Aeolus Data Innovation Science Cluster
DISC

L2A user guide

D.Trapon, T.Flament, A. Lacour, H.Stieglitz, CNRS/CNRM

DISC-Ref.: AED-TN-CNRM-L2A-036
Issue: V 2.2
Date: 20/12/2022

Document Change Log

Issue	Date	New pages	Modified pages (after introducing new pages)	Observations	Name
V 1.2	20/02/2020	4	2-3	Add section 3.2 Reading the data.	T.Flament
V 1.3	14/05/2020	20-21	-	Add sections 4.4 Radiometric correction and 4.7 User-friendly Python script.	T.Flament D.Trapon
V 1.4	11/09/2020	-	1	Add link for Python code in section 1.1 Introduction. Document format moved to DISC layout.	D.Trapon
V 1.5	20/11/2020	-	16-17	Add subsections 4.4.1, 4.4.2 and 4.4.3 in section 4.4 Instrument calibration: radiometric correction tuning.	D.Trapon
V 1.6	02/03/2021	13-14	-	Add section 3.6 on the heterogeneity index and updates of the Figure numbers.	A.Lacour
V 1.7	24/06/2021	-	4-5-6-13-22	Update Reference Documents, Table-1 and section 3.5 QC flags. Add section 4.8 Aeolus papers and L2A algorithm.	D.Trapon
V 1.8	09/07/2021	22-23	14-22	Update section 3.5 QC flags. Add sections 4.8 Cloud screening based on AMD information and 4.7 De-noising of SCA backscatter and extinction coefficients via Maximum Likelihood Estimation.	D.Trapon
V 1.9	20/12/2021	-	6-17-18-19-21-25	Update of Table 1. Update of sections 4.2, 4.3, 4.7 and 4.10. Update of Figure 6. Remove ICA section 3.3.3.	D.Trapon
V 2.0	07/04/2022	-	8-9-10-11-23-24-25	Update section 4.7 MLE: Maximum Likelihood Estimation Algorithm and section 3.2.1 CODA. Add subsections 4.7.2 The MLE physical constraints and QC flags. Update tables within section 3.3 Description of datasets. Update Referenced Documents.	D.Trapon
			7-28	Update of the section 3.1 Data structure. Update of the ref DOI in section 4.10 Aeolus papers and L2A algorithm.	A. Lacour
V 2.1	17/06/2022	-	12-16-21-26-28	Update of sections 2, 3.3.2 and 3.5. Add QC table in section 4.7.2 The MLE physical constraints and QC flags. Add section 4.7.3 MLE at sub-BRC scale	D.Trapon A. Lacour

V 2.2	20/12/2022	-	7-8-16-28	Update of sections: 2.3 Short history of AEOLUS operations, 3.2.2 “Subtleties and 3.4 Vertical scales Add QC table for MLEsub in section 4.7.3	D.Trapon A.Lacour

Table of Contents

1 Introduction

This document is aimed at helping new users of the L2A product corresponding to the baseline 16 related to the L2A processor version 3.16. This product was developed over a long time without feedback from the user community and might be challenging to start with. This guide gives a summary of the product content and some practical explanations and recommendations.

Practical questions are tackled in section 4. For information about “which product to use?”, “what are the difference between available products?”, go directly to 4.1. A Python script to handle the data is provided at:

https://www.aeolus.esa.int/confluence/display/CALVAL/Aeolus+Data+Reading?preview=/327842/12353927/Read_and_Plot_L2A.DBL.py

1.1 Compliance Statement

The L2A user guide is fully compliant with the management requirements of the DISC contract.

1.2 Applicable Documents

[AD-1] DLR (2019): DISC Project Management Plan. AED-PMP-DLR-001, V 1.2, 29/05/2019.

1.3 Reference Documents

[RD-1] CNRM (2022): ADM-Aeolus L2A Algorithm Theoretical Baseline Document – Particle spin-off products, AED-SD-CNRM-L2A-030, V6.0, 17/06/2022.

[RD-2] DLR (2022): Aeolus Level 2a Processor Input/Output Data Definition, AED-SD-DoRIT-L2A-025, V 3.15, 29/07/2022.

1.4 Acronyms & Abbreviations

See [AD-1] for a complete list of acronyms and abbreviations

2 Operational versus Prototype version

The team at Meteo France was responsible for developing the prototype of the L2A in MATLAB until version 3.16. The processor is then re-coded in C and installed on ESA's computer by the PDGS team (Payload Data Ground Segment). A period of a few months is always necessary for the operational processor to implement the latest developments. In addition, up to v3.10, prototype and operational processors had a shift in version numbers. The following table lists corresponding versions, and the date of implementation at PDGS of the operational version.

Table 1. Prototype and operational processor versions

Prototype version	Operational	Deployment date of processing baseline at PDGS	Comments and details
3.07	3.06	-baseline 2A01, 03/09/2018 to 15/10/2018 reprocessed as baseline 2A02, 03/09/2018 to 16/05/2019	
3.08	3.07	-baseline 2A03, 16/05-14/06 -baseline 2A04, 14-16/06, end of FM-A -baseline 2A05 for 28/06-05/09, reprocessed as baseline 2A06 since 28/06 (FM-B)	All data for these periods was also processed with the prototype. We recommend to use prototype data for this period.
3.09	3.08	31/10/2019, baseline 2A07	Similar configuration to prototype, may be used, contains attenuated backscatters at measurement level. Uses calibration from July 2019.
3.10	3.10	02/04/2020, baseline 2A08 01/08/2020, baseline 2A10	V3.10 recomputes radiometric calibration itself from high altitudes clear sky signal. Improvement of ~10% to the average radiometric calibration values, some shorter term variations are left.
3.11.1	3.11	08/10/2020, baseline 2A11	A new radiometric correction of the calibration coefficients K_{Ray} and K_{Mie} (i.e. coefficients characterizing the radiometric efficiency of the receiver) has been implemented into L2A processor V3.11. The coefficients estimated from signal prediction in particle-free regions of the atmosphere for mid-altitudes 6 to 16 km are now calculated per observation using a multiple linear regression based on telescope temperatures oscillations (i.e. information provided by the Accurate Housekeeping Telemetry (AHT) and Thermal Control (TC) processes) which is computed to fit the K_{Ray} and K_{Mie}.
3.12.1	3.12	26/05/2021, baseline 2A12	The extinction to backscatter ratio (i.e. so-called Lidar ratio) and a scene heterogeneity index have been added to the L2A product. The scene heterogeneity index characterizes the variability of the Rayleigh/Mie useful signal within one BRC. It is defined as the standard deviation of the Rayleigh/Mie useful signal within a basic repeat cycle (BRC) with respect to the Poisson noise. A preliminary version of attenuated backscatter

			calculation and feature classification using an ATLID (EarthCare) approach has been added to the L2A processor. The corresponding data products are available for testing, but flagged invalid. The satellite on target flag provided in the L1B product is now evaluated when data from the L1B product is selected for L2A processing (e.g. K_{Ray}/K_{Mie} calculation). Data where the flag is set to false is excluded now. Signal-to-Noise ratio for the total Mie signal is recalculated in the L2A processor (workaround for missing parameter in L1B output) and used for further processing in the L2A (error calculation).
3.13	3.13	06/12/2021, baseline 2A13	Update of QC flag (i.e. including absolute errors instead of relative errors in addition to SNR) in V3.13. Add new product Cloud Screening based on AUX_MET with reordered dataset in product writing. Update of bin invalidation. Add a test of minimum valid BRCs in aerosol-free condition for radiometric correction.
3.14	3.14	29/03/2022, baseline 2A14	Final implementation of the physical regularization scheme, i.e. called Maximum Likelihood Estimation (MLE) or Denoising Scheme, to compensate the noise contamination of SCA retrievals. It corresponds to an alternative to the Standard Core Algorithm (SCA) processing of crosstalk corrected signals. Optical properties of particles are then retrieved from a minimization using the L-BFGS-B open source algorithm constrained by pre-defined physical bounds. Since its implementation in the L1B v7.11, the L2A prototype v3.14 read the total Mie SNR directly in the L1B. To be able to read older L1B files, the L2Ap will still compute a total Mie SNR estimation based on useful signal in the case L1B version is below v7.11.
3.15	3.15	17/06/2022, baseline 2A15	Implementation of the MLE denoising scheme at sub-BRC level. Add QC flag (i.e. based on SNR and error estimates) for MLE full-BRC retrievals.
3.16	3.16	-, baseline 2A16	Cleaning of the Group product. Correction of the assignation for the group mid bin geolocation. Correction of the radiometric correction. Add QC flag for MLE sub-BRC retrievals.

2.1 Access data from the ADDF

Operational products are available from the Aeolus Data Dissemination Facility (ADDF) at: http://aeolus-ds.eo.esa.int/oads/access/

In order to select a specific processor version on the ADDF, select version starting with “ADM_L2aP”, and for dates according to Table 1.

2.2 Access prototype data

Additionally, some data was processed at Meteo France to provide a quick adjustment to user problem or calibration updates. The data is uploaded to the CalVal FTP at:

ftp://ftp.eopp.esa.int/L2A/Prototype. For registered CalVal users, user name and password can be requested at "Info aeolus-calval" <Info.aeolus-calval@esa.int>. Several subfolders are available:

- Prototype 3.07 contains the first prototype processed datasets, under a mixture of 3.07, 3.08 and 3.09.

- with_attenuated_backscatter contains data for FM-B processed with prototype v3.09, covering period 28/06/2019 to 08/10/2019. They contain attenuated backscatters at measurement level.

- v3.10 contains a small sample of orbits processed with v3.10 and the radiometric correction turned on, for users who want to check the difference in optical properties.

- 3.10_RadiometricCorrectionOff contains data from 08 October 2019 to 31 Oct. 2019 reprocessed with prototype 3.10 but without radiometric correction, in a way similar to prototype v3.09.

2.3 Short history of Aeolus operations

ALADIN was switched on in the early days of September 2018. It worked until a platform error switched it off around mid-January 2019. The laser was restarted after about one month and worked again until mid-June. At this time, laser A (also designated Flight Modal A or FM-A) had seen its power decline to ~40 mJ/pulse. It was decided to switch to FM-B. FM-B operations ended in October 2022 in order to restart the FM-A for the AEOLUS mission end of life. The FM-A switch has been completed the 29 November 2022.

The two lasers are slightly different, and the three periods of operation of FM-A had different settings and also exhibit different behaviour. It is important to use 4 different calibrations for the 4 periods.

3 Overview of product content

3.1 Data structure

The data structure uses the Aeolus nomenclature of measurements, observations and range bins.

The laser is operated with a 50 Hz pulse repetition frequency, so called « measurements » are the data element resulting from the accumulation of 20 pulses i.e. ~3 km of horizontal integration (20 pulses / 50 pulses per second * satellite speed of about 7km/s). « Observations » (sometimes named BRC, Basic Repeat Cycle, from the initial burst mode of operation) are the accumulation of 30 measurements, i.e. around 85 km horizontal integration.

Note that for some orbits, the number of observations may not be the same from one dataset to another. For instance, the number of observations within the geolocation dataset may not match the number of observations within the SCA dataset. In such situation it is advised to compare the variables start_of_observation_time to find the correspondence between the datasets.

On the vertical scale, the backscattered signal is integrated over 24 time intervals that correspond to 24 « range bins ». The range bin settings can be adjusted, usually with bins thickness from 250 m to 2 km but always with the limitation that the total number of bins is 24 (because of the hardware, this is the number of accumulation lines of the detectors and cannot be changed). More detail is given in 3.4.On user request, the attenuated backscatters are output since prototype version 3.09. They are produced at measurement level.

3.2 Reading the data

A tutorial is provided here: https://www.aeolus.esa.int/confluence/display/CALVAL/Aeolus+Data+Reading.

Further information for specific L2A subjects is given below.

3.2.1 CODA

The binary .DBL format is cumbersome to read directly. The CODA software (https://atmospherictoolbox.org/coda/ or get the code directly from https://github.com/stcorp/coda) was specifically developed to facilitate the reading. CODA can be interfaced to Python, Matlab, IDL, Fortran, C and Java. Command lines in a terminal are also possible.

Specific binary format descriptions, named coda definition files, are available for Aeolus data at: https://stcorp.github.io/codadef-documentation/AEOLUS/. These files are needed to allow CODA to “know” the binary format it is asked to read. The Codadef file you are using needs to be up to date if you read newer files. Without the update, an incompatibility could happen, especially if you try to read the most recent prototype data. https://stcorp.github.io/codadef-documentation/AEOLUS/index.html lists all the Aeolus files that can read read. A sub-page links to ALD_U_N_2A files formats (Figure 1). File descriptions give for a complete overview of data product branches and variable names.

		version	format	definition
ALD_U_N_2A	Aeolus Level 2A Product (DBL)	0	binary	ALD_U_N_2A_02_02
		1	binary	ALD_U_N_2A_03_00
		2	binary	ALD_U_N_2A_03_01
		3	binary	ALD_U_N_2A_03_02
		4	binary	ALD_U_N_2A_03_05
		5	binary	ALD_U_N_2A_03_08
		6	binary	ALD_U_N_2A_03_09
		7	binary	ALD_U_N_2A_03_10
		8	binary	ALD_U_N_2A_03_12
		9	binary	ALD_U_N_2A_03_13
		10	binary	ALD_U_N_2A_03_14
		version	format	definition
ALD_U_N_2A_HDR	Aeolus Level 2A Product (HDR)	0	xml	ALD_U_N_2A_HDR_02_02
		1	xml	ALD_U_N_2A_HDR_03_00
		2	xml	ALD_U_N_2A_HDR_03_01
		3	xml	ALD_U_N_2A_HDR_03_08
		4	xml	ALD_U_N_2A_HDR_03_12
		5	xml	ALD_U_N_2A_HDR_03_14

Figure 1. View of the S&T table listing L2A products version, and links to their description at *https://stcorp.github.io/codadef-documentation/AEOLUS/index.html*

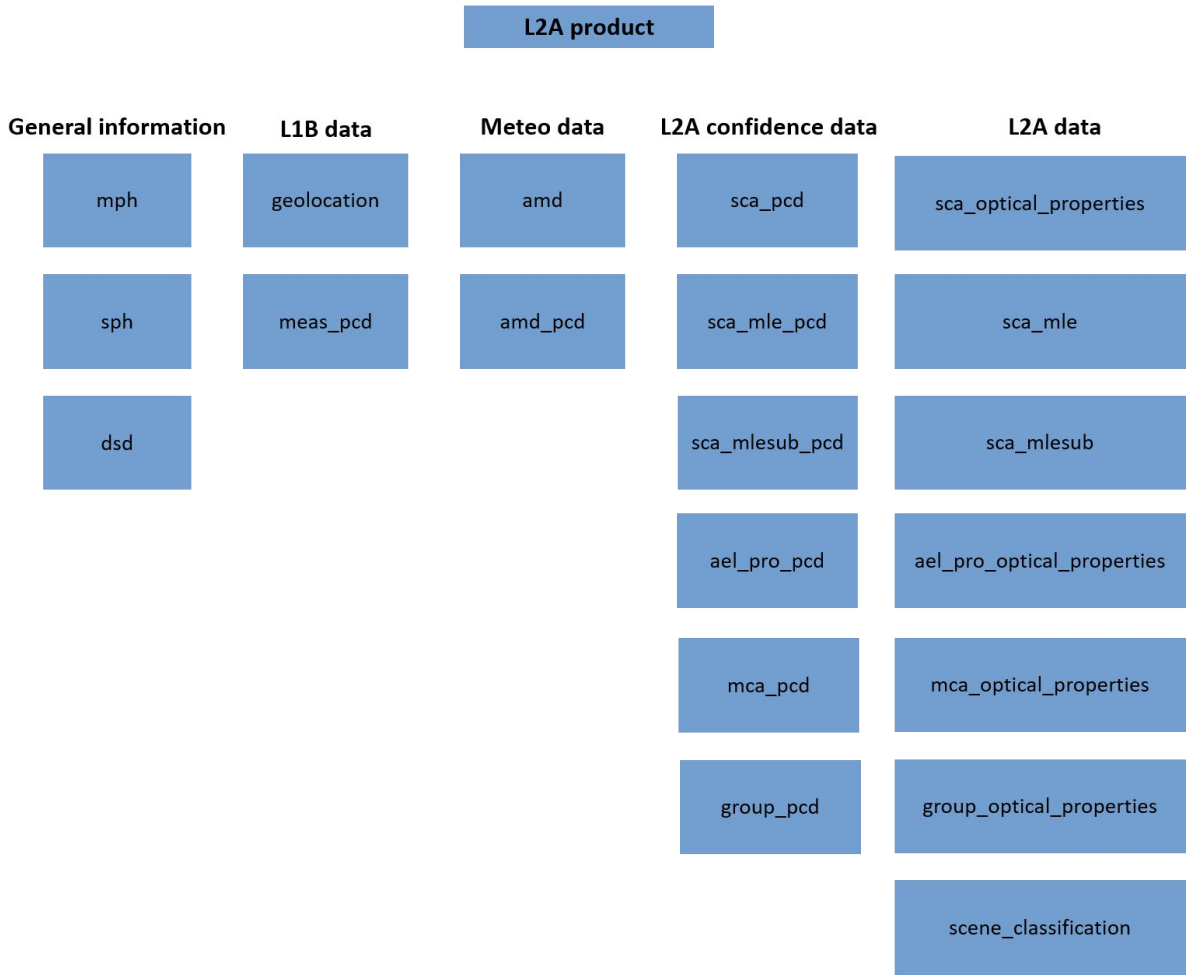
3.2.2 "Subtleties" (or tips and tricks, or hints to deal with the complex structure of the products)

The SCA product cannot be applied everywhere. Typically, a few profiles for every orbit file are not processed. This is reported in the sca_applied flag contained in the L2A dataset: Meas_PCD_ADS. On the other hand, all geolocations from the L1B are reported. The sca_applied flag must be used to sort which observations are actually present in the L2A products.

Geolocations for the mid-bin product are provided in a different place (see 3.3.4 Geolocations).

3.3 Description of datasets

The L2A product is composed of 17 datasets. The Input Output Definition Document (IODD) [RD-2] gives full details about the product structure but a more human-friendly (and high level) description is given below.



3.3.1 SCA: Standard Correct Algorithm

SCA is the algorithm of the L2A processor that makes use of the High Spectral Resolution capability of ALADIN. It uses both Rayleigh and Mie channels to perform so-called “crosstalk correction” to get separately attenuated molecular backscatter and attenuated particulate backscatter. This allows to retrieve particulate extinction and backscatter without any assumption on the lidar ratio.

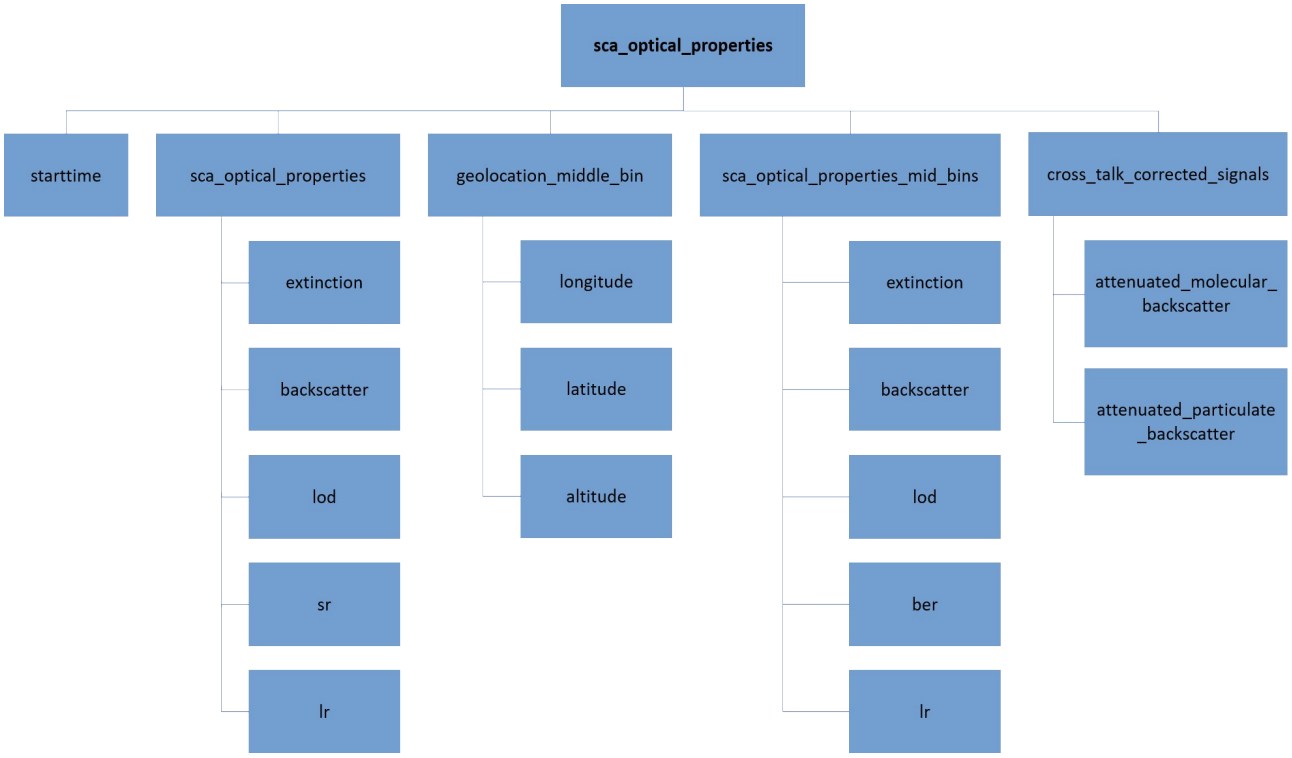
3.3.1.1 Attenuated backscatters as a product

The attenuated backscatters, separated in molecular and particulate contributions, are intermediate steps and were not provided in the first place. Since prototype version 3.09, they are output in the SCA optical properties dataset and are provided at measurement scale (~3km). See Operational versus Prototype version for information on where to find corresponding data.

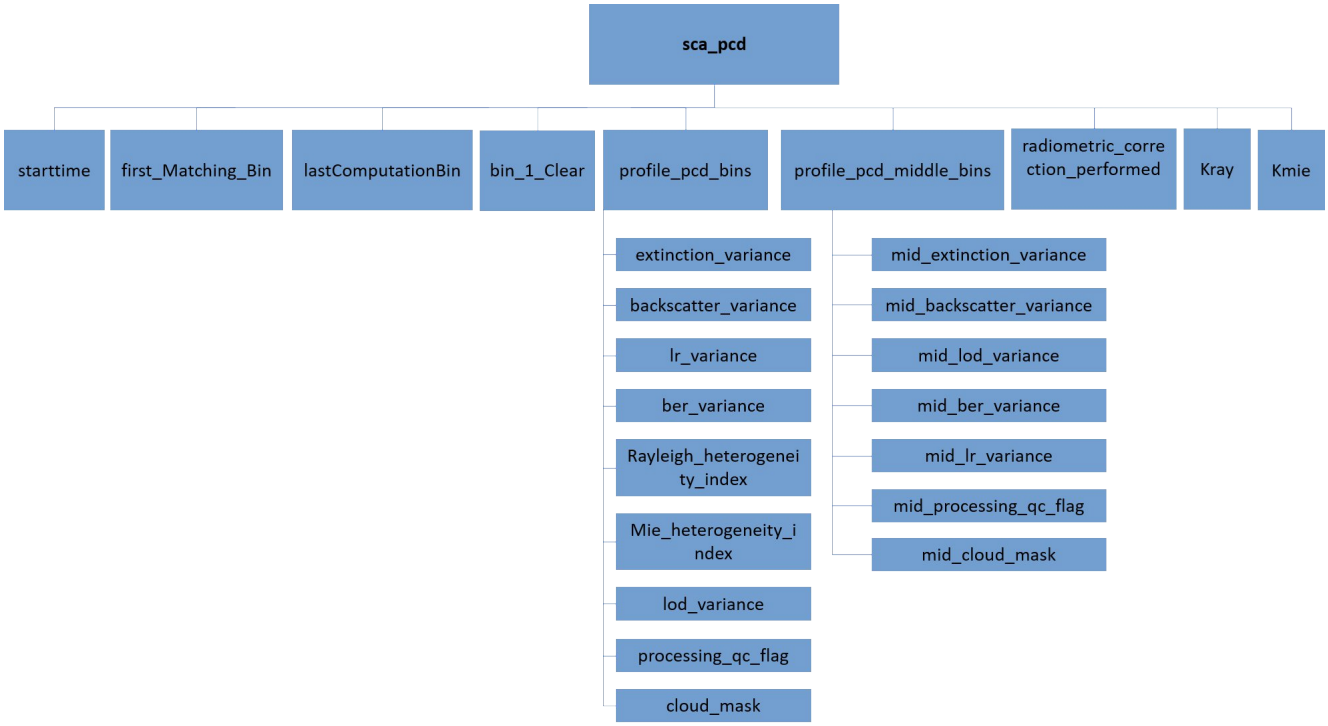
3.3.1.2 Dataset structure

The following tables give an overview of how the “SCA optical properties” and “SCA product confidence data” datasets are organised.

SCA Optical Properties



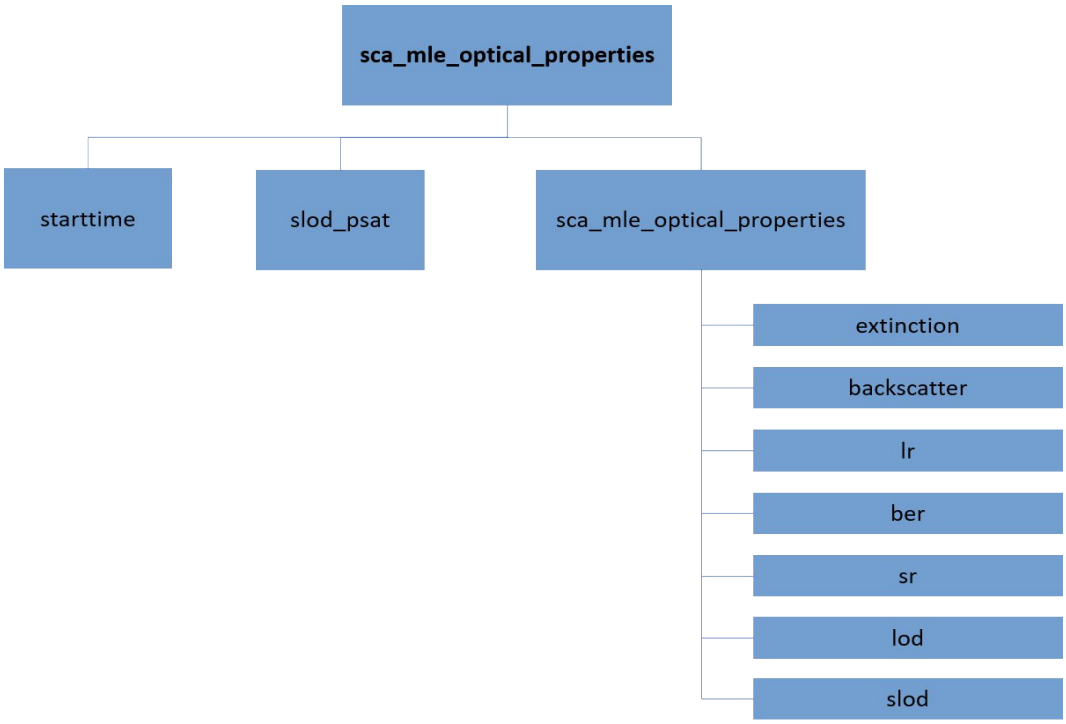
SCA Product Confidence Data



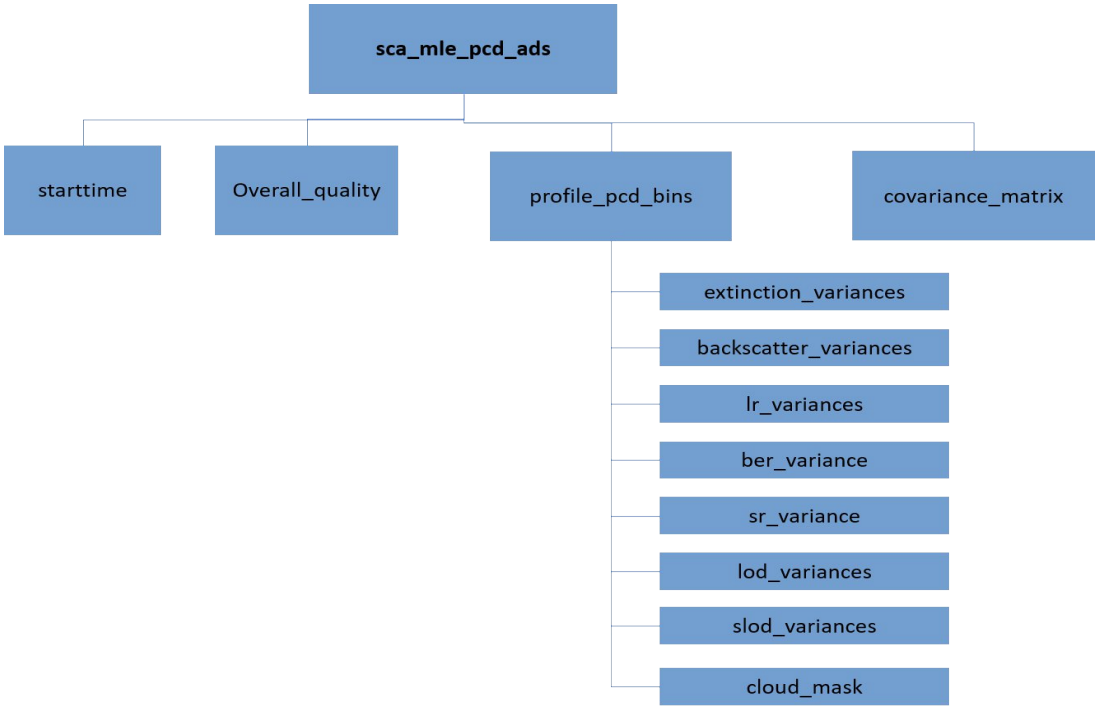
3.3.2 MLE: Maximum Likelihood Estimator (Denoising scheme)

The following tables illustrate the dataset structures of a new physical regularization scheme called MLE which has been implemented in L2A prototype v3.14. It can be seen as an alternative to the SCA. The following tables illustrate the datasets “SCA MLE optical properties” and “SCA MLE product confidence data”. A description of the algorithm principle is available in section 4.7.

SCA MLE Optical Properties

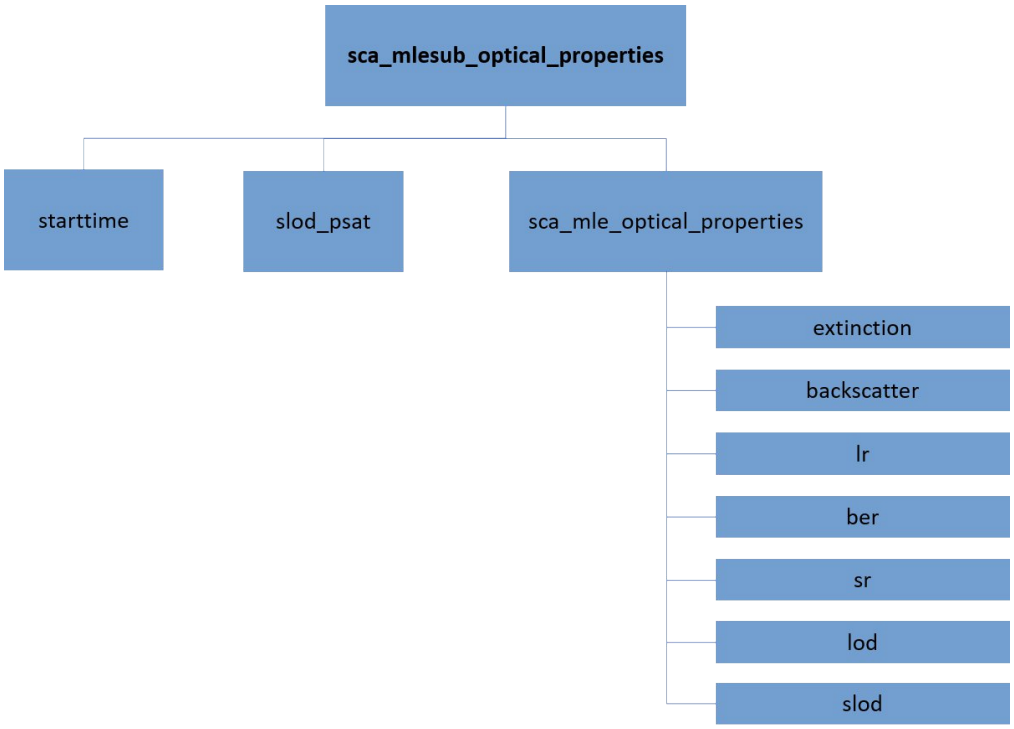


SCA MLE Product Confidence Data

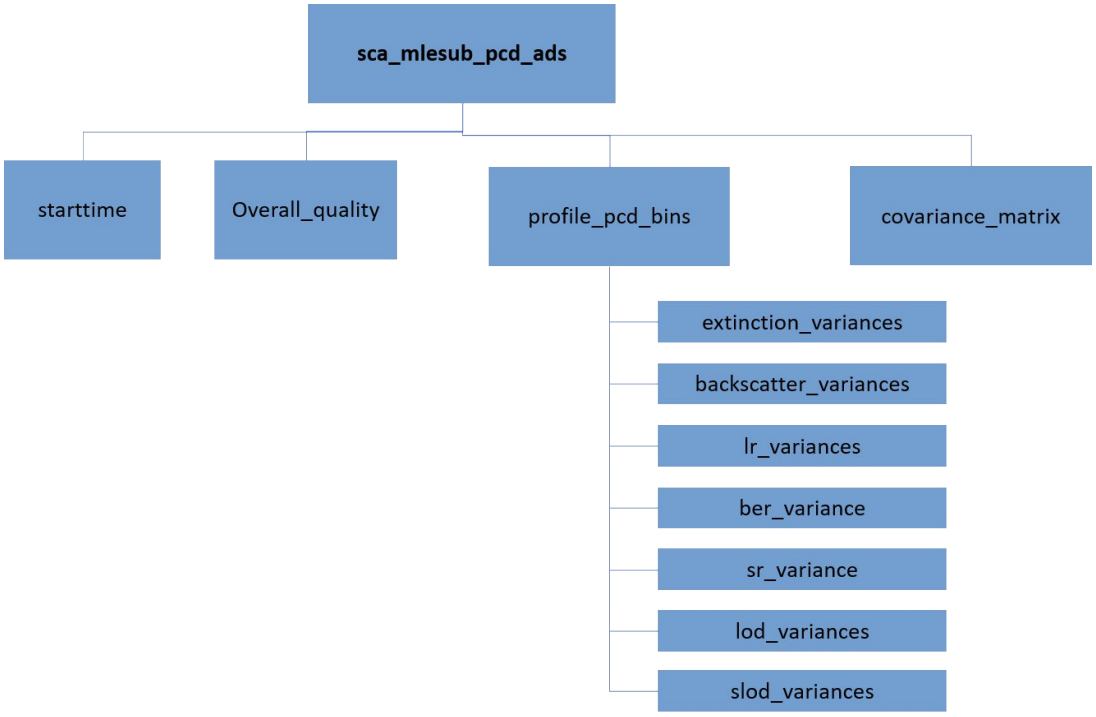


Since L2A v3.15, the MLE algorithm is executed a second time at an increased horizontal resolution. The following tables illustrate the datasets “SCA MLEsub optical properties” and “SCA MLEsub product confidence data”. A description of the algorithm principle is available in section 4.7.3.

SCA MLEsub Optical Properties

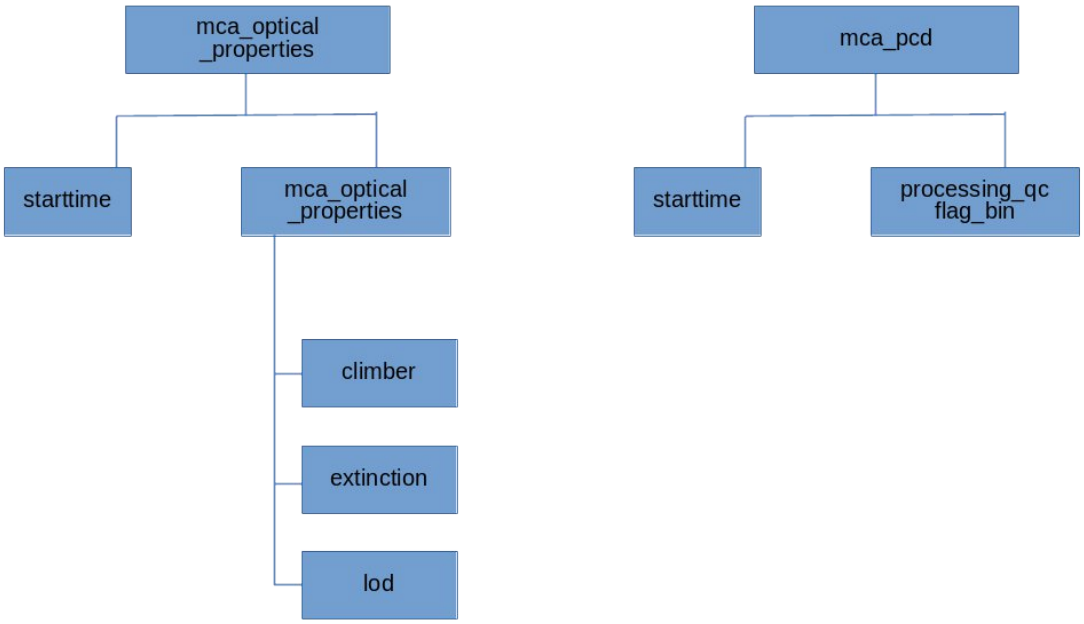


SCA MLEsub Product Confidence Data



3.3.3 MCA: Mie Channel Algorithm

MCA is only based on the Mie channel and provides particulate extinction. It relies on climatological data to make an assumption on the lidar ratio. The climatology used in the TD01 dataset uses a backscatter-to-extinction ratio of 0.07 everywhere.



3.3.4 Feature finder and Group product

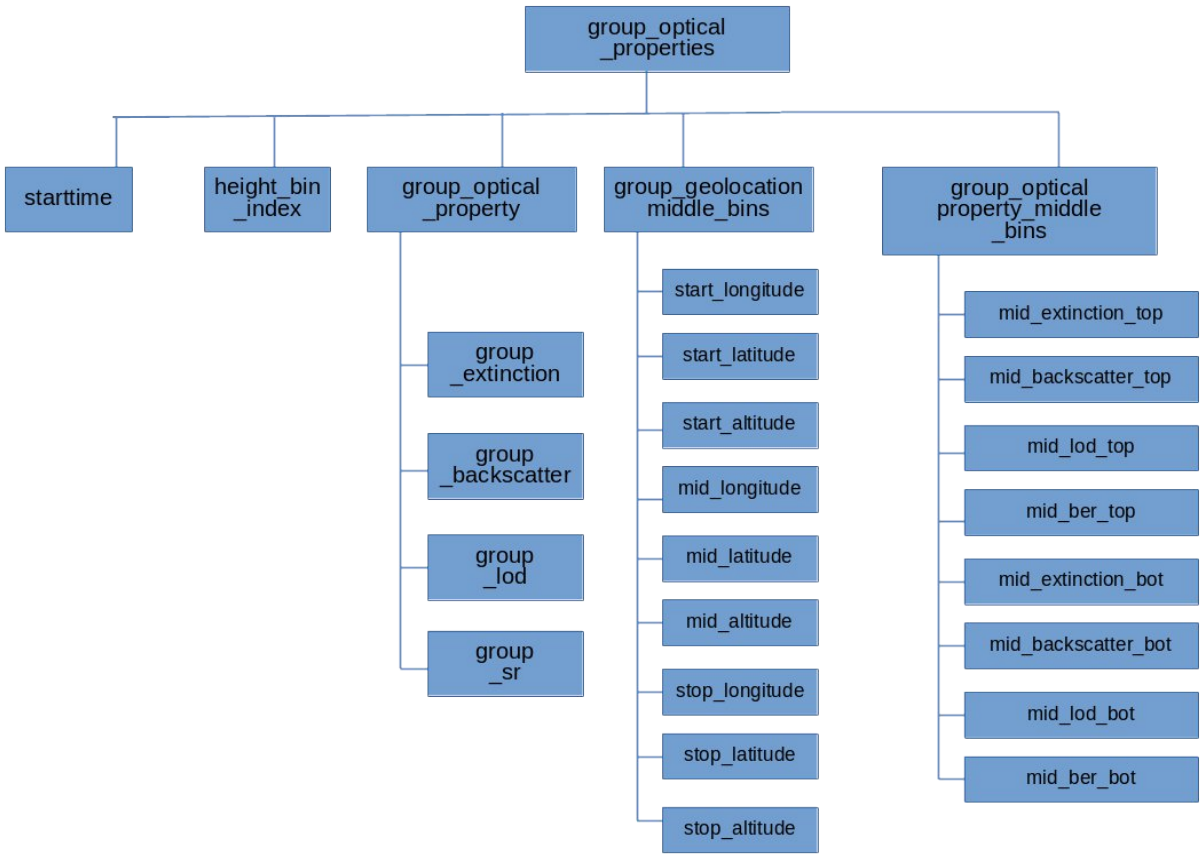
The feature finder is the first step in determining sub-BRC scale features. It is based on a median filter of the Mie channel SNR. Any group of pixels larger than a minimum size (currently 5 measurements) continuously above a predefined threshold in Mie SNR is detected as a group.

Signals at groups level are then accumulated and processed with the same SCA algorithm as the main BRC-scale product.

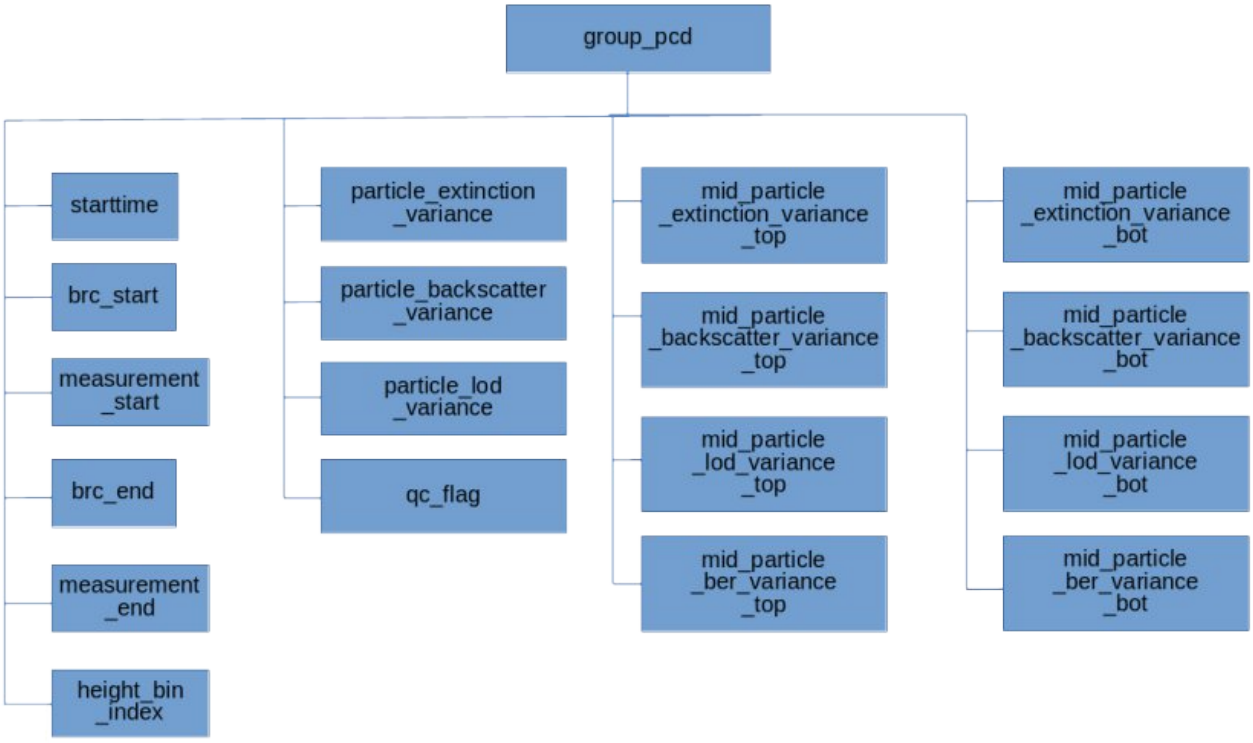
The group product might be more difficult to interpret. A group is identified by its start time, which gives the Observation to which it belongs, and its heigh bin, which gives its altitude. Several groups can be present in a given Observation for a given bin. Their geolocation is retrieved by using the group_PCD dataset, where start and stop measurement within the observation are reported.

Groud_pcd and group_optical_properties have the same number of entries. Geolocations can then be retrieved from the geolocation dataset.

Group Optical properties

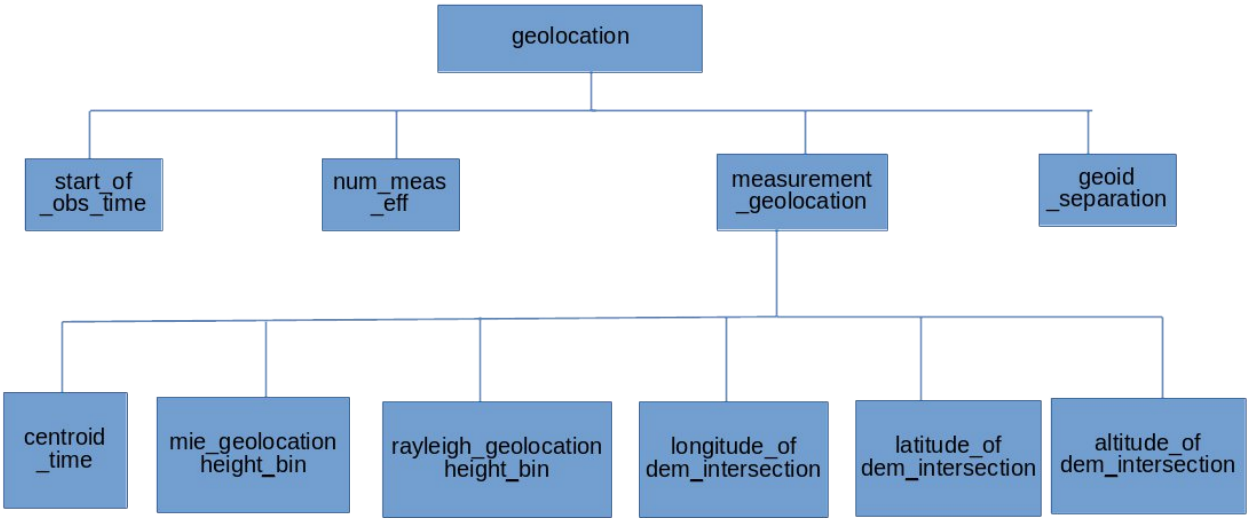


Group Product Confidence Data



3.3.5 Geolocations

Geolocations are provided directly from the L1B file for Rayleigh and Mie channel bins. Geolocations for "mid-bins", on a different vertical scale (see 3.4), are provided within the SCA_OPT_MDS.



3.3.6 Scene Classification

The Scene Classification product is currently still at draft stage and was not investigated for a long time. For instance, Relative Humidity is arbitrarily set at 0.98 in bin #14, for testing purpose.

It contains two indicators, AladinCloudFlag and NWPCloudFlag (described in the L2A IODD). AladinCloudFlag is derived from comparison of the Backscatter-to-Extinction ratio to thresholds defined in the AUX_PAR_2A configuration file. NWPClouFlag is derived from Meteorological data.

Meteorological data from the AUX_MET_12 files is provided in the AMD_ADS (Auxiliary Meteorological Data). This more basic data should be preferred by users who want to derive their own classification.

3.4 Vertical scales

The ALADIN instrument uses two different vertical scales for the Mie and Rayleigh channel respectively. Mie and Rayleigh vertical scales are both reported in the Geolocation_ADS.

The Rayleigh Range Bin Setting, or Rayleigh vertical scale is used for the SCA and Feature finder products.

The Mie vertical scale is used for the MCA product.

The SCA also contains an additional vertical scale, the « middle bin » designed to compensate an error propagation in the extinction (see the ATBD for more details about how the extinction is derived). Mid-bin extinction is not subject to the propagation of the error from bin to bin at the expense of reduced vertical resolution. Mid-bin products are derived from the average over two vertical bins and the value is affected to a mid-bin ranging from the middle of the upper bin to the middle of the lower one. The mid-bin scale is to be read from the SCA_OPT_MDS and to be used with « mid- » variables contained in the SCA.

The group product also contains results in the mid-bin vertical scale. For each feature identified in the Rayleigh scale, the optical properties are computed for the top middle bin and the bottom middle-bin. The top middle bin result is the average between the feature bin and the adjacent bin above and the bottom middle bin corresponds to the average of the feature bin and the adjacent bin below. For each features, the middle bins product provides geolocation for 3 positions: start, mid and stop. "Start" stands for the upper boundary of the top mid bin, "mid" is for the boundary between the top and the bottom mid-bin and stop is for the lower boundary of the bottom mid-bin.

Vertical scales contain one more value than the parameter they describe: they give the upper and lower limit for each range bin. Mie and Rayleigh vertical scales contain 25 values to plot 24 backscatter or extinction values. Mid-bin scale contains 24 altitudes to describe 23 mid-backscatter or mid-extinction.

3.5 QC flags

The content of Quality Check (QC) flags for SCA algorithm is described in ATBD v6.0, section "6.2.3.6 Data Quality Flag".

The QC flag is an 8-bit unsigned integer that is the sum of the results of validity checks summed up in Table 2. Each check yields 0 when invalid, 2^n when valid. The users have to split the integer into the results of the check themselves.

Table 2. Validity criteria and thresholds used in TD01 dataset for SCA

Internal name	Possible values	Condition	
alpha_valid	0, 1	Ray_SNR_valid ==1 AND alpha_error_bar_valid==1	
beta_valid	0, 2	Mie_SNR_valid == 1 AND beta_error_bar_valid == 1	
Mie_SNR_valid	0, 4	Mie_SNR > 40	
Ray_SNR_valid	0, 8	Ray_SNR > 90	
alpha_error_bar_valid	0, 16	estimated_extinction_error < $8e^{-4}$ m^{-1}	
beta_error_bar_valid	0, 32	backscatter_estimated_error < $1e^{-5}$ $m^{-1}.sr^{-1}$	
Total_attenuation_valid	0, 64	Accumulated_optical_depth < 4	
Not used	0, 127	--	

Table 3. Validity criteria and thresholds used in TD01 dataset for SCA mid-bin product

Internal name	Possible values	Condition	
mid_alpha_valid	0, 1	mid_Ray_SNR_valid ==1 AND mid_alpha_error_bar_valid==1	
mid_beta_valid	0, 2	mid_Mie_SNR_valid ==1 AND mid_beta_error_bar_valid==1	
mid_BER_valid	0, 4	0.01 < mid_BER < 0.1	
mid_Mie_SNR_valid	0, 8	mid_Mie_SNR > 40	
mid_Ray_SNR_valid	0, 16	mid_Ray_SNR > 90	
mid_alpha_error_bar_valid	0, 32	estimated_mid_extinction_error < $8e^{-4}$ m^{-1}	
mid_beta_error_bar_valid	0, 64	estimated_mid_backscatter_error < $1e^{-5}$ $m^{-1}.sr^{-1}$	
mid_Total_attenuation_valid	0, 128	Accumulated_optical_depth < 4	
Not used	0, 255	--	

3.6 Heterogeneity index

The heterogeneity index is an indicator of the signal variability within a BRC. Its computation is described in ATBD v6.0, section "6.4 Improvement of SCA products: Heterogeneity indexes". The index is defined relatively to the Poisson noise. Its value should be 1 if the variability within a BRC is dependent on the Poisson noise only. Following subsections present the heterogeneity index values obtained from simulations and real data.

3.6.1 E2S simulation

A first test is performed on a scenario simulated by the E2S. The scenario consists in 8 BRCs with 30 measurements each and is representative of convective clouds observed in the tropics. Except for the 4th BRC, each BRC represents heterogeneous scenes (Figure 2 top panel). To test the robustness of the indicator, 20 independents simulations are performed. The resulting 20 profiles of heterogeneity index prove that the indicator is reliable except in areas with dominant noise. Those areas are below optically thick features and the bins below 2 km which have a vertical resolution of 250 m.

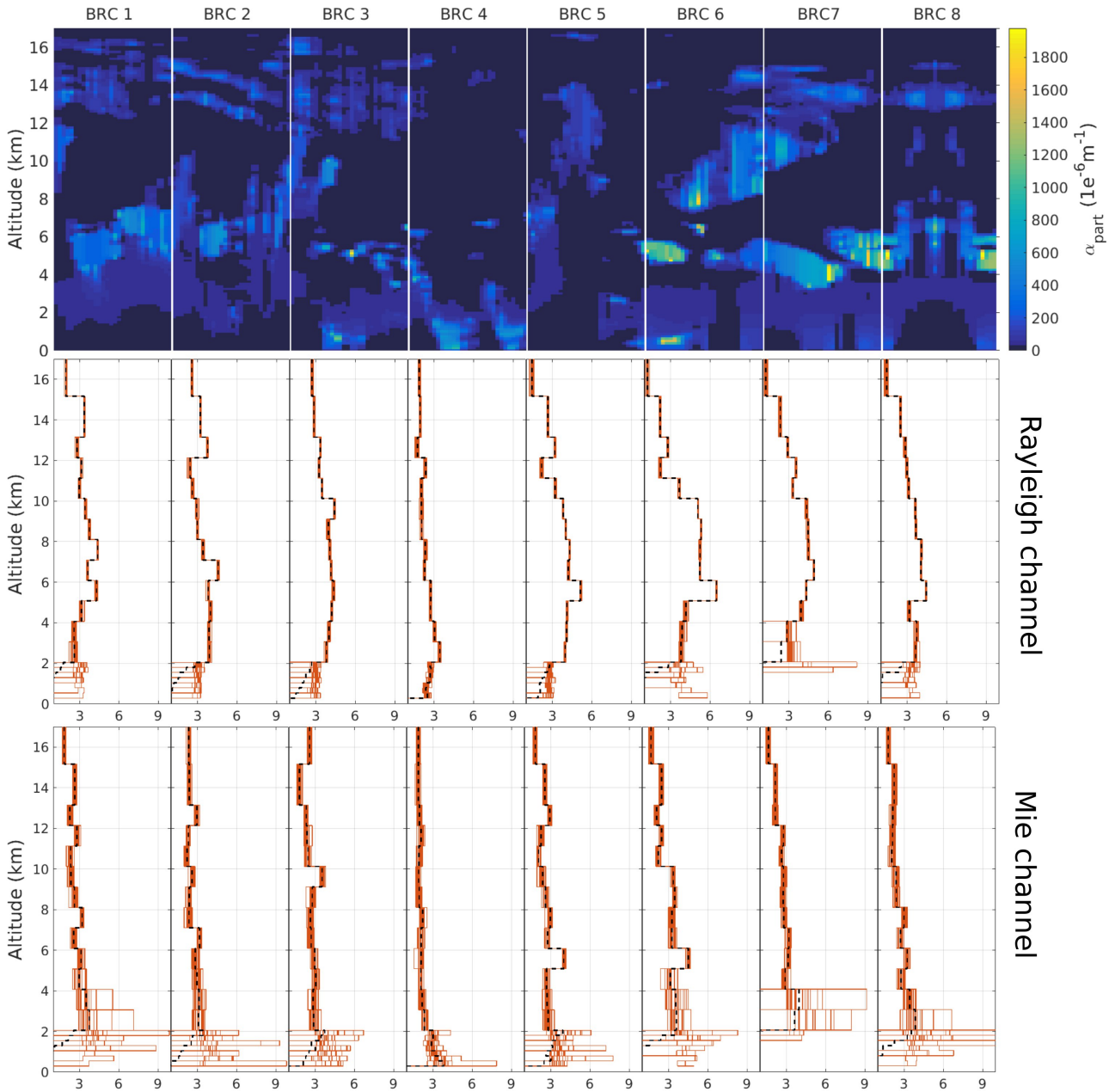


Figure 2. Particular extinction defined as input of the E2S (Top panel). Heterogeneity index profiles retrieved for the Rayleigh (middle) and Mie (bottom) channels. Each red profile corresponds to one simulation and the black dotted line the average of the 20 simulations.

It is also noticeable that in this test case, homogeneous BRC bins yield an index value around 2 for the Mie channel and between 2 and 3 for the Rayleigh channel. In both channel values above 3 denote a heterogeneous scene.

3.6.2 Existing observations

In practice the values of these heterogeneous indexes are rarely equal to 1. Figure 3 and Figure 4, indicates that homogeneous bins yield a *heterogeneity_index* ranging between 3 and 5 respectively for the Rayleigh and Mie channel.

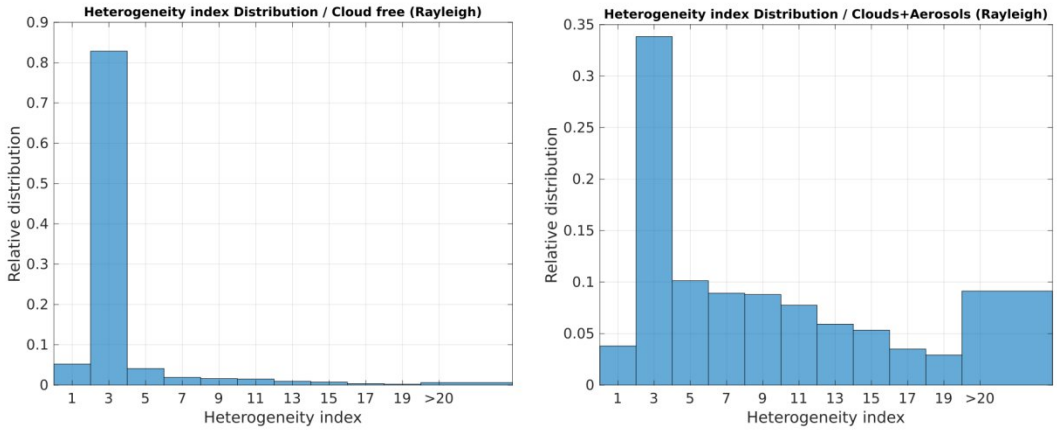


Figure 3. Distribution of the heterogeneity index computed on one orbit file and separated by cloud free (left) and cloudy (right) areas. Bins with SNRray < 50 are rejected.

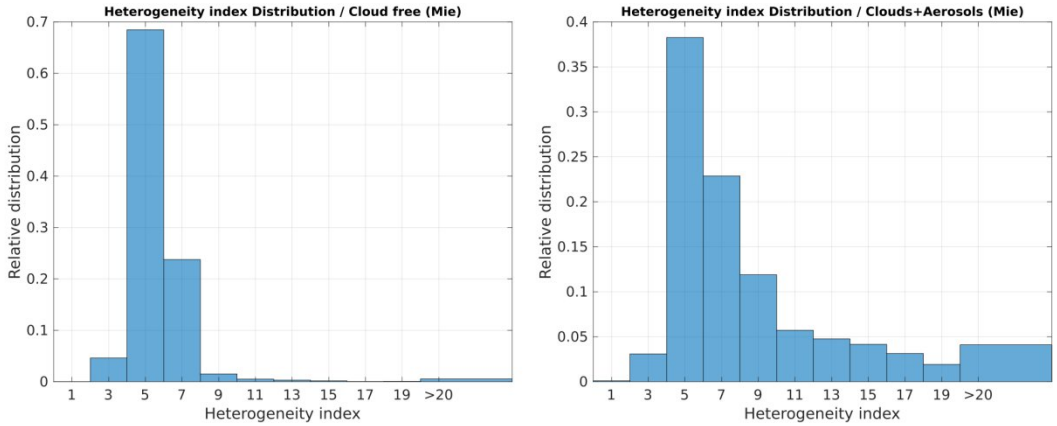


Figure 4. Same as Figure 3. Bins with SNRmie < 35 are rejected.

Figure 5 shows an example of the heterogeneity index applied on the Rayleigh channel for a simple scene. This example suggests that in cloudy areas, the bins can be considered as relatively homogeneously populated while the heterogeneously index is below 10.

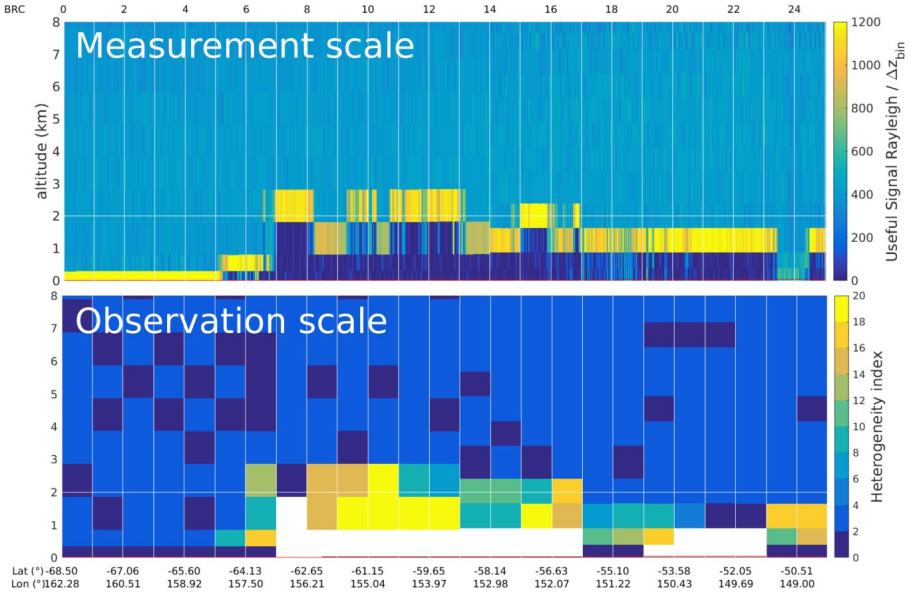


Figure 5. Comparison between the Rayleigh useful signal (top) and the associated heterogeneity index (bottom) on a scene.

4 Practical questions and recommendations

4.1 Which data to select for basic L2A data exploration

Main L2A products are SCA backscatter and extinction coefficients, on Rayleigh altitude scale:

- To get SCA backscatter (or extinction), go to:

sca_optical_properties => sca_optical_properties => backscatter (or extinction)

- To get the appropriate altitudes for the corresponding profile, go to:

geolocations => measurement_geolocation => rayleigh_geolocation_height_bin

Warning: this altitude vector contains 25 values whereas backscatter (or extinction) vector only contains 24 values. Details are given in section 3.4.

- To get corresponding backscatter (or extinction) estimated variance, go to:

sca_pcd => profile_pcd_bins => backscatter_variance (or extinction_variance)

4.2 SCA normal vs Mid-bins

Mid-bin products were designed to overcome the oscillating propagation of errors in the extinction coefficient (if the extinction is overestimated (underestimated) in a bin, it tends to be underestimated (overestimated) in the next). The error propagation calculations conducted in section 6.3.1 of L2A ATBD show that a proper averaging of two consecutive bins has a much better accuracy. Therefore, the mid-bin extinction should be used rather than the extinction if the accuracy is the first priority. It must be noted, however, that its vertical resolution is coarser. So, in high SNR regions, the extinction can be used and will provide a better vertical resolution. As shown in Figure 6 with middle plot the bin extinction is consistent above 5km with acceptable SNR but non-reliable below (i.e. overestimation of extinction coefficient for a second aerosol layer at ~4km).

Mid bin backscatter can also be used, it considers the average of the upper and lower limit of the two respective bins. However, the SCA backscatter product does not suffer the same oscillating error propagation as the extinction, so the accuracy of the mid-backscatter is not significantly better than the backscatter itself (e.g. comparable estimates shown in Figure 6 left plot). As regards the backscatter to extinction ratio (BER) or the so-called lidar ratio corresponding to the extinction to backscatter ratio the mid-bin product often provides the best estimates. Nevertheless, the bin rejection might be more severe (e.g. line gap on right plot) because of close to zero or negative averaged backscatter.

Note that Aeolus only measures the co-polar component of the backscatter hence a systematic underestimation of the backscatter signal (up to ~50% for highly depolarizing particles such as dust). As a result, the lidar ratio appears overestimated when comparing estimates with ground based lidar.

Please also note that a systematic offset affects the backscatter coefficient in low altitude. An overestimation of the backscatter signal can then be seen below 2km as shown in Figure 6 with left plot. Investigation are ongoing but it is likely to be related to non-moving earth surface (ground return) in lowest range bin (i.e. signal coming from solid targets and not from particles or molecules). The user can take a closer look at the L1B product within the Global Annotation Data Set (GADS) to see if the measurement of the bin are contaminated by ground return signal. The information can indeed be found in the L1B products called Top_Ground_Bin (i.e. highest detected bin with ground return) and Bottom_Ground_Bin (i.e. lowest detected bin with ground return) within Rayleigh_Measurement_Validity_Criteria.

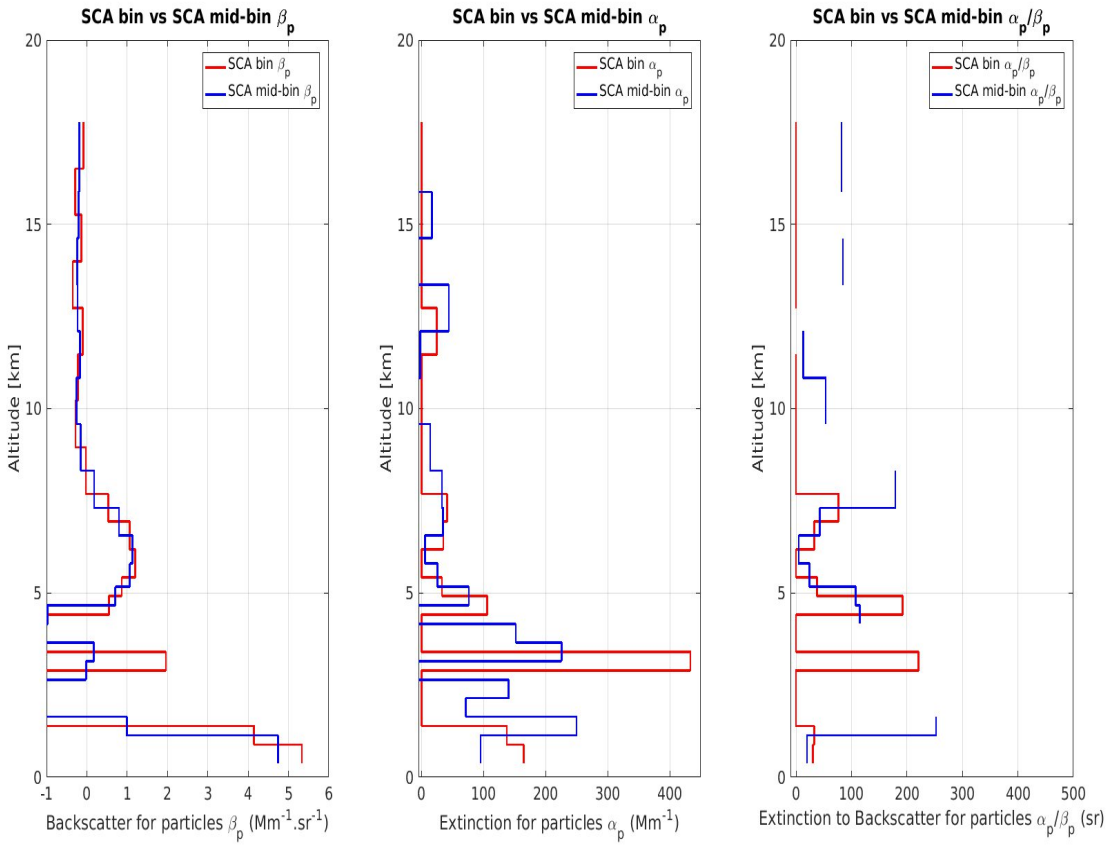


Figure 6. Comparison of SCA bin and mid-bin product for backscatter (left), extinction (centre) and extinction to backscatter (right) for selected profile (i.e. aerosol layer in low altitude above Cape Verde).

Normal bins have 24 values between 25 altitudes levels, and mid-bins have 23 values between 24 altitude levels.

4.3 SCA quality flags

See 3.5.

Mie and Rayleigh Signal to Noise Ratio (SNR) are the main proxies for quality but standard deviation of the extinction and backscatter coefficients had been added in the quality flag product.

Examples of bin rejection on SCA product extinction and backscatter coefficients are shown in Figure 7. It has to be noted that tuning of QC thresholds was designed to keep interesting features (i.e. aerosol plumes, clouds…) in final product. Therefore, key features presented in Figure 7 (i.e volcanic ash generated by Tonga eruption on January 2022) are not rejected despite noisy signal.

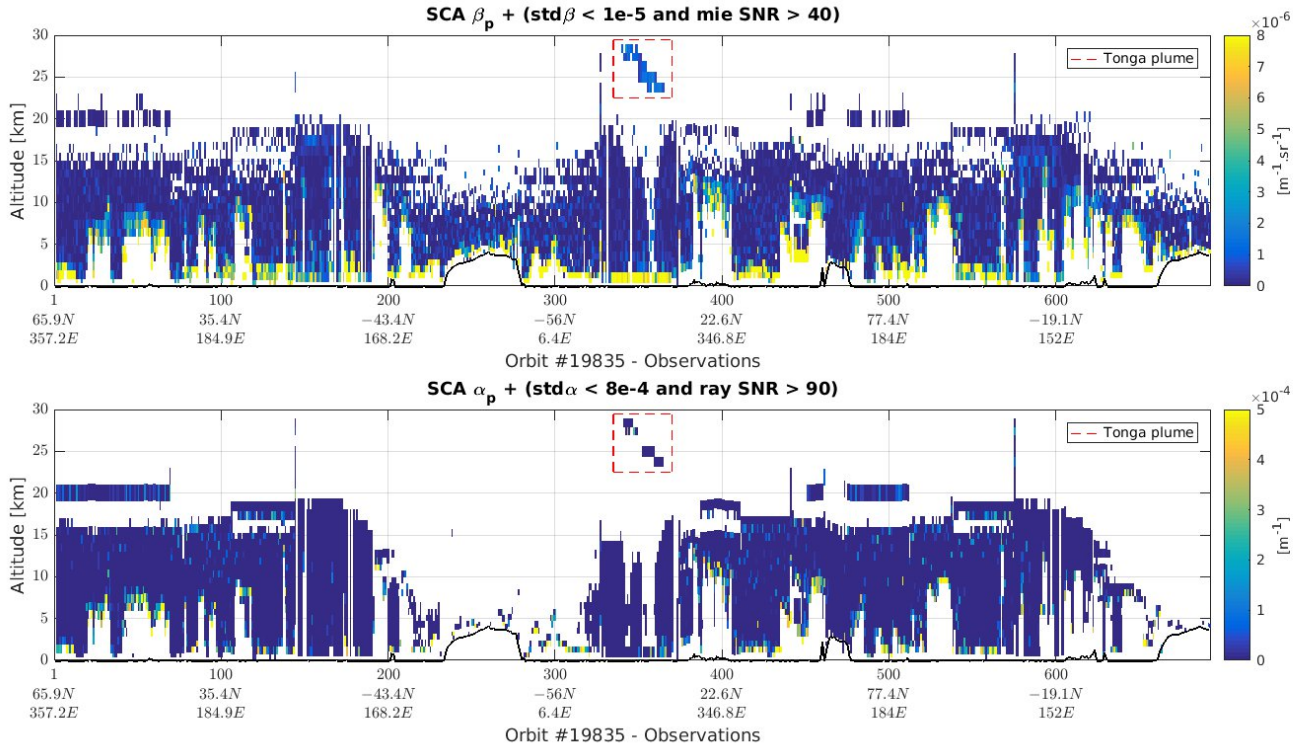


Figure 7. Extinction and backscatter retrieved by SCA with Quality Flag based on SNR and error estimates.

4.4 Instrument calibration: radiometric correction tuning

4.4.1 Initial estimation of calibration coefficients K_{Ray} and K_{Mie}

Accumulated lidar signals in vertical range bin for Rayleigh and Mie channels, respectively S_{Ray} and S_{Mie}, rely on calibration coefficients. In addition to coefficients corresponding to Rayleigh and Mie signals in Fabry-Perot and Fizeau the K_{Ray} and K_{Mie} are of particular relevance as they characterize the radiometric efficiency of the receivers. Initially they were considered as constant that can be estimated from auxiliary calibration data produced by the Instrument Response Calibration (IRC) mode when the laser is pointing at NADIR with no wind and no Doppler shift (e.g. dark dashed lines respectively for K_{Ray} in Figure 8 and K_{Mie} in Figure 9).

4.4.2 Orbit averaged correction of K_{Ray} and K_{Mie}

K_{Ray} and K_{Mie} were firstly assumed to be constant along Aeolus orbit. Nevertheless, when focusing on aerosol-free regions of the atmosphere for deriving K_{Ray} and K_{Mie} from the intensity of the molecular signal, significant deviations for both coefficients (i.e. up to 15%) were observed (e.g. light blue curve in Figure 8 for K_{Ray}).

It was decided to exploit these particle-free regions of the atmosphere for which K_{Ray} and K_{Mie} are proportional to well-known values. A corrective factor can be easily derived from signal prediction (i.e. (observed signal – predicted signal) / predicted signal) and median of relative error's deviation from zero is therefore taken as multiplicative term in order to get unique "orbit averaged" values of K_{Ray} and K_{Mie} per orbit (e.g. dark solid lines in Figure 8 and Figure 9 respectively for K_{Ray} and K_{Mie}). This corrective scheme has been included in L2A processor V3.10.

4.4.3 Fitting K_{Ray} and K_{Mie} per observation by using telescope temperatures

Other studies have demonstrated that thermal constraints (i.e. temperatures gradient between outer and inner parts of the telescope) vary along the orbit and affect the L2A radiometric performance. Moreover, a robust correlation (i.e. $R^2 > 0.7$) between deviations of the so-called Kratio (i.e. ratio between K_{Ray} and K_{Mie}) and telescope temperatures' oscillations (i.e. information given by telescope housekeeping and thermal control procedures) has been shown.

A corrective scheme (i.e. multiple linear regression) using housekeeping telemetry data has been applied: the K_{Ray} and K_{Mie} derived from the intensity of the molecular signal in particle-free regions of the atmosphere are given as input of the regression with 12 temperatures timelines at observation scale extracted from L1B data (i.e. temperatures corresponding to sensors distributed all over the telescope of the primary mirror).

Knowing that telescope temperatures are available per observation for both conditions (i.e. clear sky and aerosol contaminated regions of the atmosphere) the regression coefficients therefore allow to provide corrected values of K_{Ray} and K_{Mie} for all observation of the orbit (e.g dark blue curve without gap in Figure 8 for K_{Ray}).This corrective scheme has been included in L2A processor V3.11 in addition to the orbit correction scheme. The Aeolus Data Innovation and Science Cluster DISC can choose between the two radiometric corrections when processing the L2A by setting a flag in the auxiliary parameters file AUX_PAR_2A. The flag is given as output per observation within the final product indicating to the users which method has been applied.

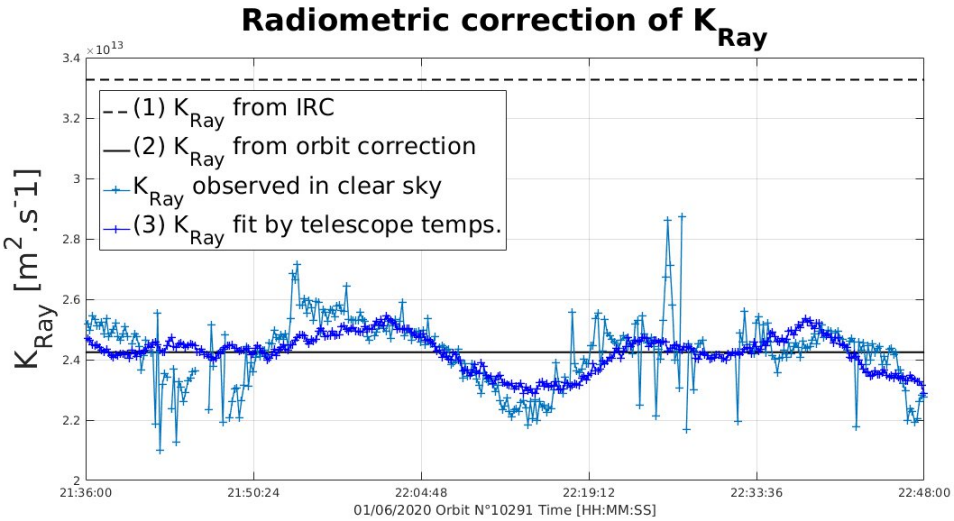


Figure 8. K_{Ray} deviation along Aeolus orbit: Example with orbit N°10291 on 1st of June 2020.

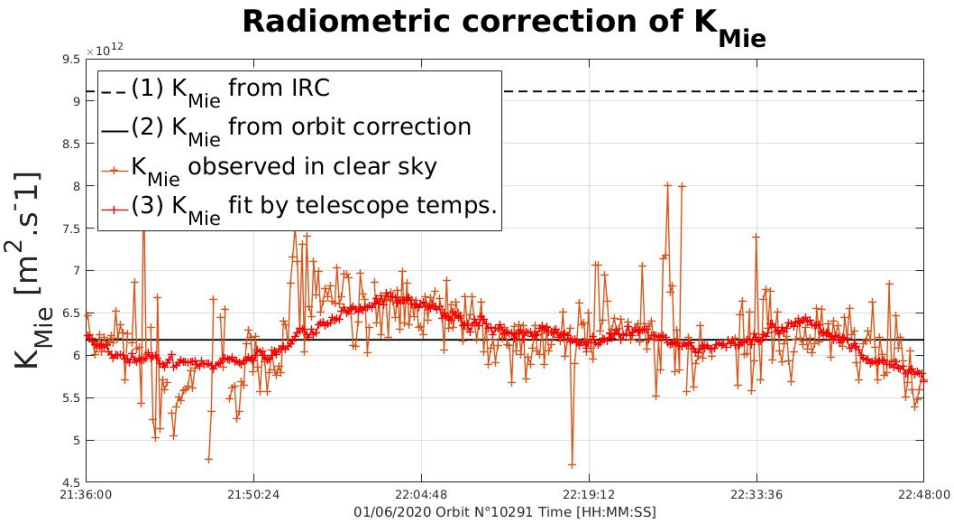


Figure 9. K_{Mie} deviation along Aeolus orbit: Example with orbit N°10291 on 1st of June 2020.

4.5 Error estimates

Error estimates are provided as variances. The standard deviation is obtained by taking the square root of these values. Error estimates are based on the SNR and the intensity of the lidar signal. The SNR is based on the hypothesis of Poisson noise which underestimates the actual noise. Also, any error on the calibration coefficients is ignored at the moment.

4.6 Known limitations

Extinction is thresholded to 0 if computation yields a negative result. This practice in the L2A processor might bias statistical studies. This strategy was chosen years ago, because the extinction is too sensitive to small errors that propagate from one bin to the next. Settings negative extinction to 0 provides more realistic results.

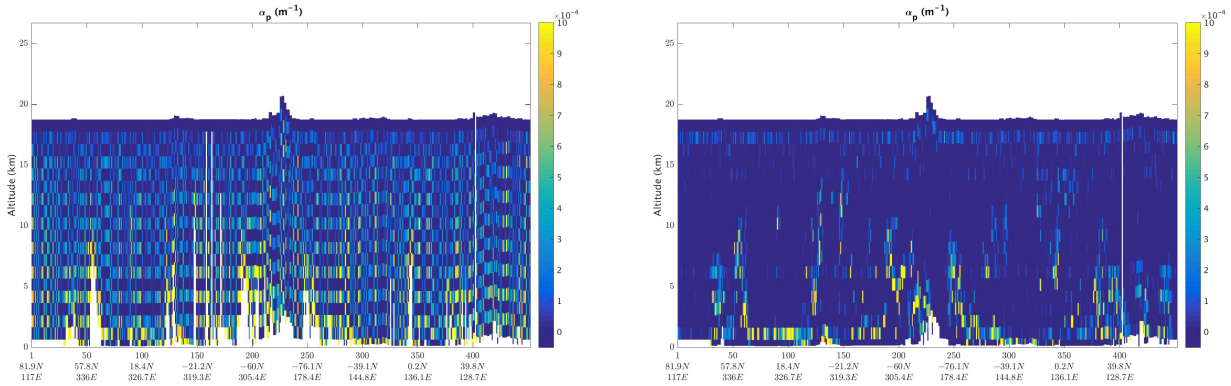


Figure 10. Illustration of the oscillation of extinction when negative values are not thresholded to 0 (left) and in the nominal product (right).¶

4.7 MLE: Maximum Likelihood Estimation Algorithm

4.7.1 De-noising of SCA backscatter and extinction coefficients

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

A physical regularization scheme has been implemented within L2A processor v3.14 to reduce the noise contamination of SCA optical product. It is called Maximum Likelihood Estimation (MLE) or Denoising Scheme. It can mainly be seen as an alternative to the Standard Core Algorithm (SCA) processing of crosstalk corrected signals. The script allows retrieval of extinction for particles and extinction to backscatter ratio (i.e. so-called lidar ratio) within pre-defined physical bounds by the use of the L-BFGS-B open source algorithm.

Examples of output products produced with the SCA and the MLE scheme are illustrated below. In Figure 11 the extinction coefficient in the first processed bin (i.e. first matching bin between Rayleigh vertical grid and Mie vertical grid) appears noisy for the SCA but not for the MLE. The Saharan Air Layer, within the orange box, also appears more homogeneous with the MLE output.

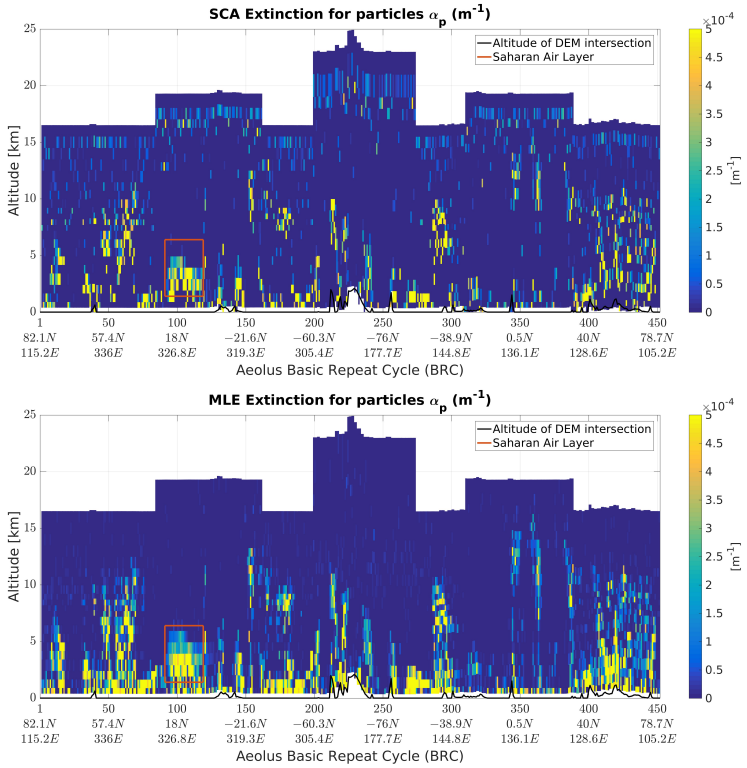


Figure 11. Extinction for particles coefficient processed by SCA algorithm (i.e. top) and de-noised by MLE scheme (i.e. bottom) within B13 prototype code.

4.7.2 The MLE physical constraints and quality flags

The denoising scheme MLE retrieves optimized optical properties corresponding to a physically constraint best variables that matches the solution of a minimization problem solved by L-BFGS-B algorithm. It then assumes that extinction and backscatter are vertically collocated and the lidar ratio retrievals are bounded in expected physical bounds, i.e. between 2 sr and 200 sr. A positivity constraint on the extinction and backscatter is also set. The MLE extinction and backscatter therefore can't be negative hence the number of bins with zero values (e.g. dark blue color code in Figure 12).

To avoid zero division when calculation of the extinction to backscatter ratio it was decided to invalidate the bins where the backscatter is equal to zero. The figure 12 illustrates the MLE extinction and backscatter retrievals for particles as the lidar ratio. The invalid bins, which are coded -1 in the MLE datasets, have been set as NaN for visualization. The lidar ratio product then appears flagged, especially in clear sky conditions.

Moreover, the profiles being initialized as an aerosol free state of the atmosphere, an initialization value of the physically constrained lidar ratio has been set at 60sr with no extinction. As a result, values close to 60 sr (i.e ranging from 55 sr to 65 sr) are given for bins in particle-free regions of the atmosphere instead of 0sr. It is recommended to use the variances for rejecting such estimates. The MLE valid profiles are also aligned with the SCA (i.e. the SCA profiles being not processed for negative Mie signal or too low SNR occurring in top range bins).

Despite being physically coherent the extinction and backscatter MLE retrievals also show high variances below clouds and dense aerosol layers. The retrievals where the minimization lacks of signal with bounded bins LR at 2 sr and 200 sr could be flagged invalid as a positive bias was also observed in low altitudes below 2km.

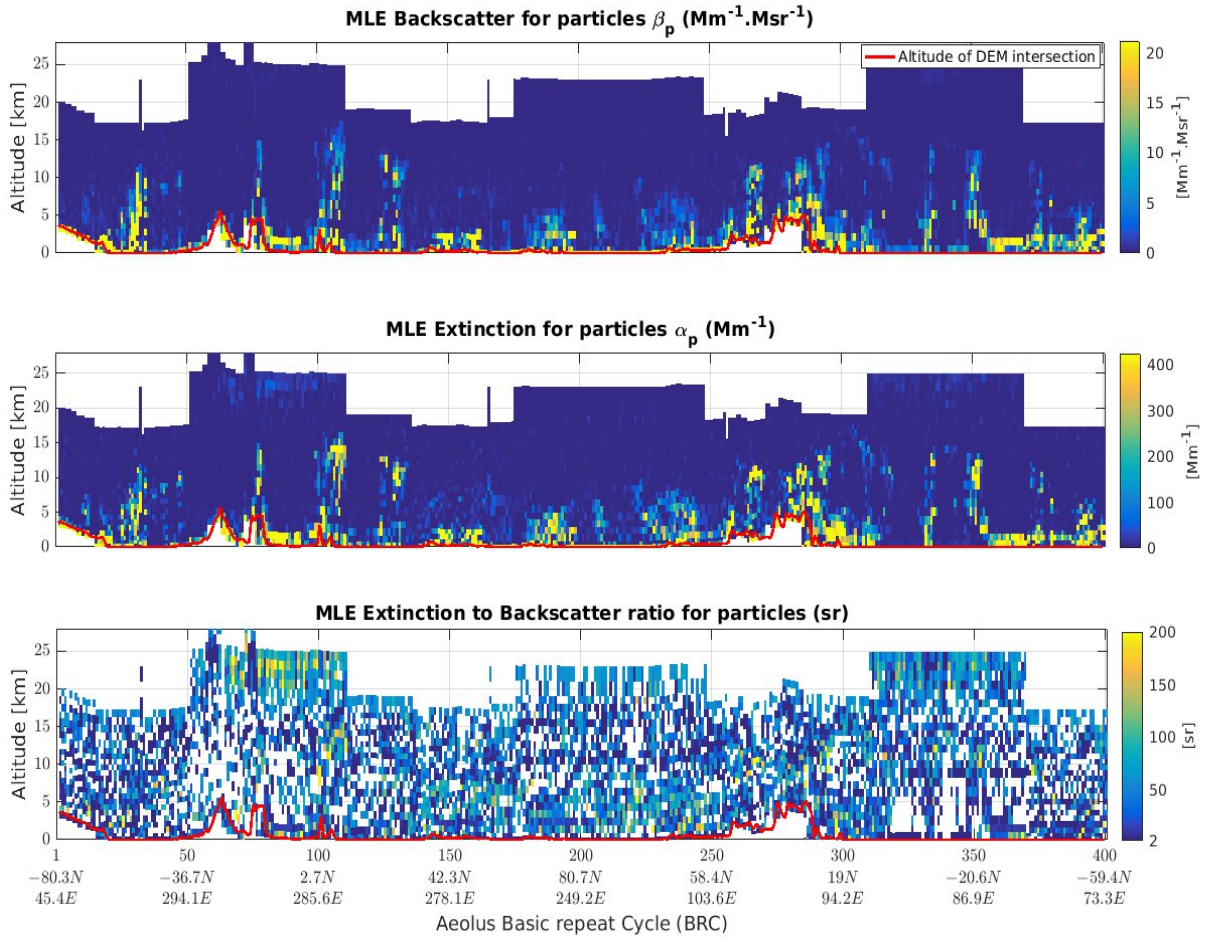


Figure 12. Optimized extinction, backscatter and lidar ratio retrieved by the operational MLE code within B14

The work on Quality Check (QC) flag based on SNR taken from L1B product and using the MLE error estimates similarly to the SCA is ongoing for removing the overestimated outputs. The Rayleigh SNR can indeed be used to remove bounded lidar ratio retrievals. The content of QC flags for MLE algorithm is described in ATBD v6.0, section "9.1.2 Data Quality Flag". A QC flag is provided for the MLE retrievals given at full-BRC scale. The QC is an 8-bit unsigned integer that is the sum of the results of validity checks summed up in Table 4. Each check yields 0 when invalid, 2^n when valid. The users have to split the integer into the results of the check themselves. Examples of MLE products with applied QC are shown in Figure 13.

Table 4. Validity criteria and thresholds used for MLE

Internal name	Possible values	Condition
alpha_valid	0, 1	Ray_SNR_valid ==1 AND alpha_error_bar_valid==1
beta_valid	0, 2	Mie_SNR_valid == 1 AND beta_error_bar_valid == 1
Mie_SNR_valid	0, 4	Mie_SNR > 30
Ray_SNR_valid	0, 8	Ray_SNR > 70
alpha_error_bar_valid	0, 16	Estimated_extinction_error < 1e^{-}2 m^{-1}
beta_error_bar_valid	0, 32	Estimated_backscatter_ error < 1e^{-}3 m^{-1}.sr^{-1}
Not used	0, 63	--

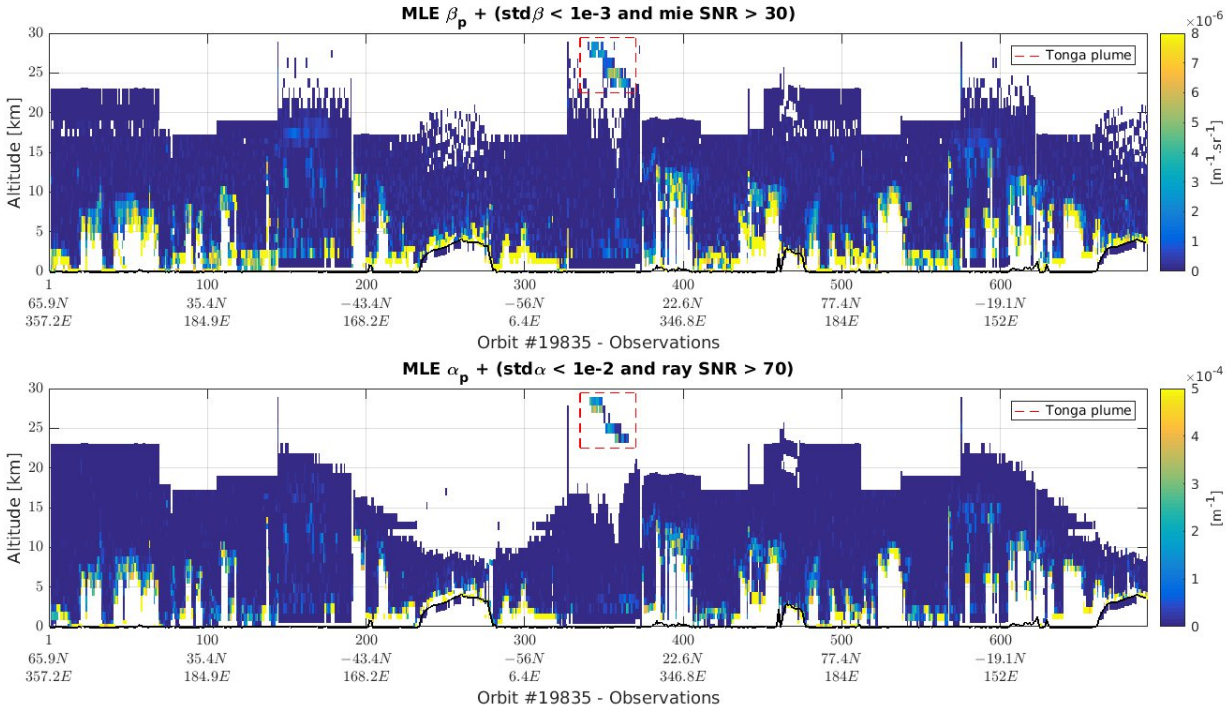


Figure 13. Extinction and backscatter retrieved by MLE with Quality Flag based on SNR and error estimates.

4.7.3 MLE at sub-BRC scale

Since L2A v3.15, the MLE algorithm is executed a second time at an increased horizontal resolution. The idea is to use the MLE approach, less sensitive to the noise, to provide more detailed product. The outputs contained in the new sub-BRC datasets are the same than in the MLE at BRC scale datasets.

The horizontal resolution is defined in the AUX_PAR_2A through a horizontal sampling factor that defined the number of sub-divisions per BRC. The resolution may be tuned depending on the N/P setting of the instrument. N being the number of measurements accumulated in one BRC and P the number of pulses accumulated within one measurement. Since April 2022, N=5 and the horizontal sampling factor has been set to 5 profile per BRC i.e. the horizontal resolution is 1 measurement (17.4 km).

For N=30 and N=15, the setting that were used before, the horizontal sampling factor can be changed in a different AUX_PAR_2A. Compatible options depending on the N/P settings are recoded in Tab. 5:

Table 5. Horizontal resolution possibilities depending on ALADIN N/P settings with 1 BRC = 87 km

	Number of measurements accumulated per sub-profile						
N/P settings	**1**	**2**	**3**	**5**	**6**	**10**	**15**
30/19 (until 2021-12-13)	2.9 km	5.8 km	8.7 km	14.5 km	17.4 km	29 km	43.5 km
15/38 (until 2022-04-04)	5.8 km	-	17.4 km	29 km	-	-	87 km
114/5	17.4 km	-	-	87 km	-	-	-

Figure 14 shows the results for the identical test case presented in Figure 11 In this example N=30 measurements, meaning that the MLE sub-BRC horizontal resolution is 10 measurements.

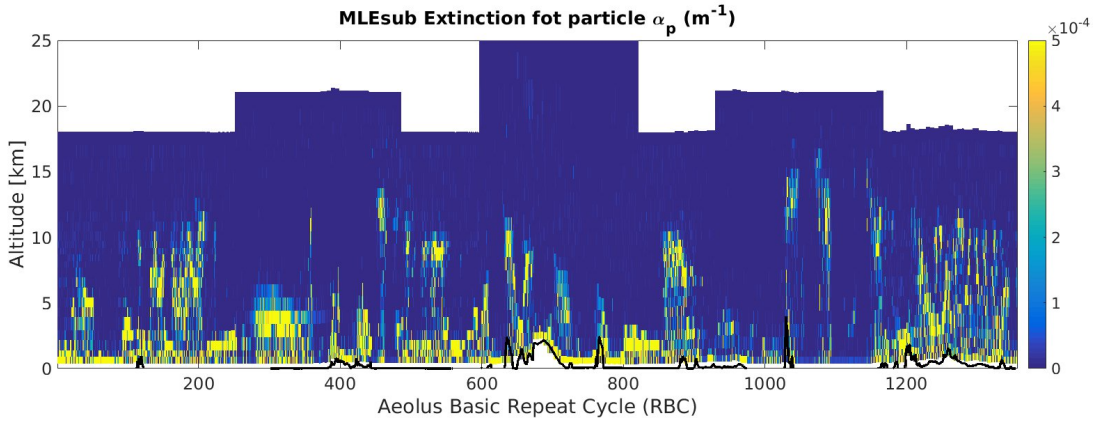


Figure 14. Extinction coefficient estimated by the MLE sub-BRC algorithm for the orbit #10568 on 19 June 2020 and starting at 0751 UTC

A QC flag has been implemented for the MLE retrievals given at sub-BRC scale within L2A prototype version 3.16. It is aligned with MLE full-BRC flagging as the same validity conditions are applied. The QC is then an 8-bit unsigned integer that is the sum of the results of validity checks summed up in Table 5 and each check yields 0 when invalid, 2^n when valid.

Table 5. Validity criteria and thresholds used for MLEsub

Internal name	Possible values	Condition
alpha_valid	0, 1	Ray_SNR_valid ==1 AND alpha_error_bar_valid==1
beta_valid	0, 2	Mie_SNR_valid == 1 AND beta_error_bar_valid == 1
Mie_SNR_valid	0, 4	Mie_SNR > 30
Ray_SNR_valid	0, 8	Ray_SNR > 70
alpha_error_bar_valid	0, 16	Estimated_extinction_error < $1e^{-2}$ m^{-1}
beta_error_bar_valid	0, 32	Estimated_backscatter_ error < $1e^{-3}$ $m^{-1}.sr^{-1}$
Not used	0, 63	--

¶

4.8 Cloud screening based on AMD information

A cloud mask has been implemented in the L2A processor from v3.13. The mask allows the users to flag the bin contaminated by clouds (i.e. without distinction between liquid water clouds, pure ice clouds and mixed clouds). The mask is based on total cloud backscatter derived from both liquid and ice water content information contained in Auxiliary Meteorological Data. A backscatter threshold has been determined and cloud contaminated bins are flagged 1 where the backscatter is superior to 1e-7 $m^{-1}.sr^{-1}$, otherwise the value 0 is assigned. The cloud mask is only provided at observation scale for SCA bin and mid-bin product aligned with coarser L2A Rayleigh grid.

The Figure 15 below shows the derived backscatter from AMD data for a whole orbit on the Rayleigh vertical grid (top) used for setting the cloud screening. The cloud mask product itself given at bin scale for the corresponding orbit is showed in bottom.

The selected case is useful as it shows very low cloud contamination for the core plume of a Saharan Air Layer (SAL) (e.g. orange box) indicating that this particulate feature is indeed purely made of aerosols.

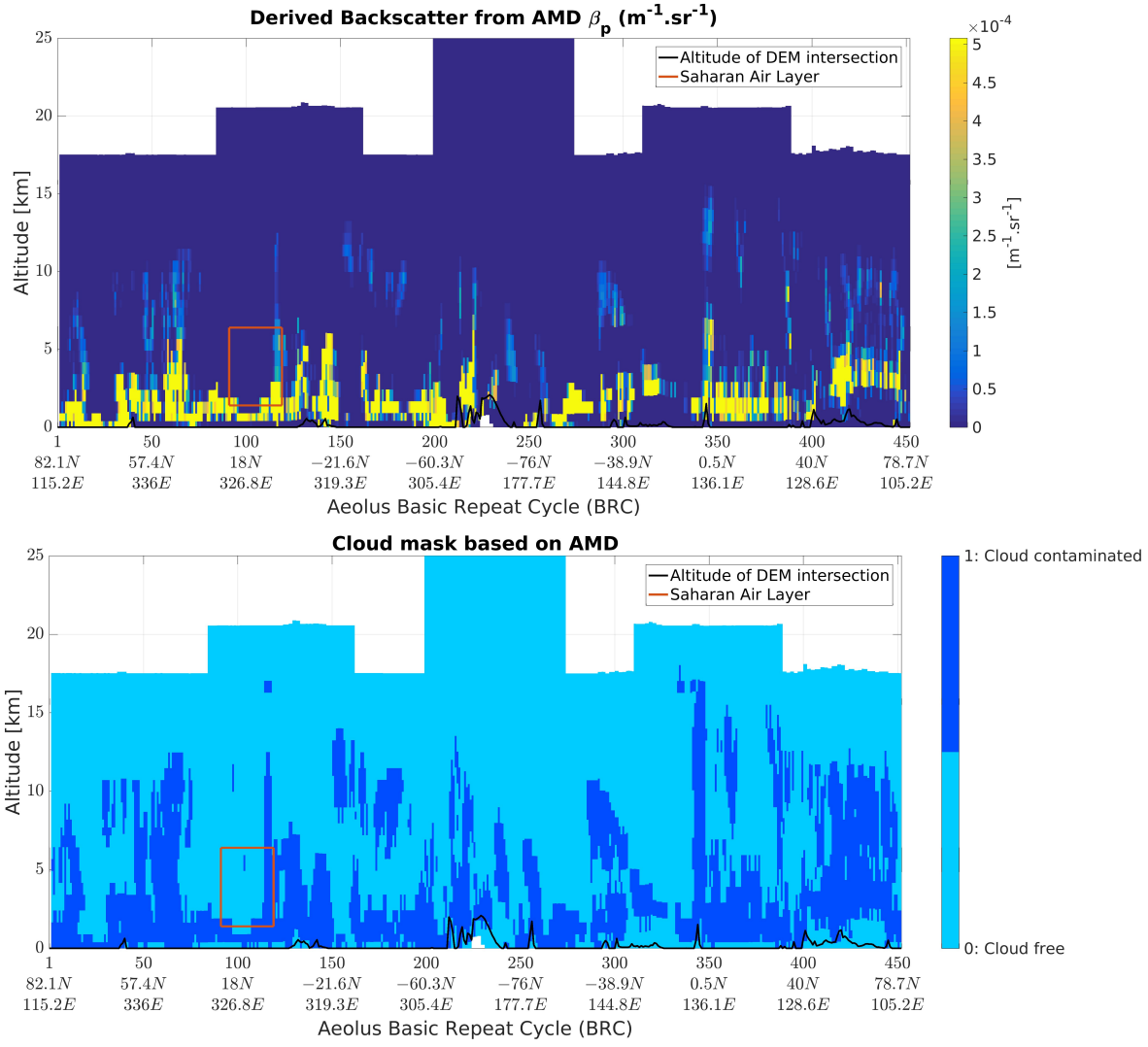


Figure 15. Derived backscatter from AMD information used for cloud flagging (top) and cloud mask product included in SCA_PCD_ADS dataset for bin scale (bottom)

4.9 User-friendly Python script

A short python script (version 2.7 or higher) is made available for extracting and reading main Aeolus Level 2A products by the use of CODA interface (i.e. direct product reading interface). The script is divided in 3 parts: 1) Set environment variables and modules 2) Products extraction through CODA interface 3) Proposal for L2A products visualization: resampling on finer grid.

4.9.1 Set environment variables and modules

This section corresponds to the import of required python modules necessary to data extraction, resampling and plotting. Moreover, it allows the user to set the CODA_DEFINITION (i.e. location of AEOLUS codadef file) to the environment variable as the CODA directory (i.e. location of the Python package). Optional path for figure saving can also be added. For more information about CODA_DEFINITION and Python package see installation instructions for CODA look at the INSTALL file provided with downloadable source package.

For more information about Python CODA interface see

https://stcorp.github.io/coda/doc/html/python/index.html.

4.9.2 Product extraction through CODA

This section makes use of CODA interface main functions coda.open and coda.fetch to open the L2A binary file (*.DBL) product file and return a file handle through which data elements are specified for extraction. Geolocation and data element of two scales (i.e. « measurements » resulting from the accumulation of 19 pulses and « observations » or « BRC » resulting from the accumulation of 30 measurements) are extracted separately. For « observations » data elements: the extinction, backscatter and lidar ratio are extracted for Standard Correct Algorithm (SCA) respectively at Bin and Mid-Bin scales (see Section 4.2 SCA normal vs Mid-bins). For « measurements » data elements only the attenuated backscatters signals are extracted for Standard Correct Algorithm (SCA). Altitude of intersection between WGS84 Digital Elevation Model (DEM) and the satellite line-of-sight for the first measurement (first measurement centroid time of the first BRC) is also extracted and displayed as red line in plot examples. It can be used to easily discriminate overflown continents versus sea.

Additional products as scattering ratio and extinction, backscatter for Mie Core Algorithm (MCA) are also extracted. The script requires the user choice for valid or non-valid data quality flag. Details of quality flag are explained in section 3.5. The extraction of quality check (QC) flag consider the particular format though which each validity check is saved (i.e. 7 useful bits within an 8-bit unsigned integer for bin and 8-bit unsigned integer for mid-bin products). Therefore, the unsigned integer are converted to boolean sequences (i.e. the split method introduced in section 3.5) and each result of the check corresponding to quality flag is assigned to extinction or backscatter signals.

4.9.3 Proposal for L2A products visualization: resampling on finer grid

The python script includes a function called resampling which reinterprets Aeolus L2A main products: one proper way to read the L2A product is to re-compute a finer altitude grid and get interpolated values in order to make color contrast perception easier between bins. Basically, the idea is to re-scale the signals by defining a finer altitude grid which is taken for interpolation. Therefore, the first step is to define a new vertical scale and get the altitude interval depending on Range Bin Settings (RBS) allocated for each observation (i.e. BRC). Then an interpolation is performed to assign values of new re-scaled color matrix. Finally, we can create a pseudocolor plot with finer grid (e.g. Figure 16). Please note that a grey background is applied to supplement default colormap viridis of matplotlib library. Moreover, a plotting procedure with logarithmic scale with SCA mid-bin extinction signal is given as example at the end of the script.

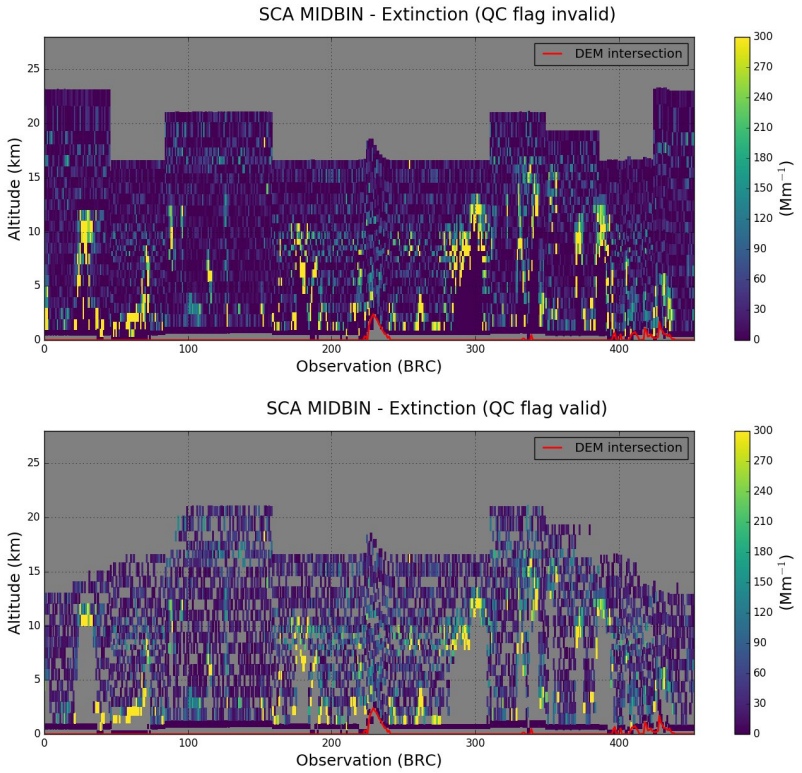


Figure 16. Illustration of SCA mid-bin extinction nominal product (top) and with valid SNR QC flag (bottom).

4.10 Aeolus papers and L2A algorithm

For more details about L2A Processor algorithm and applications, a non-exhaustive list of Aeolus papers is given below:

- Pierre Flamant, Juan Cuesta, Marie-LAURE Denneulin, Alain Dabas & Dorit Huber (2008) ADM-Aeolus retrieval algorithms for aerosol and cloud products,Tellus A: Dynamic Meteorology and Oceanography, 60:2, 273-286, DOI: 10.1111/j.1600-0870.2007.00287.x

- David G. H. Tan, Erik Andersson, Jos De Kloe, Gert-Jan Marseille, Ad Stoffelen, Paul Poli, Marie-LAURE Denneulin, Alain Dabas, Dorit Huber, Oliver Reitebuch, Pierre Flamant, Olivier Le Rille & Herbert Nett (2008) The ADM-Aeolus wind retrieval algorithms, Tellus A: Dynamic Meteorology and Oceanography, 60:2,191-205, DOI: 10.1111/j.1600-0870.2007.00285.x

- A. Dabas, M. L. Denneulin, P. Flamant, C. Loth, A. Garnier & A. Dolfi-Bouteyre (2008) Correcting winds measured with a Rayleigh Doppler lidar from pressure and temperature effects, Tellus A: Dynamic Meteorology and Oceanography, 60:2, 206-215, DOI: 10.1111/j.1600-0870.2007.00284.x

- Baars, H., Radenz, M., Floutsi, A. A., Engelmann, R., Althausen, D., Heese, B., et al. (2021). Californian wildfire smoke over Europe: A first example of the aerosol observing capabilities of Aeolus compared to ground-based lidar. Geophysical Research Letters, 48, e2020GL092194. https://doi.org/10.1029/2020GL092194

- Ehlers, F., Flament, T., Dabas, A., Trapon, D., Lacour, A., Baars, H., and Straume-Lindner, A. G.: Optimization of Aeolus Optical Properties Products by Maximum-Likelihood Estimation, Atmos. Meas. Tech., 15, 185–203, https://doi.org/10.5194/amt-15-185-2022, 2022.

- Flament, T., Trapon, D., Lacour, A., Dabas, A., Ehlers, F., and Huber, D.: Aeolus L2A Aerosol Optical Properties Product: Standard Correct Algorithm and Mie Correct Algorithm, Atmos. Meas. Tech., 14, 7851–7871, https://doi.org/10.5194/amt-14-7851-2021, 2021.