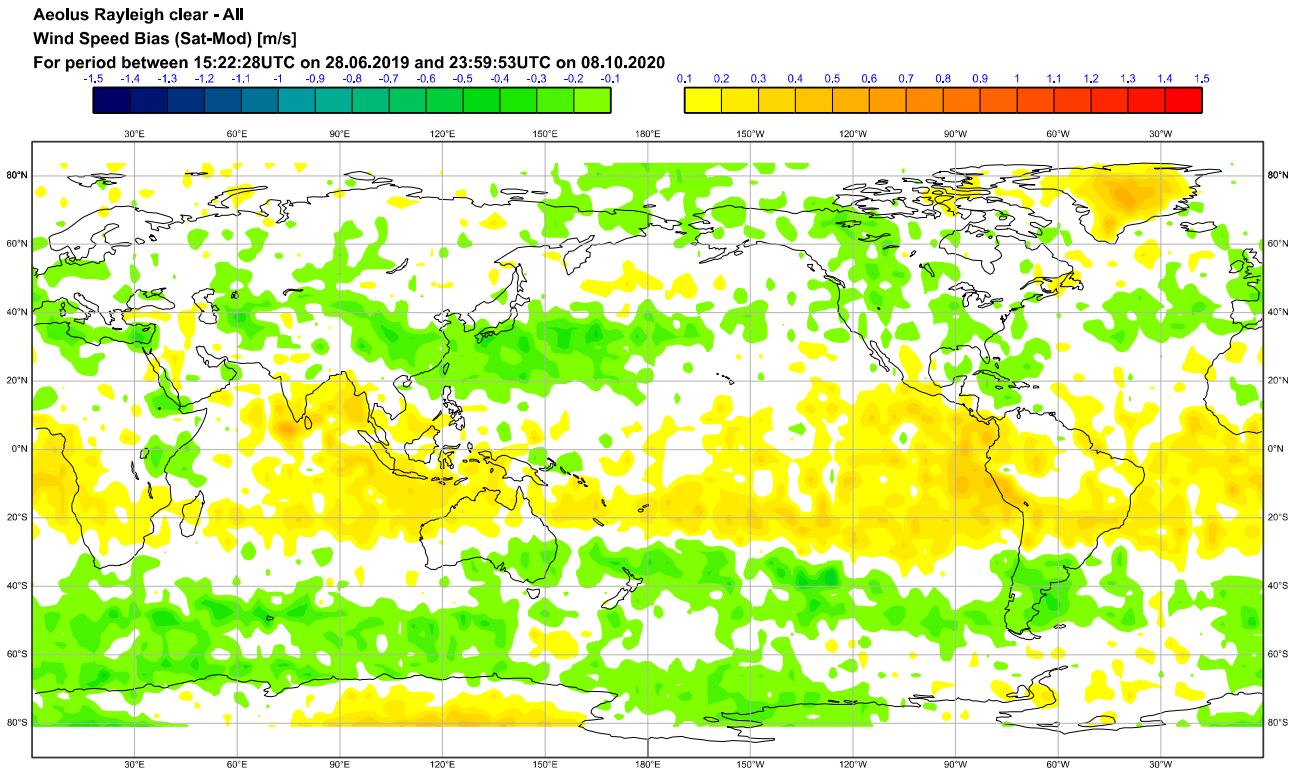


Figure 41: Geographical distribution of HLOS wind speed bias between Rayleigh clear and ECMWF model for the whole second reprocessing dataset.

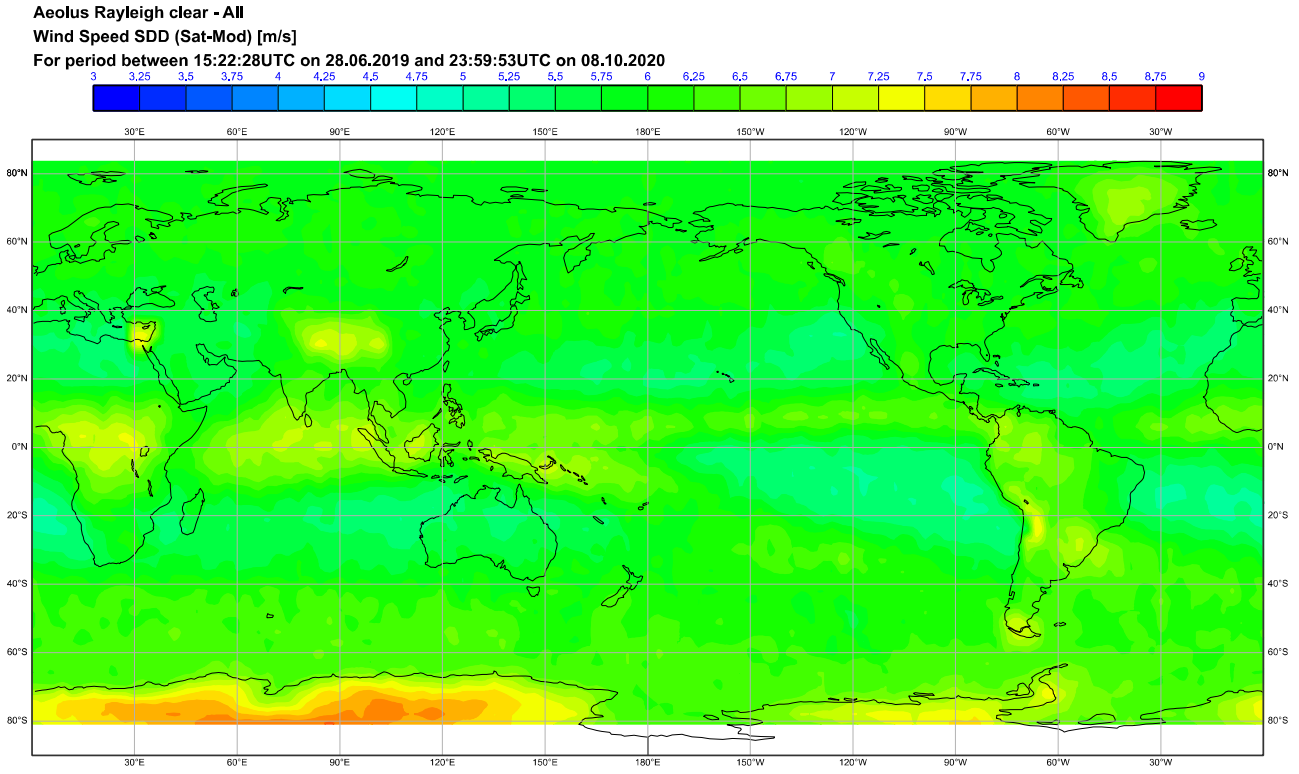


Figure 42: As in Figure 28 but for the Rayleigh clear SDD.

The whole second 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 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 also not much difference between the statistics from both orbit directions with descending orbits show marginally better fit to the model.

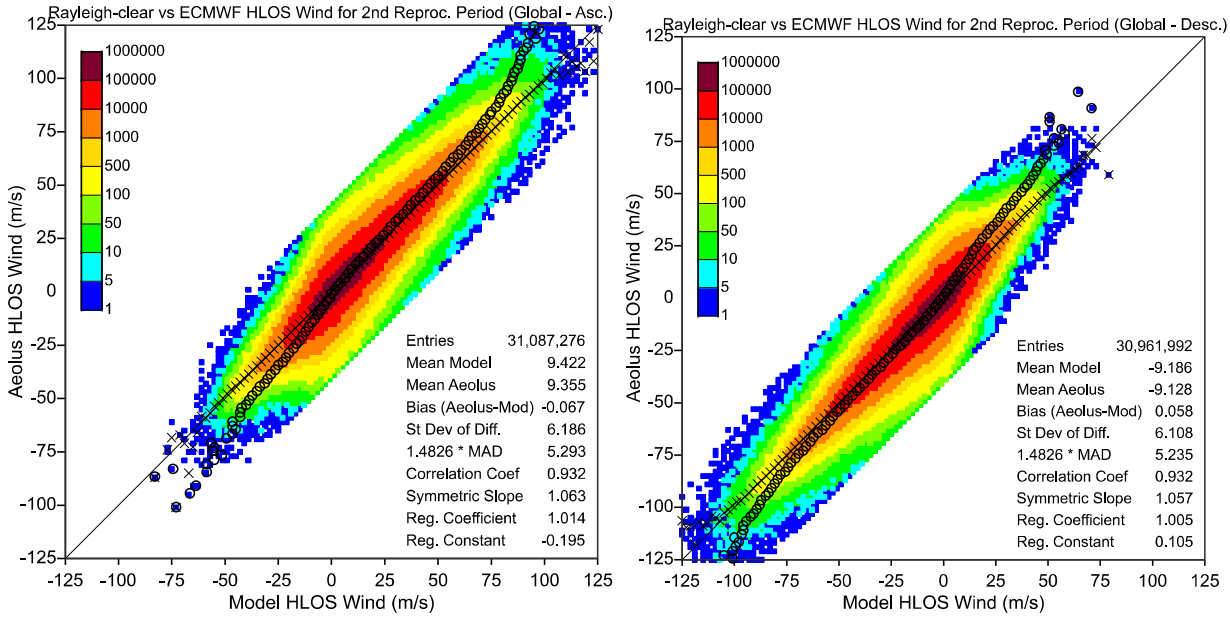


Figure 43: As in Figure 34 but with the discrimination between ascending (left panel) and descending (right panel) orbits. Refer to Figure 34 for the meaning of crosses, circles and colour-coding.

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 overall bias is rather small and positive (about +0.4 m/s) and almost independent of the altitude between altitudes 10 and 22 km. However, for the lowest and the highest parts of the profile the bias becomes negative and reaches about -1.0 m/s. The number of observations at upper part of the profile is rather small but the same cannot be said about the lower part.

The overall SMAD and SDD follow a ε-shaped (epsilon-shaped) curve. SMAD and SDD values of more than 7-8 m/s occur at high and low altitudes. Both SMAD and SDD show linear decrease with altitude when they get their minimum values of 4.5 m/s and 5.5 m/s, respectively, at an altitude of about 5 km. They start to increase to reach local maxima of 5.0 m/s and 6.0 m/s, respectively, at altitude of 9 km. Another brief decrease with altitude occurs before they change trend to increase with altitude at 12 km. The increase with altitude continues until the top of the profiles with SMAD and SDD values reach 14 m/s at the highest altitude.

The ε-shaped curve of SMAD and SDD with altitude can be attributed to the typical pattern of attenuated backscatter with altitude in clear air. Observations from descending orbits show marginal superiority when it comes to agreement with the model between 5- and 15-km altitudes.

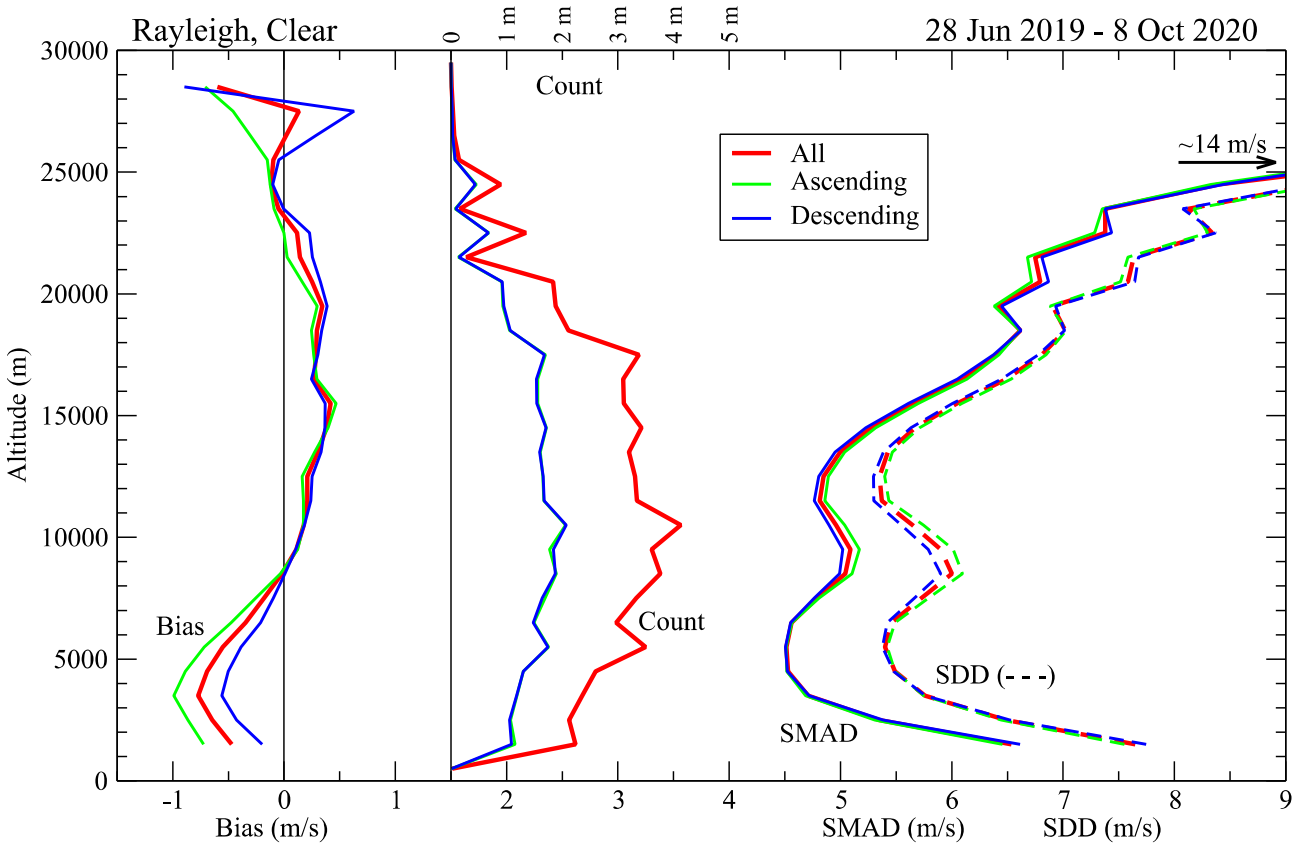


Figure 44: Profiles of L2B Rayleigh clear HLOS wind deviations from ECMWF model (bias, SDD and SMAD) together with the number of collocations at each altitude. Profiles show discrimination between ascending and descending orbits in addition to all data. SDD and SMAD values at the top of the profile,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

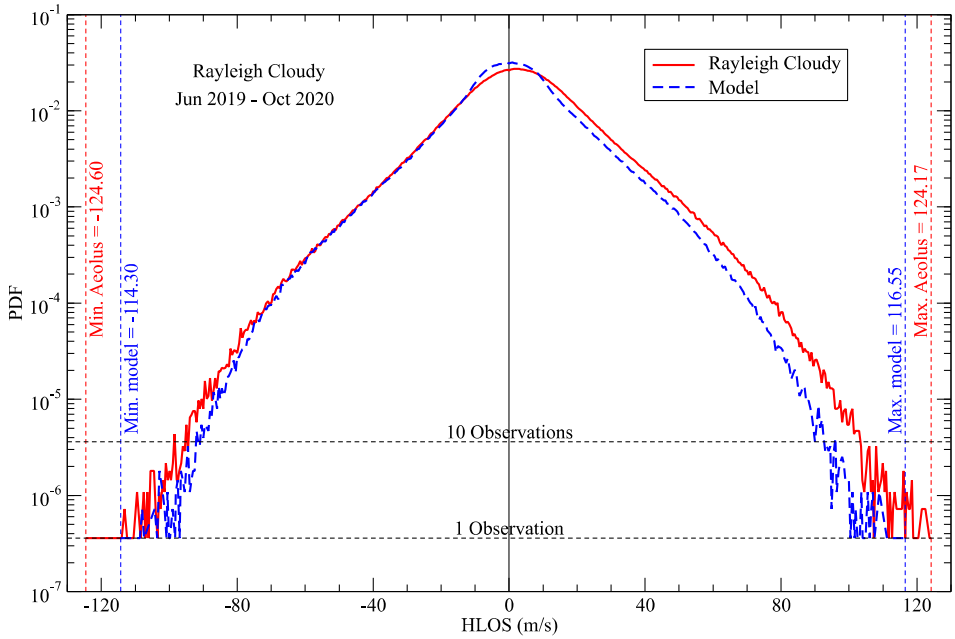


Figure 45: Probability density function (PDF) of Rayleigh cloudy HLOS winds from the whole second reprocessed dataset. The corresponding PDF of ECMWF model collocated with Rayleigh cloudy observations is also shown. The maximum and minimum (maximum of absolute negative values) for both Rayleigh cloudy 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 both first and second reprocessed L2B 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, the number of Rayleigh cloudy observations shows mild gradual decline with time. The number of valid observations from ascending and descending orbits are almost equal as expected. On the other hand, the number of valid Rayleigh cloudy wind observations in the second reprocessing is about 50% higher than that of first reprocessing.

As can be seen in the middle panel of Figure 46, the (running) mean bias fluctuates around +0.5 m/s from start until the third week of July 2019 when it jumped to 1.5 m/s. This may be attributed to the LFA and RCT and hence a slightly inappropriate RBC file being applied from that date onwards. This shows that Rayleigh-cloudy winds are more sensitive to this than Rayleigh-clear winds. This is an interesting result and should be investigated further. An additional 0.5-m/s jump occurred towards the end of third week of October 2019. From that date onward, there seems to be a slight increasing trend in bias which takes the bias towards the end of the second reprocessing period (October 2020) to slightly more than +2.0 m/s. The bias of the second reprocessing is in general higher than that of the first. Bias of ascending orbits is slightly lower than that of descending orbits.

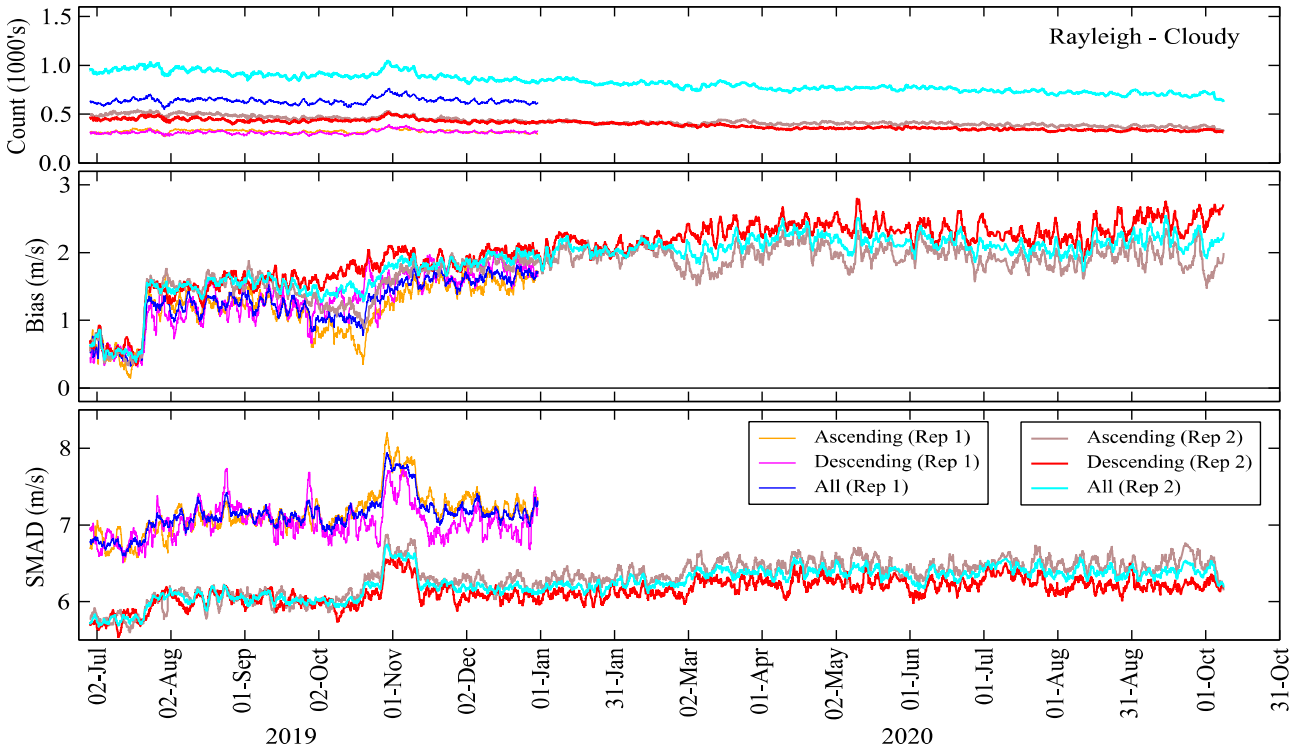


Figure 46: Time series of number of valid L2B Rayleigh cloudy wind observations, HLOS wind speed bias and SMAD between L2B Rayleigh cloudy and ECMWF IFS model winds. Running means along 30 successive files (orbits) are shown. Curves for all data as well as for ascending and descending orbits are shown for the both first (Rep 1) and second (Rep 2) reprocessings. Note that the first reprocessing curves end at the end of 2019.

According to the lower panel of Figure 46, SMAD varied initially around 5.8 m/s before it jumped by about 0.3 m/s during the third week of July 2019. This level continued till the two-week period (from 28 October till 10 November 2019) of special refined range-bin setting intended for AMV comparison which witnessed enhancement of variability as can be seen from the lower panel of Figure 46. The SMAD returned to its value at about 6 m/s afterwards. There has been very mild increasing trend since then and reached 6.3 m/s at the end of the second reprocessing period (October 2020).

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. Furthermore, the second reprocessed Rayleigh cloudy wind show significant improvement in agreeing with the model that the first reprocessing. The SMAD reduction is about 1 m/s (~14%).

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 second reprocessing. The agreement between the 5.5 million collocations of Rayleigh cloudy and model wind is acceptable with a global bias of +1.9 m/s, SDD of 7.1 m/s and SMAD of 6.3 m/s. The correlation coefficient is 0.925. Some other statistics are provided in the lower right corner of the plot. These statistics show that Rayleigh cloudy winds are not as good as those from Rayleigh clear or Mie cloudy. However, the degradation of Rayleigh clear means that towards the end of the reprocessing period, both Rayleigh products are of comparable SMAD values.

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 following a

"stretched" S-shape. This shape can be attributed either to nature of Rayleigh cloudy wind being much noisier than the model wind or possibly some model issues.

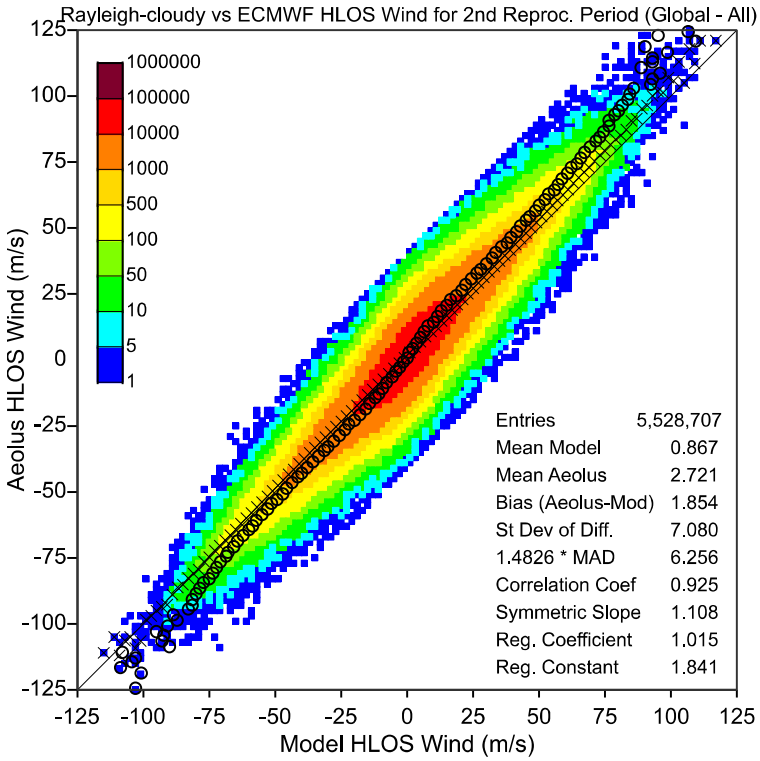


Figure 47: Density scatter plot comparing L2B Rayleigh cloudy to ECMWF model HLOS winds over the whole globe during the whole period of second reprocessing. Boxes are colour coded based on the number of collocations within the box. The circles are the average model value given L2B value while the crosses 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 second reprocessed dataset. The bias distribution does not show a symmetric shape around zero. The negative bias at the negative HLOS values is rather small compared to the positive bias at the positive side of HLOS. The bias curve does not pass through the origin. A uniform reduction of about 1.8 m/s would bring the distribution closer to the symmetry and reduce the bias in the product significantly. The bias becomes rather high (about 10 m/s) at extreme values (above 90 m/s). Note that the number of those values is rather small. The distribution of the SDD with respect to the HLOS wind is rather flat at around 8 m/s for most of the range. 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 evolution of the agreement between Rayleigh cloudy and ECMWF model is represented by four selected monthly scatter plots in Figure 49. The marginal gradual increase in SDD and SMAD values as time progressing can be easily seen by comparing the statistics.

The second reprocessed Rayleigh cloudy dataset was discriminated based on the geographical location being in extratropical Northern Hemisphere (NH), Southern Hemisphere (SH) or in the Tropics. 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 "stetched" S-shape distribution of Rayleigh cloudy versus model HLOS winds is more obvious in the Tropics as can be seen in the lower panel of Figure 50.

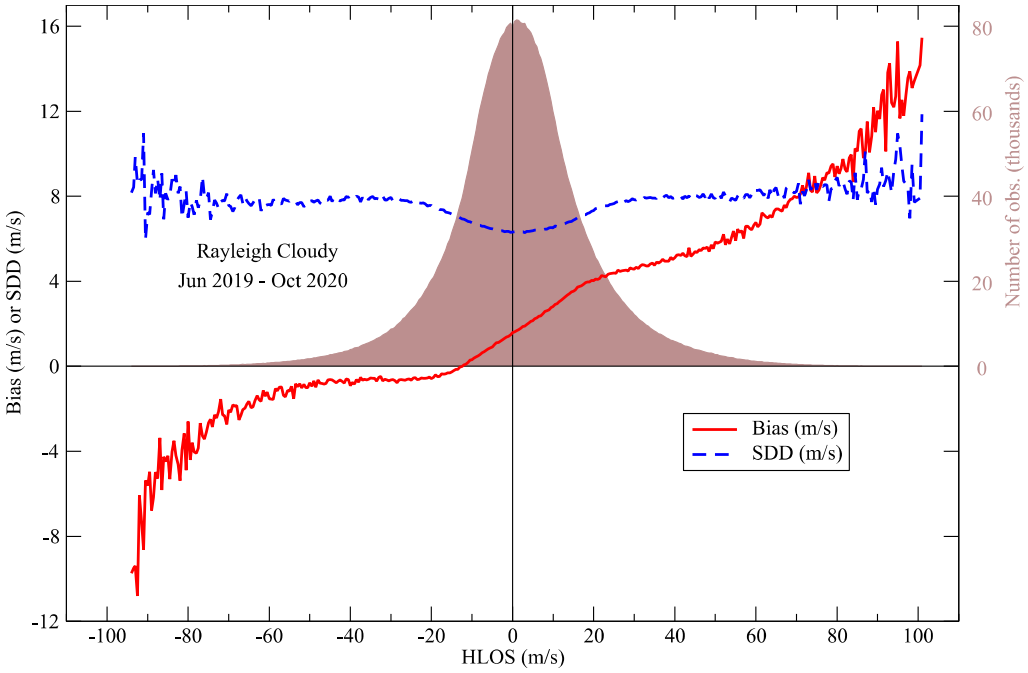


Figure 48: Number of valid Rayleigh cloudy HLOS wind observations, bias and SDD (compared to ECMWF model) as functions of HLOS wind for the whole second reprocessed dataset.

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 second 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 amount 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 second 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 the continents, including the Antarctica, ITCZ and the monsoon region. Slightly reduced SDD values can be seen on the western sides of the continents.

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 number of observations from each orbit direction (ascending vs descending) are almost equal as expected. The biases with respect to the model from ascending and descending orbits are not much different from each other before late September 2019. From that time onward, winds from descending orbits started to be consistently higher than those from ascending orbits. The deviation reached its peak during the second half of October where there was up to 1 m/s difference between the two.

Figure 46 shows also that before early October 2019, there was no systematic differences in SMAD and SDD values from descending and from ascending orbits. However, after that, observations from descending orbits are slightly better than those from ascending orbits.

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.

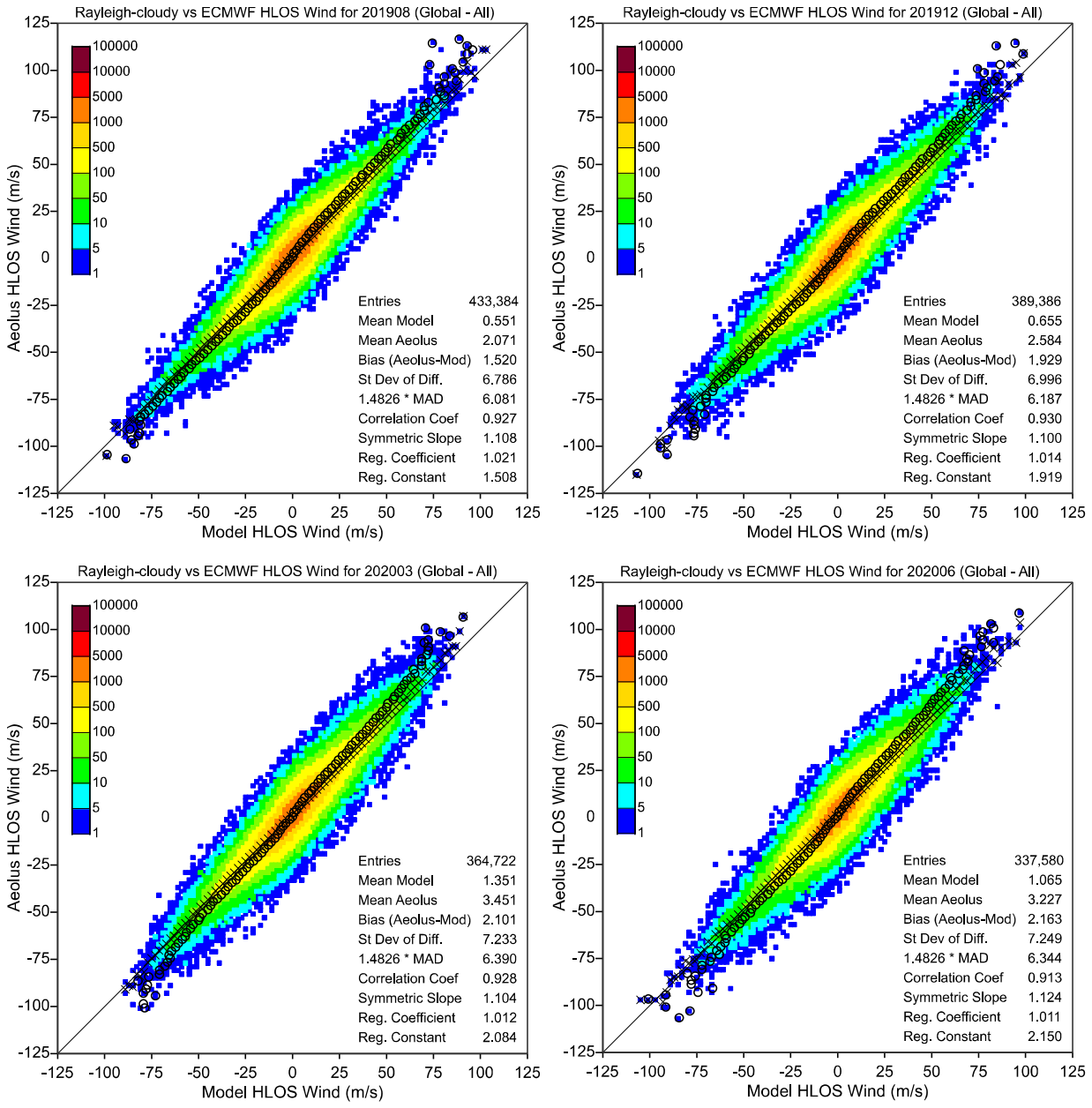


Figure 49: As in Figure 47 but each scatter plot represents the monthly data of 4 selected months: August 2019 (upper left), December 2019 (upper right), March 2020 (lower left) and June 2020 (lower right). Refer to Figure 47 for the meaning of crosses, circles and colour-coding.

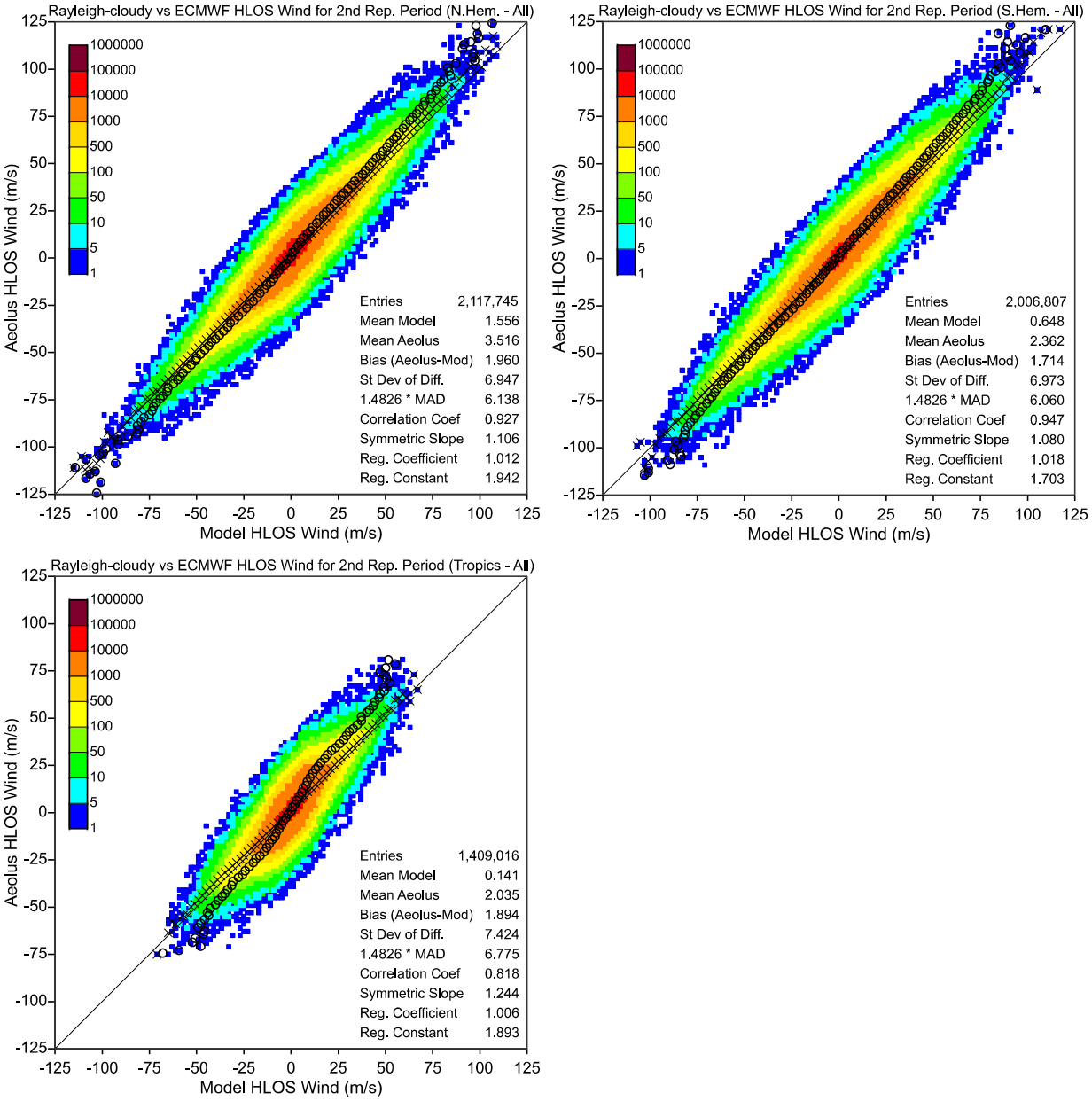


Figure 50: As in Figure 47 but with discrimination between extra tropics (upper) Northern Hemisphere (upper left) and Southern Hemisphere (upper right) and Tropics (lower). Refer to Figure 47 for the meaning of crosses, circles and colour-coding.

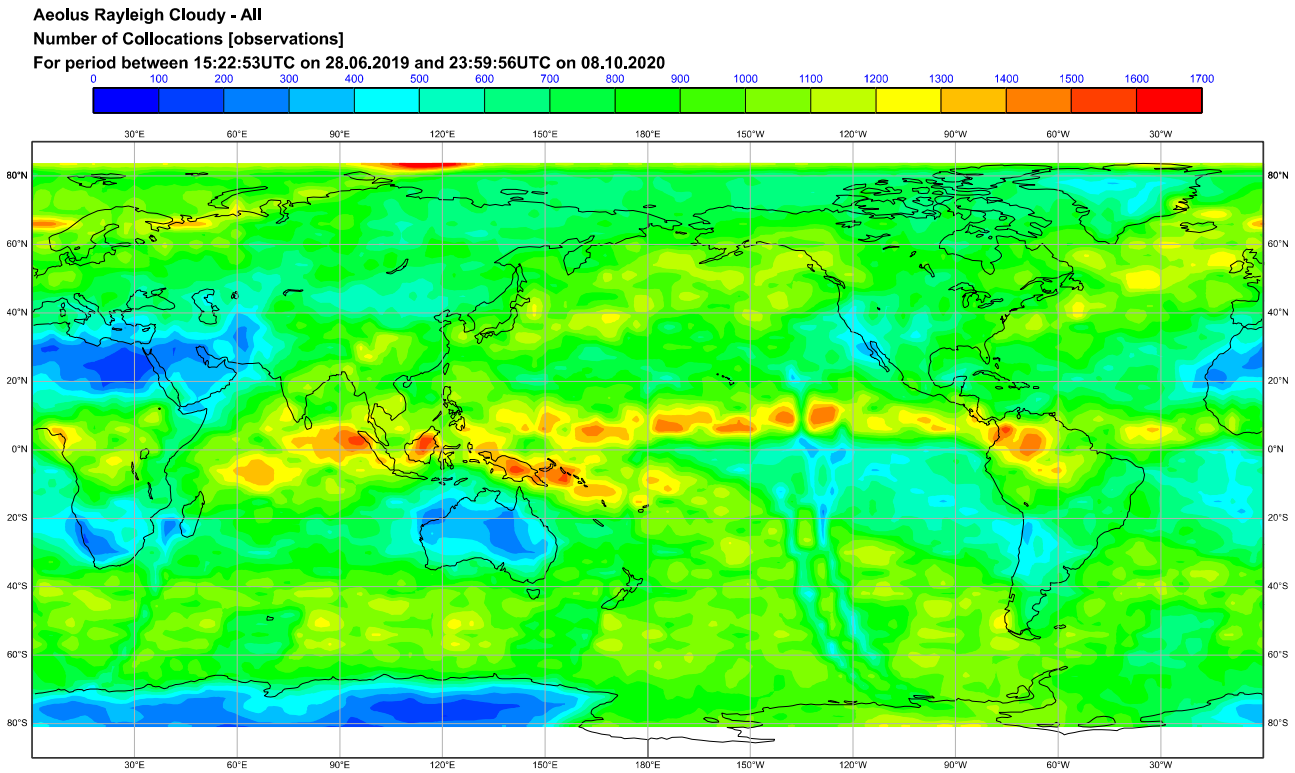


Figure 51: Geographical distribution of number of Rayleigh cloudy L2B observations per 3°X3° latitude-longitude box during the whole second reprocessing.

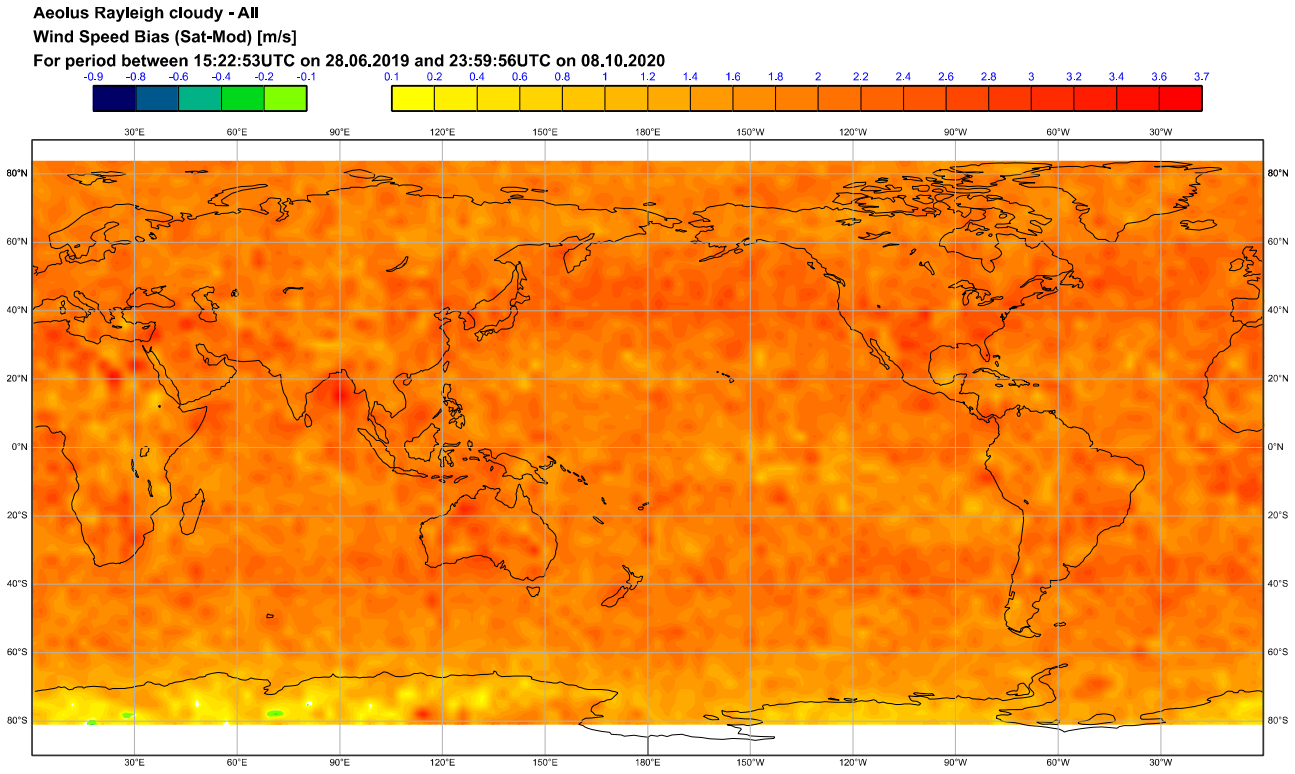


Figure 52: Geographical distribution of HLOS wind speed bias between Rayleigh cloudy and ECMWF model for the whole second reprocessing dataset.

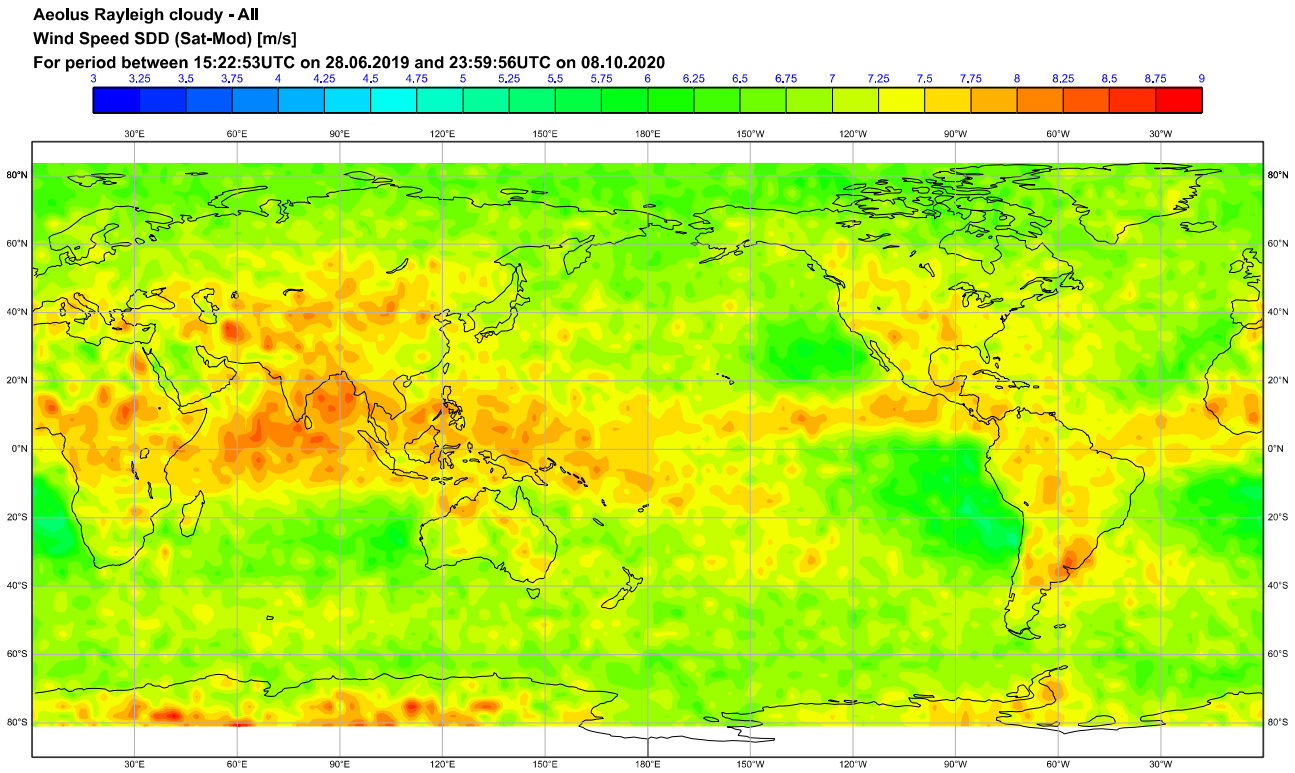


Figure 53: As in Figure 52 but for the Rayleigh cloudy SDD.

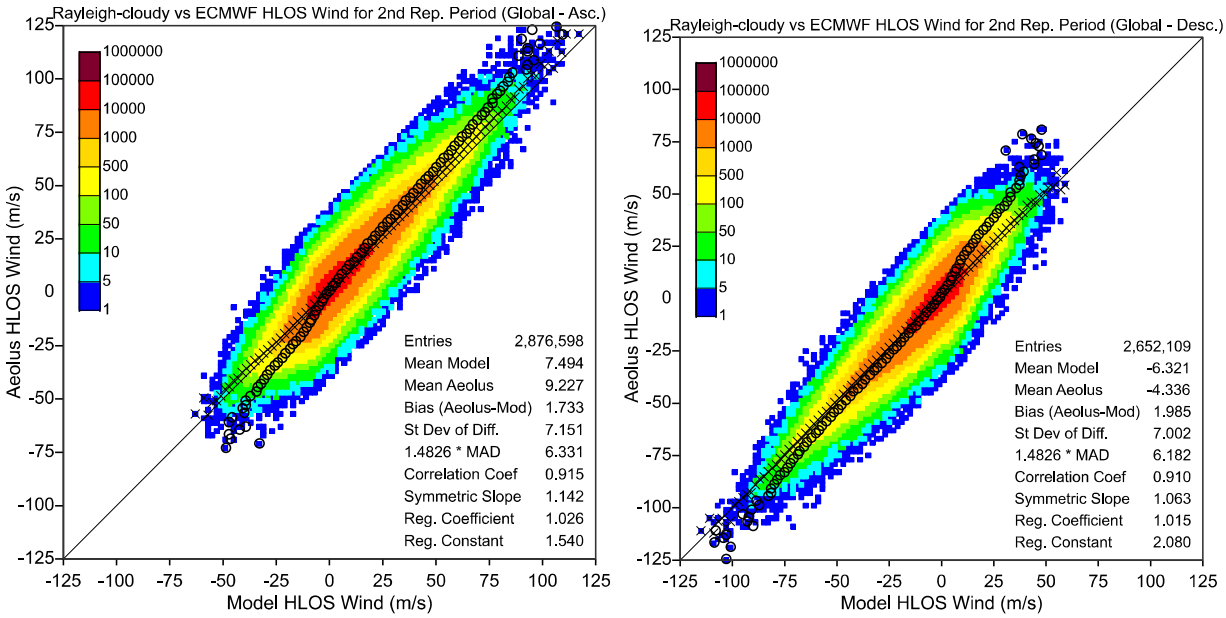


Figure 54: As in Figure 47 but with the discrimination between ascending (left) and descending (right) orbits. Refer to Figure 47 for the meaning of crosses, circles and colour-coding.

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. The overall bias is rather high. It fluctuates around +2 m/s and almost independent of the altitude.

The overall SMAD and SDD profiles in Figure 55 follow a "reflected" ξ-shaped (xi-shaped) curve. SMAD and SDD values at the lower part of the profiles are about 5.5 m/s and 6.5 m/s, respectively. The values increase linearly up to 7.0 m/s and 7.7 m/s at an altitude of 9 km. The values start to decrease briefly until they reach the 12 km altitude after which another increase start to reach 7.7 m/s and 8.5 m/s, respectively at altitude 18.5 km. After hovering around the same value for 2-3 km, SMAD and SDD values start to increase again until they reach 10.5 m/s at the top of the profile.

SMAD and SDD profiles in Figure 55 also suggest that observations from descending orbits are in better agreement with the model at the lower parts of the profiles (below 12-km altitude) and the upper parts of the profile (above 23-km altitude). In the middle (12-23 km), observations from ascending orbits are in better agreement. The SMAD differences are rather large while the SDD differences are marginal possibly due to existence of outliers.

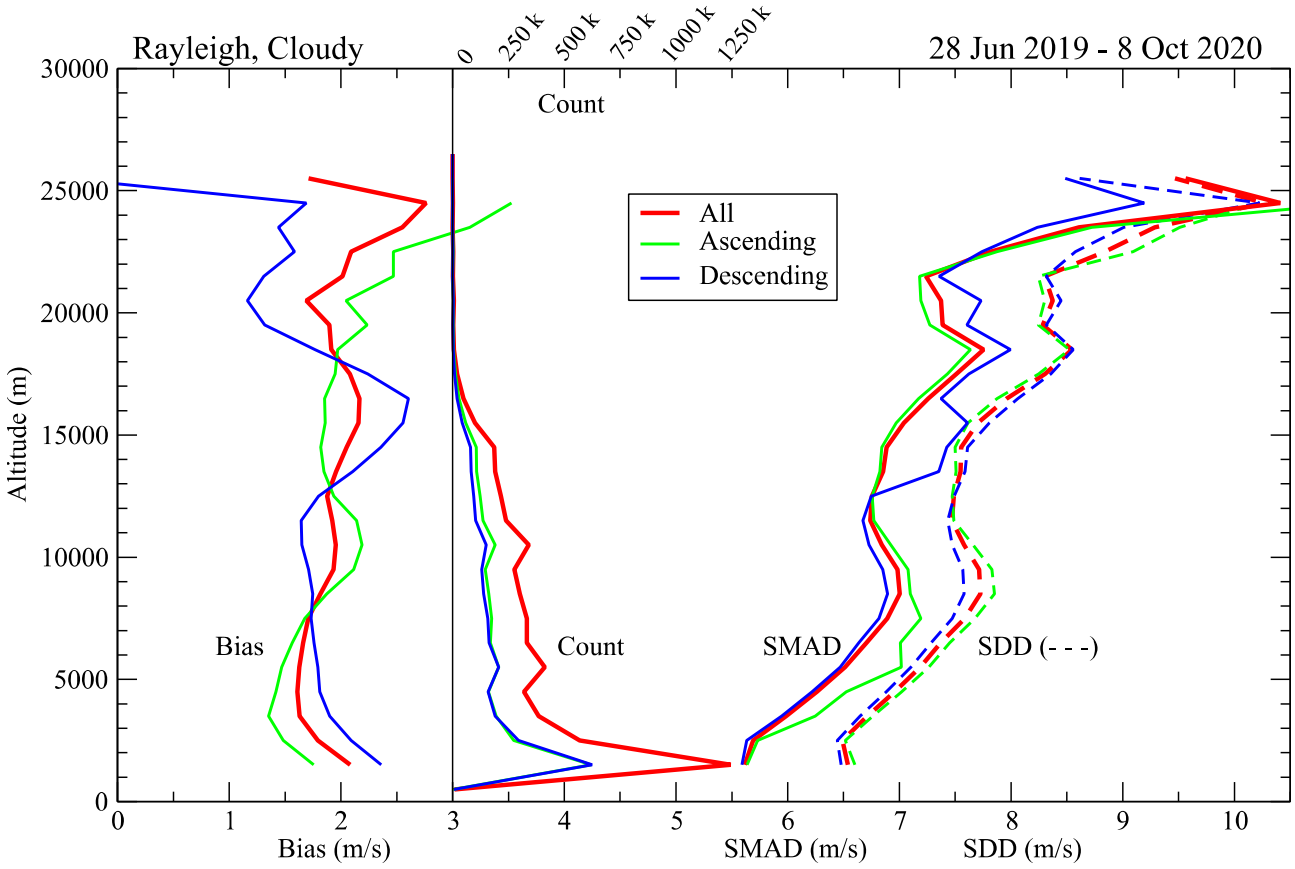


Figure 55: Profiles of L2B Rayleigh cloudy HLOS wind deviations from ECMWF model (bias, SDD and SMAD) together with the number of collocations at each altitude. Profiles show discrimination between ascending and descending orbits in addition to all data.

8 Conclusions

Second Aeolus reprocessed data set covering the period from 24 June 2019 till 9 October 2020 produced by the PDGS using the processing baseline 11 (B11) which is the processor versions deployed operationally in NRT in October 2020, namely: L1bP 7.09.1, L2aP 3.11.1, L2bP 3.40.2. From the user's point of view, data are available from 28 June 2019 as the first 4 days were used for laser instrument sensitivity tests. The data covers the operation period of the second Flight Model laser (FM-B). Homogeneous L1A, L1B, L2A and L2B data products and the respective auxiliary files were produced and verified.

Improvements:

The following main improvements were introduced compared to operational and first reprocessed data:

- Correction of wind bias in L2B product using daily updates of the correlation between ECMWF IFS O-B (observation-background) statistics and temperatures from the telescope primary mirror M1. This eliminates large wind biases for the NRT product produced between 1 January (the day just after the first reprocessing stops) and 20 April 2020 (the date on which M1 bias correction was activated in NRT processing). This correction is applied to Rayleigh and Mie L2B winds. It implicitly corrects for constant drifts in bias). Unlike the NRT M1 bias correction which uses data from the previous day to correct the biases of the day, the reprocessing considers the data of the day for that leading to even lower biases. This is one of the advantages of the reprocessing campaigns.
- Multiple linear regression based on M1 temperature is used to calibrate radiometric coefficients K_{ray} and K_{mie} along the orbit. The radiometric coefficients are needed for L2A products. Furthermore, the parameter attenuated backscatter is provided for the whole period.
- Correction of hot pixels per orbit in time periods between dark signal measurements (DUDE) for L1B product, which is an improvement compared to updating the hot-pixel correction only 4 times per day after DUDE for the NRT data products; improvement of the hot-pixel correction for the upper-most Rayleigh range bin for some DUDE calibrations of the NRT product up to 21 October 2019.
- L2B wind products were flagged invalid for periods of known instrument tests or degraded data quality, which eases the automatic use of L2B products.

The reprocessed data products have the following characteristics:

General:

- The quality of signal levels (both Rayleigh and Mie) in the L1B product is improved with respect to hot-pixel dark current levels. This leads to improvements in L2A and L2B products, since L1B signal level is used as input for them.
- The radiometric calibration of the L2A products is improved and reflects the observed signal loss in the atmospheric path of the instrument over time; this leads to lower systematic errors for the backscatter and extinction profiles, with slightly lower values than the NRT products; this results in a lower number of unrealistic, negative values for the backscatter and extinction coefficient.

Mie cloudy wind:

- The Mie cloudy wind bias with respect to the ECMWF model shows global averages around -0.25 m/s (underestimation), with most orbits showing bias within ±0.4 m/s. This is a slight improvement compared to NRT products, but slight degradation with respect to the first reprocessed dataset with average bias of -0.15 m/s.

- Mie wind bias is small everywhere with some enhancements over Antarctica and the major mountain ranges.
- Mie wind bias is almost invariant with respect to altitude except at both ends of the profile where not much data available.
- Mie-wind random error is small with SMAD (scaled median absolute difference) fluctuating around 3.5 m/s except for two weeks from 28 October till 10 November 2019 when a special refined range bin setting (for AMV comparison) was used. Mie-wind random error is small everywhere with some enhancements over the mountain ranges, the Intertropical Convergence Zone (ITCZ) and the monsoon region.
- Mie winds from ascending orbits agrees better with the model compared to that from descending orbits.
- Mie wind random error increases linearly with altitude up to about 16 km beyond which the error reduces briefly before it keeps increasing above 21 km.

Rayleigh clear wind:

- The Rayleigh clear wind is globally unbiased compared to the ECMWF model except for October 2019 (and possibly in October 2020 as well) and March 2020. This is a significant improvement compared to the bias in the NRT product before April 2020. The, already negligibly small bias in the first reprocessed dataset, has become even smaller.
- Rayleigh clear wind shows negligibly small bias almost everywhere with clear distinction between positive bias in the Tropics (in addition to Greenland and the Antarctica) and negative bias in the extra tropics.
- Rayleigh wind shows a small positive bias, which is almost invariant with respect to altitude, between altitudes 10 and 22 km. This small positive bias is balanced by small negative bias outside that altitude range.
- Random error from Rayleigh clear wind was rather small (SMAD less than 4.5 m/s) before mid-October 2019. The random error started to gradually increase afterwards until it reached its highest (SMAD of about 6.3 m/s) around September 2020. This trend was disrupted by a jump in the error during the special range-bin setting period (28 October - 10 November 2019) when SMAD value reached 5.5 m/s.
- Rayleigh-wind random error is relatively small everywhere with some enhancements over Antarctica, the mountains ranges and the monsoon area.
- Random error in Rayleigh clear winds during descending and ascending orbits is similar except for the period from December 2019 and April 2020 when observations from descending orbits show slightly lower errors compared to the model.
- Rayleigh-wind random-error shows a ε-shaped profile with two local minima at 5-km and 12-km altitudes and a local minimum at 9-km altitude. The minimum values are 4.5 m/s and 5.5 m/s.

Rayleigh cloudy wind:

- The number of valid Rayleigh cloudy wind observations in the second reprocessing is about 50% higher than that of NRT and that of the first reprocessing.
- The Rayleigh cloudy wind bias with respect to the ECMWF model at the beginning of the FM-B period (late June 2019) was about +0.5 m/s (overestimation). It kept increasing until it became more than

2.0 m/s towards the end of the second reprocessing period (early October 2020). The bias of the second reprocessing is in general higher than that of the first.

- Bias of Rayleigh cloudy wind is high everywhere. However, it is lower over Antarctica.
- Rayleigh cloudy wind show lower bias wrt the model during ascending orbits compared to that of descending orbits.
- Rayleigh cloudy wind bias from all orbits is almost invariant with respect to altitude. However, the bias from ascending and descending orbits varies with altitude.
- The random error of Rayleigh cloudy wind wrt ECMWF model shows significant reduction by 1 m/s (~14% reduction) compared to that of NRT and the first reprocessing.
- Rayleigh cloudy random error increased slowly over the whole period (SMAD increased from 5.8 m/s in June 2019 to 6.3 m/s in October 2020). The error during the two-week period (from 28 October till 10 November 2019) of the special refined range-bin setting intended for AMV comparison which witnessed enhancement of variability.
- Towards the end of the second reprocessing period, the random error from both classes of Rayleigh channel were very close to each other.
- Rayleigh cloudy wind random error is rather high everywhere with extra enhancements over the continents, the ITCZ and the monsoon region.
- Rayleigh cloudy wind from descending orbits agrees better with the model compared to the ascending orbits.
- Random error of Rayleigh cloudy wind wrt ECMWF model increases with altitude (with minor inversions).

Limitations:

However, the reprocessed data products have the following known limitations:

- Data products should be only used from 28 June 2019 onwards due to laser instrument sensitivity tests.
- There are several short gaps in data coverage due to satellite downlink problems, orbital manoeuvres or performed calibrations as outlined in: https://earth.esa.int/eogateway/instruments/aladin/quality-control-reports
- L2B data product was flagged invalid during periods of degraded data quality (mostly due to dedicated instrument test periods). The above link can be consulted for data exclusion (and unavailability) periods of L1B and L2A products.
- 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.
- The L2B Mie cloudy winds show a constant bias of -0.25 m/s (underestimation). This is a known issue and related to the choice of the Mie estimated error threshold which will be further investigated.
- There is a minor issue with the Mie winds related to “wobbly” bias behaviour wrt the wind speed value. The local bias peaks migrate with time. This is turned out to be a result of uncorrected nonlinearity. A solution was found and will be implemented in the forthcoming reprocessings.

- The L2B Rayleigh clear winds show an altitude dependent bias of about 1 m/s over 20 km with an underestimation at lower altitudes and overestimation at higher altitudes; this is related to an imperfect Rayleigh response calibration using the atmospheric temperature profile.
- As the L2B bias correction is based on global O-B statistics, there are higher regional biases; a small difference in bias is observed for ascending and descending paths.
- Both the L2B Mie and Rayleigh winds show an increase for the bias starting end September 2019 until beginning of November 2019 with a maximum in mid-October 2019 for the ascending and descending paths; maximum values in mid-October are about -0.4 m/s to 0.4 m/s (Rayleigh) and -0.4 m/s to 0.0 m/s (Mie); this is also present again with lower maximum amplitudes in March 2020 for both Mie and Rayleigh winds; the cause of this behaviour is related to a seasonal issue for the correction of the solar background; the cause will be further investigated and hopefully lead to improved bias correction for the next reprocessing campaign.
- L2B winds show higher random error during the period with AMV range-bin settings from 28 October to 10 November 2019; this is as expected, since random error is a function of altitude and range-bin size, so a globally changed altitude sampling leads automatically to a change in random error for both channels; during that period also the L2A product shows limitations: the Mie Core Algorithm (MCA) can be used, but the Standard Correct Algorithm (SCA) cannot be fully exploited, i.e. the AMV range bin settings are not being properly handled for mid-altitudes where several Rayleigh bins match one single Mie bin. As a result, SCA products are only partially valid during this period.
- L2B Rayleigh cloudy winds still show significant bias (0.5 m/s to 2 m/s) and significant higher random error than Rayleigh clear winds at the beginning of the period (5 m/s to 6 m/s) although the random error for both cloudy and clear was almost the same (6.3 m/s) during the last 7 weeks of the period. A dedicated correction algorithm for these winds has been prepared and we hope this can be improved upon for a future reprocessing.
- The continual degradation of Rayleigh clear wind over the FM-B period, unfortunately, will reduce the impact of such data on NWP. This needs to be taken into account while evaluation any long OSE that uses this dataset.

9 Suggestions for Future Work

Based on the observed limitations, the following investigation suggestions for improving the quality of the forthcoming reprocessings as well as the NRT processing:

- The L2B Mie cloudy winds show a constant bias of -0.25 m/s (underestimation). This is a known issue and related to the choice of the Mie estimated error threshold which will be further investigated. Furthermore, the shape of bias profile with altitude needs explanation.
- L2B Mie cloudy wind shows higher biases over Antarctica while Rayleigh cloudy wind shows the opposite. The reason behind that and possible correction need investigations.
- The L2B Rayleigh clear winds show an altitude dependent bias of about 1 m/s over 20 km with an underestimation at lower altitudes and overestimation at higher altitudes; this is related to an imperfect Rayleigh response calibration using the atmospheric temperature profile.
- As the L2B bias correction is based on global O-B statistics, there are higher regional biases.
- Mie winds from ascending orbits agrees better (lower random errors) with the model compared to that from descending orbits. However, Rayleigh cloudy winds show the other way around (descending is better). On the other hand, Rayleigh clear winds during descending and ascending orbits compare similarly to the model except for the period from December 2019 and April 2020 when observations from descending orbits show slightly lower errors compared to the model in agreement with Rayleigh cloudy. This needs further explanation. Does the enhanced convection at 18 local solar time has any role to play?
- The number of valid observations from ascending orbits is larger than descending in the upper troposphere. Does the enhanced convection at 18 local solar time has any role to play?
- Both Rayleigh classes show enhanced random errors wrt the model over Antarctica while Mie winds tends to show the opposite. An explanation is needed for this behaviour.
- Both the L2B Mie and Rayleigh winds show an increase for the bias starting end September 2019 until beginning of November 2019 with a maximum in mid-October 2019 for the ascending and descending paths; maximum values in mid-October are about -0.4 m/s to 0.4 m/s (Rayleigh) and -0.4 m/s to 0.0 m/s (Mie); this is also present again with lower maximum amplitudes in March 2020 for both Mie and Rayleigh winds; the cause of this behaviour is related to a seasonal issue for the correction of the solar background; the cause will be further investigated and hopefully lead to improved bias correction for the next reprocessing campaign.
- L2B winds show higher random error during the period with AMV range-bin settings from 28 October to 10 November 2019; this is as expected, since random error is a function of altitude and range-bin size, so a globally changed altitude sampling leads automatically to a change in random error for both channels. Is it possible to improve on this?
- L2B Rayleigh cloudy winds still show significant bias (0.5 m/s to 2 m/s) and significant higher random error than Rayleigh clear winds at the beginning of the period (5 m/s to 6 m/s) although the random error for both cloudy and clear was almost the same (6.3 m/s) during the last 7 weeks of the period. A dedicated correction algorithm for these winds has been prepared and we hope this can be improved upon for a future reprocessing.