TN15.3

End-to-end testing of the continuous mode L2B processor

Authors

MICHAEL RENNIE
ECMWF

ALAIN DABAS
MÉTÉO-FRANCE

CHANGE LOG

Version	Date	Comment
1.0	22 Feb 2013	MR: Initial attempt
1.1	15 April 2013	MR: Increased content significantly
1.2	30 May 2013	MR: take on suggestions of L. Isaksen and O. Reitebuch
1.3	16 Jul 2013	MR: Add new section explaining new discoveries made after the main investigation
2.0	Nov 2013	MR: following ESA review
3.0	Jan 2014	MR: outstanding calibration testing of CCN2 added
3.1	March 2014	MR: considered comments of ESA, particularly in section 5

Contents

1 Introduction

This technical note is an extension of the work done for TN15.2: "Aeolus observational requirements and the measurement grouping algorithm" [RD7]. It focusses on the quality assessment of Aeolus Level-2B (L2B) winds produced from the continuous mode (CM) chain-of-processors end-to-end testing system using realistic test scenarios, and in particular how the L2B grouping algorithm affects the results.

The CM L2B processor, incorporating the new measurement grouping algorithm, as described in [RD7], is subjected to a set of end-to-end performance tests including instrumental effects (e.g. detection noise) and variable atmospheric conditions. The statistics of the L2B HLOS wind errors are analysed, and anomalous behaviour is discussed as appropriate throughout the technical note. The verification of L2B products is of course possible with simulations, since the truth is known i.e. wind input to the Aeolus simulator. It is possible to analyse the sensitivity of the L2B results to the many input parameter settings. This is a useful exercise to confirm the processing is behaving nominally, so that the L1B/L2B processing team is well prepared for the real Aeolus data. The simulator is unlikely to agree perfectly with the eventual real Aeolus data; some aspects of the mission are very difficult to simulate accurately, meaning that probably some problems will only be encountered with real data.

Because CM Aeolus data naturally leads to spatially-adjacent observations, it is of interest to investigate the **observation error correlation** (for the Rayleigh wind results only in this report) as a function of horizontal distance — such investigations are presented in section 2.2.3.

The **accuracy of the observation error estimate** (part of the L2B processing product) is investigated in section 2.3.1 i.e. determining the consistency of the estimated standard error values with the "real" error statistics and the dependency on e.g. L2B group size.

The technical note **does not assess the expected impact of CM Aeolus data upon NWP** (as was suggested it should do in the contract), since such investigations have already been addressed as part of the work for the separate ECMWF Impact Studies [RD5] and VHAMP [RD13] contracts.

This technical note was requested by ESA for Sub-task 1.3 of [AD1] and further documented in [AD2], work package 2720.

1.1 Documents

1.1.1 Applicable documents

	Title	Ref	Ver.	Date
[AD1]	Statement of Work for Change Request #4 "Aeolus Level 2B/C Processor - Implementation of Continuous Mode Operations & Extended Pre-Launch Support". Contract 18555/04/NL/MM	AE-SW-ESA-GS-038	1.0	Dec 2010
[AD2]	ESTEC Contract No. 18555/04/NL/MM "Change request for CCN No.4"	N/A	1.0	Feb 2011

1.1.2 Reference documents

	Title	Ref	Ver.	Date
[RD1]	TN15.1 Inventory of Aeolus Target Assimilation Systems	AE-TN-ECMWF-GS-151	1.3	Oct 2012
[RD2]	ADM-Aeolus level-2B algorithm theoretical baseline document	AE-TN-ECMWF-L2BP-	2.4	Dec 2012

		0024		
[RD3]	The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study	By Z. Liu and F. Rabier, Q. J. R. Meteorol. Soc. (2002), 128, pp. 1367-1386		2002
[RD4]	TN3.1a Test cases for the L2B processor	AE-TN-KNMI-GS-0031a	1.0	Feb 2011
[RD5]	TN3 of ESA study contract 4000104080: Synthesizing of draft Aeolus observation requirements, collection of simulated observations and support to VAMP CCN2 contract studies	AE-TN-ECMWF-impact-study-003	4.1	Jan 2013
[RD6]	ADM-Aeolus Level-2B/2C Processor Input/Output Data Definitions Interface Control Document	AE-IF-ECMWF-L2BP-001_20121211_IODD_Iss 2.00	2.00	Dec 2012
[RD7]	TN15.2 Aeolus observational requirements and the measurement grouping algorithm	AE-TN-ECMWF-152-v2.0_20131121	2.00	Nov 2013
[RD8]	TN3.1b Test results for the L2B processor	AE-TN-KNMI-GS-0031b	1.0	Feb 2011
[RD9]	VAMP – Vertical Aeolus Measurement Positioning, Final Report	AE-TN-KNMI-VAMP-FR	1.0	Nov 2010
[RD10]	ADM-Aeolus System Requirements Document	AE-RS-ESA-SY-001	03.1	Sep 2010
[RD11]	Aeolus Level 1b Processor and End-to-End Simulator End-to-End Simulator Detailed Processing Model	ADM-MA-52-1801	3.01	Jul 2012
[RD12]	Generation and update of AUX_CSR	AE-TN-MFG-L2P-CAL-0003	2.0	Jan 2012
[RD13]	Final Report, Vertical and Horizontal Aeolus Measurement Positioning, CCN2 to ESA Contract 20940	AE-FR-VHAMP	1.0	September 2013
[RD14]	The definition of an atmospheric database for Aeolus	By G. J. Marseille et. al. Atmos. Meas. Tech., 4, 67-88		2011

1.1.3 Acknowledgements

Thanks to KNMI, and in particular Jos de Kloe and Gert-Jan Marseille, for lots of help with the chain-of-processors and understanding some of the processing results. Thanks to Oliver Reitebuch and Lars for very useful reviews of the document.

Acronyms

AOCS	Attitude and Orbit Control System
ASCII	American Standard Code for Information Interchange (a character coding scheme)
ATBD	Algorithm Theoretical Baseline document
BM	Burst mode
BRC	Basic Repeat Cycle
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CM	Continuous mode
CSR	Corrected Spectral Registration
DA	Data assimilation
DEM	Digital Elevation Model
DWL	Doppler Wind Lidar
ECMWF	European Centre for Medium-Range Weather Forecasts
EE	Earth explorer
EGM	Earth Gravitational Model
ERA	ECMWF re-analysis
E2S	ADM-Aeolus End-To-End Simulator
FP	Fabry-Pérot
HLOS	Horizontal Line Of Sight
IDL	Interactive Data Language
IODD	Processor Input/Output Data Definitions Interface Control Document
ISR	Instrument Spectral Registration
KNMI	Royal Netherlands Meteorological Institute
LITE	Lidar In-space Technology Experiment
L1B	Level-1B
L2B	Level-2B
L2Bp	L2B processor
N/A	Not applicable
MRC	Mie response calibration
NaN	Not a number
NWP	Numerical weather prediction
PM	Progress meeting
QC	Quality control
RMA	Reference model atmosphere
RRC	Rayleigh Response Calibration
SNR	Signal to noise ratio
SR	Scattering ratio
SRD	System requirements document

TBD	To be determined
TN	Technical note
VHAMP	Vertical and Horizontal Aeolus Measurement Positioning
WGS	World Geodetic System
XML	Extensible Markup Language
ZWC	Zero wind correction

2 End-to-end testing results

This technical note assesses the quality of the Aeolus operational continuous mode (CM) processing chain with a variety of simulated scenarios that were generated with the Aeolus End-to-End Simulator (referred to as E2S). The joining of the simulator to the actual ground segment processing chain was developed by KNMI: this testing makes use of Python scripting also used for the "chain-of-processors" test report (TN3.1b [RD8]).

The following processing chain is run: KNMI's atmospheric database → E2S (v3.02) → L1BP (v6.01) → L2BP (v2.00) → IDL→ verification plots. The plotting and statistics for this technical note are produced with newly developed tools written in the IDL programming language rather than the Matlab tools which are used for TN3.1b. The IDL program reads the ASCII dumps of L2B EE-format files (via a feature available in the L2Bp package: `Write_L2B_Text_Product`), it also reads the E2S input scenario data (Segment*.xml files, providing meteorological and optical properties inputs). From this data the verification statistics presented throughout the technical note can be calculated.

The IDL verification tools were developed because the already available Matlab verification (at the time of investigation) was not compatible with the continuous mode L2B file format and it was not trivial to update; it was easier to start afresh, since the author was not familiar with Matlab at the time. Creating new software also helped to comprehend the new CM L2B file format. The IDL tools have developed, during late 2012/early 2013, into more detailed analysis tools than the Matlab counterparts, and are preferred for this investigation.

This investigation uses a number of informative test cases for assessing the quality of the L2B results. The tests are a selection of those used in TN3.1b but with some modifications as required (note that a TN3.1b with all the test cases for the CM processors will be produced at a later date). For most of the presented test cases, the E2S settings have been selected for the realistic assessment of random errors i.e. all noise sources are switched "on". **The default E2S noise settings are used, unless otherwise stated.** The E2S simulated random noise terms are: Poisson noise, Laser frequency variation, AOCS position and orientation errors (i.e. random positioning and pointing errors), and ACCD readout noise, see [RD11] for an explanation of these sources of error. The laser energy is set to 110 mJ for the majority of tests; the assessment of 80 mJ i.e. the recently proposed lower laser energy for the start of the mission will be documented as part of another work package at a later date and is also used in chapter 5 of this report (version 3.0 onwards).

The E2S atmospheric inputs are provided by the KNMI atmospheric database; see [RD14]. The E2S inputs of optical parameters (particulate and molecular backscatter and extinction) and meteorological fields (temperature, pressure, HLOS wind) are reasonably realistic for the majority of test cases presented herein. For example, some tests have **optical parameters** (both cloud and aerosol) applicable at Aeolus' wavelength, as retrieved **from space-borne lidar missions** (LITE and CALIPSO[1], see [RD9] for details). The E2S meteorological and molecular backscatter/extinction inputs are derived from ECMWF NWP model fields. **The meteorological fields which are co-located with LITE data are obtained from the ECMWF ERA-40 reanalysis, which has a 120 km horizontal grid and 60 vertical levels, i.e. fairly poor horizontal resolution compared to state-of-**

[1] The processed level-1 CALIPSO data provides a global coverage of lidar particle backscatter at high vertical and horizontal resolution (horizontal resolution is 3300 m, the vertical resolution is 125 m)

the-art NWP systems. **Higher resolution ECMWF data**, from the operational analysis in 2007, which has ~25 km horizontal grid and 91 vertical levels, **is co-located with the CALIPSO test cases**. **In the future, higher resolution ECMWF data may be incorporated into the database** (e.g. 10 km horizontal grid, estimated effective horizontal resolution 40-80 km, see [RD5], and 137 vertical levels are available to ECMWF research staff) for more realistic simulations due to more realistic wind variability.

The simulation of **some of the known sources of Aeolus' systematic errors has not been considered in this investigation i.e. those arising from imperfect calibration.** The calibration files used by the L2Bp (AUX_RBC_L2) and L1Bp (MRC, subsequently used in the L2Bp via the L1B product) were generated by E2S calibration procedure simulations with noise terms switched off (using ISR and MRC files provided by the L1B team). For the Rayleigh channel, an ISR file with noiseless Fabry-Perot transmission functions is converted into an auxiliary Rayleigh-Brillouin correction file (AUX_RBC_L2). The AUX_RBC_L2 file was made using a L2B processor tool developed in 2008 by P. Poli. This procedure skips some of the operational calibration chain steps, i.e. where a noisy RRC is input to the AUX_CSR updater to update the ISR to correct for atmospheric channel étendue effects; see [RD12]. The E2S settings for the tests in this report did not simulate the étendue, so the L2Bp generated AUX_RBC_L2 file is valid. For the Mie winds the MRC is also simulated without noise and is therefore idealistic in terms of the systematic errors that will result.

Another source of systematic error that is ignored in this report is the zero wind correction (the use of ground returns, which act as a zero wind calibration reference, as a bias correction for various affects e.g. satellite-to-earth relative velocity). In the test cases of this technical note **systematic miss-pointing (AOCS) errors are switched "off"** in the E2S. This is to avoid overcomplicating the investigation into L2B wind quality (i.e. the focus is on getting decent wind results without these complications first). Therefore the zero wind correction does not require application in the L2Bp for the tests performed.

The chosen Aeolus vertical range-bins for some of the cases tested here are shown in Figure 1. These range-bins aim at maximising Mie and Rayleigh range-bin overlap: see page 34 of [RD9]. Note that the Mie range-bins reach 23 km which is important for cross-talk correction of the Rayleigh channel (i.e. removing the effect of the Mie signal from the Rayleigh channel). Be aware that these are not the same range-bins as specified in ESA's System Requirements Document [RD10] for the performance requirements.

In the following sections, a series of increasingly realistic test cases are presented. The quality of the Rayleigh and the Mie L2B HLOS wind is assessed and any unexpected behaviour is discussed when appropriate.

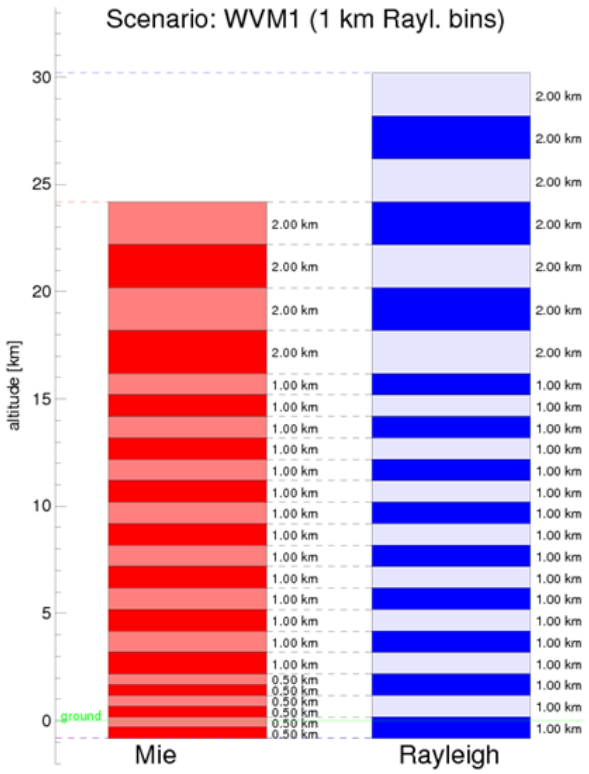


Figure 1. Aeolus vertical range-bins used in some of the end-to-end testing, for both channels. This selection is referred to as WVM_1km, see VAMP [RD9] for more details.

2.1 Simple scenario with artificial clouds

The first test case has a simplistic atmosphere: the E2S input HLOS wind is constant at 50 m/s. The E2S input temperature, pressure and particle backscatter profiles are homogeneous in the horizontal, but they do vary realistically in the vertical. Artificial clouds are generated (in the test scenario set-up) at three levels: 5, 10 and 15 km. The clouds stand out clearly in Figure 2 which shows a cross-sectional view (altitude against along-track distance) of the E2S input scattering ratio (SR). The E2S input SR (ρ_{true}) is calculated from the E2S input molecular and particulate (aerosol) backscatter profiles as follows:

$$\rho_{true} = \frac{\beta_p + \beta_m}{\beta_m} \qquad \textbf{(eq. 1)}$$

where β_p is the particle backscatter and β_m is the molecular backscatter.

This test was run to ensure the L2B results are good with a relatively simple case, before moving on to more realistic atmosphere test cases. The test case number is #0027 as described in [RD4], but it has been modified: the 4 BRCs are repeated ten times (with different random noise in each sample), to generate more reliable error statistics.

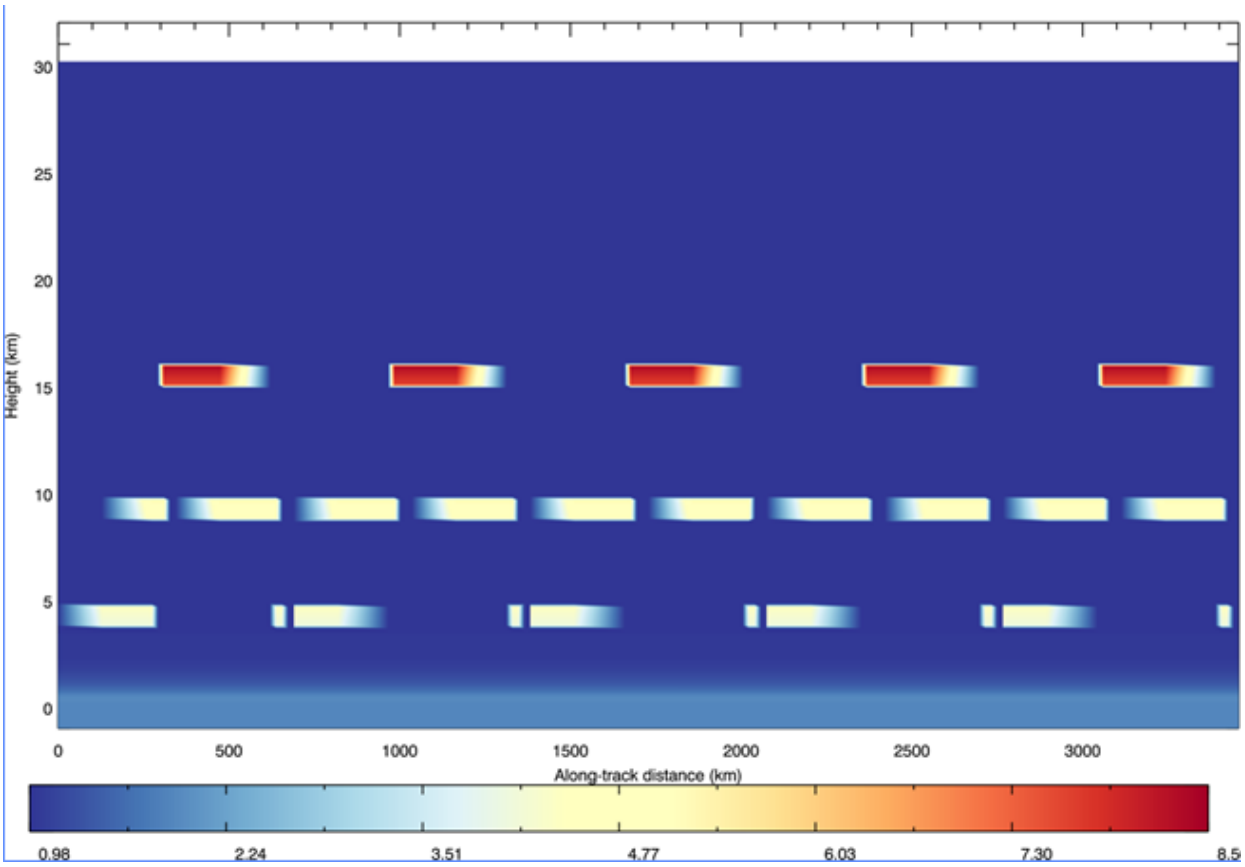


Figure 2. Cross-sectional view of E2S input SR along the orbital path. SR as derived from the E2S particle and molecular backscatter profiles. Units: dimensionless ratio.[2]

Some of the processing settings are now described. The realistic random noise sources are switched "on" in the E2S for this test (section 2 lists those settings). The L2B grouping algorithm is set to "advanced" (see [RD7] for an explanation) with a maximum horizontal accumulation length of 100 km for both the Mie and Rayleigh observations; hence it matches ESA's System Requirements Document (SRD) [RD10] for observations below 14 km. It should further be kept in mind that the ESA SRD requirements are valid for a horizontal averaging of 140 km above 14 km. The L2B processor was run using the L1B refined SR (as chosen in the AUX_PAR_2B file) to determine the classification of measurements into clear and cloudy. The classification threshold is 2.0 at 0 km height, linearly varying to 3.0 at 30 km. The L1B SR values at measurement level are averaged over the measurements that form a L2B wind, so the noise is reduced with observation level SR values. The Appendix (section

[2] There are a few artifacts in the plot due to the IDL contouring procedure not working ideally with the sharp changes in scattering ratio at the edges of the artificial clouds.

6.6) explains why the refined rather than the nominal SR was chosen; in summary the nominal SR was found to be too large for small SR values, which introduced large wind-speed dependent biases into Rayleigh winds through the cross-talk correction. The refined SR was much less biased for small SR values.

The L2B Rayleigh-clear wind results are assessed with error statistics such as those shown in Figure 3, which shows profiles of the mean error, standard deviation of error and number of observations of the L2B HLOS wind observationsplotted as function of altitude (referenced to the geoid)[3]. Appendix section 6.1 defines how these and other error statistics are calculated. **The error bars displayed on the mean error profile (blue) in Figure 3 are the 68% confidence interval on the mean error**. The green line shows ESA's System Requirements Document [RD10] performance requirements for the random error for Aeolus winds[4]. The orange line, associated with the orange axis, shows the number of observations used in the statistics on each altitude bin (24 altitude ranges to match the Rayleigh range-bins are chosen). The Mean(all errors) and Stdev(all errors) values noted on the top left of Figure 3 show the overall statistics (total error, independent of altitude).

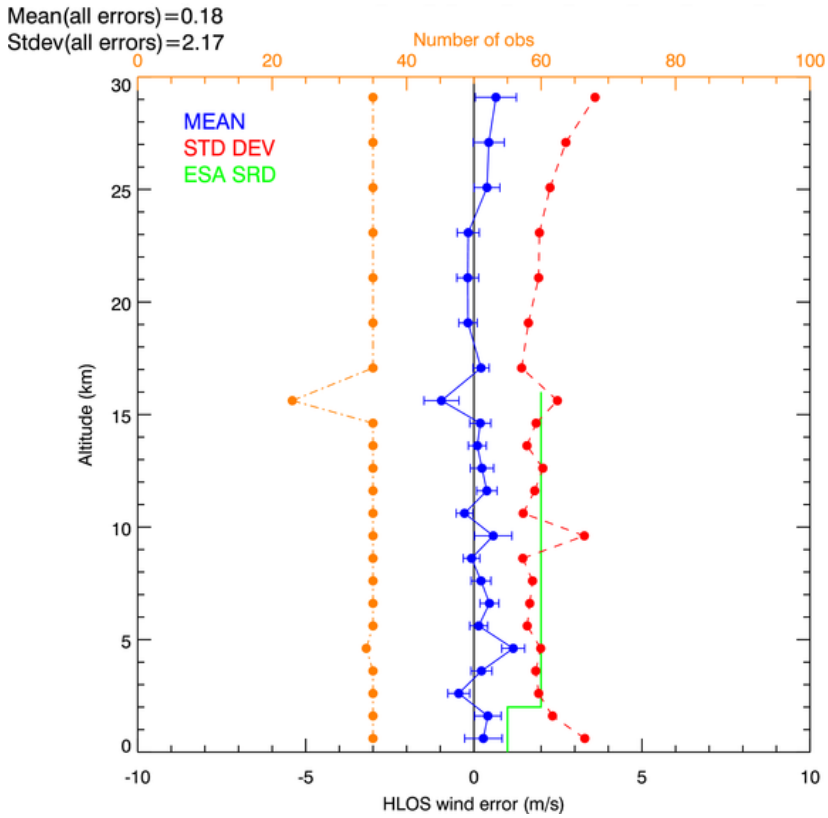


Figure 3. Statistics of the HLOS wind errors as a function of altitude (referenced to geoid) for the Rayleigh-clear L2B wind results for test case #0027 (repeated 10 times).

The L2B Rayleigh-clear wind results of Figure 3 appear to be of good quality, with standard deviation generally similar to the SRD threshold and with reasonably small biases. There appears to be a tendency for bias in the Rayleigh-clear results when clouds are present (e.g. at 15 km), implying that

[3] The same format of plotting error statistics is used in other test cases throughout the technical note.
[4] The requirements are defined for an idealistic atmosphere, which is horizontally homogeneous over the 100 km integration length.

the scattering ratio is erroneous or there is an issue in the cross-talk correction. Also, notice that the standard deviation is increased (at 10 km and 15 km) and number of Rayleigh-clear observations is reduced at 15 km due to many measurements being classified as cloudy, and hence used in the Rayleigh-cloudy observations (with much reduced measurements forming an observation, the noise increases). There is also a hint of a change in bias (to positive bias) for Rayleigh range-bins above the highest Mie range-bin (i.e. above 23 km in this case), which again perhaps indicates some sensitivity to the scattering ratio estimate (for the Rayleigh range bins above the highest Mie range bin, the scattering ratio defaults to 1.0)

It is important to note that quality control (QC) of the HLOS winds is not applied when producing the statistics in Figure 3; nor throughout the report (unless stated otherwise). That is, none of the L2B wind results are rejected from the statistics calculation, other than those with missing values (e.g. due to gross errors in the processing) in the L2B output.

The L2B Mie-cloudy results are shown in Figure 4. The Mie wind results are of good quality at around the 5, 10 and 15 km levels, where the Mie backscatter is strong due to the presence of the artificial clouds. However, a negative bias is apparent (-1 to 0 m/s) for the binned data just below these levels. The near-surface Mie results are clearly gross errors, which is because of ground contamination of a few of the measurements (there are relatively few resultant observations), which should draw the 50 m/s winds towards 0 m/s (no miss-pointing) and hence produce the observed negative bias. The Ground Wind Detection flag of the L1B measurements the L2Bp has been shown to reject such measurements in earlier versions of the software, and a bugfix of the current version is discussed below in section 4.

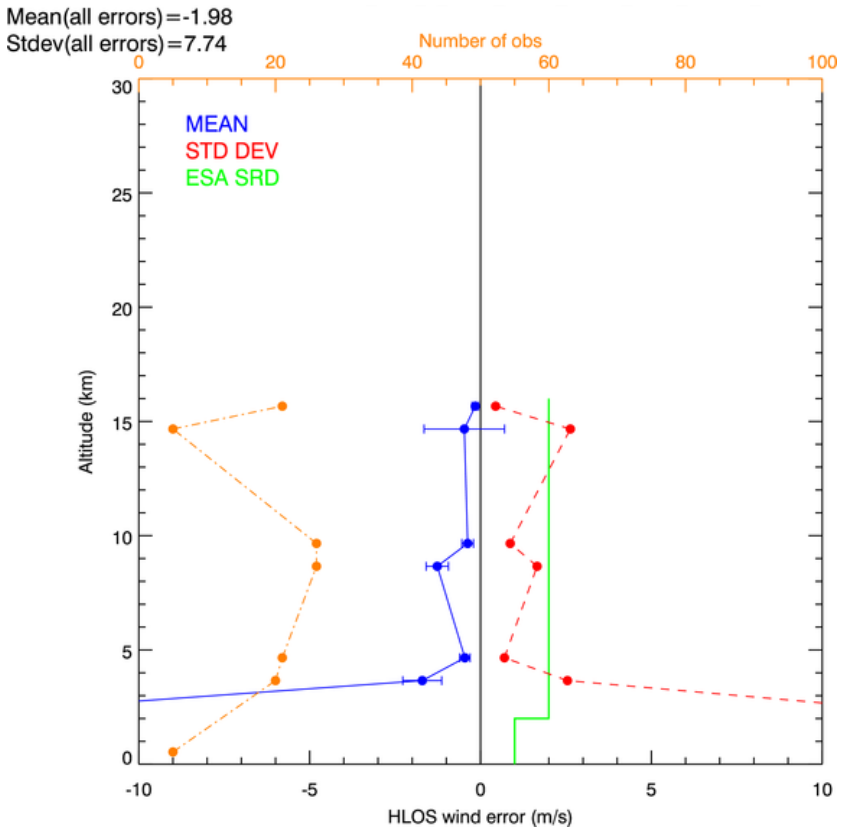


Figure 4. Statistics of the HLOS wind errors in Mie-cloudy L2B wind results.

Note that the negative bias cannot be related to vertical wind shear, since the wind is a constant 50 m/s for this test case. By analysing the wind results in detail, it was seen that winds contributing to the negative bias are retrieved from low signal levels i.e. relatively few measurements contributing to the observation and hence large predicted observation errors. The tendency to negative bias may be explained by gross errors which can be produced by the Mie core algorithm for low signal-to-noise (note that the L2B estimated SNR is not being used properly in the L2Bp at present, see section 4 about the L2B processor treatment of Mie SNR). By applying QC based on the Mie observation error estimate (i.e. rejecting observations with estimated errors greater than 0.8 m/s (this is very low because the error estimate values are underestimated, see section 2.3.1) then the bias was removed, since the low SNR biased observations are rejected. Note that very few observations were rejected for the levels which did not have bias using this QC.

The conclusion for this simple test case is that the Rayleigh CM L2B wind results are good in terms of noise and bias given sufficient signal strength (i.e. sufficient backscatter) and avoiding ground contaminated winds.

2.2 Increased realism: a tropical scenario with cirrus cloud (LITE)

This section investigates a more realistic atmospheric scenario. The test case is 8 BRCs long and is listed as test number #0103 in [RD4]. Figure 5 shows an altitude versus along-track distance cross-sectional view of the E2S input SR, which is derived from LITE data (see [RD9] for details of the method). The SR is clearly more realistic than the previous test case, being heterogeneous in the horizontal and vertical directions. The LITE data is sampled with ~3 km horizontal resolution and 125 m vertical resolution. A notable feature is the strongly scattering tropical cirrus cloud at 15 km altitude for the first half of the scene. Some particulate backscatter can be seen below 8 km, although not very highly scattering.

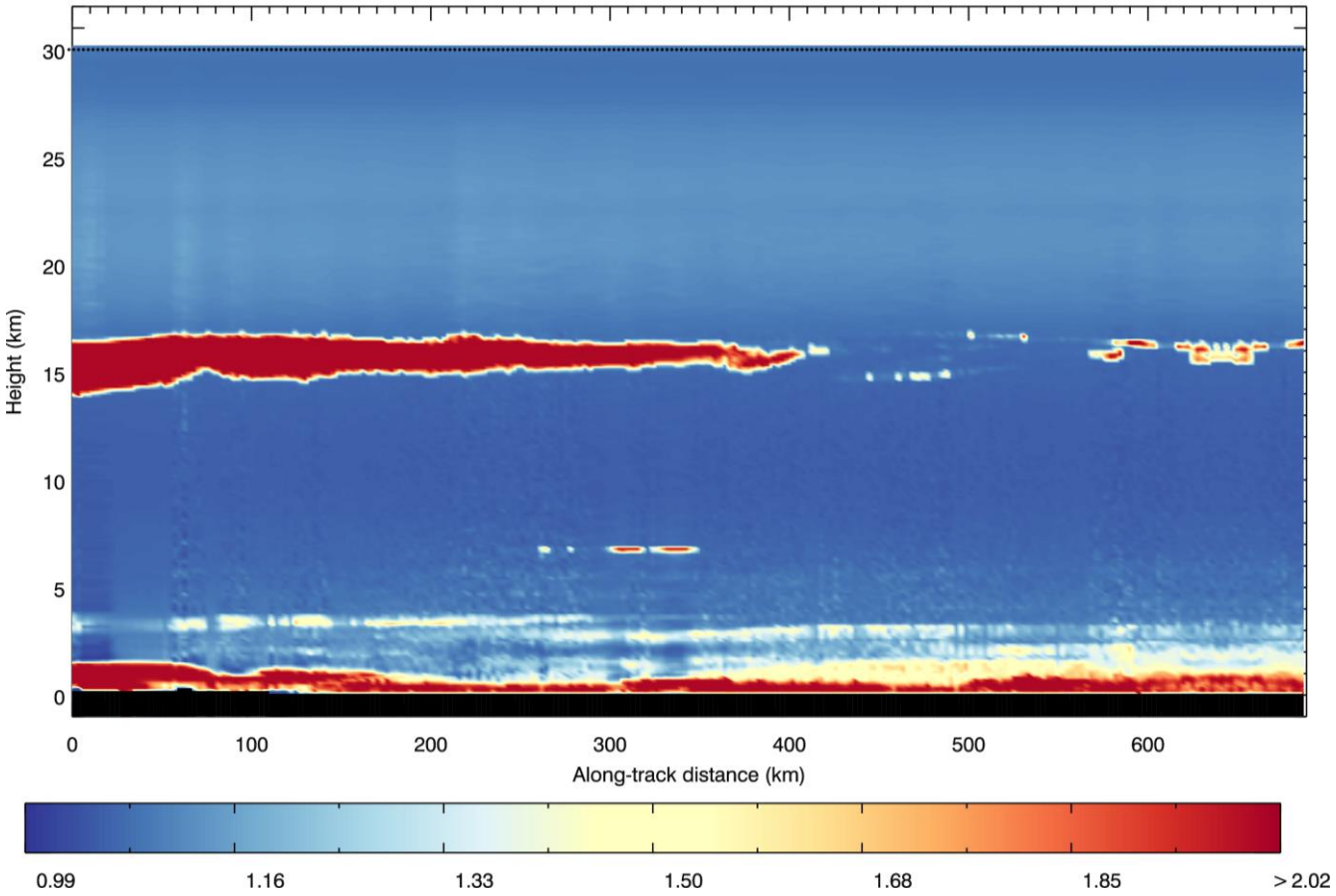


Figure 5. A cross-sectional view of the E2S input scattering ratio ("true" SR), as derived from the LITE mission for test case #0103. The horizontal extent covers the 8 BRCs of the test case.

The E2S orbital ground-track has no correspondence with the LITE orbital ground track (and co-located ECMWF data), therefore the terrain of the atmospheric inputs does not match that of the digital elevation model (DEM) used by the E2S (and hence reported in the L2B results). To overcome this mismatch, the input atmospheric profiles (KNMI atmospheric database) are always extrapolated to -1 km to ensure the profiles are always available for the E2S, despite LITE/model orography. This extrapolation causes the vertical stripes as visible in Figure 5 below 0 km.

The E2S input wind, temperature and pressure are also horizontally and vertically heterogeneous; ECMWF re-analysis model (ERA-40, 120 km, and 60 levels) fields are interpolated to be co-located with the much higher resolution LITE data. The E2S input HLOS winds are shown in Figure 6. The ERA-40 winds appear to be very smooth in the along-track dimension (and also in the vertical according to [RD9], since it significantly underestimates vertical wind shear). In the future, it would be interesting to use the latest ECMWF model setup (e.g. 10 km, 137 levels) as input to some test cases to see how the increased meteorological variability affects the results. This would not slow the E2S processing since, as mentioned above, the E2S inputs are already over-sampled to a 3 km horizontal resolution, despite the underlying meteorological fields being of much poorer resolution. Note that higher resolution model fields are used in the CALIPSO test cases, shown in section 2.3. However only on a 25 km grid (bear in mind that the effective resolution will be approximately 6 times this, see [RD5]).

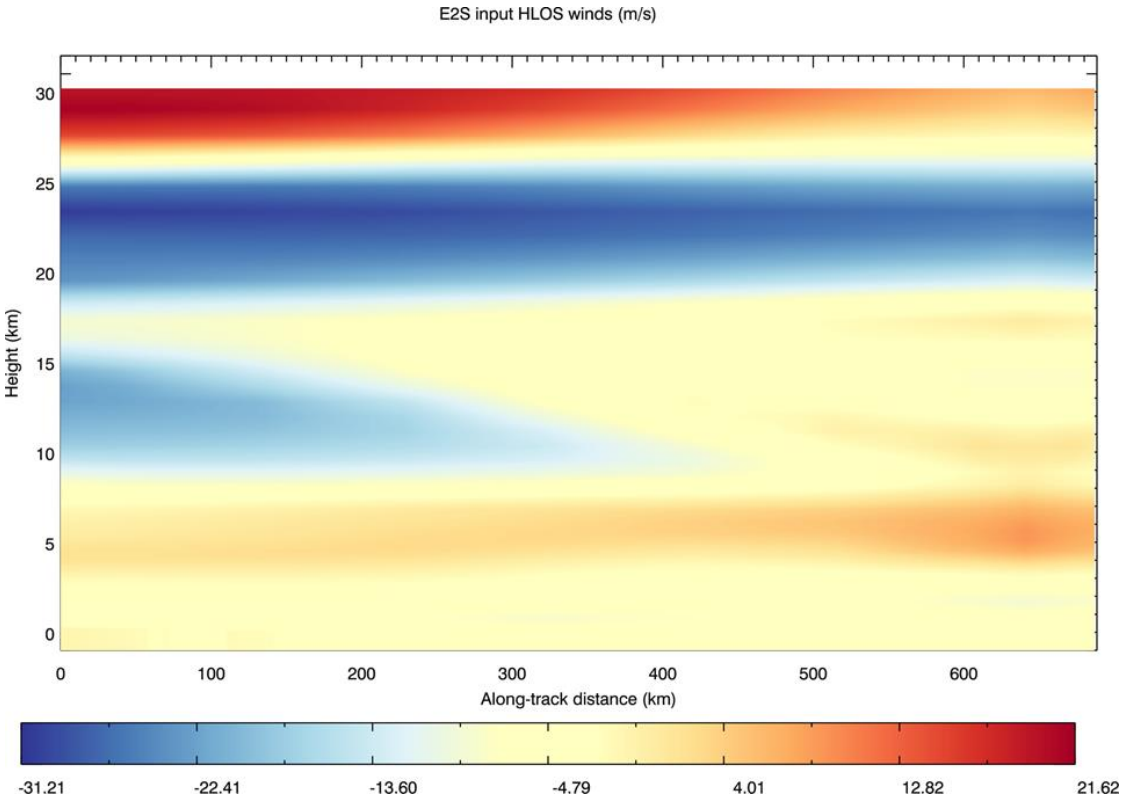


Figure 6. A cross-sectional view of the E2S input HLOS wind ("true" HLOS wind), units: m/s, from the ERA-40 model for test case #0103. The colour bar indicates the value.

The L2B processor is set to "advanced" measurement grouping with a 100 km maximum accumulation length (for both Mie and Rayleigh channels); this leads to 7 observations per range-bin level (given the 8 BRCs, each being ~86 km) with the last group being shorter than 100 km (expected behaviour of the grouping algorithm, see [RD7]).

The verification of L2B HLOS results is made against E2S input winds ("truth") that are closest to the centre-of-gravity (COG) of the L2B observations. It was found that changing the definition of the true wind to be an average of the input winds over a rectangle representative of the observation dimensions gave only negligible changes in the statistics for this test case — this is because the interpolated ERA-40 model HLOS winds have little variance on the 100 km horizontal scale (the chosen observation size) and less than realistic wind variance in the vertical dimension.

Figure 7 shows an altitude against along-track distance cross-sectional view of the true HLOS winds used for verification of the L2B results (i.e. E2S inputs), but on the grid of the L2B observations: they are displayed with a box indicating the horizontal and vertical extent of the measurements[5] that went into the observation (this can be compared to the raw input wind cross-section in Figure 6). White areas amongst the coloured rectangles (observations) indicate an absence of observations (e.g. because of particles affecting the classification).

[5]Most of the time the measurements that go into the retrieval of a HLOS wind observation are a contiguous group of measurements; however occasionally (due to partially cloudy scenes) it is possible to have e.g. only 2 widely separated measurements forming an observation. The plotting of Figure 7 can give the impression of a contiguous block in such cases, since the rectangles are based on the first and last measurement of the group.

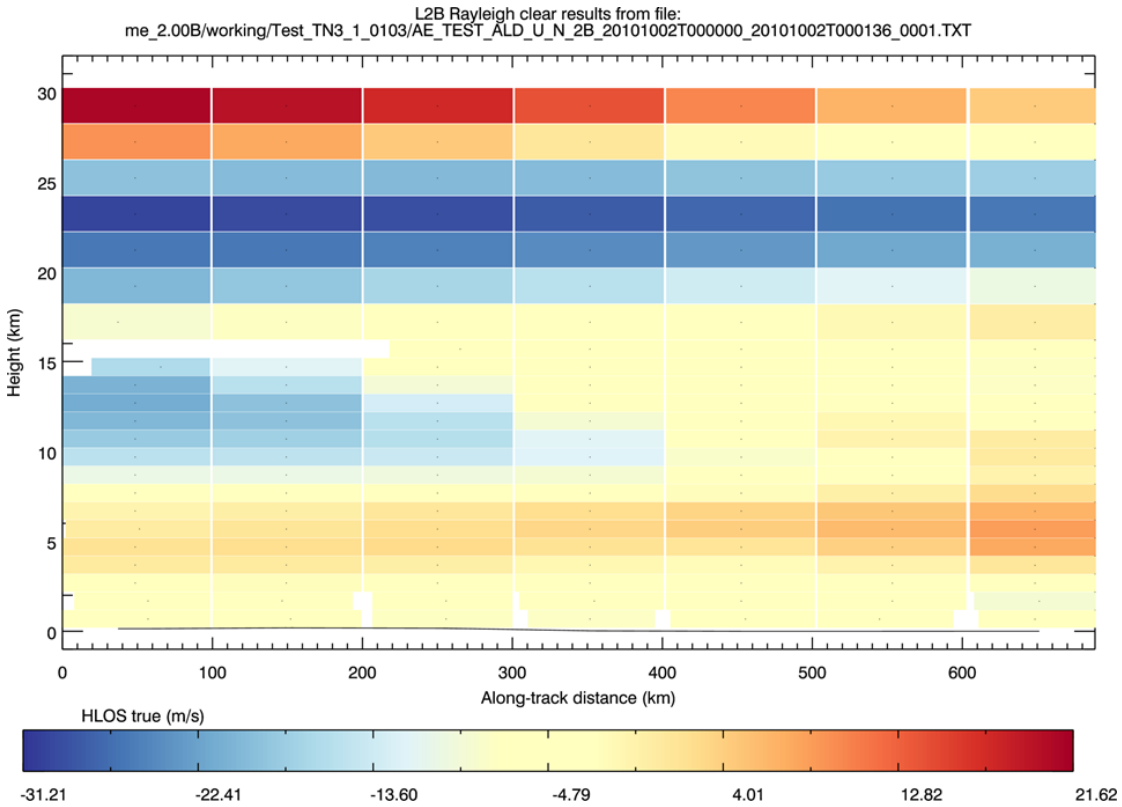


Figure 7. A cross-sectional view of the "true" HLOS winds at the resolution of L2B Rayleigh-clear wind results which are displayed in Figure 8. The colour bar indicates the value.

The L2B Rayleigh-clear HLOS observations are shown in Figure 8, with the observation "resolution box". The HLOS error statistics are constructed from differences between the winds of Figure 7 and Figure 8 according to the method described in the Appendix, section 6.1.

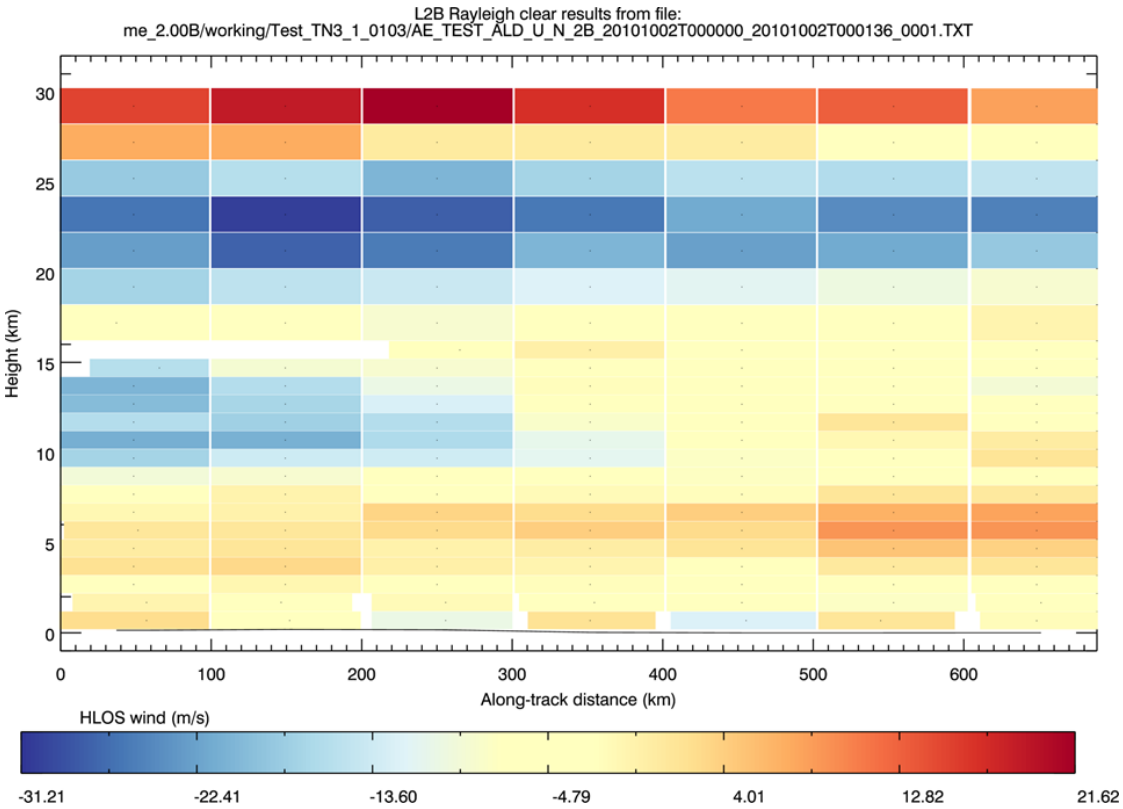


Figure 8. A cross-sectional view of the L2B Rayleigh-clear HLOS wind results from test case #0103. The noise of Aeolus winds is apparent when compared to the truth on the same grid in Figure 7. The colour bar indicates the value.

The L2B observations of Figure 8 appear to capture the greater than 100 km scale structure of the true HLOS wind field very well; however there is clearly some noise in the L2B results, as is expected. The continuous nature of the winds with CM helps to form a contiguous picture of the wind-component field as compared to BM where only 25% of the cross-section would be sampled (albeit with increased accuracy).

Figure 9 shows the Rayleigh-clear HLOS error statistics for this sample of 7 observations per range-bin. Judging by the large confidence intervals on the mean errors, the sample size is not large enough to be statistically reliable, so in the next section a much larger sample size is produced to investigate the test case.

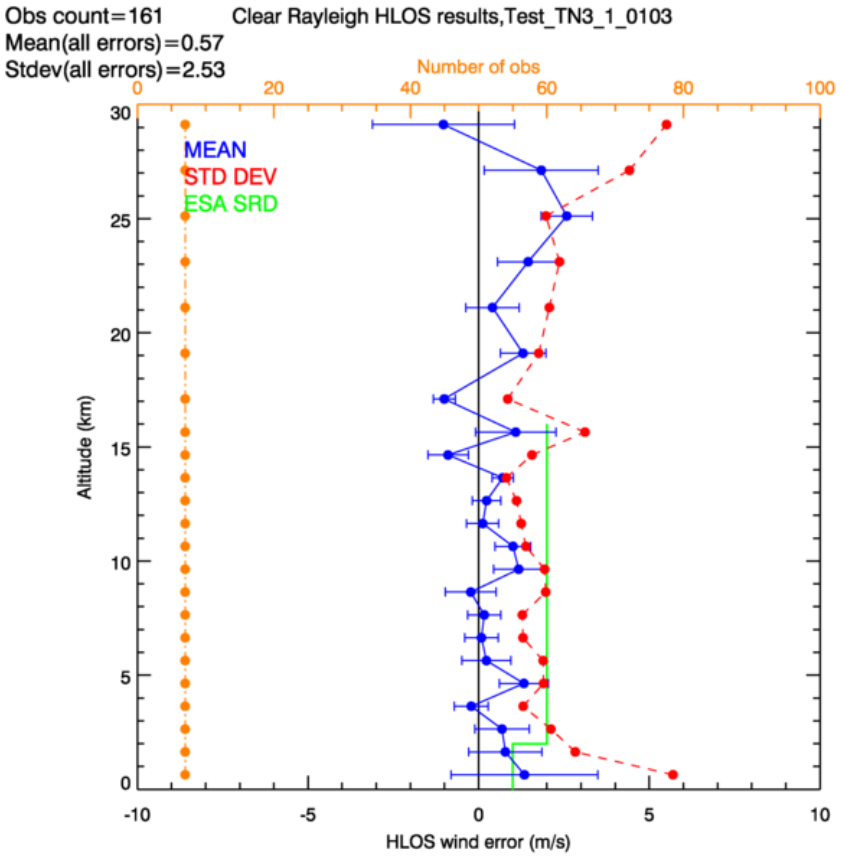


Figure 9. The statistics of the L2B Rayleigh-clear HLOS error for the one sample of L2B winds from test case #0103.

2.2.1 Increasing the sample size

To increase the reliability of the statistics, the same test case (#0103) was stochastically repeated 32 times (i.e. 32 runs of the processing chain, with independent noise in each run, taking about 42 hours run time; mostly due to the E2Sv3.01 (before performance improvements)). Figure 10 shows the equivalent statistics to Figure 9, but with the much larger sample size. As was the aim, the estimate of the mean errors in Figure 10 has much smaller confidence intervals. This indicates that the positive bias in HLOS error seen for most range-bins, of typically 0.5-1 m/s, is statistically significant. The standard deviation of HLOS error is typically just less than 2 m/s for the range 4 - 15 km, therefore meeting ESA's requirement, despite this test case being a variable atmosphere (albeit with less than realistic atmospheric variability).

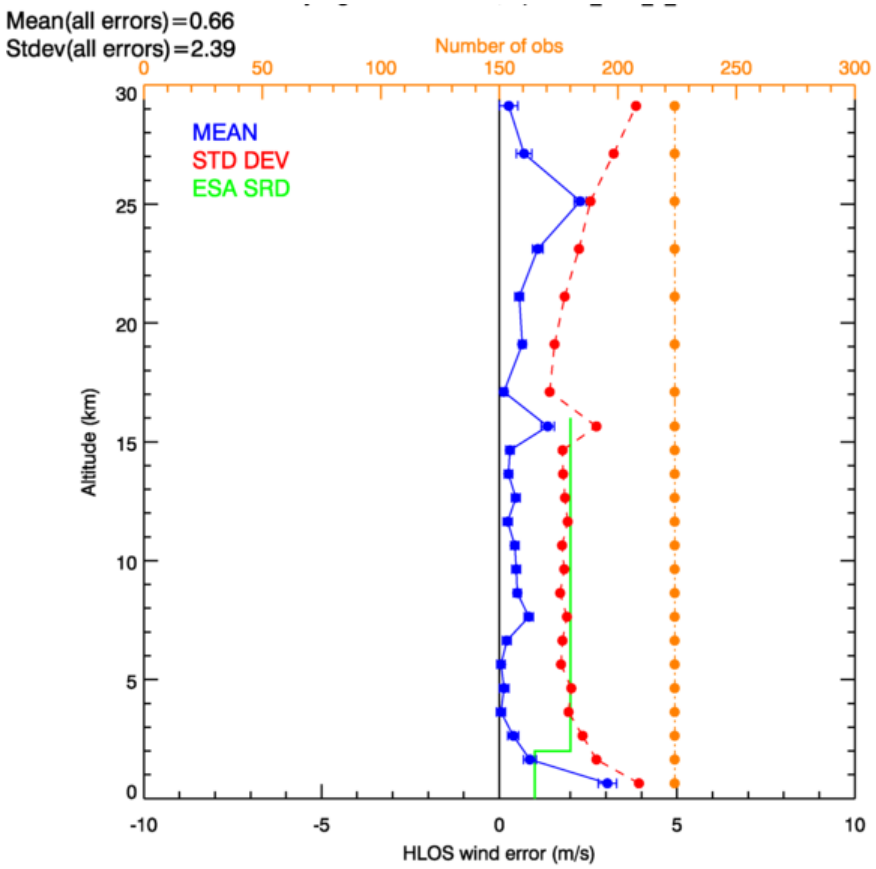


Figure 10. The statistics of the Rayleigh-clear L2B HLOS wind errors from the accumulated results of 32 independent samples of test case #0103. The number of winds used at each range-bin is shown in orange. The L2B group size is restricted to 100 km.

The positive bias in HLOS error was investigated. Initial tests indicated (not in the verification shown above) that positive bias was found to be correlated with vertical wind shear. This is because the vertical placement of a Rayleigh wind, with its skewed vertical weighting function, can create an apparent bias when verified against the high-resolution E2S input winds, as is done for verification in this report. This effect is more pronounced for thicker vertical range-bins (i.e. higher altitude range-bins) and for larger vertical wind shear across the range-bin.

However, the correlation between vertical wind shear and HLOS error can be minimised by changing the vertical placement of the Rayleigh wind to be a fraction of 0.45 within a range-bin (set by `Rayleigh_Height_Weight_Upper` in the AUX_PAR_2B file). This optimized setting was used for the results of Figure 10; for which the correlation is indeed very small, 0.013, as shown in Figure 11. Therefore the positive bias that remains, as in Figure 10, is not thought to be due to the vertical wind shear.

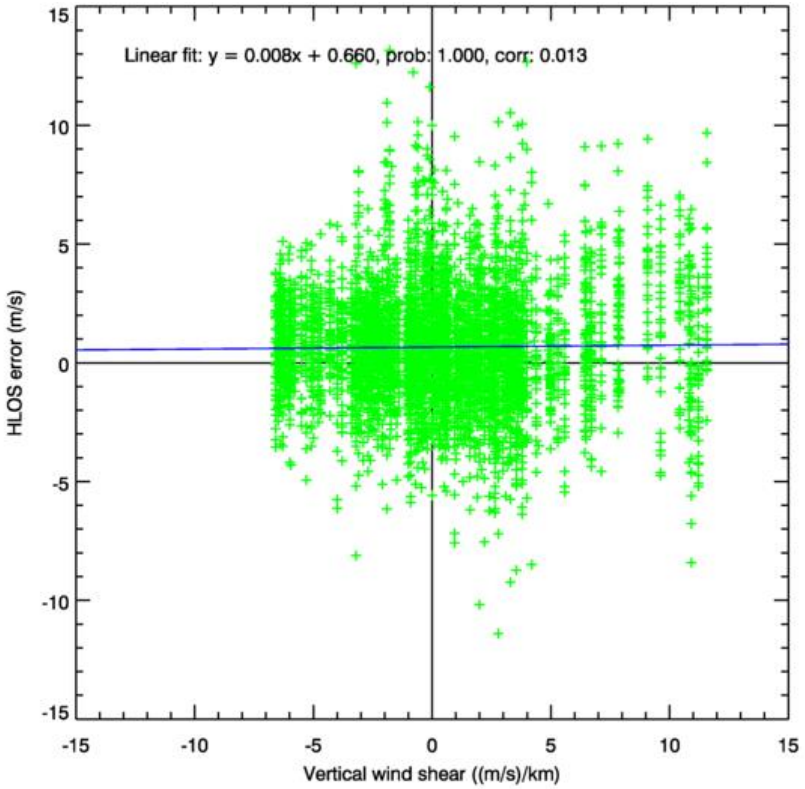


Figure 11. A scatter plot of HLOS error vs. vertical wind shear (individual results shown by green crosses) for the L2B Rayleigh-clear results. The line of best linear fit to the data is shown in blue, indicating negligible correlation (which agrees with the very small correlation value of 0.013).

The 0.45 vertical placement is a little lower than a theoretical estimate, which was determined to be around 0.47-0.49 (see the Appendix, 6.7) for 2 km range-bins assuming a linear vertical wind shear, so perhaps this tuned vertical placement is compensating for some other errors.

After discounting vertical wind shear as the cause of the positive bias, it was discovered that the Rayleigh wind bias has some dependency of HLOS value, becoming more positively biased for negative HLOS values. The HLOS dependence of the bias is shown in Figure 12. This particular test case has a relatively larger frequency of negative HLOS winds compared to positive HLOS winds, which tends to produce an overall positive bias as seen in Figure 10.

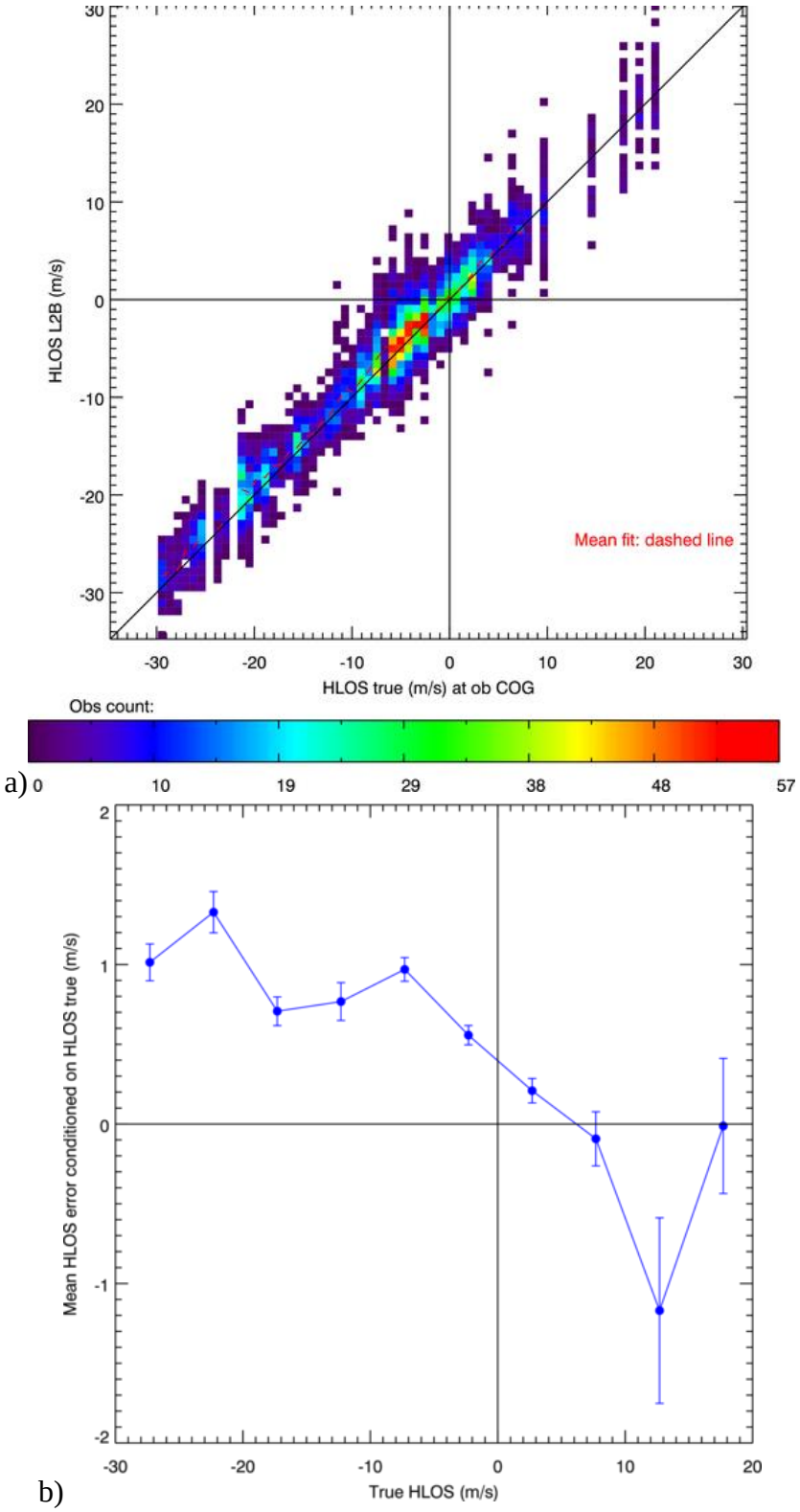


Figure 12. The statistics of the L2B Rayleigh-clear HLOS errors for test case #0103 (repeated 32 times). a) Density plot of all HLOS error vs. true HLOS wind. The red-dashed line shows the mean of the y given x. b) Mean HLOS wind error conditioned on the true HLOS wind.

The HLOS dependent bias appears to be also common to all the other realistic test cases that have been run (this will be shown in other test cases later in the report). Note that the winds were not particularly biased for the simple test scenario shown earlier in Figure 3 perhaps by coincidence, in that the wind was a constant 50 m/s, a region of the slope where bias is smaller. The source of the bias will be analysed further in section 2.3.

It was considered whether the AUX_RBC file or L2B processing using the file was leading to the wind-dependent bias. The AUX_RBC file used in the processing was generated by the L2B team, as explained at the start of section 2. One way to confirm this would be to use an ideal AUX_RBC file generated by the L1B team using independent code. Such a file became available, and the slope bias was very similar using the independent file, therefore the AUX_RBC file is not to blame. It was noticed that there is almost no correlation between HLOS error and the E2S input temperature. The other parts of the processing chain (E2S→L1B→L2B) cannot yet be ruled out as being the source of the bias.

It is of concern that bias is present with supposedly ideal calibration files. With knowledge of the truth (as in these simulations) it would clearly be easy to correct the bias, however for the real Aeolus data the truth will of course be unavailable.

The Mie-cloudy results (from the larger sample) shown in Figure 13 are reasonable in quality for the 15.5 km range-bin, with standard deviation at just less than 2 m/s and zero bias. Most Mie results are in this 15.5 km range-bin (~110 results), due to the presence of the strongly scattering cirrus cloud covering half of the scene at that level. There are also some results for the adjacent range-bin below (~30 results), but this shows a tendency towards negative bias which is similar behaviour to the simpler test case (see Figure 4), which was shown to be due to lower SNR.

The Mie-cloudy results near the surface are biased by about +4 m/s, this is because there is ground contamination of some of the measurements (the true wind is ~-4m/s and the ground velocity is 0 m/s since there is no mis-pointing). There are no Mie-cloudy results from just above the surface to 5 km altitude despite Figure 5 showing that there is particulate scattering (with SR ~1.5, probably aerosol) in this height-range. The SR values of 1.0-2.0 of these areas are lower than the classification threshold used, hence the measurements are not available for the Mie-cloudy winds observations. However, even for the Mie-clear winds (which will try to form retrieve Mie winds from the measurements classified as clear) there are no wind results in this area. This behaviour was explained after the main investigation, see section 4.

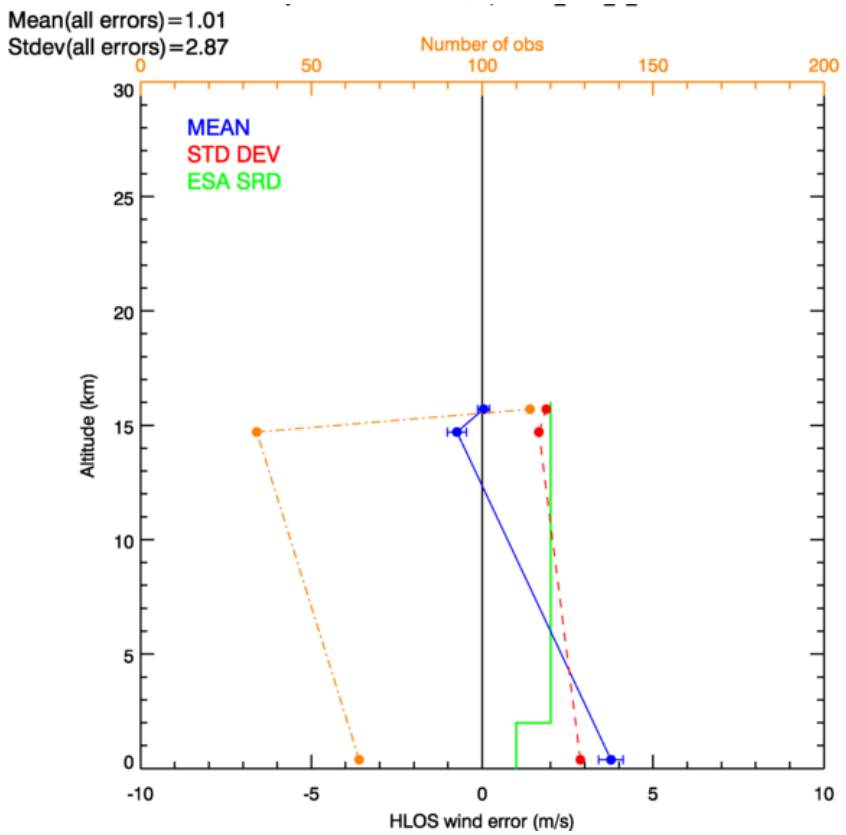


Figure 13. The statistics of the Mie-cloudy L2B HLOS wind errors from 32 independent samples of test case #0103.

To gain a better understanding of the Mie-cloudy results of Figure 13, Figure 14 shows a cross-sectional view of the first six 8-BRC samples (from the total of 32 samples). Each sample is about 700 km long. It can be seen that the near-surface Mie wind observations (in close proximity to the DEM model line) are fairly short in horizontal extent. It would seem that some measurements that are contaminated with ground returns have been processed to L2B winds. This must be because the L1B flagging of ground wind contaminated measurements is not working perfectly (the L2B processing will discard the flagged measurements). The reason for this software bug and the bugfix was discovered after this test was run and analysed, as explained in section 4 below.

It can also be seen from Figure 14 that the number of Mie-cloudy wind observations per 8-BRC sample on the 15.5 km range-bin fluctuates from 3 to 4, due to the L2Bp classification into clear/cloudy measurements being based on noisy L1B refined scattering ratio (this noisiness is expected). The Mie results at ~14 km are also formed from relatively few measurements, perhaps accounting for the tendency towards poorer quality (negative bias) — this may be similar to the simple test case of section 2.1. There is no evidence that wind shear is an issue.

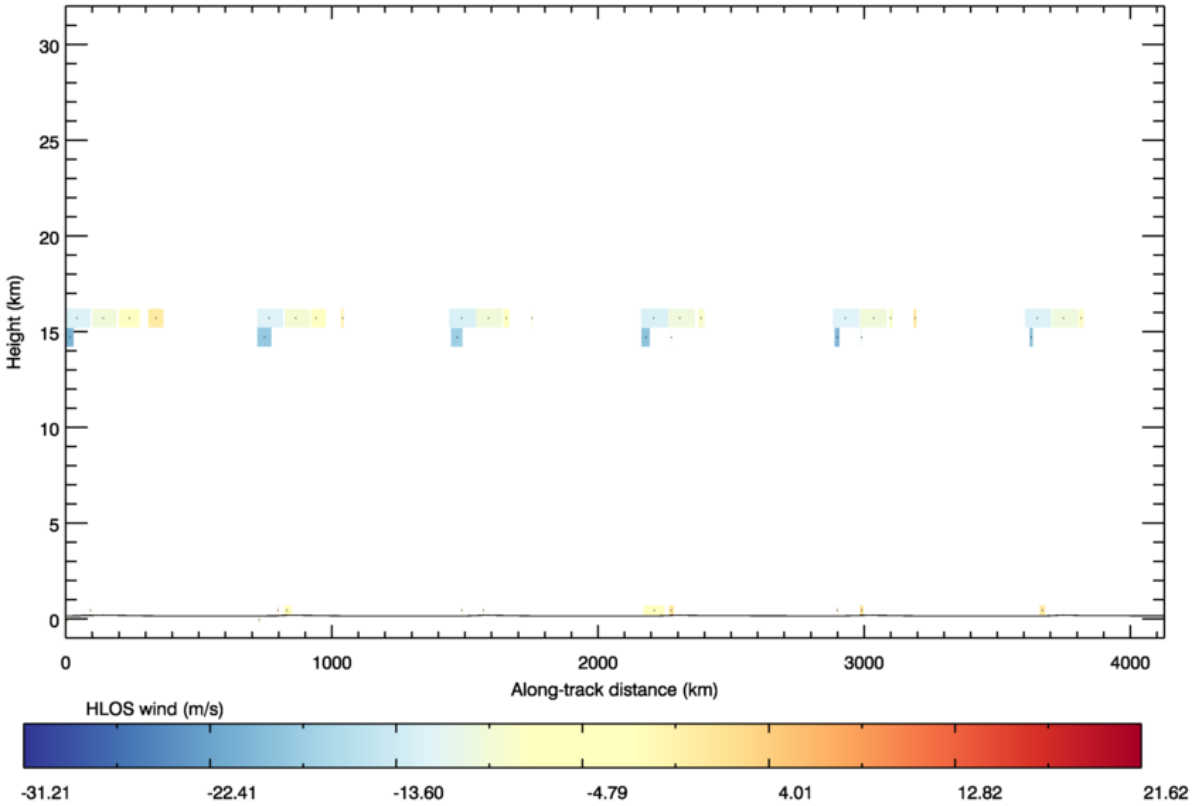


Figure 14. A cross-sectional view of the L2B Mie-cloudy HLOS winds from the first six samples of test case #0103. The DEM along the track is shown by black horizontal line (i.e. close to 0 km).

The other wind classification types that the L2B processor produces (i.e. Mie-clear and Rayleigh-cloudy) have not been shown, since the results are of much poorer quality and as such are unlikely to be used in NWP data assimilation, and are therefore of less interest (although they have been investigated in detail).

2.2.2 Changing the grouping algorithm accumulation length

This section investigates what happens as the L2Bp grouping algorithm accumulation length is altered. With the grouping algorithm maximum accumulation length set to 50 km (rather than 100 km, as presented earlier), the Rayleigh-clear statistics shown in Figure 15 are produced, which can be compared to Figure 10.

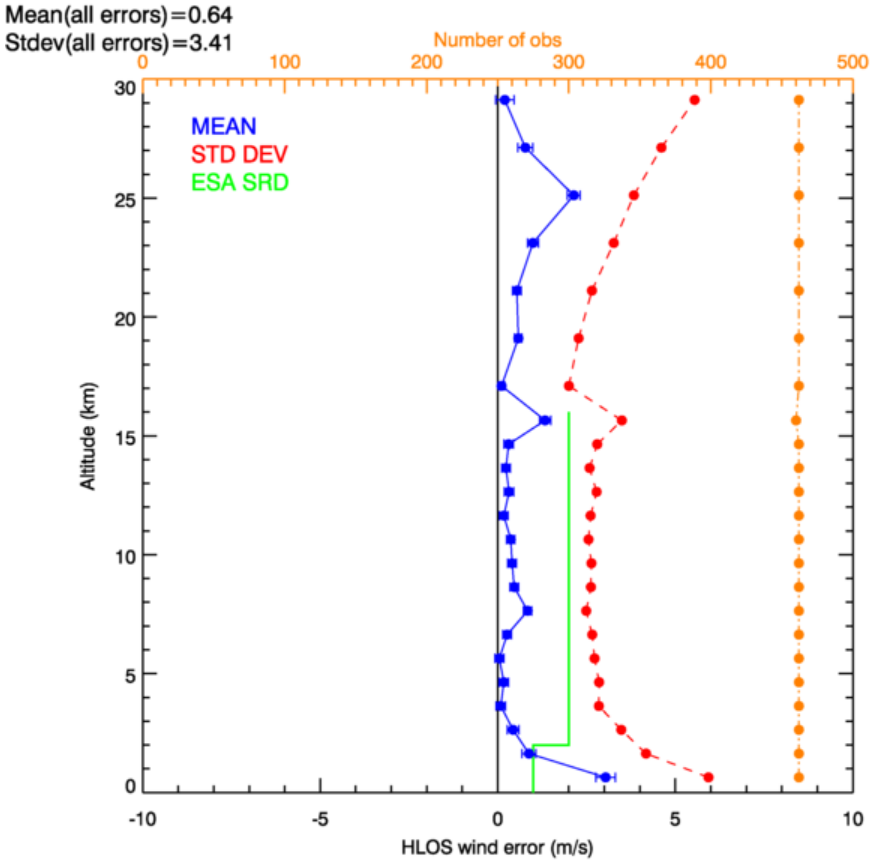


Figure 15. The statistics of the Rayleigh-clear L2B HLOS wind errors from 32 independent samples of test case #0103. The number of winds used at each range-bin is shown in orange. The group size is restricted to 50 km.

It should be noted that the number of observations per level is doubled as expected. Due to shorter averaging the random error the standard deviation increases as expected. However the bias is almost identical, indicating that it is not a function of the number of measurements that form an observation (something inherent to each measurement).

Assuming that the standard deviation of HLOS observation error has the form (as a result of averaging) as shown in equation 2 below (see section 6.4 of the Appendix for an explanation):

$$\sigma_{HLOS} = \frac{B}{\sqrt{d}} \qquad \textbf{(eq. 2)}$$

where d is the observation horizontal accumulation distance and B is a constant. With d=100 km and σ_{HLOS}=1.8 m/s at 10 km altitude (see Figure 10), then a value of B=18.0 (units not relevant) is obtained. So if the grouping changes to d=50 km then σ_{HLOS}=2.6 m/s is the expected standard deviation. This calculated value is very close to the value observed in Figure 15, providing confidence that the grouping algorithm is working correctly i.e. there are not adding any unexpected sources of error.

The Mie-cloudy results with 50 km grouping accumulation length are shown in Figure 16. The reduction in group size to 50 km increases the standard deviation compared to the 100 km grouping (shown in Figure 13), however the relative increase in standard deviation is smaller than for the Rayleigh clear results, so not in line with the $1/\sqrt{d}$ model. Perhaps this is because the actual Mie horizontal accumulation lengths are generally not changed from 100 to 50 km, as is possible with contiguous Rayleigh signals in the horizontal, but to a distribution of averaging lengths due to the sporadic spatial distribution of the particulate backscatter. The biases at ~14.5 km and near the surface

remain very similar as with 100 km grouping, perhaps suggesting they are consistently present in the measurement level data.

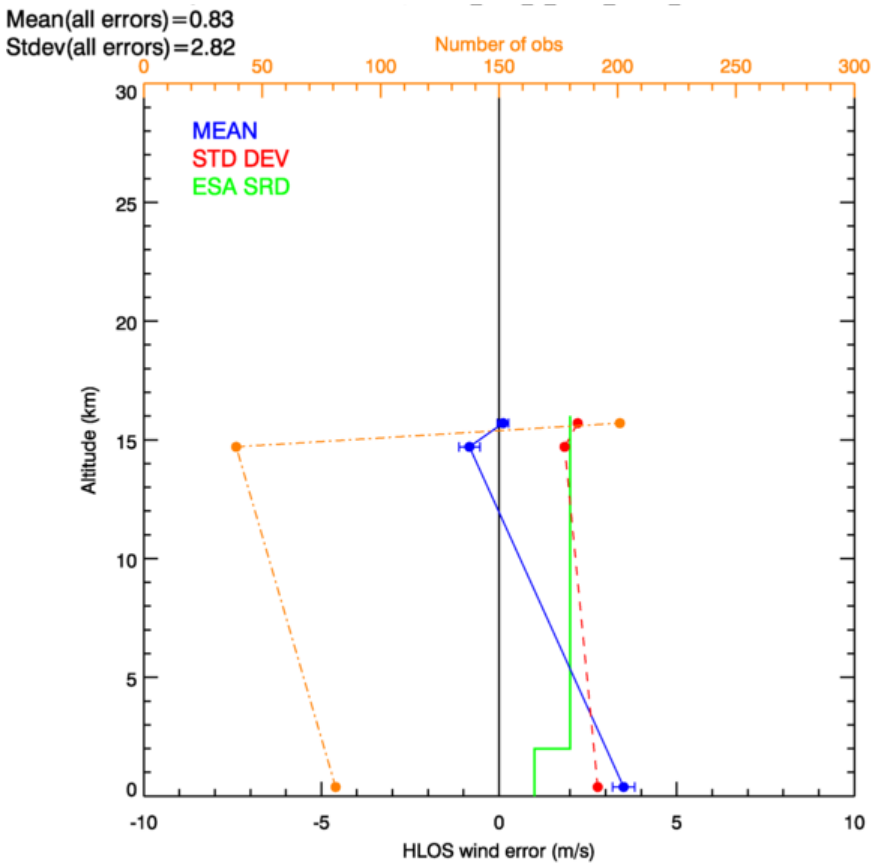


Figure 16. The statistics of the Mie-cloudy L2B HLOS wind errors from 32 independent samples of test case #0103. The number of winds used at each range-bin is shown in orange. Group size is restricted to 50 km.

2.2.3 Estimating the horizontal observation error correlation

With the repetition of the same test scenario #0103 (section 2.2.1), with randomly generated errors in each sample, it is possible to estimate the error correlation between observations. The horizontal error correlation is examined in this section.

The correlation is horizontal in the sense that the error correlation between observations on the same range-bin is estimated, but at varying horizontal distances (along the orbital track) from one another. The method is explained in the Appendix, see section 6.2. It was found for test case #0103 that 32 samples are enough for a stable error covariance matrix, and hence stable error correlation matrix.

Covariance (and hence correlation) matrices have only been produced for the Rayleigh-clear results. Unfortunately, it was not possible for the Mie-cloudy results due to the number of Mie observations varying from sample to sample due to the effect of noise producing an "on-off" nature to the wind observations — it was not clear how to deal with missing values in the matrix calculations. The varying observation numbers is illustrated for the Mie-cloudy wind was shown in Figure 14.

Figure 17 plots the L2B Rayleigh-clear HLOS horizontal error correlations at four arbitrarily selected range-bins (at 7, 10, 15 and 24 km).

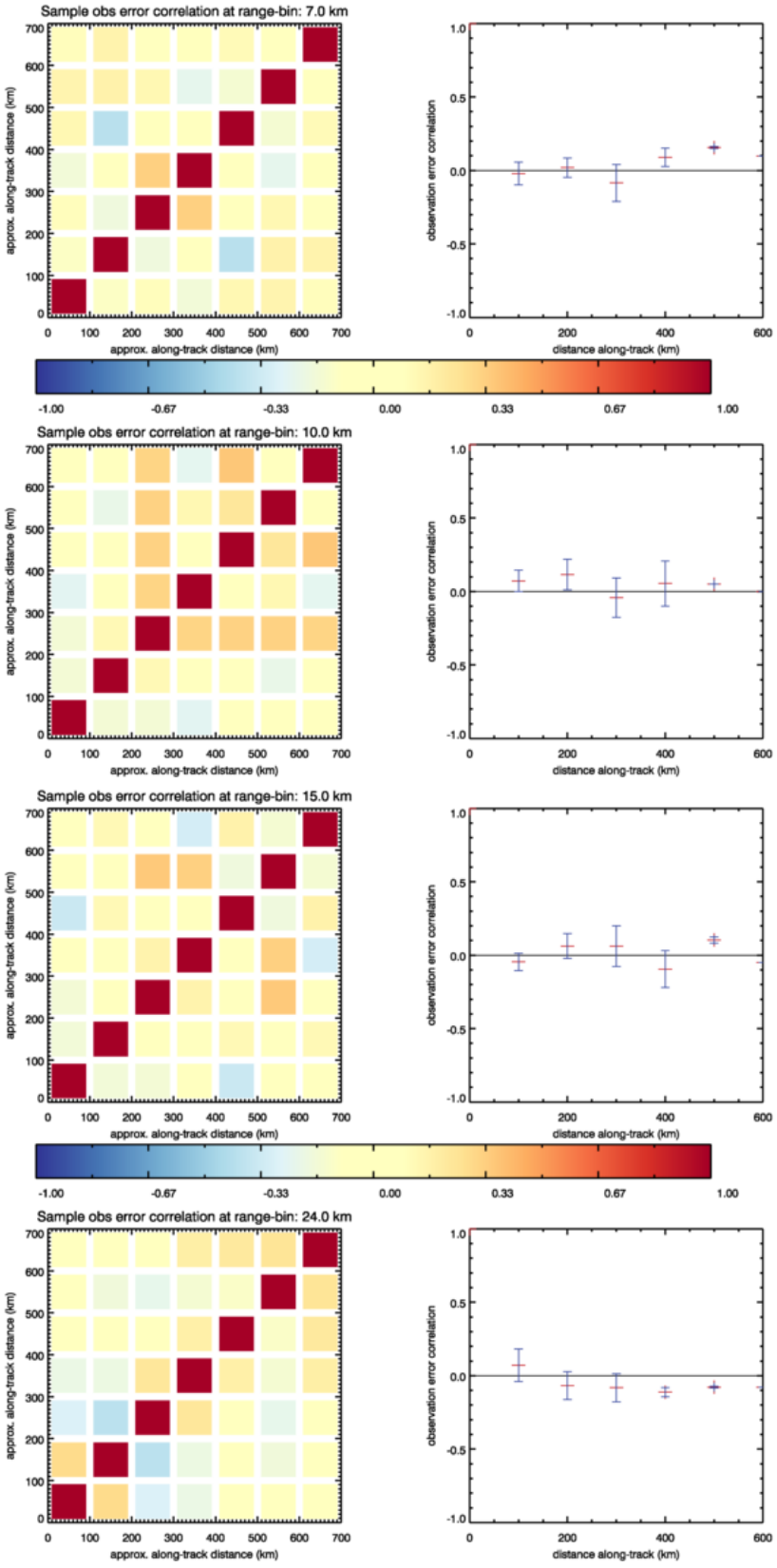

Sample obs error correlation at range-bin: 7.0 km
approx. along-track distance (km)
observation error correlation
distance along-track (km)
Sample obs error correlation at range-bin: 10.0 km
Sample obs error correlation at range-bin: 15.0 km
Sample obs error correlation at range-bin: 24.0 km
-1.00
-0.67
-0.33
0.00
0.33
0.67
1.00

Figure 17. Left column: Rayleigh-clear HLOS (100 km horizontal averaging) error correlation matrix as a function of horizontal distance, for 4 range-bins (top to bottom: 7, 10, 15 and 24 km). Right column: average correlation vs. horizontal distance.

The horizontal grouping maximum accumulation length is set to 100 km. Each sample is 7 observations wide (in this 8 BRC test case), which determines the dimension of the correlation matrices (7 by 7 matrices). Generally it is seen in Figure 17 that the off-diagonal elements of the error correlation matrices are close to zero (left-hand side plots, colour indicating correlation value via the colour bar). This is clearly shown by the right-hand side plots where the error correlation value versus horizontal distance between observations is plotted by averaging the correlation matrix elements over nearest neighbours, second nearest neighbours etc. So the conclusion is that **correlated error is negligible with 100 km grouped Rayleigh-clear observations, at least for this test case.** This small error correlation is good news, since correlated error is potentially very damaging for data assimilation in NWP, unless procedures to mitigate the effects are implemented (such as horizontal thinning of observations, or explicit representation of the correlations within the data assimilation R matrix) which have their costs.

Next, it was tested whether smaller L2B grouping would make any difference to the result. Rayleigh-clear 50 km grouping results are shown in Figure 18 (notice the increase to 14 by 14 matrices). Unfortunately the 15 km range-bin is not available due the varying number of observations per sample (i.e. some missing) with 50 km groups, due the effect of the cloud on the classification at that level. **It is seen that the error correlation for the 50 km grouped Rayleigh observations is also very small, as for the 100 km grouped observations**.

Prior to these results it was presumed that observation horizontal error correlations would exist, since horizontal error correlations in the measurement level data (~2.85 km horizontal resolution) which are averaged to produce observations could propagate to the observations, made worse by the lack of gaps in CM data, unlike BM. If this was the case, then **perhaps one would expect correlations between observations to increase as the observation horizontal resolution is decreased** via reducing the grouping length, **but this is not seen for the Rayleigh-clear winds in the test case presented**, implying the E2S has no mechanism of generating horizontal correlation of measurement error (which perhaps is correct) for Rayleigh-clear winds (at least with the miss-pointing biases off).

The mechanisms for producing measurement level error correlation are thought to be predominantly atmospheric scene-dependent i.e. coherent atmospheric structures across many measurements generating similar errors, which will tend to be correlated. Therefore perhaps the observation error correlation is more of an issue for the Mie winds, due to vertical placement of the wind issues e.g. not knowing where within the range-bin the particles are present, yet assigning the wind to the centre of the range-bin, creating error in wind shear scenarios. As mentioned, the error correlations of Mie winds have not yet been investigated.

Recently it has been pointed out by KNMI that with significant laser frequency variation (e.g. in case of significant sensitivity to microvibrations), vertically correlated errors for the Rayleigh signal below optically thin clouds with horizontally varying backscatter (such as cirrus) could become a problem (an effect not simulated in the version of E2S used in this study).

The errors are furthermore too optimistic in the sense that the E2S is not simulating an orbit-location dependent systematic error (e.g. residual bias after the harmonic bias correction on the BRC level) which may in effect produce correlated error between neighbouring observations. For example, if such a systematic error varies very quickly (e.g. randomly varying from one BRC to the next) along the orbit it would appear as increased random error, but if it varied very slowly along the orbit (e.g. effectively constant over say 10 BRCs - ~860 km) then it would appear as approximately as a constant

bias (which hopefully, but not necessarily, could be corrected for), however if it varied at an in-between rate, then it would in effect generate horizontally correlated observation error, which would be very difficult to detect. This should be simulated accurately in future work due to its potentially damaging effects in NWP. Also, the vertical correlation of observation errors should be investigated.

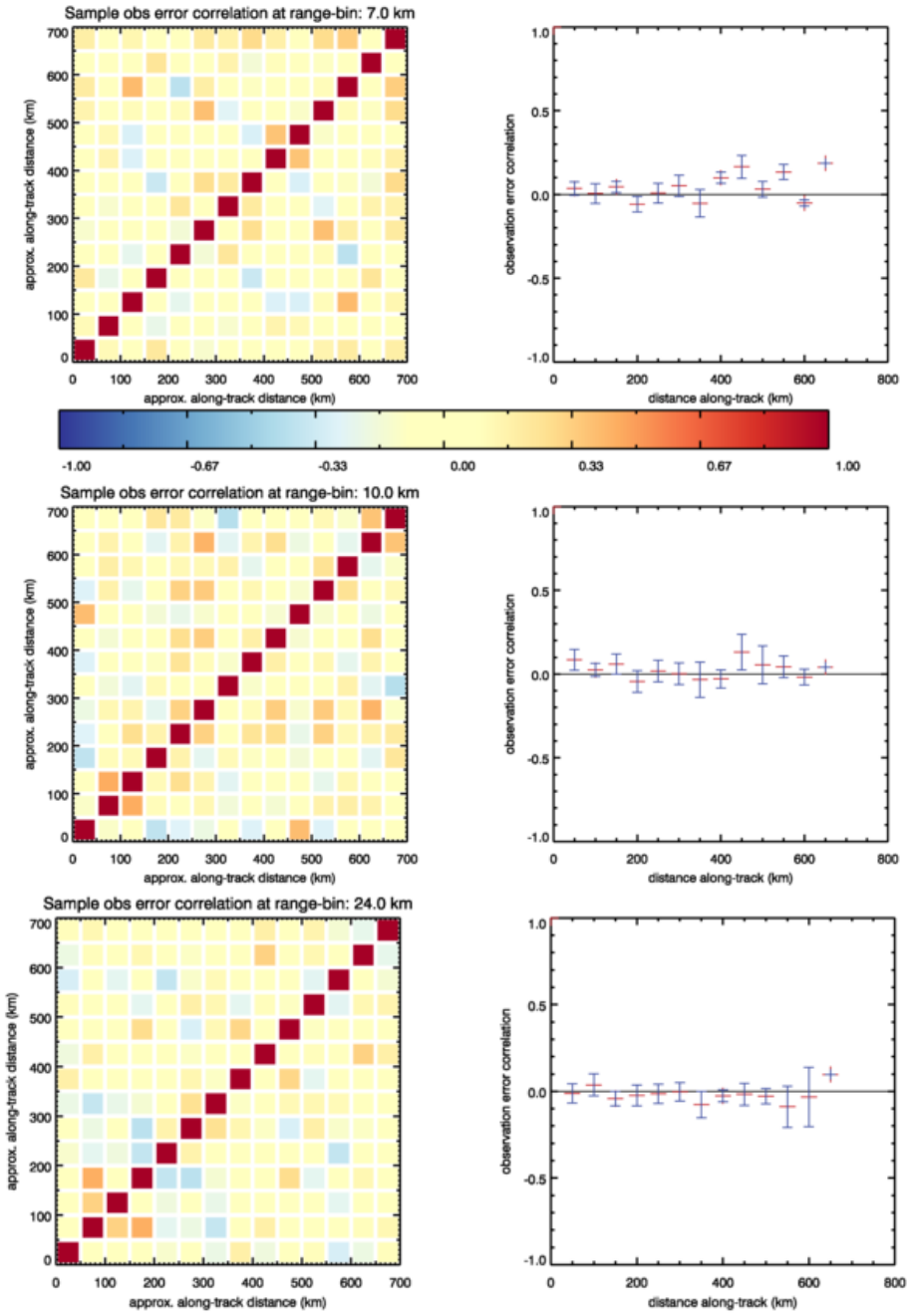


Figure 18. Left column: Rayleigh-clear HLOS (50 km horizontal averaging) error correlation matrix as a function of horizontal distance, for 3 range-bins (top to bottom: 7, 10 and 24 km). Right column: average correlation vs. horizontal distance.

2.3 Realistic half-orbit test scenario: CALIPSO

This section presents the most realistic test cases available in the KNMI database. The chosen test case is number #0204 as listed in TN3.1a [RD4], however with a modification: the random sources of noise are switched on (as done for previous test cases in this report). E2S particle backscatter inputs are retrieved from CALIPSO level-1 532 nm attenuated backscatter data, from a night-time half-orbit in 2007, see [RD9] (it is actual Orbit 1 of the 15 night-time orbits available in the KNMI database on 1/1/2007; this samples from northern latitudes across the equator in Africa). This test case uses higher resolution ECMWF model data (co-located with the CALIPSO data) than used in section 2.2 (which was co-located with LITE data) and should therefore be more realistic in terms of meteorological variability. The L2B processor was run with a 100 km maximum accumulation length for the grouping algorithm, and the L1B refined scattering ratio was used, for the same reasons explained in section 2.1. The aim of this test case was to see if the higher resolution meteorological data and large variability of scenes over a half-orbit made any difference to the results and to have large samples over a large variety of atmospheric scenes.

The range-bin definition for this test case, see Figure 19, differs from that of previous test cases, by having higher resolution Mie-bins closer to the ground (since this test case was originally designed for testing zero wind correction), this results in the top Mie bin being relatively low at around 16 km. Like in previous test cases the miss-pointing biases are switched off.

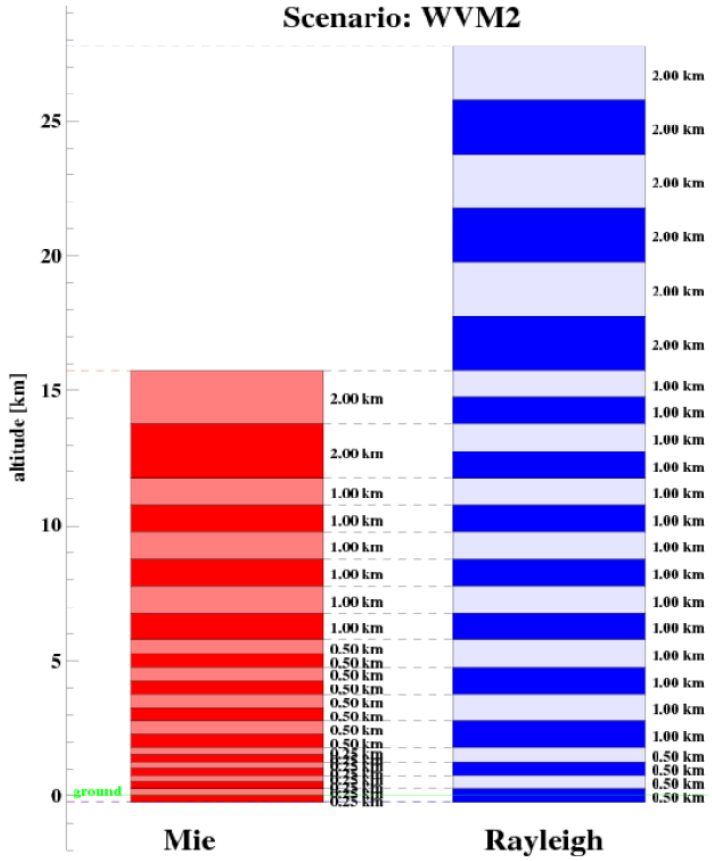


Figure 19. Aeolus vertical range-bins used in test case TN3.1 #0204 for both channels. This selection is referred to as WVM_2km, see VAMP [RD9] for more details.

The E2S v3.02 required some modifications to run this test case successfully at ECMWF. Firstly we had to incorporate a patch as described in the anomaly report AE-IPF-151, which stopped a spurious vertical shift in the wind results when there was significant orography (high DEM). Before correcting this, the verified results look much worse when orography was high (as seen for sections of the CALIPSO test case). Also it was found that the E2S produced NaN (not a number) values in the output for the CALIPSO test cases, and so the processing chain failed. A suggested modification (by ECMWF) in the interpolation of range-bins times as documented in AE-IPF-159 fixed this and allowed it to run successfully at ECMWF (a proper fix has been implemented by Dorit Huber (DoRIT) since).

Figure 20 and Figure 21 show cross-sections of the E2S input HLOS wind and scattering ratio fields respectively for the CALIPSO test case (first half-orbit). This has 219 BRCs, a sufficiently large sample for reliable error statistics. Qualitatively, Figure 20 indicates a fair amount of wind variability in the scenario from the ECMWF fields.

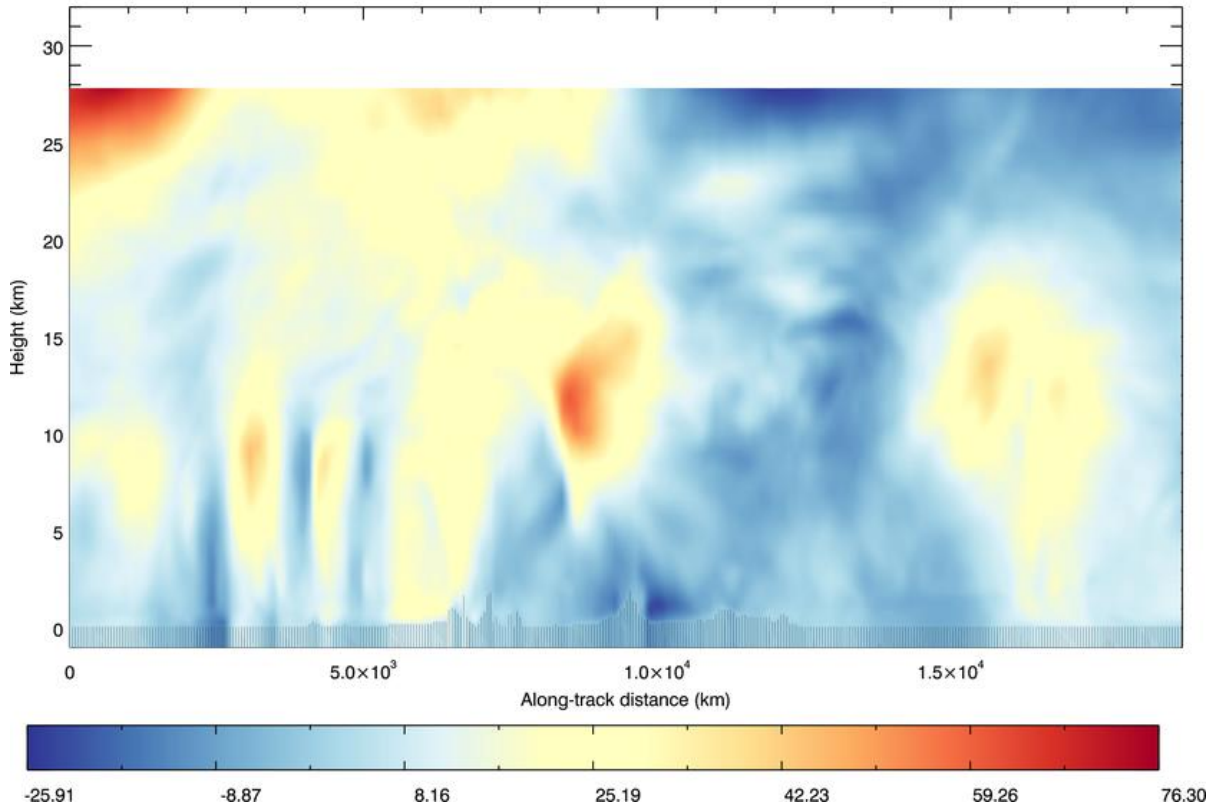


Figure 20. A cross-sectional view of the E2S input HLOS winds (m/s) for test case #0204. The light shading near 0 km altitude shows the terrain of the atmospheric inputs (N.B. this is not equal to the DEM used in E2S).

The CALIPSO scene provides a wide distribution of clouds in the vertical, hence a better assessment of Mie results should be possible than for the one LITE test case of section 2.2.1. Figure 21 shows that the SR retrieved from CALIPSO is fairly noisy, especially at higher altitudes (it appears to be much noisier than the LITE scene of Figure 5, although note the plot scale has changed). This CALIPSO scene does not have a lot of elevated aerosol, which can be spotted by relatively low SR values (e.g. 1.1-1.5) (this was confirmed by looking at the CALIPSO website: http://www-calipso.larc.nasa.gov/products/lidar/browse_images, there is some polluted dust below 4 km altitude at

around 12,000 km along the orbital track).

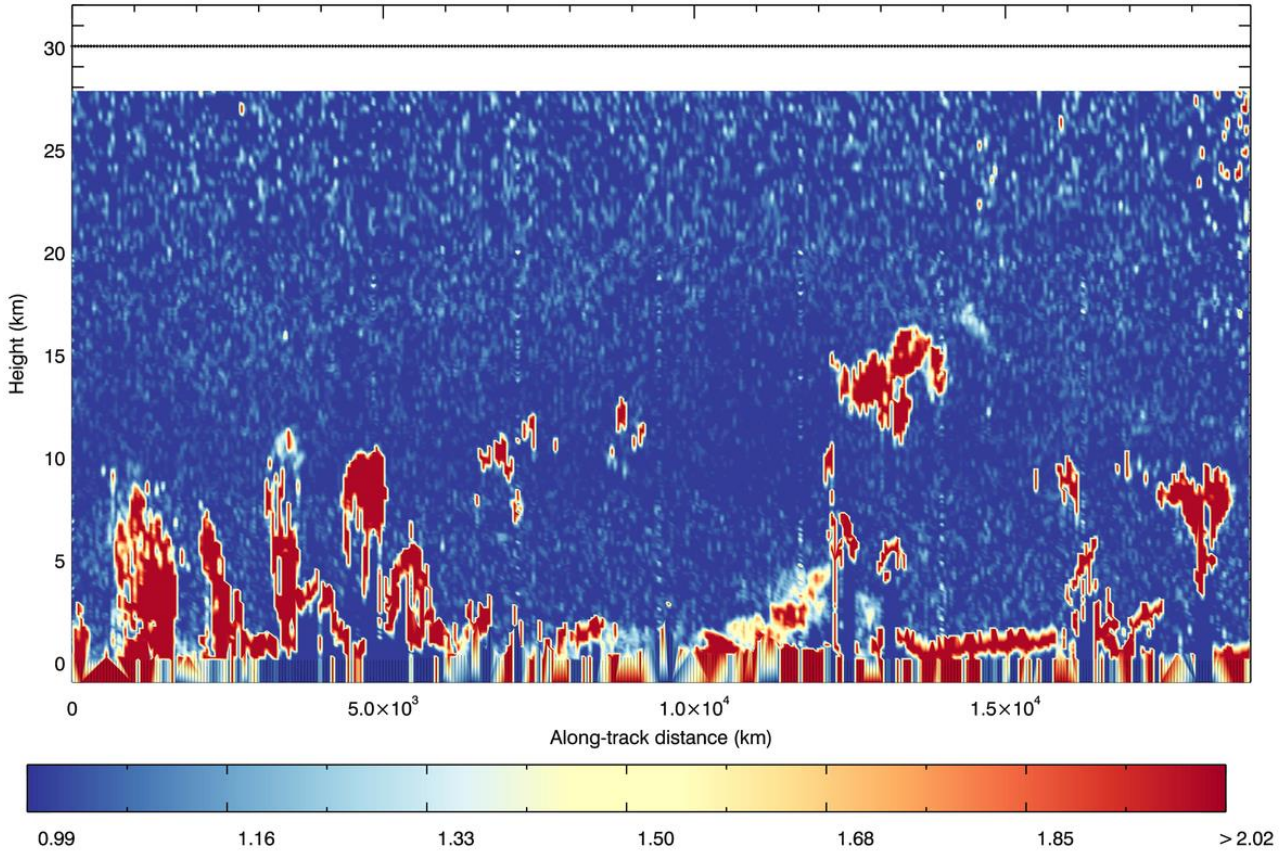


Figure 21. A cross-sectional view of the E2S input scattering ratio (as derived from CALIPSO data, one night-time half-orbit on 1/1/2007 sampling from Northern Russia, through Europe to Africa) for test case #0204.

Figure 22 and Figure 23 are cross-sectional plots of the retrieved L2B Rayleigh-clear observations and the true wind (ECMWF model input winds at the same geolocations) respectively. It is clear that the L2B Rayleigh CM results are capturing the large-scale features of the wind, e.g. jet streams (around 10-15 km altitude) very well in comparison to true wind on the same grid.

The Rayleigh-clear winds have noticeably larger errors below strongly attenuating clouds, e.g. at 14,000 km along-track distance and below 5 km altitude as shown by Figure 22 and more clearly in Figure 24 (which shows a cross-section of the actual retrieved wind errors). The L2B product does a good job of predicting the increased noisiness as can be seen in Figure 25 which shows the L2B estimated HLOS standard error (1 standard deviation); this clearly highlights the larger errors below the strongly attenuating clouds.

Another feature that stands out in the L2B Rayleigh-clear results of Figure 22 are the gross errors in winds below the DEM (as shown by the black line) when sufficiently elevated. This is seen at e.g. the 5,000 and 18,000 km along-track distance marks, where the DEM reaches several kilometres altitude (the first significant high ground is due to Antarctica). It is thought to be a problem with the L1B processing in that Rayleigh data can still be produced below the DEM, when there should be no processing, nor any signal to process. This is further explained in section 4 below.

The below-DEM problem is not apparent for the L2B Mie-cloudy results cross-sectional view as shown in Figure 26; Figure 27 has the corresponding true winds. There is a large sample of Mie

results, as is expected given the E2S input scattering ratio with its many cloud/aerosol features. Also see how the Mie results are complementary to the Rayleigh results, by filling some of the gaps present in cross-section the Rayleigh-clear observations. However, there is a lack of Mie-cloudy results in the elevated dust area pointed out earlier which could potentially provide good Mie results, this is a result of the particular L2B settings chosen (e.g. classification, see section 4).

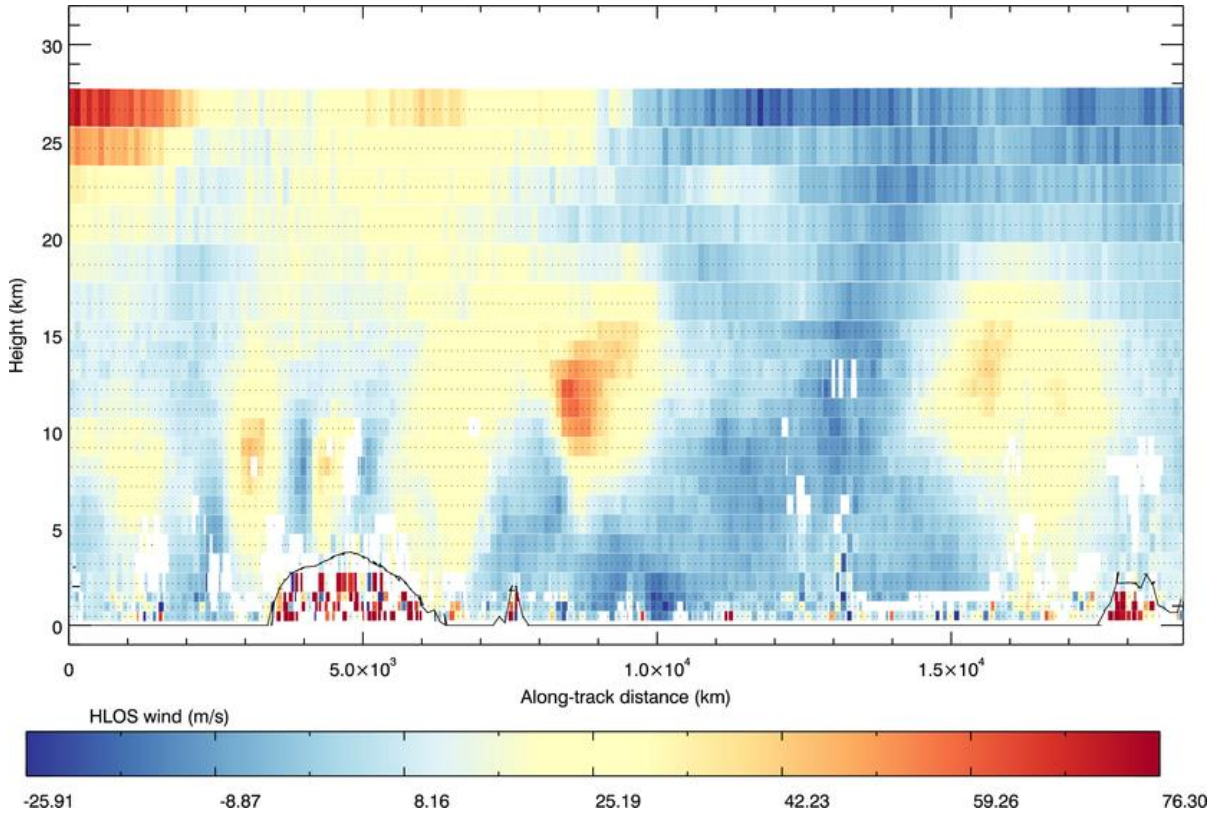


Figure 22. L2B Rayleigh-clear HLOS wind results (m/s) for test case Mispointing_3_1_orb1_with_noise. The black line shows the DEM.

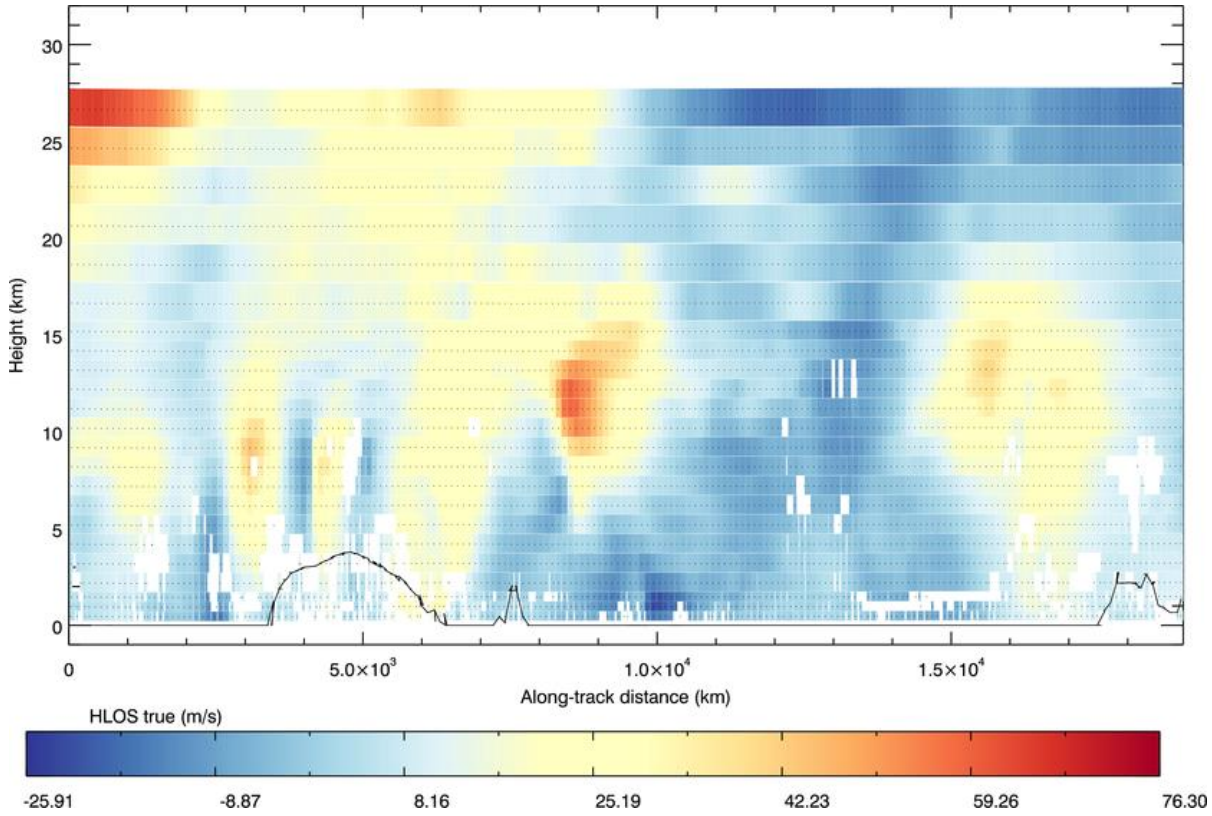


Figure 23. True HLOS winds (E2S input) shown to match the observation resolution/sampling shown in Figure 22.

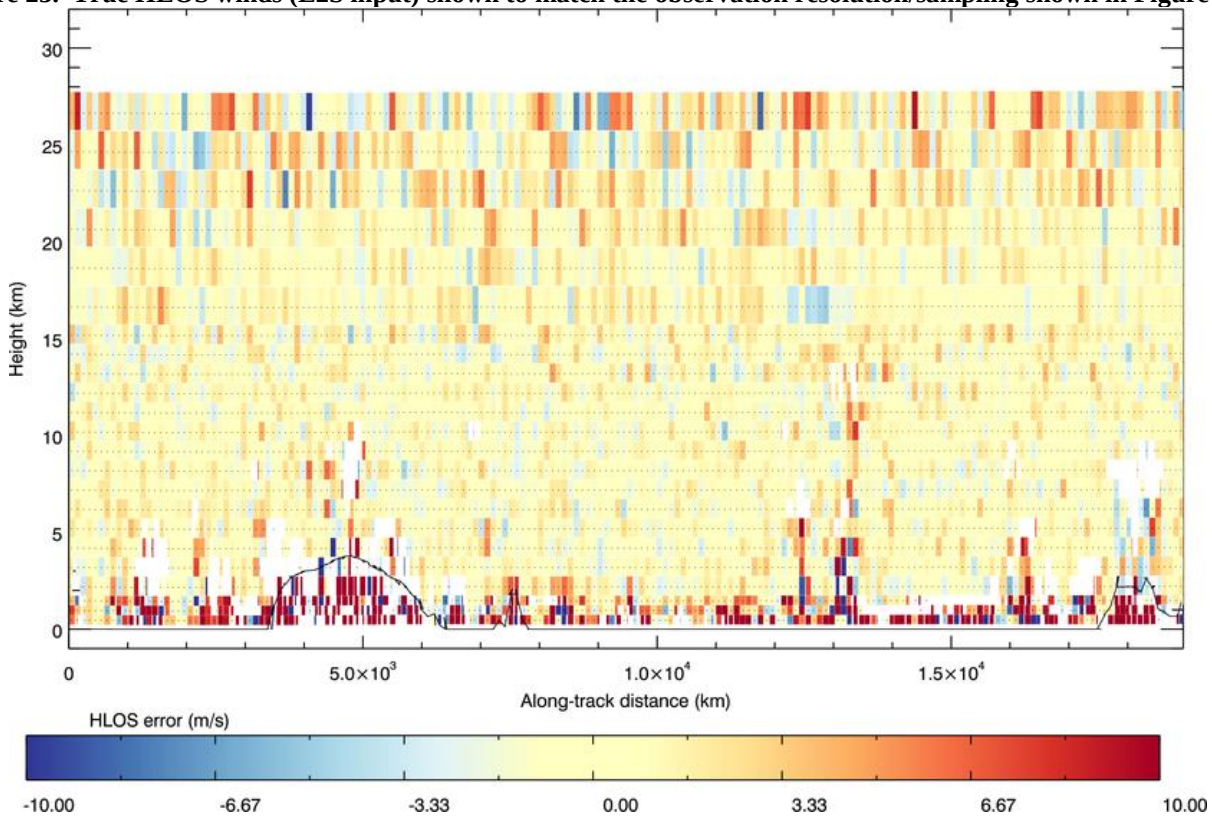


Figure 24. The individual HLOS errors for the L2B Rayleigh-clear observations i.e. Figure 22 minus Figure 23.

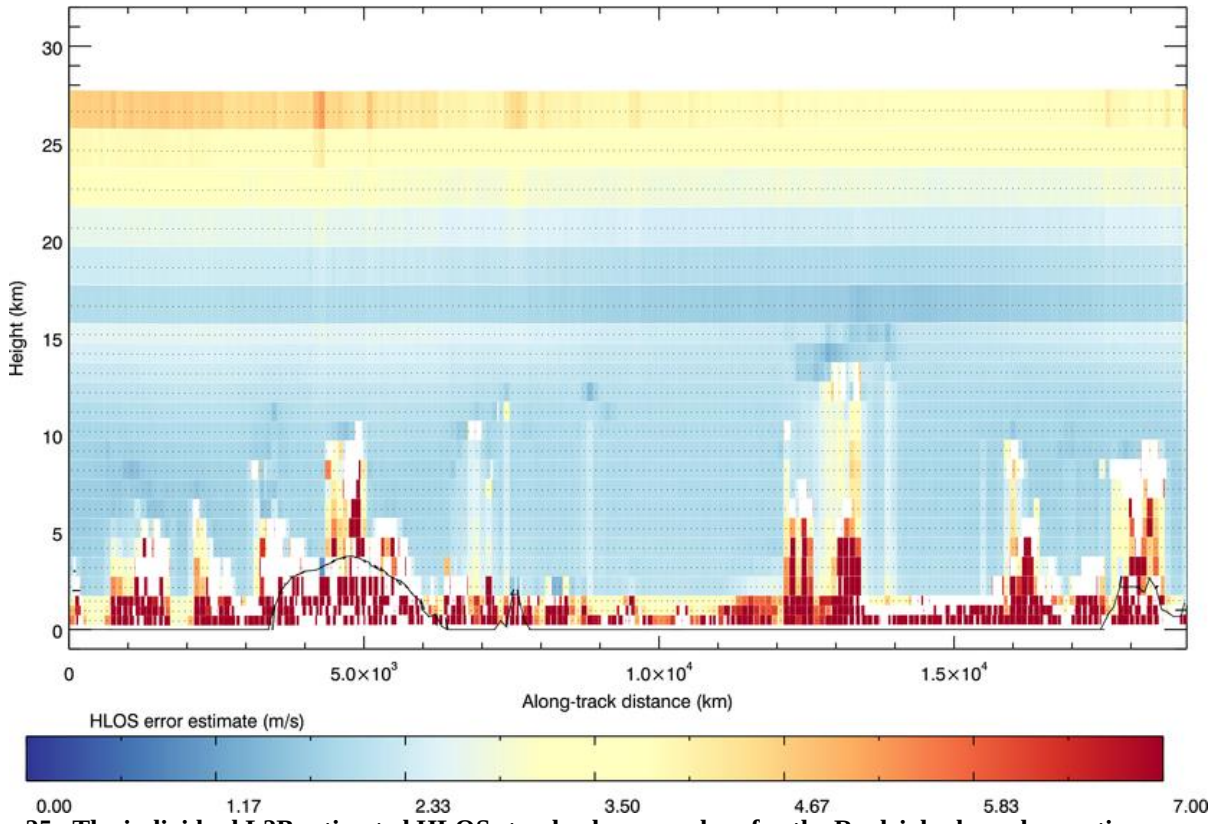


Figure 25. The individual L2B estimated HLOS standard error values for the Rayleigh-clear observations.

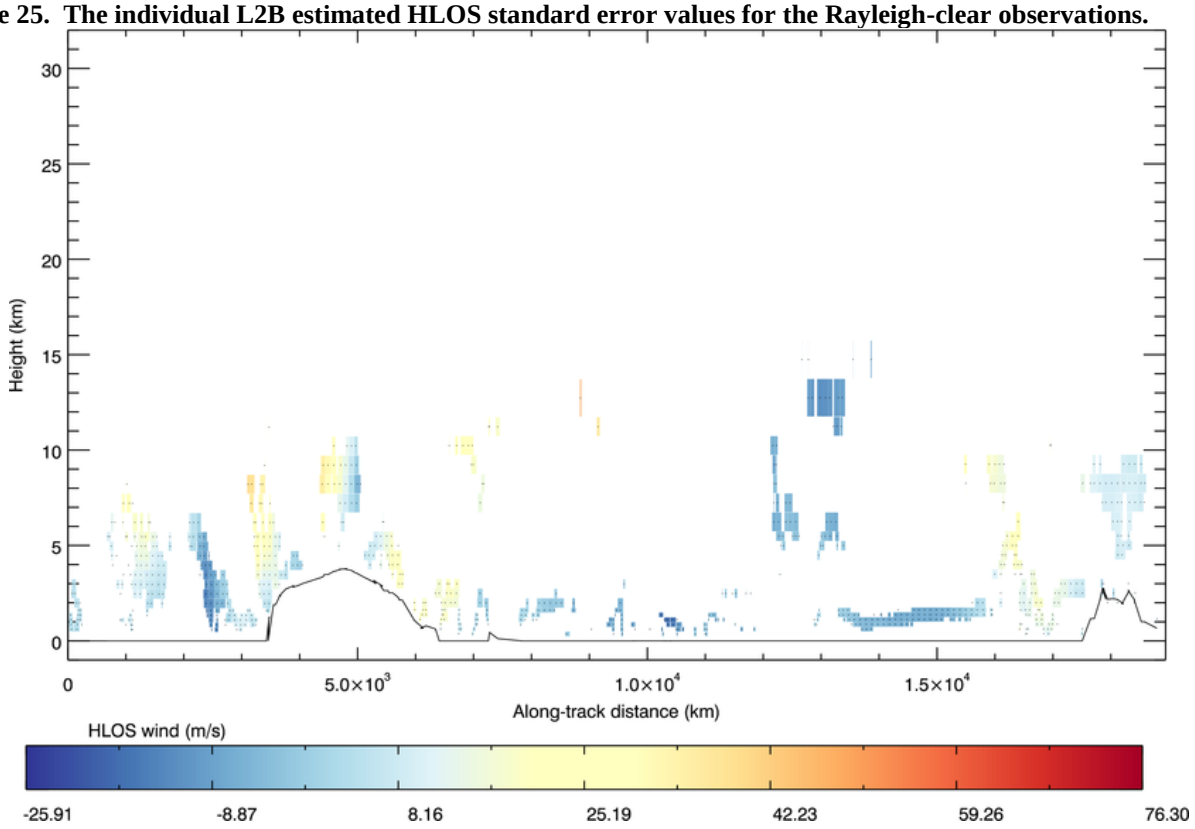


Figure 26. L2B Mie-cloudy HLOS wind results (m/s) for test case Mispointing_3_1_orb1_with_noise. The black line shows the DEM.

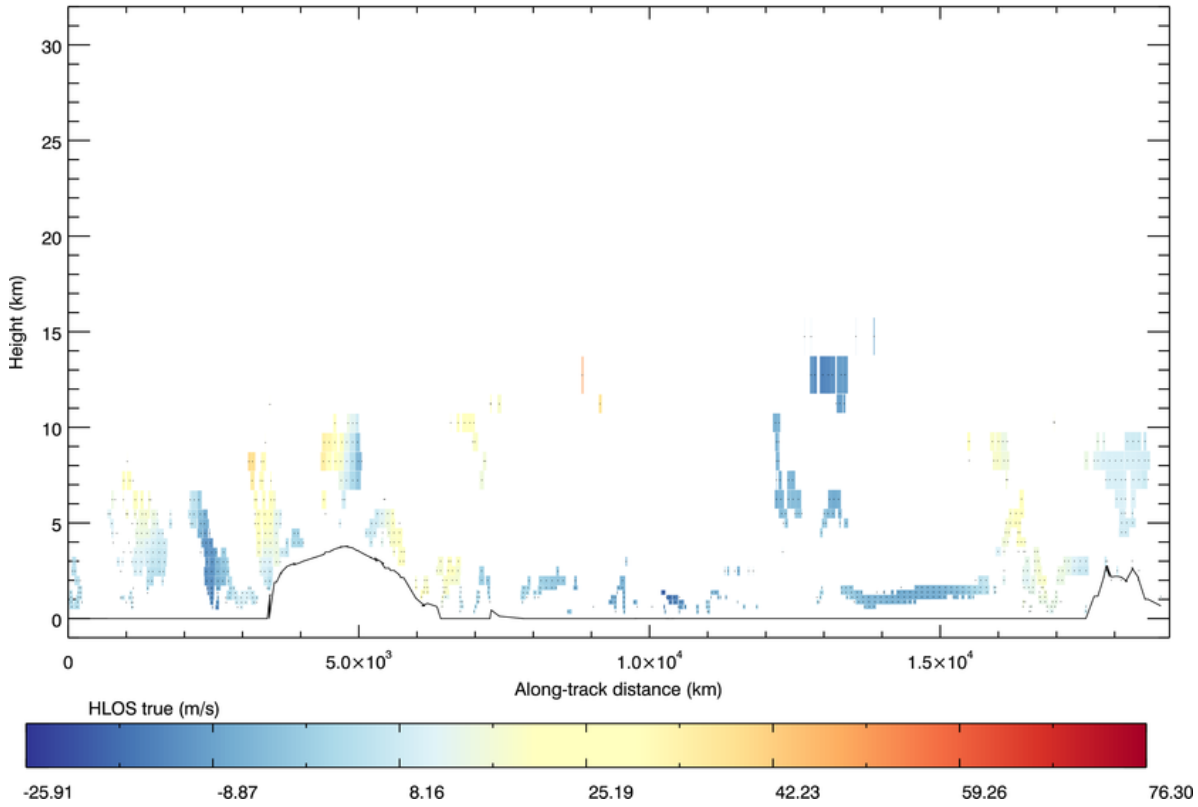


Figure 27. E2S input (true) HLOS winds (E2S input) shown to match the resolution shown in Figure 26.

The verification statistics of this CALIPSO half-orbit test case are presented in the following paragraphs. Figure 28 shows the Rayleigh-clear results, in the same manner as shown earlier. To mitigate the effects of the spurious Rayleigh-clear results below the DEM, as seen in Figure 22, wind results which lie below the DEM have been rejected from the statistics (this improved the error statistics below ~3 km i.e. the maximum altitude of the DEM for this test case). This is the only QC applied for the error statistics.

As seen for the LITE test case earlier, there is positive bias in the Rayleigh-clear winds, although not as large as in the Figure 10 test case (tropical cirrus). But note that the tropical cirrus test case was a shorter scenario repeated many times, whereas the CALIPSO test case sampling a much broader array of scenes in terms of wind and particle distribution, so perhaps some bias can average out to some extent. There is also a large positive bias near the surface where signals have been significantly attenuated due to near surface cloud/aerosol; it is unclear why it should be positively biased, note that the standard deviation of error is also very large (an explanation for this was found after the main investigation, see sections 2.3.2 and 4).

Like in previous test cases the bias is found to have some dependency on HLOS such that more positive bias occurs for negative HLOS winds: Figure 29, showing the HLOS dependence of the HLOS error, demonstrates that for this test case that positive HLOS values are more frequent than in the section 2.2 test case, therefore the bias is less positive. Similar behaviour was observed for the second half-orbit CALIPSO test case (results not shown here). See 2.3.2 for the proposed explanation of the Rayleigh channel bias.

The increased variability of the input E2S wind (ECMWF) has apparently not degraded the quality of the L2B results by any significant margin, at least when comparing the error statistics above 5

km with those to the LITE tropical cirrus case. Below 5 km the results are clearly degraded, but this is because the scenario is much cloudier than the tropical cirrus case. The results which lead to the large positive bias and standard deviation near the surface could be rejected in NWP by using the L2B error estimate. The L2b Quality Control (QC) can be set to remove observations with high standard deviations, which is clearly the case below 2-5 km as seen in Figure 25, where the variability goes above 7 m/s.

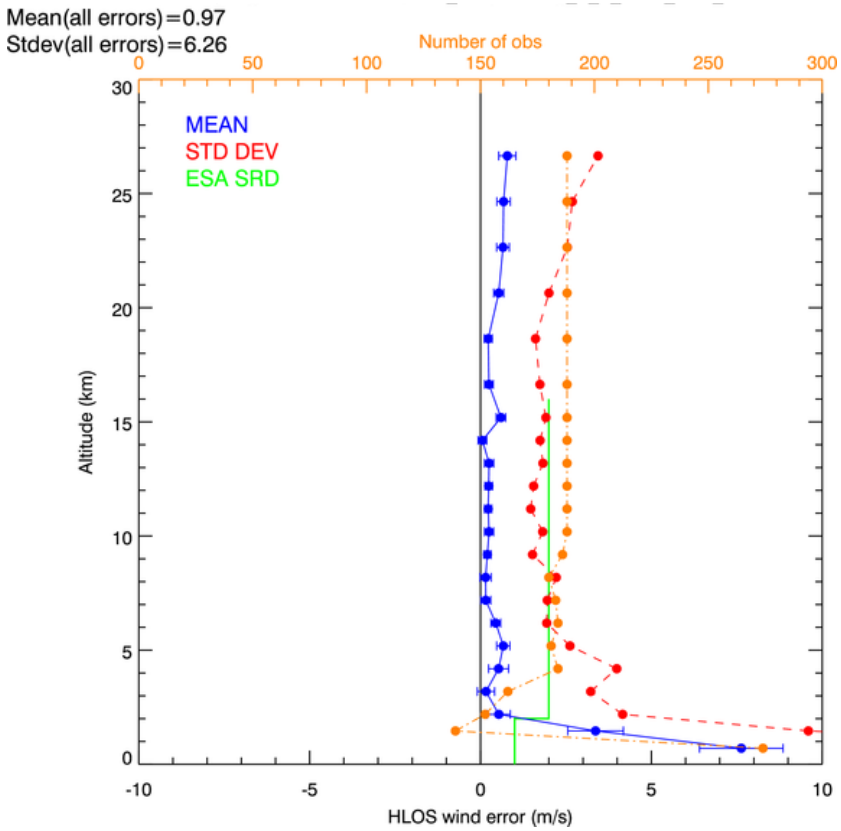


Figure 28. The statistics of the Rayleigh-clear L2B HLOS wind errors from CALIPSO test case #0204 (first half-orbit). The number of winds used at each range-bin is shown in orange. Group size is restricted to 100 km.

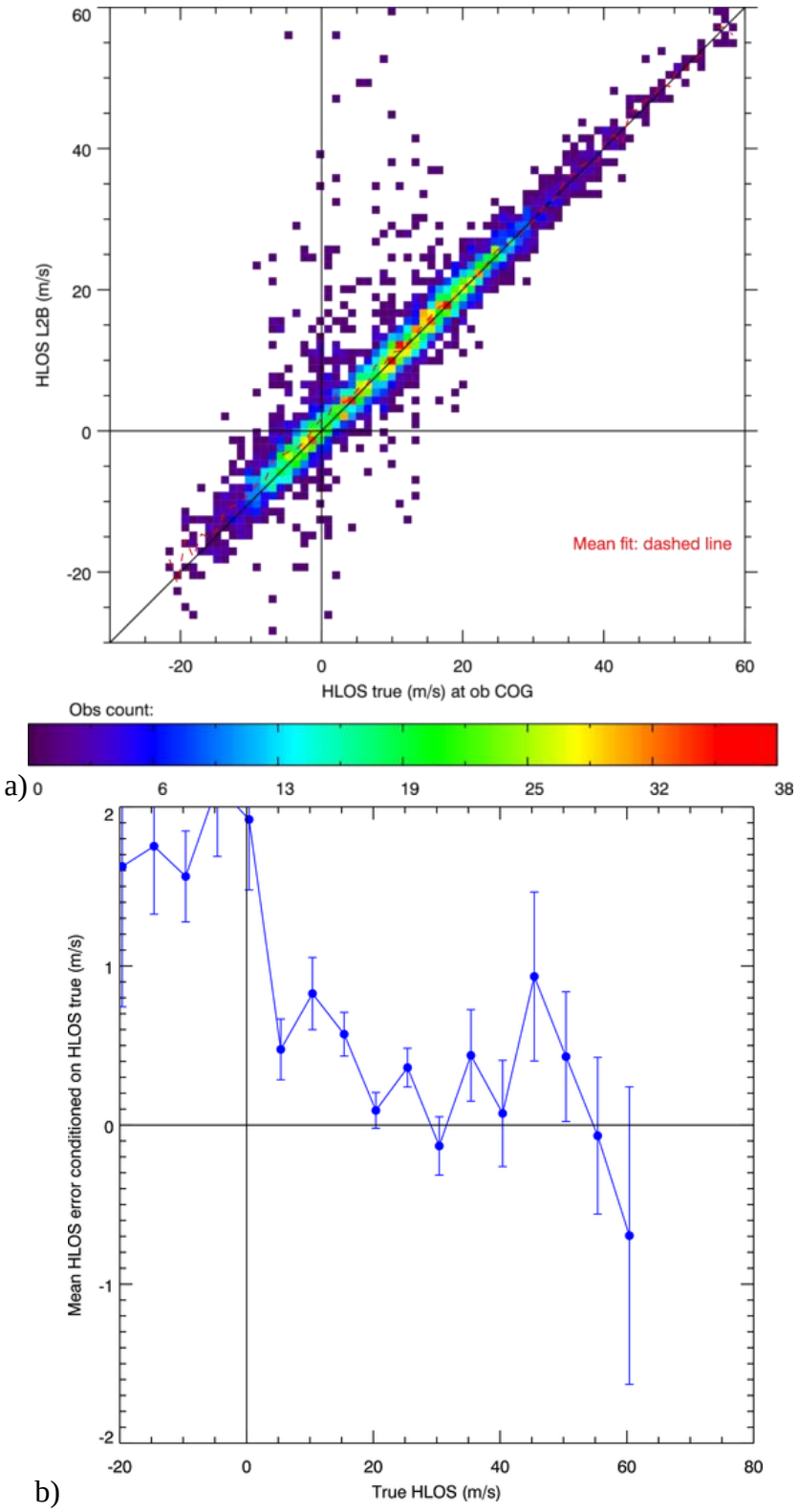


Figure 29. The statistics of the Rayleigh-clear HLOS errors for test case #0204. a) Density plot of all L2B HLOS error vs. true HLOS wind. The red-dashed line shows the mean of the y given x. b) The mean HLOS wind error conditioned on the true HLOS wind.

The Mie-cloudy results are given in Figure 30. As mentioned earlier, the Mie observations have a much greater vertical coverage due to the cloudiness of the scene compared to the short LITE tropical cirrus test case — there are many more results towards the surface, but fewer at 15 km (this test case lacks tropical cirrus). The Mie-cloudy results appear to be of good quality, i.e. they have standard deviation of 1-2 m/s and bias less than 1 m/s, although there is perhaps again a tendency to negative bias, especially between 3-7 km. Some HLOS dependence to the Mie bias was observed, however it is unclear how reproducible it is (it is certainly not a linear dependence, like for the Rayleigh). The stricter ESA SRD of 1 m/s is not met for the Mie winds (nor is it for the Rayleigh) in the lowest few kilometres. This indicates that accuracies of 1 m/s cannot be achieved for realistic scenarios, with very different distributions of cloud/aerosol to the ESA SRD RMA aerosol profile.

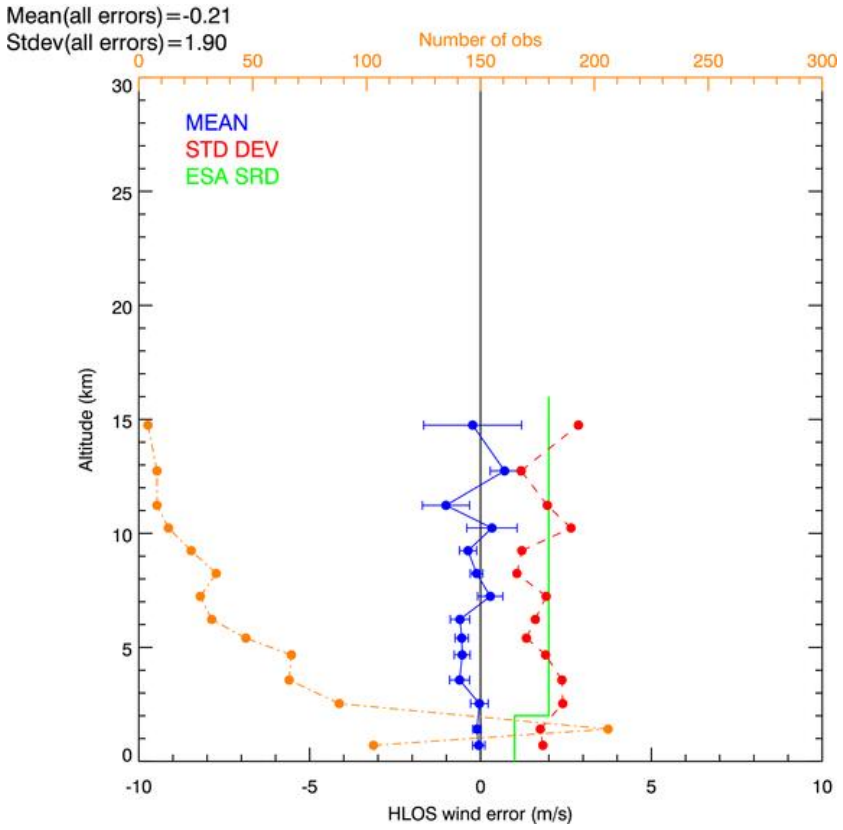


Figure 30. Statistics of Mie-cloudy L2B HLOS wind errors from CALIPSO test case #0204 (first half-orbit). The number of winds used at each range-bin is shown in orange. Group size is restricted to 100 km.

In general, from the CALIPSO test cases (the most realistic investigated in the report), the Aeolus L2B wind results are of good quality (relative to the expectations for Aeolus i.e. ESA SRD) in terms of random error above 2 km. The main concern lies with the wind dependent bias in the Rayleigh results and some tendency for negative bias in the Mie results, and also that the Mie L2B error estimate appears to be too small. The reason for the negative bias is investigated further in section 2.3.2 and 4 below.

2.3.1 Quality of L2B estimated errors

The verification results of the second CALIPSO half-orbit available in the KNMI database for 1/1/2007 is given by Figure 31 and Figure 32 (note that the E2S inputs are not shown for the second half-orbit, but they are similar to the first half-orbit in terms of cloud distribution and wind variability). The second half-orbit has very similar error statistics to the first half-orbit. A modification to the verification plots is the plotting of the mean of the estimated L2B observation error (in purple) for each altitude bin. The mean observation error is provided in the L2B product and was plotted for the first CALIPSO half orbit in Figure 25). This average error is the mean of the L2B error estimates for all available observations on each altitude level. This L2Bp estimated error line should ideally match the standard deviation of true HLOS error standard deviation line (shown in red).

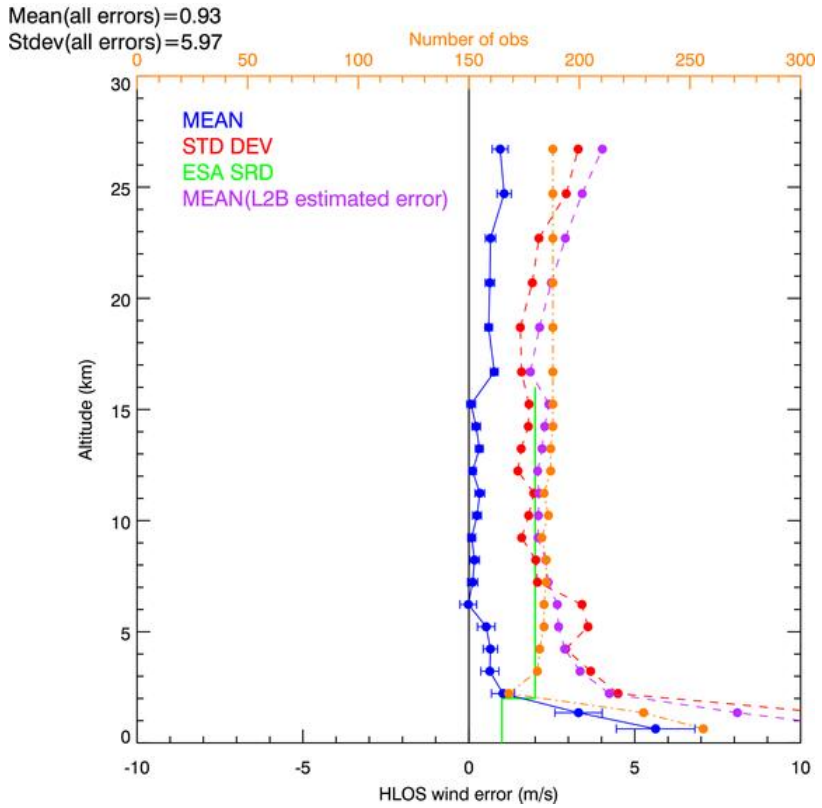


Figure 31. The statistics of the Rayleigh-clear L2B HLOS wind errors from a second CALIPSO half-orbit available on 1/1/2007. The mean of the L2B observation estimated standard errors are shown in purple. Group size is restricted to 100 km.

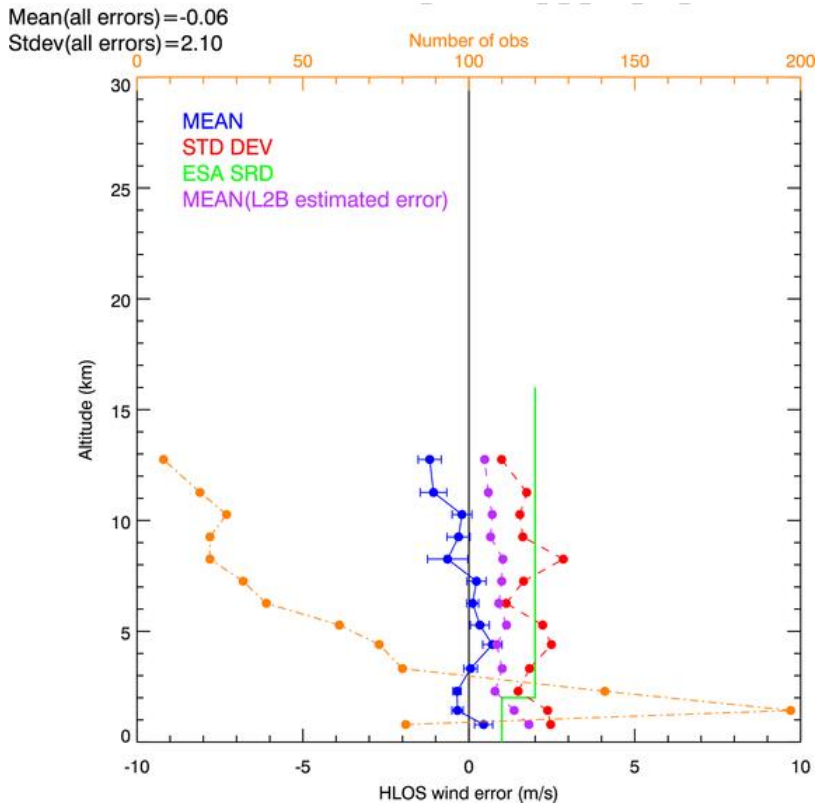


Figure 32. The statistics of the Mie-cloudy L2B HLOS wind errors from a second CALIPSO half-orbit available on 1/1/2007. The mean of the L2B observation estimated standard errors are shown in purple. Group size is restricted to 100 km.

For the Rayleigh-clear winds, the L2B estimated errors are a little too large above 6 km and too small below; but they are typically within 20% and therefore are thought to be very good estimates. One should bear in mind that most observation types used in NWP are not able to provide accurate dynamic error estimates like this. The Mie-cloudy HLOS error estimates are unfortunately underestimated, being around 40-80% of the real HLOS error standard deviation. The processor should be checked to ensure the calculation is being done correctly i.e. to check if any error terms are missing, to see if this estimate can be improved upon.

2.3.2 Further investigation into the Rayleigh bias

The Rayleigh-clear bias was investigated further using the second CALIPSO half-orbit test case. It was found that increasing the laser energy from 110 mJ to 1000 mJ in the E2S significantly reduced the positive bias, particularly for range-bins with relatively low signal levels i.e. the upper range-bins (16-27 km) and those lower down which appear to be affected by particle induced attenuation (compare Figure 33 to Figure 34). This confirms that the AUX_RBC file is not the source of the bias, since it remains constant in both tests.

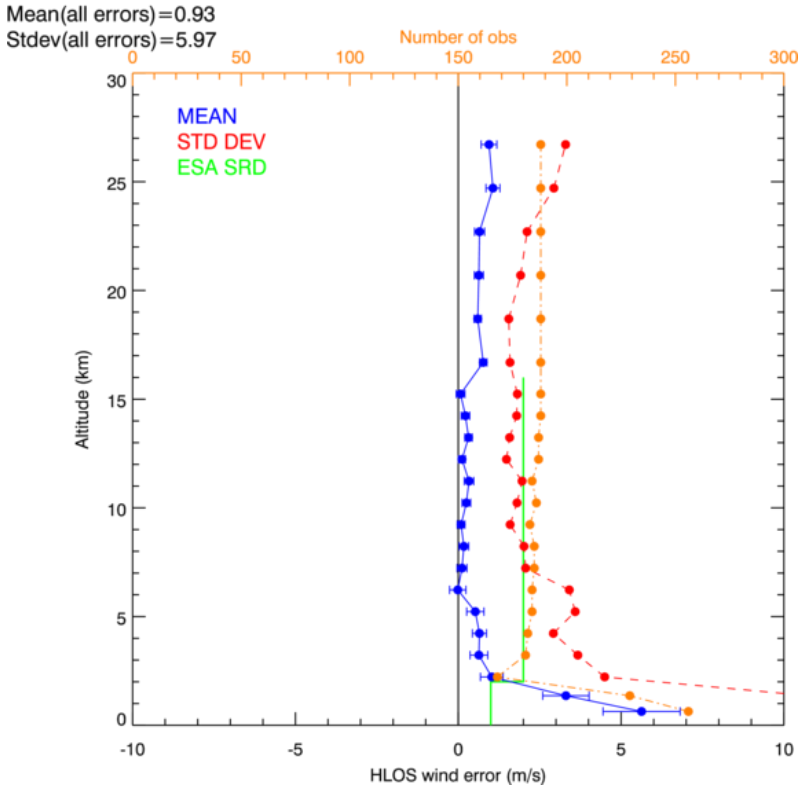


Figure 33. Statistics of Rayleigh-clear HLOS error for the second half-orbit of the CALIPSO test case. Using 110 mJ laser energy.

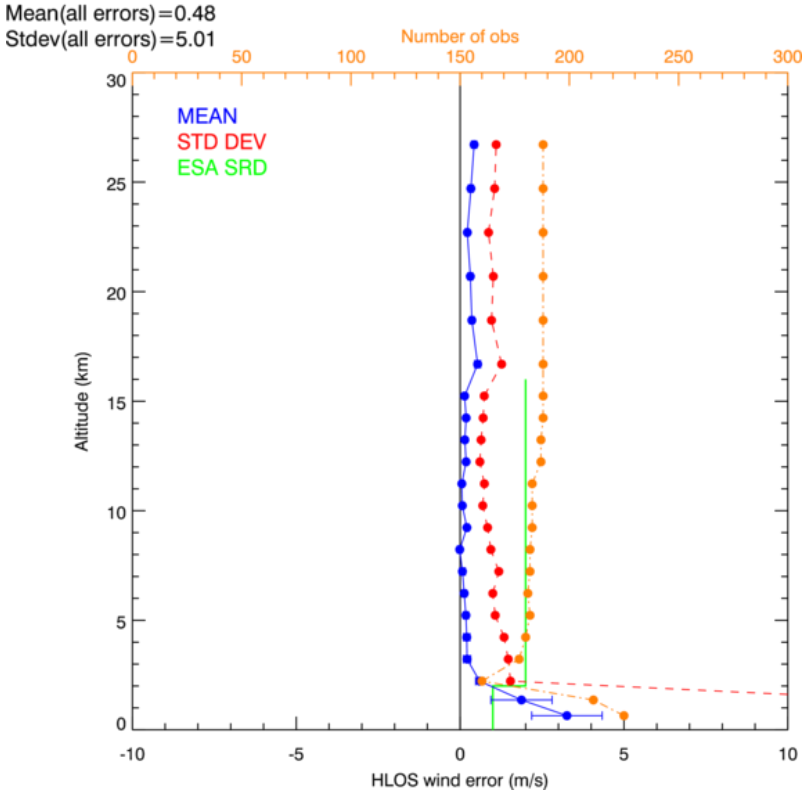


Figure 34. Statistics of Rayleigh-clear HLOS error for the second half-orbit of the CALIPSO test case. Using 1000 mJ laser energy.

By running a number of simpler test cases (e.g. test case #0016), which have no particles in the scene, the evidence suggests that perhaps the positive bias is coming from two sources: low Rayleigh signal levels at the measurement level and residual scattering ratio estimate errors.

Some aspects of the effect of Rayleigh signal levels on the bias:

- The bias appears to become more positive as the Rayleigh signal levels decrease (as a result of lower laser energy or lower signal due to the nature of atmospheric backscatter and attenuation). Figure 35 below shows how the bias tends to become more positive as the L2B estimated error (approximately inversely proportional to square root of signal strength, see Appendix section 6.5) increases. This figure is from a test case where the E2S input scattering ratio is fixed at 1.0 and the L2B processor is forced to use 1.0 (and laser energy is 110 mJ), thus eliminating SR errors. This effect seems to be responsible for a residual positive bias of order 0.5-1 m/s when the laser energy is set to 110 mJ. Note in this test case there is no HLOS value dependency (slope error).

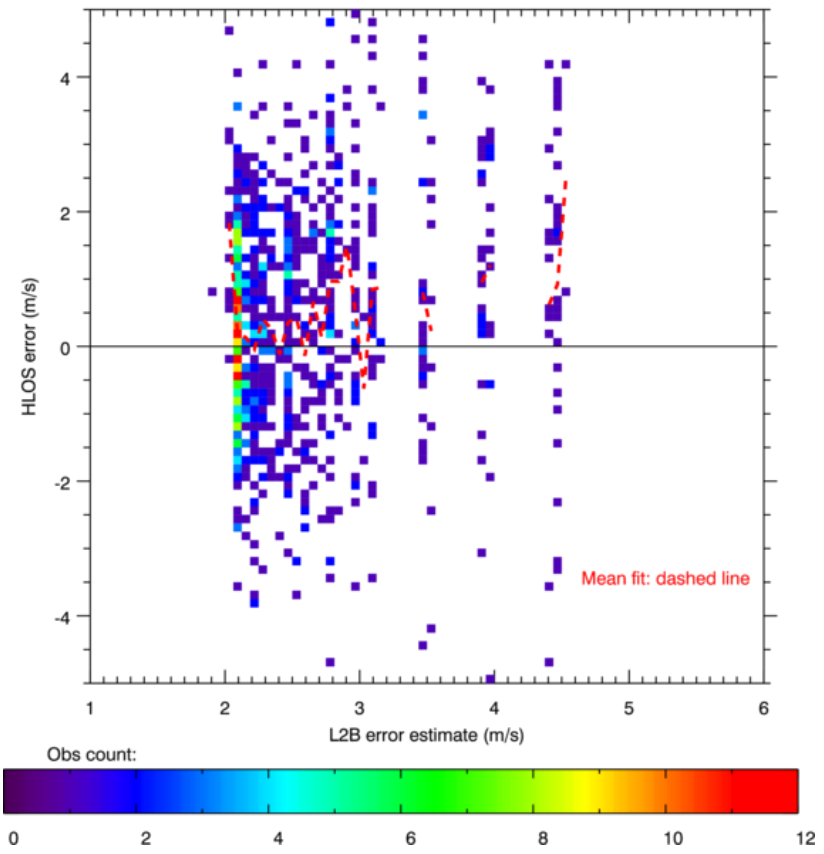


Figure 35. HLOS error distribution (and mean: red dashed line) as a function of the L2B error estimate. The red-dashed line shows the mean of the y given x. This is for a simple test case with winds constant per BRC and with no particulate backscatter.

The same tendency to more positive bias as the signal strength decreases is also seen for the first half-orbit CALIPSO results, see Figure 36.

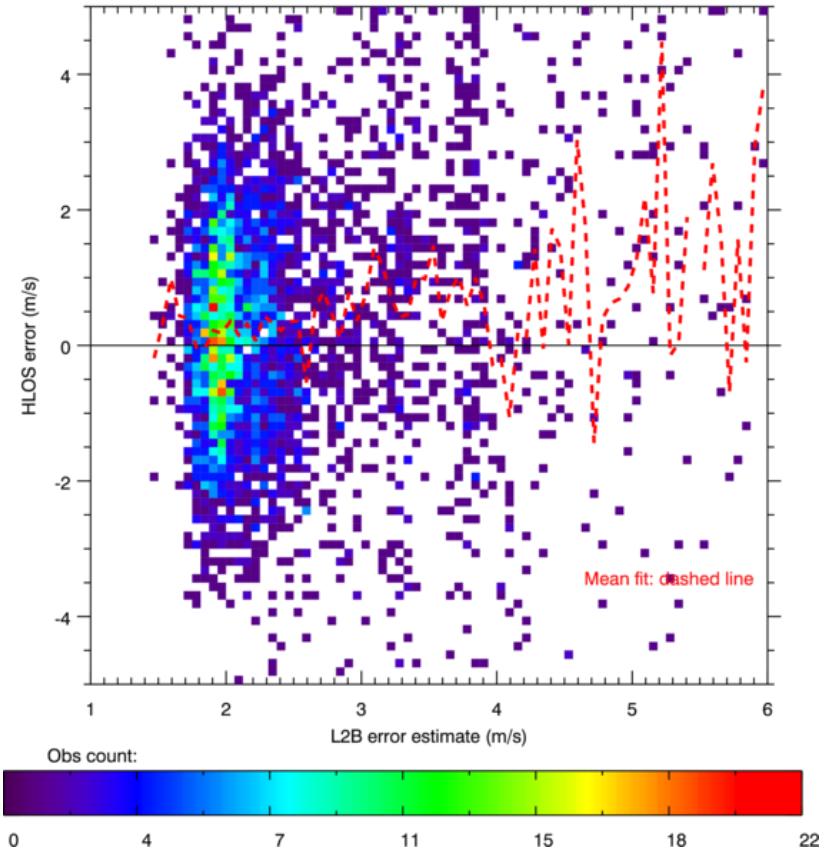


Figure 36. HLOS error distribution (and mean: red dashed line) as a function of the L2B error estimate. The red-dashed line shows the mean of the y given x. This is for the more realistic CALIPSO first half-orbit test case.

- The bias does not change with the Rayleigh observation horizontal averaging length; if the bias is systematic to the measurement data then averaging measurements will not remove the bias.
- Tests were run with the background solar spectral radiance set to zero (compared to the default 1813 $Wm^{-2}\mu m^{-1}$) to see if this affected the bias — it did not.

A hypothesis for the cause of the bias is that the rounding of FP signals A and B which results from 16-bit analogue-to-digital convertor (as simulated in E2S) for low signal counts may be producing a "quantisation" error that becomes apparent in CM with 110 mJ e.g. Rayleigh spectrometer counts can be 200 or 201, not some fractional value, which affects creates error in the Rayleigh response and hence HLOS wind. However, it is unclear why this leads to a systematic error. The root cause for the systematic nature of the "quantisation" error needs to be understood. **After the main investigation, the real source of the bias was found, see section 4.**

Some points on the effect of scattering ratio residual errors:

- If the scattering ratio is positively biased i.e. the L1B derived SR values are larger than the E2S input true SRs then this reduces the resultant L2B HLOS values. A small positive SR bias is evident when using the L1B refined SR, see the Appendix section 6.6. This effect can therefore reduce and compensate for the positive bias that is apparently coming from

low Rayleigh signal bias.

- However in some test cases the SR used by the L2B processor can be slightly negatively biased. This occurs for the Rayleigh range-bins higher than the uppermost Mie range-bins. Here the SR is assumed to be 1.0 (since there is no Mie channel range-bins are available to estimate the value). In CALIPSO test cases there is often input E2S SR > 1.0 values at these levels (which mostly looks like noise in the signals). Therefore there is a negative SR bias, which increases L2B HLOS values, adding to the positive bias from low Rayleigh signals. It is a concern that this effect will produce very large bias when polar stratospheric clouds are encountered (where Mie range-bins are absent) for real Aeolus data.

Therefore a hypothesis for the HLOS bias resulting from **SR systematic errors** is that they **are larger, the further away from the Rayleigh FP filter cross-points** (around +40 m/s in HLOS space) the Doppler shift is. This will induce a wind bias that depends on the true HLOS when the SR has systematic errors (i.e. a slope error). See the Appendix section 6.6 for further detail on this hypothesis.

It is thought that the improvement in bias going from 110 mJ to 1000 mJ in Figure 33 to Figure 34 is due to a combination of improving (a) the SR estimates (slope error improves) and (b) the smaller effect of "quantisation" errors (bias on highest range-bins reduces) as the spectrometer counts increase.

The Aeolus SRD has been defined for an ideal atmosphere with particle scattering only taking place in the PBL. It defines requirements for absolute and slope-error accuracies. The operational tool available today to translate the SRD requirements into the L2b MRD requirements is the E2E, L1b, calibration and L2b chain of processors, run on realistic test cases. Currently, the MRD and SRD bias requirements are identical, which means that biases in realistic atmospheres should not exceed those of an ideal atmosphere. This investigation has shown that great care has to be taken in minimizing potential biases through appropriate design, on-ground and in-orbit characterization, and data processing and testing if this is to be achieved.

The tendency of the Mie results to negative bias should be investigated in future.

3 Conclusions

The quality of the HLOS winds produced via the continuous mode chain-of-processors, up to the L2B product stage has been investigated. Generally, the L2B products are performing well, particularly in terms of random error. However some unexpected biases were discovered, which appear to have become more prominent for continuous mode data. The Rayleigh results have wind-speed dependent biases up to 1- 2 m/s, whereas the Mie have a tendency to negative biases (less than 1 m/s). The Rayleigh bias has a sensitivity to signal level and small errors in the estimated scattering ratio, perhaps explaining why it became more prominent for CM data rather than for BM due to effective lower energy delivered to the atmosphere per unit time. It is concerning that biases are already apparent without considering errors in the calibration process.

Some tweaking of the L2B processor settings was required to obtain the good results. For example, the use of the refined L1B scattering ratio for classification and Rayleigh channel Mie cross-talk correction was critical, and the more appropriate vertical placement of the Rayleigh winds to mitigate representativeness issues.

In terms of random error the L2B results (with 110 mJ laser energy and 100 km horizontal averaging) are meeting the ESA system requirements for the 2-16 km vertical range. However,not below 2 km. This discrepancy should be investigated further. The horizontal observation error correlation for Rayleigh winds was found to be very small with testing using one LITE scenario. However, this result from a very limited test case may not be representative for global data over all seasons.

The new grouping algorithm of the L2B processor has been shown to work well in terms of random error calculation on observation level for both the Rayleigh and the Mie channels.

Finally the L2B observation error estimates have been shown to agree well with the true error statistics for the Rayleigh channel. However, some further investigation is required for the Mie channel, since they are roughly half the true error standard deviation.

4 Important issues found after the investigation

This section explains some important issues that were discovered after the main investigations were completed:

- **CALIPSO scenarios**

  Whilst performing the investigation it was assumed that the CALIPSO test cases (half-orbit scenarios) used the full-horizontal resolution KNMI processed CALIPSO profiles i.e. spaced every 3.5 km (the old BM measurement spacing). However, it was later discovered that profiles spaced every 50 km were used (KNMI created these test cases with lower sampling to save E2S processing time for half-orbits, which could take days to run). With the recent E2S updates (in 2013) which have significantly improved the E2S processing time, the high-resolution CALIPSO scenes can be used (profiles every 3.5 km). With the high resolution scenarios, the Rayleigh-clear error statistics remained similar; however a positive effect was that more Mie results were produced, due to the increased scattering ratio variability on the small-scales (less than 50 km).

- **E2S interpolation**

  For all test cases in this report, E2S interpolation was switched "on". This is a potential source of error for the results shown, since the effective E2S input winds are modified by the interpolation i.e. we no longer verify against the "true" E2S wind inputs which lead to simulations of the backscattered spectra.

  An investigation into the difference between interpolation "on" and "off" for a CALIPSO half-orbit test case, with profiles spaced every 50 km was done. With interpolation "off" (and still verifying against the KNMI atmospheric database inputs) the HLOS error standard deviation improved slightly for the Rayleigh-clear results, but was 20% worse for the Mie-cloudy results. The better Mie results with the interpolation "on" can be explained by the interpolation smoothing the input SR and hence increasing the measurements with sufficiently large Mie backscatter, and hence increasing the number of Mie winds. The biases remained the same. In future testing, interpolation will be switched off by default.

  With high-resolution CALIPSO data (i.e. profiles every 3.5 km) the interpolation should make little difference, since the CM measurements are spaced every 2.85 km.

- **Mie Quality control**

  The L2B processor is using the Downhill-simplex Mie core algorithm. The behaviour of the algorithm can be monitored by the QC flags. However in the test cases investigated in this report the QC flags were ignored. If a "non-missing" HLOS wind value was produced by the Mie core algorithm, then it was used as part of the statistics.

  Some odd behaviour for the L2B Mie QC flagging needs to be investigated: many flags are being triggered all the time, hence this is why they were ignored. For example, when the L2B processor uses the L1B Mie core settings (as controlled via the AUX_PAR_2B file) we consistently get the following flags being triggered: `MaxItLorFit_meas; ResErrThresh_meas; MaxItNonLinOpt_meas; MaxItLorFit_ref; ResErrThresh_ref; MaxItNonLinOpt_ref; SNR_threshold_failed`.

  Also, it was discovered that the L2B processor SNR threshold flag is not meaningful, it is actually set to be the L1B "measurement level" SNR value in the middle of a L2B group of measurements (that are accumulated for the retrieval). The SNR should be that of the accumulated observation, not the SNR of one measurement in the middle of the group. Therefore we ignore the SNR flag for the L2B Mie results. To the best of our understanding, there is no QC being applied to the Mie results for the error statistics of this report. See section 6.8 for an explanation of the L2Bp Mie quality control procedure.

  Using the L2B estimated error values (which are correctly calculated at the observation level) as a QC criterion; we find that the standard deviation can be reduced by setting the threshold to e.g. 2 m/s.

- **Clarification on the classification**

  The classification into cloudy or clear measurements is based on the L1B measurement-level SR values (refined SR was chosen). This choice is controlled in the L2B AUX_PAR file. The SR thresholds are provided as a linear function with height, i.e. currently SR of 2.0 at 0.0 km and 3.0 at 30 km. These parameters are not necessarily optimal for CM.

The measurements classified as "clear" go into retrieving a Rayleigh-clear and Mie-clear winds, and the measurements classified as being "cloudy" go into retrieving a Rayleigh-cloudy wind and a Mie-cloudy wind (on a given range-bin). In this report we disregard the Rayleigh-cloudy and Mie-clear winds, since they are generally of significantly poorer quality in terms of bias and standard deviation. However, the Mie-clear winds do not seem too bad however (standard deviation around 2-3 m/s). Perhaps a more sophisticated classification is needed to get the most from the Mie.

- **Rayleigh bias**

It was discovered that the wind-dependent bias of the Rayleigh-clear results was mostly due to an effect called Dark Current in Memory Zone. This is a shot-noise which is accumulated and added to the Fabry-Perot A and B signals in the E2S. The effect of DCMZ is to induce a wind-speed dependent bias due to the nature of the Rayleigh response calculation; the size of the effect is larger in magnitude for lower signal levels. The bias is positive for winds less than 40 m/s (the filter cross-point) i.e. -100 to +40 m/s and negative for winds greater than 40 m/s, hence there is an overall tendency to positive bias, due to the larger number of winds in this range.

Systematic errors in SR can also induce a wind-dependent bias (as has been demonstrated); however it is believed that the main contributor to the wind-speed dependent bias seen in the report is due to DCMZ. QC based on L2B HLOS wind estimated error (e.g. reject > 6m/s) certainly improves the bias significantly by removing the results with lowest signal level and hence largest bias.

- **Accuracy of Mie below 1 km**

It was noted that the Mie winds do not achieve a 1 m/s random error quality below 2 km for realistic atmospheres (the ESA SRD requires this, although for an idealistic atmosphere) — 2 m/s is achievable however.

It was queried by a reviewer of the report if this was because the boundary layer aerosol Mie results are not being generated in the L2Bp due to classification thresholds i.e. rejecting too many low SR measurements (i.e. those with SR < 2.0) from the Mie-cloudy results, which could provide the high accuracy winds. Comparisons with L1B Mie results showed that the L1Bp is producing reasonable quality Mie results in low SR areas (as more commonly seen for the LITE scenes, but not at 1 m/s accuracy it would appear) where the L2B Mie-cloudy results are absent. Also, the L2Bp Mie-clear results (which would contain the low SR results) statistics were spoiled by outliers.

An investigation to improve the L2B Mie results by altering the L2B settings was undertaken. After varying some of the processing settings it was found that more Mie results could be obtained from the low SR areas with the following L2B settings:

1. Ignore the L1B measurement-level QC flag - which caused unnecessary rejection of lots of measurements which have low Mie signals (but still viable when averaged).

2. Allow the screening of measurement-level L1B refined SR to have minimum value of 0.8 (logically it was previously set to 1.0) — this allows many more measurements with potentially low Mie contributions to enter the wind retrieval.

3. Change the L2B SR threshold for classification from 2.0 to 1.15 (thus allowing many of the low SR values to become part of the good quality "Mie-cloudy" results), in practice bringing in any aerosol backscatter regions. The classification applies to both Mie and Rayleigh i.e. there are not separate thresholds for each.

4. With the above settings there are many more outliers in the Mie-cloudy and Rayleigh-clear error statistics; hence the use of QC becomes much more important. Using the L2B Mie observation overall validity flag for QC became necessary (with the old settings this flag was never raised, whereas it is quite regularly with the new settings). Also QC based on the L2B error estimates became necessary. Reasonable criteria are: reject Rayleigh-clear with > 6 m/s error estimate (to avoid biases from the low signal levels), reject Mie-cloudy with error estimate > 2.8 m/s (remember the Mie error estimate is too small, see section 2.3.1, so this is not as strict as it would appear) to avoid gross errors resulting from low signal levels.

Testing with a variety of CALIPSO half-orbit test cases (valid on the 1/1/2007) revealed the following. The L2B **"Rayleigh-clear"** is much less biased than the L1B Rayleigh due to L2Bp application of the temperature and cross-talk corrections. With the new settings, the L2B Rayleigh-clear is less positively biased due to the rejection of low signal levels results (which are affected by the DCMZ effect) and smaller SR values (hence less reliance on cross-talk correction). Observation numbers have reduced significantly below 12 km as a result of low signal levels resulting from signal attenuation due to clouds, but the error statistics are considerably improved.

The L2B "**Mie-cloudy**" looks better than the L1B Mie, mainly due to a L1B problem at ~4km (large positive bias and increased standard deviation). This is because ground contaminated observations were not flagged (overall quality flag) in L1Bp v6.01. There are high mountains over the Himalayas, hence the bad statistics for L1B at 4 km. **The L2Bp successfully uses the ground wind detection on measurement level provided by the L1Bp output and so avoids this** (however, in the study above a bug in the L2bP lead to this detection not working – see below). The comparison of L2Bp Mie-cloudy for the new and old settings shows the results have lower error standard deviation with the new settings (closer to 1 m/s, but still above) and critically they are produced in low SR areas (i.e. aerosol). **Mie-cloudy observation numbers are overall increased with new settings**. However, the 1 m/s accuracy near the surface based on statistics for half-orbits is still not achieved, mainly because PBL aerosol with clear air above is relatively infrequent for the CALIPSO data on 1/1/2007. Locally, where such conditions exist, it can be seen that the Mie results are of high quality.

- **Rayleigh Ground wind detection**

  The L2B processor used in this study contained a bug, which meant that ground return contaminated Rayleigh measurements were not being rejected. This led to very poor (biased) Rayleigh winds for range bins in contact with the surface; in fact this introduces a wind-speed dependency to the bias. This bug has now been fixed. This can significantly improve the near-surface Rayleigh error statistics in some test cases.

- **Terrain model**

  It was discovered after the investigations that the terrain model can be switched “on” in E2S which allows the Aeolus vertical rangebins to shift vertically in line with the DEM. This is a

more realistic simulation of Aeolus. Note that the current DEM implementation leads to updates in the bin commanding every 200 km.

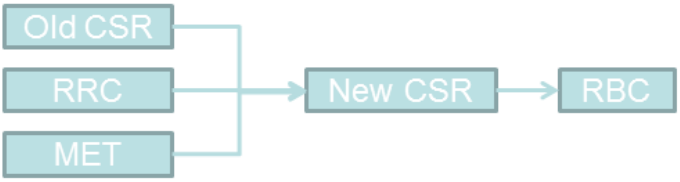
5 Rayleigh calibration testing for WP2250

Sensitivity testing of the L2B Rayleigh wind results to calibration errors is a task remaining from the L2B contract (in particular CCN2: work package 2250). This chapter documents the testing done in 2013 to complete this task. Note that all tests are done with a defined laser output energy of 80 mJ. In earlier chapters, 110 mJ was assumed. Alain Dabas (L1B team/Météo-France) has generated a variety of AUX_RBC_L2 files for the L2B team for testing the impact of a range of error sources in the calibration processing chain. This work does not deal with the Mie calibration errors, only the Rayleigh. More in depth calibration testing will be performed as part of CCN5 work.

A list of suggested tests was provided by Alain Dabas in a presentation given at L2B/C PM 27 (in January 2013). The tests were partitioned in two types as follows:

- Two chains:
  - Initial: generation of CSR from ISR

  - Routine: update of CSR every time an IRC is done.



Initial chain

settings: Default E2S instrument settings.

Reference: AUX_RBC from ISR generated with noise option off; use of Gaussian Rayleigh spectra.

Tests:

1. Generate ISR (Instrument Spectral Registration) with noise option on and test impact of corresponding AUX_RBC on L2B winds. For this, use of a few standard scenarios (selection of L2B team).
2. Compute AUX_RBC from ISR with a wrong value of FSR (Free Spectral Range) in AUX_PAR_CS and AUX_PAR_RB (note: keep same value in both).

Update chain

settings: Default E2S settings except for an étendue effect (width = 500MHz, no tilt).

Reference: AUX_RBC from CSR updated with IRC_Std (Instrument Response Calibration) with no noise. Also noise-free ISR.

Tests:

3. AUX_RBCs from updated CSR several times with several realizations of IRC_Std_noise (=IRC_Std + noise on) and see the impact on wind products.

4. AUX_RBCs updated CSR several times with several realizations of IRC_Std_noise_aocs (=IRC_Std_noise + AOCS errors).
5. AUX_RBCs from updated CSR several times with several realizations of IRC_Std_noise with noisy AUX_MET data.
6. AUX_RBC from updated CSR several times with IRC data on scene with aerosols, but perfect AUX_MET data.
7. Update CSR several times with IRC data generated with the height-bin overlap option on IRC_Std_noise, and then a scenario with aerosols.

It was agreed at the L2B/C PM29 (October 2013) that not all of these tests are required to complete the CCN2 work. For example, test 7 is not possible until the E2S is updated to include the height-bin overlap option, so it was agreed that this test is not required to complete the CCN2 calibration testing. The latest release of E2S V3.04 (Feb 2014) simulates the height-bin overlap and this will be switched on for the end-to-end testing under CCN5 of this contract. Test 4 including AOCS errors led to some issues in the CSR updater stage which need to be investigated (as reported on L2b PM29), and will also not be tested for CCN2 work. Also, the magnitude of possible AOCS errors currently specified by the E2S is being consolidated with Astrium at the time of writing for v3.0 of this document.

The results from the remaining tests are given in the following sections.

5.1 Test 1: ISR noise

This test investigates how noise in the ISR (instrument spectral registration) calibration mode affects the AUX_RBC_L2 product and hence the L2B Rayleigh winds. **ISR is a calibration mode of Aeolus which measures the Fabry-Perot transmission filters for channels A and B (i.e. transmission versus frequency).** This mode was intended by ESA/Astrium for monitoring the characteristics of the FP spectrometer, and not as it is now used in the generation of Rayleigh calibration inputs for the L2B processor: the ISR is used for the calculation of Rayleigh response calibration curves; in particular the generation of the internal reference Rayleigh response and the atmospheric response calibration look-up tables.

A number of AUX_RBC files derived from independent realisations of the ISR noise were produced. Figure 37 shows two example L2B Rayleigh-clear HLOS mean error (green and blue) versus true HLOS results using AUX_RBC files from two independent realisations of ISR noise (both with 80 mJ laser energy in the E2S wind mode). There is systematic error (relative to truth) in both cases, but one result (green) is notably more positively biased than the other (blue) by a constant offset i.e. independent of wind speed. Similar constant offsets were demonstrated for other noisy ISR realisations. Both sets of L2B winds started from the same L1B product (as simulated from a WVM LITE test scenario, #0103, with 8 BRCs and default noise conditions, E2S had étendue off), but used the two different AUX_RBC files to produce two L2B products to assess the effect of the ISR noise.

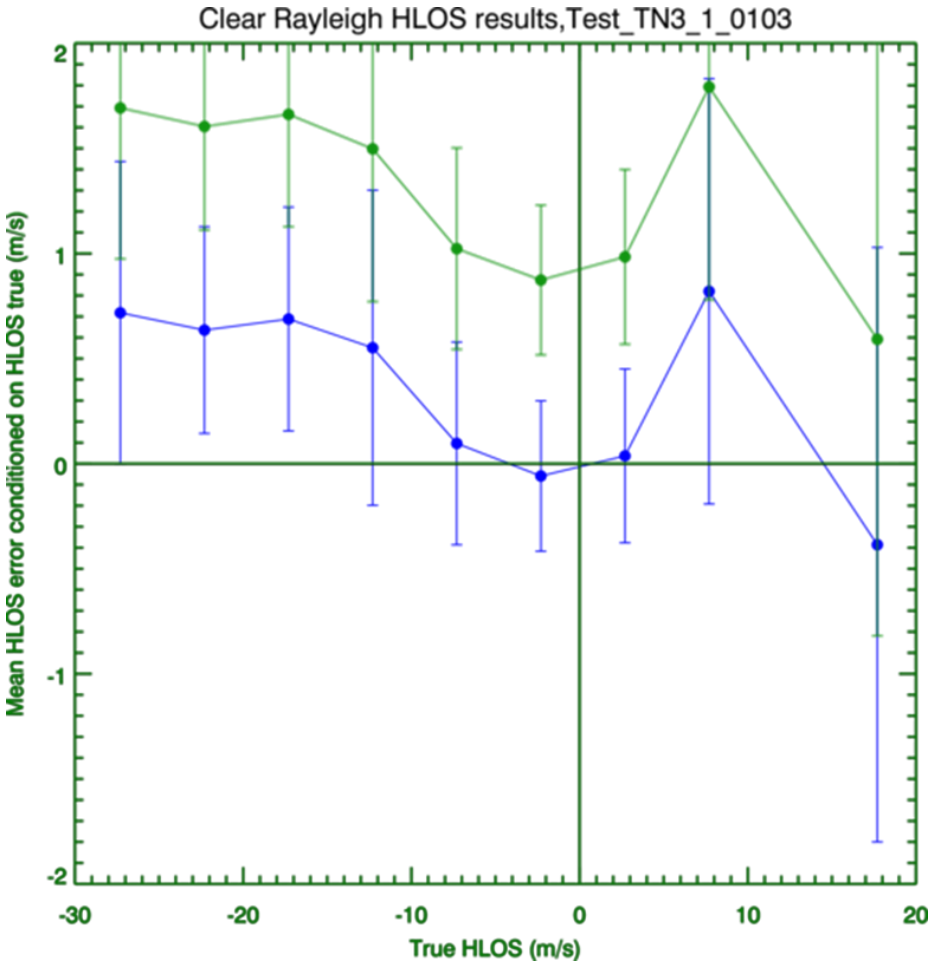


Figure 37. Systematic HLOS errors resulting from two independent AUX_RBC files derived from independent noisy ISRs, used with the same WVM scenario.

The offset between the mean error is ~1 m/s, which is a worrying large bias. The cause of this bias was determined and it is explained in the following paragraphs.

Figure 38 shows the transmission functions from three ISRs (see legend for which is which) over the full spectrometer frequency range. At this large frequency-scale the transmissions functions from different ISR files are indistinguishable.

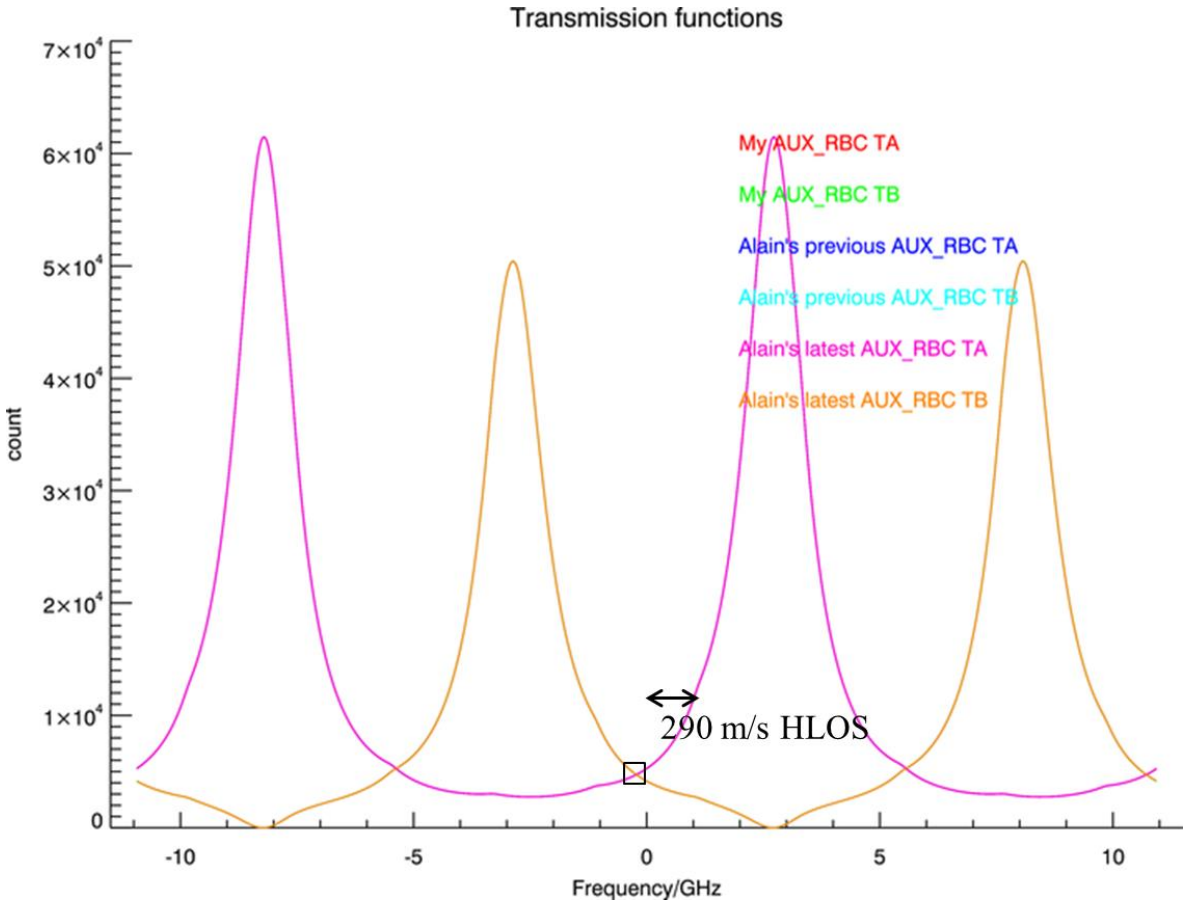


Figure 38. Fabry-Perot transmissions T_A and T_B (in LSB units) as a function of frequency (relative to a reference). The arrow length indicates a frequency spacing equivalent to roughly 290 m/s HLOS wind. The small black rectangle is the zoom area used for Figure 39.

However, if we zoom into the data around the black rectangle of Figure 38 (around the cross-point of filters A and B nearest to 0 GHz frequency offset) we get Figure 39. There are clearly differences between the noisy and noise-free ISRs at this scale. The E2S simulated noise in the ISR is of the following types: Poisson, read-out noise, and frequency variation noise. **The laser energy was set to 110 mJ for this one example.**

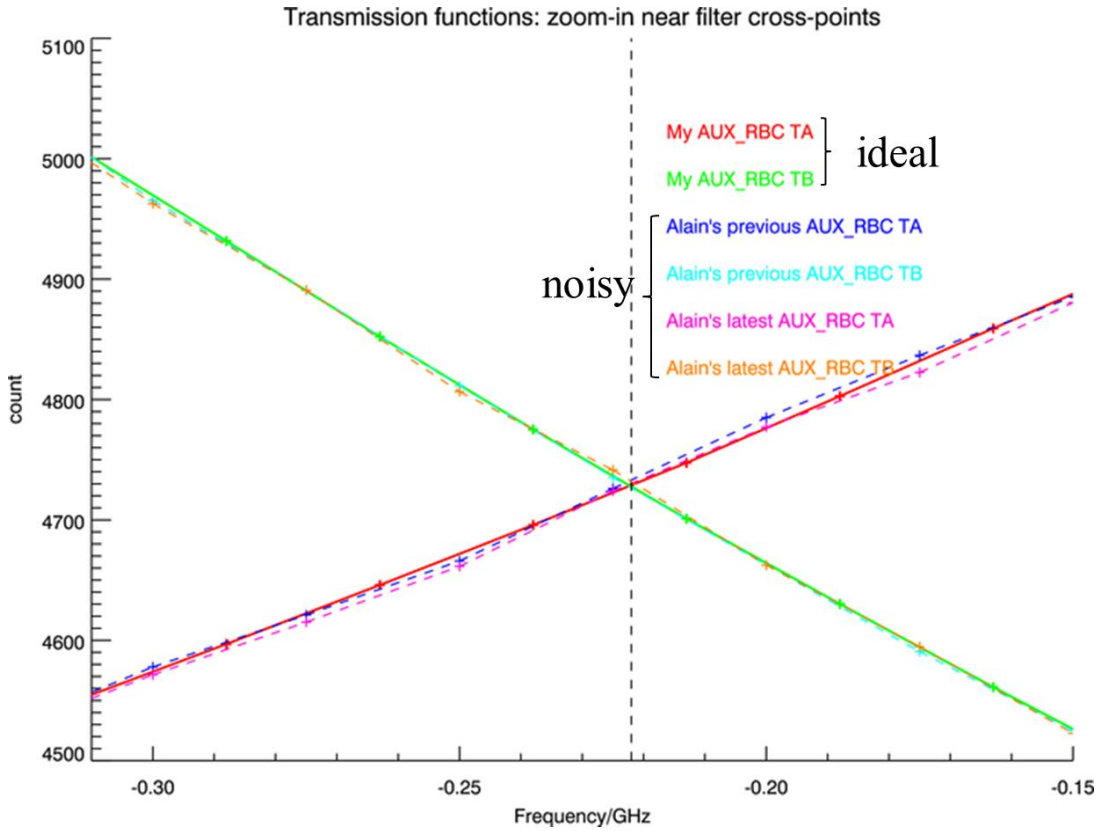


Figure 39. Fabry-Perot transmissions T_A and T_B (in LSB) across a much smaller frequency range. The noisy ISRs can be distinguished from the ideal ISR.

The ISR transmission functions are converted to an internal reference Rayleigh response calibration look-up table by calculating:

$$RR_{\text{int\_}calib}(f) = \frac{T_A(f) - T_B(f)}{T_A(f) + T_B(f)}$$

This assumes the internal reference bandwidth is very small i.e. such that a Dirac delta function can be assumed. Any noise in the transmission functions will be propagated into this look-up table, as demonstrated by Figure 40.

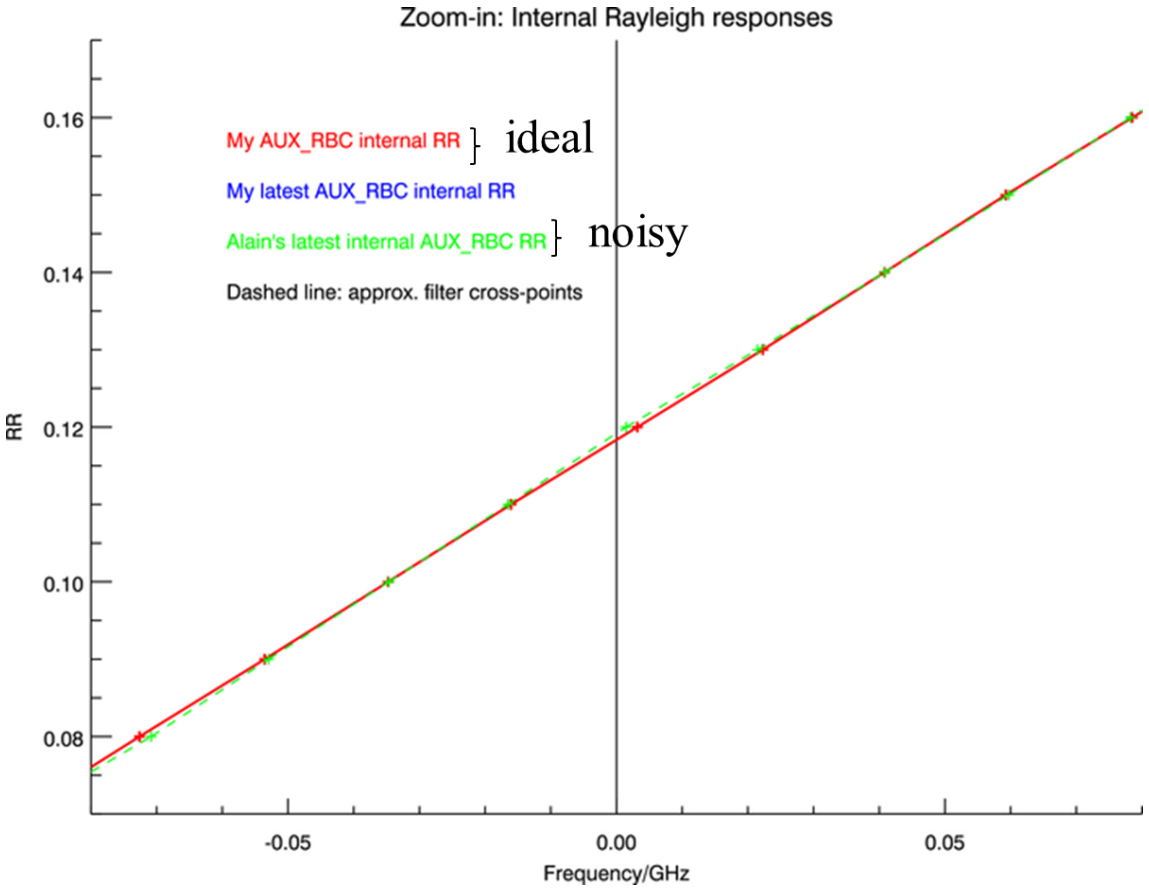


Figure 40. Internal reference calibration: Rayleigh response versus frequency as calculated from the ideal and a noisy ISR.

The apparently small errors of Figure 40 in the noisy internal Rayleigh response calibration curve (when the green differs from red) can translate into large HLOS errors due to the very sensitive relation between HLOS wind error and frequency error:

$$\Delta v_{HLOS} = \frac{\lambda_0 \Delta f}{2\sin\theta} = 291.1\Delta f\,[GHz]$$

This means for a fixed internal RR value a frequency error of 1 MHz (0.001 GHz) from the ideal curve leads to a 0.29 m/s HLOS error. Plotting the implied HLOS error (from this formula) as a function of frequency due to the errors in the internal Rayleigh response leads to Figure 41; the errors look random in nature.

Typically the measured internal Rayleigh response remains constant (for a fixed output laser frequency and fixed spectrometers), therefore a HLOS error of the same value is repeatedly invoked from the internal calibration look-up table which leads to systematic error. For example, if the internal RR is repeatedly measuring 0.118 (green dashed line, see Figure 41) then a positive bias of ~0.5 m/s will result for all Rayleigh winds (independent on HLOS value) — however it could be much lower by chance if e.g. internal RR is 0.4.

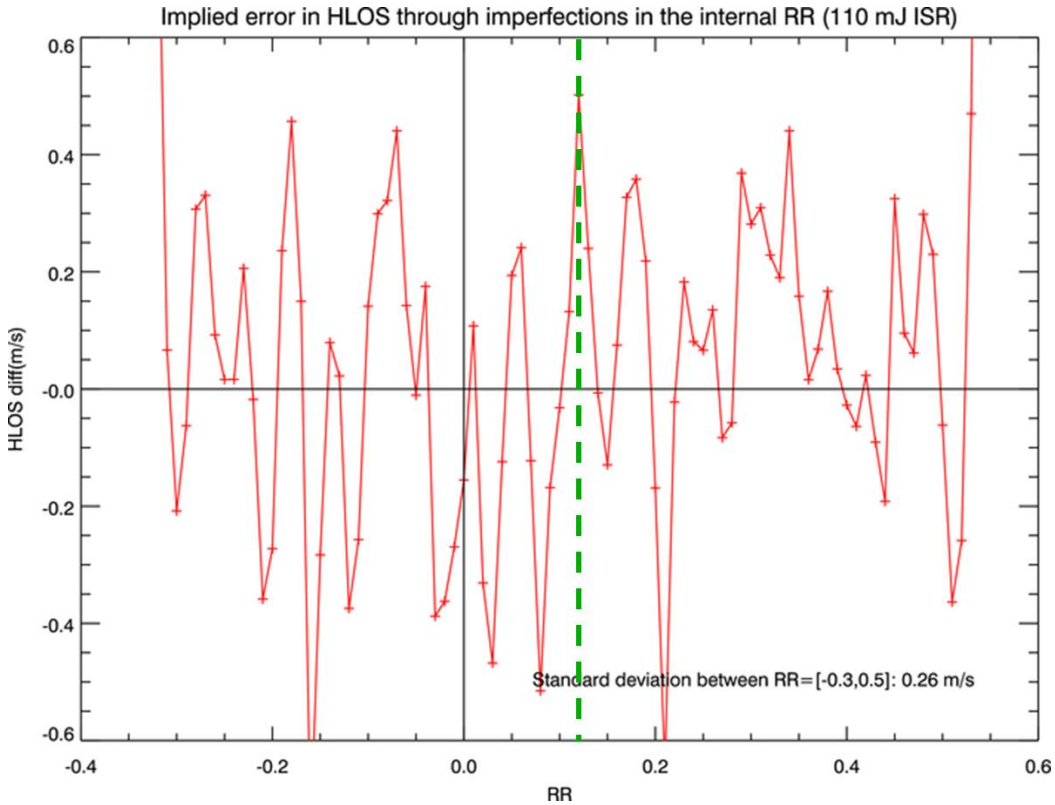


Figure 41. Implied HLOS wind error from one noisy ISR as a function of internal Rayleigh response value.

Clearly, the larger the ISR noise is, the greater the chance of a larger systematic error. Assuming the internal reference RR remains constant, then the bias will change every time a new ISR is used. The operational baseline is that the ISR will be characterized early in-flight, and then repeated when necessary. It is currently assumed that this will happen only a few times during mission lifetime due to the stable environmental conditions in-orbit. Since the bias is independent of wind speed, it should be relatively easy to correct using e.g. NWP model output at the risk of biasing towards any NWP model bias.

Thankfully the noise in the ISR transmission functions **did not affect the atmospheric Rayleigh response calibration** look-up table**.** This is because the integral over frequency of a noisy FP transmission curve multiplied with the broad Rayleigh atmospheric backscatter spectrum (sigma ~1.6 GHz) averages out the noise, which therefore becomes small. The internal reference signal is in comparison to the Rayleigh backscatter a very frequency-narrow function (50 MHz FWHM), and hence does not lead to much averaging of the transmission function noise.

It may be possible to reduce the effect of noise from the ISR by fitting a function to the transmission curves close the appropriate filter cross-point. This will be investigated by the L1B team. Also ESA/Astrium is investigating strategies to improve the signal-to-noise ratio of the ISR.

It was noticed that the level of noise simulated in the ISR by the E2S is smaller than expected. The following shows why it is smaller than expected. The ACCD count (LSB) is obtained from the electron count as shown (given a radiometric gain of 0.429 LSB/e^-):

$$c_{ACCD} = \beta c_{e^-} = 0.429 c_{e^-}$$

The standard error of of the ACCD counts, assuming only Poisson noise on the electron count, can be expressed as follows:

$$\sigma_{c_{ACCD}} = \beta\sigma_{c_e} = \beta\sqrt{c_e} = \sqrt{\beta c_{ACCD}} = 0.65\sqrt{c_{ACCD}}$$

If we insert ACCD counts at ~4700 LSB (close to the filter cross-point), the standard deviation of the noise should be 45 LSB units according to this formula. However, in Figure 39, the noise is smaller than this at around 10-20 LSB. Also this calculation did not consider frequency variation and read-out noise, only Poisson noise, so the E2S error should in fact be larger than 45. This suggests there is an issue to be resolved in the E2S. This is a concern, because the level of internal reference bias due to the noisy ISR is already a significant source of systematic error.

5.2 Test 2: incorrect FSR value

Two AUX_RBC_L2 files were produced with:

- The CSR updater and AUX_RBC are produced assuming a FSR = 10968MHz (+30MHz, larger than the true value).
- The CSR updater and AUX_RBC are produced assuming a FSR = 10908MHz (-30MHz, smaller than the true value).

±30MHz is the magnitude of the uncertainty on the FSR according to Benjamin Witschas (DLR).

The effect on the L2B winds using the same WVM scenario as Test 1:

- On occasional results the Rayleigh winds differ by 0.01 m/s – so this is a negligible source of error and this will not need further investigation.

5.3 Test 3: IRC noise

With the ISR kept the same (a noise free ISR), this test case investigates the effect of noise in the IRC (Instrument Response Calibration). **For the Rayleigh channel calibration the IRC is done with the satellite in nadir pointing mode for atmospheric returns within the range 6-16 km altitude; the result is the Rayleigh Response Calibration (RRC).** The RRC output is used by the AUX_CSR updater to update the ISR (by convolution with a top-hat function) to account for atmospheric path effect (étendue).

The IRCs tested are generated from a standard atmosphere E2S input scenario (i.e. no particulate contamination):

- with a top-hat width of w=500MHz with no tilt.
- The E2S noise sources are: Poisson, read-out, frequency variation.
- AOCS errors are not considered in this test case.

The IRCs were simulated and processed with the latest versions of the E2S (3.03) and L1P (6.02).

The WVM E2S simulation and L1B processing used earlier versions i.e. 3.02 and 6.01 respectively, however this is not an issue, since the Rayleigh spectrometer properties have not changed

with the updated software. The WVM scenario chosen was the first CALIPSO half-orbit test case. It was simulated with étendue "on" with a width of 500 MHz i.e. matching the E2S settings used in generating the calibration data. This scenario has a sufficient range of HLOS values to assess the changes in HLOS dependent mean error.

Three AUX_RBC products were produced, all of which are computed from AUX_CSR products derived by the CSR updater code applied to three different noisy IRCs. They are called version numbers 21, 22 and 23. The Rayleigh-clear HLOS mean error versus true HLOS results are shown in Figure 42. The three noisy IRCs (blue lines) tests have a small positive bias in HLOS relative to **no-noise IRC** (**red dots**, AUX_RBC test 40). The bias is around 0.2 m/s and is small compared to other sources of bias. The bias seems to be a constant offset relative to the no-noise case and hence the gradient of the wind dependent bias does not seem to be affected. This implies that the atmospheric Rayleigh response calibration curve intercept is offset relative to the ideal.

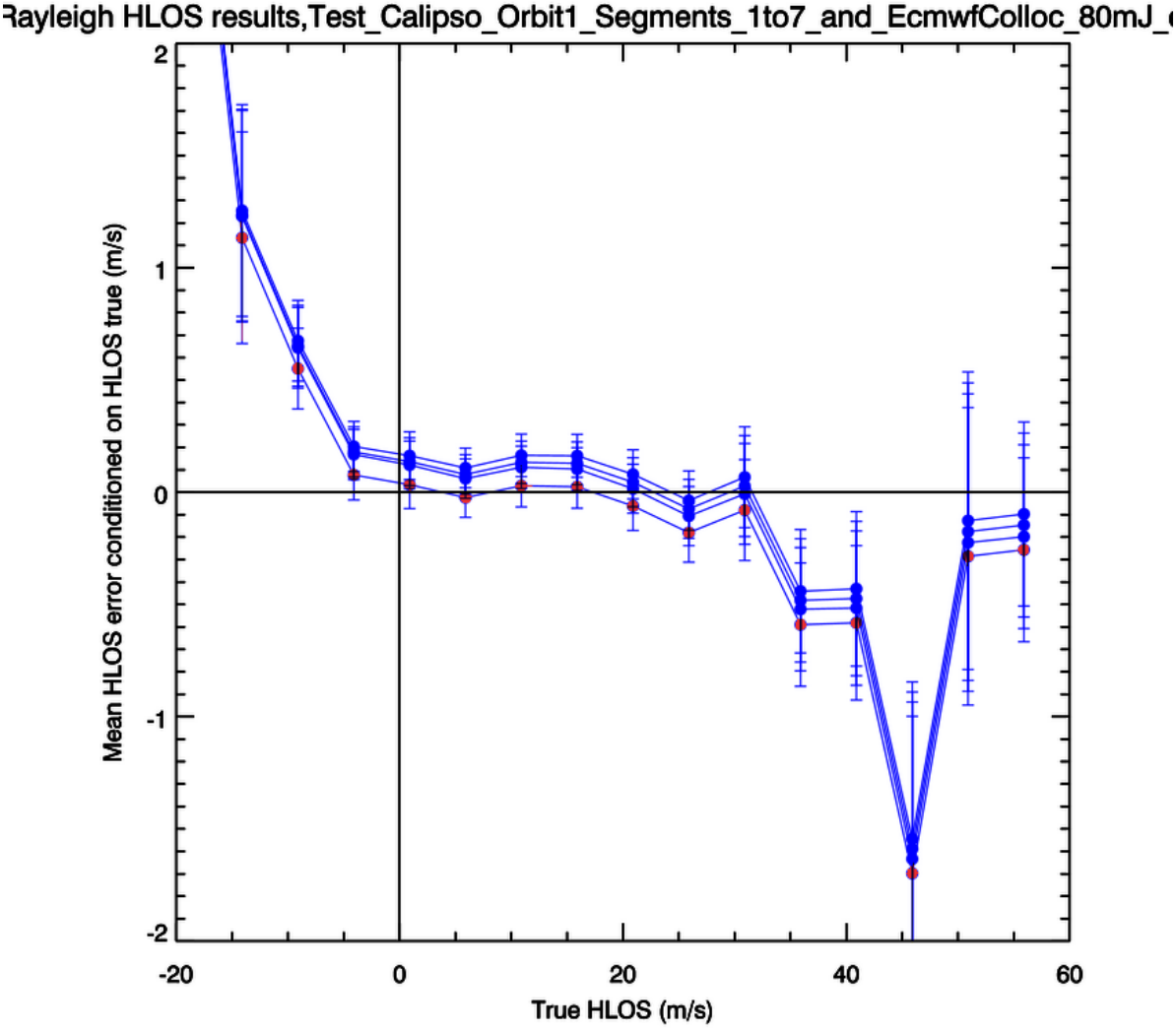


Figure 42. Rayleigh-clear HLOS mean error versus true HLOS curves for the no-noise IRC in red (AUX_RBC file 40) and the three noisy IRCs in blue (AUX_RBC files 21, 22 and 23).

So it would appear that IRC noise is not a major concern in terms of systematic errors, when compared

to e.g. the effect of Dark Current Noise in Memory Zone (at least for the simple atmospheric scenario used here in calibration testing), or the ISR noise.

5.4 Test 5: AUX_MET noise

Six AUX_RBC_L2 products were produced to investigate temperature errors in AUX_MET files. AUX_MET files provide a priori temperature information from an NWP model that are used by the calibration processor (AUX_CSR updater) and so they can affect the AUX_RBC_L2 files i.e. errors in the AUX_MET data will propagate to the AUX_RBC_L2 calibration product.

All the calibration mode scenarios useda standard homogeneous atmosphere without particulate contamination (clouds/aerosol) They were simulated and processed with the E2S v3.03 and L1P v6.02. A top-hat width of 500MHz has been commanded for all IRCs in the E2S.

The WVM E2S simulation and L1B processing used earlier versions i.e. 3.02 and 6.01 respectively compared to the calibration modes. However, this is not an issue since the Rayleigh spectrometer properties have not changed. The pulse energy is 80mJ for the WVM. A top-hat width of 500MHz has been commanded in the E2S in WVM to match the calibration files.

Case 40: The ISR and IRC have no noise. The UpdateCSRfromIRC procedure found that the ISR did not characterize the double FP properly but was able to correct it. It found that the best match with the actual RRC is obtained with a top-hat width = 600MHz and a tilt = 0.0415 (so very close to zero).

Cases 41, 42 and 43: these three cases are similar to the previous, but the photon counting noise is now switched on in the IRC scenarios and random errors were added to the associated temperature profiles. Each case has a different realisation of the IRC and temperature noise. The temperature errors are vertically and horizontally uncorrelated. The vertical resolution is 150 m. The errors follow Gaussian statistics with a zero mean and a standard deviation of 2 K (an upper limit as expected for ECMWF model noise). The UpdateCSRfromISR detected the presence of an étendue effect and found a good correction:

- Case 41: est. width = 626MHz, est. tilt = 0.0484.
- Case 42: est. width = 570MHz, est. tilt = 0.0494.
- Case 43: est. width = 606MHz, est. tilt = 0.0481.

Case 44 and 45: This is identical to the previous ones, but this time temperature errors are constant (i.e. temperature bias): -2K for case 44 and +2K for case 45. This level of bias is larger than expected with ECMWF model profiles (at short-range forecast). The CSR corrections are as follow:

- Case 44: est. width = 790MHz, est. tilt = 0.038.
- Case 45: est. width = 415MHz, est. tilt = 0.0716.

Clearly, temperature biases in AUX_MET profiles, as defined in these two test cases, will have an impact. This is not a surprise. If the temperature in the AUX_MET file is constantly lower than the real temperature, the calibration processor will broaden the transmission curves of the FPs because the

width of the molecular spectra it generates are narrower than in reality.

The WVM scenario chosen to assess the AUX_RBC files was the first CALIPSO half-orbit test case simulated with:

- étendue on
- a width of 500 MHz i.e. matching the E2S settings for generation the calibration data, i.e. the same scenario used in Test 3.

The effect of the IRC noise combined with random temperature errors (from calibration use of AUX_MET, cases 41-43) in terms of L2B wind systematic errors is shown in Figure 43. The bias difference is on the order of 0.2 m/s HLOS, compared to a noise-free bias curve shown in red and appears as a constant offset. This is a very similar result to the IRC noise only (Figure 42). In fact it seems that the Gaussian temperature errors with standard deviation of 2 K do not have a significant effect on the calibration — the difference relative to the no-noise reference (red) is mostly due to the IRC noise only.

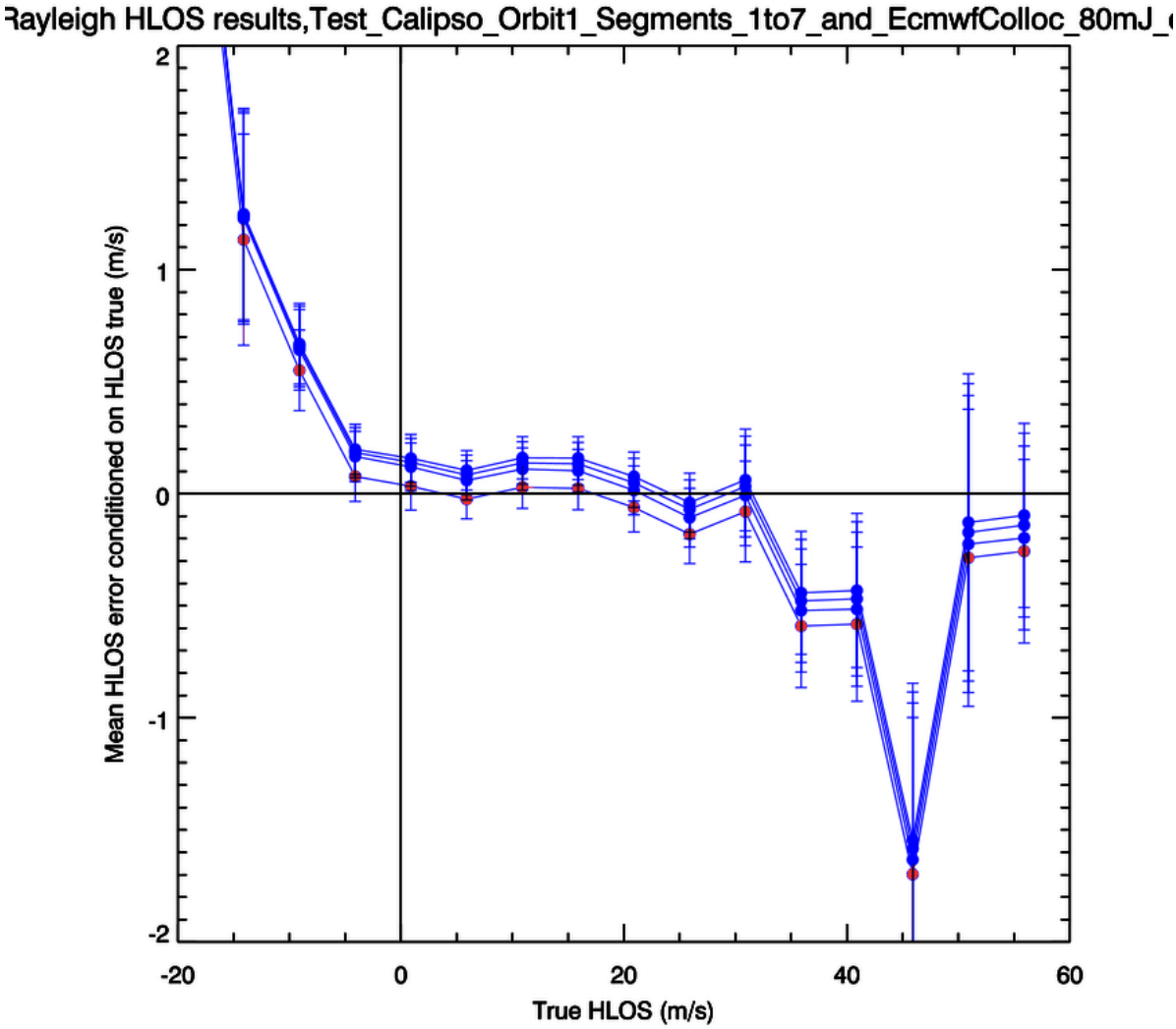


Figure 43. Rayleigh mean error versus true HLOS curves for AUX_RBC file 40 (red) which is without noise and for 41-43 (blue lines) which have IRC noise and AUX_MET profile noise (with standard deviation of 2 K).

The wind results when using the AUX_MET file with temperature bias (cases 44 and 45) in the calibration is shown in Figure 44. The biases have a noticeable effect since it appears to induce a wind speed dependent bias relative to the ideal calibration (red); this reaches 0.4 m/s at the extreme winds. The "slope" bias is of opposite signs for the -2 K and +2 K temperature biases.

In practice a constant bias over the horizontal domain of one RRC (which takes 20 mins i.e. about one quarter of an orbit) in the AUX_MET temperature data is unlikely from the ECMWF model. But large biases can occur in specific weather conditions, e.g. in the winter stratosphere during a stratospheric warming event. ECMWF model biases were assessed (for the period January to March 2014) against radiosonde temperatures and it was found that the biases varied from region to region and from season to season, but the biases rarely exceed 0.4 K and were typically around 0.2 K (this assumes that the radiosondes are not themselves the source of the bias). The random noise level in radiosonde O-B departure statistics is around 1-2 K standard deviation, so the value of 2 K used for the AUX_MET testing is about the upper limit for the expected noise of ECMWF temperature forecasts.

In summary for the temperature error testing, the good news is that it appears that AUX_MET temperature errors apparently lead to reasonably small systematic errors in Rayleigh calibration. Pressure errors were not considered since the E2S did not at the time of the study simulate the effect on the spectrum. This is currently being implemented and will be tested during CCN5 of this contract. Also pressure tends to be of lesser importance than temperature for the Rayleigh-Brillouin spectrum.

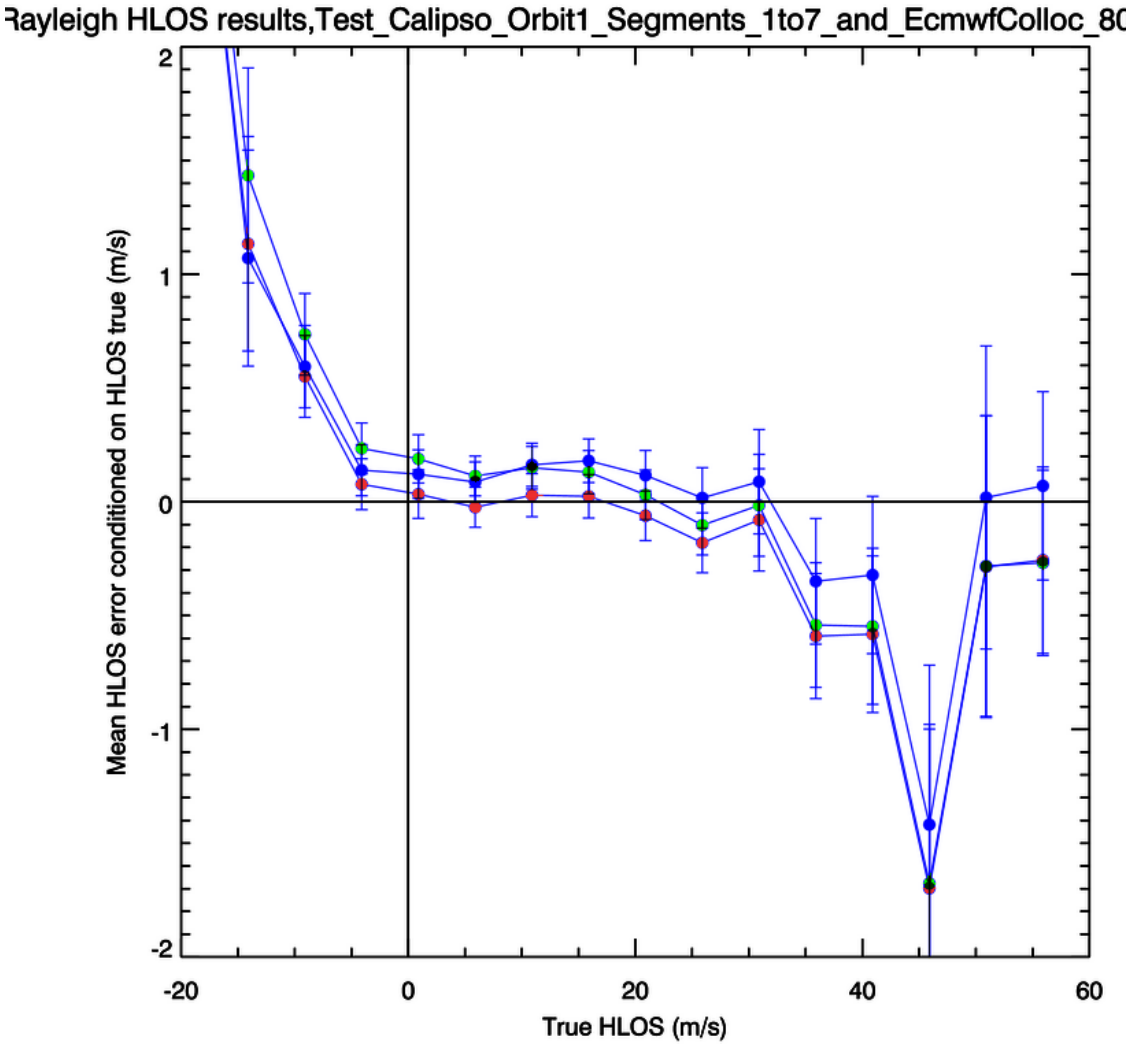


Figure 44. Rayleigh mean error versus true HLOS curves for AUX_RBC file 40 (red) which is without noise and for 44 and 45 (blue and green respectively) which have IRC noise and AUX_MET profile bias (with bias of -2 and +2 K respectively).

5.5 Test 6: Particulate contamination of the calibration scene

One AUX_RBC file was derived using a noise-free ISR and with the IRC (noise on) containing a 1 km thick cloud with a base at 8.5 km in the first half of the scenario followed by clear air. The E2S uses a top-hat width of 500MHz. The IRC has 81 BRCs, with the cloud in the first 40. The cloud backscatter was set to 7×10^{-6} $m^{-1}sr^{-1}$ and the extinction to 0. The backscatter is about two times the molecular backscatter at the same altitudes, so the scattering ratio is about 3. The extinction is unrealistic, but it does not matter for the test, since the cloud is relatively thin and in the lower part of the RRC calibration vertical extent (6-16 km). The calibration processor does not reject cloud contaminated height bins but uses L1B scattering ratio estimates to include the aerosol contribution in the RRC predictions.

The CSR updater detects that the ISR cannot predict the IRC and searches for a tilted top-hat that does it. It returns a top-hat width of 663MHz and a tilt of 0.04.

The effect on the systematic errors (using the same half-orbit WVM scenario as in Test 3 and 5) can be seen in Figure 45. The calibration affected by cloud (and IRC noise) is shown in blue, the red one is the noise-free IRC without cloud. The cloud causes change in the HLOS dependency of the systematic error i.e. changes the "slope" error (which happens to improve the DCMZ dominated bias by coincidence here!). The difference in bias between the ideal and cloud contaminated calibration is however small at around +0.1 m/s at the positive HLOS end of the range.

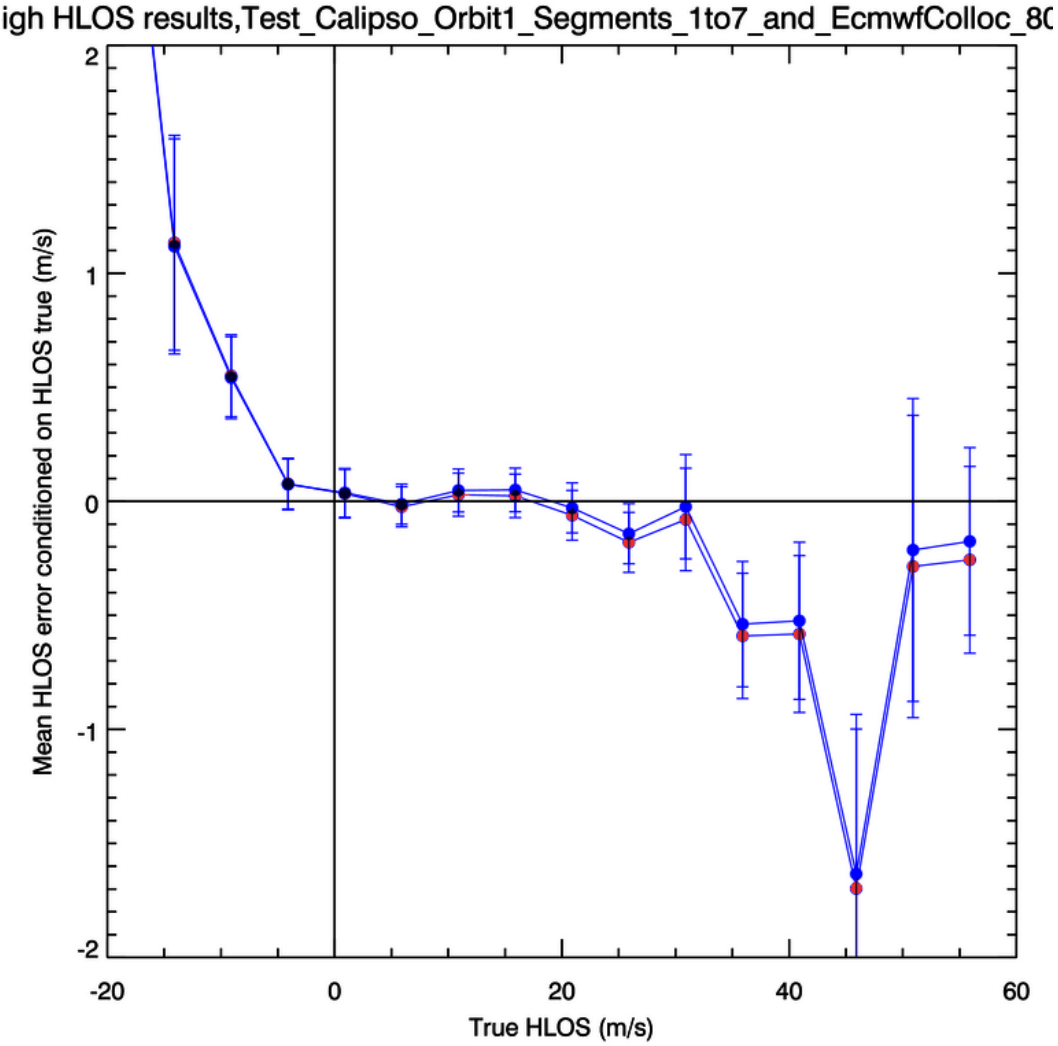


Figure 45. Rayleigh mean error versus true HLOS curves for AUX_RBC file 40 (red) which is the clear scene (without noise) and for 46 (blue) which has IRC noise and contains cloud for half of the calibration scene.

It would appear that (at least for this scenario) **partial cloud contamination** of the RRC is not a major issue for the L2B calibration inputs. However, a real test to be done in CCN5 would be to use more realistic cloud scenarios e.g. based on CALIPSO data or ECMWF model clouds.

5.6 Conclusions fromRayleigh calibration testing

It appears that the most significant source of systematic error resulting from error propagation in the Rayleigh calibration procedure is from ISR noise affecting the internal reference calibration — this

generates bias up to 1 m/s. Noise occurring in the IRC and AUX_MET data is of lesser importance, leading to biases of order 0.2 m/s. Partial cloud contamination of the calibration scene and AUX_MET temperature bias induces "slope" errors which can lead to bias of 0.4 m/s at the extreme range of HLOS winds.

Clearly the internal reference bias is the most pressing issue to resolve given these results, and indeed the L1B team will look into improving this in upcoming contract work. However with some bad luck the other sources of bias could in practice combine to reach a bias level which is unacceptable level for NWP use. A caveat to this research is that only a limited number of fairly idealistic calibration atmospheric scenarios have been tested. More realistic scenarios (with a wide range of scenes) will be investigated as part of CCN5 work on calibration testing.

6 Appendix

6.1 Definition of error statistics used for verification

The HLOS wind error of a L2B HLOS wind observation (label *ij*, where *i* refers to horizontal position, and *j* to vertical range-bin) is defined to be:

$$\varepsilon_{ij} = o_{ij} - t_{ij} \qquad \textbf{(eq. 3)}$$

where *o* is the L2B HLOS wind observation, and *t* is the E2S input HLOS wind (*t* for truth). The true HLOS wind is chosen to be the input E2S wind which is closest to the centre-of-gravity geolocation of the L2B wind observation (in practice a closest match in time and altitude is a sufficient criterion).

Such a definition of truth is suitable when assuming that Aeolus is to be interpreted as an in-situ point-measurement wind. One can be fairer to Aeolus by taking into account its resolution, i.e. by averaging the true wind field in a similar manner to the way Aeolus is constructed, both in the vertical and the horizontal. For the majority of test cases in this technical note, where the E2S input wind does not vary significantly over the resolution "cell" of Aeolus, the improved representation of truth is not required. It was found that with the CALIPSO test cases (using 25 km ECMWF horizontal grid fields) that an almost negligible improvement in the standard deviation of error could be obtained using the averaged truth (E2S winds). As the NWP model resolution increases relative to the Aeolus observation resolution then this effect will be become more noticeable.

The sample mean HLOS wind error on a given range-bin is given by:

$$\bar{\varepsilon}_j = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{ij} \qquad \textbf{(eq. 4)}$$

where *N* is the number of observations available on range-bin *j* (i.e. the sample size). The sample standard deviation on a given range-bin is:

$$\sigma[\varepsilon]_j = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\varepsilon_{ij} - \bar{\varepsilon}_j)^2} \qquad \textbf{(eq. 5)}$$

The standard error of the mean is estimated as follows:

$$\bar{\varepsilon}_j \pm \frac{\sigma[\varepsilon]_j}{\sqrt{N}} \qquad \textbf{(eq. 6)}$$

6.2 Estimating horizontal error covariance/correlation statistics

Considering only HLOS wind errors on a given range-bin, a matrix of observation errors (like in equation 4) **E** can be formed from *n* HLOS observations (e.g. number of observations on a given range-bin) and *m* random samples of the observations (i.e. each sample is a run of the chain of processors to obtain a randomly generated sample):

$$\mathbf{E} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & .. & \varepsilon_{1n} \\ \varepsilon_{21} & \varepsilon_{22} & .. & \varepsilon_{2n} \\ : & : & .. & : \\ \varepsilon_{m1} & \varepsilon_{m2} & .. & \varepsilon_{mn} \end{bmatrix} \qquad \textbf{(eq. 7)}$$

The vector of the means of the observation errors is:

$$\bar{\boldsymbol{\varepsilon}} = \begin{bmatrix} \overline{\varepsilon_1} & \overline{\varepsilon_2} & .. & \overline{\varepsilon_n} \end{bmatrix} \qquad \textbf{(eq. 8)}$$

The means are subtracted from each column of **E** to give the matrix **F**:

$$\mathbf{F} = \begin{bmatrix} \varepsilon_{11} - \overline{\varepsilon_1} & \varepsilon_{12} - \overline{\varepsilon_2} & .. & \varepsilon_{1n} - \overline{\varepsilon_n} \\ \varepsilon_{21} - \overline{\varepsilon_1} & \varepsilon_{22} - \overline{\varepsilon_2} & .. & \varepsilon_{2n} - \overline{\varepsilon_n} \\ : & : & .. & : \\ \varepsilon_{m1} - \overline{\varepsilon_1} & \varepsilon_{m2} - \overline{\varepsilon_2} & .. & \varepsilon_{mn} - \overline{\varepsilon_n} \end{bmatrix} \qquad \textbf{(eq. 9)}$$

The sample observation error covariance matrix **S** (of size *n* by *n*) can be obtained by:

$$\mathbf{S} = \frac{1}{m-1}\mathbf{F}\mathbf{F}^T \qquad \textbf{(eq. 10)}$$

This is an estimate of the population error covariance matrix. The sample correlation matrix **C** can be calculated from the sample covariance matrix **S** by:

$$C_{ij} = S_{ij} / \sqrt{S_{ii}}\sqrt{S_{jj}} \qquad \textbf{(eq. 11)}$$

The correlation matrix gives the observation error correlations between an observation and its neighbouring (adjacent, adjacent but one, etc.) observations.

6.3 Aeolus noise at the measurement level

Considering an Aeolus measurement (e.g. number of signal electrons from accumulations at the 2.85 km horizontal scale), the number of signal electrons (n) will be affected by Poisson noise as follows:

$$n \pm \sigma_n = n \pm \sqrt{n} \qquad \textbf{(eq. 12)}$$

Next we consider the conversion from electrons to LSB (least significant bit, number of ACCD counts) and the so called "radiometric gain".

The ACCD count (LSB) is obtained from the electron count as shown (given a radiometric gain β of e.g. 0.429 LSB/e^-):

$$c_{ACCD} = \beta c_{e^-} = 0.429 c_{e^-}$$

The standard error of the ACCD counts, assuming only Poisson noise on the electron count, can be expressed as follows:

$$\sigma_{c_{ACCD}} = \beta \sigma_{c_e} = \beta \sqrt{c_e} = \sqrt{\beta c_{ACCD}} = 0.65 \sqrt{c_{ACCD}}$$

The relative error (inverse of SNR) on the signal ACCD counts is:

$$\frac{\sigma_{c_{ACCD}}}{c_{ACCD}} = \frac{\sqrt{\beta}}{\sqrt{c_{ACCD}}} = \frac{\sqrt{\beta}}{\sqrt{CE_{laser}}} \qquad \textbf{(eq. 13)}$$

In the above equation it is assumed that the number of ACCD counts is proportional to the LIDAR laser energy E_{laser}.

The HLOS wind retrieval is a function of the spectrometer counts. Therefore, the HLOS wind relative error has the same functional form as above if the wind retrieval is a linear function of the signal electron count (probably valid for Rayleigh, less so for the Mie). Therefore HLOS wind relative error:

$$\frac{\sigma_{HLOS}}{HLOS} \propto \frac{1}{\sqrt{E_{laser}}} \qquad \textbf{(eq. 14)}$$

So, for example, if the laser energy drops from 120 to 80 mJ then the relative error changes by a factor:

$$\frac{\sqrt{E_{laser}}}{\sqrt{E_{laser\_new}}} = \sqrt{\frac{3}{2}} = 1.22 \qquad \textbf{(eq. 15)}$$

That is, the expected standard deviation of HLOS random error increases by 22% going from 110 to 80 mJ.

6.4 Aeolus noise on observations

An observation is effectively a mean of the measurement level spectrometer counts, followed by the HLOS wind retrieval.

Defining the following:

o = observation level spectrometer count
m= measurement level spectrometer count
t=true measurement level spectrometer count
ε=error on measurement spectrometer count

The observation level spectrometer count is a mean of the measurement scale spectrometer counts. The measurement error can be propagated to the observation error as follows:

$$o = \frac{1}{N}\sum_{i=1}^{N} m_i = \langle \underline{m} \rangle = \frac{1}{N}\sum_{i=1}^{N}(t_i + \varepsilon_i)$$

$$\mathrm{cov}(\underline{\varepsilon}) = \mathrm{cov}(\underline{m} - \underline{t}) = \sigma_{ij}^2$$

$$\sigma_o^2 = \left(\frac{\partial o}{\partial \underline{m}}\right)\mathrm{cov}(\underline{\varepsilon})\left(\frac{\partial o}{\partial \underline{m}}\right)^T$$

The observation error variance is a mapping of measurement error covariance by the observation operator tangent-linear model. For example, see below for N=3 (observation made of 3 measurements).

$$\sigma_o^2 = \begin{pmatrix} \frac{\partial o}{\partial m_1} & \frac{\partial o}{\partial m_2} & \frac{\partial o}{\partial m_3} \end{pmatrix} \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_{33}^2 \end{pmatrix} \begin{pmatrix} \frac{\partial o}{\partial m_1} \\ \frac{\partial o}{\partial m_2} \\ \frac{\partial o}{\partial m_3} \end{pmatrix}$$

$$\sigma_o^2 = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \frac{1}{N} \end{pmatrix} \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_{33}^2 \end{pmatrix} \begin{pmatrix} \frac{1}{N} \\ \frac{1}{N} \\ \frac{1}{N} \end{pmatrix} = \frac{1}{N^2}\sum_{i,j=1}^{N}\sigma_{ij}^2$$

If all measurement errors are equal and if no error correlations exist then the observation variance is:

$$\sigma_{ii}^2 = \sigma_m^2 \;\&\; \sigma_{ij}^2 = 0$$

$$\therefore \sigma_o^2 = \frac{1}{N}\sigma_m^2$$

This is the familiar 1/√N drop in the standard deviation of error from super-obbing (averaging). If all measurement errors are equal, but they have error correlations then extra terms come into the equation

which increases the variance relative to the previous formula as follows:

$$\sigma_{ii}^2 = \sigma_m^2 \,\&\, \sigma_{ij}^2 \neq 0$$
$$\therefore \sigma_o^2 = \frac{1}{N}\sigma_m^2 + \frac{1}{N^2}\sum_{i,j=1(i\neq j)}^{N}\sigma_{ij}^2$$

6.5 HLOS error dependence on Rayleigh signal strength

The section derives how the HLOS wind error is related to signal strength for the Rayleigh channel. The LOS wind velocity depends on the Doppler-shifted frequency (*f*) as follows:

$$v_{LOS} = \frac{\lambda_0 (f - f_0)}{2} \qquad \textbf{(eq. 16)}$$

The Rayleigh response can be approximated as a linear function of the Doppler shifted frequency:

$$R = \alpha f + d \qquad \textbf{(eq. 17)}$$

Therefore the error variance in LOS wind space can be written:

$$v_{LOS} = \frac{\lambda_0 (R - R_0)}{2\alpha} \qquad \textbf{(eq. 18)}$$

$$\sigma_{v_{LOS}}{}^2 = \left|\frac{\partial v_{LOS}}{\partial R}\right|^2 \sigma_R{}^2 = \left(\frac{\lambda_0}{2\alpha}\right)^2 \sigma_R{}^2 \qquad \textbf{(eq. 19)}$$

The Rayleigh response is calculated from the useful signal A and B as:

$$R = \frac{A - B}{A + B} \qquad \textbf{(eq. 20)}$$

Therefore the error variance of the Rayleigh response is:

$$\sigma_R{}^2 = \left|\frac{\partial R}{\partial A}\right|^2 \sigma_A{}^2 + \left|\frac{\partial R}{\partial B}\right|^2 \sigma_B{}^2 = \frac{4(B^2 \sigma_A{}^2 + A^2 \sigma_B{}^2)}{(A + B)^4} \qquad \textbf{(eq. 21)}$$

Making the approximation A=B, $\sigma_A = \sigma_B$ and with Poisson noise $\sigma_A{}^2$=A then the HLOS wind error standard deviation is inversely proportional to the size of the signal on the Fabry-Perot:

$$\sigma_R{}^2 = \frac{\sigma_A{}^2}{2A^2} = \frac{1}{2A} \qquad \textbf{(eq. 22)}$$

$$\sigma_{v_{HLOS}} \propto \frac{1}{\sqrt{A}} \qquad \textbf{(eq. 23)}$$

6.6 L1B scattering ratio

To produce good quality CM L2B Rayleigh-clear wind observations, it was found that the L1B "refined" scattering ratio (SR) values should be used for L2B classification and Rayleigh cross-talk correction, rather than the L1B nominal SR. The L1B nominal SR was apparently used successfully in the past for the BM chain of processors.

The L1B nominal SR estimates produced by the L1B processor v6.01 (with the default L1B settings, not the new ones suggested in anomaly report AE-IPF-149) are biased too large for scenes where they should be 1.0 (i.e. clear scenes, where good Rayleigh results are expected). An ARTs anomaly report: AE-IPF-149 was created on this issue. The problem becomes evident whenever realistic noise is switched "on" in the E2S.

When the E2S input SR (true SR) is very close to 1.0 (i.e. zero or only small amounts of particulate backscatter), the L1Bp is estimating the nominal SRs to be typically 1.5-2.0. This is shown in Figure 46a): a plot comparing the true SR to the L2B estimated SR (the observation-scale averages of the L1B SR on measurements) used to produce Rayleigh-clear winds. This test case was the second CALIPSO half-orbit (see section 2.3). Note that the L2B Rayleigh-clear used SR values are cut-off above ~2 for this figure because the classification selects them to be used in the Rayleigh-cloudy observations (according to the chosen L2B parameters). The L1B refined SR distribution is shown in Figure 46b); the refined SR is almost unbiased when the true estimates are close to 1.0 (the vast majority of results, and the ones of concern for Rayleigh-clear winds). However the results are poor for larger values (which are of lesser concern for the Rayleigh cross-talk correction, since good Mie results should be possible in replacement).

Biased SR values create bias in the L2B Rayleigh-clear winds, since the Rayleigh-Brillouin correction (ILIAD) of the L2Bp uses the SR estimate to try to correct for small amounts of Mie contamination in the Rayleigh signal (cross-talk).

The HLOS bias when using the nominal L1B SR is about -2 m/s for the Rayleigh-clear winds. This is shown in Figure 47 a) which plots the Rayleigh-clear HLOS error against the SR error (estimate minus truth). There is a cluster of results with SR errors of around +0.3-0.8, which have a corresponding HLOS bias around -2 m/s, which tends to dominate the overall bias seen in the error statistics vs. altitude plots. Figure 47 b) shows the results when using the refined L1B SR: the SR estimates are much better (although there is a tendency for small positive SR bias), and hence the HLOS bias is small. So there is a clear improvement in the L2B Rayleigh-clear winds by switching to the L1B refined SR, hence why it has been the chosen setting in this technical note.

Remember that also the L2Bp uses the SR to differentiate between clear and cloudy measurements (in a procedure called classification). The poor quality L1B nominal SR leads to too many Rayleigh-cloudy measurements (since the SR values are too high) and hence too many Rayleigh-cloudy winds are produced. Also there is degradation in the Mie-cloudy results, since they are trying to retrieve a wind from signals that in reality do not have a Mie peak.

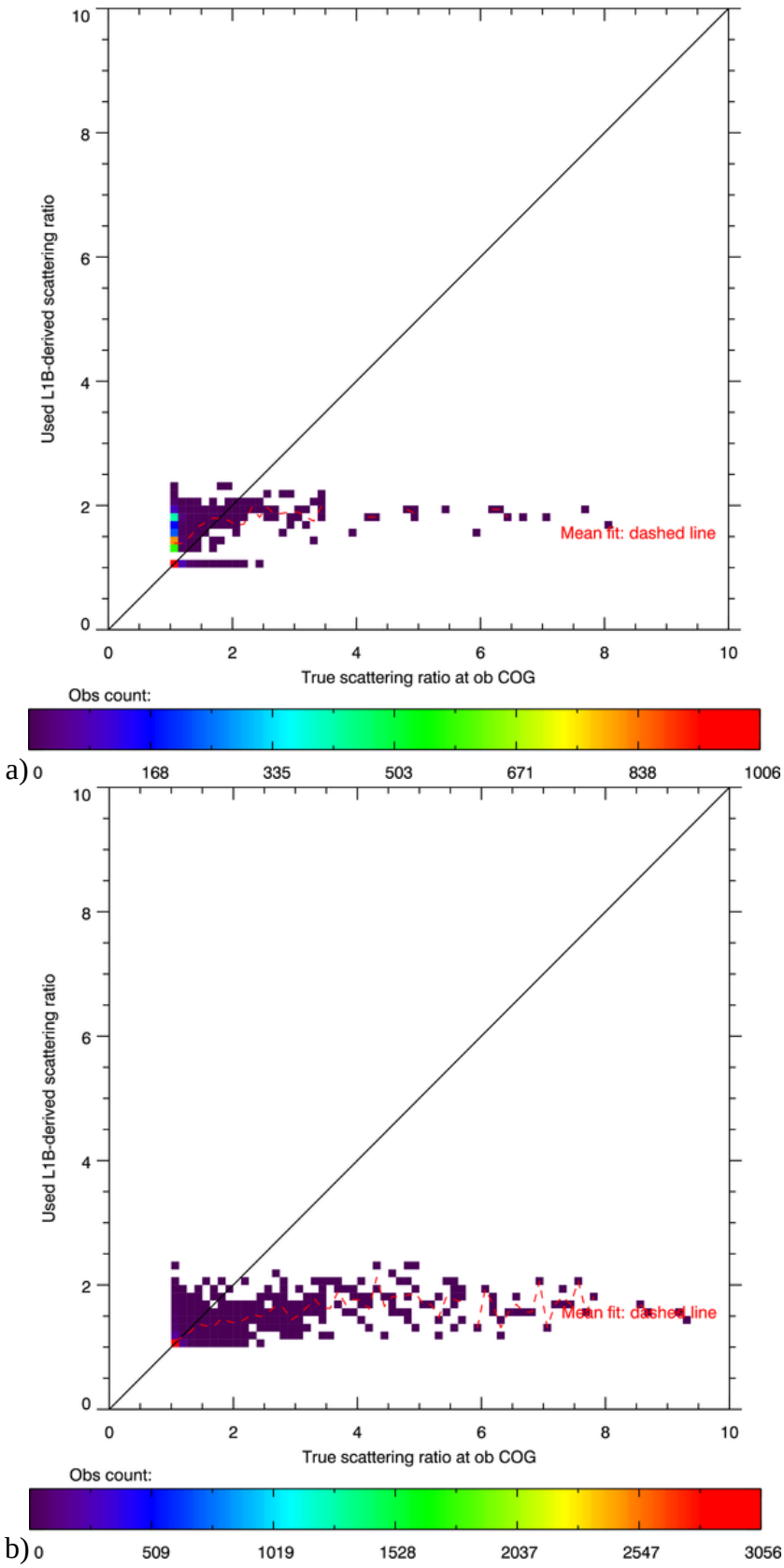


Figure 46. Density plots of used L1B scattering ratio versus the true scattering ratio. a) For the nominal L1B SR b) For the refined L1B SR. The red-dashed line shows the mean of the y given x.

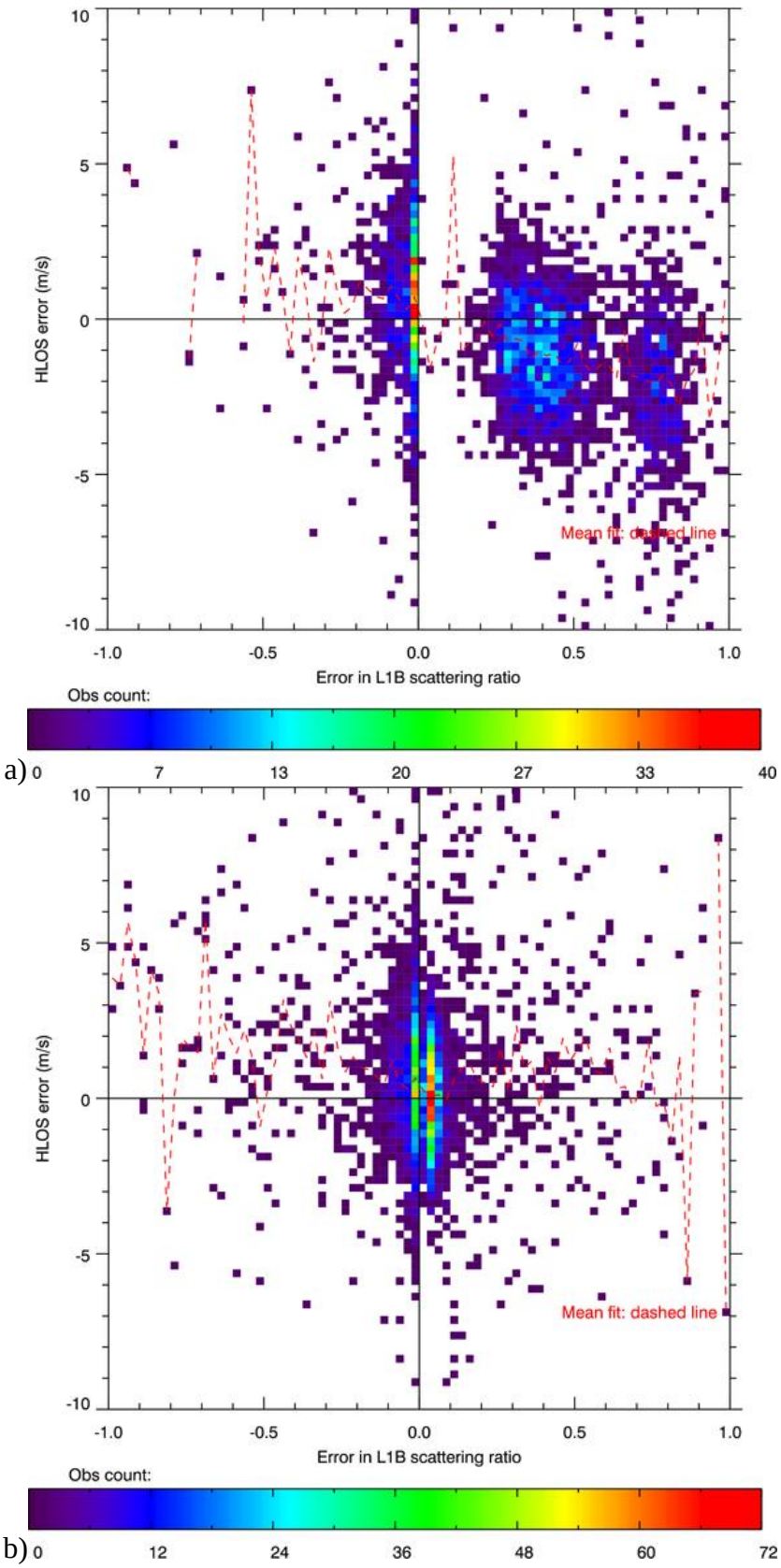


Figure 47. Density plots showing how the Rayleigh-clear HLOS error depends on the error in the L1B SR estimate. a) For L1B nominal SR b) for L1B refined SR. The red-dashed line shows the mean of the y given x.

Following the presentation by Alain Dabas (Météo-France) in L1B PM 28 (25/04/2013), the sensitivity of the HLOS winds to the scattering ratio can be investigated as follows.

The atmospheric backscatter intensity spectrum is the addition of the molecular backscatter spectrum (I_m, Rayleigh-Brillouin scattering) and the particulate backscatter spectrum (I_p, Mie scattering) as follows:

$$I(f, f_d) = \beta_p I_p(f, f_d) + \beta_m I_m(f, f_d) = \beta_m \left[I_m(f, f_d) + (\rho - 1) I_p(f, f_d) \right] \quad \textbf{(eq. 24)}$$

This is a function of the frequency (f) over which the spectrum exists and the Doppler shift frequency (f_d) coming from the atmospheric wind. The spectra are normalised such that:

$$\int_{-\infty}^{\infty} I_p(f, f_d) df = \int_{-\infty}^{\infty} I_m(f, f_d) df = 1 \quad \textbf{(eq. 25)}$$

The SR is given by:

$$\rho = 1 + \frac{\beta_p}{\beta_m} \quad \textbf{(eq. 26)}$$

Since the Mie backscatter spectrum has a relatively narrow band-width compared to the Rayleigh, it can be approximated as a Dirac delta function:

$$I_p(f, f_d) \approx \delta(f_d) \quad \textbf{(eq. 27)}$$

The spectrometer count on Fabry-Perot channels A and B can then be written:

$$\begin{aligned} A, B &= \int_{-\infty}^{\infty} T_{A,B}(f) I(f, f_d) df = \int_{-\infty}^{\infty} T_{A,B}(f) \beta_m \left[I_m(f, f_d) + (\rho - 1) I_p(f, f_d) \right] df \\ A, B &= \beta_m \left[\int_{-\infty}^{\infty} T_{A,B}(f) I_m(f, f_d) df + (\rho - 1) T_{A,B}(f_d) \right] \end{aligned} \quad \textbf{(eq. 28)}$$

where T_A and T_B are the channel transmission functions.

The Rayleigh response is:

$$R = \frac{A - B}{A + B} = \frac{(A_m + A_p) - (B_m + B_p)}{(A_m + A_p) + (B_m + B_p)} \quad \textbf{(eq. 29)}$$

where:

$$A_m, B_m = \int_{-\infty}^{\infty} T_{A,B}(f) I_m(f, f_d) df \quad \textbf{(eq. 30)}$$

$$A_p, B_p = (\rho - 1) T_{A,B}(f_d) \quad \textbf{(eq. 31)}$$

Dropping the indication of dependence on f_d for clarity, it can be shown that the partial derivative of the Rayleigh response with respect to SR is:

$$\frac{\partial R}{\partial \rho} = \frac{2(T_A B_m - T_B A_m)}{\left(A_p + B_p + A_m + B_m\right)^2} \quad \textbf{(eq. 32)}$$

When ρ=1, the change in R following a change in SR is:

$$\Delta R\big|_{\rho=1} = \frac{2(T_A B_m - T_B A_m)}{(A_m + B_m)^2}\Delta\rho \qquad \textbf{(eq. 33)}$$

Note that eq. 24 can be rearranged to give the following equation when ρ=1:

$$\left.\frac{\partial R}{\partial \rho}\right|_{\rho=1} = \frac{T_A - T_B - R_{\rho=1}(T_A - T_B)}{A_m + B_m} \qquad \textbf{(eq. 34)}$$

which is very similar to the formula of the derivative used in the L2Bp cross-talk correction, see [RD2].

The behaviour with the CM Fabry-Perot transmission functions (which are shown in Figure 48) can be used in a sensitivity calculation into the size of the HLOS error due to SR error. Figure 49 shows the sensitivity of Rayleigh response to SR for an atmospheric temperature of 200 K (defining the width of the Rayleigh spectrum) and a HLOS wind speed of 50 m/s (defining the f_d). The corresponding change in HLOS wind with the SR (given the RR slope is ~0.6 GHz^{-1}, as shown in Figure 48) is shown in Figure 49. The magnitude of the HLOS error (effectively a bias if the SR is systematically wrong) is the same magnitude as that seen in the L2B processor when using the L1B nominal SR.

Since the behaviour is roughly linear, the following approximation (only applicable at HLOS wind=50 m/s and temperature=200K) can be made for the HLOS error (in units: m/s) dependence on SR error:

$$\varepsilon_{HLOS} \approx -5.6\Delta\rho \qquad \textbf{(eq. 35)}$$

So if the SR is systematically positively biased by 0.5, then a Rayleigh HLOS wind bias of around -2.8 m/s results. The size of the coefficient (e.g. -5.6) is larger for lower atmospheric temperatures, as the *R* curve becomes a steeper function of frequency (more sensitive). The example above is therefore at the worse-case end of the possible errors (since 200 K is at the low end of the atmospheric temperature range).

Note also, that the HLOS error becomes negligible if a wind of +50 m/s is selected, which apparently is because the transmission functions (T_A and T_B) cross around that point (equal at that corresponding frequency, actually this happens at closer to +40 m/s), meaning the Mie peak adds similar counts for the A and B channels around that point, whereas at -50 m/s the A has a much larger transmission for the B for a Mie peak. Similar characteristics were shown with the VAMP study [RD9], but with smaller biases (perhaps due to the chosen HLOS value or differing spectrometer characteristics back then).

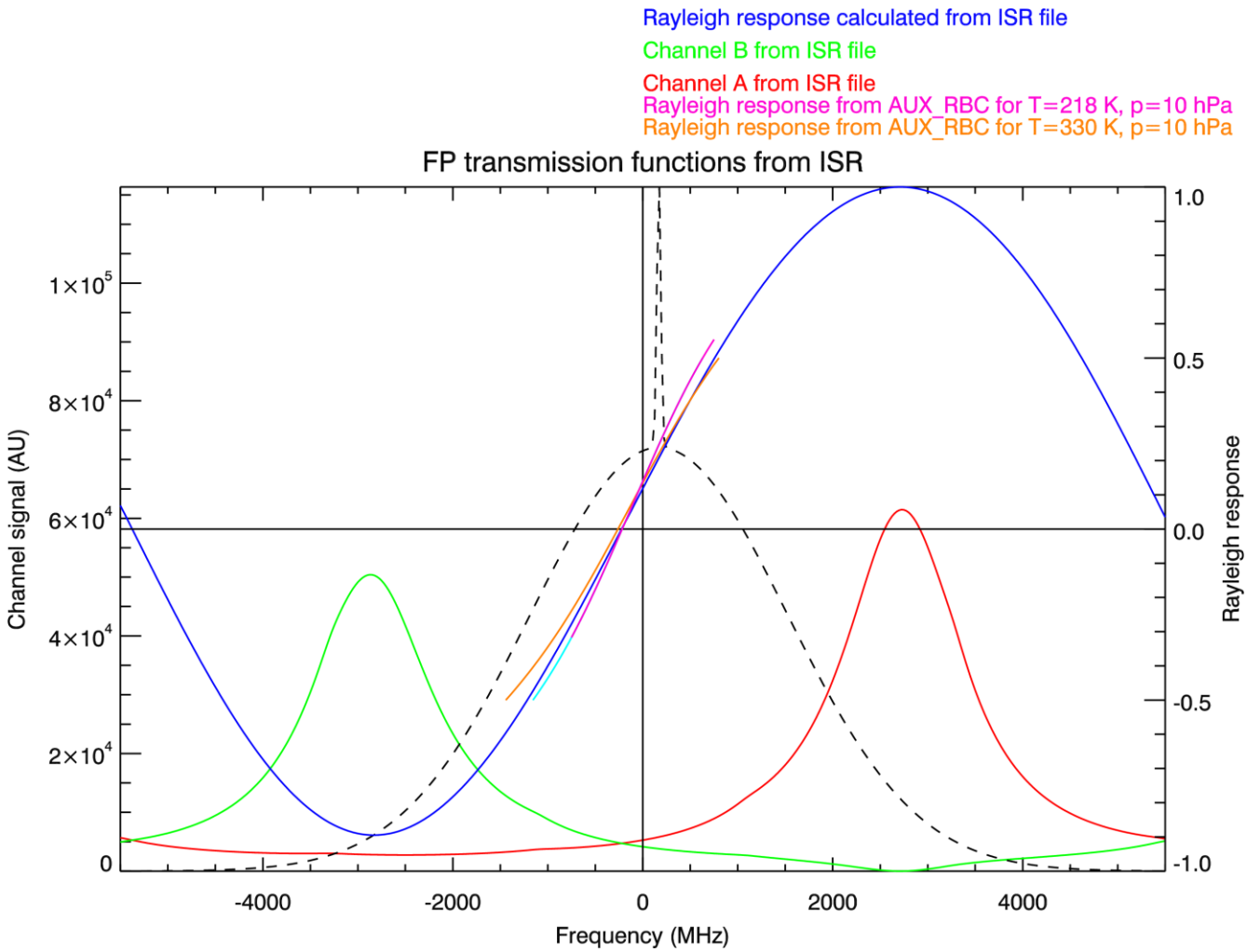


Figure 48. T_A (red) and T_B (green) as a function of frequency. The Rayleigh response given a delta-function input is shown in blue. The other (pink, orange) curves show example RR curves from the AUX_RBC file. An example atmospheric backscatter spectrum is shown by the dashed line (Rayleigh + Mie peak) with a -50 m/s HLOS wind.

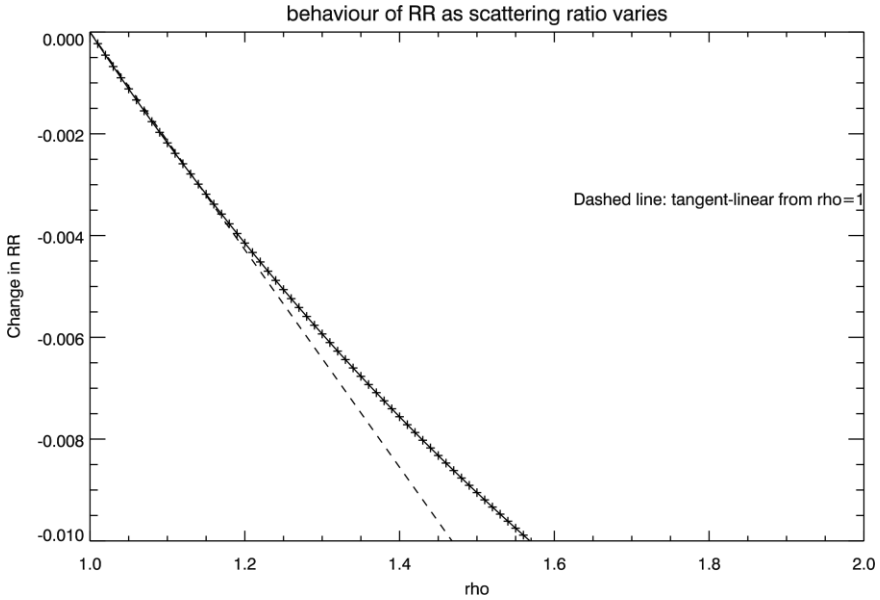


Figure 49. How the Rayleigh response changes as the Mie spectrum is increased via rho=scattering ratio increasing. The crosses are results from numerical calculations, and the dashed line is the tangent-linear model from rho=1.

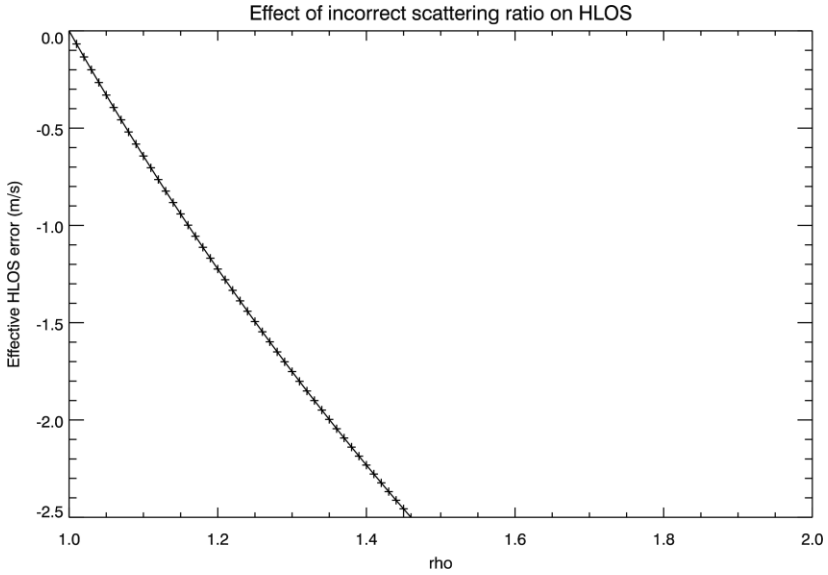


Figure 50. The approximate change in HLOS wind (HLOS error) as the scattering ratio changes (is in error).

6.7 Derivation of optimal placement of the Rayleigh wind

Figure 51 shows the optimal placement (orange line) of a Rayleigh wind within a 2 km range-bin (i.e. the upper Rayleigh bins) as a function of altitude. It is optimal in the sense that it should reduce the bias when comparing the Rayleigh wind to the true point-like wind given linear vertical wind shear.

This information was derived by assuming that an Aeolus Rayleigh wind represents the normalized, integrated, weighted true wind across the altitude range (z) of the range-bin, that is:

$$u_{aeol} = \frac{\int_{z_1}^{z_2} w(z) u_{true}(z) dz}{\int_{z_1}^{z_2} w(z) dz} \qquad \textbf{(eq. 36)}$$

where z_1 and z_2 refer to the range-bin bottom and top altitudes. The weighting function $w(z)$ is chosen to be the attenuated molecular backscatter (relevant for Rayleigh winds in a clear atmosphere):

$$w(z) = \beta_m(z) T^2(z) = \beta_m(z) \exp\left(-2\int_z^{z_{sat}} \alpha_m(z) dz\right) = \beta_m(z) \exp\left(-2\int_z^{z_{sat}} 8.4\beta_m(z) dz\right) \qquad \textbf{(eq. 37)}$$

T is the two-way transmission function, and α_m is the molecular extinction coefficient. The molecular backscatter is approximated via an exponentially decreasing density with altitude (with a scale-height of 7 km). The term $K/R(z)^2$ can be ignored from the LIDAR equation, since this varies relatively little within the range.

Assuming that the true wind varies linearly with height, there are analytical solutions to the integrals (not shown here) to produce the results of Figure 51. The height where the true wind is equal to the Aeolus average wind is then selected as the most appropriate placement for the Aeolus average wind, which typically turns out to be 0.47-0.49 of a fraction of a 2 km range-bin (varying with altitude).

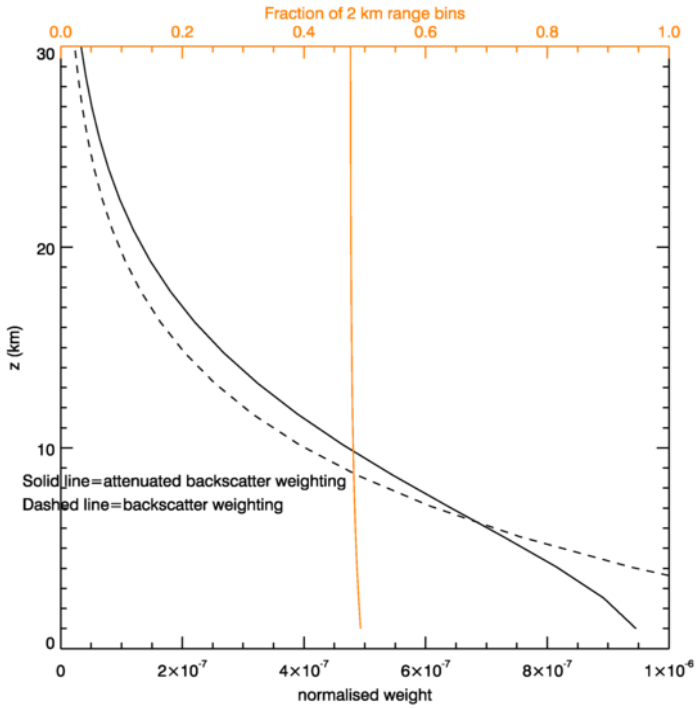


Figure 51. The fraction of the way up a 2 km range-bin that is optimal for Rayleigh wind vertical placement (in orange) when assuming a linearly varying wind with height (z). The solid black line shows the normalized attenuated backscatter profile (the weighting) which led to the result. The dashed line is the backscatter profile without attenuation.

6.8 L2Bp Mie Quality Control

The L2B processor method is to run the Mie core algorithm irrespective of the SNR (signal to noise ratio) value for the given set of Mie ACCD counts. This decision was based on VAMP study suggestions (according to comments in the L2B source code).
A SNR check is still performed, but after the Mie core has run. This SNR check sets a PCD bit flag to indicate whether a threshold SNR was exceeded or not (using a SNR threshold defined in the AUX_PAR_2B). This SNR QC flag does not affect the overall validity flag of the Mie wind result (which depends on thresholds on the peak height, FWHM and peak position relative to the maximum ACCD count). The user of the L2B data can decide to pay attention to the SNR flag if they wish to.

However, with the L2Bp version (v2.00) used in this TN, the SNR value used is not a valid value for QC anyway. It is using a measurement level SNR value in the middle of an observation as a placeholder; a value from the L1B data. The L2Bp will be updated (in v2.1) to calculate the observation level SNR value - then meaningful SNR QC can be done. At present the SNR_threshold is set to 0.0 because the comparison is meaningless - therefore no winds are flagged as exceeding SNR threshold in L2Bp.

It should be noted that the SNR QC check is somewhat redundant however since the L2Bp also does a calculation of the estimated error of the Mie observation. It propagates ACCD errors through the Mie core algorithm method to produce an error estimate (68% confidence interval, if Gaussian errors) on HLOS (in m/s). Practically this is more useful to NWP users than SNR value QC flags. NWP user can for example reject Mie observations if error estimate is > 3m/s, this should eradicate any outliers which can spoil the observation error statistics.