Royal Netherlands
Meteorological Institute
Ministry of Infrastructure
and Environment.

TROPOS

Freie Universität Berlin

Algorithm Theoretical Basis Document (ATBD)

Processor: *A-L2A A-PRO*
Products: *A-AER, A-EBD, A-ICE, A-TC*
Software Version: 8.1

Prepared by: D.P. Donovan, G-J van Zadelhoff, J. Williams (KNMI)
U. Wandinger (TROPOS)

Reference	***A-PRO_ATBD***
Issue	***8***
Revision	**1**
Date of Issue	***12/03/2021***
Document Type	**ATBD**
Distribution	

CHANGE LOG

Reason for change	Issue	Revision	Date
V8.1 Replaced some faulty “true” plots. depol values of 0 were being set to white. Fixed broken references. Addressed [ACAPRIL-224]	8	1	11/03/2021
Chapter 6and the scene Appendix are still far from properly finished, in the sense that much work remains to be done. However, they are now “complete” enough to be self-standing. [ECAPRIL-225]	8	1	07/03/2021
V8.0 Updated calculation of a-priori errors in log space to correct form. Worked on the validation section. Still (very) incomplete. Moved Scene descriptions to an Appendix. Will also include info on the GEM adjustment work.	8	0	22/01/2021
V7.0 -Fixed typos in into equations. -Flagged obsolete material in Section 6. The whole of that section needs a re-write!	7	0	04/06/2020
Update for AR6: GJvZ: Updated classification sections to reflect changes associated with addition of stratospheric typing. D.D. Typo fixes and updated config tables to reflect recent changes	6	0	20/06/2019
V5.01 -D.D. Added material addressing the proper accounting of the Rayleigh depolarization. -D.D. Updated a number of algorithm steps based on recent code changes. (some inaccuracies may still remain though).	5	01	05/02/2019
Update for V5.0 D.D. Note: It has been found that the neglect of the Rayleigh depol signal in a number of algorithm steps is inaccurate for low depolarizing and low backscatter targets (e.g. optically thin spherical aerosols). This finding means that the SW needs to be corrected as well as the relevant parts of the ATBD. Due to lack of time before the delivery this has NOT been included in this update but will be included in the next revision.	5	0	21/11/2018
Major updates in preparation for AR4	3	2	06/11/2017
Update for AR4: Much work still needs to be done though	3	1	25/09/2017
D.D. Fixed wrong equations in x-talk discussion. Added notes here and here.	2	09	29/07/2017
D.D Some minor fixes. (mainly in the ice section)	2	08	17/07/2017
D.D Note: much scientific work still needs to be included.	2	0	06/02/2017
D.D. and D.M.D: Edits in preparation of TRR2 Section 6.4 and 6.5 will need to be updated/expanded for AR2 with latest results.	1	4	15/12/2016
D.D. Major post-AR1 edits.	1	1	30/05/2016

D.M.D. changes based on the TRR1 comments and the AER algorithm	1	0	14/03/2016
D.D. Editorial fixes	0	3	28/10/2015
D.M.D. added several parts related to the A_AER and A_ICE	0	2	30/09/2015
D.D. Creation (Merging and revision of much ATLAS material)	0	1	09/03/2015

Table of contents

List of Figures

1 PURPOSE AND SCOPE

This document describes the algorithms within the EarthCARE L2a Lidar Extinction, Backscatter and Depolarization processor. Within this processor four distinct algorithms exist, a procedure aimed at deriving the large-scale aerosol (and thin cloud) extinction and backscatter (A-AER), an optimal estimation Extinction and Backscatter retrieval algorithm (A-EBD), a lidar classification procedure (A-TC) and an ice microphysical property estimation algorithm (A-ICE). Several output products are generated by this processor. Collectively, these algorithms output high (1km) and low (10km+) horizontal resolution profiles of lidar extinction, backscatter, optical depth, particle type, ice effective radius and ice water content. This document presents the theoretical background of the algorithms that comprise the processor as well as presenting various examples.

2 APPLICABLE AND REFERENCE DOCUMENTS

2.1 Applicable documents

Reference Documents	
[AD1]	A-FM PDD , Version 8.1, Mar 2021
[AD2]	A-AER PDD , Version 8.1, Mar, 2021
[AD3]	A-EBD PDD , Version 8.1, Mar, 2021
[AD4]	A-TC PDD , Version 8.1, Mar, 2021
[AD5]	A-ICE PDD , Version 8.1, Mar, 2021
[AD6]	Earth Observation Ground Segment File Format Standard, PE-TN-ESA-GS-0001, issue 2.0, 3 May 2012
[AD7]	ATLID L1b PDD, EC.ICD.GMV.ATL.00001_01.00_ATL L1 Product Definition Document - Volume A (ATL_PDD_N)_2015-05-15
[AD8]	ATLID L1b ATBD, EC.TN.ASF.ATL.00034, Issue 06_Rev00 ATBD_ECGP

2.2 Reference documents

Keyword	Reference
Achert et al. 2014	Achtert, P., and M. Tesche (2014), Assessing lidar-based classification schemes for polar stratospheric clouds based on 16 years of measurements at Esrange, Sweden, J. Geophys. Res. Atmos., 119, 1386–1405, doi:10.1002/2013JD020355.
Baars et al. 2012	Baars, H., A. Ansmann, D. Althausen, R. Engelmann, B. Heese, D. Müller, P. Artaxo, M. Paixao, T. Pauliquevis, and R. Souza, 2012: Aerosol profiling with lidar in the Amazon Basin during the wet and dry season. *J. Geophys. Res.,* **117**, D21201, doi:10.1029/2012JD018338.
Donovan et al., 2015	Donovan, D. P., Klein Baltink, H., Henzing, J. S., de Roode, S. R., and Siebesma, A. P.: A depolarisation lidar-based method for the determination of liquid-cloud microphysical properties, Atmos. Meas. Tech., 8, 237–266, https://doi.org/10.5194/amt-8-237-2015, 2015.
Erfani, and Mitchell ,2016	Erfani, E. and Mitchell, D. L.: Developing and bounding ice particle mass- and area-dimension expressions for use in atmospheric models and remote sensing, Atmos. Chem. Phys., 16, 4379–4400, https://doi.org/10.5194/acp-16-4379-2016, 2016.
Foot, 1988	Foot, J. S. (1988), Some observations of the optical properties of clouds. Part II: Cirrus,

	Q. J. R. Meteorol. Soc., 114, 145 – 164.
Groß, S., et al, 2011	Groß, S., M. Tesche, V. Freudenthaler, C. Toledano, M. Wiegner, A. Ansmann, D. Althausen, and M. Seefeldner, 2011: Characterization of Saharan dust, marine aerosols and mixtures of biomass-burning aerosols and dust by means of multi-wavelength depolarization and Raman lidar measurements during SAMUM 2. *Tellus B*, **63**, 706–724.
Groß, S., et al, 2012	Groß, S., V. Freudenthaler, M. Wiegner, J. Gasteiger, A. Geiß, and F. Schnell, 2012: Dual-wavelength linear depolarization ratio of volcanic aerosols: Lidar measurements of the Eyjafjallajökull plume over Maisach, Germany. *Atm. Env.*, **48**, 85–96.
Gross et al, 1995	Michael R. Gross, Thomas J. McGee, Upendra N. Singh, and Patrick Kimvilakani, "Measurements of stratospheric aerosols with a combined elastic–Raman-backscatter lidar," Appl. Opt. 34, 6915-6924 (1995)
Heintzenberg, 2009	The SAMUM-1 experiment over Southern Morocco: Overview and introduction. Tellus 61B, 2-11, DOI: 10.1111/j.1600-0889.2008.00403.x
Heymsfield et al., 2005	Heymsfield, A. J., D. Winker, and G.-J. van Zadelhoff (2005), Extinction-ice water content-effective radius algorithms for CALIPSO, *Geophys. Res. Lett.*, 32, L10807, doi:10.1029/2005GL022742.
Heysmfield et. al.,2013	Heymsfield, A.,J., C. Schmitt ,and A. Bansemer, Ice Cloud Particle Size Distributions and Pressure-Dependent Terminal Velocities from In Situ Observations at Temperatures from -8 to -86C. J. Amos. Sci., DOI: 10.1175/JAS-D-12-0124.1 (2013)
Hogan 2006	Fast, approximate calculation of multiply scattered lidar returns, Applied Optics, 45, 5984-5992 (2006).
Hogan and Kew 2005	Hogan, R. J., and S. F. Kew: A 3D stochastic cloud model for investigating the radiative properties of inhomogeneous cirrus clouds. Q. J. R. Meteorol. Soc., 131, 2585-2608 (2005)
Hu. 2007	Hu, Y., 2007: Depolarization ratio–effective lidar ratio relation: Theoretical basis for space lidar cloud phase discrimination. Geophys. Res. Lett., 34, L11812, doi:10.1029/2007GL029584.
Hu. et al. 2009	Hu Y., D. Winker, M. Vaughan, B. Lin, A. Omar, C. Trepte, D. Flittner, P. Yang, S.L. Nasiri, B. Baum, W. Sun, Z. Liu, Z. Wang, S. Young, K. Stamnes, J. Huang, R. Kuehn and R. Holz, "CALIPSO/CALIOP Cloud Phase Discrimination Algorithm", JTECH, DOI: 10.1175/2009JTECHA1280.1
Kittaka et al. 2008	Kittaka, C.; Winker, D.; Omar, A.; Liu, Z.; Vaughan, M.; Trepte, C., Global Aerosol Distributions Derived From the CALIPSO Observations, American Geophysical Union, Fall Meeting 2008, abstract #A41A-0078
Klett 1985	Klett, J.D. Lidar inversion with variable backscatter/extinction ratios Applied Optics, Vol. 24, Issue 11, pp. 1638-1643 (1985) doi:10.1364/AO.24.001638
Muller et al. 2011	Muller, D. Ansmann, A. and Petzold, A , ICAROHS final report; Recommendations for future single and multi-wavelength HSRL instruments
Nocedal and Wright (2006)	Nocedal J. and Wright S.J., Numerical Optimization, Springer, New York, NY, USA (2006).
Noel et al. (2009)	Noel, V., A. Hertzog, and H. Chepfer (2009), CALIPSO observations of wave-induced PSCs with near-unity optical depth over Antarctica in 2006 – 2007, J. Geophys. Res., 114, D05202, doi:10.1029/2008JD010604.
Omar et al. 2009	Omar, Ali H., and Coauthors, 2009: The calipso automated aerosol classification and lidar ratio selection algorithm. J. Atmos. Oceanic Technol., 26, 1994–2014. doi: 10.1175/2009JTECHA1231.1
Platt 1973	Platt, C. M. R., 1973: Lidar and radiometric observations of cirrus clouds, J. Atmos. Sci., 30, 1191–1204.
Press et al. 1992	Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannerty, B.P. (1992), Numerical Recipes in C: The Art of Scientific Computing, (2nd Edition), Cambridge University Press, New York.

Shipley 1983	Shipley, S.T., D.H. Tracy, E.W. Eloranta, J.T. Trauger, J.T. Sroga, F.L. Roesler and J.A. Weinman, "A High Spectral Resolution Lidar to Measure Optical Scattering Properties of Atmospheric Aerosols, Part I: Instrumentation and Theory", Applied Optics, 23, 3716-3724, 1983
Rodgers 2000	Rodgers, C.D., “Inverse methods for Atmospheric Sounding; Theory and Practice”, Series on Atmospheric, Oceanic and Planetary Physics-Vol. 2, World Scientific, ISBN 981-02-2740-X, 2000.
Weinzierl and van Zadelhoff 2011	Weinzierl, B. and van Zadelhoff, G-J, Technical Note 2: ICAROHS project, 2011

3 SCIENTIFIC BACKGROUND

Algorithm history

The A-PRO processor is comprised of a number of separate interacting algorithms namely

- **A-AER** : Aerosol oriented large-horizontal scale (10+ km) retrievals of aerosol (and thin cloud) optical properties.
- **A-EBD** : Cloud and aerosol optical property retrievals at high (1-km) horizontal resolution.
- **A-TC :** Target classification and aerosol type discrimination using the A-AER and A-EBD as inputs
- **A-ICE:** Empirically based estimation of ice cloud effective radius and Ice-Water-Content (IWC) using retrieved extinctions and ancillary temperature data.

The algorithms comprising the A-PRO processor (the A-EBD, A-AER, A-TC and A-ICE procedures) are mainly built upon the corresponding algorithms developed during the ESA EarthCARE ATLAS project. However, significant changes to the A-EBD algorithm core have been made. In particular, a novel treatment of lidar Multiple-Scattering (MS) has been formulated and used in the present processor.

3.1 Algorithm introduction

3.1.1 Measurement Principle

ATLID is a so-called High Spectral Resolution Lidar (HSRL). In the case of ATLID an etalon is used to (partially)-separate the spectrally narrow return from aerosols and clouds ('Mie [1]') from the thermally broadened return from atmospheric molecules ('Rayleigh'[2]). This is semantically depicted in Figure 1. In order to separately quantify the Mie and Rayleigh scattering contributions a cross-talk correction procedure is applied as part of the L1 processing (See Figure 2).

---

[1] Mie scattering, properly, refers only to scattering by perfect spheres. Here (and indeed through much of EarthCARE related documentation) we use the term rather loosely to broadly cover what should be termed `particulate' scattering.

[2] As is the case with our use of `Mie' scattering, the term 'Rayleigh' scattering is used loosely. A more accurate term would be `molecular' or `Rayleigh-Brillouin' scattering.

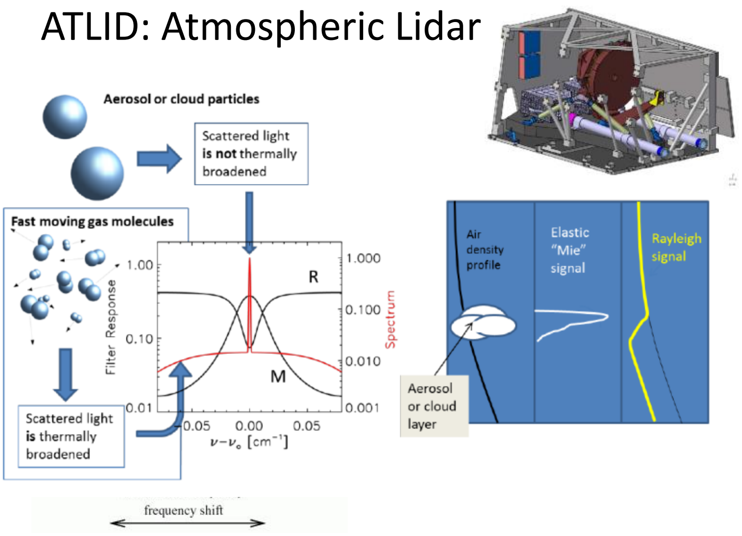


Figure 1: ATLID measurement principle

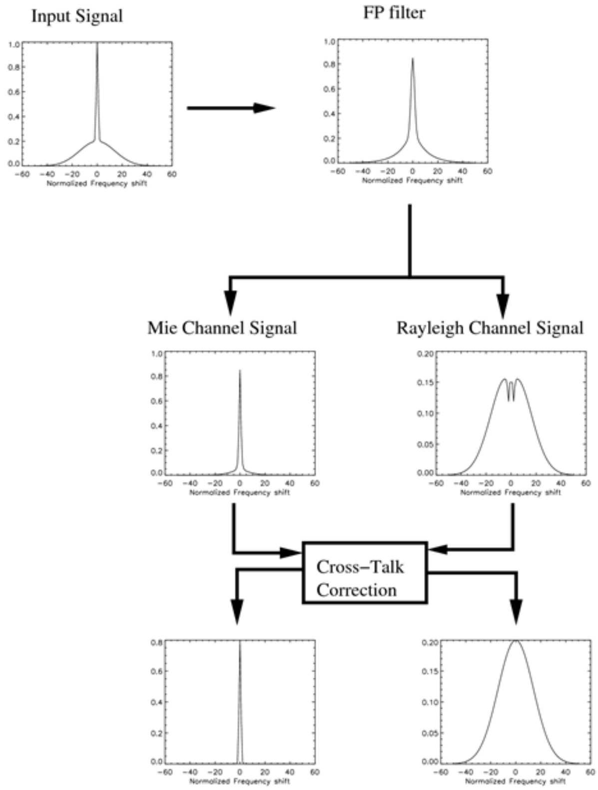


Figure 2: Schematic depiction of the ATLID Cross-talk correction procedure. The cross-polarization total (Mie+Rayleigh) channel is not depicted.

After cross-talk correction and calibration, the three ATLID channels can be related to the atmospheric extinction and backscatters (neglecting multiple-scattering effects) as:

$$b_{R,\|}(z) = p_{R,\|}(z)r(z)^2 = \beta_{R,\|}(z)\exp\left[-2\int_{z_{lid}}^{z}(\alpha_M(z')+\alpha_R(z'))dr'\right] \tag{2.2.1}$$

$$b_{M,\|}(z) = p_{M,\|}(z)r(z)^2 = \beta_{M,\|}(z)\exp\left[-2\int_{z_{lid}}^{z}(\alpha_M(z')+\alpha_R(z'))dr'\right] \tag{2.2.2}$$

$$b_{T,\perp}(z) = p_{T,\perp}(z)r(z)^2 = \left(\beta_{M,\perp}(z)+\beta_{R,\perp}(z)\right)\exp\left[-2\int_{z_{lid}}^{z}(\alpha_M(z')+\alpha_R(z'))dr'\right]. \tag{2.2.3}$$

Where $b_{R,\|}$ is the Rayleigh attenuated backscatter, $b_{R,\|}$ is the Mie attenuated backscatter and $b_{T,\perp}$ is the total cross-polar attenuated backscatter., $p_{R,\|}$ is the Rayleigh co-polar power, $p_{M,\|}$ is the Mie co-polar power and $p_{T,\perp}$ is the total cross-polar power. z is the atmospheric altitude and $r(z)$ is the range from the lidar. α_M is the aerosol and cloud extinction and α_R is the atmospheric Rayleigh extinction. $\beta_{M,\|}$ is the co-polar Mie backscatter, $\beta_{R,\|}$ is the co-polar Rayleigh backscatter, $\beta_{M,\perp}$ is the cross-polar Mie backscatter and $\beta_{R,\perp}$ is the cross-polar Rayleigh backscatter.
Note: The equations above are appropriate for calibrated signals (i.e. explicit calibrations factors are all assumed to have a value of 1).

ATLID does not directly measure the Mie-only cross polarized attenuated backscatter, nor the total Rayleigh or total Mie attenuated backscatters, however, these quantities (which will be seen to be necessary for the inversion of the lidar signals) can be deduced from Eqs. (2.2.1)-(2.2.3) . e.g.

$$b_{M,\perp}(z) = b_{T,\perp}(z) - b_{R,\|}(z)\delta_R \tag{2.2.3}$$

$$b_R(z) = b_{R,\|}(z)\left(1+\delta_R\right) \tag{2.2.3}$$

$$b_M(z) = b_{M,\|}(z) + b_{M,\perp}(z) = b_{M,\|}(z) + b_{T,\perp}(z) - b_{R,\|}(z)\delta_R \tag{2.2.3}$$

where $b_{M,\perp}$ is the Mie cross-polar attenuated backscatter, b_R is the total (co- + cross-polar) Rayleigh attenuated backscatter, b_M is the Mie total (co- + cross-polar) attenuated backscatter and δ_R is the Rayleigh linear depolarization ratio. The linear depolarization ratio for Rayleigh scattering in the case of EarthCARE where the receiver band-width is small enough such that the Raman wings are largely excluded is on the order of 0.4%.

The primary function of this processor is to invert the lidar signals to obtain estimates of backscatter and extinction, and using these values together with the particle linear depolarization ratio ($\delta_M = \left(\beta_{M,\perp} / \beta_{M,\|}\right) = \left(b_{M,\perp} / b_{M,\|}\right)$) to classify the extinction and backscatter values as being associated with ice cloud, water cloud or aerosol (type).

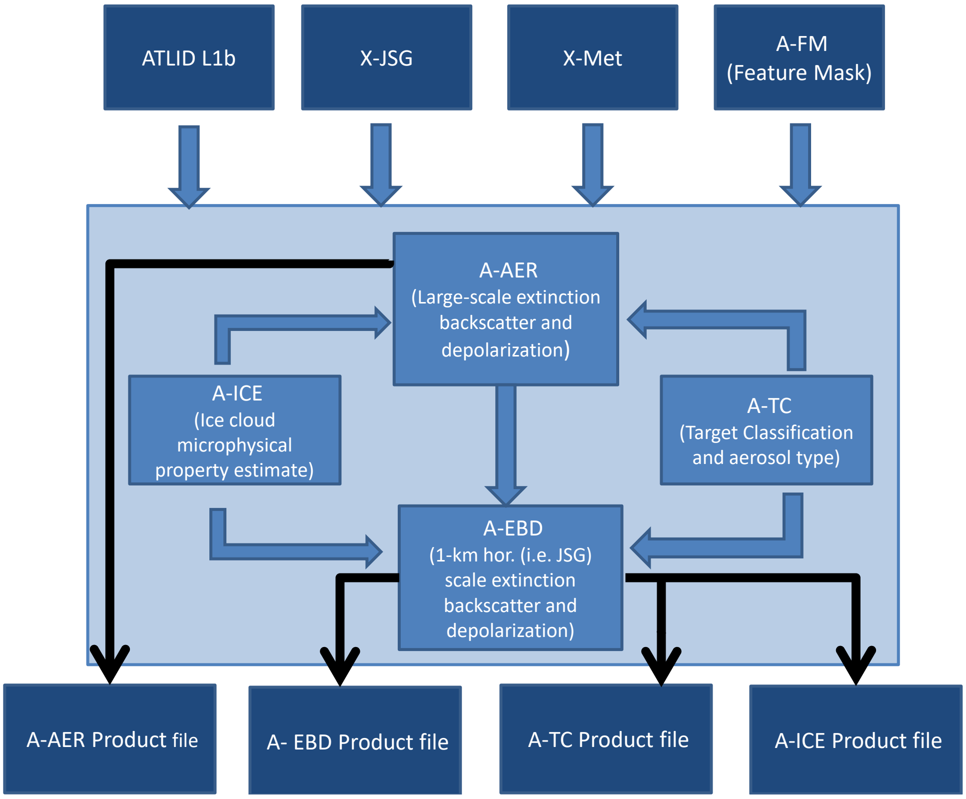


Figure 3: Schematic overview of the relationship between the algorithms and products within the A-PRO processor

The relationship between the algorithms within the A-Prof processor is shown in Figure 3.It should be noted that aspects of the A-TC and A-ICE algorithms will be also "called" from the A-AER algorithm. For both A-AER and A-EBD correcting for the effects of multiple-scattering in the lidar signals requires an a priori estimate of multiple-scattering parameters related to the effective particle size. This ice cloud effective radius (based on a parameterization which uses temperature as an input) will be provided via the A-ICE algorithm insuring consistency between A-AER and A-EBD.

3.1.2 A-AER

The APRIL A-AER algorithm is based on mainly conventional HSRL lidar approaches . At its core, this algorithm relies on a fairly direct method to derive the aerosol extinction and backscatter from the Rayleigh and Mie signals e.g. for the extinction. Including a calibration factor and a correction for Multiple-scattering, the equation describing the Rayleigh attenuated backscatter can be written as:

$$b_R(z) = (1+\delta_R)\,p_{R,\|}(r)\,r(z)^2 C_R^{-1} = \beta_R(z) M_R(z) \exp[-2\int_0^z (\alpha_M(z)+\alpha_R(z))dz'] \qquad (2.2.4)$$

where M_R is the multiple-scattering correction (MS) factor appropriate for the Rayleigh channel and $\mathrm{B}_R(z)$ is the attenuated Rayleigh channel backscatter). The corresponding equation for the Mie channel can be written as

$$b_M(z) = r(z)^2\left[C_{M,\|}^{-1} p_{M,\|}(r) + C_\perp^{-1} p_{T,\perp} - C_{R,\|}^{-1}\delta_R\, p_{R,\|}\right] = \beta_M(z) M_M(z) \exp[-2\int_0^z (\alpha_M(z)+\alpha_R(z))dz'].$$

(2.2.5)

where M_M the multiple-scattering factor appropriate for the Mie channel. Starting from Eq. (2.2.4), by taking logs and differentiating, it can be shown that (note the C_R and δ_R terms fall away since they are effectively constant).

$$\alpha_M(z) = \frac{1}{2}\left[\frac{1}{\beta_R(z)}\frac{d}{dz}(\beta_R) + \frac{1}{M_R}\frac{d}{dz}(M_R) - \frac{1}{p_R r(z)^2}\frac{d}{dz}\left(p_R r(z)^2\right)\right] - \alpha_R(z) . (2.2.6)$$

Assuming that M_R is negligible for a given situation and noting that β_R and α_M are known functions of the (known) atmospheric density, then, in principle, Eq.(2.2.6) provides a direct means for estimating the Mie extinction.

The total particulate backscatter profile can also be rather directly estimated using Eqns. (2.2.4) and (2.2.5), i.e.

$$\beta_M(z) = \beta_R(z)\left(\frac{b_M(z)}{b_R(z)}\right)\left(\frac{M_R^{-1}(r)}{M_M^{-1}(r)}\right)\exp[2\int_0^z (\alpha_M^e(z) + \alpha_R(z))dz'] . \qquad (2.2.7)$$

The methods described above (apart from the complications arising from Multiple-scattering) are straight forward in principle. However, these methods are particularly affected by the presence of noise. Thus, they can only be applied if the signal profiles are averaged to achieve a seasonable SNR. Due to the non-linear nature of the lidar equation this averaging must be done in an intelligent manner taking into account the presence of thick clouds and variations in the surface height. Thus the specification of the cloud/surface screening and the averaging strategy are major practical components of this algorithm. For completeness, lidar multiple-scattering factor appear in the above discussion. However, in this present version of the A-AER algorithm, MS effects are assumed to be negligible. This is a limitation and may be addressed in a future version of this algorithm. MS effects are however, extensively treated within the companion A-EBD algorithm.

3.1.2.1 Cloud and Surface Screening

Since the relationship between extinction and lidar signal non-linear, “blindly averaging” the lidar signal over a significant horizontal domain will lead to biased results if large horizontal variations in the optical properties are present. Thus, in order to accurately retrieve aerosols properties, the algorithm must be able to deal with any potential cloud (or in general any “strong” feature) contamination inside the averaging domains. In essence, we want to avoid averaging cloud and aerosol signals together.

As a first step, the attenuated backscatter signals are smoothed using box-car averaging. The smoothing window is configurable but should not be more than a few horizontal or vertical pixels. The lidar scattering ratio

$$R(z) = \left(\frac{\beta_M(z) + \beta_R(z) + \beta_\perp(z)}{\beta_R(z)}\right) \approx \left(1 + \frac{b_\perp(z) + b_M(z)}{b_M(z)}\right) \qquad (2.2.8)$$

is then calculated using the smoothed signals. Here small averaging domains are appropriate since this scattering ratio is only used to identify “strong” features and is not used further for quantitative inversion purposes. A threshold is applied to the smoothed scattering ratio (together with a threshold on the input A-FM field) to generate a “strong-feature” mask. The mask is then extended to exclude surface points as well as altitudes below any “strong feature” to generate a horizontal averaging mask. This mask is then used in the next step to guide the horizontal averaging of the attenuated backscatter profiles.

3.1.2.2 Averaging in the horizontal dimension to achieve the required SNR

As previously mentioned, in order to perform the HSRL extinction and backscatter inversion, a minimum SNR is required. The signals are smoothed using a box-car window for each altitude level respecting the horizontal averaging mask described in the previous section . Note that the width can be height dependent. The SNR ratio profile of the average data is calculated using the standard deviation of the cloud screened signal within the window. The width of the box is increased until a minimum average SNR threshold is reached or a maximum window width is reached.

It should be noted that increasing the size of the sliding window in the vertical dimension can compensate for reducing the width of the window in the horizontal dimension.

3.1.2.3 Evaluation of the Rayleigh Channel logarithmic derivative

Once the signals have been smoothed appropriately Eq. (2.2.6) can be evaluated. The key-step in this evaluation is the estimation of the $\frac{1}{p_R r(z)^2}\frac{d}{dz}\left(p_R r(z)^2\right) \equiv \frac{d}{dz}\ln\left(p_R r(z)^2\right)$ term. This is accomplished by the least squares fitting of a polynomial of degree 1 using a sliding window in the vertical. The slope of the line is used to estimate the derivative and the slope uncertainty is used to estimate the derivative uncertainty. Special care must be taken to estimate the derivative within half the window width from the ground. This is accomplished by using a small window width close to the ground down to a minimum size of 3 range gates.

3.1.2.4 MS effects

As previously noted, MS effects are not included in the current implantation of A-AER, however, MS effects are relevant and should be accounted for in a future version of the algorithm. A-AER is restricted to optically “thin” targets and it may be (incorrectly) supposed that MS effects are not-significant. Even in the case of optically thin targets MS effects can be important. In particular, two situations in which MS effects on the retrieved aerosol optical properties may require correction are:

1. The presence of large (e.g. dust) aerosols which causes the effective extinction within the aerosol to be less than the actual extinction. Within aerosol layers (and within ice clouds) this effect could possibly be accounted for using the MS extinction correction factor approach i.e. (Platt, 1981)

2. Aerosols underlying thin cirrus clouds. In this case there can be enough signal from the below cirrus aerosol regions to conduct an inversion. However, the “tail” of multiply scattered photons below the higher altitude cirrus clouds may bias the retrieval results to a degree where it becomes necessary to implement a correction. This effect is expected to be greatest for ice clouds with small particle sizes (e.g. high cold cirrus clouds: common in tropical areas).

   An example of the MS “tail” under an idealized cirrus cloud is shown in Figure 4. Here the Rayleigh return directly under the cloud is elevated due to MS. As the lidar pulse propagates downwards the return “relaxes” towards the single-scattering return. There is no aerosol under the cirrus but the relative impact on the Rayleigh signal below cloud in the presence of aerosol is expected to be the same to a good approximation. The tail affects the slope of the Rayleigh signal and will lead to an over estimation of the aerosol extinction if it is not accounted for. Further, the tail cannot be modelled by Platt’s approach and a more accurate MS forward model must be used.

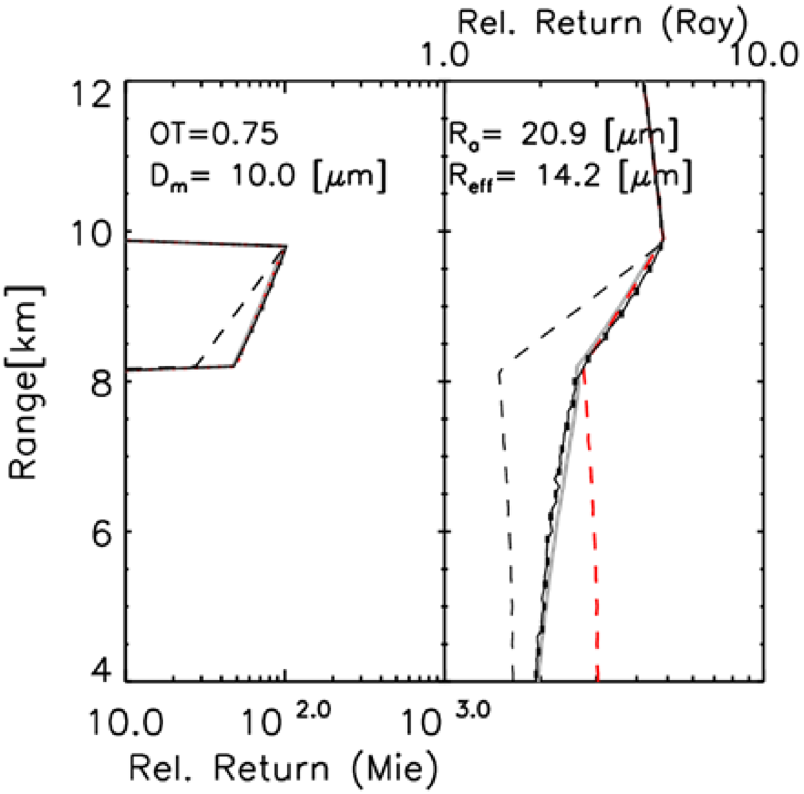


Figure 4: Mie co-polar and (Right) Rayleigh channel co-polar returns for an ice cloud of an optical thickness of 0.75 and an effective radius of 14.2 microns. Black solid: ECSIM MC results. Dashed Black: Single scattering results. Solid Grey: Hogan's model results and Red-Dashed the results of Platt's method with a MS factor of 0.45.

For both cases where MS may have to corrected for it will be necessary to assume an effective size of the aerosol and cloud particles. For the cirrus particle effective radii a parameterization based on temperature will likely be sufficient. For the large dust aerosols the effective size will be a function of the aerosol type produced by the A-TC procedure. Multiple scattering effects will also be discussed in more detail within the A-EBD algorithm description.

3.1.3 A-EBD

The Rayleigh signal from a HSRL lidar can be used to estimate the extinction profile in a rather direct manner [Shipley, 1983] by estimating the derivative of the range-corrected logarithmic signal. However, the applicability of this method is limited due to a high required SNR ratio. Thus, one can assert that accurate but low-precision extinction information is, in general, provided by the Rayleigh signal. In contrast, extinction information can also be extracted from the Mie signal channel which, in general, may be viewed as less accurate (since factors such as the extinction-to-backscatter ratio must be specified in order to "invert" the signal [Klett 1985]) but more precise (since the SNR ratio of the input data does not impact the derived extinction product to the same degree.).

Taking into account these two observations, it is advantageous to formulate a retrieval procedure which simultaneously uses both the Rayleigh and Mie signals in order to combine the accurate but less precise Rayleigh channel derived information with the less-accurate but more precise information derived using the Mie channel. The essence of the algorithm described in this ATBD, is to perform a Klett-like retrieval using an S profiles which yields an extinction profile which, in turn, enables an optimal reconstruction of the observed Rayleigh channel. It is thought that an optimal-estimation based variational approach is best suited for this purpose. By including the cross-talk and calibration factors into the state-vector the approach is tolerant (to a certain degree) to calibration and cross-talk correction errors and estimates of the relevant parameters are an auxiliary output of the procedure. . Multiple-scattering (MS) effects are taken into account using a novel extension of Platt's approach.

In this section, we give an overview of the general background theory and major techniques that the algorithm relies on. In Section 5, more detailed specific information is given.

The algorithm makes use of the total backscatter signal as well as the Rayleigh signal profile. It should be noted that the total signal is not directly measured by ATLID. However, as discussed earlier in Section 3.1.1, the total signal can be reconstructed by appropriately combining the separate Mie co-polar, Rayleigh co-polar and Total cross-polar channels. This issue is addressed in more detail later in Section 3.1.4.5.

3.1.3.1 Inversion of the two-component lidar equation with variable extinction-to-backscatter values

For simplicity, as a starting point for the presentation of the background theory we first consider the case where we assume that the ratio between the single-scattering and multiple-scattering return is known a-priori. Later we will present the treatment of multiple-scattering in detail.

The total (calibrated) backscatter signal measured by ATLID can be written as

$$b_t(z) = \left(\beta_M(z) + \beta_R(z)\right) M^{-1}(z) \exp\left[-2\int_{z_{lid}}^{z} (\alpha_M(z') + \alpha_R(z'))dr'\right] \tag{2.2.9}$$

where z is altitude (above MSL), $r(z)$ is the range from the lidar, z_{lid} is the lidar altitude, β denotes backscatter, α denotes extinction, and the subscripts M and R are used to distinguish between Mie (aerosol+cloud) and Rayleigh scattering extinction and backscatter respectively. M is an effective multiple scattering correction factor that must be calculated with the aid of a model.

The Rayleigh terms in Eq .(2.2.9) are simply related to the atmospheric density profile which is assumed to be known. However, the presence of two unknowns ($\alpha_M(z)$, $\beta_M(z)$) and only one measured quantity ($b_t(z)$) means that Eq. (2.2.9) cannot be directly solved (inverted).

Without loss of generality, we can write:

$$\alpha_M(z) = S(z)\beta_M(z) \tag{2.2.10}$$

where $S(z)$ is the so-called extinction-to-backscatter ratio and depends on the characteristics of the scatterers being probed. Eq .(2.2.9). If we use Eq. (2.2.10), multiply through by $M(z)$ and the inverse two-way Rayleigh attenuation we can write

$$\begin{aligned} \mathrm{B}_t(z) &= b_t(z)M(z)\exp\left[2\int_{z_{lid}}^{z} \alpha_R(z')dr'\right] \\ &= \frac{1}{S(z)}\left(\alpha_M + S(z)\beta_R(z)\right)\exp\left[-2\int_{z_{lid}}^{z} \alpha_M(z')dr'\right] \end{aligned} \tag{2.2.11}$$

We further note that if we define

$$\mathrm{B}'_t(z) = exp\left[-2\int_{z_{lid}}^{z} S(z')\beta_R(z')dr'\right]\mathrm{B}_t(z)S(z). \tag{2.2.12}$$

Then, using Eqs.(2.2.12) and (2.2.11) it can be seen that

$$\mathrm{B}'_t(z) = \alpha'(z)\exp\left[-2\int_{z_{lid}}^{z} \alpha'(z')dr'\right] \tag{2.2.13}$$

where

$$\alpha'(z) = \left(\alpha_M(z) + S(z)\beta_R(z)\right) \tag{2.2.14}$$

At this point, the two-component lidar equation with variable S values has now been re-cast into a more manageable form. Eq. (2.2.13) now has the same form as the well-known simple single-scatter lidar equation in a single component atmosphere. This equation is in fact a differential equation whose solution can be written as

$$\tau'(z_1, z_2) = -0.5\ln\left[1 - 2\int_{z_1}^{z_2} \mathrm{B}'_t(z')dr'\right] \tag{2.2.15}$$

where $\tau'(z_1, z_2) = \int_{z_1}^{z_2} \alpha'(z')dr'$. Equivalently, in terms of α' the solution can be written as

$$\alpha'(z) = \frac{\mathrm{B}'_t(z)}{\dfrac{\mathrm{B}'_t(z_m)}{\alpha'(z_m)} - 2\int_{z_m}^{z} \mathrm{B}'_t(z')dr'} \tag{2.2.16}$$

where z_m is some boundary range where $\alpha'(z_m)$ is prescribed. Using Eq. (2.2.13) the boundary value term can be expressed as

$$\frac{\mathrm{B}'_t(z_m)}{\alpha'(z_m)} = \exp\left[-2\int_{z_m}^{z_{lid}} \alpha'(z')dr'\right]. \tag{2.2.17}$$

If z_m is chosen to be an altitude high enough so that no non-negligible amounts of aerosol/cloud exist between z_m and z_{lid} we can write

$$\frac{\mathrm{B}'_t(z_m)}{\alpha'(z_m)} = \exp\left[-2\int_{z_m}^{z_{lid}} S(z')\beta_R(z')dr'\right]. \tag{2.2.18}$$

Thus, if profiles of $M(z)$ and $S(z)$ are specified[3], then Eqns. (2.2.16) and (2.2.18) can then be used to estimate the extinction profile. In the next section (Section 3.1.3.2) we will show how using the Rayleigh signal provides a means whereby suitable profiles of $S(z)$ may be estimated. In Section 3.1.4 the methods used to account for multiple-scattering used in this work are described.

In this work, Eq.(2.2.16) is not used directly (since another formulation is used which accounts for MS effects in an intrinsic manner as described in Section 3.1.4.4), however, certain partial derivatives stemming from Eq.(2.2.16) are useful for error calculations.

3.1.3.2 Use of the Rayleigh Channel

Up to this point we have used only the total backscatter signal. Using the Rayleigh channel gives us a means, in effect, to help constrain the values of $S(z)$ necessary in order to apply Eqn.(2.2.16). In particular, we seek to find the profile of $S(z)$ that allows us (with the additional aid of a suitable multiple-scattering model to estimate $M(z)$) to accurately forward model the Rayleigh signal. In broad terms, the general approach that this retrieval procedure embodies is schematically shown in
Figure 5.

---

[3] Note that for altitudes where we expect $\alpha(z)$ to be small compared to $\beta_R(z)$ (i.e. on the basis of e.g. an input Target Mask) then the corresponding values of $S(z)$ can be set to any non-zero constant (in this work we use 1) and the derivation presented here remains valid.

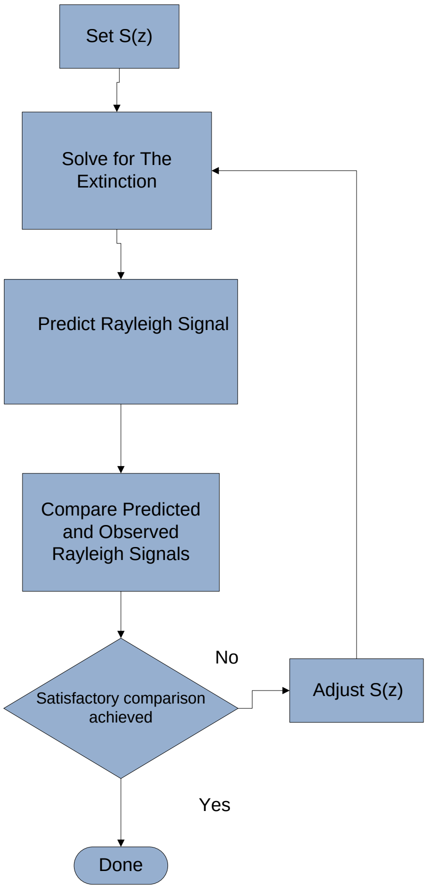


Figure 5: High level simplified schematic of the core approach of the algorithm Note that within this diagram multiple-scattering is not explicitly treated.

The forward model for the (calibrated) Rayleigh signal can be written as

$$\begin{aligned} p_R(z) &= \frac{\left(\beta_R(z)\right)}{r(z)^2}\frac{1}{M_R(z)}\exp\left[-2\int_{z_{lid}}^{z}(\alpha_M(z')+\alpha_R(z'))dr'\right] \\ &= \frac{\left(\beta_R(z)\right)}{r(z)^2}\frac{1}{M_R(z)}\exp\left[-2\left(\tau_R(z)+\tau'(z)-\tau_\beta(z)\right)\right] \end{aligned} \qquad (2.2.19)$$

where, $\tau_R(z)=\int_0^z \alpha_R(z')dz'$ $\tau_\beta(z)=\int_0^z S(z')\beta_R(z')dz'$, M_R is the multiple scattering factor for the Rayleigh signal (which in general may be different from that appropriate for the Mie channel signal). The Rayleigh scattering and extinction profile are known functions of the laser wavelength and the atmospheric density profile while $\alpha(z)$ could be supplied via e.g. Eqn. (2.2.16) together with Eq. (2.2.14). In this work, an alternate novel approach for determining the extinction profile while accounting for multiple-scattering is used and is described in the following section.

3.1.4 *Lidar Multiple Scattering*

In general, for lidar cloud remote sensing multiple scattering effects must be taken into account. This is especially true for space-based observations. Monte-Carlo (MC) simulations are highly accurate. However, in general, they are too computationally expensive to use in an inversion procedure. Here, for completeness, we outline two non-Monte-Carlo approaches that were used during the algorithm development phase and present a novel approach which was ultimately settled upon for inclusion in the EBD algorithm code.

3.1.4.1 Analytical model of Hogan

A fast and reasonably accurate approach is the approach developed by Hogan (2006). Using Hogan's model the multiple scattering factors seen by the telescope $M_t(z)$ and range $M_R(z)$ can be predicted as functions of the extinction profile and a few other parameters. In particular,

$$M_{t,R} = M_H\left(\alpha_M(z); R_a(z); \alpha_R(z); \rho_l; \rho_t; r(z)\right) \tag{2.2.20}$$

where $M_H(z)$ is used to denote the particular values of $M_t(z)$ or $M_R(z)$ retuned by Hogan's model which is a are functions of
ρ_l i(the laser e^{-1} width), ρ_t (the telescope field-of-view) and R_a (the "equivalent area" radius which for a collection of particles each with characteristic size D is defined such that

$$\pi R_a^{\ 2} = \alpha^{-1}\int_0^\infty N(D)\sigma(D)Ac(D)dD \tag{2.2.21}$$

where $N(D)$ is the particle size distribution). In Eq.(2.2.21) σ is the single particle extinction cross-section and Ac is the particle cross-sectional area. For further details of Hogan's model the reader is referred to Hogan [2006

Hogan's model can be used to predict $M(z)$ in Eq.(2.2.11) so that $b_t(z)$ can be defined. Then, with an assumed $S(z)$ profile and boundary value Eqn. (2.2.16) can be solved to yield the extinction profile. However, $M_t(z)$ is itself a function of the extinction profile. Thus, the solution of the system must proceed iteratively. $M_R(z)$, which is related to $M_t(z)$ by a multiplicative, constant is used in Eq.(2.2.19) to account for the effect of MS on the Rayleigh channel.

A sample comparison between Hogan's model and exact MC results is shown in Figure 6. Here, by comparing the total return and the single-scatter only return that MS effects are significant. Further, it can be seen that, the influence of particle size within the cloud is limited. However, the Rayleigh scattering "tail" under the cloud is sensitive to the cloud particles sizes.

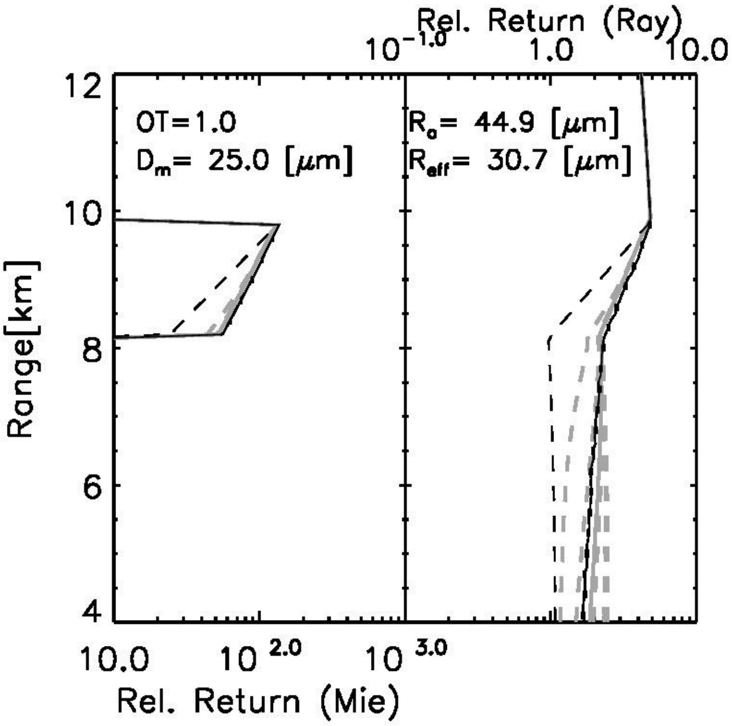


Figure 6. Sample comparison between the ECSIM lidar Monte-Carlo multiple scattering model and the analytical model due to Hogan (2006). (Left) Mie co-polar and (Right Rayleigh channel co-polar returns for an ice cloud of OT 0. and an effective radius of 30.7 microns. Black solid: ECSIM results. Dashed Black: Single scattering results. Solid Grey: Hogan's model results for the true value of Ra. Dashed Grey from left to right: Hogan's model results for Ra=10, 25.0, 50, 100 and 200 microns respectively.

Hogan's approach is orders of magnitude faster than MC calculations. However, it is still much slower than the corresponding single-scattering case. Moreover, we also require an efficient means to initialize $M(z)$ when solving Eq. (2.2.16). We first discuss Platt's effective extinction approach and discuss how this can be extended to handle the phenomenon of decaying tails.

3.1.4.2 Platt's approach

Within Figure 6, it can also be noted that, within the cloud, that the observed signal closely resembles a less attenuated version of the single-scattering signal. This is to be expected when the particles are large compared to the wavelength of the laser light so that half the scattered energy is scattered forward in a narrow diffraction lobe and largely stays within the lidar receiver file-of-view. This result was noted by Platt (1973) and forms the basis of a simple method for accounting for Multiple-scattering effects.

If we define

$$M_p(z) = \exp\left[-2\int_0^z \left(1-\eta(z)\right)\alpha_M(z)dr'\right] \qquad (2.2.22)$$

where η is the Platt coefficient which physically describes the fraction of scattered energy that remains within the lidar filed-of-view (and thus behaves like it has not be scattered). Approximating $M_t(z)$ by $M_p(z)$ in Eq.(2.2.9) gives

$$p_t(z) = \frac{\left(\beta_M(z) + \beta_R(z) + \beta_\perp(z)\right)}{r(z)^2} \exp\left[-2\int_{z_{lid}}^{z} (\eta(z')\alpha_M(z') + \alpha_R(z'))dr'\right] \qquad (2.2.23)$$

Compared to the approach of Hogan, Platt's approach is faster and simpler but is limited. In Figure 7, a comparison between a MC calculation, the results of Hogan's model and Eq.(2.2.23) are shown. Here it can be seen that, within the cloud Platt's approach performs well. However, under the cloud Platt's approach cannot capture the decaying structure of the "tail".

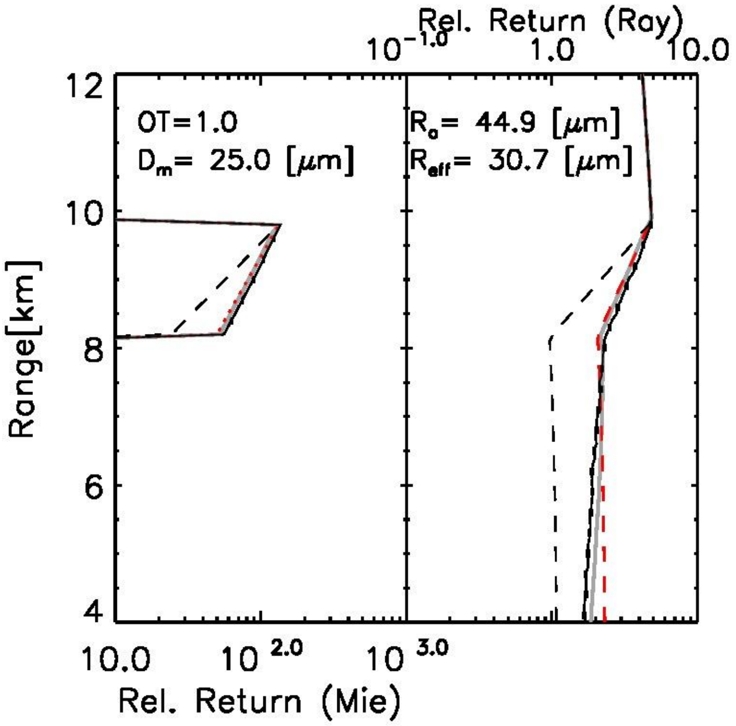


Figure 7: As Figure 6, however the results for Hogan's model are shown only for Ra=25.0 microns and the Red lines shown the result of predicting M using Eq. (2.2.22) with η =0.55.

3.1.4.3 The origin of MS tails

In previous work it was found that even though the model of Hogan is many orders of magnitude faster than MC approaches it is still slow when compared to Platt's effective extinction approach. On the other hand, as was demonstrated in the previous sections, Platt's effective extinction approach cannot describe the decay of the return towards single-scatter values below layers. Hence an alternative approach for accounting for MS effects was formulated which, in terms of complexity, lies between the approaches of Platt and Hogan. We begin by considering an heuristic description of the effect of multiple scattering as a pulse passes through a scattering medium and beyond.

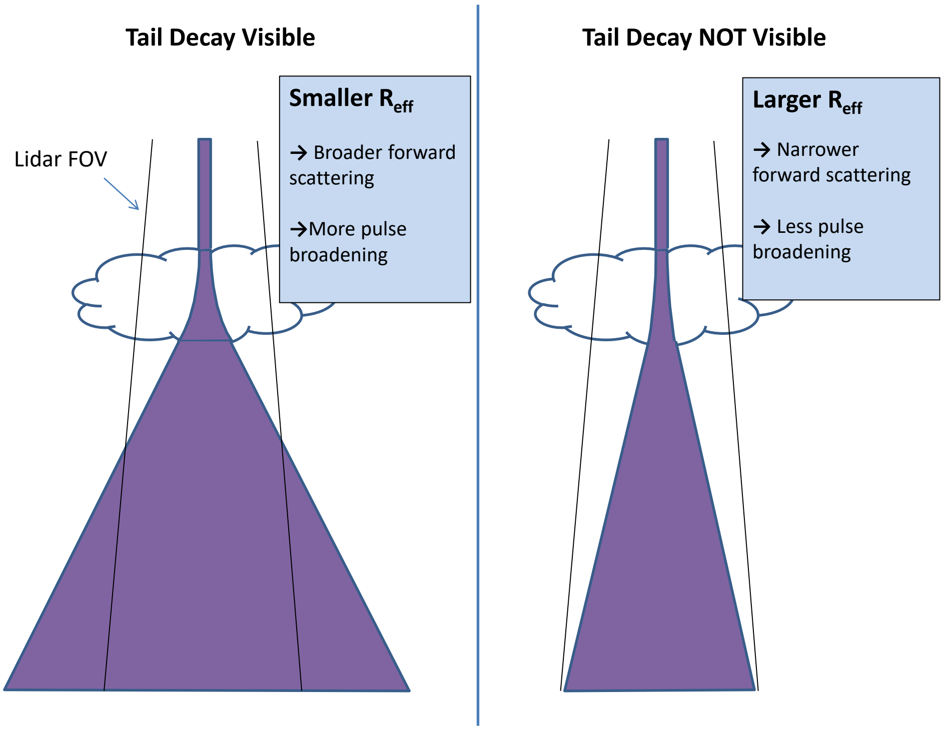


Figure 8: Schematic depiction of the mechanisms behind the occurrence of decaying tails below scattering layers. Here the Purple regions represent the broadened laser pulse extent.

Referring to Figure 8, within the cloud the low mean-free-path of the photons ensures that the multiply-scattered light that contributes to the detected signal tends to be confined to within the field-of-view of the lidar. However, the angular variance of the lidar beam will be broadened as it propagates downwards through the cloud with more and more photons undergoing scattering events.

At cloud base the lidar beam emerges with an effective angular divergence which increases with the optical thickness of the cloud and decreases with the size of the cloud particles. This is due to that fact that the angular-width of the cloud phase function forward lobe increases with decreasing particle size i.e.

$$\theta_{sc} \simeq \left(\frac{\lambda}{\pi R_a} \right) \tag{2.2.24}$$

Below cloud base the lidar beam will continue to propagate with a given divergence. However, the horizontal spread of the photons is no longer constrained by the presence of the cloud. As the beam continues to propagate downwards, depending of the lidar receiver footprint more and more of the multiple-scattered photons will travel outside of the receiver cone.

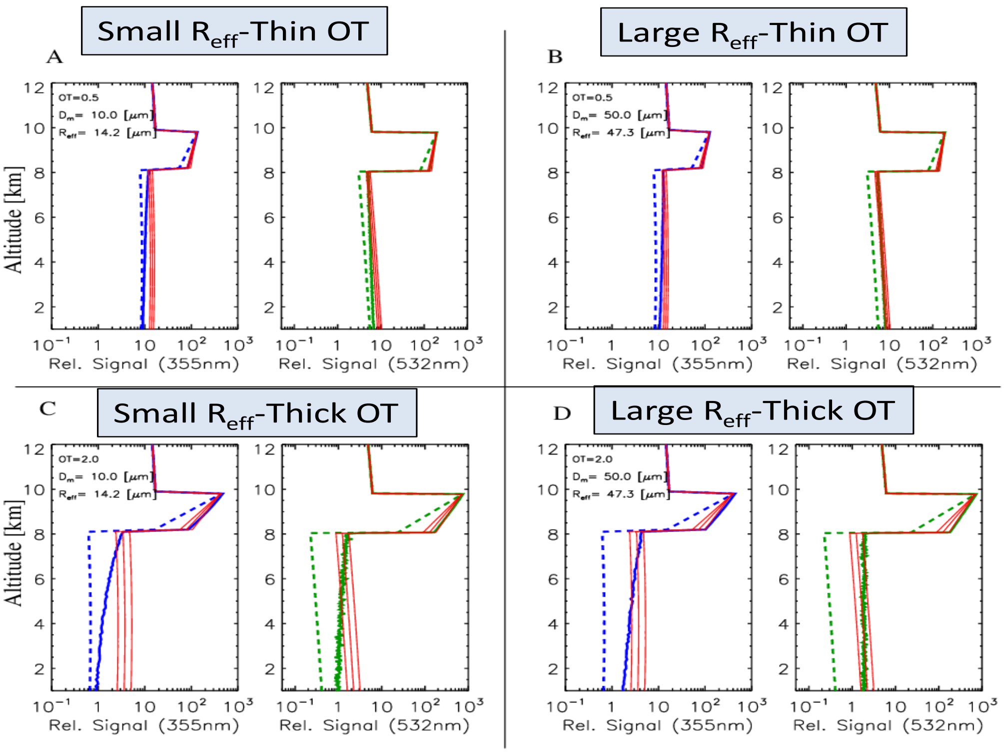


Figure 9: Monte-Carlo simulations of the total return for ATLID and CALIPSO for different idealized layers. The dashed lines correspond to the single-scattering return and the Red-lines correspond to the results of Platt's approach with three different MS coefficient values (0.4,0.5,0.6).

It is instructive to compare the expected effects of multiple-scattering on the below layer signal for ATLID with the case of CALIPSO. Since the effective footprint of CALIPSO is greater than that of ATLID the resulting "decay" below cloud base should be much more noticeable for ATLID than CALIPSO. In fact for CALIPSO, for the ice cloud cases shown the "decay" length for CALIPSO is so long that it is only notable for relatively thick cirrus with relatively small particle sizes (as illustrated in Figure 9, esp. the lower Left panel).

Even though the below cloud tails in CALIPSO observations may are much less apparent than those expected for ATLID, such cases can be readily found in the CALIPSO data record. A, by no means atypical, example of high tropical cirrus as measured by CALIPSO is shown in Figure 10. Here the Left-Highlighted cirrus cloud has a temperature in the range of -50C to -70C and thus is expected to be comprised of crystals with effective radii on the regions of 10-20 microns [Donovan et al., 2003]. Further by comparing the signal levels above and below the cloud it appears that the cloud is similar in optical thickness to the example shown in the Bottom-Left panel of Figure 9 . Thus, if our calculations are realistic, then a decay of the signal should be evident below the cloud base. In fact such a decay is indeed visible (see the Left-Hand Panels) and (qualitatively) well-matches our calculations in terms of magnitude and rate of decay. For comparison, average signals taken from a nearby region (Right-Highlighted area in Figure 10) with much thinner upper level cirrus show no such decay. The CALIOP 532 nm detectors are known to have a 'non-ideal transient response', and after measuring a large signal they do not immediately recover, but instead an shown an exponential decay [McGill et al., 2007; Hunt et al., 2009]. The 1/e decay-constant associated with this non-ideal response (on the order of 0.4-0.5 km or about 3.0 msec see e.g. Figure. 18 of Hunt et al., 2009), however, is very fast compared with the decay time seen here. Thus, we feel that in this case the observed signal decay below the cirrus is primarily due to MS effects.

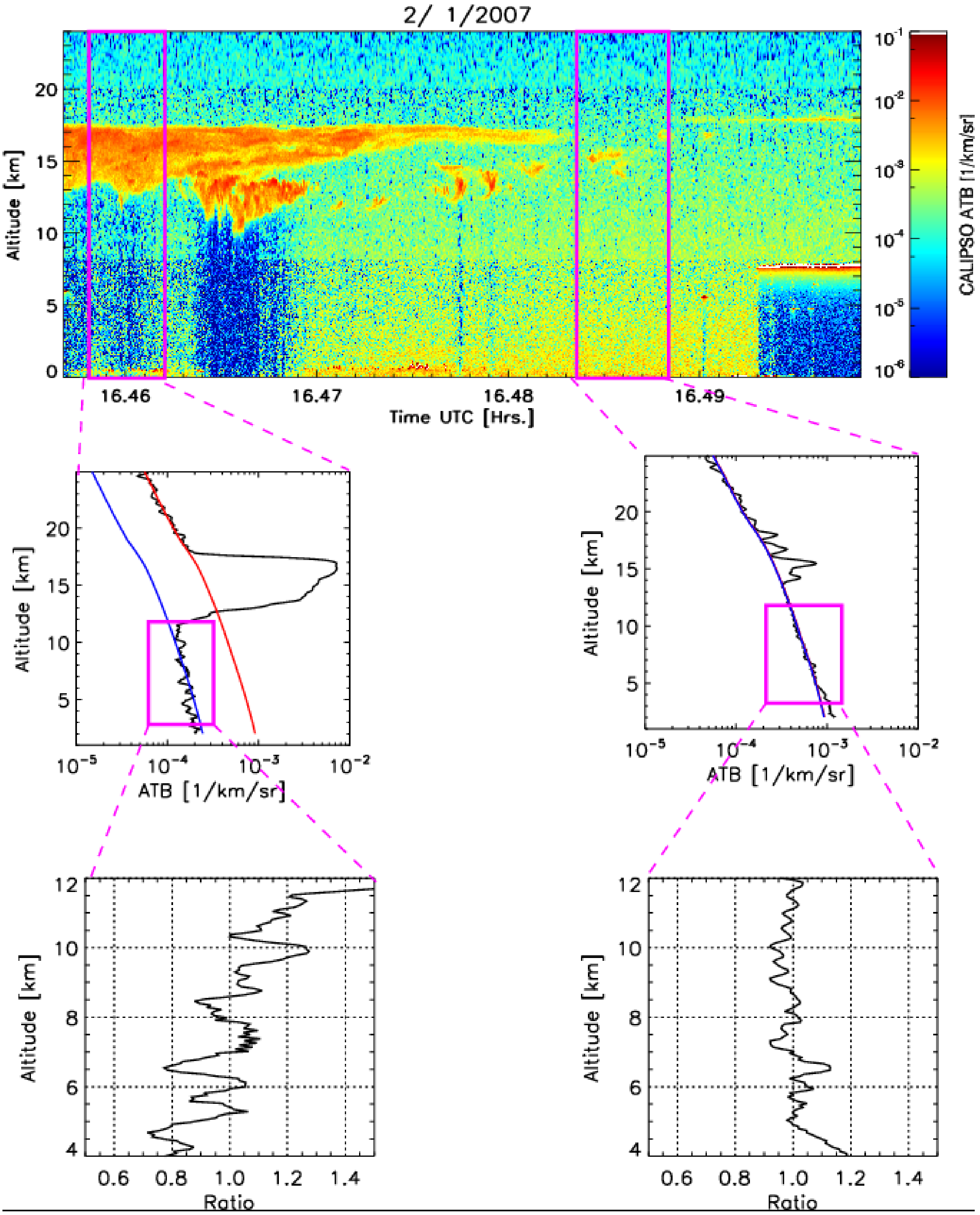


Figure 10: (Top-Panel) CALIPSO attenuated backscatter from UTC 16.48 to 16.5 Hrs UTC on Jan 02 2007 (Lat/Long from -9.37,137.7 Deg to -10.29,137.54 Deg). The middle panels show average profiles corresponding to the boxed regions shown in the top-Panel along with theoretical cloud/aerosol free attenuated backscatter profiles normalized to the observed average signals both below (Blue) and above (Red) the cloud layer (note in the Middle-Right Panel) the red and Blue curves overlap on this scale). The Bottom Panels show the respective ratios of the below-cloud observed attenuated backscatter profiles to the expected cloud/aerosol free signal.

3.1.4.4 An Extension to Platt's approach.

In the previous section it has been explained that while the forward model approach of Hogan is accurate and fast compared to Monte-Carlo methods, it is slow compared to method like Platt's effective extinction method. Moreover, using Hogan's model, which cannot be analytically inverted (unlike Platt' modified lidar equation), means that within our framework MS effects would have to be accounted for in an iterative fashion. This means the MS forward model would have to be called many time (experience shows between 5-10 times) per optimization step. This is undesirable. However, it also has been demonstrated that Platt's approach is inadequate for ATLID since it cannot cope with the existent of MS tails. Here building upon the ideas presented in the previous sections, we present an extension of Platt's approach which is suitable for our needs.

As a first step, it is important to realize that the tails are fundamentally a consequence of the return signal decaying towards single-scatter values. Accordingly, we can write

$$b_M = \beta_M(z)e^{-2\tau(z)}\left[(1-f(z)) + f(z)e^{2\eta\tau(z)}\right], \tag{2.2.25}$$

where η is the Platt multiple-scattering extinction factor (here for simplicity assumed to be constant per layer) and $f(z)$ is the return signal multiple scattering fraction. The above equation is particular to the Mie channel, however, fundamentally similar equations can also be written for the Rayleigh and cross-polar channel. When $f(z)=1$, Eq. (2.2.25) effectively reduces to Eq. (2.2.23) and when $f(z)=0$: it reduces to the normal single-scattering lidar equation. What is required is a method to determine the profile of $f(z)$.

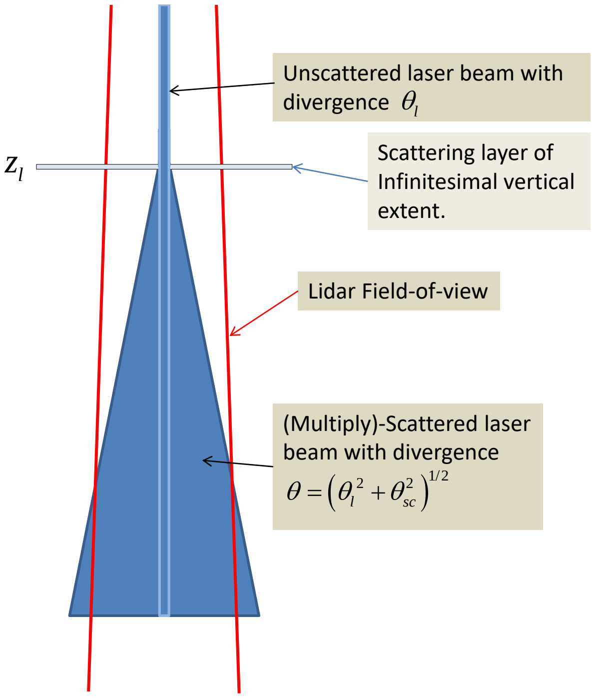


Figure 11: Schematic depiction of the angular broadening experienced by a lidar pulse as it interacts with a physically (but not optically) thin scattering layer at altitude z_l .

We start by considering the case of a physically thin scattering layer. If we assume that the beam has a Gaussian profile and model the forward-scattering lobe of the effective layer phase function by a Gaussian [Eloranta 1998], then the divergence of the forward scattered light will also be Gaussian with a divergence given by the convolution of the incoming beam divergence (θ_l) with the effective scattering forward-lobe width (θ_{sc}) so that the effective width of the multiply scattering radiation emerging from the layer bottom is given by

$$\theta_{eff} = \left(\theta_{sc}^2 + \theta_l^2\right)^{1/2} \qquad (2.2.26).$$

The fraction of multiply scattered lidar that then remains within the lidar fov can easily be calculated and is given by

$$f(z, z_l) = 1.0 - \exp\left[-\left(\frac{\rho_t^{\,2}(z - z_{Sat})^2}{\theta_{sc}^{\,2}(z - z_l)^2 + \rho_l^{\,2}(z - z_{Sat})^2}\right)\right] \qquad (2.2.27)$$

where ρ_t is the receiver telescope fov, ρ_l is the laser divergence, z_{sat} is the satellite altitude and z_l is the altitude of the scattering layer. This expression is only valid for a single thin scattering layer and so, by itself, is not so useful. However, we can generalize this expression to the case of a general profile in a heuristic approximate fashion. A rigorous calculation of $f(z)$ would be involved and would result in a similar formalism as the QSA model of Hogan. Here we will use the information present in the signal itself to calculate the effective $f(z)$ under general conditions. Since the observed signal itself contains information on the location and relative strength of the scattering at each level, we postulate the form

$$f_e(z) = \frac{\int_{z_{sat}}^{z} f(z, z_l) b_M(z_l) dr_l}{\int_{z_{sat}}^{z} b_M(z_l) dr_l} \qquad (2.2.28)$$

where $\mathrm{B}_{at,Mie}$ is the Mie channel attenuated backscatter. That is, we use the observed Mie channel backscatter itself as a weighting factor to determine the effective $f(z)$ profile. In the limit of a single thin scattering layer this expression yields the correct result.

An example comparison between exact Monte-Carlo calculations, Platt's approach and the "Platt+Tails" approach (Eqn. (2.2.28) together with Eqn. (2.2.25)) is shown in Figure 12. Here a fitting procedure was used to find the best values of η and θ_{sc} for each of the two layers. It can be seen that the extended Platt approach provides a very good match to the MC results for the entire profile while the normal Platt approach is deficient.

As a further refinement, in order to explicitly account for the fact that the effective backscatter coefficient for particulate backscattering (which may be strongly peaked around 180 Deg) may, in general, be lower than the effective backscatter coefficient associated with molecular Rayleigh scattering (which is smooth in the backscattering direction) an additional factor is added which acts to adjust Eq.(2.2.28) based on the relative amount of particulate scattering i.e.

$$f_e(z) \rightarrow f_e(z)\left[1 - \left(R - 1\right) + (R - 1) f_{MSp}\right] \qquad (2.2.28)$$

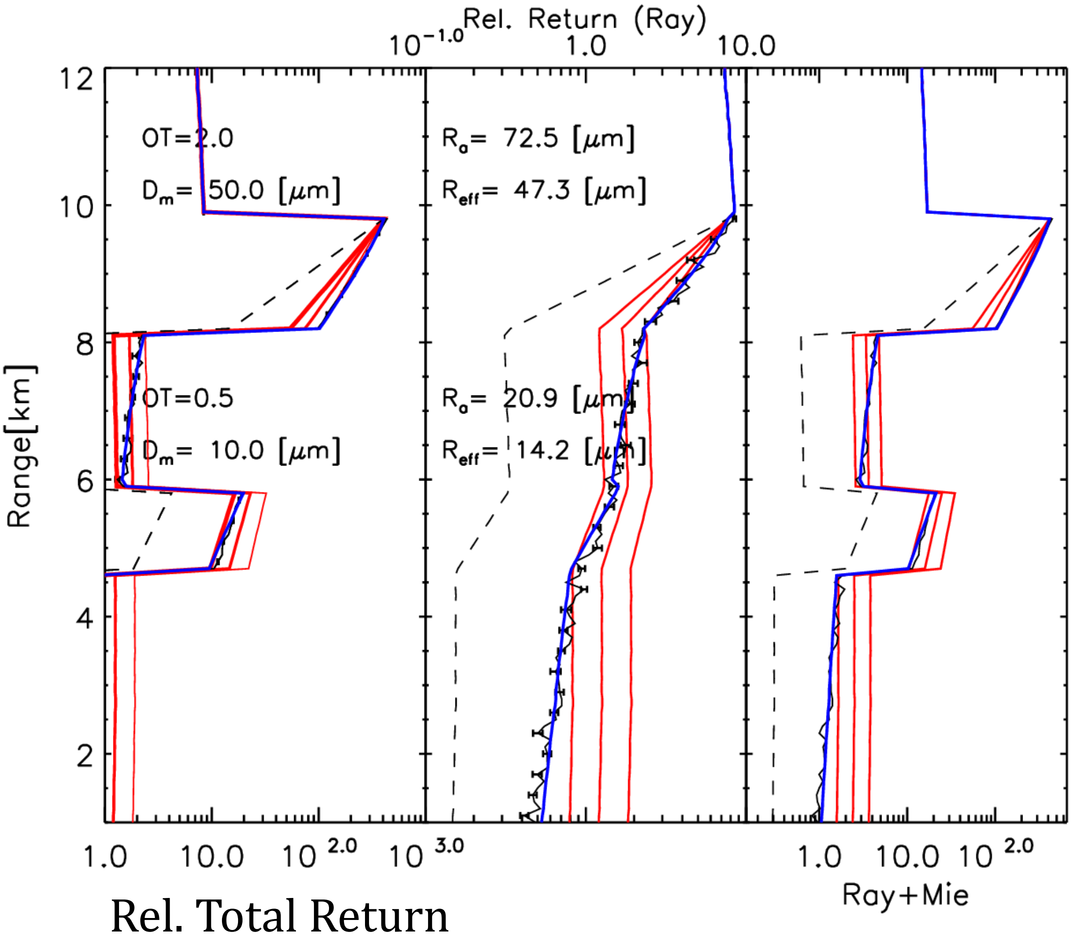


Figure 12: From Left to Right: Total simulated total attenuated backscatter, Rayleigh attenuated backscatter profile and the total simulated attenuated backscatter profile (difference scale than the first columns) corresponding to a two-layer ice cloud system with the given layer parameters. The Black lines are the product of lidar Monte-Carlo radiative transfer calculations. The Dotted back lines are the single-scatter return values, the Red lines are the results of Platt's approach using tree different values of η (0.4,0.5,0.6). The Blue-lines show the result of applying Eqs. approach (Eqn. (2.2.28) together with Eqn. (2.2.25)) with optimally chosen of values of η and θ_{sc} for each of the two layers.

3.1.4.4.1 Inversion of the Platt+Tails Lidar equation

An important feature of the Platt equation is that it can be inverted in a similar fashion as the normal single scattering lidar equation as described earlier in Section 3.1.3.1 . All that is required is to replace the extinction by the effective extinction (i.e. $(1-\eta(z))\alpha(z)$). The Platt+Tails equation can also be inverted. However, a different approach is required. We start by considering a form of Eq.(2.2.25) where the molecular and particulate extinction and backscatter contributions are made explicit, i.e.

$$b_t(z) = C_{lid}\left(\beta_{M,\pi}(z)+\beta_{R,\pi}(z)\right)e^{-2\int_{z_{sat}}^{z}(\alpha_R(z')+\alpha_M(z'))dr'}\left[\left(1-f(z)\right)+f(z)e^{2\int_{z_{sat}}^{z}\eta(z)\alpha_M(z')dr'}\right]. \quad (2.2.29)$$

Then by correcting for the Rayleigh attenuation and making use of the fact that $\alpha_M(z)=S(z)\beta_{\pi,M}(z)$ the above expression can be re-arranged to yield:

$$B'_t(z)=\alpha'(z)\left[\left(1-f(z)\right)e^{-2\tau'(z)}+f(z)T_{Rs\eta}e^{-2\tau'_\eta(z)}\right], \quad (2.2.30)$$

where :

$$\alpha'(z)=\left(\alpha_M(z)+\beta_{\pi,R}(z)S(z)\right) \quad (2.2.31)$$

$$\tau'(z)=\int_{z_{sat}}^{z}\alpha'(z')dr' \quad (2.2.32)$$

$$\tau'_\eta(z)=\int_{z_{sat}}^{z}\left(1-\eta(z)\right)\alpha'(z')dr' \quad (2.2.33)$$

$$B'_t(z)=C_{lid}^{-1}b_t(z)S(z)e^{-2\int_{z_{sat}}^{z}\beta_{\pi,R}(z')S(z')dr'}\;e^{+2\int_{z_{sat}}^{z}\alpha_R(z')dr'} \quad (2.2.34)$$

Noting that:

$$d\tau'(z)=\alpha'(z) \quad (2.2.35)$$

and

$$d\tau'_\eta(z)=\left(1-\eta(z)\right)\alpha'(z)=\left(1-\eta(z)\right)d\tau'(z) \quad (2.2.36)$$

Eqn.(2.2.30) can be written as

$$B'_t(z)=d\tau'\left(1-f(z)\right)e^{-2\tau'(z)}+d\tau'_\eta\frac{T_{Rs\eta}f(z)}{(1-\eta(z))}e^{-2\tau'_\eta(z)}. \quad (2.2.37)$$

Except for the cases when either $f(z)=1$ or $f(z)=0$, then the above equation does not have a Klett-type analytical solution. However, a useful discrete gate-by-gate solution form can be obtained. If we assume that f and η are constant across a measurement grid-cell of width Δz then

$$\begin{aligned}
B'_{t,i}\,\Delta z_i &= \frac{1}{2}\left(1-f_i\right)e^{-2\tau'(z)}\Big|_{r_i+\frac{\Delta r}{2}}^{r_i-\frac{\Delta r}{2}} + \frac{1}{2}\frac{T_{Rs\eta,i}\,f_i}{(1-\eta_i)}e^{-2\tau'_\eta(z)}\Big|_{r_i+\frac{\Delta r}{2}}^{r_i-\frac{\Delta r}{2}} \\
&= \frac{1}{2}\left(1-f_i\right)\left[e^{-2\tau'_{i-1}}-e^{-2\tau'_i}\right]+\frac{1}{2}\frac{T_{Rs\eta,i}\,f_i}{(1-\eta_i)}\left[e^{-2\tau'_{\eta,i-1}}-e^{-2\tau'_{\eta,i}}\right] \qquad (2.2.38) \\
&= \frac{1}{2}\left(1-f_i\right)e^{-2\tau'_{i-1}}\left[1-e^{-2\alpha'_i\Delta r_i}\right]+\frac{1}{2}\frac{T_{Rs\eta,i}\,f_i}{(1-\eta_i)}e^{-2\tau'_{\eta,i-1}}\left[1-e^{-2(1-\eta_i)\alpha'_i\Delta r}\right]
\end{aligned}$$

Where, here, i refers the altitude range-gate and $B'_{t,i}$ is the discrete form of Eq. (2.2.34) constructed using the observed attenuated total backscatter. The above expression cannot be analytically solved for α_i. However, α_i can be efficiently numerically solved for by (e.g. using Brent's method) finding the real root of

$$0 = 2B'_{t,i}\,\Delta z_i + \left(1-f_i\right)e^{-2\tau'_{i-1}}\left[e^{-2\alpha'_i\Delta r_i}-1\right]+\frac{T_{Rs\eta,i}\,f_i}{(1-\eta_i)}e^{-2\tau'_{\eta,i-1}}\left[e^{-2(1-\eta_i)\alpha'_i\Delta r_i}-1\right] \qquad (2.2.39)$$

on a gate-by-gate basis.

3.1.4.5 Construction of the Total Signals

The extinction algorithm makes use of the total (Rayleigh+Mie both co-and cross-polar) signal and the Rayleigh (co-+cross-polar) signals. However, these signal s are not directly available as an L1 inputs. Instead we must construct them from the available input L1 signal products. Eqns. (2.2.3)-(2.2.3) form the basis of this procedure.

Further, in this work we allow for a degree of uncertainty in the applied Mie an Rayleigh bulk calibration coefficients by introducing multiplicative adjustment factors (f_{C_M}, f_{C_R}) for the cross-talk and calibration coefficients such that the *adjusted total attenuated Rayleigh extinction corrected backscatter* is given by

$$\mathrm{B}_t^f(z) = \left(\frac{1}{f_{C_R}}\right)\left(\mathrm{B}_M + \mathrm{B}_\perp\right) + \left(\frac{1}{f_{C_M}}\right)\left(\mathrm{B}_R\right) \qquad (2.2.40)$$

Note here that the adjustment factors are not-range dependent and we have assumed a fixed cross relative polarization calibration (i.e. $C_M / C_\perp$)

Similarly the equation for the *adjusted* Rayleigh channel calibrated attenuated backscatter can be written as

$$\mathrm{B}_R^f(z) = \left(\frac{1}{f_{C_R}}\right)\mathrm{B}_R(z) \qquad (2.2.41)$$

Note: In future work robustness against cross-talk correction errors could be built in by including explicit cross-talk adjustment factors into the state-vector and taking these into account during the construction of the total and Rayleigh signal profiles.

3.1.4.6 Optimal Estimation procedure.

Here we will discuss the means by which we will achieve our goal of solving Eq, (2.2.39)such that the Rayleigh signal predicted by Eq.(2.2.19) matches the observed Rayleigh signal (i.e. solving the system depicted in Figure 5 except using Eq. (2.2.39) instead of Eq.(2.2.16)).

In this work we employ the principle of Optimal Estimation [Rodgers 2000]. In general terms, we formulate a cost-function that characterizes the likelihood of the measurements being what they were given a particular parameter configuration of an appropriate forward model combined with our expectations concerning the forward-model parameters. The desired solution is then obtained by minimizing this function.

Our cost function can be written as

$$\begin{gathered}\chi^2 = [\mathbf{y} - \mathbf{F(x)}]^{\mathbf{T}} \mathbf{S}_{\mathbf{e}}^{\mathbf{-1}} [\mathbf{y} - \mathbf{F(x)}] + \\ [10^{\mathbf{x}^{\mathbf{f}}} - \mathbf{I}]^{\mathbf{T}} \mathbf{S}_{\mathbf{a}}^{\mathbf{f}\,-1} [10^{\mathbf{x}^{\mathbf{f}}} - \mathbf{I}]\end{gathered} \tag{2.2.42}$$

1. $\mathbf{y}_{i_g}$ is the *observation vector* (related to the actually observed signal via Eq.(2.2.41) and transformed to Rayleigh corrected attenuated backscatters)

$$\mathbf{y} = (\mathrm{B}_{R,1}, \mathrm{B}_{R,2} \ldots \mathrm{B}_{R,nz})^T \tag{2.2.43}$$

2. $\mathbf{x}$ is the *state-vector* and $\mathbf{x}^{\mathbf{f}}$ is *the log of the normalized state-vector*. Here defined such that

$$\log_{10}(\mathbf{x}^{\mathbf{f}}) = (f_{C_M}, f_{C_R}, f_{f_{MSp}}, f_{\eta,1}, f_{\theta_{sc},1}, f_{S,n_l} \ldots\ldots\ldots f_{\eta,n_l}, f_{\theta_{sc},n_l}, f_{S,n_l})^T \tag{2.2.44}$$

where n_l is the number of layers used in the retrieval. In this work we have chosen to cast the problem in terms of the log (to insure positivity) of the normalized variables (the f_{R_a}, f_η, f_S 's) such that

$$\mathbf{x} = 10^{\mathbf{x}^{\mathbf{f}}} \mathbf{x}_{\mathbf{a}} \tag{2.2.45}$$

(here multiplication between two column or two row vectors is assumed to be element-wise) where $\mathbf{x}_{\mathbf{a}}$ is the *a priori* matrix

$$\mathbf{x}_{\mathbf{a}} = ({C_M}^a, {C_R}^a, f_{MSp}^a, {\eta^a}_1, {\theta^a}_{sc,1}, {S^a}_1 \ldots\ldots\ldots {\eta^a}_{n_l}, {\theta^a}_{sc,nl}, {S^a}_{n_l})^T \tag{2.2.46}$$

where, for simplicity, we have written, the expressions for the state and a priori vectors corresponding to the case where the cross-talk associated elements are not range-dependent. f_{MSp} is the multiple-scattering backscattering peak factor.

3. $\mathbf{F(x)}$ is the forward model (Eq.(2.2.19)).

4. $\mathbf{S}_{\mathbf{e}}$ is the observation error covariance matrix. Using Eqn.(2.2.41) and the fact that the observed signals are

uncorrelated gives

$$
\begin{aligned}
[S_e]_{i,j} &= \left(\frac{\delta b_R}{b_R}\right)^2 \mathrm{B}_R \quad i = j \\
&= 0 \quad\quad\quad\quad\quad\; i \neq j
\end{aligned}
\tag{2.2.47}
$$

5. $\mathbf{x}_a$ is the a priori state vector corresponding to the normalized variables so that, by definition, $\mathbf{x}_\mathbf{a}^\mathbf{f} = \log_{10}(1,1,1,\ldots1) = (0,0,0\ldots.0)$

6. $\mathbf{S}_\mathbf{a}^\mathbf{f}$ is the a priori error covariance matrix corresponding to the log of the normalized variables.

It may be noted that the form of the cost function related to the distance from the a priori is somewhat non-standard. This has to do with the choice of working in terms of $\mathbf{x}^\mathbf{f}$ instead of using $\mathbf{x}$ .Using $\mathbf{x}^\mathbf{f}$ insures that that all the minimization variables are scaled to the same magnitude which aids in the numerical minimization process. The second reason has to do with the nature of the state-variables themselves and their variances. It is more physically reasonable to express the a priori distribution of S and R_a in a multiplicative rather than linear sense. For example, observations show that S for a certain aerosol type is more accurately characterized by stating 20 $\times / \div$ 0.5 (giving a 1-sigma range in this case of 10-40) rather than say 20 ± 20 (giving a 1-sigma range in linear space of 0-40). Our choice of functional form for the a priori components of the cost function reflects this idea. Thus, here we assume that the difference between the true and the a priori values of the components of the state-vector are lognormaly distributed so that e.g.

$$
S^f_{a,i,i} = \log_{10}\left(1 + \left(\frac{\sigma_{x_i}}{x_i}\right)^2\right) \tag{2.2.48}
$$

3.1.4.7 Determination of inversion layers and Specification of a priori values

In order to minimize Eqn. (2.2.42), the a priori values of $\mathbf{S}_\mathbf{a}^\mathbf{F}$ and $\mathbf{x}_a$ must be specified. This procedure is linked to the determination of the inversion layers. Each output resolution volume is classified as being ice cloud, water cloud or aerosol on the basis of the backscatter, extinction and the depolarization ratio. In addition, the temperature and the estimated boundary layer height from a suitable operational analysis may ultimately also be used.

For the volumes identified as being `aerosols' the classification is then refined using the aerosol classification algorithm. The classification procedures are algorithms in their own right and are described in detail in

Once the classification, including aerosol typing, has been achieved then the inversion layers are determined. These layers are determined both on the basis of the Target Mask and the structure of the **A_AER** retrieved scattering ratio, extinction-to-backscatter ratio and the depolarization ratio together. This process is described in more detail in Sections 5.4 and 5.5. At this point in time, the state variables are assumed to be uncorrelated between different inversion layers.

For layers classified as aerosol, the a priori values are generated using the output of the *large-scale aerosol-only backscatter, extinction and depolarization ratio product* (A-AER) together with the output of the aerosol classification algorithm. Corresponding a priori values of θ_{sc} and η are based directly on the aerosol type while in the case of the extinction-to-backscatter ratio they are computed from the quantitative large-scale extinction-backscatter product. In particular, for a given layer extending in altitude between z_l and z_u we have

$$S_a = \frac{<\alpha_{ls}>}{<\beta_{ls}>} = \frac{\sum_{i=l}^{u} \alpha_{ls}(z_i)}{\sum_{i=l}^{u} \beta_{ls}(z_i)} \tag{2.2.49}$$

where the ls subscript is used to denote products generated by the large-scale aerosol-only algorithm and the brackets denote averaging. We assume that the large scale extinction and backscatter products are uncorrelated so that the corresponding variance in S_a is then

$$\sigma_{S_a}^2 = S_a{}^2 \left(\frac{\sigma_{<\alpha_{ls}>}^2}{<\alpha_{ls}>} + \frac{\sigma_{<\beta_{ls}>}^2}{<\beta_{ls}>} \right) \tag{2.2.50}$$

Since the large-scale extinction product is, in general, strongly correlated in the vertical dimension, the vertical correlation must be taken into account. This is done by assuming N effective number of samples equal to the ratio of the layer vertical extent to the window used in the determination of the derivative in Eq.(2.2.6)

$$\sigma_{<\alpha_{ls}>}^2 = \left(\frac{\left(\sum_{i=i_{bot}}^{i_{top}} \sigma_{\alpha,i}^2 \right)}{N_{eff}} \right)^{1/2} \tag{2.2.51}$$

where:

$$N_{eff} = \frac{z_{top} - z_{bot}}{\Delta z_{fit}}$$

Here Δz_{fit} is the length of the fitting window used in the determination of the derivative in Eq.(2.2.6) . The variance of the layer averaged backscatter is obtained in a similar manner.

For ice and water clouds separate a priori S, η and θ_{sc} a priori values and variances are used. These values are set in the algorithm configuration file.

3.1.4.8 Optimization procedure

In this work the minimum of the cost function is obtained numerically using a variant of the *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) approach [Press et al. 1992]. The approach follows that described in Press et al. except for the fact that numerical derivatives are used. This was done largely due to save time during the development process. The algorithm speed seems to be sufficient using numerical derivatives and is more flexible during the development process as a change to the cost-function form does not in turn require a laborious re-calculation of the gradient.

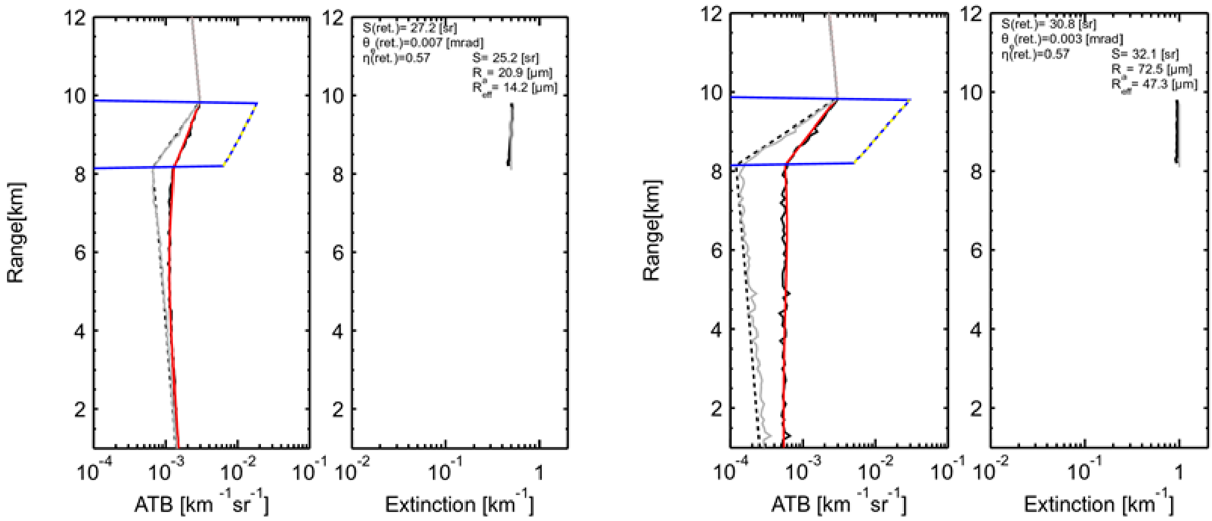


Figure 13: Sample retrieval results applied to idealized ice clouds (Left OT 0.25, Right OT 0.5) with the indicated values of effective radius and equivalent area radius. The Blue lines are the simulated cross-talk corrected Mie signal and the Black-solid line is the corresponding Rayleigh channel attenuated backscatter. The Red-line is the Rayleigh attenuated backscatter forward predicted by the retrieval procedure. The Dotted-Black line is the simulated Single-scatter attenuated backscatter and the Grey-line is the single-scatter only Rayleigh attenuated backscatter forward predicted by the retrieval procedure. The Right panels show the true extinction profiles (Black) and the retrieval results (Grey).

Two example retrieval results are shown in Figure 13. Here simulated observations were generated using the ECSIM lidar Monte-Carlo model . No simulated instrument noise was applied, what noise there is present is Monte-Carlo noise. Here the extinction profiles and the extinction-to-backscatter values are retrieved with good precision. The retrieved values of θ_{sc} can be related to the layer equivalent area radii (R_a) using Eq.(2.2.24). For the case depicted in the Left-panel the retrieved value for θ_{sc} of 0.007 mrads corresponds to an equivalent area radius of 16 microns which corresponds fairly well to the true value of 14 microns. For the Right panel the estimated retrieved value of R_a is 38 microns which also compares well to the true value, which in this case, is 47 microns. Example cross-sections of the cost-function corresponding to the Right-hand Panel of Figure 13 are shown in Figure 14. Here it can be seen that the deepest minimum is reached when η is closest to its true value of 0.57. The long-narrow valley structure around the minimum indicates that errors in the retrieval products will be correlated with the error in S expected to be less than the estimated error in θ_{sc} .

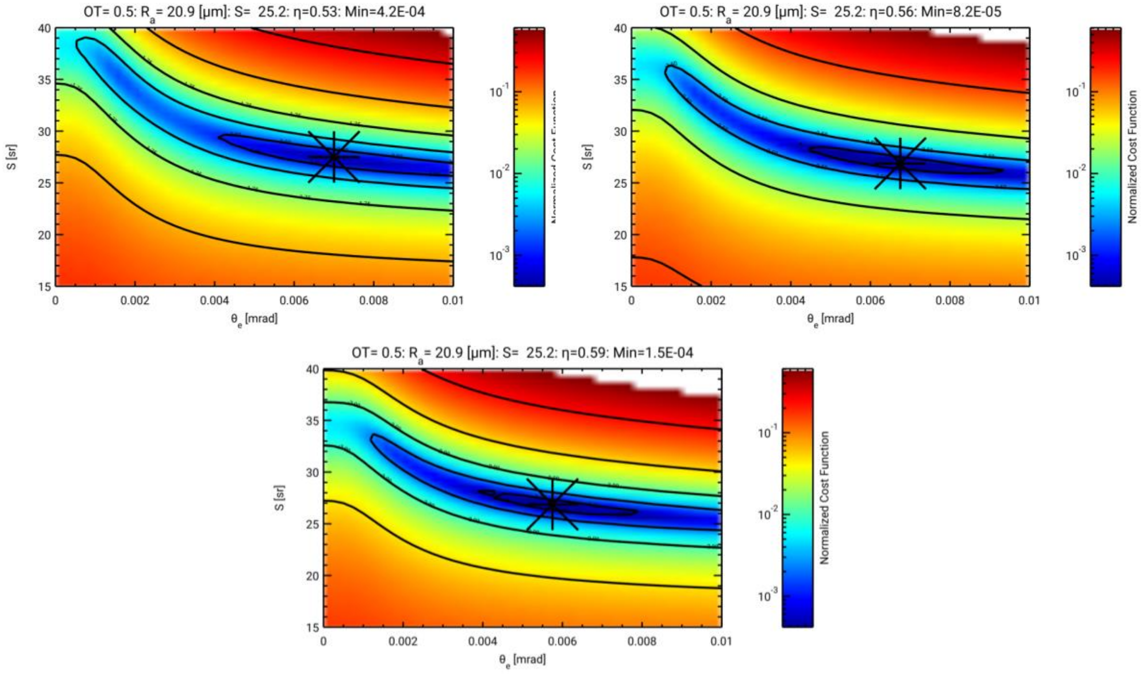


Figure 14: Cross-sections of the cost function assuming perfect calibration and cross-talk correction for a single-layer simulated ice cloud as a function of θ_{sc} and S for three different values of η. The * symbols mark the location of the minimum for each slice.

3.1.4.9 Error Estimates

Following Press et al., the formal covariance matrix of the retrived state-vector can be derived from the curvature-matrix

$$\mathbf{C} = \frac{1}{2}\left[\frac{\partial^2 \chi^2}{\partial x^f{}_i \partial x^f{}_j}\right]^{-1} \qquad (2.2.52)$$

where the partial derivatives are taken at the minimum point. Applied to Eqn. (2.2.42), this will yield the covariance matrix for the calibration factor, the extinction-to-backscatter ratios, etc.

The variances and covariances involving the retrieved extinction values and optical thicknesses are then derived using the state-vector covariance matrix and the signal error estimates appropriately combined with the partial derivatives of the extinction and optical thickness with respect to the state variables and the signals themselves. More details are given in Section 5.5.2.

3.1.5 A-TC

The A-TC procedure is comprised of a number of layer-determination and classification procedures and is based on previous developments made during the ATLAS and IRMA projects. The classification procedures have been realised as subroutines which are called from either A-AER or A-EBD. The layering determination procedures are called by the A-AER algorithm code but are read in and used in the A-EBD procedure.

3.1.5.1 Layering Determination

A-TC includes the identification of what are termed significant layers. The basic idea is to apply simple statistical tests to continuous layers where targets are have been identified as being present in order to identify significant sub-layers based on differences in the depolarization ratio and/or lidar-ratio and/or backscatter-ratio (See Figure 15).

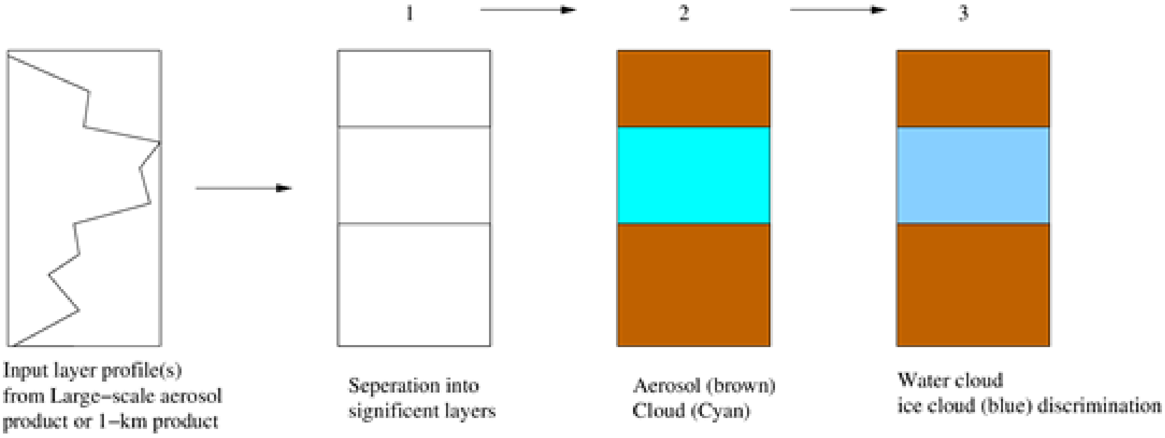


Figure 15: Schematic depiction of the concept of the significant layer determination procedure.

Once the sub-layering structure processing has been completed, the aerosol/ice-cloud/water-cloud discrimination is performed. This determination is based on conditional thresholds applied to extinction, backscatter and depolarization levels cast into a decision tree.

3.1.5.1.1 Detection of significant height boundaries.

The approach used here is somewhat “brute-force”. For a given trial number of layers the goodness-of-fit parameter for all possible layering structures is evaluated. The goodness of fit parameter (or reduced chi-squared parameter) here is defined (for 2 layers) as

$$\chi^2_{r,x} = \frac{1}{nz-1-nl}\left[\sum_{i=1}^{nl_1}\sum_{j=i+1}^{nl}\left(\left(\frac{x_i-\overline{x_1}}{\sigma_{x_i}}\right)^2+\left(\frac{x_j-\overline{x_2}}{\sigma_{x_j}}\right)^2\right)\right] \qquad (2.2.53)$$

where nz is the number of altitude bins being considered, nl is the number of layers being considered (2 in this example) and x_i can be, for example, the depolarization ratio δ or the Backscatter ratio R and the barred quantities represent error weighted averages. i.e.

$$\bar{x} = \frac{\sum \frac{x_i}{\sigma_{x_i}}}{\sum \sigma_{x_i}^{-1}} \tag{2.2.54}$$

where the summations are carried out over the appropriate layer indices. After all possible goodness-of-fit parameters are determined for a fixed number of layers the best value is chosen. This is done for different number of allowed sub-layers up to a set maximum. Finally, the optimal number of layers is selected by using a criteria that involves trading-off the improvement in goodness-of-fit parameters with increasing number of layers.

3.1.5.2 Cloud-Aerosol identification

The physical information available using the lidar measurements is strictly not enough to unambiguous separate clouds from aerosol. Thus, there is no alternative but to appeal to the fact that clouds tend to be optically thicker than aerosol and to augment the decision process by using auxiliary information such as expected tropopause height from, for example, ECMWF forecast or analysis fields. Accordingly, the procedure described here uses a number of thresholds applied primarily to the measured backscatter.

3.1.5.2.1 Cloud Phase Determination

Water/ice discrimination will be performed using a procedure similar to that used for CALIPSO. Based on recent ATLID ECSIM simulations, the expected relationship between layer-integrated attenuated backscatter and layer depolarization is shown in the Right-Panel of Figure 16. Here it can be seen that the relationship observed from actual CALIPSO measurements can be reproduced using the ECSIM lidar multiple-scattering code applied to water clouds with various optical depths and effective cloud particle sizes. The same lidar multiple-scatter code applied to ATLID shows that a similar relationship for water-clouds will be observed.

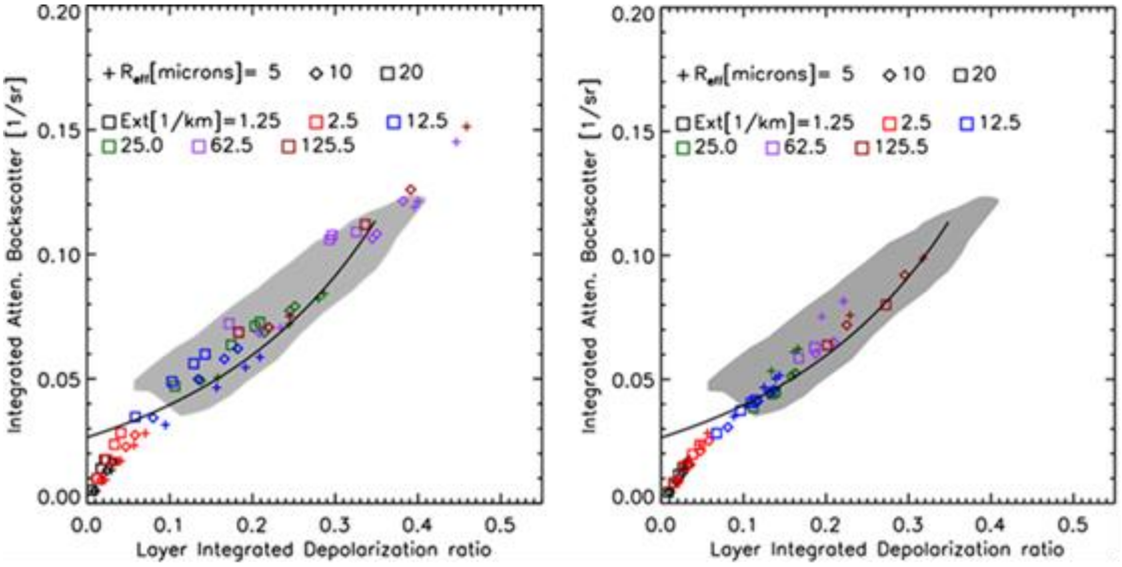


Figure 16: Integrated-layer-attenuated-backscatter vs. layer-depolarization for CALIPSO (Left) and ATLID (Right). Here the Grey-Area corresponds to the range of actual CALIPSO observations (Hu et al. 2007) and the points correspond to Monte-Carlo lidar radiative transfer simulation results using the indicated extinctions and particle sizes. The solid lines correspond to Eq.(6) of Hu et al. (2007).

3.1.5.3 Aerosol Type Determination

For areas classified as being aerosols, A-TC will contain procedures for assigning probabilistic aerosol type. This type is being based primarily on the observed relationship between the aerosol lidar-ratio and the depolarization ratio. Aerosols are classified (e.g. dust, marine etc.) according to a rule-based system. Auxiliary information such as the temperature and Tropopause height (from X-MET) are also used in the decision process.

Up to now, there is no specific aerosol typing scheme available in the literature that considers a classification based on measurements at 355 nm. Therefore, we have collected what is to our knowledge the largest data set of aerosol observations involving measurements of the EarthCARE-relevant parameters of particle lidar ratio and particle linear depolarization ratio at 355 nm. The complete discussion on this is presented in the EarthCARE L2 coordination document[AD9] and is summarized here.

For the aerosol classification, a suitable set of basic aerosol types must be defined. This basic set must be complete enough to reasonably encompass the range of types encountered in nature but should not be more extensive than required so that the basic number of types is tractable. The classification should allow the separation of natural and anthropogenic aerosols. The types need to be described consistently in terms of microphysical properties (size, shape, refractive index), which are used to represent them in scattering models, and of optical and radiative properties, which are either observed with the EarthCARE instruments and used for the classification (lidar ratio, linear depolarization ratio, Ångström exponent) or applied for radiative-transfer calculations and closure studies (single-scattering albedo, asymmetry parameter).

The classification illustrated in Figure 17 together with Table 1 serves as the basis for the aerosol typing scheme for ATLID.

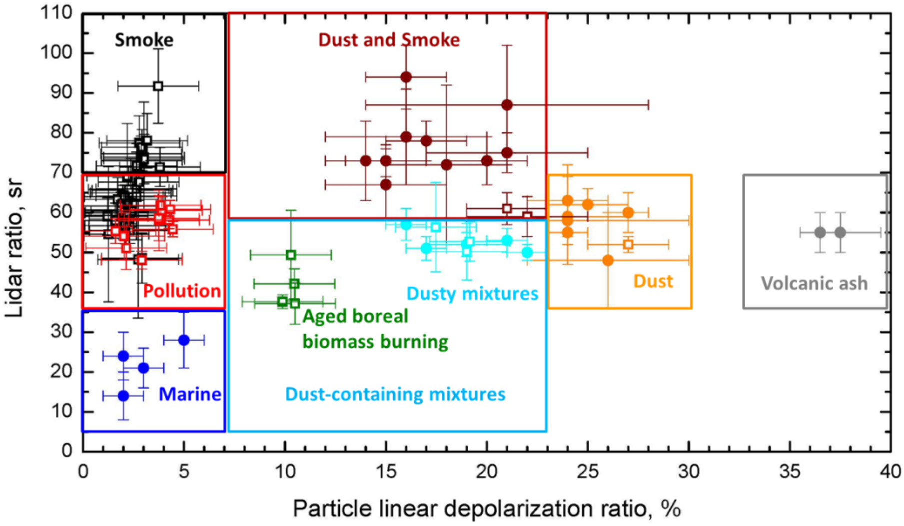


Figure 17: Aerosol classification from measurements of lidar ratio and particle linear depolarization ratio at 355 nm. Ground-based observations were performed with the Raman-polarization lidars POLIS (University of Munich, dots) and PollyXT (Leibniz Institute for Tropospheric Research, open squares) at Cape Verde (dust, marine, dust and smoke, dusty mixtures; dots; Groß et al., 2011), Leipzig, Germany (pollution, aged boreal biomass-burning aerosol, dusty mixtures; open squares), Munich, Germany (volcanic ash; dots; Groß et al. 2012), in the Amazon Basin (smoke; open squares; Baars et al., 2012), and over the North Atlantic (dust, dust and smoke; open squares; Kanitz et al., 2013).

For usage for the EarthCARE mission the end-to-end aerosol classification model (HETEAC) has been developed which will:

- deliver the required theoretical description of aerosol microphysics that is consistent with experimentally derived optical properties (hybrid approach) and
- close the loop from aerosol microphysics to radiative properties (end-to-end approach).

HETEAC provides the microphysical description and the derived optical and radiative properties of a number of pre-defined pure aerosol types as well as mixtures of these types. The definition of aerosol types follows the common understanding of aerosol typing and be consistent with the CALIPSO approach. The proposed types are listed in the Table below.

Table 1: Aerosol types for the EarthCARE aerosol classification scheme.

Aerosol Type	Typical Optical Depth at 355 nm	Typical Lidar Ratio at 355 nm	Typical Linear Depolarization Ratio at 355 nm	Typical Ångström exponent (355 nm/670 nm)	Remarks
Clean marine	0.05 – 0.15	15 – 30 sr	<0.05	0.0 – 0.5	Contains sea salt and components of oceanic origin, hygroscopic, relatively large droplets
Clean continental	0.05 – 0.15	30 – 50 sr	<0.05	1.0 – 2.0	Contains an unspecific mixture of background aerosol particles from natural and anthropogenic sources
Mineral dust	0.3 – 1.5	50 – 60 sr	0.20 – 0.30	0.0 – 0.5	Contains large mineral, non-spherical particles originating from deserts and dry soils, typically hydrophobic
Biomass-burning smoke	0.1 – 1.0	60 – 80 sr	<0.10	0.5 – 1.5	Contains mainly small, absorbing particles with variable properties depending on type of fire and aging processes
Pollution	0.15 – 1.0	50 – 70 sr	<0.05	1.0 – 2.0	Contains sulphates, nitrates, soot, and other components of anthropogenic origin, typically small and mainly hygroscopic particles
Stratospheric aerosol	<0.2	30 – 60 sr	<0.05	1.0 – 2.0	Contains mainly small sulphuric-acid droplets

Characteristic bimodal log-normal size distributions, spectrally resolved complex refractive indexes and a shape parameter (distribution) per aerosol mode shall describe the microphysical properties. These properties will be adapted such that

- consistency with literature values is given as far as possible,
- consistency with other retrieval schemes of space-borne and ground-based remote-sensing activities is reached as far as possible, and

characteristic intensive properties (lidar ratio, depolarization ratio, Ångström exponent) as derived from advanced ground-based remote-sensing observations are reproduced. Based on a detailed comparison of literature and computational studies, as shown in Table 2 the following recommendations are made for the individual components.

Table 2: Recommendations for the modeling of dust in HETEAC when optimizing for only 355nm or when fitting to both the 355 and 532nm observations.

	Dust model optimized for 355 nm	**Dust model optimized for 355 and 532 nm**
r_{eff}, µm	**1.94**	**2.2**
$r_{0,\mathrm{V}}$, µm	**2.32**	**2.634**
$\ln \sigma^*$	**0.6**	**0.6**
m_{R} (355, 532 nm)	**1.54, 1.50**	**1.54, 1.48**
m_{I} (355, 532 nm)	**0.006, 0.004**	**0.006, 0.004**
Shape	**Gasteiger’s spheroid distribution**	**Gasteiger’s spheroid distribution**
S (355, 532 nm), sr	**57.9, 39.2**	**65.7, 46.3**
δ (355, 532 nm), %	**25.1, 31.9**	**23.9, 32.6**
ω_0 (355, 532 nm)	**0.780, 0.871**	**0.765, 0.860**
g (355, 532 nm)	**0.825, 0.776**	**0.837, 0.793**
$å_{\mathrm{ext}}$ (355, 532 nm)	**−0.115**	**−0.106**
$å_{\mathrm{bsc}}$ (355, 532 nm)	**−1.079**	**−0.974**

Table 3: Microphysical and optical properties at 355 nm of four basic aerosol components used to simulate multimodal particle distributions in as proposed by HETEAC.

	Coarse mode Nonspherical (Dust)	**Coarse mode Spherical (Sea salt)**	**Fine mode Non-absorbing (Water soluble)**	**Fine mode Absorbing (Soot)**
r_{eff}, µm	1.94	1.94	0.14	0.14
$r_{0,\mathrm{N}}$, µm	0.788	0.788	0.07	0.07
$r_{0,\mathrm{V}}$, µm	2.32	2.32	0.1626	0.1626
$\ln \sigma^*$	0.6	0.6	0.53	0.53
m_{R}[4]	1.54	1.36	1.45	1.50
m_{I}	0.006	1.e−8	1.e−5	0.043
Shape	Gasteiger’s spheroid distribution	Spherical	Spherical	Spherical
S, sr	57.9	18.1	59.9	117.3
δ, %	25.1	0.0	0.0	0.0

Positions in the lidar ratio-depolarization diagram can be described by mixtures of three aerosol types.
The depolarization ratio depends very sensitively on the contribution of fine-mode particles. The dependence is non-linear in terms of volume, i.e. the mixing of dust with fine-mode particles or sea salt can be well resolved as long as the dust contribution dominates. Vice versa, a relatively large amount of dust is needed to cause a considerable depolarization ratio.

---

[4] m_{R} & $m_{\mathrm{I}}i$ refer to the real and imaginary part of the aerosol refractive index

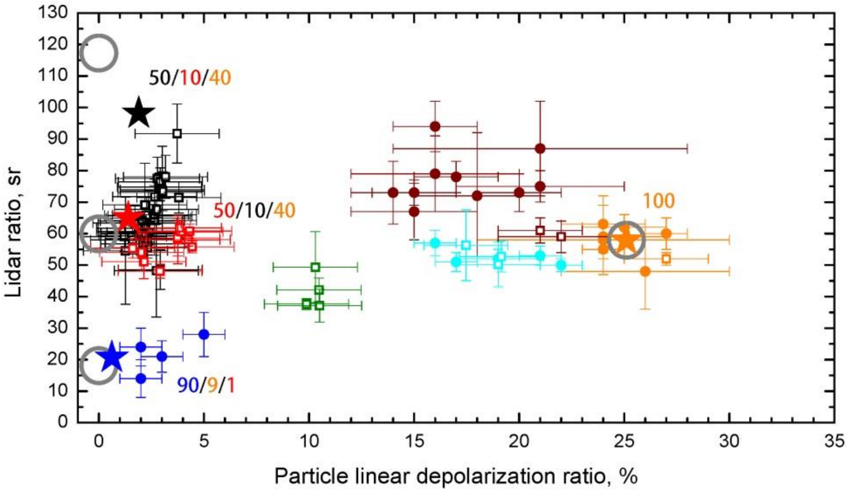


Figure 18: Proposed S-δ center positions of aerosol types (stars) composed of one or several basic aerosol modes (gray circles) in comparison to experimental data. The numbers indicate the volume mixing ratios of the modes (in %).

Similarly, very large and very small lidar ratios are only obtained when the soot-like or the sea-salt components dominate, respectively. It can be seen in that a considerable change of the optical parameters is only observed when one of the components starts dominating the mixture. Many of the mixtures produce lidar ratios of 50–70 sr. Such values are indeed most often observed in nature and are considered to be typical for continental, polluted conditions.
In view of these findings, the definition of aerosol types becomes a bit easier, since the classification can be based on the assumption that one typical particle mode is dominating, but “natural” contributions of particles with different properties will also be possible.

3.1.5.3.1 Retrieving Aerosol, Stratospheric Aerosol and Stratospheric Cloud probabilities and mixtures

The target classification focuses on retrieving a number of mixtures based on the pure types described above, which conditions span the observable parameter space. The types and values should, when applicable, be the same for the AM-COL, ACM-TC and A-TC algorithms. Ice crystals are included as an aerosol type since they span part of parameter space not occupied by aerosols and can thereby be used to separate thin ice clouds from aerosol fields and assist in a better cloud mask and typing.

The used aerosol mixtures currently are provide in , including the parameters defining the 2D-Gaussian probability distribution as described by Eq. (2.2.54). All parameters are defined in the configuration file to make the validation and updating of the algorithm to new parameterizations straightforward. The 1-sigma distribution values for each of the particle types are estimated to span parameter space but will need to be validated in a future effort.

Table 4: Aerosol, Stratospheric cloud and Stratospheric aerosol types and the parameters defining the probability distributions for each of these types. The tropospheric and stratospheric aerosol values have been chosen to be consistent with the HETEAC scheme and (see Figure 17) and Gross et al, 1995. The stratospheric cloud depolarization related values are loosely based on Noel et. al (2009) and Achert et al. (2014). The PSC lidar-ratio variables are consistent with those associated with the PSC scene (seen Appendix-A) and are only considered placeholders at this point. It is expected that the values will be updated based on post-launch observations.

#	**Name**	**Angle (deg)**	δ_o	σ_δ	S	σ_S
Troposphere						
1	Marine	20	0.03	0.04	20	12
2	Continental Pollution	-3	0.03	0.04	55	15
3	Smoke	-5	0.03	0.04	88	12
4	Dusty-Smoke	-15	0.14	0.06	73	15
5	Dusty-Mix	15	0.14	0.06	43	15
6	Dust	0	0.22	0.05	55	15
7	Ice	0	0.40	0.10	15	10
Stratosphere Aerosols						
6	Ash/Coarse mode dust	0	0.45	0.08	55.	10
7	Sulfate	0	0.03	0.05	40.	15
8	Smoke	-5	0.03	0.04	70	13
Stratosphere Clouds						
6	Stratospheric Supercooled Ternary (STS)	90	0	0.2	55	15
7	Nitric Acid Trihydrate (NAT)	0	0.15	0.08	40	15
8	Liquid-Ice mixture	0	0.40	0.25	30	15

Each of the probability distributions is described using bivariate Gaussian probability densities and are described by:

$$\begin{aligned} A &= \frac{\cos(\theta)^2}{2\sigma_\delta^2} + \frac{\sin(\theta)^2}{2\sigma_S^2} \\ B &= \frac{-\sin(2\theta)}{4\sigma_\delta^2} + \frac{\sin(2\theta)}{4\sigma_S^2} \\ C &= \frac{\sin(\theta)^2}{2\sigma_\delta^2} + \frac{\cos(\theta)^2}{2\sigma_S^2} \\ P(\delta,S) &= P_0 \exp(-1(A(\delta-\delta_0)^2 + B(\delta-\delta_0)(S-S_0) + C(S-S_0)^2)) \end{aligned} \tag{2.2.54}$$

with θ the rotation of the 2D-Gauss, assuming the depolarization on the x-axis. S is the lidar ratio, δ the depolarization σ_s and σ_δ their respective variances. Each distribution is normalized to a maximum of 1. In Figure 19, the distributions described by the parameters for the tropospheric types in Table 1 are depicted.

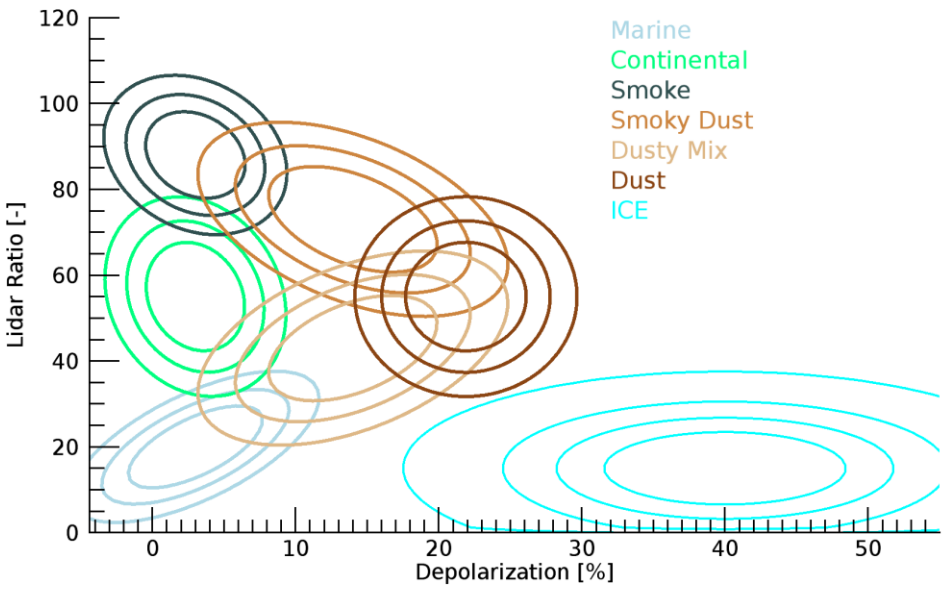


Figure 19: Probability density distributions for the six aerosol types and ice crystals as is currently being used.

3.1.5.3.2 *Calculating the type probabilities*

For each observed combination of lidar ratio and depolarization, the probability can be calculated for all the types being considered. If this calculation were to be done rigorously, one would need to calculate the convolution of the 2D observed Gaussian with the 2D bivariate Gaussian probability density functions. This calculation becomes an enormous expression when solved analytically and therefore the calculation is carried out approximately by retrieving the probability 49 times (7x7) between plus and minus 3σ in both the lidar ratio and depolarization (see Figure 20 for an example). The effective probability (Peff,d) is calculated by adding the weighted 49 probabilities (P) according to:

$$P^{l}_{eff,d}(\delta_k,S_k)=\sum_{i=1}^{7}\sum_{j=1}^{7}P(\delta_k+(i-4)\sigma_{\delta,k},S_k+(j-4)\sigma_{S,k})\bullet W(i,j)$$

$$W(i,j)=\frac{1}{2\pi}\exp(-\frac{(i-4)^2}{2}-\frac{(j-4)^2}{2}) \tag{2.2.54}$$

with S_k and δ_k the observed lidar ratio and depolarization with their respective error estimates ($\sigma_{S,k}$, $\sigma_{\delta,k}$) and $P^{l}_{eff,d}$ the probability of the observation k for type l . The weighting function (W) assumes an even spread between -3σ and +3σ and P is the probability distribution as was defined by Eq. .

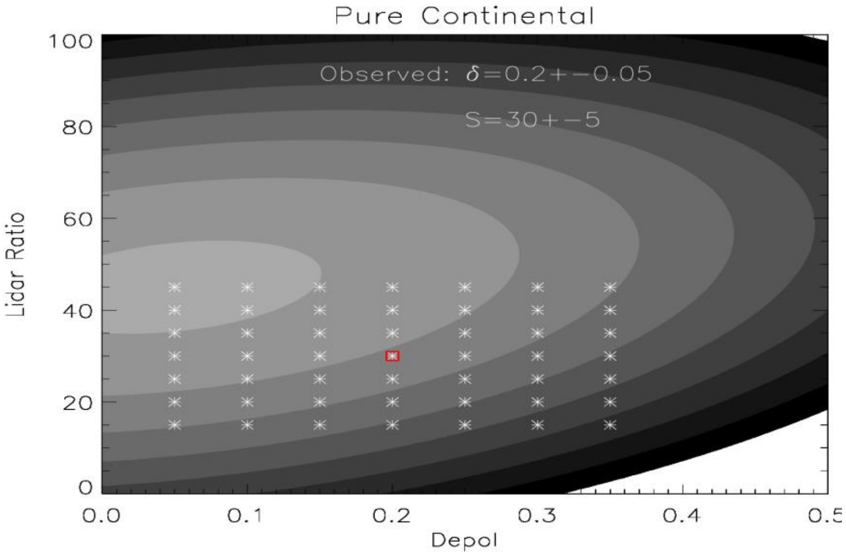


Figure 20: Calculating the clean continental type probability for an observed point with a lidar ratio of 30 and depolarization of 0.2. The calculation is performed at all 49 points and weighted according to the distance to the observed value [Equation]

The resulting probabilities for all the relevant types is the first direct result and is be reported in the A-TC product file.

3.1.5.3.3 Assignment of aerosol type

Even though the type probabilities are the main results from the procedure, the most probable type will be provided in the output. Defining a single aerosol type will only be possible in those cases where a single type probability is much higher than any of the other types. In Figure 21, three slices through the tropospheric parameter space are plotted, each at a constant depolarization. Shown are the percentages of the aerosol probabilities to the sum of all the probabilities at each point. As all the aerosol types considered here (except for marine aerosols) are centered around relatively high lidar ratios the marine aerosol type is the preferred type for lidar ratio's below 20 for low depolarizations. At higher depolarizations all low lidar ratio measurements are assigned as ice crystals (light blue). Since an aerosol type can dominate a point in parameter space relatively it can have a negligible absolute value. To ensure that a relevant retrieval of type is made a minimum probability is used. As the algorithm deals with Gaussian distributions the minimum is set to the 3σ level (which includes 99.73% of all data), or a probability of ($e^{-(9/2)} = 0.0111$).

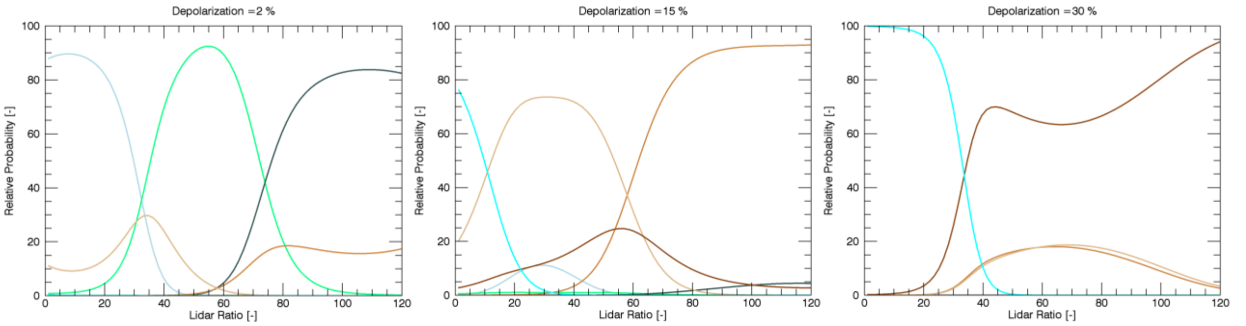


Figure 21: Normalized probability fractions for the 5 aerosol types. for three slices of depolarization. The observations are assumed to have a negligable error estimate in these cases. The line colors folow the definition as shown in Figure 19.

In the described parameters, there are three type of dust(y) mixtures. It is to be validated if all three should be retrieved or if for instance the dusty-mix and smoky-dust type should be assigned a single dusty mixture type. A future effort should also be used to see if the a volcanic (at larger depolarization and lidar ratio than the current dust definition) would be beneficial.

Similar plots can be created for the stratospheric aerosols and clouds, however in the latter case the distinction in lidar ratio will be near impossible due to the large σ_S assumed for the three PSC types. Effectively the PSC is separated on depolarization only until a better PSC parameterization at 355nm can be made based on data.

3.1.5.4 Aerosol and Ice horizontal homogeneity

The lidar ratio and depolarization and therefore all the subsequent retrievals are based on vertical layers. Even though there is substantial horizontal smoothing the layer averaged values from subsequent JSG pixels can vary enough to change the aerosol/ice typing from one type to the next and back to the former, i.e. smoky dust to dust and vice versa. Except for observations very close to sources, aerosol fields are more homogenous. To homogenise the aerosol and fields and fill in gaps for which no retrieval was possible, despite a good extinction retrieval a hybrid-median filter type calculation is performed.

The hybrid median filter checks the entire image pixel by pixel using an n x 3 box, where n is an odd integer of 5, 7,9 etc.. The centre pixel is calculated using the two ‘diagonal’ and the horizontal and vertical rows within this box (Figure 22). For each of the rows the median value is calculated, after which the median value of these four median values is taken. This type of algorithm is very effective in removing single noise events and filling small gaps and is one of the two main routines used in the ATLID Featuremask processor (AD1). Classifications use integers and a value can have a very different meaning from the next. The median is therefore not the appropriate tool and instead the mode of the four lines is calculated.

The hybrid-mode-algorithm is run iteratively five times to ensure that the image has converged, e.g. there are no more changes in the image between this iteration and the next. As only mode values are used, there are no smoothing edge effects, hence the need for only a few iterations to converge. This procedure is performed for only aerosol and ice cloud pixels and not for liquid droplets, surface and attenuated regions.

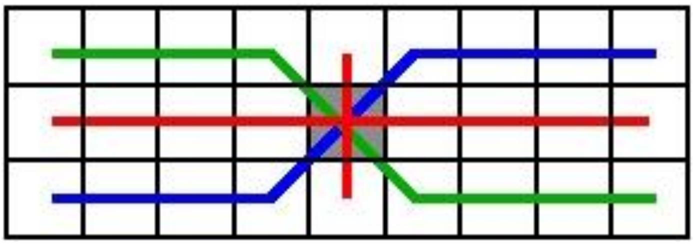


Figure 22: The hybrid mode filter example uses a 9 x 3 box to determine the value of the center box (shaded). For each line (2x red, blue and green) the first mode is calculated. From the resulting 4 modes, the first and second modes are calculated and used for homogenising the classification.

3.1.6 A-ICE

The APRIL A-ICE algorithm will be based on the ATLAS algorithm. In terms of coding and algorithm complexity, this algorithm is arguably the least complicated algorithm of the suite of algorithms to be developed and implemented within APRIL. The basic approach is to use the ice cloud extinction values from A-EBD along with the temperature to and apply empirical relationships (e.g. Heymsfield et al., 2005) in order to estimate the ice cloud water content and effective radius.

The parameterization for the ice-water content has the form

$$IWC[\mathrm{gm}^{-3}] = C_0 \left(\frac{\alpha[\mathrm{km}^{-1}]}{1000} \right)^{C_1} \tag{2.2.55}$$

where C_o and C_1 depend on the temperature.

Using the fact that, in general, $R_{eff} = {}^3\!/_8\, \rho_i^{-1} IWC \alpha^{-1}$ the corresponding ice cloud effective radius can also be estimated in a consistent manner. i.e.

$$R_{eff} = \frac{3}{8} \frac{C_o}{\rho_i} \left(\frac{\alpha}{1000} \right)^{C_1 - 1} \tag{2.2.56}$$

The above coefficients depend on temperature in the following manner

$$C_0 = C_{0,0} + C_{0,1} T[^{\circ}\mathrm{C}] \tag{2.2.57}$$

and

$$C_1 = C_{1,0} + C_{1,1} T[^{\circ}\mathrm{C}] \tag{2.2.58}$$

4 JUSTIFICATION FOR SELECTION OF THE ALGORITHMS

The main justifications for the selection of the algorithms with respect to the criteria specified by ESA are given in the following table.

Criteria	A-AER	A-EBD	A-TC	A-ICE
Scientific importance with respect to the EarthCARE scientific mission goals	Aerosol extinction and backscatter at the 10km+ scale are essential products. Output of A-AER essential for A-EBD.	Lidar-only Extinction and backscatter profiles at the 1-km scale are essential products.	Lidar-only Identification of targets as being water cloud, ice cloud or aerosol (type)are essential products.	A priori estimates of ice cloud particle size are essential to the operation of A-EBD and A-AER
Demonstrated scientific accuracy and precision	Direct HSRL method Is well-understood and is the standard method for the retrieval of backscatter and extinction from a HSRL system.	The approach is new. However previous work conducted during the ATLAS project indicates the approach is vastly superior in terms of accuracy and precision compared to more conventional HSRL approaches.	Approach has been used for Calipso and various ground and airborne lidar systems.	Relevant parametrizations based on a very large body of in situ measurements.
Reliability	The direct approach is reliable in terms of convergence and stability.	Extensive testing will insure the reliable operation of the algorithm in terms of convergence and stability.	Decision tree approach is reliable	Decision tree approach is reliable
Maturity and heritage, including validation status	Direct HSRL method Is well-understood and is the standard method for the retrieval of backscatter and extinction from a HSRL system.	Approach based on earlier ATLAS work. Simulations support the validity of the basic approach.	Approach has been used for Calipso and various ground and airborne lidar systems.	Relevant parametrizations based on a very large body of in situ measurements.
Capability to estimate data product reliability and errors and produce respective data product error descriptors	Error estimates can be derived.	A-EBD is a OE based approach so that detailed error estimates are available	Target probabilities are reported as well as simple discrete classifications.	A priori error bounds are well-known.
Variational techniques, such as Optimal Estimation, shall be	The direct approach is not variational. However, A-AER output does serve as a priori input to A-EBD (which is a variational	A-EBD is a OE based approach	N/A	N/A

preferred whenever possible	approach)			
Performance (i.e., execution speed)	The direct HSRL method is fast.	The newly developed Platt+tails approach is perhaps the fastest conceivable treatment of MS affects with the required accuracy.	Decision tree approach is very fast	Parameterization based approach is very fast.

5 MATHEMATICAL ALGORITHM DESCRIPTION

5.1 Input parameters

5.1.1 Input Products

Input	Remarks
A-NOM (L1b Lidar data)	The source of the lidar attenuated backscatters at native resolution along with necessary instrument parameter (e.g. cross-talk coefficient) values and error estimates
X-JSG	The source of the information necessary the mapping of the L1b measured grid to the output JSG grid
X-MET	The source of the atmospheric temperature and density fields necessary for the inversion process
A-FM	Feature mask product used to help identify altitudes where clouds and/or aerosol are present as well as ground altitude and fully attenuated regions.

5.2 Configuration parameters

Here the configuration parameters are sub-divided into different groups. In the symbol entry is empty this means that the symbol is the same as the variable name.

Variable	Symbol	Size	Type	Description	Unit
General parameters					
logging_output_file_aer		1	string	file name for logging messages	1
logging_output_file_ebd		1	string	file name for logging messages	1
logging_level		1	int	DEBUG = 0, INFO = 1, PROGRESS = 2, WARN = 3, ERROR = 4	1
laser_full_angle_divergence		1	float	full-angle laser divergence	mrad
telescope_full_angle_field_of_view		1	float	full-angle telescope fov	mrad
adjust_input_feature_mask		1	int	Adjust the FM based on smoothed R and Retrieved extinction	1
A-AER and A-EBD					
vertical_derivative_determination_window		1	float	vertical window for determining S and Extinction	km
maximum_horizontial_smoothing_window		1	int	sliding expanding horizontal maximum window	km
horizontial_smoothing_target_SNR		1	float	Target signal-to-noise ratio for the horizontial smoothing window limits	km
smooth_R_cutoff_for_horizontial_smoothing		1	float	Discard points with Scattering ratios greater than this when doing horizontial averaging	1
feature_mask_cutoff_for_horizontal_smoothing		1	float	Discard points with FM values greater than this when doing horizontial averaging	1
feature_mask_difference_layer_threshold		1	float	If the FM changes by this much or greater in one cell then start a new layer	1
dR2_difference_layer_threshold		1	float	If the absolute scattering ratio difference changes by this much or greater in one cell then start a new layer	1
preliminary_smoothing_horizontial_window		1	int	Number of along_track pixels to smooth for initial classification determination	1
preliminary_smoothing_vertical_window		1	int	Number of vertical pixels to smooth for initial classification determination	1
layer_splitting_chisq_relative_step_stop		1	float	Stop layer splitting if chi2 decreases to less than this relative value for any one step	1
layer_spltting_chisq_relative_mi		1	float	Stop layer splitting is chi2 is within this factor of the minimum level reached	1

n_stop					
medium_ resolution_ number_of_ jsg_pixels		1	int	Number of along_track pixels to smooth for medium resolution classification determination	1
low_resolution_number_of_jsg_pixels		1	int	Number of along_track pixels to smooth for low resolution classification determination	1
lidar_feature_ mask_threshold		1	int	Threshold to apply to the target mask	1
lidar_feature_ mask_threshold_force_cloud		1	float	Threshold to apply to the target mask..above this value targets are almost certainly clouds	1
extend_feature_ mask_ standard_ deviation_factor		1	float	If the smoothed Mie signal in a_aer is greater than a factor of extend_fm_sd_fac than the estimated sd then force the internal feature mask to 1 [used only if apply_class_mod_filter =1]	1
censor_feature_ mask_ standard_ deviation_factor		1	float	If smoothed Mie signal and CRO signal in a_aer is less than a factor of censor_fm_sd_fac than the estimated sd then force the internal feature mask to 0 (or -1) [used only if apply_class_mod_filter =1]	1
class_mode_ filter_window		1	int	Width of hybrid mode filter to apply when class filtering	Jsg_pixels
class_mode_ SNR_minimum_threshold		1	float	When class filtering remove points whose ext SNR is below this threshold	1
minimum_cloud_ R_threshold		1	float	Threshold on the backscatterratio (R) for cld detection	1
maximum_ altitude_to_fill_with_aerosol		1	float	Below this altitude BL aerosols exist	1
ice_water_ seperation_ temperature		1	float	Above this temperature [Celsius] clouds will be classified as 'water-clouds', below as 'ice-clouds'	C
homogeneous_ freezing_ temperature		1	float	Below this temperature [Celsius] clouds can only be ice'	C
integrated_ depolarization_ to_integrated_ beta_ratio_ threshold		1	float	Ratio of int depol to int beta used to distinguish ice from water	sr m
force_class_to_water		1	int	Force high scattering ratio targets to water 1==> yes	1
water_cloud_R_threshold		1	float	Scattering ratio above which we assume only water clouds	1
maximum_ allowed_layer_ extent		1	int	Maximum allowed layer extent	1
apply_class_ mode_filter		1	int	Apply class filtering to aer classes within A-EBD 1==> yes	1
apply_class_ mode_filter_to_ebd_input		1	int	Apply class filtering to output of A_AER (input to A_EBD) types 1==> yes	1
class_mode_ filter_window	$nhor_{cl}$	1	int	Width of hybrid mode filter to apply when class filtering	jsg pixels
class_mode_ SNR_minimum_threshold	snr_{cl}	1	float	When class filtering remove points whose ext SNR is below this threshold	1

use_last_gate_on_bad_solution_threshold		1	float	If this threshold is exceeded in the zero finding step in the OE routine, then use solution from previous range gate: - ==> disable this check	$m^{-1}sr^{-1}$
minimum_relative_error_in_cost_func_mie		1	float	Minimum allowed relative error to Mie_para when calculating inversion cost-function	1
minimum_relative_error_in_cost_func_ray		1	float	Minimum allowed relative error to Ray_para when calculating inversion cost-function	1
minimum_absolute_error_in_cost_func_mie			float	Minimum allowed absolute error to Mie_para when calculating inversion cost-function (- to disable use of sse_mie)	$m^{-1}sr^{-1}$
minimum_absolute_error_in_cost_func_ray		1	float	Minimum allowed absolute error to Ray_para when calculating inversion cost-function" units	$m^{-1} sr^{-1}$
gradient_convergence_tolerance		1	float	Convergence tolerance in dfpmin (gradient)	1
step_convergence_tolerance		1	float	Convergence tolerance in dfpmin (deltaD)	1
maximum_number_of_iterations		1	int	Max # of iterations	1
step_factor_used_in_line_search		1	float	Maximum step factor in dfpmin	1
SSE_limit		1	float	If initial guess sse is above this limit then do not attept an inversion	1
delta_used_in_numerical_jacobian_determination	Δ	1	float	Step size to use in for numerical Jacobian determination in error calculation procedure	1
a_priori_mie_MS_f_factor	f^{a}_{MSp}	1	float	A priori Factor to account for less MS backscatter for Mie vs Ray multiple scattering	1
starting_value_mie_MS_f_factor	f^{o}_{MSp}	1	float	Starting Factor to account for less MS backscatter for Mie vs Ray	1
sigma_a_priori_mie_MS_f_factor	$\sigma f_{MSp} / f_{MSp}$	1	float	Associated fractional AP sd	1
a_priori_water_S_ratio	S_{water}	1	float	A priori S_ratio for Water	sr
starting_value_water_S_ratio	S^{o}_{water}	1	float	starting S_ratio for Water	sr
sigma_a_priori_water_S_ratio	$\sigma_{S_{water}} / S_{water}$	1	float	The A priori fractional sd S ratio for Water	1
a_priori_water_eta_MS	η_{water}	1	float	A priori eta for Water	1
starting_value_water_eta_MS	η^{o}_{water}	1	float	starting eta for Water	1
sigma_a_priori_water_eta_MS	$\sigma_{\eta_{water}} / \eta_{water}$	1	float	The A priori fractional sd eta ratio for Water	1
a_priori_water_theta_MS	$\theta_{sc,water}$	1	float	A priori theta for Water	mrads
starting_value_water_theta_MS	$\theta^{o}_{sc,water}$	1	float	starting theta for Water	mrads

sigma_a_priori_water_theta_MS	$\sigma_{\theta_{sc,water}}/\theta_{sc,water}$	1	float	The A priori fractional sd theta ratio for Water	1
a_priori_ice_S_ratio	S_{ice}	1	float	A priori S_ratio for Ice	sr
starting_value_ice_S_ratio	S^{o}_{ice}	1	float	starting S_ratio for ice	sr
sigma_a_priori_ice_S_ratio	$\sigma_{S_{ice}}/S_{ice}$	1	float	The A priori fractional sd S ratio for ice	1
a_priori_ice_eta_MS	η_{ice}	1	float	A priori eta for Ice	1
starting_value_ice_eta_MS	η^{o}_{ice}	1	float	starting eta for Ice	1
sigma_a_priori_ice_eta_MS	$\sigma_{\eta_{ice}}/\eta_{ice}$	1	float	The A priori fractional sd eta ratio for Ice	1
a_priori_ice_theta_MS	$\theta_{sc,ice}$	1	float	A priori theta for Ice	mrads
starting_value_ice_theta_MS	$\theta^{o}_{sc,ice}$	1	float	starting theta for Ice	mrads
sigma_a_priori_ice_theta_MS	$\sigma_{\theta_{sc,ice}}/\theta_{sc,ice}$	1	float	The A priori fractional sd theta ratio for Ice	1
a_priori_aerosol_S_ratio	S_{aer}	1	float	A priori S_ratio for aerosol	sr
starting_value_aerosol_S_ratio	S^{o}_{aer}	1	float	starting S_ratio for aerosol	sr
sigma_a_priori_aerosol_S_ratio	$\sigma_{S_{aer}}/S_{aer}$	1	float	The A priori fractional sd S ratio for aerosol	1
a_priori_aerosol_eta_MS	η_{aer}	1	float	A priori eta for aerosol	1
starting_value_aerosol_eta_MS	η^{o}_{aer}	1	float	starting eta for aerosol	1
sigma_a_priori_aerosol_eta_MS	$\sigma_{\eta_{aer}}/\eta_{aer}$	1	float	The A priori fractional sd eta ratio for aerosol	1
a_priori_aerosol_theta_MS	$\theta_{sc,aer}$	1	float	A priori theta for aerosol	mrads
starting_value_aerosol_theta_MS	$\theta^{o}_{sc,aer}$	1	float	starting theta for aerosol	mrads
sigma_a_priori_aerosol_theta_MS	$\sigma_{\theta_{sc,ice}}/\theta_{sc,aer}$	1	float	The A priori fractional sd theta ratio for aerosol	1
sigma_C_lidar_rayleigh	δC_{R}	1	float	Assumed relative error in C_lid_ray	1
sigma_C_lidar_mie	δC_{M}	1	float	Assumed relative error in C_lid_mie	1
A-TC parameters					
tropopause_definition		1	character	“wmo” or “calypso”	
min_height_abo		1	float		m

ve_tropopause					
max_temp_psc		1	float	Maximum temperature for which PSCs are allowed	C
beta_cloud_ layer_threshold	$\beta_{thres,std}$	1	float	Default layers are seen as aerosols when average backscatter smaller than beta cld def	m-1 sr-1
beta_cloud_ layer_threshold _in_boundary_ layer	$\beta_{thres,bl}$	1	float	Boundary layers are seen as aerosols when average backscatter smaller than beta cld bl	m-1 sr-1
beta_cloud_ threshold_above _270K	$\beta_{thres,liq}$	1	float	Layers at temperatures above 270 K are seen as aerosols when average backscatter smaller than beta cld temp	1
beta_cloud_ threshold_in_ stratosphere	$\beta_{thres,strat}$	1	float	Stratospheric layers are seen as aerosols when average backscatter smaller than beta cld strat	1
minimum_PSC_ height		1	float	minimum height for which layers can be seen as Polar Stratospheric cloud	1
number_of_ temperature_ aerosol_ice_ seperation_ points	NT_{ice}	1	int	number of temperatures for ice/aerosol separation thresholds	1
number_of_ height_aerosol_ ice_seperation_ points	NH_{ice}	1	int	number of altitudes points for	1
temperature_ aerosol_ice_ seperation_ points	T_{ice}	5	float	min. temperature for which the respective threshold extinction holds	K
height_aerosol_ ice_seperation_ points	H_{ice}	2	float	height ratio (with respect to tropopause) for which the extinction threshold holds	1
extinction_ thresholds_for_ aerosol_ice_ seperation	TH_{ice}	2 x 5	float	threshold extinction for ice-aerosol separation as a function of the above height-temperature points	m-1
a_depolarizatio n_beta_ coefficient	$A_{\delta,\beta}$	1	float	A(dep,beta) value of the depolarization parameterization (Eq. 5.5.44)	1
b_depolarizatio n_beta_ coefficient	$B_{\delta,\beta}$	1	float	B(dep,beta) value of the depolarization parameterization (Eq. 5.5.44)	1
minimum_ probability_for_ ice_test		1	float	Min prob value to test for ice (Eq. 5.5.47)	1
supercooled_ water_lower_ temperature_ limit	T_{ice}^{thr}	1	float	Liquid layers with temperature lower than supercool_low are set to inconsistent liquid	1
supercooled_ water_upper_ temperature_ limit	T_{sc}^{thr}	1	float	Liquid layers with temperature lower than supercool_up are set to supercool	1
ice_wet_bulb_ upper_ temperature_ threshold		1	float	If the wet-bulb temperature of an ice layer is larger than wet_bulb_up ice is set to inconsistent ice	1
number_of_ aerosol_classes	N_{aer}	1	int	number of aerosol types	1
minimum_ aerosol_	P_{min}^{aer}	1	float	Minimum aerosol probability to be used in classification (exp(-9./2.))	1

probability_ threshold					
minimum_ aerosol_ probability_ single_type_ threshold	P_{min}^{first}	1	float	The minimum prob value for a cell to be defined as a single aerosol type (0.55)	1
minimum_ aerosol_ probability_ three_type_ limit	P_{min}^{sec}	1	float	There can only be three aerosol types when P1 lt min_P1 and P2 lt min_P2 (0.3)	1
Tropospheric Aerosol Type Description (7 times)					
name		1	string	Dust/Sea_Salt/Continental_pollution/Fine_mode_ Smoke/Dusty_smoke/Dusty_mix/Ice	1
shape		1	string	Assumed shape distribution used	1
lambda		1	float	Wavelength for which the aerosol parameters are valid	microns
linear_depolarization_ratio	δo	1	float	Central depolarization value	%
lidar_ratio	σδ	1	float	Central lidar ratio value	%
standard_deviation_linear_depolarization_ratio	So	1	float	Std deviation in the depol	%
standard_deviation_lidar_ratio	σS	1	float	Std deviation in the lidar ratio	%
angle	θ	1	float	Orientation of the 2D gaussian distribution with respect to depol axis (righthanded)	degrees
Stratospheric Aerosol Type Description (3 times)					
name		1	string	strat_aerosol_parameters1, strat_aerosol_parameters2 etc..	1
shape		1	string	Assumed shape distribution used	1
lambda		1	float	Wavelength for which the aerosol parameters are valid	microns
linear_depolarization_ratio	δo	1	float	Central depolarization value	%
lidar_ratio	σδ	1	float	Central lidar ratio value	%
standard_deviation_linear_depolarization_ratio	So	1	float	Std deviation in the depol	%
standard_deviation_lidar_ratio	σS	1	float	Std deviation in the lidar ratio	%
angle	θ	1	float	Orientation of the 2D Gaussian distribution with respect to depol axis (righthanded)	degrees
Stratospheric CloudType Description (3 times)					
name		1	string	strat_cloud_parameters1, etc..	1
shape		1	string	Assumed shape distribution used	1
lambda		1	float	Wavelength for which the aerosol parameters are valid	microns
linear_depolarization_ratio	δo	1	float	Central depolarization value	%
lidar_ratio	σδ	1	float	Central lidar ratio value	%
standard_deviation_linear_depolarization_ratio	So	1	float	Std deviation in the depol	%
standard_deviation_lidar_ratio	σS	1	float	Std deviation in the lidar ratio	%

angle	θ	1	float	Orientation of the 2D gaussian distribution with respect to depol axis (righthanded)	degrees
Ice properties					
C_00		1	float	Constant value from Heymsfield 2005	1
C_01		1	float	Constant value from Heymsfield 2005	1
C_10		1	float	Constant value from Heymsfield 2005	1
C_11		1	float	Constant value from Heymsfield 2005	1
ice_density	ρ_i	1	float	Ice density from Heymsfield 2005	g m-3

5.3 Output parameters

Described in [AD2], [AD3],[AD4],[AD5]

5.4 Algorithm flow charts

5.4.1 *A-AER*

5.4.1.1 Flowchart Ae1: Overall Algorithm Structure

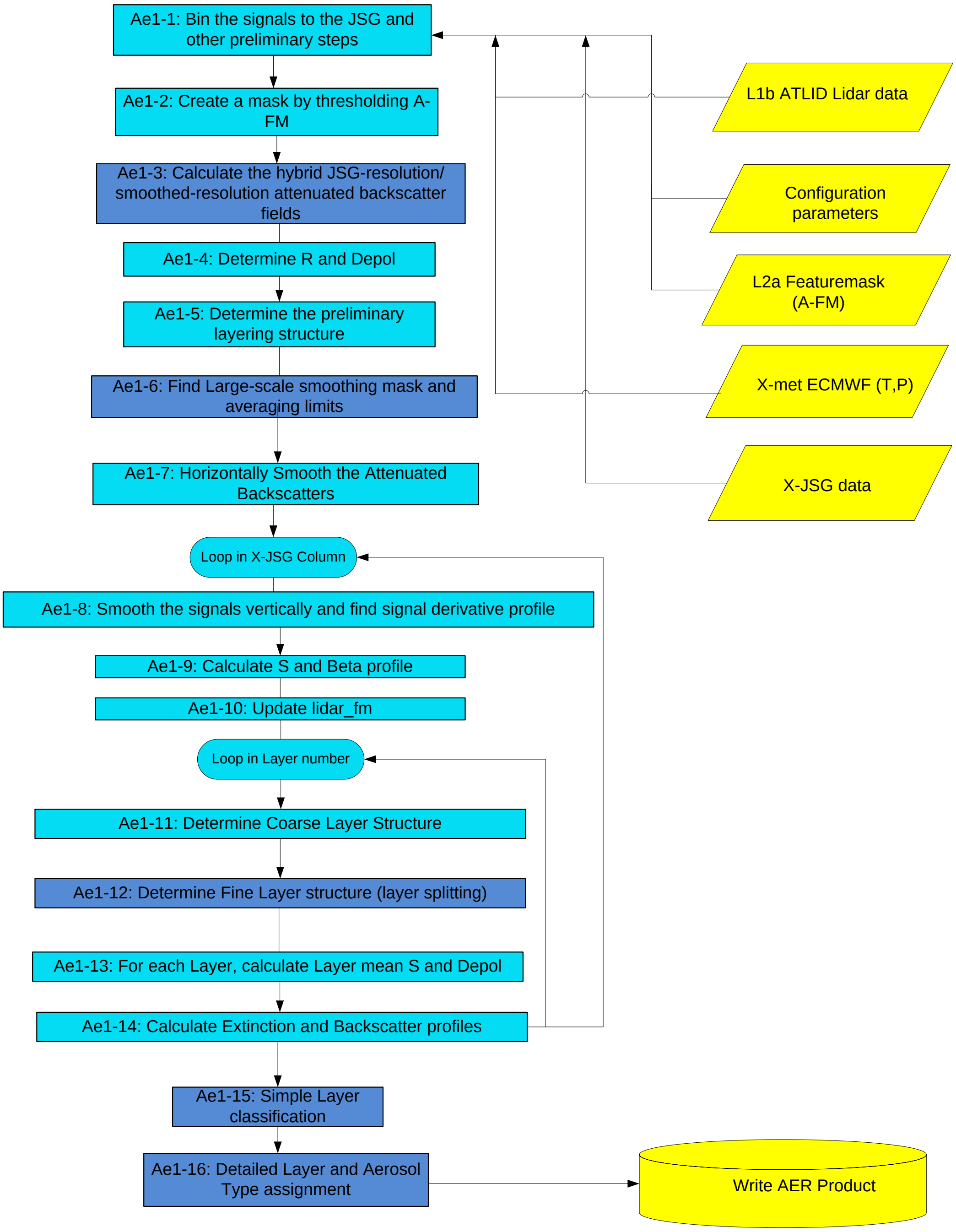


Figure 23: Schematic of the main components of the A-AER algorithm.

5.4.1.2 Preliminary smoothing procedure (Flowchart Ae1-3)

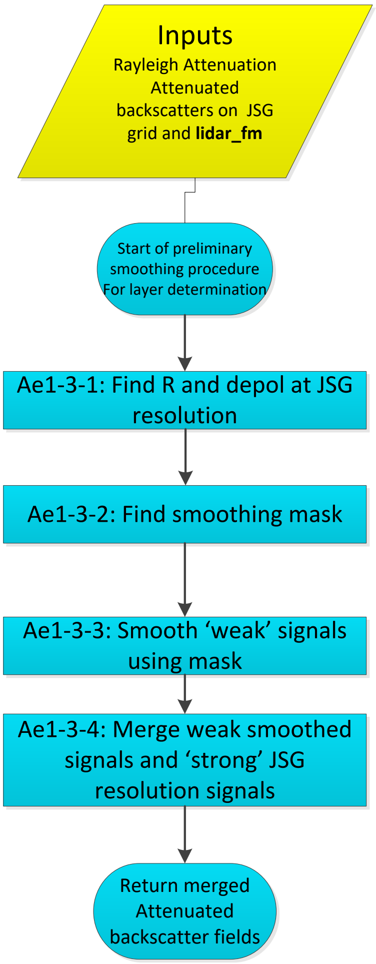


Figure 24: Preliminary smoothing procedure (Flowchart Ae1-3)

5.4.1.3 Averaging limits Determination (Flowchart Ae1-6)

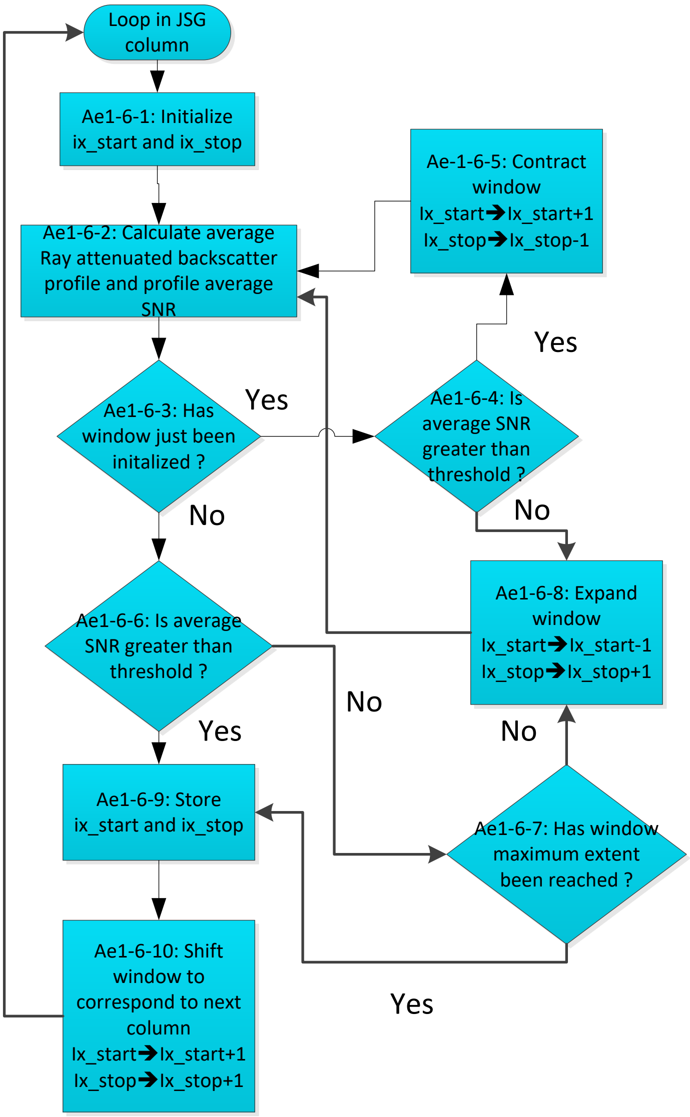


Figure 25: Flowchart corresponding to the averaging limit determination (Flowchart Ae1-6)

5.4.1.4 Layer Splitting Procedure (Flowchart Ae1-12)

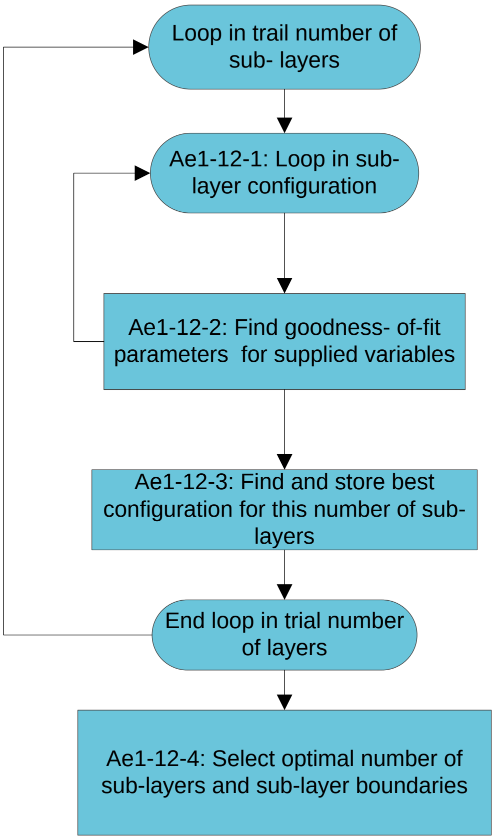


Figure 26: Flowchart of the Layer Splitting Procedure (Flowchart Ae1-12)

5.4.1.5 Simple Classification Procedure (Flowchart Ae1-15)

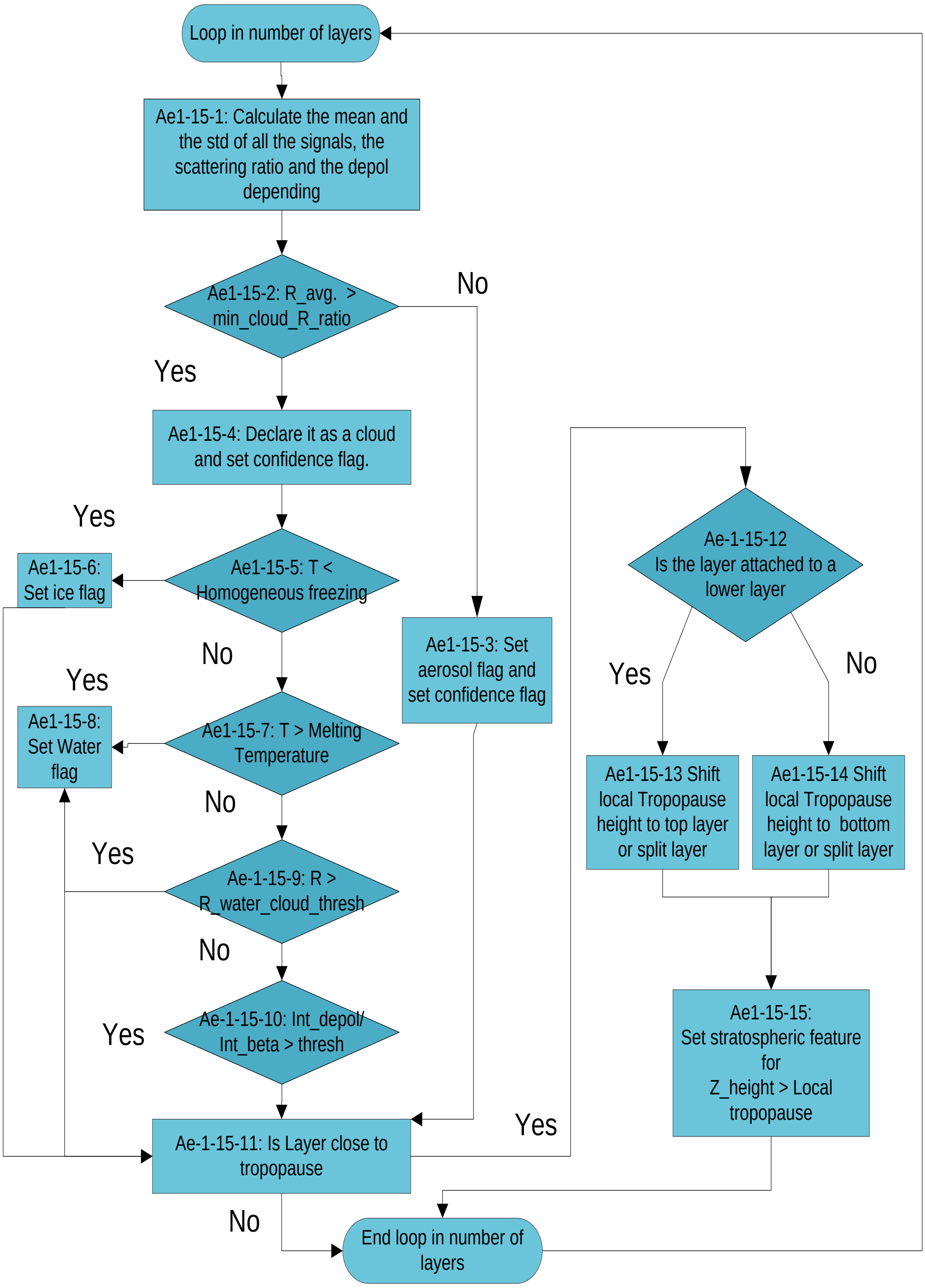


Figure 27: Simple layer classification procedure (Flowchart Ae1-15)

5.4.2 *A-EBD*

5.4.2.1 Overall Algorithm Structure: Flowchart Eb-1

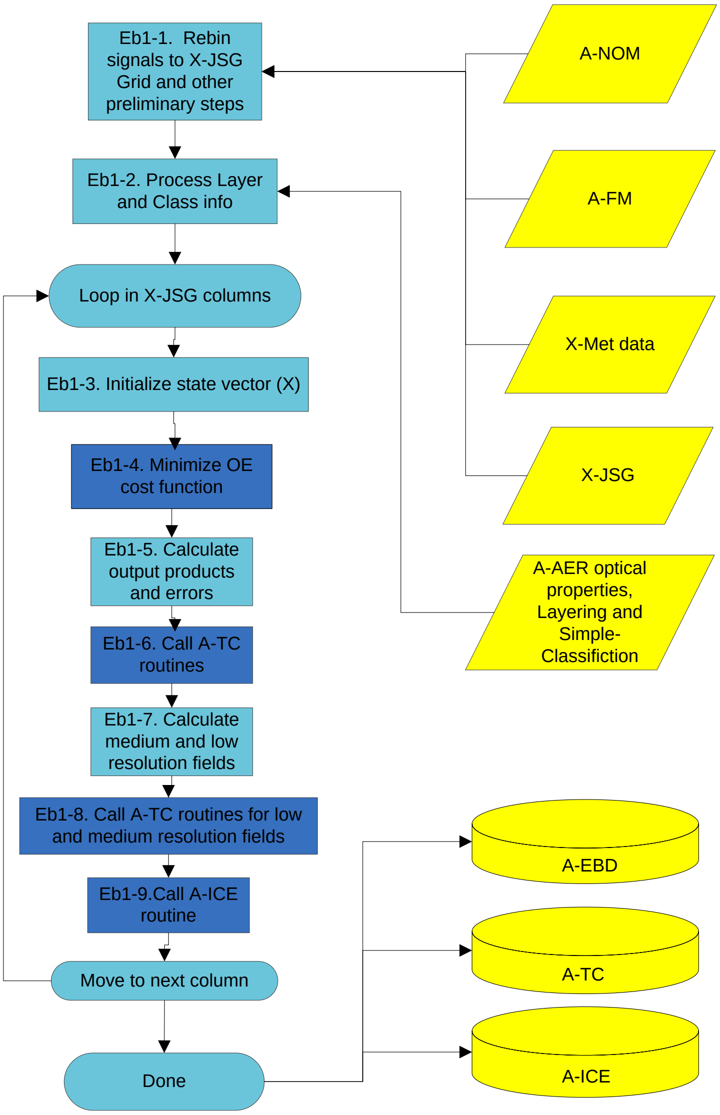


Figure 28: Main A-EBD structure (Flowchart Eb1).

5.4.2.2 Optimal Estimation Cost-function Calculation: Flowchart Eb1-4

The processes associated with the cost-function minimized in the **C5** process box in are shown in Figure 29

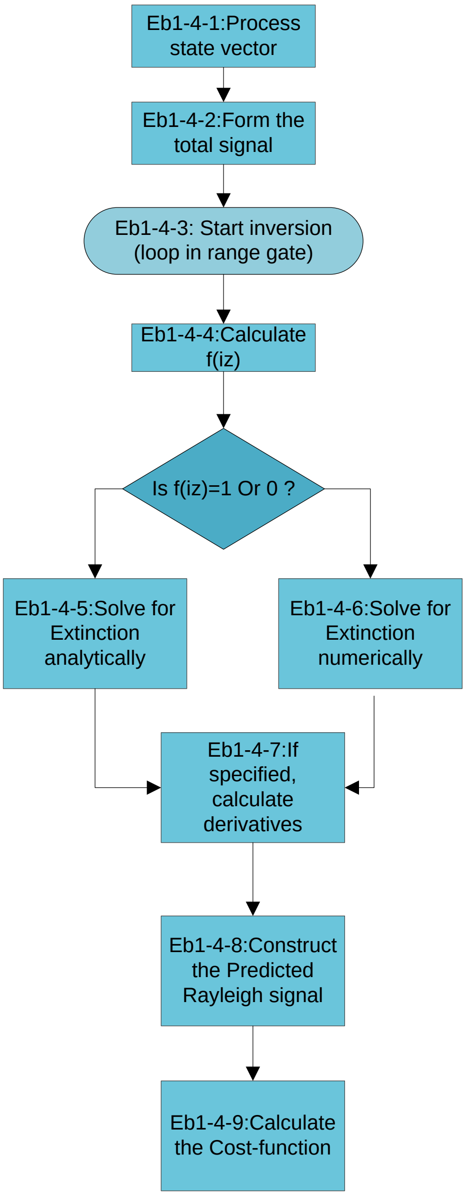


Figure 29: Optimal estimation cost-function calculation: Flowchart E1-4

5.4.3 A-TC

5.4.3.1 Detailed Lidar Classification Procedure: Flowchart TC1

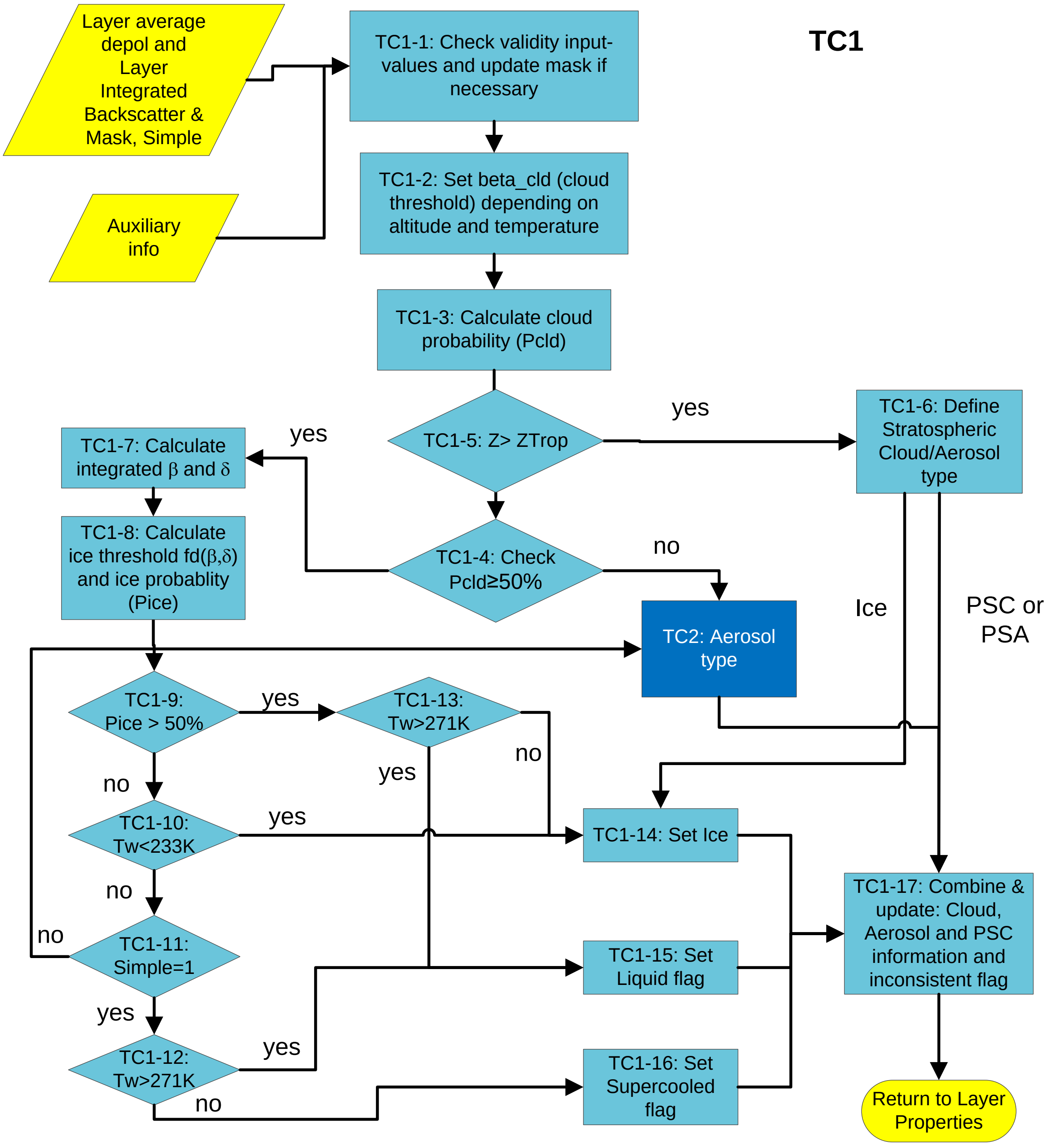


Figure 30: Aerosol cloud discrimination and water cloud/ice cloud discrimination procedure.

5.4.3.2 Aerosol Typing : Flowchart TC2

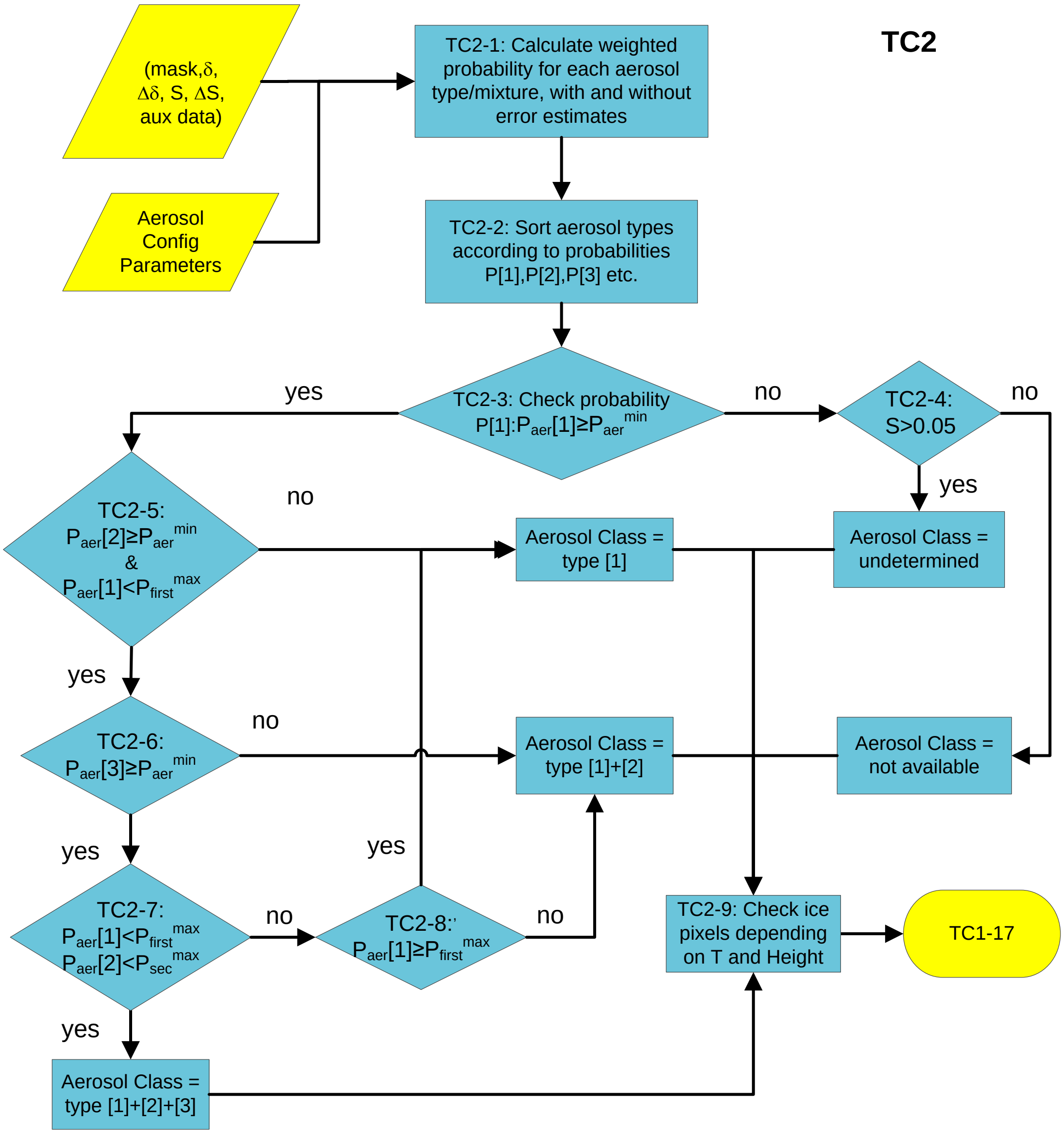


Figure 31: Aerosol Typing flow diagram (TC2). Note that the PSC and stratospheric aerosol classifications are calculated using the same routine.

5.4.3.3 Hybrid mode smoothing: Flowchart TC3

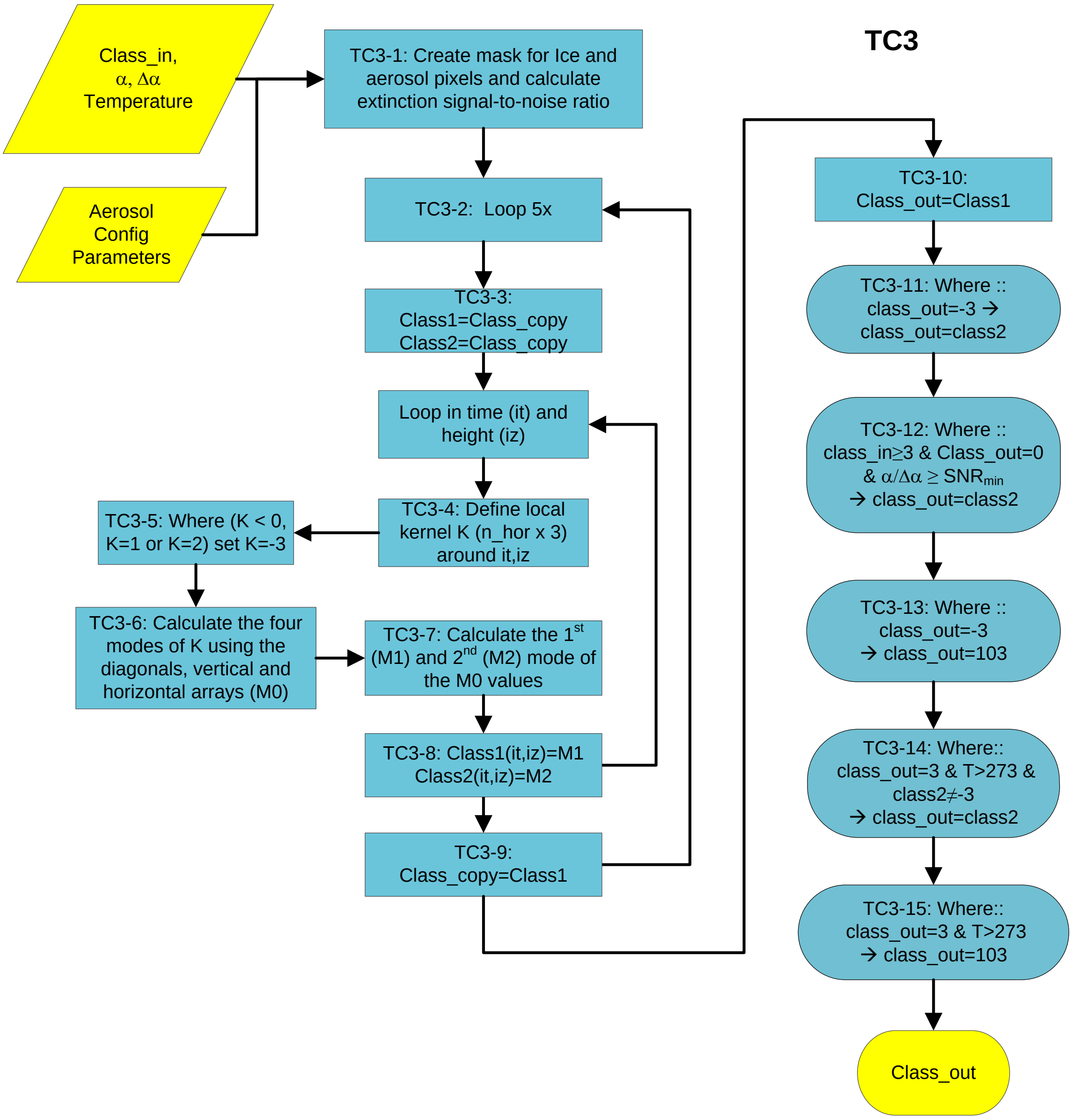


Figure 32: Hybrid mode smoothing of the aerosol and ice types

5.4.4 A-ICE: Flowchart AI1

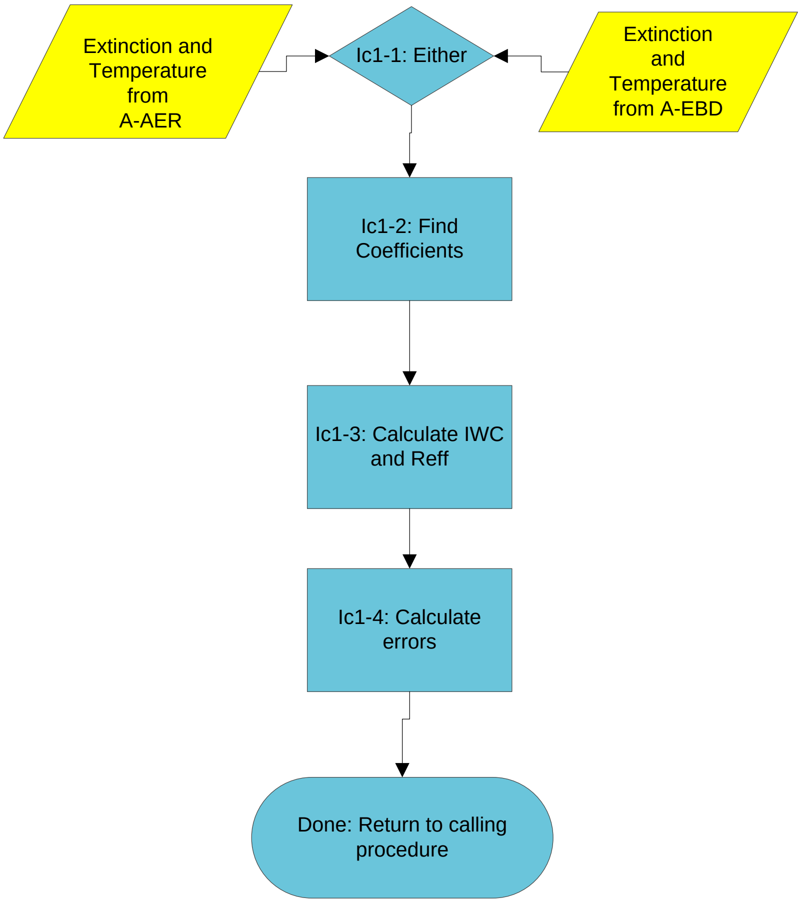


Figure 33: A-ICE algorithm flow-chart.

5.5 Algorithm definitions

5.5.1 A-AER

The main algorithm flowchart is shown in Figure 23: Schematic of the main components of the A-AER algorithm. . Here the Yellow Trapezoids represent input or output data sets. The different major algorithms steps are described below.

A1-1: Bin the signals to the JSG and other preliminary steps

The input met data as well as the signals are binned to the JSG coordinate system. For the signals, a simple linear averaging procedure is used and the standard-deviation of the signals is also calculated and stored. For each JSG column, the surface altitude is taken as the maximum surface altitude reported in the corresponding group of A-FM columns. For the A-FM featuremask the maximum value within the corresponding A-FM pixels is used.

The cross-polar attenuated Mie backscatter ($b_{M,\perp}(z)$) and the total Rayleigh attenuated backscatter ($b_R(z)$)are calculated following Eqns. (2.2.3) and (2.2.3). The (effective) depolarization ratio necessary for these calculations is itself estimated by dividing the sum of the first 100 highest in altitude elements of the time averaged feature-free cross-polar signal attenuated backscatter profile (the **cro_free_height** variable from the A-FM product) by the sum of the corresponding Rayleigh feature-free cross-polar signal attenuated backscatter profile (the **ray_free_height** variable from the A-FM product i.e.

$$\delta_R = \frac{\sum_{itop}^{itop-100} \text{cro\_free\_height(i)}}{\sum_{itop}^{itop-100} \text{ray\_free\_height(i)}} \tag{4.3.0}$$

At this stage the attenuated backscatter profiles are corrected for the effects of Rayleigh attenuation. That is, each of the Mie, Rayleigh and Cross-polar attenuated-backscatter profiles is divided by the two-way Rayleigh transmission (T^2_{Ray})

$$T_{Ray,i} = exp\left[-\sum_{i_z}^{i_{toa}} \sigma_{Ray} N_{a,i} \Delta r_i\right], \tag{4.3.0}$$

where i_{toa} is the range index corresponding to the top-of-atmosphere, σ_{Ray} is the Rayleigh scattering extinction cross-section per molecule, $N_{a,i}$ is the atmospheric number density and Δr_i is the range-gate width. Accordingly, the Rayleigh transmission corrected attenuated backscatter at the JSG grid resolution (along with their standard deviations) are generated, namely,

$$\mathrm{B}_{M,jsg}, \mathrm{B}_{R,jsg}, \mathrm{B}_{\perp,jsg}$$

and their corresponding standard deviation fields

$$\delta\mathrm{B}_{M,jsg}, \delta\mathrm{B}_{R,jsg}, \delta\mathrm{B}_{\perp,jsg}$$

Note: If only one native range-time pixel falls within a given JSG pixel then the standard deviation cannot be calculated and the uncertainty estimate from the L1 data is used instead.

Finally, the Featuremask data and lidar signals for X-JSG columns where no lidar data is present are filled with a "no-data_fill" value. This will aids in dealing with data gaps in a consistent manner.

A1-2: Create a mask by thresholding A-FM

A target mask (**lidar_fm**) is generated by applying a user-defined threshold to the re-gridded A-FM Featuremask data.

lidar_fm =1 ➔X-JSG Feature mask > Threshold ➔Target present.
lidar_fm=0➔X-JSG Feature mask < Threshold ➔ Clear atmosphere.
lidar_fm=-1➔X-JSG Feature mask reports fully attenuated ➔ No signal.
lidar_fm=-2➔X-JSG Feature mask reports (below-)surface.

In addition , if the separation between the ground and the first good target pixel is less than the **maximum_altitude_to_fill_with_aerosol** configuration parameter then **lidar_fm** is set to 1.

Ae1-3: Smooth the attenuated backscatter fields (for layering determination)

This procedure produces hybrid smoothed fields of the Rayleigh attenuation corrected attenuated backscatter fields ($\mathrm{B}_{M,h}, \mathrm{B}_{R,h}, \mathrm{B}_{\perp,h}$) along with their respective standard deviations. This is done in a multistep procedure in order to avoid smoothing over strong and weak features (which would result in e.g. Mixing cloud and aerosol returns). See flowchart Ae1-3 (Figure 24). ` Weak' features are smoothed using a box-car with dimensions of **preliminary_smoothing_horizontial_window** by **preliminary_smoothing_vertical_window** pixels, while `strong' features are unsmoothed.

Ae1-4: Determine R and Depol

The backscatter ratio corresponding to the fields found in the previous step are calculated as :

$$R_h = 1.0 + \frac{\mathrm{B}_{M,h} + \mathrm{B}_{\perp,h}}{\mathrm{B}_{R,h}} \tag{4.3.0}$$

and the corresponding estimate of the standard deviation of as

$$\delta R_h = \left(R_h - 1.0\right)\sqrt{\left(\frac{\delta \mathrm{B}_{M,h}}{\mathrm{B}_{M,h}}\right)^2 + \left(\frac{\delta \mathrm{B}_{\perp,h}}{\mathrm{B}_{\perp,h}}\right)^2} \tag{4.3.0}.$$

The depolarization ratio is calculated using

$$D_h = \frac{\mathrm{B}_{\perp,h}}{\mathrm{B}_{M,h}} \tag{4.3.0}$$

and the corresponding uncertainty as

$$\delta D_h = D_h\sqrt{\left(\left(\frac{\delta \mathrm{B}_{M,h}}{\mathrm{B}_{M,h}}\right)^2 + \left(\frac{\delta \mathrm{B}_{\perp,h}}{\mathrm{B}_{\perp,h}}\right)^2\right)} \tag{4.3.0}$$

Ae1-5: Find crude layer structure

At his point the preliminary layering structure mainly based on the **lidar_fm** is determined. Here**, lidar_fm** is used in a column-by-column manner to find discrete layer tops and bottoms.

For a given column layer boundaries are assigned to height index **iz** if any if of the following conditions are met:

1. **(lidar_fm(iz)**=0 **and lidar_fm(iz+1)**=1) or (**lidar_fm(iz)**=1 and **lidar_fm(iz+1)**=0)
2. ((abs(**FM(iz)-FM(iz-1)**).ge. **feature_mask_difference_layer_threshold**).and.(**lidar_fm(iz).**ge.1).and.(**lidar_fm(iz-1).**ge.1))). Here **FM_jsg** is the featuremask (A-FM) data rebinned to the JSG grid from Step Ae1-1.

The layering structure calculated here is not directly used in later steps (however the number of layers is) but is calculated for here mainly for debugging purposes.

Ae1-6: Find Large-scale smoothing mask and averaging limits

A mask representing data that is valid to average horizontally is built by applying a threshold (**smooth_R_cutoff_for_horizontial_smoothing**) to the **R** field generated in Step Ae1-4. Values above this threshold are assumed to be associated with strongly attenuating targets as are pixels associated with high values of the feature mask variable **FM_jsg**.

Accordingly, the smoothing mask is true when the following condition is met (and set to **false** otherwise):

acc_data_mask=**true** if (**lidar_fm**.ge.0).and.(R_h .lt.**smooth_R_cutoff_for_horizontial_smoothing**- δR_h).and.(**FeatureMask_jsg**.lt.8)

Averaging attenuated regions (e.g. below significant clouds) together with unattenuated regions would lead to inaccurate extinction and backscatter profiles. In order to avoid this, the **acc_data_mask** variable set to **false** for that all altitudes in each respective column that are below the highest **false** occurrence.

In addition, single isolated true pixel values (i.e. a true value completely surrounded by false values) are set to false.

The height-and-time dependent horizontal averaging limits are then found. This is depicted in Flowchart Ae1-6-1 and described in Section

Ae1-7: Smooth the Attenuated Backscatters

Here the mask and column-and-height dependent window limits calculated in Step Ae1-7 are used to average the Rayleigh transmission corrected attenuated backscatters horizontally. The associated standard deviations of the average quantities are also calculated using standard error propagation techniques. For example, for the Mie Rayleigh transmission corrected attenuated backscatter we have:

$$\mathrm{B}_{M,hsm} = \frac{1}{N}\sum_{i=i_{jsg}-iw/2}^{i=i_{jsg}+iw/2} \mathrm{B}_{M,jsg}(i,j_z)[mask=.true.] \qquad (4.3.1)$$

and

$$\delta\mathrm{B}_{M,hsm} = \frac{1}{N^{1/2}}\left(\sum_{i=i_{jsg}-iw/2}^{i=i_{jsg}+iw/2}\left[\frac{\left(\mathrm{B}_{M,jsg}(i,j_z)-\mathrm{B}_{M,hsm}\right)}{N}\right]^2[mask=.true.]\right)^{1/2} \qquad (4.3.1).$$

where *iw* is the width in jsg pixels of the altitude dependent horizontal smoothing window and **N** is the number of **true** mask pixels (here mask referees specifically to the **acc_data_mask** variable**)** within the averaging interval.

Similar expressions for the horizontally smoothed Rayleigh and cross-polar fields are used.

Ae1-8: Smooth the signals vertically and find signal derivative profile

Here the horizontally smoothed signals from Step Ae1-7 are smoothed vertically using a sliding linear least-squares fitting procedure at the same-time the standard deviation profile of the smoothed attenuated backscatters along with the vertical signal derivatives and associated uncertainty profiles. The width of the sliding window is controlled by the **vertical_derivative_determination_window** configuration parameter. Near the ground (and the top of the profile) the window full-width is reduced to a minimum of 3 points. For the remaining points the slope are assumed to be constant and the end points are found by extrapolation.

The vertically and horizontally smoothed signals returned are denoted as $\mathrm{B}_{M,hvsm}, \mathrm{B}_{R,hvsm}, \mathrm{B}_{\perp,hvsm}$.

Ae1-9: Calculate S and Beta profile

In subsequent steps the algorithm will separate and classify layers using the Extinction-to-backscatter ratio (along with the scattering ratio and the Depolarization ratio). S is also a primary input to the A_EBD algorithm. Thus here we will focus on accurate retrievals of S.

Here S is calculated directly from the vertically and horizontally smoothed signals in one step (as opposed to calculating the extinction and backscatter separately and then dividing to obtain S). Eqs. (2.2.4) thru (2.2.7) can be re-arranged to show that

$$S(z) = -0.5\frac{\frac{d\mathrm{B}_{R,hvsm}}{dr}}{\mathrm{B}_{M,hsvm} + \mathrm{B}_{\perp,hsvm}} \qquad (4.3.2).$$

Thus S can be retrieved rather directly avoiding issues with the division of the separately derived extinction and backscatter profiles. Along with the S profile the backscatter profile is derived i.e.:

$$\beta_M = \left(\frac{\mathrm{B}_{M,hvsm} + \mathrm{B}_{\perp,hvsm}}{\mathrm{B}_{R,hvsm}}\right)\beta_{Ray} \qquad (4.3.2)$$

$$\delta\beta_M = \beta_M\sqrt{\left(\frac{\delta\mathrm{B}_{R,hvsm}}{\mathrm{B}_{R,hvsm}}\right)^2 + \left(\frac{\delta^2\mathrm{B}_{M,hvsm} + \delta^2\mathrm{B}_{\perp,hsvm}}{\left(\mathrm{B}_{M,hvsm} + \mathrm{B}_{\perp,hsvm}\right)^2}\right)} \qquad (4.3.2)$$

where β_{Ray} is the molecular Rayleigh backscatter coefficient.

Ae1-10: Update lidar_fm

This (optional) processing step attempts to `fill-in' regions missed by A-FM and eliminate false-positive regions.

If the configuration parameter **adjust_input_feature_mask** is set to one then:

If (**lidar_fm**.eq.0).and ($\mathrm{B}_{M,hvsm}$ .gt.4 $4\delta\mathrm{B}_{M,hvsm}$).and.(**acc_data_mask**.eq.true).and.($\mathrm{B}_{M,hvsm}$ .gt.5e-5).and.

($\mathrm{B}_{M,hvsm} + \mathrm{B}_{\perp,hvsm}$ .ne.0) then

lidar_fm=2 (a 2 is used here to distinguish new extended areas of the featuremask).

Further:

If (**lidar_fm**.ge.1).and.($S\beta$).le.2.0e-6).and.(**acc_data_mask**.ge.1) then
	lidar_fm=0.

where β is the backscatter derived in Step Ae-10 and S is the corresponding lidar ratio.

Ae1-11: Determine coarse layer structure

Here the coarse layer structure is determined.

In this Step the **R** values and corresponding error estimates used are those calculated in Step Ae1-4 (since they are relatively high resolution compared to the backscatter ratio fields calculated using the more heavily smoothed fields used in subsequent steps).

For a given column layer boundaries are assigned to height index **iz** if any of the following conditions are met:

1. **(lidar_fm(iz)**=0 **and lidar_fm(iz+1)**=1) or (**lidar_fm(iz)**=1 and **lidar_fm(iz+1)**=0)
2. ((abs(**FM(iz)-FM(iz-1)**).ge. **feature_mask_difference_layer_threshold**).and.(**lidar_fm(iz).**ge.1).and.(**lidar_fm(iz-1).**ge.1))). Here **FM_jsg** is the featuremask (A-FM) data rebinned to the JSG grid from Step Ae1-10.
3. ($\left(R(iz)+3\delta R(iz)\right)$ .le. **water_cloud_R_threshold**)).and.($\left(R(iz-1)+3\delta R(iz-1)\right)$).gt. **water_cloud_R_threshold**)))
4. ($\left(R(iz)+3\delta R(iz)\right)$ .ge. **water_cloud_R_threshold**).and.($\left(R(iz-1)+3\delta R(iz-1)\right)$ lt. **water_cloud_R_threshold**)))
5. $abs\left(abs(\Delta R(iz))-abs(\Delta R(iz+1))\right)-3\sqrt{\delta\Delta R(iz)^2+\delta\Delta R(iz+1)^2}$ .ge. **dR2_difference_layer_threshold** where:
   $\Delta R(iz)=R(iz)-R(iz-1)$, $\Delta R(iz+1)=R(iz+1)-R(iz)$,
   $\delta\Delta R(iz)=\sqrt{\delta R(iz)^2+\delta R(iz-1)^2}$ and
   $\delta\Delta R(iz+1)=\sqrt{\delta R(iz+1)^2+\delta R(iz)^2}$.

Steps 1 and 2 produce a layer when gaps in the thresholded featuremask exist or a large jump in the JSG gridded featuremask index exists. Step 3 and 4 produce layer boundaries when a strong (e.g. water cloud) feature is exited or entered. Step 5 aims to assign a layer boundary when the absolute value of the change in R exceeds a given threshold (e.g. near the top and bottom of layers but not around layer peaks).

Ae1-12: Determine Fine Layer structure (Layer splitting).

This step uses the smoothed signals generated in Step Ae1-3 along with the S ratio found in Step Ae1-9.
This process was briefly described earlier in Section 3.1.5.1 and is described by flowchart Ae1-12 (Figure 26)

At the end of this processing step the layering structure is updated by effectively now treating each sub-layer identified in the previous step as its own layer.

Ae1-13: For each Layer, calculate Layer mean S and Depol
The layer average values of S and the depolarization ratio are calculated along with their respective uncertainties.

$$S_{lay}(i_{bot}:i_{top}) = \frac{1}{\left(i_{top}-i_{bot}+1\right)}\sum_{i_{bot}}^{i_{top}}\left[\frac{\dfrac{d\mathrm{B}_{R,hvsm}(i)}{dr}}{\mathrm{B}_{M,hvsm}(i)+\mathrm{B}_{\perp,hvsm}(i)}\right] \tag{4.3.3}$$

$$\delta S_{lay}(i_{bot}:i_{top}) = \left(\frac{S_{lay}}{N_{eff}^{1/2}}\right)\left[\frac{\sum\delta^2\left(\dfrac{d\mathrm{B}_{R,hvsm}(i)}{dr}\right)}{\left(\sum\left(\dfrac{d\mathrm{B}_{R,hvsm}(i)}{dr}\right)\right)^2}+\frac{\sum\left(\delta^2\mathrm{B}_{M,hvsm}(i)+\delta^2\mathrm{B}_{\perp,hvsm}(i)\right)}{\left(\sum\left(\mathrm{B}_{M,hvsm}(i)+\mathrm{B}_{\perp,hvsm}(i)\right)\right)^2}\right]^{1/2} \tag{4.3.4}$$

$$D_{lay}(i_{bot}:i_{top}) = \left(\frac{\sum\mathrm{B}_{\perp,hvsm}(i)}{\sum\mathrm{B}_{M,hvsm}(i)}\right) \tag{4.3.5}$$

$$\delta D_{lay}(i_{bot}:i_{top}) = \left(\frac{D_{lay}}{\sqrt{N_{eff}}}\right)\sqrt{\frac{\sum\delta^2\mathrm{B}_{\perp,hvsm}}{\left(\sum\mathrm{B}_{\perp,hvsm}\right)^2}+\frac{\sum\delta^2\mathrm{B}_{M,hvsm}}{\left(\sum\mathrm{B}_{M,hvsm}\right)^2}} \tag{4.3.5}$$

and, in order to (approximately) account for the correlated nature of the vertically data in the error calculations the effective number of samples N_{eff} is introduced:

$$N_{eff} = \frac{i_{top}-i_{bot}}{\mathrm{int}(\mathbf{vertical\_derivative\_determination\_window}/\Delta R)}, \tag{4.3.5}$$

where Δr is the range resolution.

Ae1-14: Calculate Extinction profiles

To insure consistency between the extinction, backscatter and S fields the backscatter (from set A8) is used together with the layer averaged S values to compute the extinction profiles. i.e.

$$\alpha_M = \beta_M S_{lay} \tag{4.3.6}$$

$$\delta\alpha_M = \alpha_M\left(\left(\frac{\delta S_{lay}}{S_{lay}}\right)^2+\left(\frac{\delta\beta}{\beta}\right)^2\right)^{1/2} \tag{4.3.6}$$

Ae1-15: Simple Layer classification

A simple classification system is applied to the layers. The primary use of this classification is to separate cloud and aerosol regions and acts as a backup/reference for the more advanced layer classification and aerosol typing procedure. This procedure is detailed in flow-chart Ae1-15 (Figure 27).

A14: Detailed Layer and Aerosol Type assignment

This procedure is associated with the A-TC algorithm and is described in Sections 5.4.3.

5.5.1.1 Ae1-3: Smooth the attenuated backscatter fields (for layering determination)

Ae1-3-1: Find R and Depol. at JSG resolution

The scattering and the depolarization ratio are calculated at the JSG resolution (R_{jsg}, D_{jsg}). The corresponding errors are also calculated. The equations used are the same as Eqns. (4.3.0)-(4.3.0) except that the native resolution JSG signals are used.

Ae1-3-2: Find smoothing mask

In preparation for smoothing, `strong' features are masked.

Accordingly **mask_sm** is defined as being .true. if:

(**lidar_fm.ge**.0).and.
(**FM_jsg**.lt.**feature_mask_cutoff_for_horizontal_smoothing**).and.
(R_{jsg} .lt. **smooth_R_cutoff_for_horizontial_smoothing** - δR_{jsg})

Ae1-3-3: Smooth `weak' signals using mask

Using the mask defined in the previous step the `preliminary smoothed signal fields' $(\mathrm{B}_{M,sm}, \mathrm{B}_{R,sm}, \mathrm{B}_{\perp,sm})$ are found along with their standard deviations. A simple box-car averaging technique is used.

Ae1-3-4: Merge `weak' and `strong' fields

The hybrid smoothed fields and their corresponding standard deviation fields are found e.g.

$\mathrm{B}_{M,hsm} = \mathrm{B}_{M,sm}$ if **mask_sm=.true.,**
$\mathrm{B}_{M,hsm} = \mathrm{B}_{M,jsg}$ if **mask_sm=.false.,**

with similar rules for the other attenuated backscatter fields and associated standard deviation fields.

5.5.1.2 Ae-1-6: Averaging limits Determination

Ae1-6-1: Initialize ix_start and ix_stop
For each column an initial window (represented by **ix_start** and **ix_stop**) is defined.

Ae1-6-2: Calculate average Rayleigh attenuated backscatter profile and profile average SNR
The average Rayleigh attenuated backscatter is calculated and the profile average SNR ratio is calculated. To increase the execution efficiency a buffer approach is used to enable rapid calculation of the means profile and associated standard deviation. That is, instead of averaging the profiles from **ix_start** to **ix_stop** upon each invocation of this step, the profile from the previous call is stored and only the differences from the previous window boundaries are either added to or subtracted from the previous result.

Ae1-6-3: Has the window been initialized ?
If the window boundaries have just been initialized then then if the average SNR is greater than **horizontial_smoothing_target_SNR** it is possible that window is larger than required (**Ae1-6-4**). In this case the window is repeatedly contacted (**Ae1-6-5**).

Ae1-6-4: If the window limits have not just been initialized then the window is repeatedly expanded (**Ae1-6-8**) so that the average SNR is greater than or equal to tSNR or the maximum allowed window extent (**maximum_horizontial_smoothing_window**) is reached **(Ae1-6-7).**

Ae1-6-9: When the suitable window limits are found then the limits are stored and the window boundaries are both shifted correspond to the next column (**Ae-6-10**).

5.5.1.3 Ae1-12: Layer Splitting Procedure

Ae1-12-1: Loop in trial number of sub-layers
The number of trial layers starts with one and goes to a set maximum or the number of height bins in the layer, whichever is smaller

Ae1-12-2: Loop in sub-layer configuration
For the specified number of layers the routine loops through all possible layering possibilities. For example, for two layers one layer would span from $iz_1 = 1, nz$ while the second layer will span from $iz_2 = iz_1, nz$. For 3 layers then we have $iz_3 = iz_2, nz$ etc. A maximum of 4 sub-levels is considered

Ae1-12-3: Find Goodness of fit parameters
For each layer the Goodness-of-fit parameter is corresponding to S,D, and R are calculated and the values are summed across the layers. For uncorrelated quantities, for example the depolarization ratio, for a single layer we have

$$\chi^2_{red,\delta} = \frac{1}{nz-1-nl}\left[\sum_{i=iz_{bot}}^{iz_{top}}\left(\frac{\delta_i-\overline{\delta}_1}{\sigma_{\delta_i}}\right)^2\right] \qquad (4.3.7)$$

where

$$\overline{\delta} = \frac{\sum_{i=iz_{bot}}^{i=iz_{top}} \frac{\delta_i}{\sigma_{\delta_i}}}{\sum_{i=iz_{bot}}^{iz_{top}} \delta_{x_i}^{-1}}. \tag{4.3.8}$$

For quantities, such as the layer extinction-to-backscatter ratios, the correlation in the supplied extinction profiles must be accounted for,. i.e. The layer Extinction-to-Backscatter ratio is given by

$$S_l = \frac{OT_l}{I\beta_l} \tag{4.3.9}$$

where α is the extinction coefficient and β is the backscatter coefficient

$$OT_l = \sum_{i=z_{bot}}^{z_{top}} \alpha_i \Delta z_i \tag{4.3.10}$$

and

$$I\beta_l = \sum_{i=z_{bot}}^{z_{top}} \beta_i \Delta z_i \tag{4.3.11}$$

The extinction errors will be correlated in altitude due to the vertical smoothing applied at Step **A7**. The effect of this correlation is (approximately) accounted for by adjusting the uncorrelated OT error estimate by a factor related to the effective-number-of samples Eq.(4.3.5).

$$\sigma_{OT_l}^2 = \left(\frac{1}{N_{eff}}\right) \sum_{i=z_{bot}}^{z_{top}} \sigma_{\alpha}^2(z) \Delta z_i \tag{4.3.12}$$

and (assuming that the error in the integrated the backscatter is uncorrelated with the extinction error) we will have

$$\sigma_{S_l}^2 = S_l^{\,2} \left(\left(\frac{\sigma_{OT_l}^2}{OT_l} \right) + \left(\frac{\sigma_{I\beta_l}^2}{I\beta_l} \right) \right) \tag{4.3.13}$$

For each supplied quantity the layer integrated quantities and their error estimates are stored for use in later steps

E4: Find and store best configuration

For each number of trial layers the layering configuration with the best total summed (across R, S and D) goodness-of-figure is identified and stored in an array. The values and error estimates for the layer integrated backscatter, depolarization etc. corresponding to the best configuration are also stored in an array.

E5: Select optimal number of layers

The array of best goodness-of-fit values is normalized by its smallest value (which almost always will be the value associated with the highest number of levels. The optimum number of layers is chosen as follows:

- If increasing the number of sub-layers decreases the summed goodness-of-fit figure by a factor of **layer_splitting_chisq_relative_step_stop** then this number of sub-layer is selected. Otherwise…
- If increasing the number of sub-layers decreases the summed goodness-of-fit figure to within **layer_splting_chisq_relative_min_stop** then this number of layers is selected.

5.5.1.4 Simple Classification Procedure

Ae1-15-1: Calculate the mean and the std of all the signals etc..
Per-layer layer-mean backscatter ratio, backscatter, depolarization values and their associated errors are calculated.

Ae1-15-2: R_avg. > min_cloud_R_ratio
If the mean scattering ration is below the **minimum_cloud_R_threshold** configuration parameter then the layer is classified as aerosol (**Ae-15-3**) and an associated confidence flag is set:
1➔mean R+ estimated error < **minimum_cloud_R_threshold**
2➔ mean R+ 2 x estimated error < **minimum_cloud_R_threshold**
3➔ mean R+ 3 x estimated error < **minimum_cloud_R_threshold**

Ae1-15-4: Declare it as a cloud and set confidence flag.
If the mean scattering ratio is above the **min_cloud_R_threshold** configuration parameter then the layer is classified as cloud (**F4**). An associated confidence flag is also set similar to step **Ae1-15-3**:
1➔mean R- estimated error > **minimum_cloud_R_threshold**
2➔ mean R- 2 x estimated error > **minimum_cloud_R_threshold**
3➔ mean R- 3 x estimated error > **minimum_cloud_R_threshold**

Ae1-15-5: T < Homogeneous freezing
If the layer mid-point temperature is below **homogeneous_freezing_temperature** then the cloud is classified as ice (**Ae1-15-6**).

Ae1-15-7: T > Melting Temperature
If the layer mid-point temperature is above **ice_water_seperation_temperature** then the cloud is classified as ice (**Ae1-15-8**).

Ae1-15-9: R > R_water_cloud_thresh
If configuration parameter flag **force_class_to_water** has been set to 1 and the scattering ratio is above **water_cloud_R_threshold** then the water cloud flag is set.

Ae1-15-10: Int_depol/Int_beta > thresh
If the ratio of the integrated depolarization-to-integrated backscatter is above **integrated_depolarization_to_integrated_beta_ratio_threshold** then the ice flag is set along with an associated confidence flag.

Ae1-15-11: Layer_lower_height< Tropopause height < Layer_upper_height
If the local tropopause lies within the layer boundaries a check is made to shift the local tropopause or split the layer into two layers with a boundary at the input tropopause. If not the next layer will be assessed

Ae1-15-12: A check is made to see if the layer is attached to a lower layer.
If it is attached the layer most likely belongs to an overshoot from a tropospheric feature:
Ae1-15-13 : If the top of the layer is within **tropo_dheight** of the tropouase height, the tropopause height is set to the top of the layer, if the difference is greater than **tropo_dheight** the layer is split into two layers with a break at the input tropopause height.
If it is not attached to the layer below it belongs most likely to a stratospheric feature moving into the troposphere:
Ae1-15-14: If the bottom of the layer is within **tropo_dheight** of the tropouase height, the tropopause height is set to the bottom of the layer, if the difference is greater than **tropo_dheight** the layer is split into two layers with a break at the input tropopause height.

Ae1-15-15: For all classified layers with a height > Tropopause height the simple classification is set to tropospheric feature

5.5.2 A-EBD

The overall high-level structure of the main algorithm is shown in Flowchart Eb1 (Figure 28). Here the Yellow trapezoids represent input or output data sets. Within the EBD algorithm, unless specifically noted, the attenuated backscatter signals used correspond to the JSG resolution variants.

Eb1-1: Rebin signals and other preliminary steps
Similar to step Ae1-1 of A-AER the attenuated backscatter signals are read in a gridded to the X-JSG and the signals are corrected for the effects of Rayleigh attenuation etc.

Eb1-2: Process Layers and Class info
The layering variables and simple classification variables generated by A-AER are read in.

Eb1-3: Initialize state vector
The a priori state vector ($\mathbf{x_a}$) described in sections 3.1.4.6 and 3.1.4.7 are defined here. This is done on a layer-by-layer basis. If retrieved **S** values (and uncertainties) are available from the A-AER product they are used, otherwise, the a-priori ice cloud, water cloud and aerosol values as specified in the processor configuration file are used. Which values are assigned to each layer is based upon the **classification** product from A-AER. If the type from A-AER is unknown then the assignment is done using the A-AER **simple_classification** variable.

As well as the a priori state vector, the initial state-vector is defined using the appropriate `starting' values specified in the configuration file are used. This enables the minimization starting point to be different from the a priori state. This can be used to investigate e.g. the sensitivity of the retrieved state-vector to the minimizations starting point independent of the a priori state.

Eb1-4: Minimize OE cost function
The minimization routine which finds the minimum of the cost-function is called. The procedure uses a Broyden-Fletcher-Goldfarb-Shanno variant of the Davidon-Fletcher-Powell method which uses a numerical Gradient approach. The routine returns the value of the minimum of the cost function, the state-vector corresponding to this minimum and an approximation to the Hessian matrix of the cost-function at the returned minimum. The Hessian is defined as

$$\mathbf{H} = \left| \frac{\partial^2 \chi^2}{\partial x_i \partial x_j} \right| \qquad (4.3.14)$$

and is related to the curvature matrix, which in turn is related to the inversion error-covariance matrix (see Eq.(2.2.52)).

The Cost-function routine is diagrammed in Section 5.4.2.2 with more information is given in Section 5.5.2.1.

In order for the minimization routine to return a useful approximation of the Hessian it is necessary to supply an initial guess for the inverse Hessian matrix ($\mathbf{H_o}$). In this work we use

$$\mathbf{H_o} = \frac{0.1}{\left\| \nabla \chi^2 \right\|_o} \mathbf{I} \qquad (4.3.15)$$

i.e. we use the identify matrix scaled by a factor appropriate to the scale of the problem [Nocedal and Wright (2006)] namely the inverse of one tenth of the length of the gradient of the cost function at the first iteration step.

The configuration parameter **maximum_number_of_iterations** controls the maximum number of iterations allowed in the minimization procedure. **step_factor_used_in_line_search** helps set the maximum allowed step size in along any one dimension in the line-search step of the minimization procedure. **gradient_convergent_tolerance** and **step_convergence_tolerance** set the gradient tolerance for convergence and the tolerance related to the increment between two full iterations respectively. If the cost function at the first iteration step is above **SSE_limit (**and **SSE_limit** >0**)** then no minimization is attempted.

Eb1-5: Calculate output products and errors
The error estimates for the quantitative product variables are calculated by applying standard error propagation techniques using the appropriate derivatives and entries of the retrieval covariance matrix ($\mathbf{C}$).

First, following [Press et al.], the covariance matrix is approximated using the inverse of the approximation of the inverse Hessian returned by the BFGS routine i.e.

$$\mathbf{C} = \frac{1}{2}\mathbf{H} \qquad (4.3.15).$$

The error in the state-vector variables can be directly calculated as:

$$\delta C_M = \log(10)\mathbf{x}_{a,1}10^{\mathbf{x}_1}\sqrt{\mathbf{C}_{1,1}} \qquad (4.3.15)$$

$$\delta C_R = \log(10)\mathbf{x}_{a,2}10^{\mathbf{x}_2}\sqrt{\mathbf{C}_{2,2}} \qquad (4.3.15)$$

$$\delta f_{MS,p} = \log(10)\mathbf{x}_{a,3}10^{\mathbf{x}_2}\sqrt{\mathbf{C}_{3,3}} \qquad (4.3.15)$$

$$\delta\eta_{i_{lay}} = \log(10)\mathbf{x}_{a,4+3(i_{lay}-1)}10^{\mathbf{x}_{4+3(i_{lay}-1)}}\sqrt{\mathbf{C}_{4+3(i_{lay}-1),4+3(i_{lay}-1)}} \qquad (4.3.15)$$

$$\delta\theta_{sc},_{i_{lay}} = \log(10)\mathbf{x}_{a,5+3(i_{lay}-1)}10^{\mathbf{x}_{5+3(i_{lay}-1)}}\sqrt{\mathbf{C}_{5+3(i_{lay}-1),5+3(i_{lay}-1)}} \qquad (4.3.15)$$

$$\delta S_{i_{lay}} = \log(10)\mathbf{x}_{a,6+3(i_{lay}-1)}10^{\mathbf{x}_{6+3(i_{lay}-1)}}\sqrt{\mathbf{C}_{6+3(i_{lay}-1),6+3(i_{lay}-1)}}\,. \qquad (4.3.15)$$

The calculation of the error estimates associated with the extinction, backscatter and optical depth is more complicated and requires the propagation of the uncertainty due to the uncertainty in the retrieved state-vector components as well as the uncertainty estimates in the $\mathrm{B}_M, \mathrm{B}_\perp, \mathrm{B}_R$ profiles.

As a first set of steps, the Jacobians with respect to the state-vector are evaluated numerically, i.e.

$$\frac{\partial\alpha_i}{\partial x_j^f} \simeq \frac{\alpha_i(\mathbf{x}^f+\Delta_j)-\alpha_i(\mathbf{x}^f-\Delta_j)}{2\Delta} \qquad (4.3.16)$$

$$\frac{\partial\beta_i}{\partial x_j^f} \simeq \frac{\beta_i(\mathbf{x}^f+\Delta_j)-\beta_i(\mathbf{x}^f-\Delta_j)}{2\Delta} \qquad (4.3.17)$$

$$\frac{\partial\tau_i}{\partial x_j^f} \simeq \frac{\tau_i(\mathbf{x}^f+\Delta_j)-\tau_i(\mathbf{x}^f-\Delta_j)}{2\Delta}, \qquad (4.3.18)$$

where the step-size Δ ,is set by the **delta_used_in_numerical_jacobian_determination** configuration parameter.

Next, the Jacobians elements necessary for linking the uncertainties in the extinction, backscatter and optical depth profiles to the uncertainty estimates in the total attenuated backscatter signal signals are calculated. An efficient manner of doing this is to calculate the multiple-scattering factor, use this to calculate the single-scatter signals and then to use analytical derivatives derived from Eq. (2.2.16).

From Eq.(2.2.25) it is apparent that the multiple-scattering factor appropriate to the total signal can be written as

$$M_t(z) = \left(1 - f_e(z)\right) + f_e(z)\exp\left[2\int_{z_{sat}}^{z} \eta(z)\alpha(z')dr'\right] \tag{4.3.19}$$

allowing B'_t to be calculated using Eq.(2.2.11) together with Eq. (2.2.12).

Starting with a discrete form of Eq.(2.2.16) the necessary partial derivatives can then be calculated. In discrete form Eq.(2.2.16) can be written as (where the indices increase with range away from the lidar)

$$\alpha'_i = \frac{\mathrm{B}'_{t,i}}{\dfrac{\mathrm{B}'_{t,i_{top}}}{\alpha'_{i_{top}}} + 2\sum_{k=i}^{i_{top}} \mathrm{B}'_{t,k}\,\Delta r_k} \tag{4.3.20}$$

where the normalization range has been set to be the top of the lidar profile. Using Eq.(4.3.20) it can be shown that

$$\frac{\partial\alpha_i}{\partial \mathrm{B}'_j} = 2\frac{\alpha'^2_i}{\mathrm{B}'_i}\Delta r_j \qquad i \neq j \qquad j < i \tag{4.3.21}$$

$$\frac{\partial\alpha_i}{\partial \mathrm{B}'_i} = -2\frac{\alpha'^2_i}{\mathrm{B}'_i}\Delta r_j + \frac{\alpha'_i}{B'_j} \qquad i = j \tag{4.3.22}$$

$$\frac{\partial\alpha_i}{\partial \mathrm{B}'_i} = 0 \qquad i > j\,. \tag{4.3.23}$$

For the backscatter coefficient we have

$$\frac{\partial\beta}{\partial B'_t} = \frac{1}{S_i}\frac{\partial\alpha}{\partial B'_t}, \tag{4.3.24}$$

and for the optical depth we have (using Eq.(2.2.15))

$$\frac{\partial\tau_i}{\partial B'_{t,j}} = \exp[2\tau'_i]\Delta r_j \qquad i < j \tag{4.3.25}$$

and zero otherwise.

The extinction error covariance matrix can now be calculated as

$$\mathbf{C}_\alpha = \mathbf{J}_\alpha \mathbf{C}'\mathbf{J}_\alpha^{\,t} \tag{4.3.26}$$

where

$$\mathbf{J}_\alpha(i,j) = \frac{\partial\alpha_i}{\partial\mathbf{x}_j} \qquad j \le size(\mathbf{x}) \tag{4.3.27}$$

$$\mathbf{J}_{\alpha}(i,j) = \frac{\partial \alpha_i}{\partial \mathrm{B}_t{}'_j} \qquad j > size(\mathbf{x}) \tag{4.3.28}$$

and

$$\mathbf{C}' = \begin{pmatrix} \mathbf{C} & . & . & 0 \\ . & \delta^2 B'_{t,1} & 0 & 0 \\ . & 0 & . & 0 \\ 0 & 0 & 0 & \delta^2 B'_{t,nz} \end{pmatrix} \tag{4.3.29}$$

i.e. C' combines the inversion covariance matrix with the diagonal total signal variance matrix. Expressions similar to Eqs.(4.3.26)-(4.3.29) are used for the calculation of the backscatter and optical depth error covariance matrices.

Eb1-6: Call A-TC routines

On a layer-by-layer basis the A-TC detailed classification routine is called to produce the JSG resolution A-TC product (See Flowchart TC1).

Eb1-7: Calculate low and medium resolution fields

The extinction, backscatter and Rayleigh attenuation corrected attenuated backscatter fields are smoothed horizontally to form the corresponding *medium* and *low resolution* fields. This step is similar to the A-AER processing Step Ae1-3). Before smoothing the fields are masked so that only points where:

((**classification** .ge. 3) .and. (**FM_jsg** < **feature_mask_cutoff_for_horizontal_smoothing**) .and..(**extinction** < 1.0e-2 1/m).

After the hybrid merged fields are found, the layer values of the lidar-ratio and depolarization ratio are recalculated and their respective error estimates are updated.

Eb1-8: Call A-TC routines for Low and Medium Resolution fields.

On a layer-by-layer basis the A-TC detailed classification routine is called to produce the **A-TC** products for the low and medium resolution product streams (See Flowchart TC1).

Eb1-9: Call A-ICE

For the layers classified as ice, the A_ICE routine is called to produce the **A-ICE** product for the JSG, Low- and Medium-resolution output streams (see Sections 5.4.4 and 5.5.4).

5.5.2.1 Optimal Estimation Cost-function Calculation

Eb1-4-1: Process state vector
The relevant parameters are extracted from the state-vector (see Eq.(2.2.44)).

Eb1-4-2: Form the total signal
The total signal is formed from the component Mie, Ray and CRO attenuated backscatter signals (see Section 3.1.4.5).

Eb1-4-3: Start inversion (loop in range gate)
The inversion process (the gate-by-gate solution of Eq.(2.2.37)) is started looping in range starting at the gates closest to the lidar.

Eb1-4-4: Calculate f(iz)
The effective f factor for this range gate is calculated (see Eq.(2.2.28)).

D5: Solve for Extinction analytically
If f is equal to 1 (or zero) then Eq.(2.2.38) can be solved directly for the effective extinction (α'_{iz}).

D6: Solve for Extinction numerically
If f is not equal to 1 (or zero) then Eq.(2.2.38) must be solved numerically for α'_{iz} . A Brent type zero finding routine (see Press et al. (1992)) is employed to do this. If no good solution is possible, i.e. the minimum of the function is greater than the **use_last_gate_on_bad_solution** parameter, then the extinction of the value of the previous range gate is used. This is necessary in order to prevent the solution from "blowing" up which can otherwise happen in cases were the optical thickness is greater than about 2.

D7: If specified, calculate derivatives
If required the cost-function routine will calculate and return the profiles of $\frac{\partial \alpha}{\partial B_t}$ and $\frac{\partial \alpha}{\partial B_t}$ this is done, typically only after the calling minimization procedure has converged. The partial derivatives are used later to help generate error estimates.

D8: Construct the Predicted Rayleigh signal
On the basis of the retrieved extinction and f profile the predicted Rayleigh attenuated backscatter is calculated. i.e.

$$B_{Ray}(z) = \beta_{Ray}(z) e^{-2\int_{z_{sat}}^{z} (\alpha_R(z') + \alpha_M(z'))dz'} \left[\left(1 - f(z)\right) + f(z) e^{2\int_{z_{sat}}^{z} \eta(z)\alpha_M(z')dz'} \right]. \qquad (4.3.30)$$

D9: Calculate the Cost-function

The various components of the cost-function (χ^2) are calculated and summed (see Eq.(2.2.42)). Minimum relative and absolute errors in the attenuated backscatters are filtered using the **minimum_relative_error_in_cost_func_mie**, **minimum_relative_error_in_cost_func_ray, minimum_absolute_error_in_cost_func_mie** and **minimum_absolute_error_in_cost_func_ray** parameters.

5.5.3 *A-TC*

5.5.3.1 Target Classification

The target classification procedure is described in three flow diagrams, i.e. TC1[, TC2 and TC3. TC1 describes the separation between clouds and aerosols and the cloud subtyping, TC2 shows the aerosol and ice typing and TC3 shows the spatial consistency testing of the resulting aerosol and ice cloud fields.

The inputs beside the layer averaged auxiliary data (i.e. temperature [T_l]) to TC1 and TC2 are:

- layer averaged depolarization and associated error (δ_l and $\Delta\delta_l$)
- layer averaged lidar ratio and associated error (S_l and ΔS_l)
- layer thickness (H_l) and central altitude (Z_l)
- layer averaged extinction and associated error (α_l and $\Delta\alpha_l$)
- layer averaged unattenuated backscatter and associated error (β_l and $\Delta\beta_l$)
- profile mask (*mask*); set to 1 for all layers for which a type needs to be retrieved
- simple liquid classification (*simple*) set to 1 if at least a single pixel within a respective layer is classified as liquid within the simple classification
- profile mask (*mask*); set to 5 for all layers for which simple says it is stratospheric.

TC1-1: Check input values
Check input values of δ_l, $\Delta\delta_l$, S_l and ΔS_l and set mask to 0 when one of these show unrealistic values (i.e. $|\delta_l|$>1e4).

TC1-2: Set Beta_cld
Set the value of the layer average backscatter coefficient used to distinguish clouds based on the altitude

$\beta_{thres} = \beta_{thres,std}$
Z_l < Boundary layer height → $\beta_{thres} = \beta_{thres,bl}$
T_l > 273K → $\beta_{thres} = \beta_{thres,liq}$
Z_l > Tropause → $\beta_{thres} = \beta_{thres,strat}$

TC1-3: Calculate the cloud probability (P_{cld})

Layers are classified as being aerosol layers if their associated average backscatter is below the **Beta_Cld** threshold. The associated probability given the estimate error in the layer mean beta (calculated earlier as part of the layering procedure by standard quadratic summation of the individual beta errors) that the threshold has been exceeded is also calculated assuming Gaussian statistics. In particular, the cloud occurrence probability is giving by

$$P_{cld}(\beta_{thres};\beta_l,\sigma_{\beta_l}) = 1 - 0.5\left[1.0 + erf\left(\frac{\beta_{thres} - \beta_l}{\sigma_{\beta_l}\sqrt{2}}\right)\right] \tag{4.3.30}$$

where the layer averaged backscatter and associated errors were calculated in the layering procedure using

$$\beta_l = I\beta\left(z_{top} - z_{bot}\right)^{-1} \tag{4.3.30}$$

and

$$\sigma_{\beta_l} = sqrt\left(\frac{\sum_{i=z_{bot}}^{i=z_{top}} \sigma_{\beta_i}^2}{\left(\frac{(z_{top} - z_{bot})}{\Delta z} \right)} \right) \qquad (4.3.30)$$

where Δz is the range resolution.

TC1-4: Separate cloud & aerosol pixels

The pixels with $P_{cld} \geq 50\%$ are dealt with in the cloud typing routines starting in **TC1-7**, whereas the other pixels are checked for their height Z with respect to the local tropopause (**TC1.-5**) and are dealt with as aerosols (**TC2**) or tropospheric clouds/aerosols (**TC1-6**).

TC1-6: Stratospheric Clouds and Aerosols

The stratospheric pixels where $P_{strat} \geq 50\%$ and the Temp < **max_temp_psc** are defined as polar stratospheric clouds they are separated on their depolarization and lidar ratio similar to how the aerosols are typed (TC2).
When $P_{strat} \geq 50\%$ and the Temp > **max_temp_psc** and $P_{cld} \geq 50\%$ the pixel is set to an ice cloud (either the tropopause has been incorrectly set or a Cirrus cloud has been pushed within the stratosphere)
For stratospheric pixels where $P_{strat} < 50\%$ the pixel is set to stratospheric aerosols. The typing is performed on their depolarization and lidar ratio similar to how the aerosols are typed (TC2).

TC1-8: Water/ice Depol Beta Test

A threshold between water-cloud and ice-cloud is constructed by using a linear relationship between layer integrated depolarization ratio and the layer integrated backscatter **(TC1-7),** i.e.

$$\delta_{thres} = A_{\delta,\beta} \cdot \beta_l \cdot H_l + B_{\delta,\beta} \qquad (4.3.30)$$

$$\delta_{thres2} = (\delta_l \cdot H_l - \delta_{thres})/\Delta\delta_l \qquad (4.3.30)$$

Both equations are evaluated for the layer integrated backscatter and depolarizations. If the observed value of the layer depolarization is greater than the resulting value of δ_{thres2}than the ice probability is calculated:

$$P_{ice} = 1. -0.5 \left[1 + erf\left(- {\delta_{thres2}}/{\sqrt{2}} \right) \right] \qquad (4.3.30)$$

Pixels with mask=1 and $P_{ice} > 0.5$ are classified as ice **[TC1-9].**

TC1-10 to TC1-12: Checking $P_{ice} < 0.5$ values

If the layer has a value smaller than 0.5 it is tested if this is indeed a liquid cloud. This is done in a series of three tests.

1. T_l is checked for temperatures below 233 K (**TC1-10**). The pixel is set to ice (**TC1-14**).
2. The layer is checked if *simple* shows that at least a single pixel within the layer is set to liquid (**TC1-11**). If this is not the case the integrated backscatter is too small for liquid droplets and the pixel is set to aerosols (**TC2**).
3. The liquid pixel is subsequently tested for T_l > 270K (**TC1-12**). If this is the case the pixel is set to liquid droplets (**TC1-15**) and otherwise supercooled droplets (**TC1-16**).

TC1-17: Combining aerosol and cloud results.

Within **TC1** the aerosol typing is called (**TC2**: Section 5.5.3.2). The aerosol typing also returns tenuous ice layers as an aerosol type. To check for internal consistency and combining the different aerosol types and cloud types into a single classification several checks are performed:

The main target signals coming from TC1 is the *target_class*, from TC2 is the *aerosol_type*
First the classification matrix is constructed and set to clear sky (0)

Where construct on input	Reason	Classification
Simple_classification =-3	There is missing data, this has been tested already within the simple classification.	-3
Simple_classification =-2	The pixel is at or below the lidar surface (as detected in A-FM). This has been masked in the simple classification	-2
Simple_classification =-1	The pixel has been fully attenuated (as detected in A-FM in the Rayleigh channel). This has been masked in the simple classification	-1
Target_class=1	Target classification is (warm) liquid droplets	1
Target_class=2	Target classified as supercooled liquid droplets	2
Target_class=3	Target classified as ice crystals	3
Aerosol_type=2^{iaer}	Check for retrievals of all aerosol types (1 to n_{aer})	10+iaer
Aerosol_type=ice=$2^{(iaer=iice)}$	The aerosol type is detected as ice	3
Strat_cloud_type=2^{ipsc}	Check for retrievals of all stratospheric clouds types (1 to n_{psc})	20+ipsc
Strat_aerosol_type=2^{istrat_aer}	Check for retrievals of all stratospheric aerosols (1 to n_{strat_aer})	25+i_strat_aer
Aerosol_type=-1	The highest aerosol probability is too low, no specific class can be assigned	101
Aerosol_type=-2	Observation is outside of parameter space	102
Strat_Aerosol_type=-1	The highest stratospheric aerosol probability is too low, no specific class can be assigned	104
Strat_Aerosol_type=-2	Stratospheric aerosol Observation is outside of parameter space	105
Strat_cloud_type=-1	The highest stratospheric cloud probability is too low, no specific class can be assigned	106
Strat_cloud_type=-2	Stratospheric cloud Observation is outside of parameter space	107

5.5.3.2 Aerosol Typing (TC2)

The procedures within the aerosol typing module are mathematically relatively simple. There is a straightforward route to follow as is shown in the flow diagram. The most important task, related to this algorithm, will be to check the probabilities of the each of the aerosol types and mixtures. The algorithm is set up to both handle pure types (thereby creating mixtures) or starting with mixtures. In the former case, up to three types are returned which can be interpreted as a 'mixture'. In the latter case only a single mixture is retrieved. In the discussion below the word 'aerosol-type' is used for the types as defined by the configuration parameters, including ice-crystals.

TC2-1: Calculate weighted probability for each aerosol type:

Each aerosol type is defined by the 5 configuration parameters for each type ($\delta_0, S_0, \sigma_\delta, \sigma_S$ and θ), describing a two-dimensional Gaussian probability distribution with a maximum of 1.

$$\begin{aligned} A &= \frac{\cos(\theta)^2}{2\sigma_\delta^2} + \frac{\sin(\theta)^2}{2\sigma_S^2} \\ B &= \frac{-\sin(2\theta)}{4\sigma_\delta^2} + \frac{\sin(2\theta)}{4\sigma_S^2} \\ C &= \frac{\sin(\theta)^2}{2\sigma_\delta^2} + \frac{\cos(\theta)^2}{2\sigma_S^2} \\ P(\delta,S) &= P_0 \exp(-1(A(\delta-\delta_0)^2 + B(\delta-\delta_0)(S-S_0) + C(S-S_0)^2)) \end{aligned} \qquad (4.3.30)$$

with θ the rotation of the 2D-Gauss, assuming the depolarization on the x-axis.

Calculating the type probabilities

For each observed combination of lidar ratio and depolarization, the probability can be calculated for the aerosol types being considered. If this calculation were to be done rigorously, one would need to calculate the convolution of the 2D observed Gaussian with the 2D bivariate Gaussian probability density functions. This calculation becomes an enormous expression when solved analytically and therefore the calculation is carried out approximately by retrieving the probability 49 times (7x7) between plus and minus 3σ in both the probability and depolarization. The effective probability ($P_{eff,d}$) is calculated by adding the weighted 49 probabilities (P) according to:

$$\begin{aligned} P^{l}_{eff,d}(\delta_k,S_k) &= \sum_{i=1}^{7}\sum_{j=1}^{7} P(\delta_k + (i-4)\sigma_{\delta,k}, S_k + (j-4)\sigma_{S,k}) \bullet W(i,j) \\ W(i,j) &= \frac{1}{2\pi}\exp(-\frac{(i-4)^2}{2} - \frac{(j-4)^2}{2}) \end{aligned} \qquad (4.3.30)$$

with S_k and δ_k the observed lidar ratio and depolarization with their respective error estimates ($\sigma_{S,k}, \sigma_{\delta,k}$) and $P^{l}_{eff,d}$ the probability of the observation k for type l . The weighting function (W) assumes an even spread between -3σ and +3σ and P is the probability distribution as was defined above.

TC2-2: Sort aerosols

Sort the aerosols according to their values starting with the highest Probabilities → P1, P2, P3 and their respective types (T1,T2,T3). In the table below the different routes are described with in light grey the possible routes in case when determining aerosol mixtures:

Probability test	Secondary test	Aerosol type Result	TC2 order	Classification value
P(1) < P_{min}^{aer}	$S_l > 0.05$	Aerosol types outside of parameter space	2-3 , 2-4	101
P(1) < P_{min}^{aer}	$S_l \leq 0.05$	No aerosols	2-3 ,2-4	0
P(1) ≥ P_{min}^{aer}, P(2) <P_{min}^{aer}		Single aerosol type only	2-3, 2-5	T[1]
P(1) ≥ P_{min}^{aer}, P(2) ≥P_{min}^{aer}	P(1) ≥P_{min}^{first}	Single aerosol type only	2-3, 2-5	T[1]
P(1) ≥ P_{min}^{aer}, P(2) ≥P_{min}^{aer}, P(3) <P_{min}^{aer}		Mixture of two aerosol types	2-3, 2-5, 2-6	T[1]& T[2]
P(1) ≥ P_{min}^{aer}, P(2) ≥P_{min}^{aer}, P(3) ≥P_{min}^{aer}	P(1) ≥P_{min}^{first}	Single aerosol type only	2-3, 2-5 2-6, 2-7, 2-8	T[1]
P(1) ≥ P_{min}^{aer}, P(2) ≥P_{min}^{aer}, P(3) ≥P_{min}^{aer}	P(1) <P_{min}^{first} P(2) <P_{min}^{sec}	Mixture of three aerosol types	2-3, 2-5, 2-7	T[1] & T[2] & T[3]
P(1) ≥ P_{min}^{aer}, P(2) ≥P_{min}^{aer}, P(3) ≥P_{min}^{aer}	P(1) <P_{min}^{first} P(2) ≥P_{min}^{sec}	Mixture of two aerosol types	2-3, 2-5, 2-7, 2-8	T[1] & T[2]

TC2-9: Check ice pixels

Ice is one of the possible 'aerosol types' retrieved within the procedure. To test whether the ice retrieval is consistent with the retrievals, the extinction is compared to a temperature and height based threshold. In version 4.0 only the temperature relationship varies whereas the height relationship is not variable at this point.

A linear interpolation in temperature (T_{ice}) and height (H_{ice}) parameter space (TH_{ice}) is performed with the layer central Temperature (T_l).
For all layers where:

$\alpha_l < TH_{ice}(T_l, H_l)$ and the highest aerosol type probability > 1.e-4

The ice pixels are changed to aerosol pixels and the aerosol type is updated to the next highest aerosol type probability.

The final results from the TC2 procedure is combined in TC1-17 with the cloud retrievals resulting in the final target classification.

5.5.3.3 Ice and aerosol horizontal homogeneity (TC3)

In this procedure, the horizontal and vertical homogeneity of the aerosol and ice clouds field is checked and smoothed using the hybrid-mode routine described in Section 3.1.5.4. Liquid layers, attenuated regions and surfaces are not touched, nor removed by this procedure.

TC3-1: Create an aerosol and ice mask

Check input classification and set for all aerosol (class≥10) and ice clouds (class=3) the mask to 1
Copy the input classification to *class_copy*
Calculate the extinction signal to noise ratio:

$$SNR_{\alpha} = {}^{\alpha_l}/_{\Delta\alpha_l}$$

TC3-2: Loop the hybrid mode calculation 5x to fill gaps and smooth areas (ends at TC3-9)

TC3-3: Copy classes

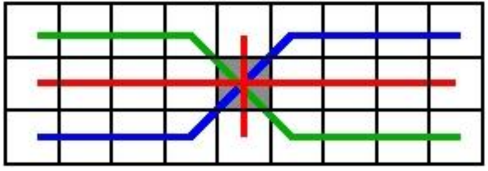
Copy *class_copy* to two dummy variables (*class1, class2*) after which the array is looped through in time (it) and height (ih)

TC3-4: Define kernel

For each pixel (it,ih) copy a Kernel (K) with size $nhor_{cl}$ x 3 around it,ih from class_copy
Set all liquid (K=1), supercooled (K=2), surface (K=-2) and attenuated (K=-1) to K=-3 [**TC3-5**]

TC3-6: Calculate the Kernel four modes

Retrieve the mode (highest occurrence) of the four colored lines depicted below.



From these four values calculate the first and second mode (**TC3-7**) and update class1 and class2 with these values respectively (**TC3-8**).

TC3-9: Update class_copy

Update class_copy with class1 and return to **TC-3-2**, if the hybrid-mode has been performed 5 times go to **TC3-10**.

TC3-10 - 3-15: Create output classification and update values

Create class_out=class1
The consistency of the outputs are checked and updated using a number of where constructs:

Where construct	Reason	TC3	Change into
Class_out=-3	Original pixel was surrounded by liquid, attenuated or surface pixels. Class2 might have a valid neighbor	3-11	Class2
Class_in≥3, Class_out=0, $SNR_{\alpha} \geq 3 \geq SNR_{min}$	The input classification had and assigned pixel, this has been set to clear sky. If the SNR is large enough try to see if class 2 is more appropriate	3-12	Class2
Class_out=-3	For all pixels which had both class1 and class2 as -3 set to unknown.	3-14	103
Class_out=3, T>273, Class2≠3	If the class_out is set to ice but the temperature > 273K and class2 is not equal to ice. Set class2	3-14	Class2
Class_out=3, T>273	For all remaining ice pixels above 273K set type to unknown	3-15	103

5.5.4 A-ICE

Ic1-1:
The procedure is called using the ice extinction and temperature information from either the main A-AER or A-EBD procedure

Ic1-2: Find coefficients

$$C_0 = C_{0,0} + C_{0,1} T[^{\circ}\mathrm{C}] \qquad (4.3.31)$$

$$C_1 = C_{1,0} - C_{1,1} T[^{\circ}\mathrm{C}] \qquad (4.3.32)$$

Ic1-3: Calculate IWC and Reff

$$IWC[\mathrm{gm}^{-3}] = C_0 \left(\alpha[m^{-1}]\right)^{C_1} \qquad (4.3.33)$$

$$R_{eff}[\mu m] = \frac{3}{8}\frac{C_o}{\rho_i}(\alpha)^{C_1-1}.10^6 \qquad (4.3.34)$$

Ic1-4: Calculate Errors

The errors due to the uncertainty in the input extinction are calculated. Using standard error propagation theory the above equations leads to

$$\delta IWC = \left|\frac{C_1 IWC}{\alpha}\delta\alpha\right| \qquad (4.3.34)$$

and

$$\delta R_{eff} = \left|R_{eff}\left(C_1 - 1\right)\frac{\delta\alpha}{\alpha}\right| \qquad (4.3.34)$$

Note that the errors calculated here do not include the uncertainty inherent in the parameterization itself (which will often be the dominant error source).

6 ALGORITHM PERFORMANCE, SENSITIVITY STUDIES, LIMITATIONS

A-PRO has been evaluated using several detailed simulated frames. It should be noted that the evaluation presented here remains preliminary. Not all the relevant parameters have been assessed yet and the sensitivity of the results to e.g. a priori values remains a work in progress. Future work will be requited in order to optimize the algorithm configuration in terms of e.g. averaging scales, a priori error estimates. However, A-PRO is mature enough so that a useful evaluation can be conducted.

6.1 Testing and Development Scenes

Several frame sized scenes have been constructed. ECSIM RT and instrument models have been used to create corresponding simulated realistic ATLID data. A-PRO has been applied to these data and the results have been compared to the input "truth".

Five main test scenes have been constructed so far and are briefly described below in Table 5.

Scene Name	Orbit (lat,lon➔lat,lon)	Data Source	Brief Description
MACC	-11.8,19.5➔42.9,0.7	CAMS cloud and aerosol fields	Transit over Africa into southern Europe. An aerosol rich scene. The path does not follow an EarthCARE orbit. The orbit was selected on the basis of Calipso observations.
Halifax	68.9,-50.1➔21.36,-70.0	ECCC Gem outputs merged with CAMS aerosol fields	Southern Greenland to the Caribbean, passing cloud to Halifax N.S. Canada during winter. A storm system passed over, Broken clouds and cloud-free aerosol regions are passed over in the southern section of the frame.
Baja	68.6,-96.7➔21.1,-115.3	ECCC Gem outputs merged with CAMS aerosol fields	A frame staring in Northern Canada passing over the Rocky Mountains, the Baja peninsula, and ends in the Pacific. Large topographical variations are present in this scene.
Hawaii	23.86,-161.0➔023.81,-170.3	ECCC Gem outputs merged with CAMS aerosol fields	A tropical Pacific scene passing near the Hawaiian Islands.
PSC	-66.2,69.7➔-66.4,-75.23	Built using Calipso+CSAT DARAR products	A winter transact over Antarctica with both tropospheric and stratospheric features (including PSCs) present. This scene does not (yet) correspond to a EarthCARE orbit.

Table 5: Primary test scenes.

In the following sections examples of the application to of A-PRO to each of the above scenes will be presented and discussed. Details of how the scenes were constructed are given in Appendix A.

6.2 MACC Scene Results

Simulated ATLID attenuated backscatter fields as well as the corresponding A-FM product (which serves as one of the inputs to A-PRO) are shown in Figure 34 for the MACC scene. Of particular note is the depolarizing returns near the ground between 10 and 20 Deg N. This area corresponds to the presence of dust-aerosol (See Appendix-A).

Sample retrievals along with selected true fields are shown in Figure 35. Here the “low-resolution” hybrid resolution (e.g. 150km horizontal resolution for weak-features and 1-km for clouds and other ‘strong’ features) retrieval results are shown for the extinction, lidar-ratio, and classification fields. These arguably are the fields most useful for the end-users. It can be seen that the extinction and lidar-ratio retrievals follow the true fields on the whole although differences of varying amounts are present. Of particular note is the fact that about 40N under 5km that clear-sky is reported, this is due to the A-FM product being in error here. Another notable observation is the fact that the classification procedure picks up the fact that the low-level aerosol between about 5 and 25N is mainly Dust.

More quantitative detail is shown in Figure 36. Here it can be seen that for aerosols with an extinction over about 10^{-4} m^{-1} that the extinction is generally retrieved with an accuracy on the order of 10%. For extinction values between $1\text{-}5 \times 10^{-5}$ m^{-1} the accuracy decreases leading to differences on the order of up to 50% and the differences tend to be correlated in altitude. Errors are also seen to sharpy increase as the lidar signals are attenuated by strong features. Errors in the retrieved lidar ratios tend to be around -/+ 15 sr and also increase sharply when the lidar signals are strongly attenuated.

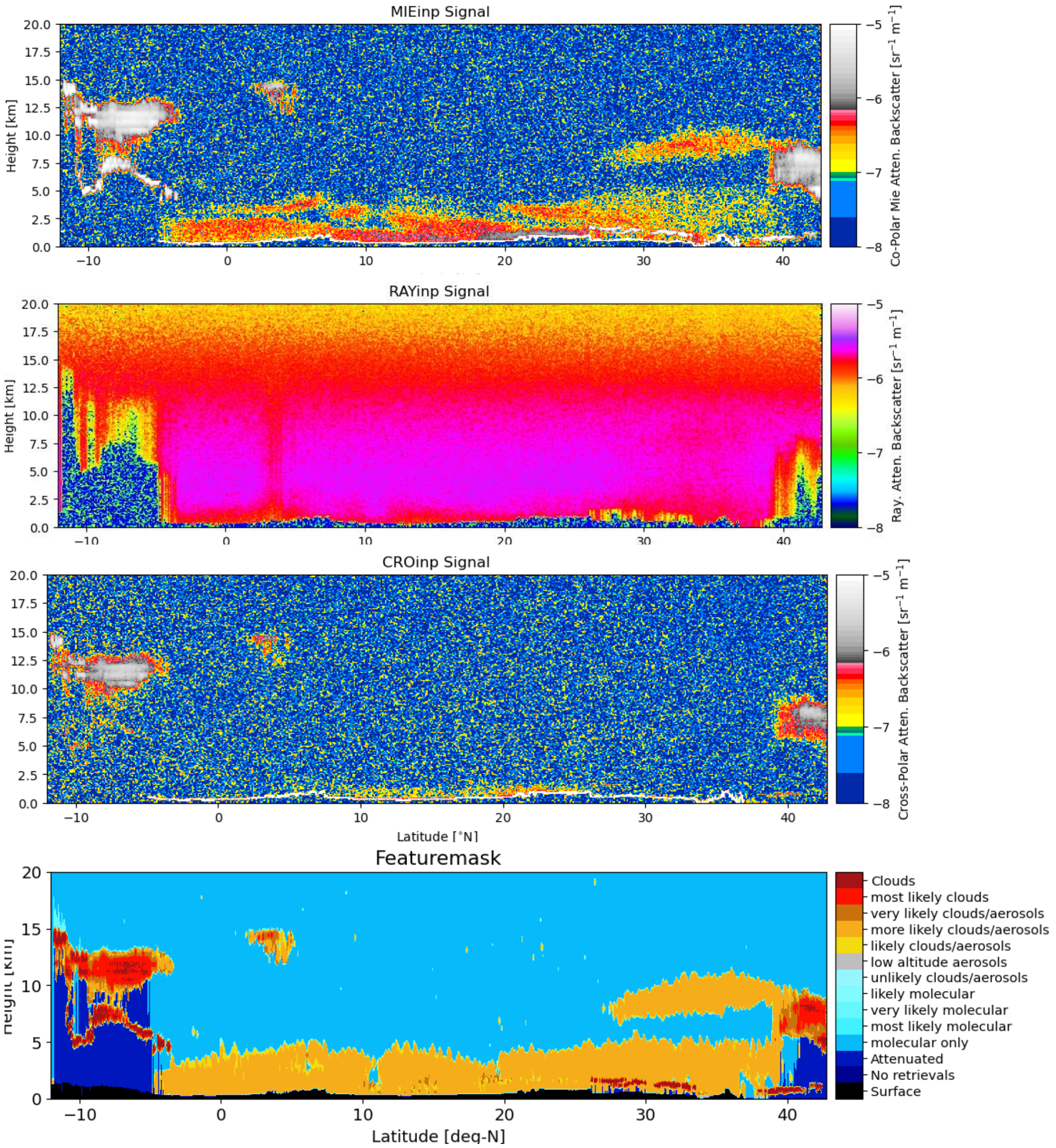


Figure 34: Simulated Mie, Rayleigh and Cross-polar attenuated backscatter fields along with the corresponding A-FM product. The signals have been smoothed using a 10 pixel along-track window

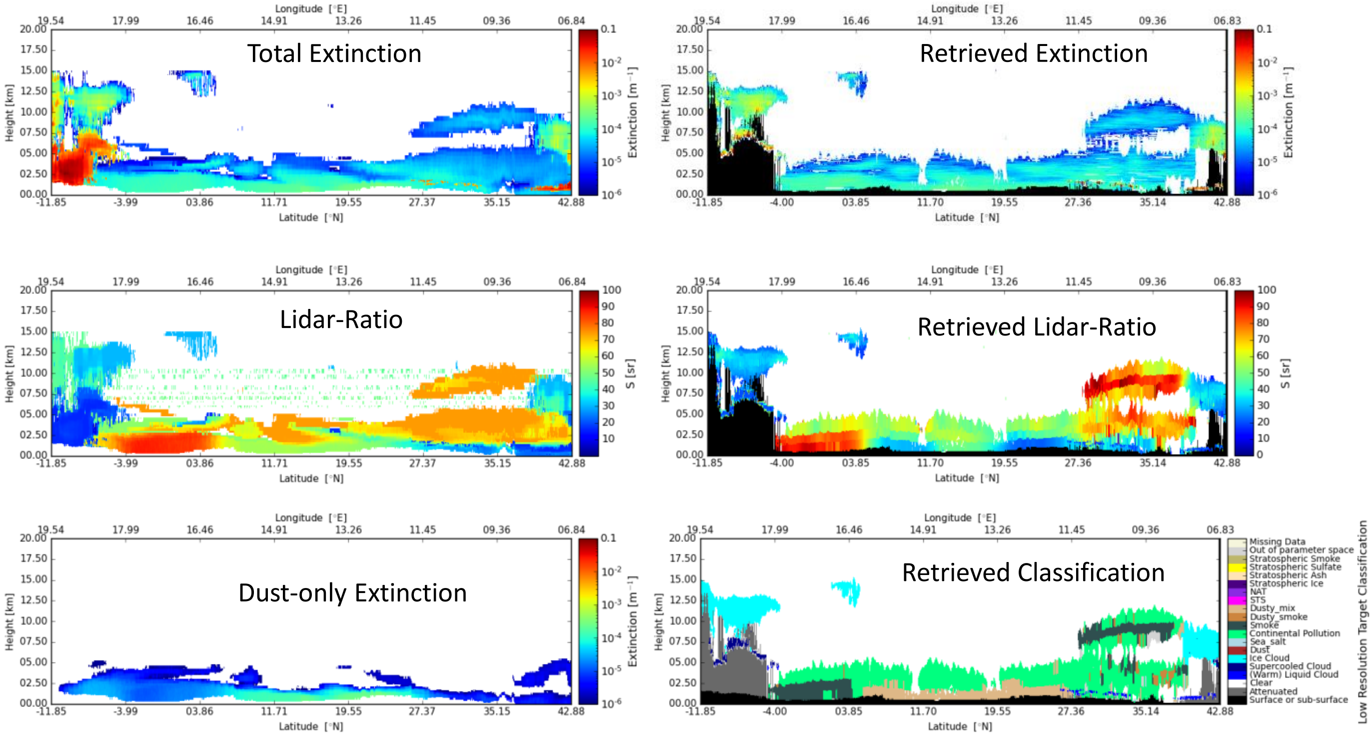


Figure 35: Sample retrieval results (Right) along with true fields (Left) for the Macc scene. Black areas in the retrieval plots correspond to fully attenuated areas.

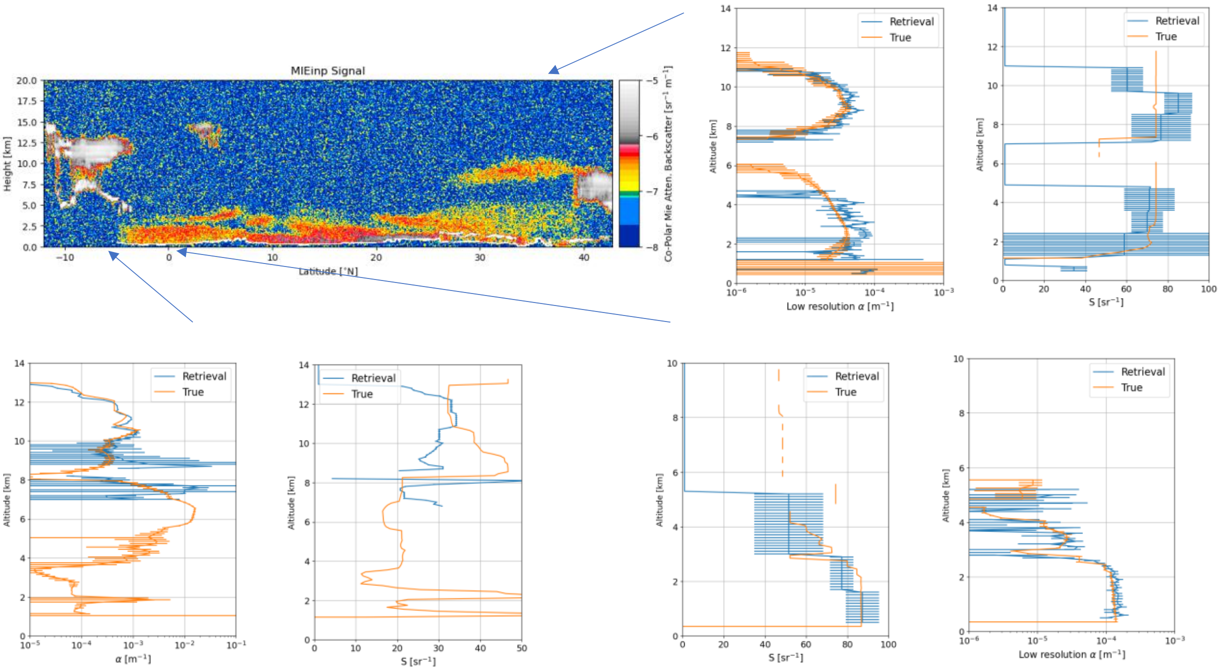


Figure 36: Sample profiles of retrieved extinction and lidar-ratio as well corresponding model truth profiles. Here the model truth profiles have been horizontally smoothed to match the resolution of the retrievals.

6.3 Halifax Scene Results

The first of the GEM derived scenes is the Halifax scene. It is a more complex scene than the Macc scene. Here supercooled water, ice, snow, and aerosol are present. Referring to Figure 38 it can be seen that the retrieved lidar ratios in the ice clouds are lower than the true values. This may be due to the a priori ice cloud lidar-ratio being too low and/or weighted too strongly. Most features (except in cases of strong attenuation) are well retrieved however it can be noted that the elevated aerosol layer at about 60 N at 7.5 km is mostly missed. The extinction of this aerosol region is on the order of $2x10^{-6}$ m^{-1} which, considering the limited extent of the region, is to small to be retrieved. The aerosol layers in the southerly section of the scene are less well retrieved than the more optically thick layers in the Macc scene. Never-the-less the lidar-ratio is retrieved with enough accuracy to discern that two distinct layers (a lower altitude marine layer and a higher altitude continental pollution layer) are present.

Two example retrieval profiles are shown in Figure 39. Here for the cirrus case, it can be seen that the lidar-ratio and extinction retrievals are generally too low. In this case it appears that the a-priori lidar-ratio generated by the AER step are too low and/or weighted to heavily in the EBD step. The reason for this behaviour is not clear at this stage, but may be due to the handling of multiple-scattering (e.g. the a priori values of theta and eta used here). For the aerosol example, the extinction values are retrieved with an accuracy on the order of 10% below 2 km and approaching 50% at about 4km. As noted earlier the aerosol extinction values in this scene are lower than the ones in the Macc scene leading to comparatively less accurate (but still useful) retrievals.

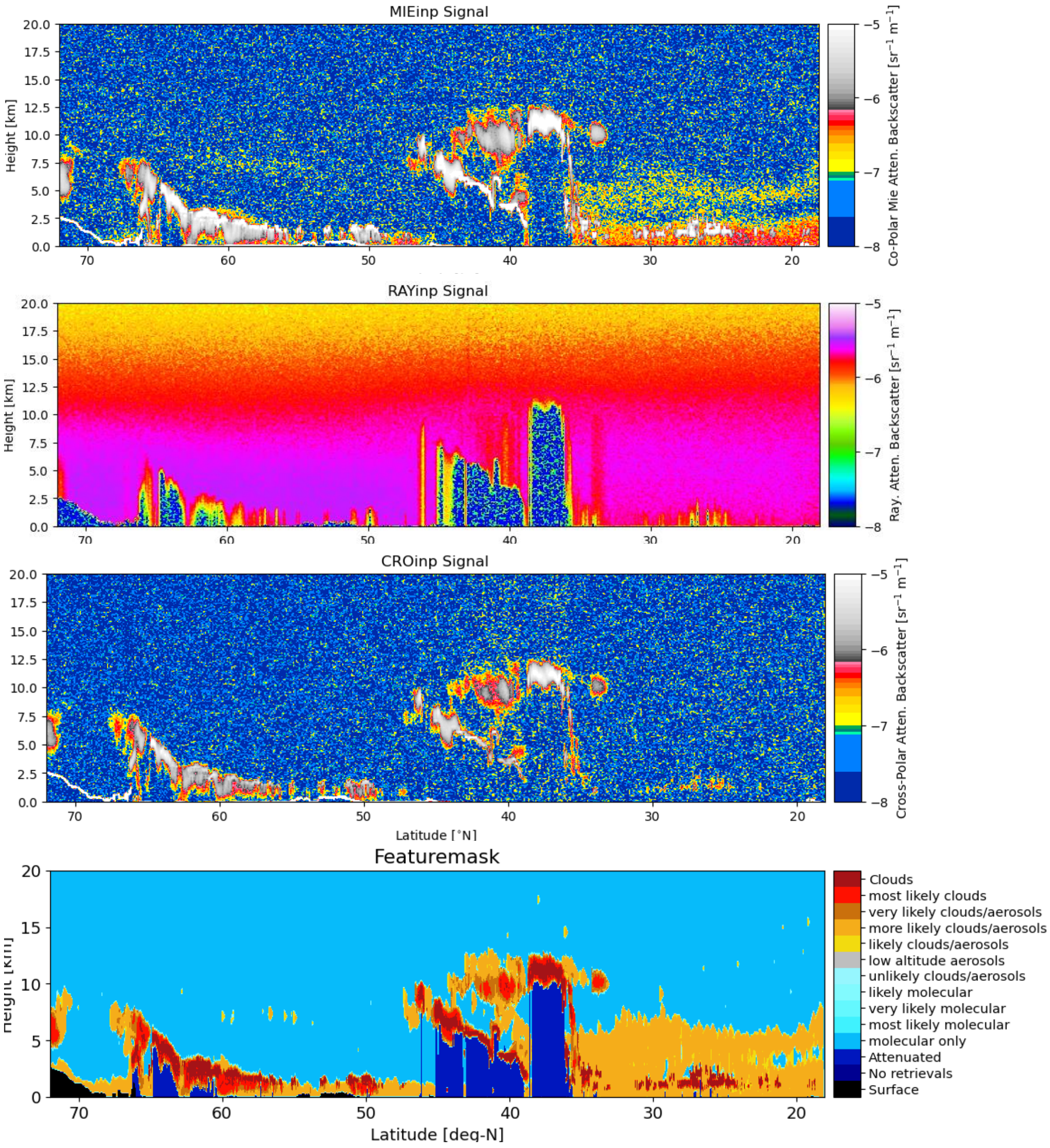


Figure 37: Simulated Mie, Rayleigh and Cross-polar attenuated backscatter fields along with the corresponding A-FM product for the Halifax scene. The signals have been smoothed using a 10 pixel along-track window

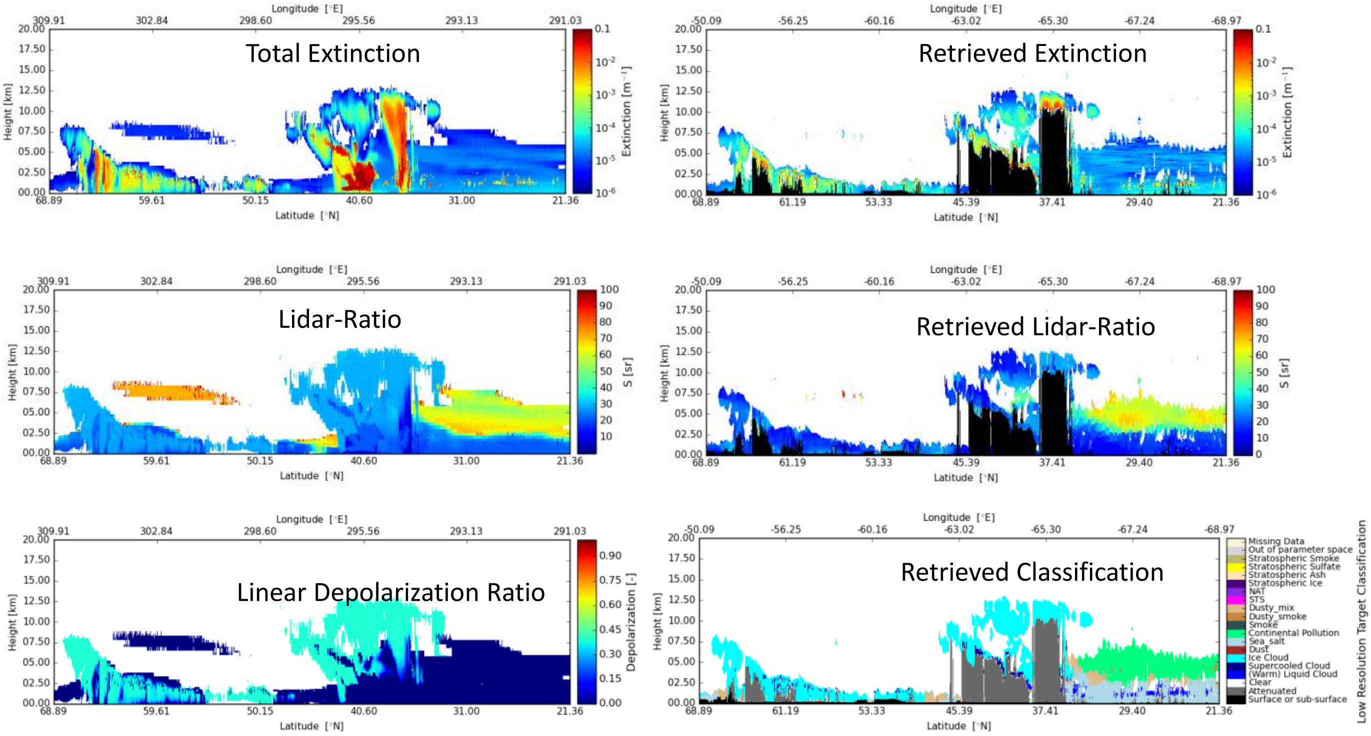


Figure 38: Sample retrieval results (Right) along with true fields (Left) for the Halifax scene.

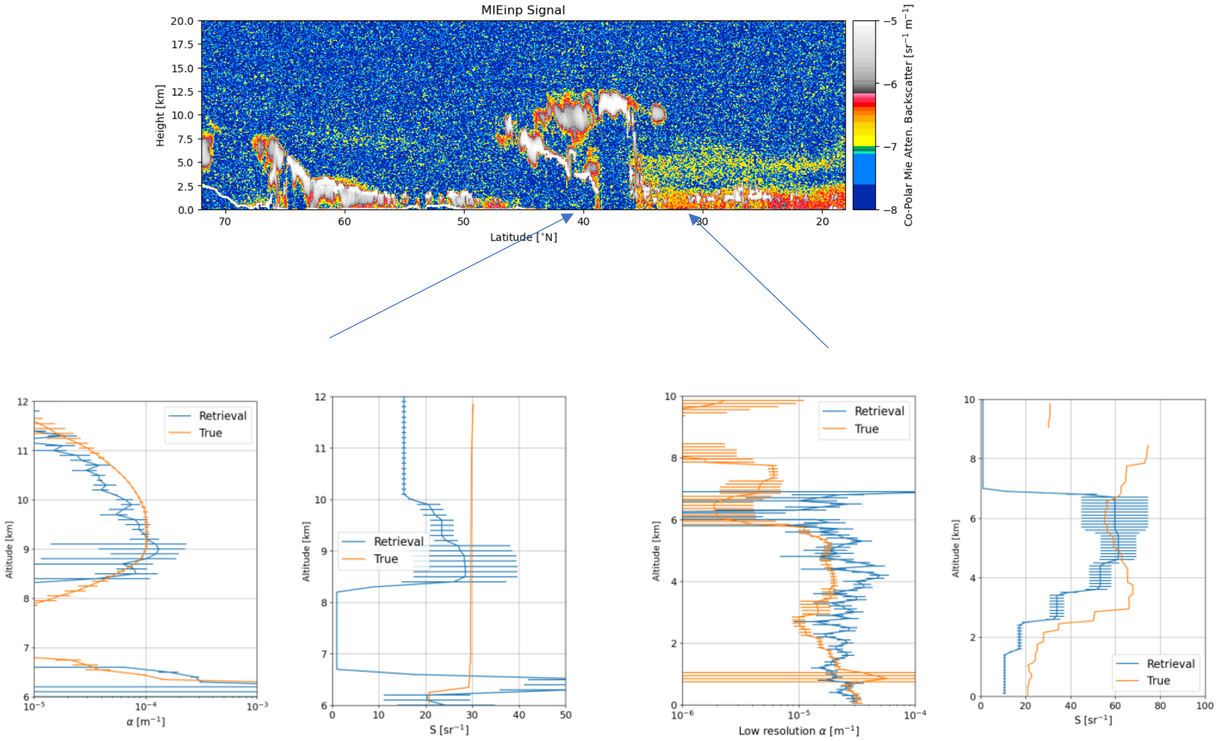


Figure 39: Sample profiles of retrieved extinction and lidar-ratio as well corresponding model truth profiles. Here the model truth profiles have been horizontally smoothed to match the resolution of the retrievals.

6.4 Baja Scene Results

The Baja scene is the 2nd GEM derived scene. This scene has a lot of topographical variation compared to the Halifax scene and contains large regions of thinly distributed aerosols. Considering the results shown in Figure 41 and Figure 42 it can be scene that the results are in order with the previously considered scenes. In particular the underestimation of lidar-ratio and extinction as was scene in the Halifax results is also evident in the results for this scene. It is also evident that the varying topography is handled in an appropriate fashion. As may be expected, the more disuse aerosol fields (mainly north of 60 DegN between 2 and 8 km or so) are not well retrieved.

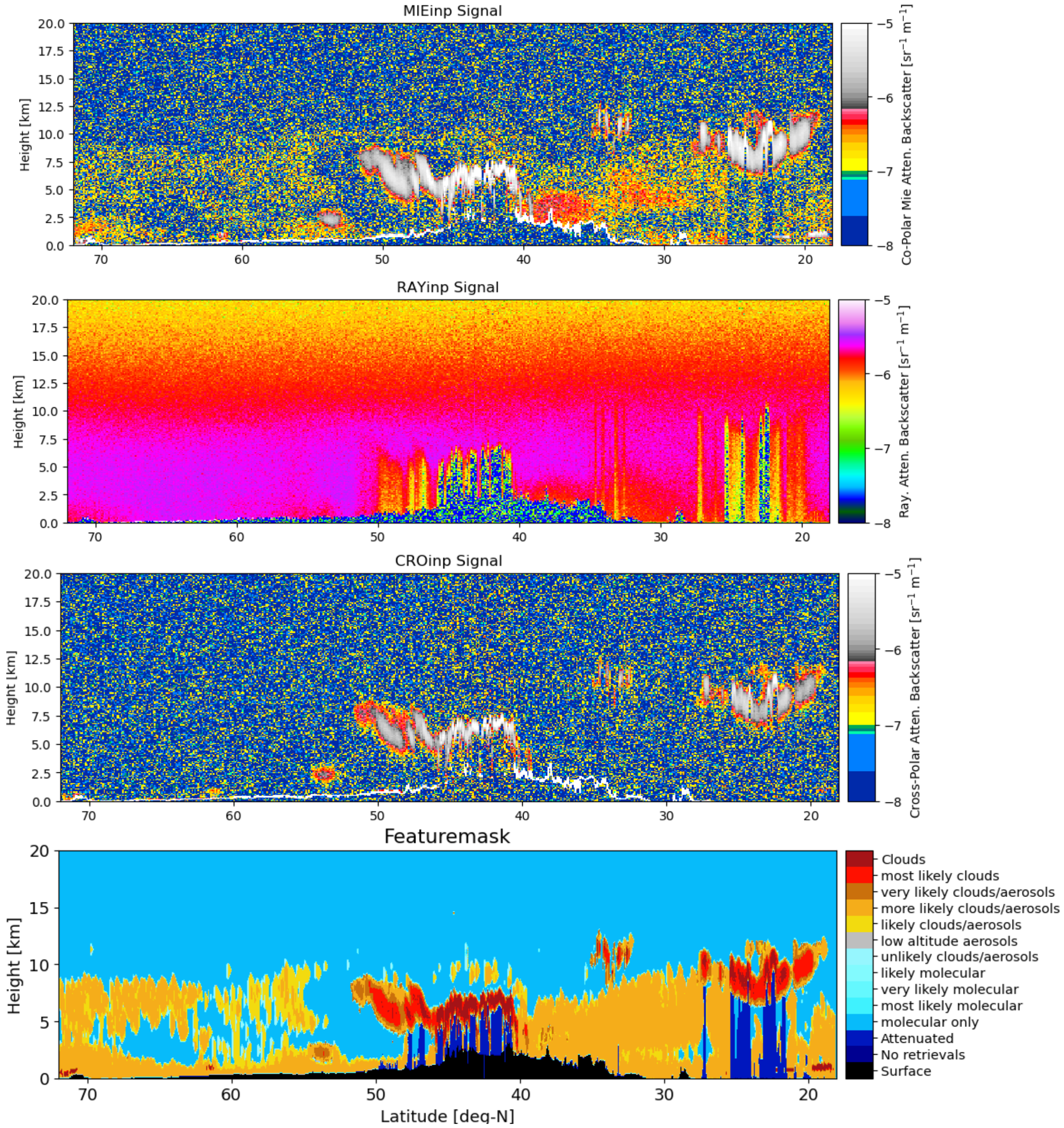


Figure 40: Simulated Mie, Rayleigh and Cross-polar attenuated backscatter fields along with the corresponding A-FM product for the Baja scene. The signals have been smoothed using a 10 pixel along-track window.

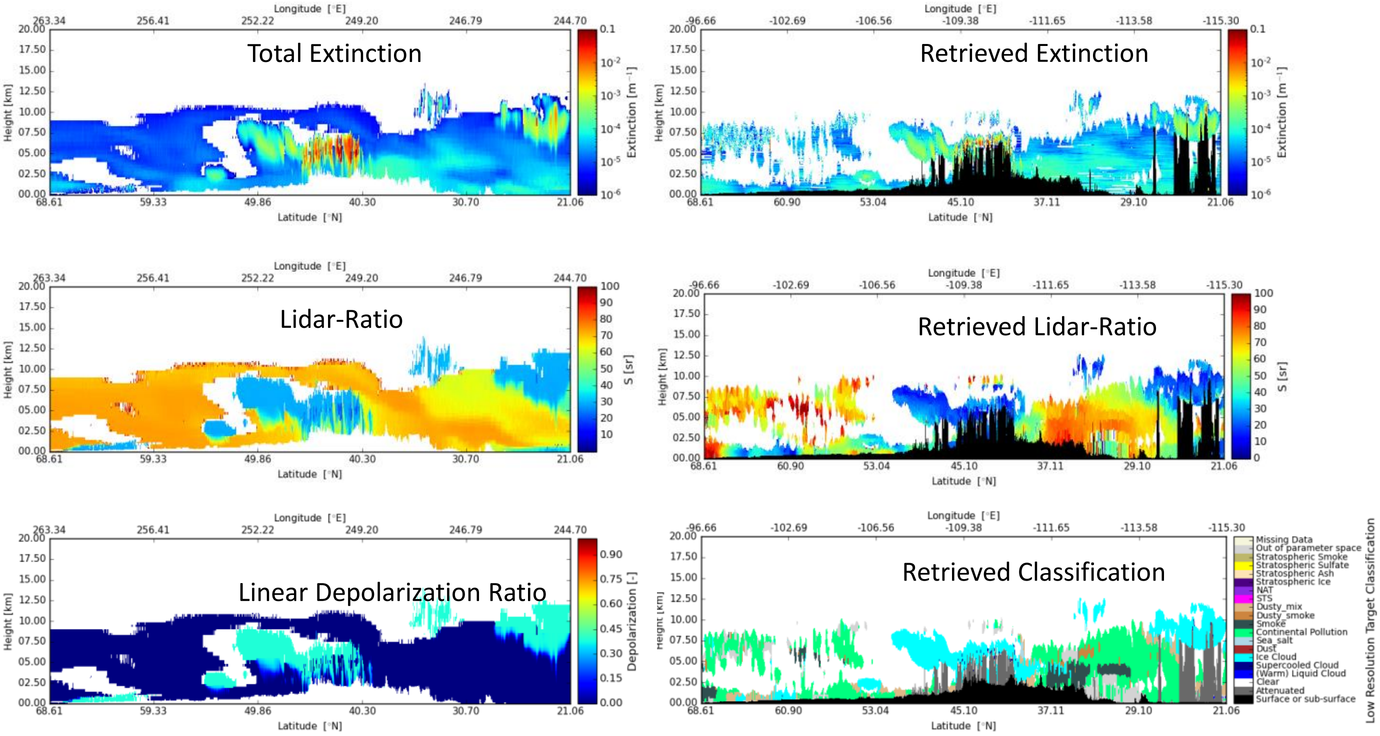


Figure 41: Sample retrieval results (Right) along with true fields (Left) for the Baja scene.

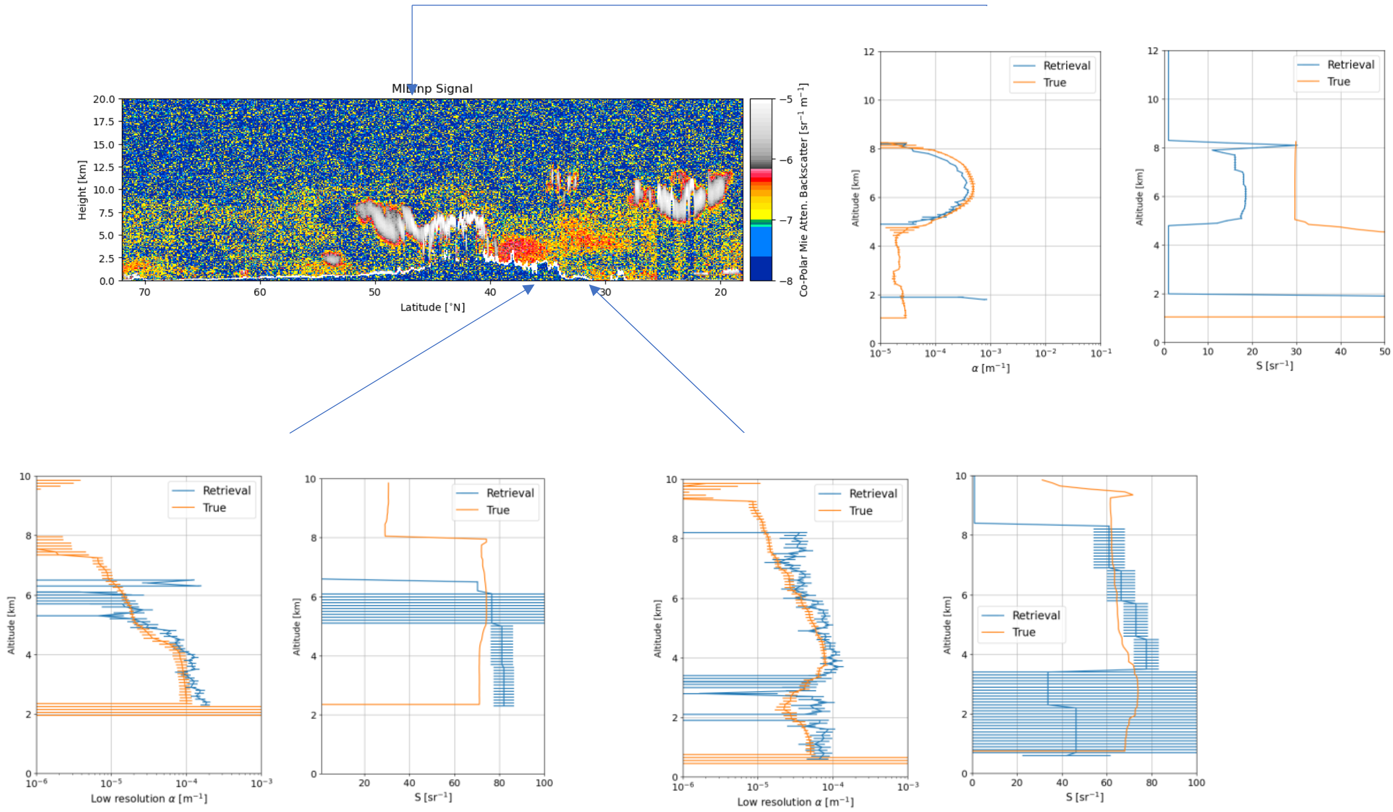


Figure 42: Sample profiles of retrieved extinction and lidar-ratio as well corresponding model truth profiles. Here the model truth profiles have been horizontally smoothed to match the resolution of the retrievals.

6.5 Hawaii scene Results

The 3rd GEM based scene is the Hawaii scene it is almost completely over ocean (nadir track is completely over ocean while a few of the smaller Hawaiian Islands are within the MSI track). Results are shown in a similar format to the previous scenes in Figure 43-Figure 45. The retrieval accuracies are consistent with the other scenes. However, here the classification algorithm tends to falsely detect Dust aerosols. The reason for this is unclear but is likely related to the large errors in the associated observed linear depolarization ratio. This low SNR is, in turn, related to the occurrence of higher-level cirrus (which attenuates the signal from the lower level aerosols) and the diffuse nature of the aerosols present in this scene.

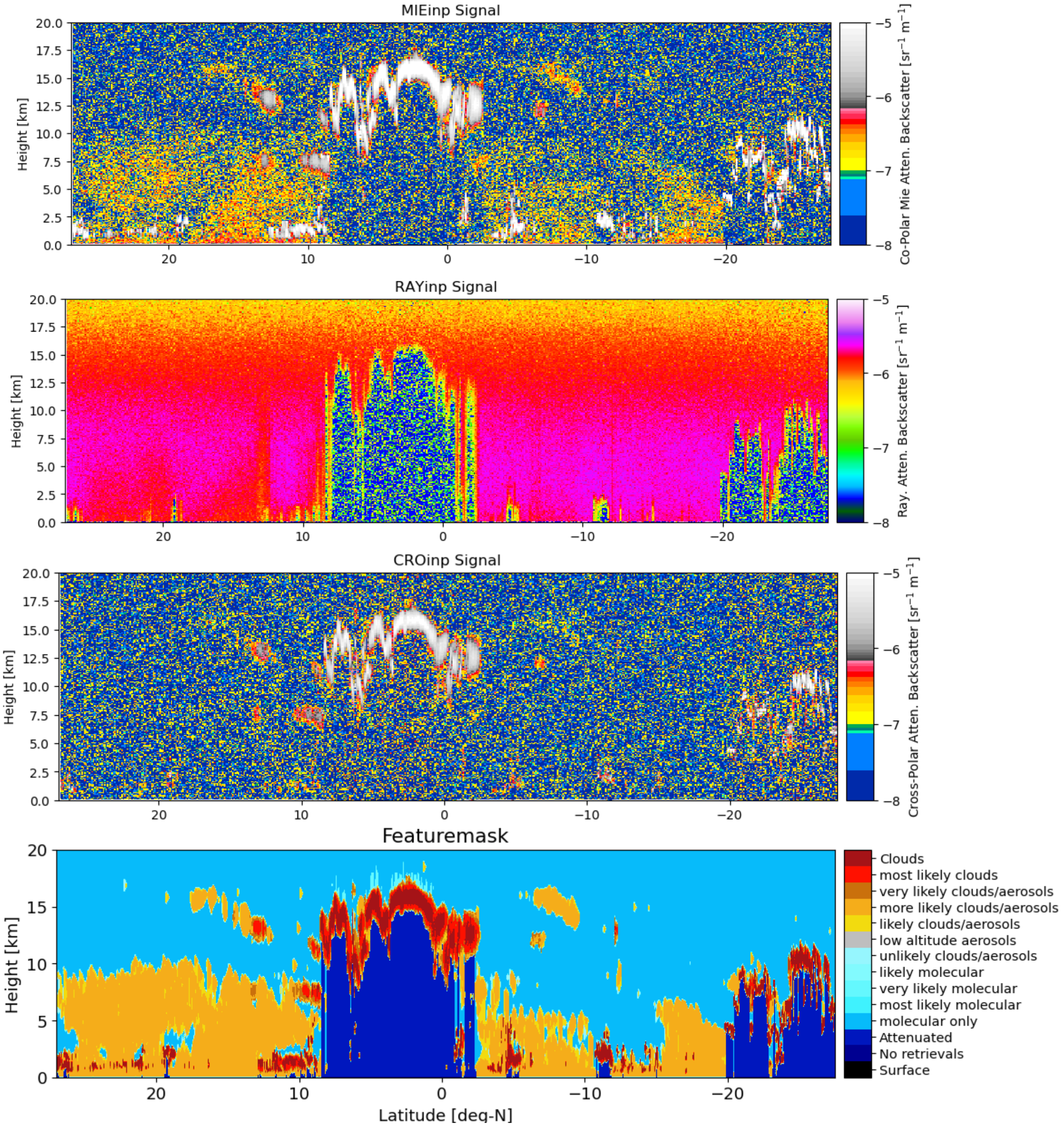


Figure 43: Simulated Mie, Rayleigh and Cross-polar attenuated backscatter fields along with the corresponding A-FM product for the Hawaii scene. The signals have been smoothed using a 10 pixel along-track window

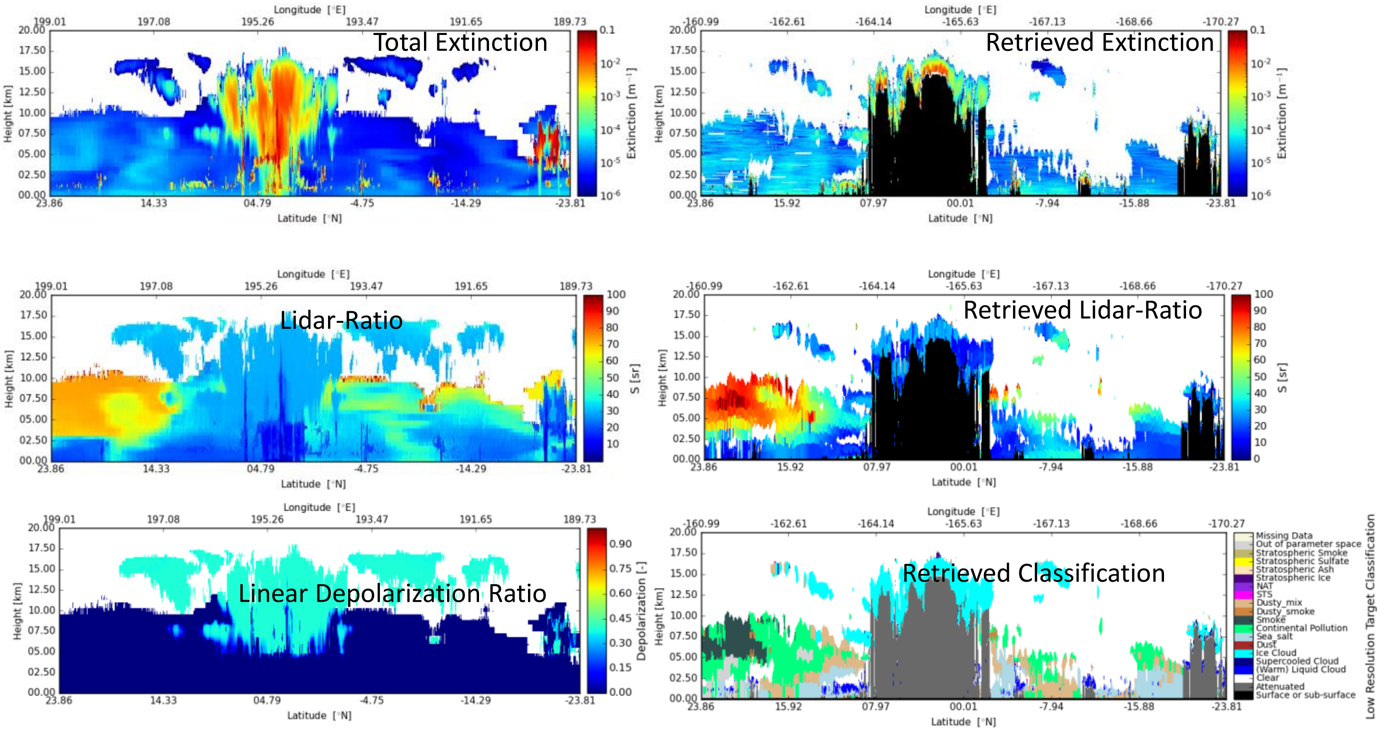


Figure 44: Sample retrieval results (Right) along with true fields (Left) for the Hawaii scene.

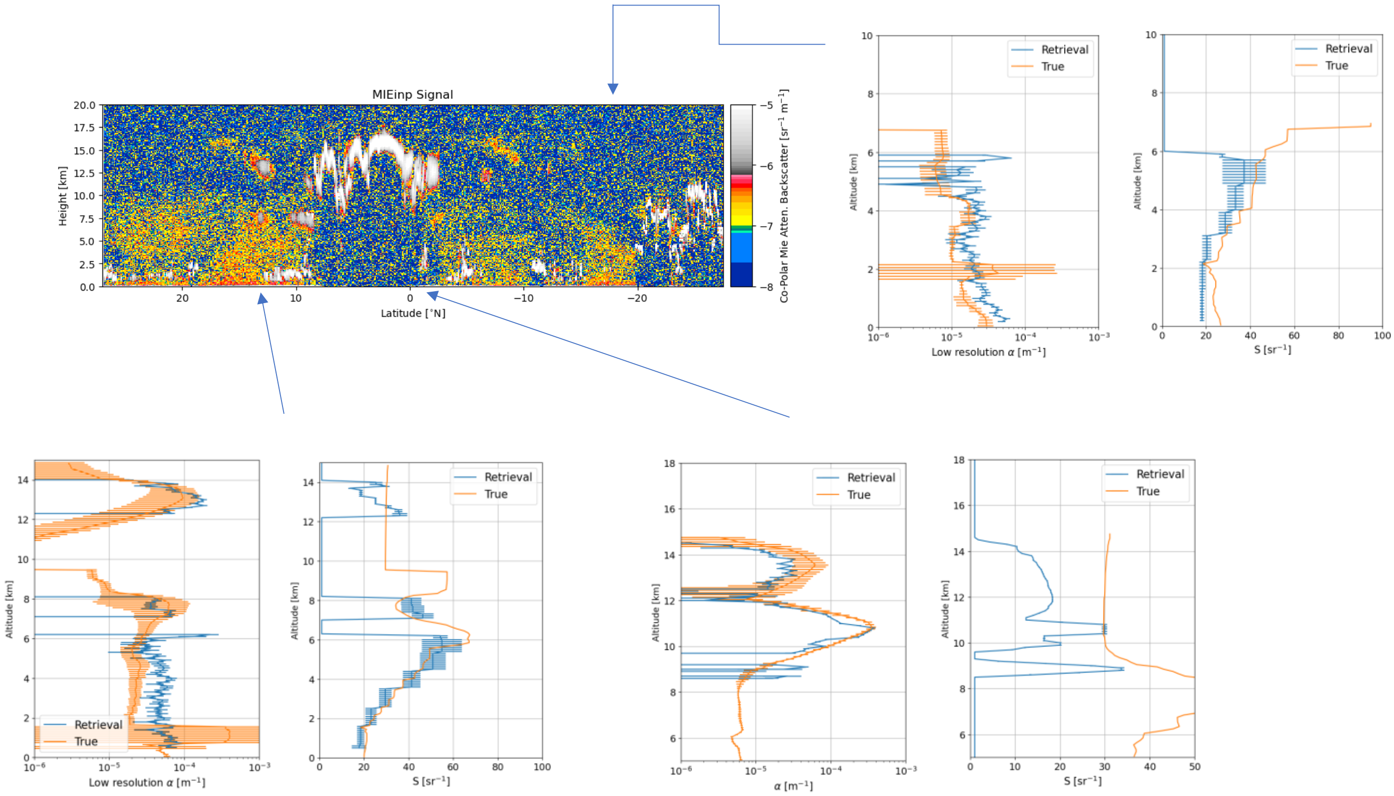


Figure 45: Sample profiles of retrieved extinction and lidar-ratio as well corresponding model truth profiles for the Hawaii scene. Here the model truth profiles have been horizontally smoothed to match the resolution of the retrievals.

6.6 PSC Scene Results

The last tests scene considered is not built upon GEM output. Rather, it is built using CALIPSO+CSAT retrieval products. It is quite unlike the other tests scenes in that it contains many stratospheric targets. Sample results for this scene are shown in Figure 46Figure 48. It can be seen that the classification product successfully discriminated between (spherical) STS and depolarizing PSCs. It can also be noted that the retrieved extinction can become much larger than the model truth in the last few range gates before the lidar signal is extinguished. This type of behavior is often associated with forward inversions. Here it should be noted that the reported errors in the extinction (see the extinction error bars in the lower left panel in Figure 48) are appropriately large.

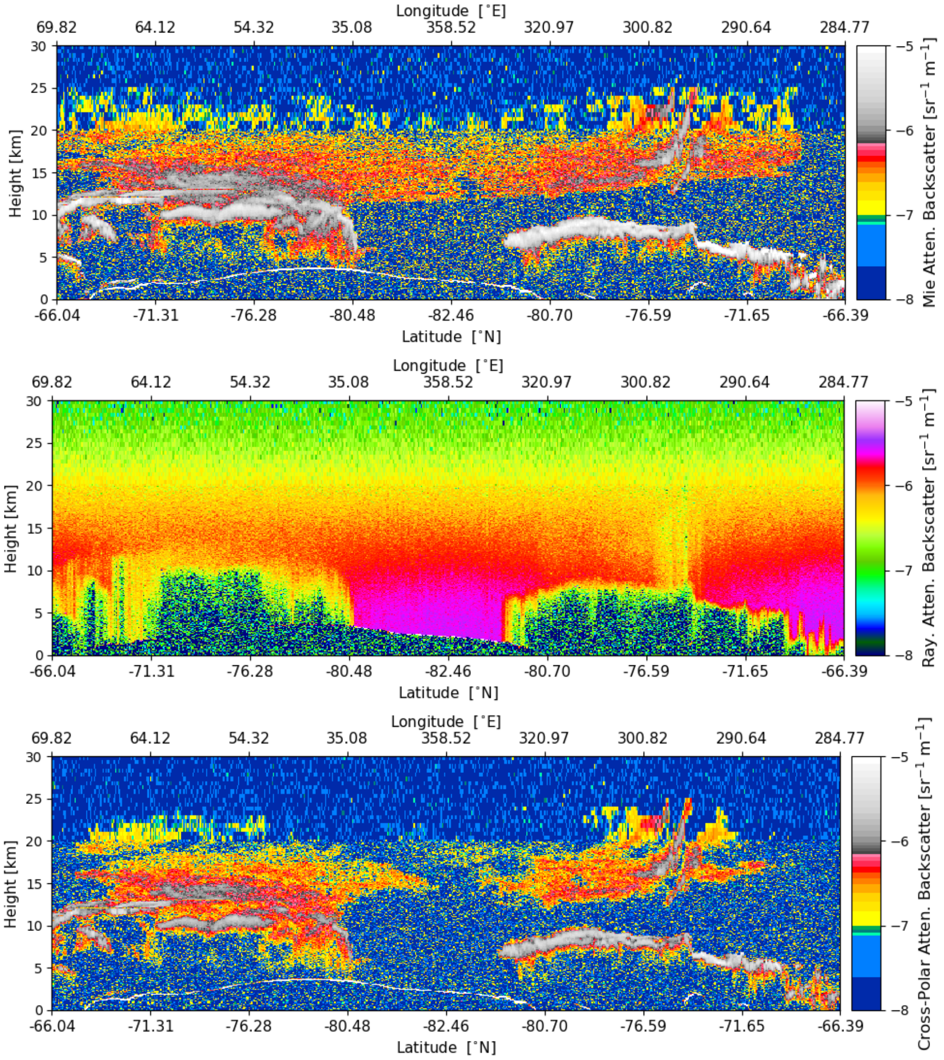


Figure 46: Simulated Mie, Rayleigh and Cross-polar attenuated backscatter fields for the PSC scene. The signals have been smoothed using a 10 pixel along-track window

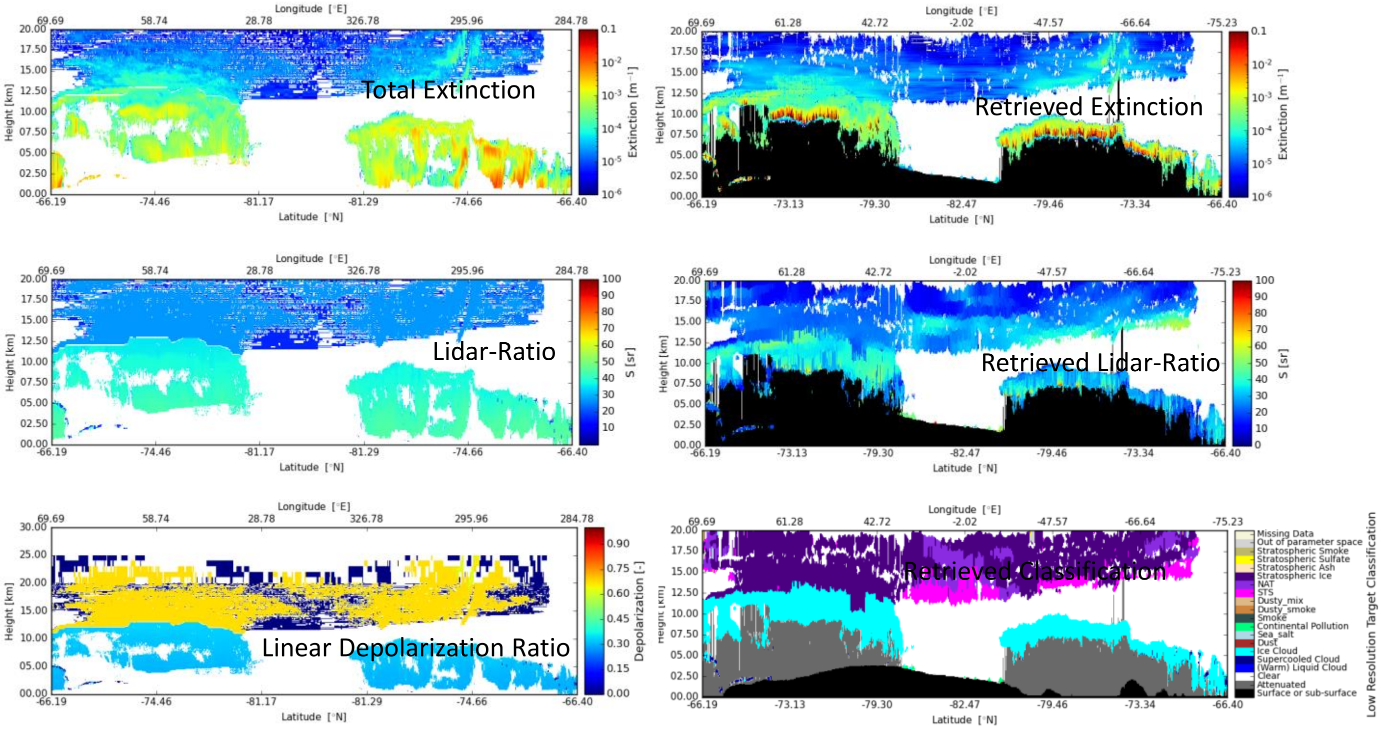


Figure 47: Sample retrieval results (Right) along with true fields (Left) for the PSC scene.

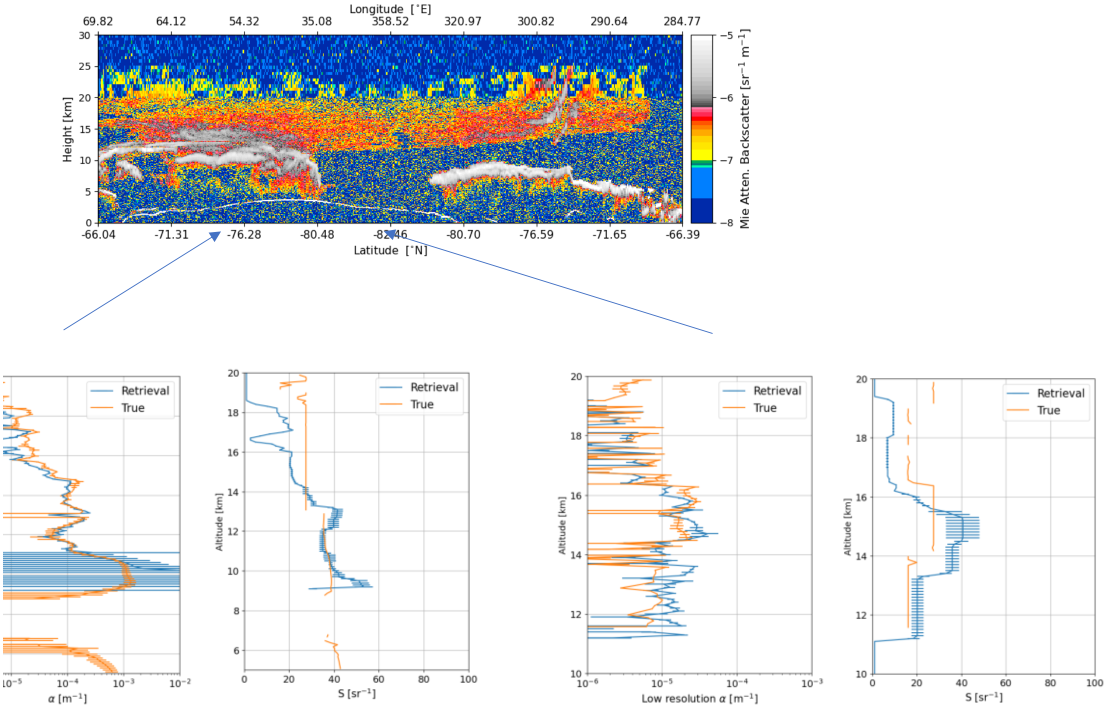


Figure 48: Sample profiles of retrieved extinction and lidar-ratio as well corresponding model truth profiles for the PSC scene. Here the model truth profiles have been horizontally smoothed to match the resolution of the retrievals.

6.7 Discussion

As mentioned at the start of this section the analysis presented here is to be considered preliminary. Those, preliminary, the comparison of the A-PRO retrieval results and the model truth has demonstrated that the procedure produces useful results and is robust across many situations (e.g. SNR, multi-layer, multi-target, degrees of horizontal variability etc.). In general, the lidar ratios are often retrieved within about -+ /10 Sr and extinction errors are on the order of 10-15% for values of extinction above 0.1 km^{-1}. However, biases seem to exist, particularly with respect retrievals in cirrus clouds. It seems this particular issue may be linked to too-low values of the lidar-ratio being estimated in the low-resolution AER part of the procedure. This, in turn may be due to unintended biases as a result of unsuitable averaging windows, multiple-scattering related issues, or even mundane bugs in the code. This issue along with others, continue to be investigated and further evolution/improvements of the A-PRO retrieval procedure will result.

7 APPENDIX A: TEST SCENES

Here a brief overview of the methods and data sources used to generate the tests scenes used in the ATBD is given.

7.1 Radiative Transfer Calculations

To generate the lidar signals the ECSIM lidar Monte-Carlo model was used (Donovan et al, 2015). This model treats polarization and its interaction with multiple-scattering. After the Monte-Carlo calculations were completed an instrument model that accounts for e.g. the instrument collection efficiency, the effects of background and dark current noise, was applied.

To generate the MSI and BBR radiances DISORT was applied. To save computational time not every pixel was calculated. Instead, a large number of sub-domains were established and only about 1 in 10 pixels were selected per sub-domain. The selection was carries out such that the full range of optical depths present in each respective sub-domain was spanned and each distinct surface type was covered. The pixels that were not selected were then filled in using the nearest sampled pixel.

For the DISORT calculations, a Cox-Muck type ocean surface parameterization was used and an RTTOV like surface scheme implemented.

7.2 The MACC(CAMS) scene

ECMWF MACC(CAMS) fields were used to construct a Frame sized (approx. 6000x200km) scene. The original fields are at a low 0.5x0.5 Deg. resolution. Using pseudo-random number and then smoothing to achieve “realistic” correlations scales the ECMWF fields were downscaled to a 500 meter horizontal resolution. These processed are schematically depicted in Figure 49 and Figure 50.

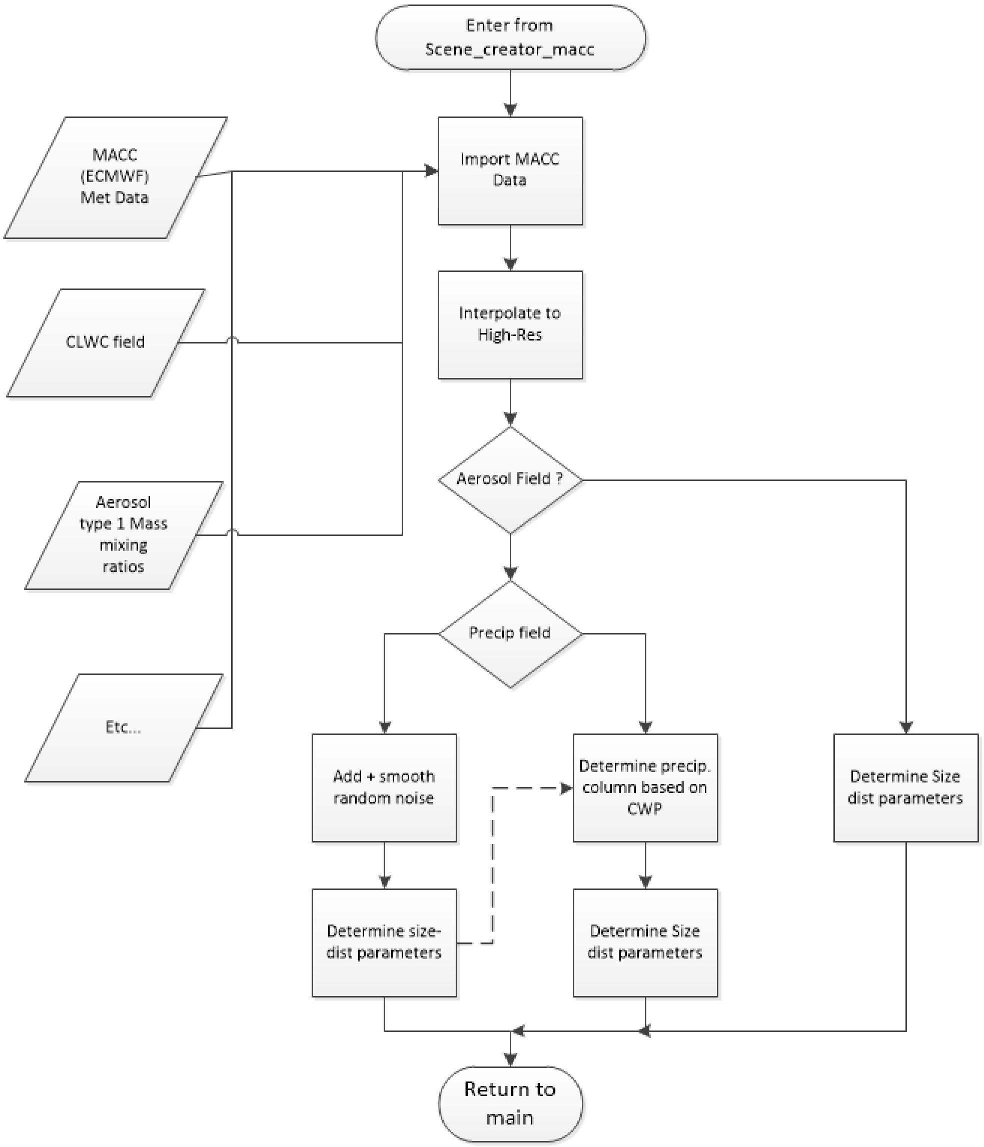


Figure 49: Schematic depiction of the process of building the MACC scene.

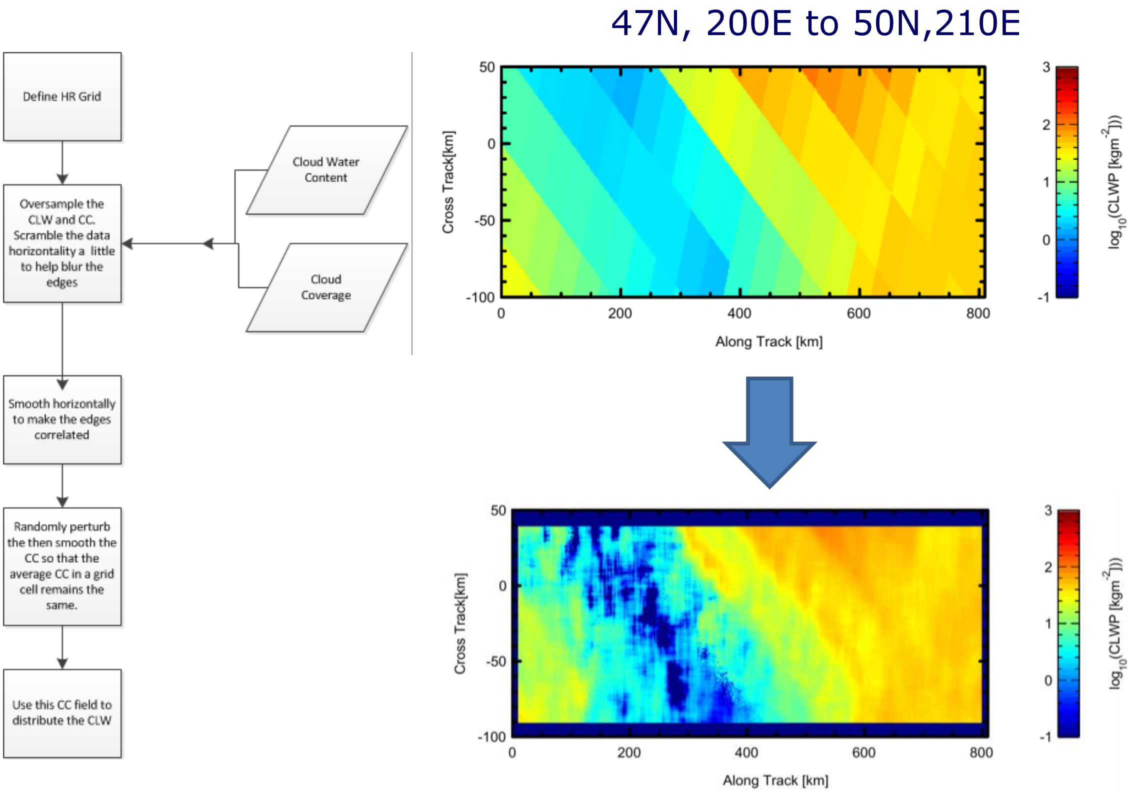


Figure 50: Schematic depiction of the downscaling of the ECMWF fields.

ECSIM simulated ATLID L1 results are shown in Figure 51 along with CALIPSO attenuated backscatter observations. An impressive coincidence between the features observed by CALIPSO and those that are visible in the ECSIM ATLID Mie channel is evident. It is interesting to note that this correspondence exists in in spite of the fact that CALIPSO data was not ingested into the ECSIM simulations or the input MACC data. The correspondence observed lends a degree of confidence in both the input MACC data as well as the ECSIM simulation process itself.

In Figure 52 the extinction fields corresponding to the different scattering types contained in the MACC scene are presented and the corresponding Lidar-Ratios are shown in Figure 53. The scattering species contained in the scene are listed in Table 6: Scattering types present in the MACC scene.. In order to build the scene a mapping between the MACC species and the ECSIM scene, a mapping between the MACC mass content fields and the ECSIM scattering types has to be made. Details of this mapping are given in Table 7. Within this table, the various MACC fields used as a basis to define the ECSIM scattering fields are as follows:

1. CIWC➔Cloud Ice Water Content
2. CSWC➔Cloud Snow Water Content
3. CLWC➔Cloud Liquid Water Content
4. CRWC➔Cloud Rain Water Content
5. DD1-DD3➔Dust (in different size intervals)
6. SS1-SS3➔Seat Salt (in different size intervals)
7. SO4➔Sulphate aerosol
8. BCB➔Fine mode strongly absorbing aerosol
9. OMB➔Weakly absorbing aerosol

The extinction and lidar ratios per component are presented in Figure 52 and Figure 53 respectively.

No.	Scattering type	ECSIM scattering type
1	Ice Cloud	Baum Aggregated solid columns
2	Snow	Baum General Habit mixture
3	Cloud Water	Water
4	Rain	Rain
5	Coarse mode non-spherical absorbing aerosol	Heteac Coarse Dust (spheroids)
6	Coarse mode non-absorbing aerosol	Heteac Coarse Salt
7	Fine-mode weakly absorbing aerosol	Heteac Fine Weak
8	Fine-mode strongly absorbing aerosol	Heteac Fine Strong

Table 6: Scattering types present in the MACC scene.

MACC Field	CIWC	CSWC	CLWC	CRWC	DD1-DD3	SS1-SS2	SO4	BCB	OMB
Dry-to-Wet mass factor	1	1	1	1	1	2	2	2	2
Reff (sigma,gama)									
Ice Cloud	5um(-40C)➔100um(0C) Gamma=3								
Snow		10um(10C)➔1000um(0C) Gamma=8							
Water Cloud			Reff determined via No and CLWC Gamma=8						
Rain				20um (0C)➔650um (40C) Gamma =3					
Coarse Dust					1.0um Sigma=0.3	1.1um Sigma=0.3	1.1um Sigma=0.3	1.1um Sigma=0.3	1.1um Sigma=0.3
Coarse Salt						1.94um Sigma=0.6			

Fine Weak						0.14um Sigma= 0.53	0.14um Sigma= 0.53		
Fine Strong								0.14um Sigma= 0.53	0.14um Sigma= 0.53
No Rmin,Rmax									
Ice Cloud	No determined via Rm and IWC 0.0—2000.0 um								
Snow		No determined via Reff and IWC 0.0—2000.0 um							
Water Cloud			Fixed at 500/cm3						
Rain				No determined via Reff and RWC 0.0—2000.0 um					
Coarse Dust					No determined via Reff and Aerosol mass 0.0—2000.0 um	Via Reff 0.0→100 um	Via Reff 0.0→100 um	Via Reff 0.0→100um	Via Reff 0.0 →100um
Coarse Salt						Via Reff 0.0→100 um			
Fine Weak						Via Reff 0.0→100 um	Via Reff 0.0→100 um		
Fine Strong								Via Reff 0.0→100um	Via Reff 0.0 →10 um

Vol. Fraction									
Ice Cloud	1								
Snow		1							
Cloud Water			1						
Rain				1					
Coarse Dust					1	9	5	3	3
Coarse Salt						90			
Fine Weak						1	95		
Fine Strong								97	97

Table 7: Scattering type size characteristics and their relationship with the MACC mass fields.

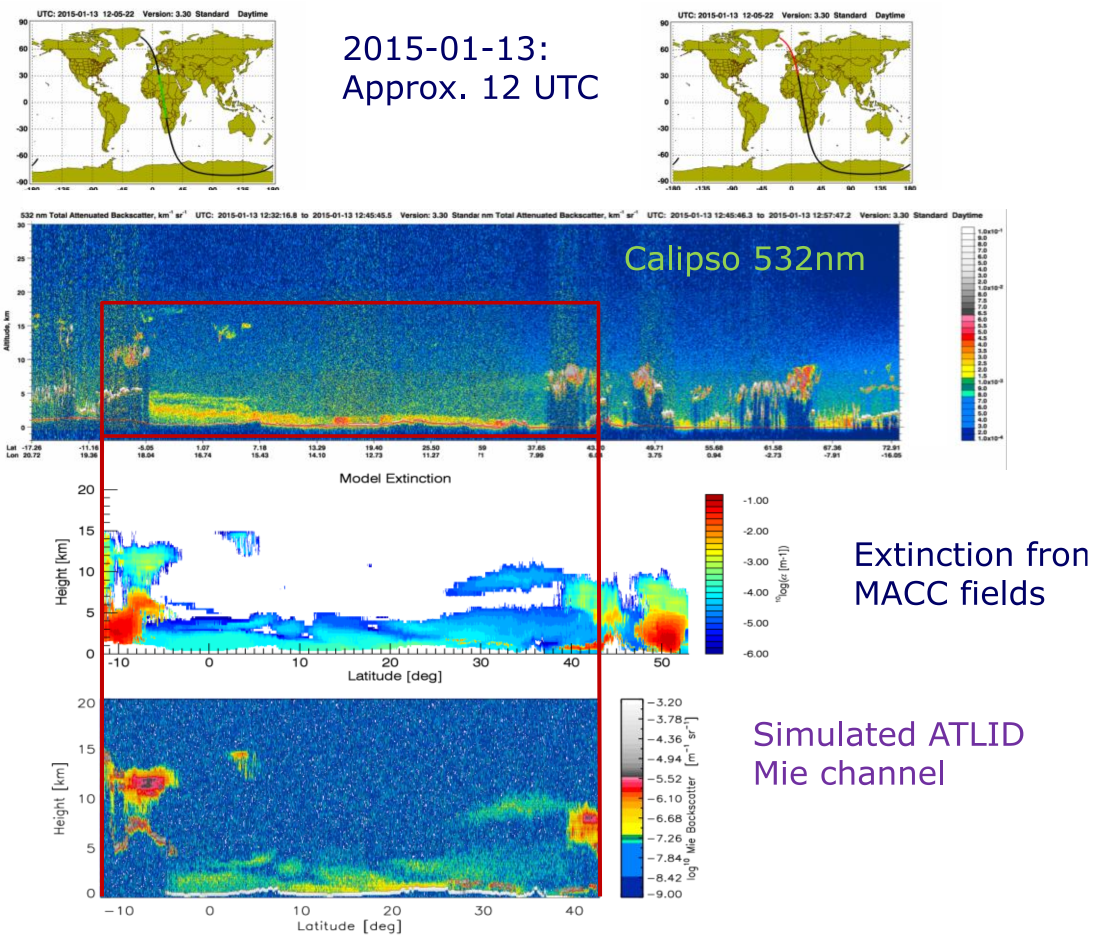


Figure 51: Calipso data following the selected extracted MACC data and the corresponding scene total extinction field at 355nm and the corresponding simulated ATLID Mie channel attenuated backscatter field.

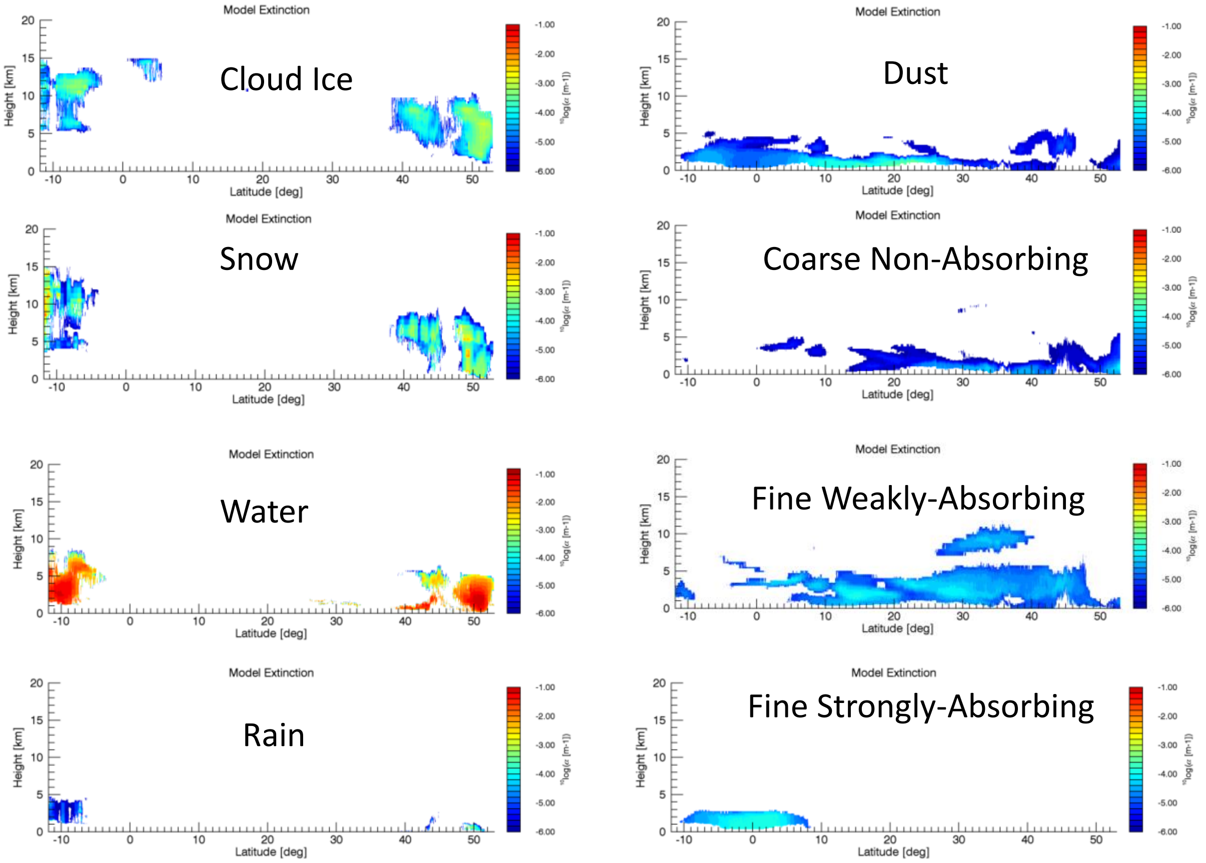


Figure 52: True extinction fields for the different scattering species contained in the MACC scene.

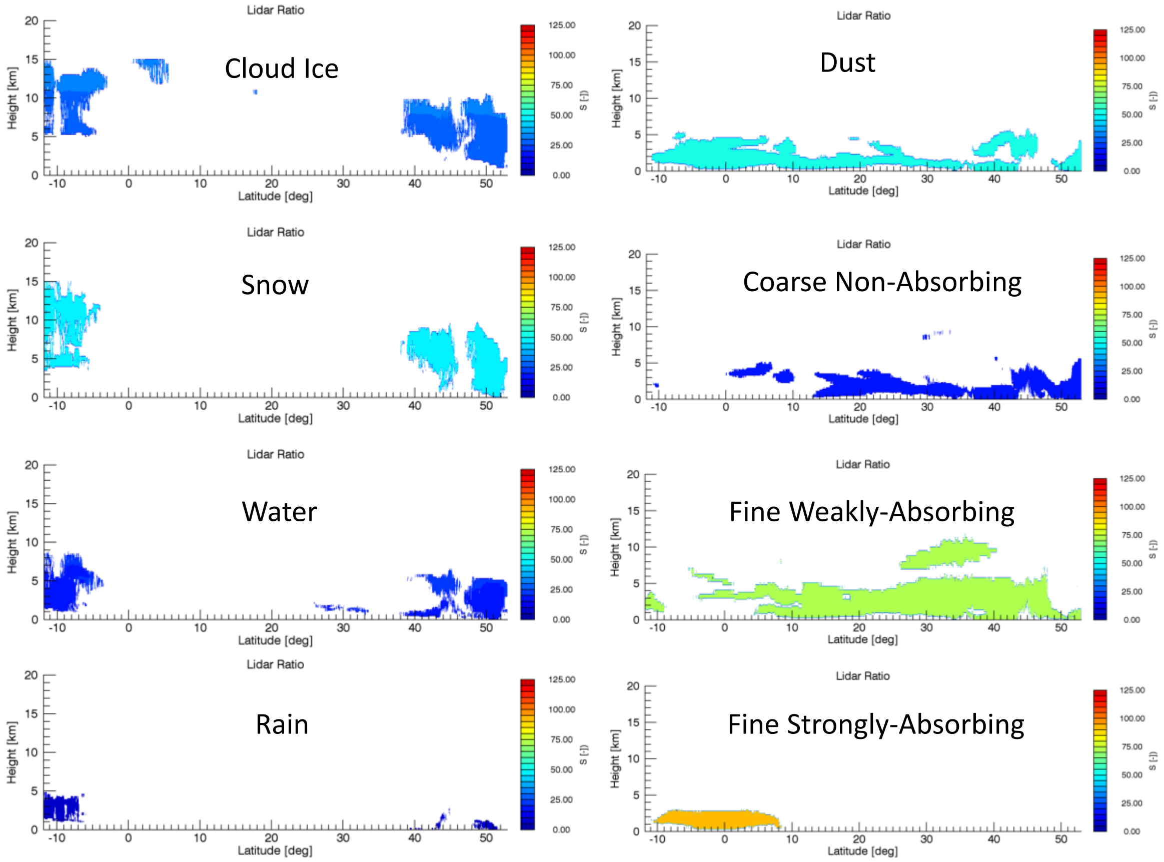


Figure 53: Lidar-ratio for the different scattering species contained in the MACC scene.

7.3 GEM derived SCENES

In this section the three scenes built using ECCC GEM data are described. Simulations based on the ECCC global forecast mode (GEM) along with CAMS aerosol fields were used as the basis for creating simulated scenes. Multiple high-resolution (0.25 km hor. Res.) GEM segments were `stitched together' to form 6000km+ simulated tracks.

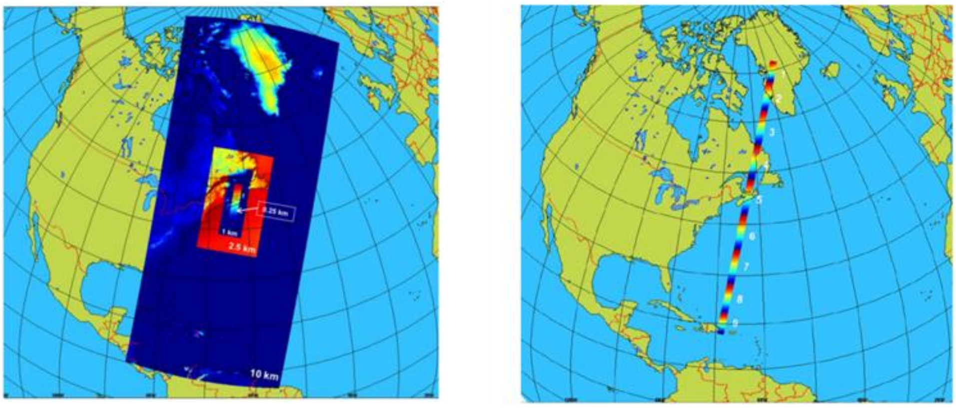


Figure 54: The ECCC GEM model is capable of very high resolution, however only for a limited domain (Left panel). To create the scenes the high-resolution sections of several runs were spliced together (Right panel).

In the GEM model, Clouds/precip. are handled by a two-mode bulk (Milbrandt-Yau) scheme. In order to include realistic aerosol fields the GEM information was merged with corresponding CAMS(MACC) aerosol model data using a procedure similar to that used in the construction of the MACC scene. The phase function for the aerosol and ice/snow were the same as those used in the MACC scene.

7.3.1 Adjustments to Hydrometer fields

During the course of development, it was found that strictly following e.g. the mass-vs-size relationships and size distribution assumptions embedded in the GEM ice/snow fields lead to e.g. too many too-small ice particles and e.g. unrealistic particle fall-speeds and IWC -vs- radar reflectivity.

Thus, a number of adjustments to the MACC hydrometer fields were carried out.

- Ice size distributions were multiplied by D^4 with the number density scaled to preserved mass.
- M(D) and A(D) relationships adjusted to Erfani & Mitchell (2016) for Ice and Heymsfield et al., (2013) for Snow mass.
- The 'implicit' ice fields were deleted.

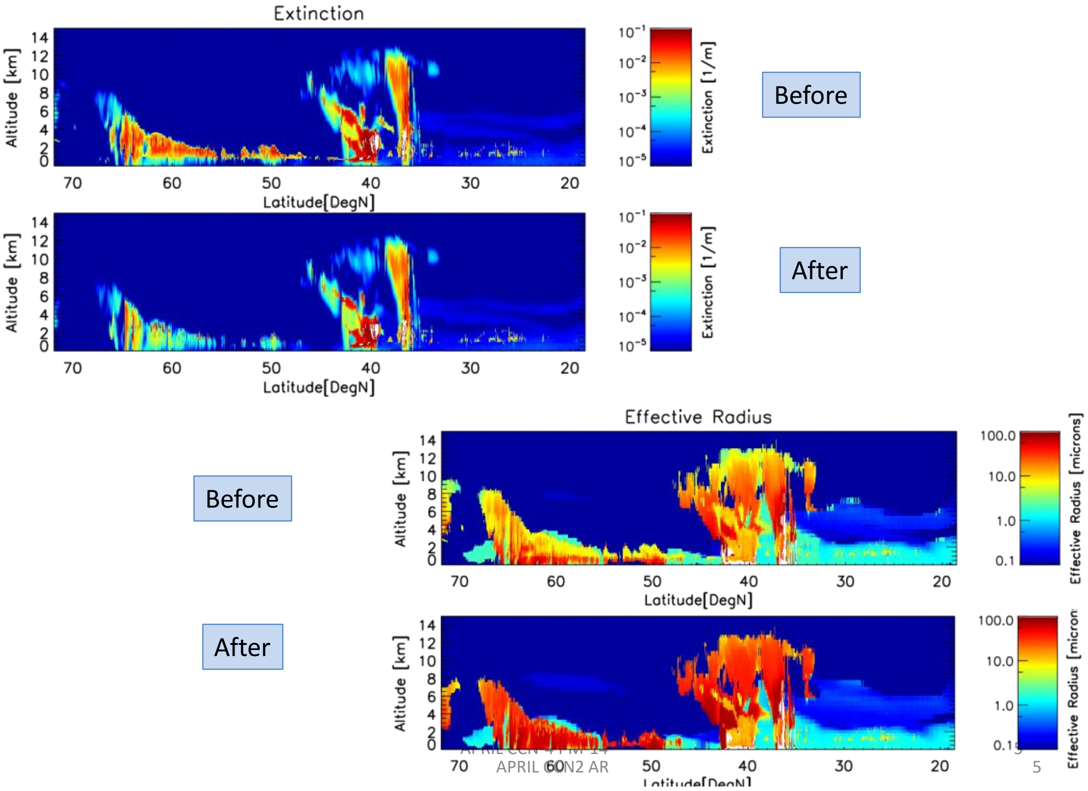


Figure 55: Extinction and effective particle radius before and after adjustment for the Halifax scene.

As a result of these adjustments, e.g. the distribution the ice particle size-vs-temperature was brought much more in line with distributions observed in nature.

7.3.2 The Halifax Scene

The following figures given an overview of the horizontal, vertical, and per-species structure of the Halifax scene.

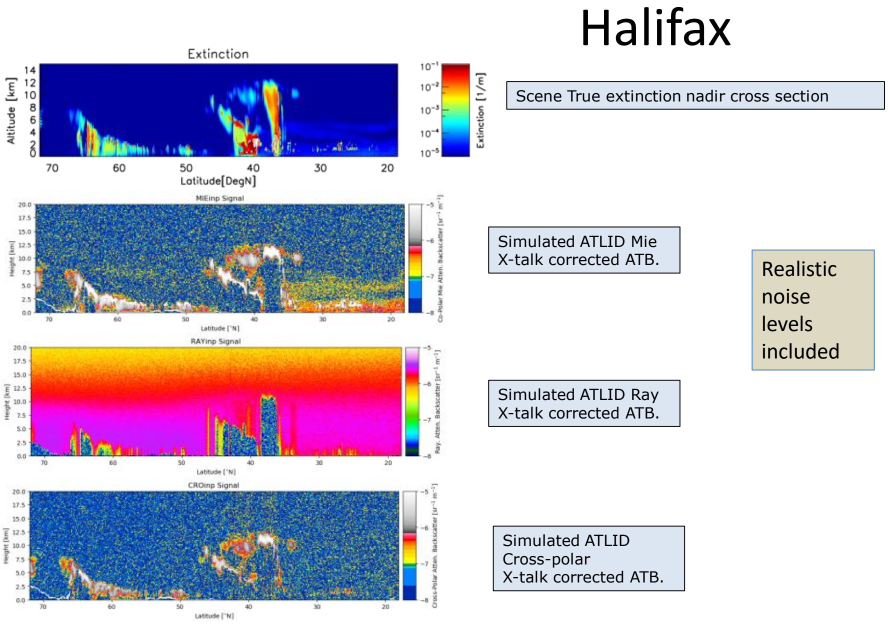


Figure 56: Nadir 355nm extinction and the simulated ATLID attenuated backscatters for the Halifax scene.

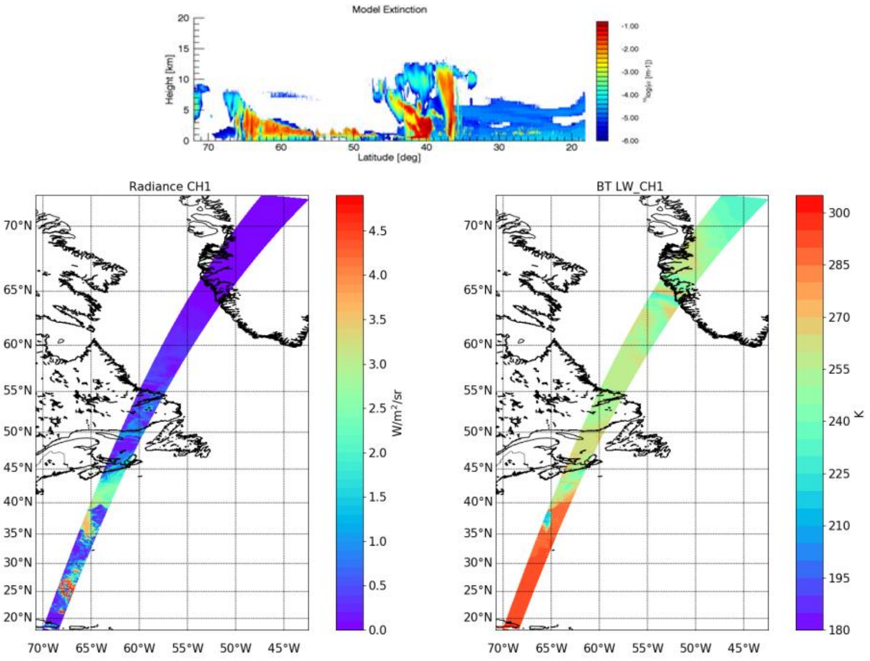


Figure 57: Nadir extinction and MSI CH1 radiances and LW-CH1 brightness temperatures.

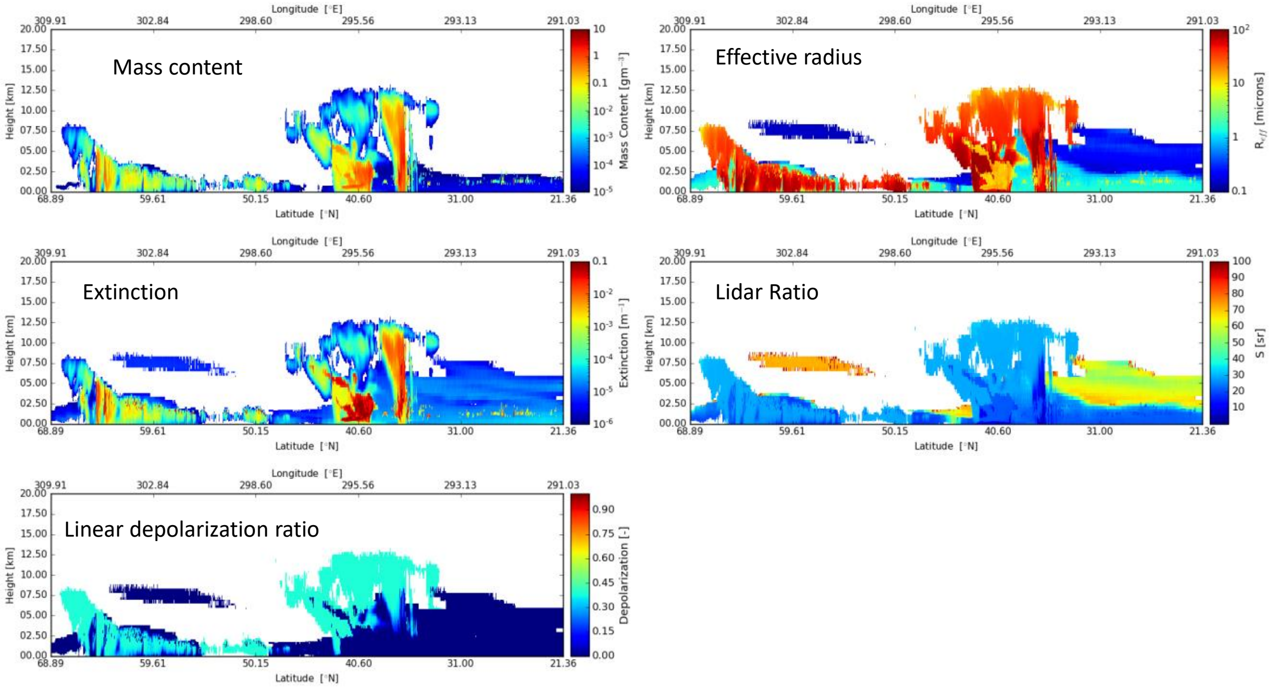


Figure 58: Total Mass content, effective radius, Extinction (355nm), lidar-ratio (355nm) and linear depolarization ratio (355nm) for the Halifax scene at Nadir.

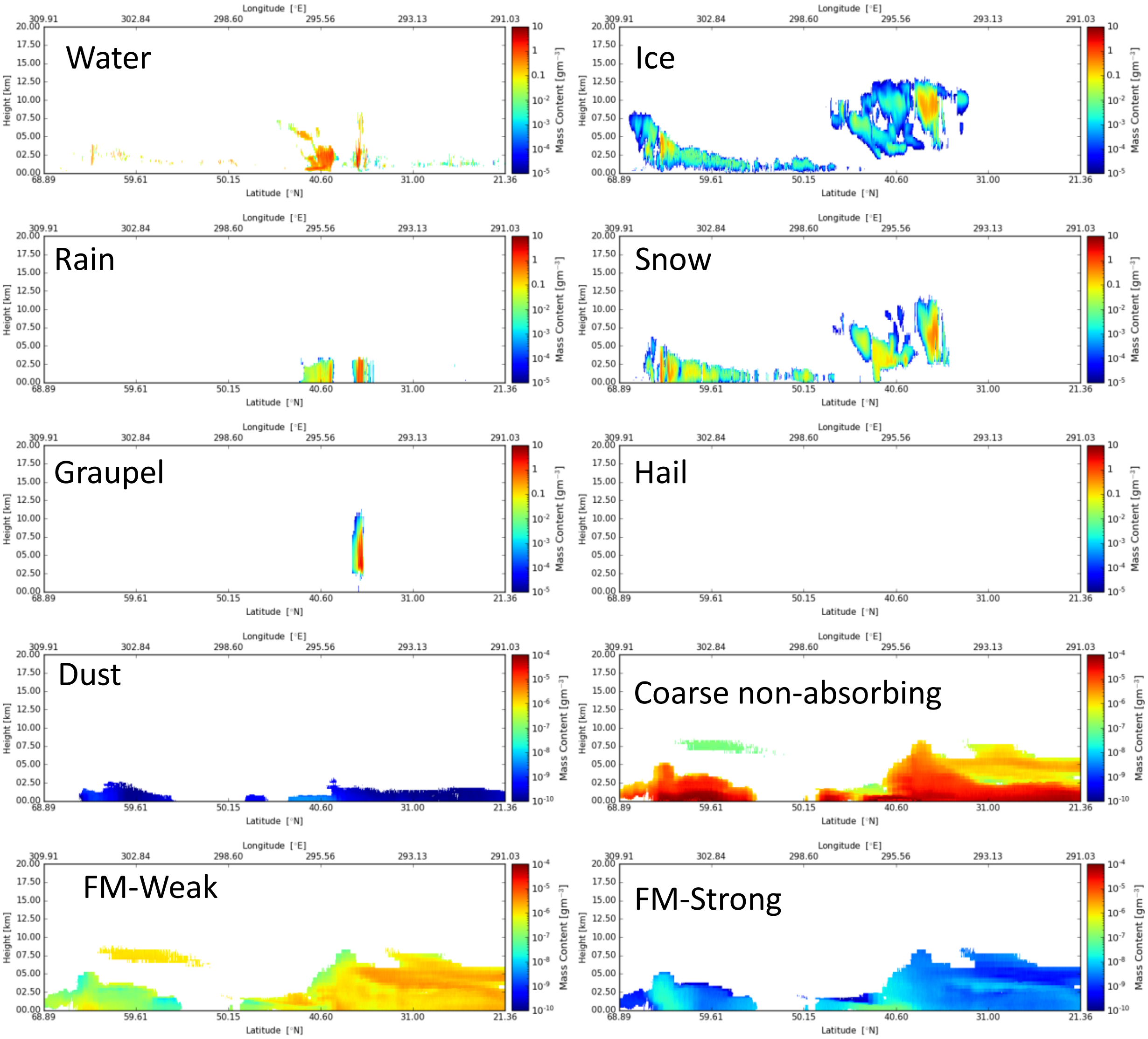


Figure 59: Mass content per species for the Halifax scene.

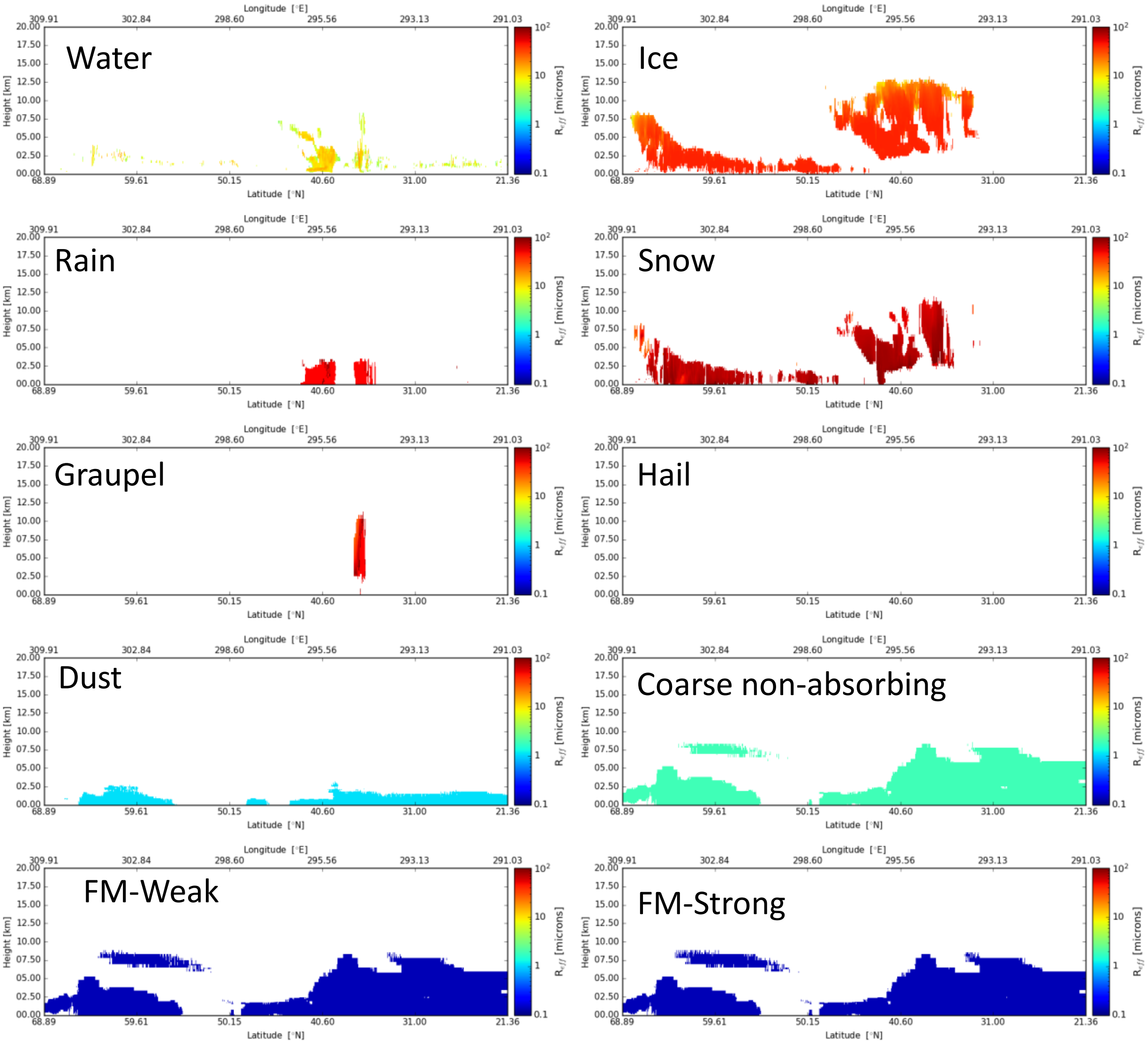


Figure 60: Effective Radius per species for the Halifax scene.

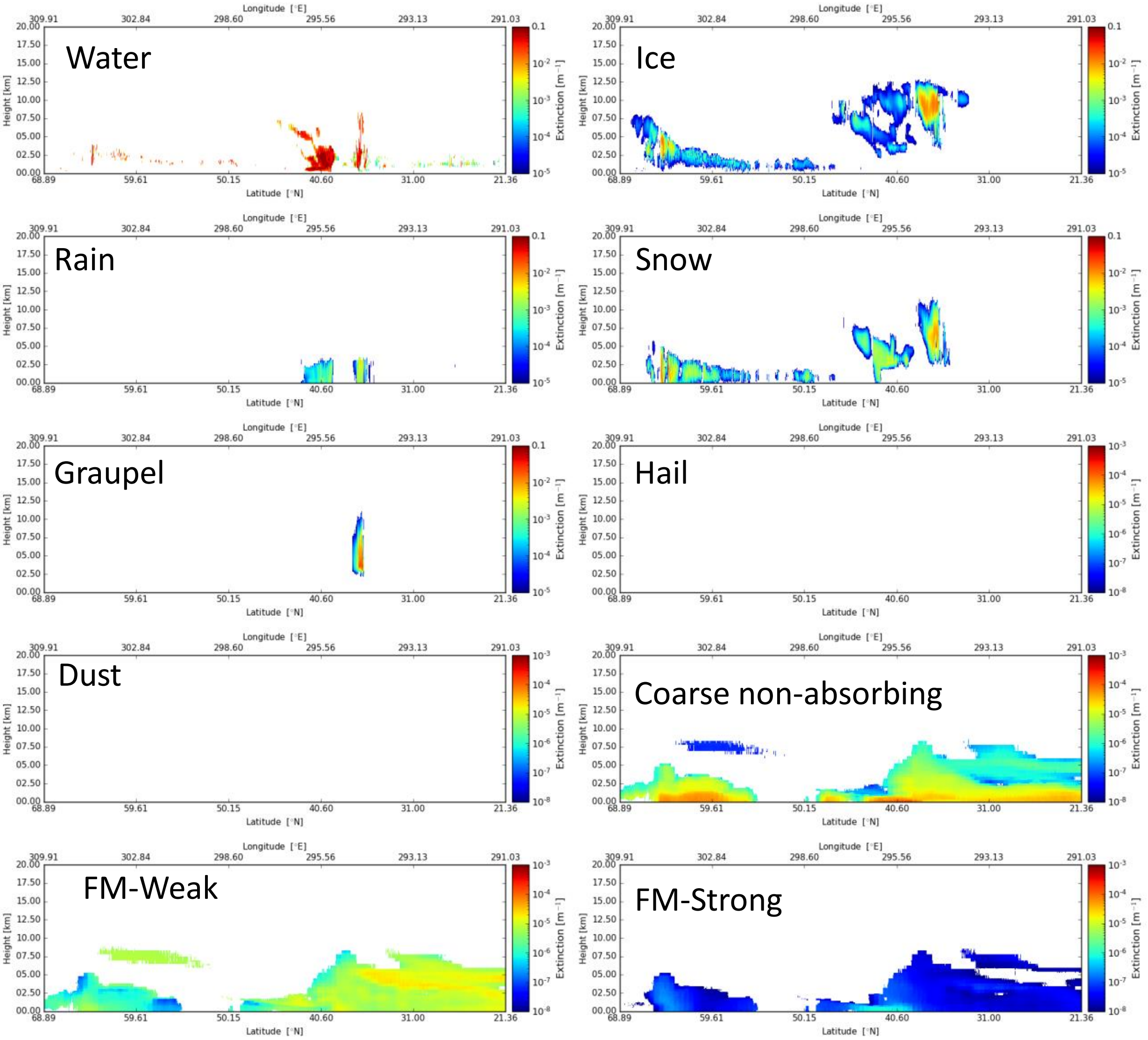


Figure 61: Extinction at 355nm per species for the Halifax scene.

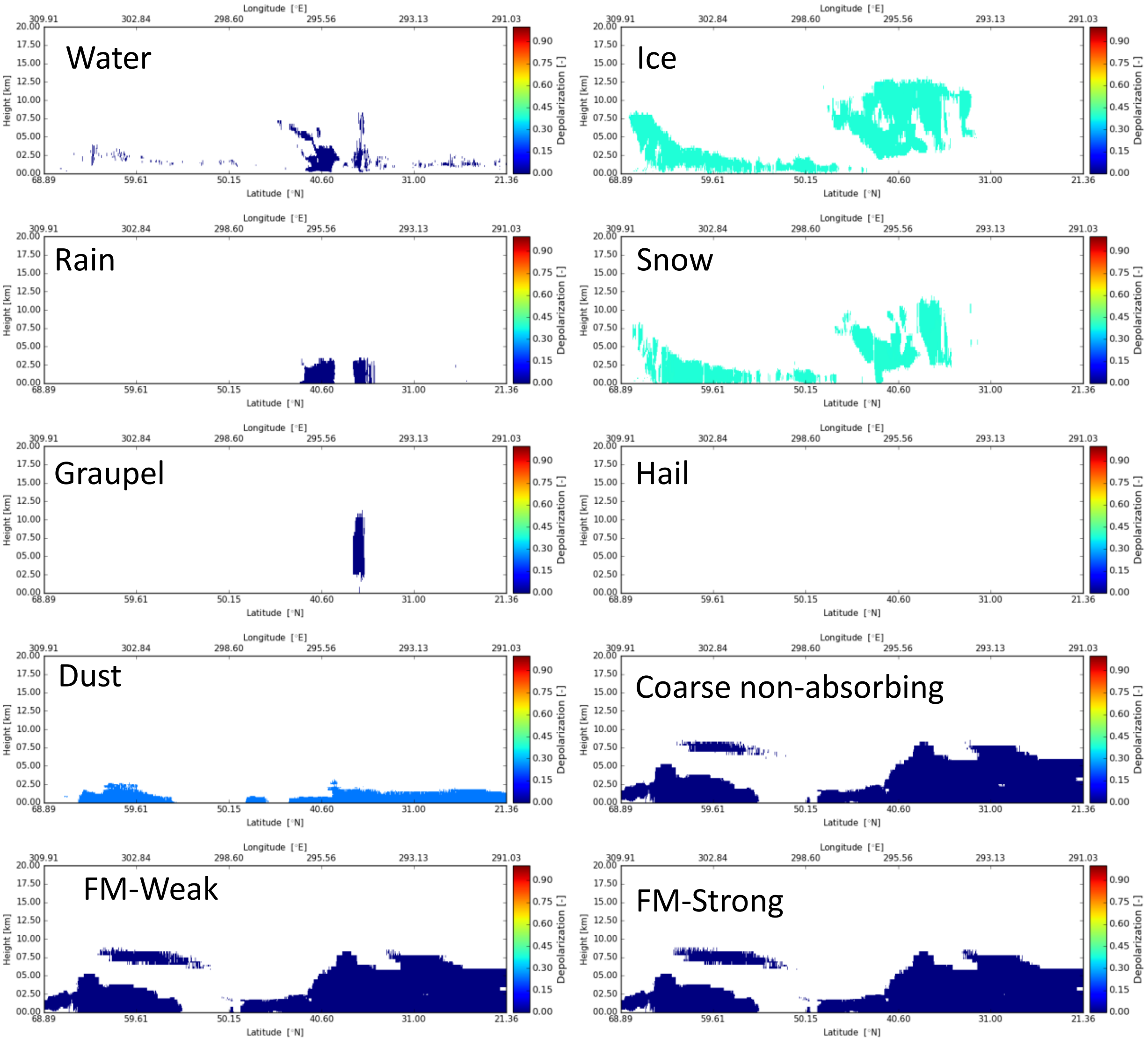


Figure 62: Linear depolarization ratio at 355nm per species for the Halifax scene.

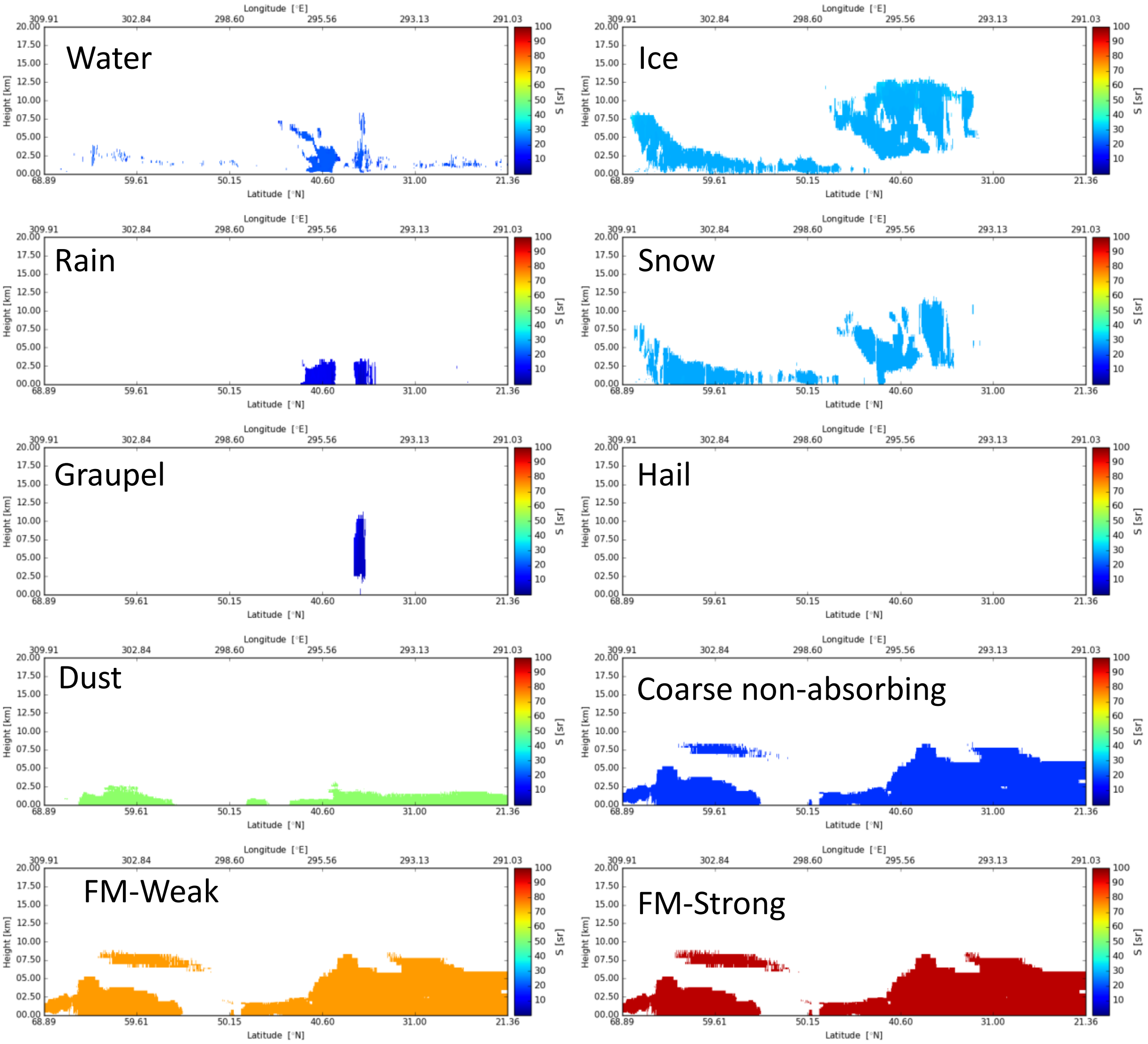
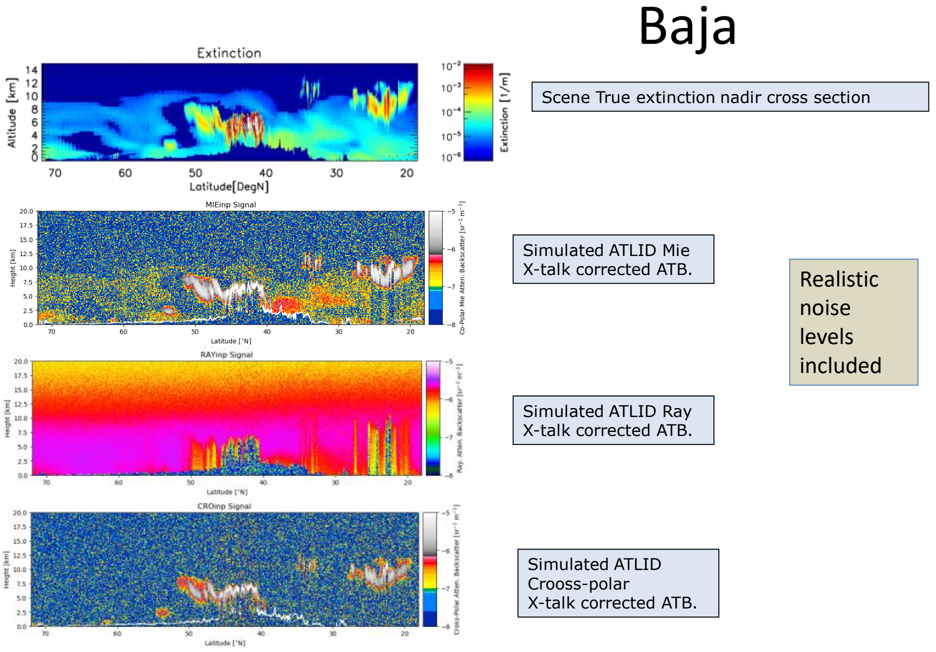


Figure 63: Lidar-ratio per species for the Halifax scene.

7.3.1 *The Baja Scene*

The following figures given an overview of the horizontal, vertical, and per-species structure of the Baja scene.


Baja

Scene True extinction nadir cross section

Simulated ATLID Mie X-talk corrected ATB.

Simulated ATLID Ray X-talk corrected ATB.

Simulated ATLID Crooss-polar X-talk corrected ATB.

Realistic noise levels included

Figure 64: Nadir 355nm extinction and the simulated ATLID attenuated backscatters for the Baja scene.

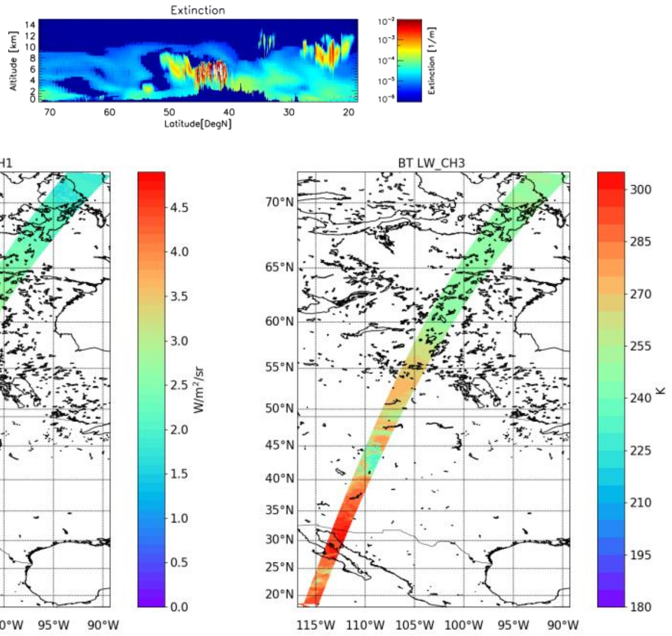


Figure 65: Nadir extinction and MSI CH1 radiances and LW-CH1 brightness temperatures.

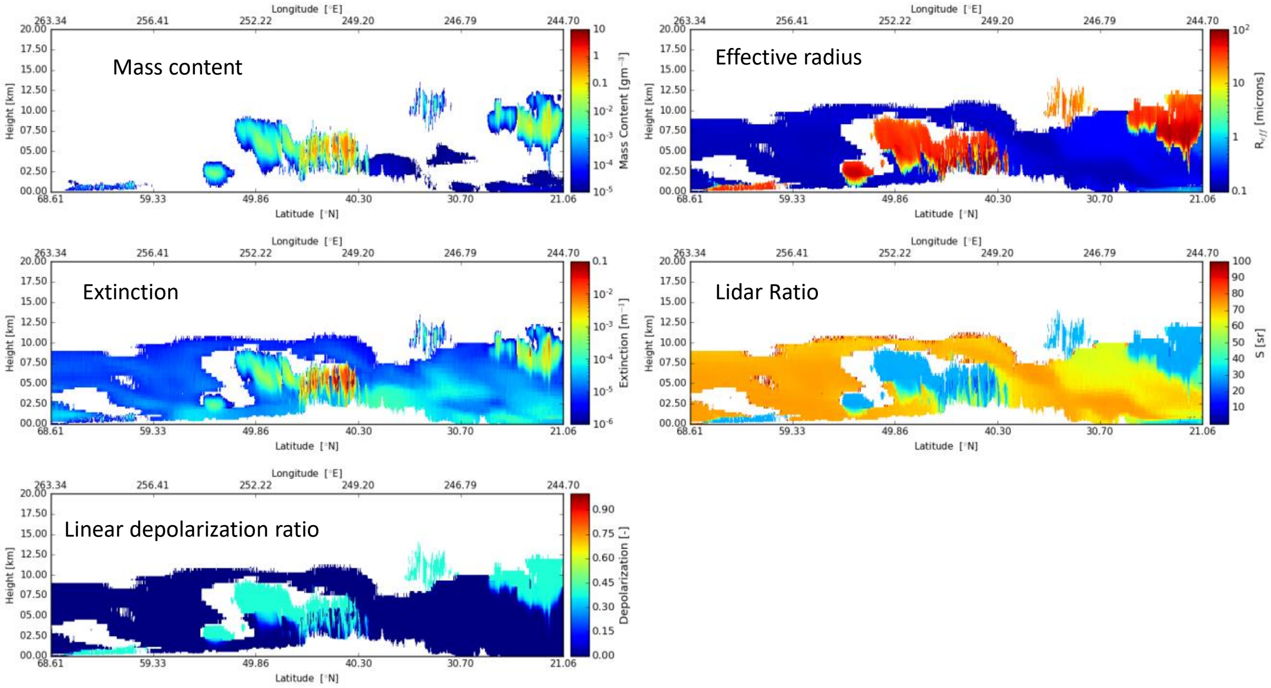


Figure 66: Total Mass content, effective radius, Extinction (355nm), lidar-ratio (355nm) and linear depolarization ratio (355nm) for the Baja scene at Nadir.

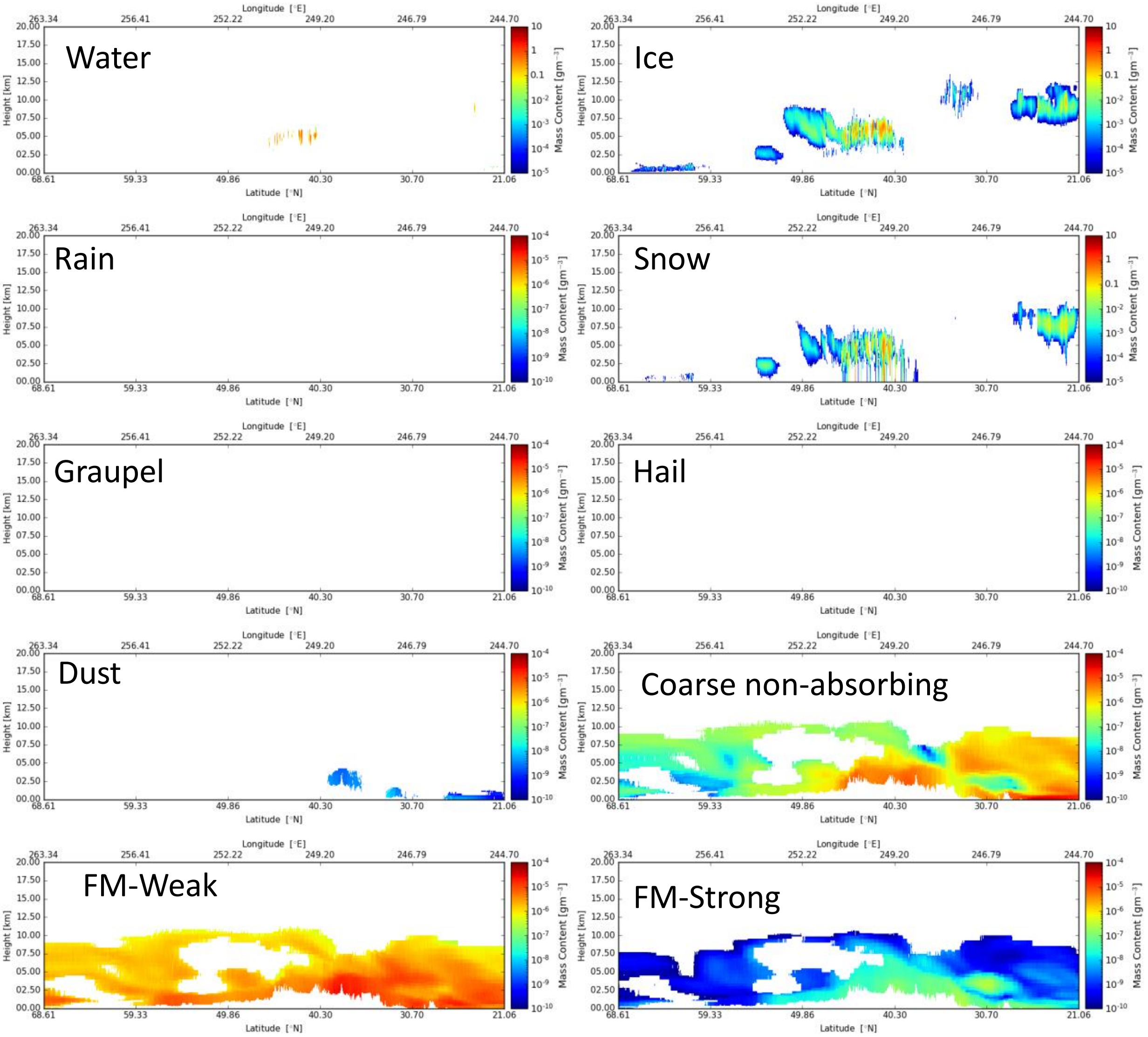


Figure 67: Mass content per species for the Baja scene.

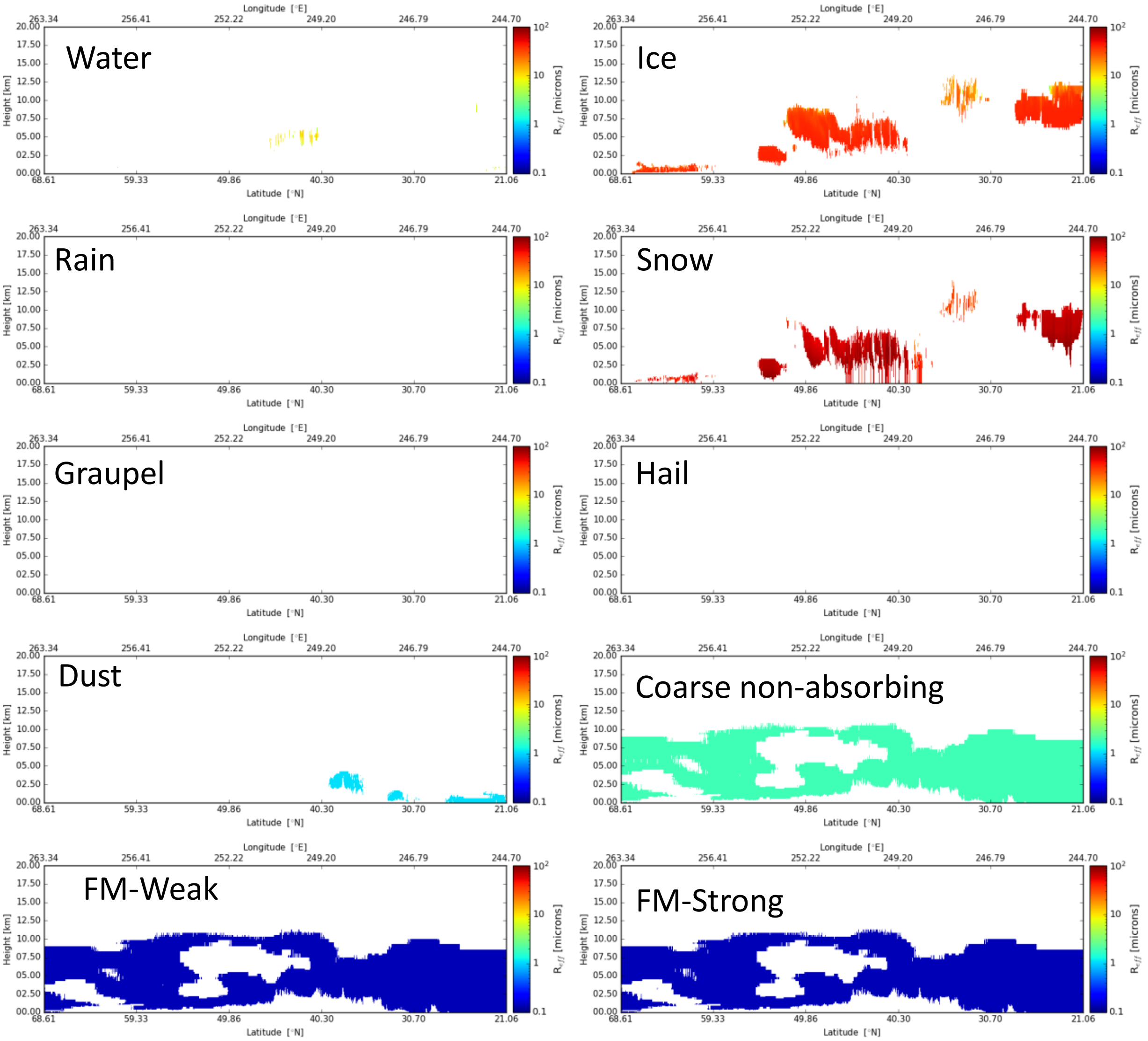


Figure 68: Effective Radius per species for the Baja scene.

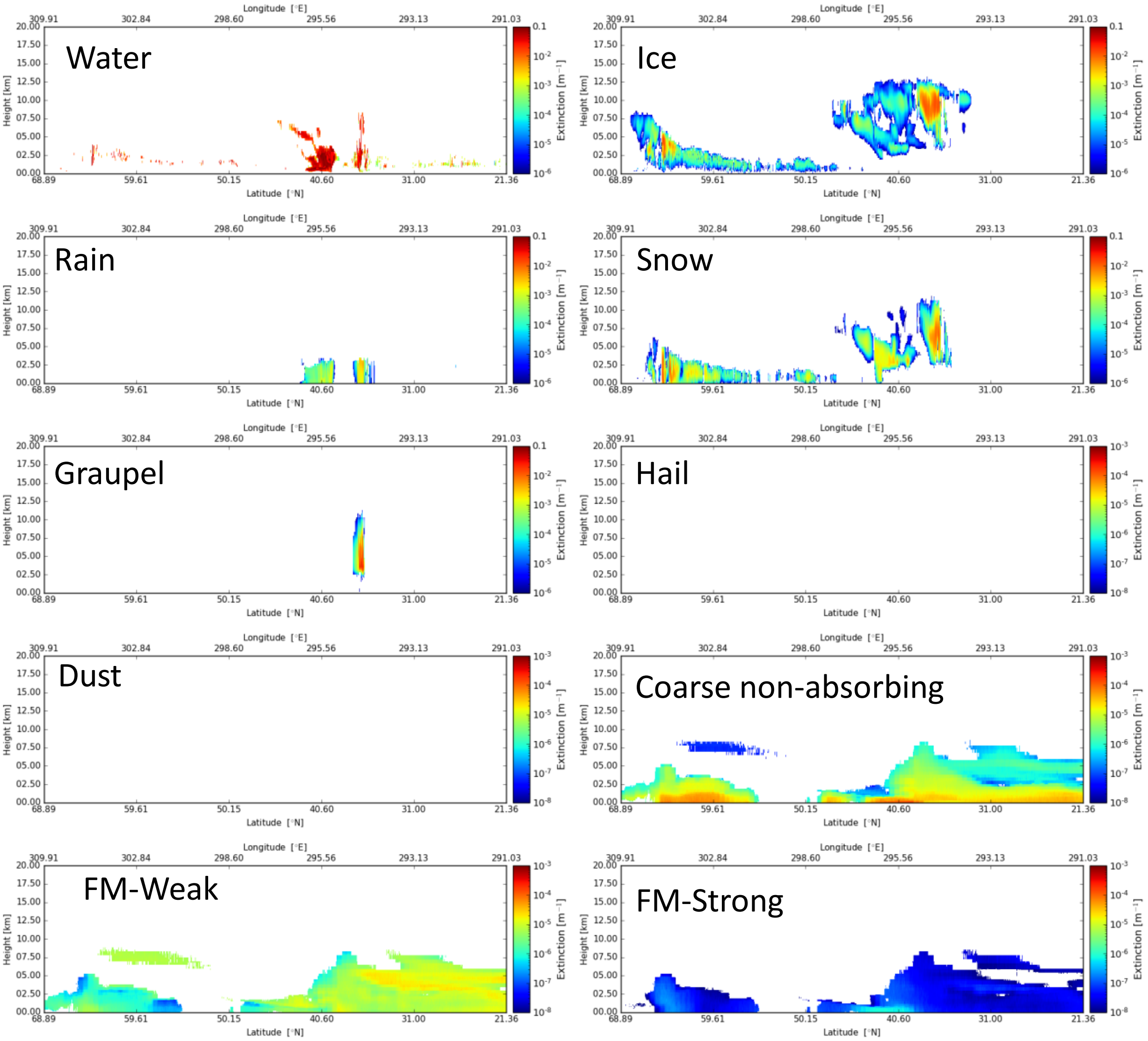


Figure 69: Extinction at 355nm per species for the Baja scene.

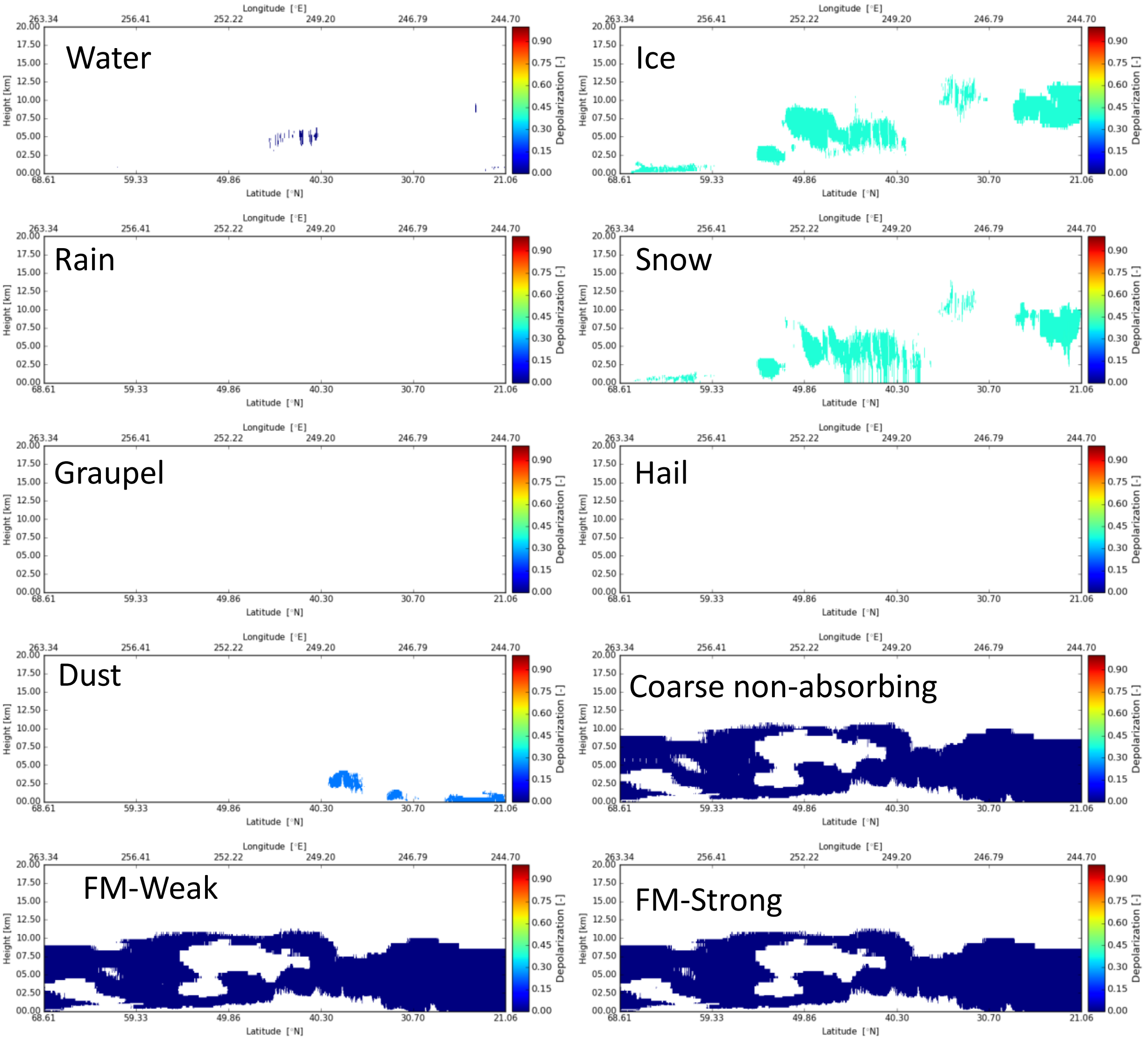


Figure 70: Linear depolarization ratio at 355nm per species for the Baja scene.

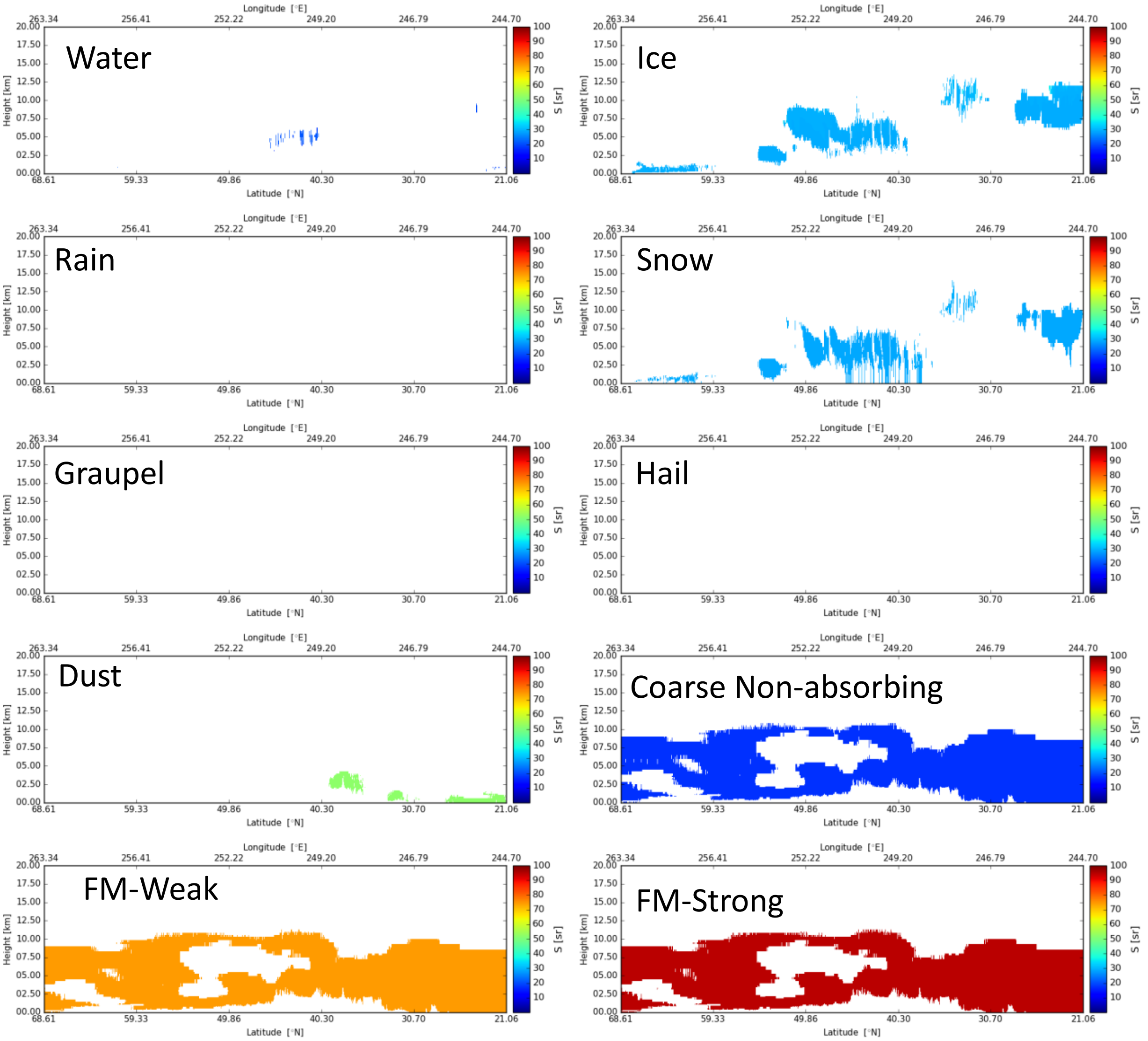


Figure 71: Lidar-ratio per species for the Baja scene.

7.3.2 The Hawaii Scene

The following figures given an overview of the horizontal, vertical, and per-species structure of the Baja scene.

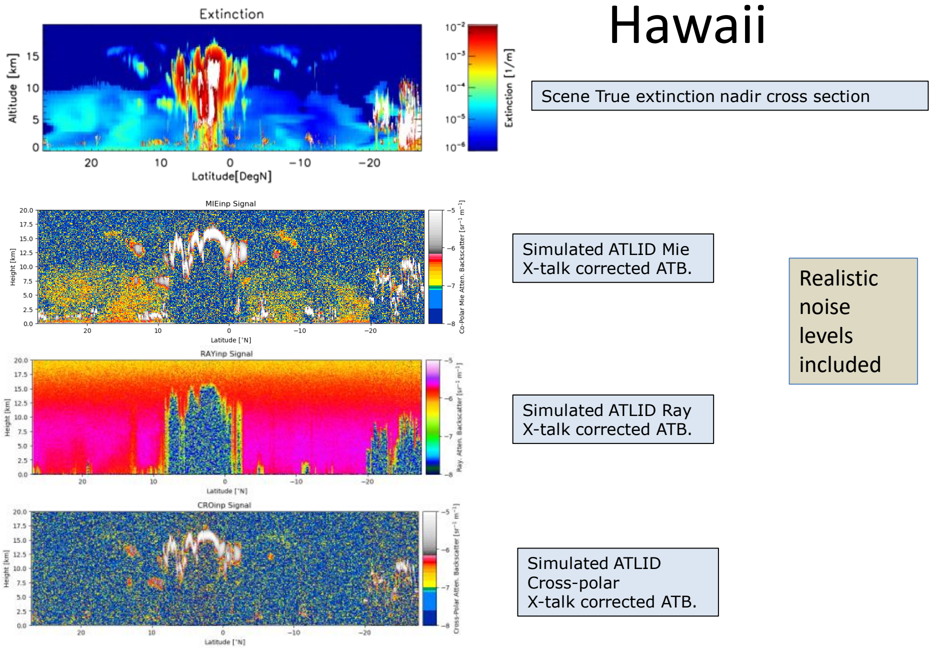


Figure 72: Nadir 355nm extinction and the simulated ATLID attenuated backscatters for the Hawaii scene.

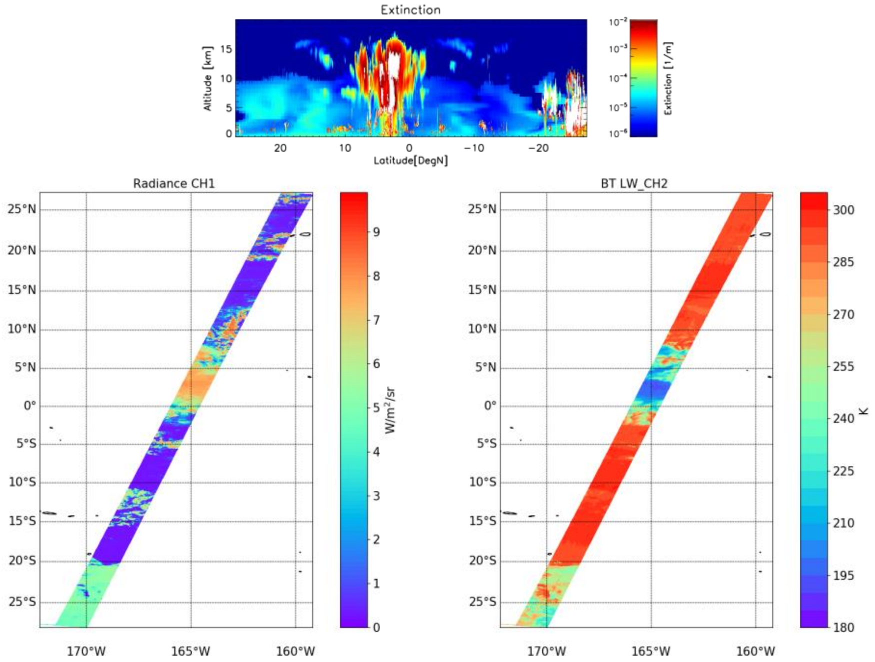


Figure 73: Nadir extinction and MSI CH1 radiances and LW-CH1 brightness temperatures.

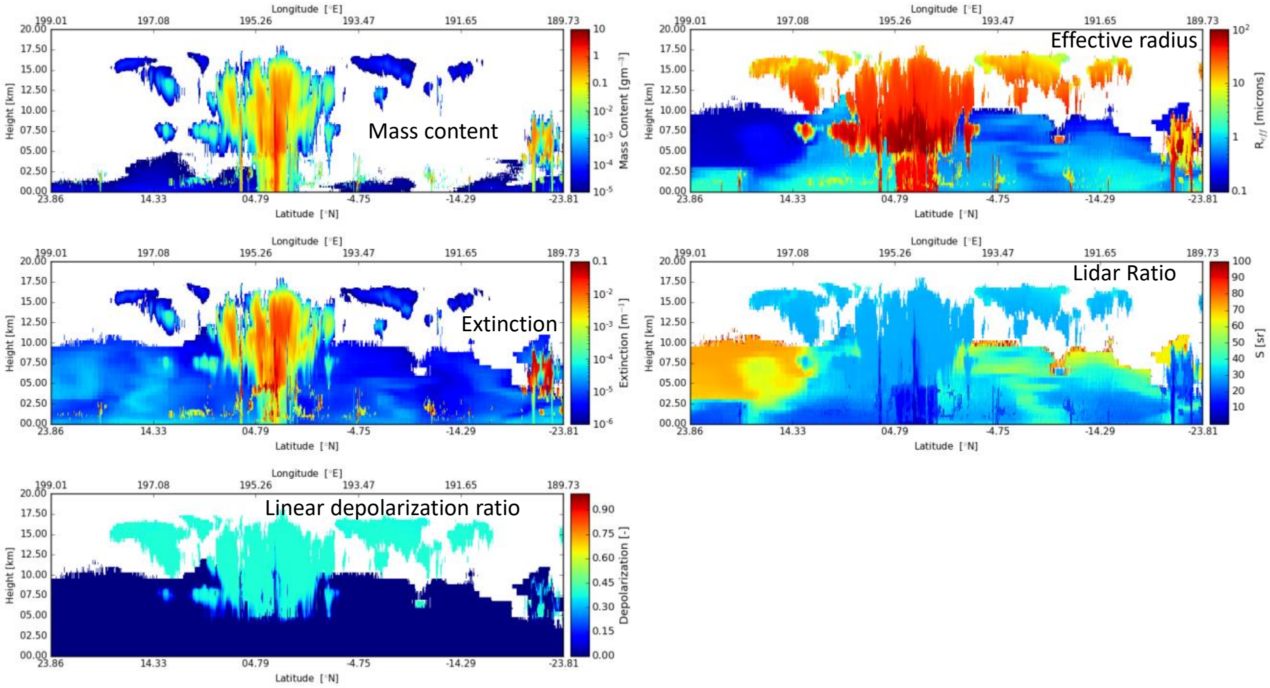


Figure 74: Total Mass content, effective radius, Extinction (355nm), lidar-ratio (355nm) and linear depolarization ratio (355nm) for the Hawaii scene at Nadir.

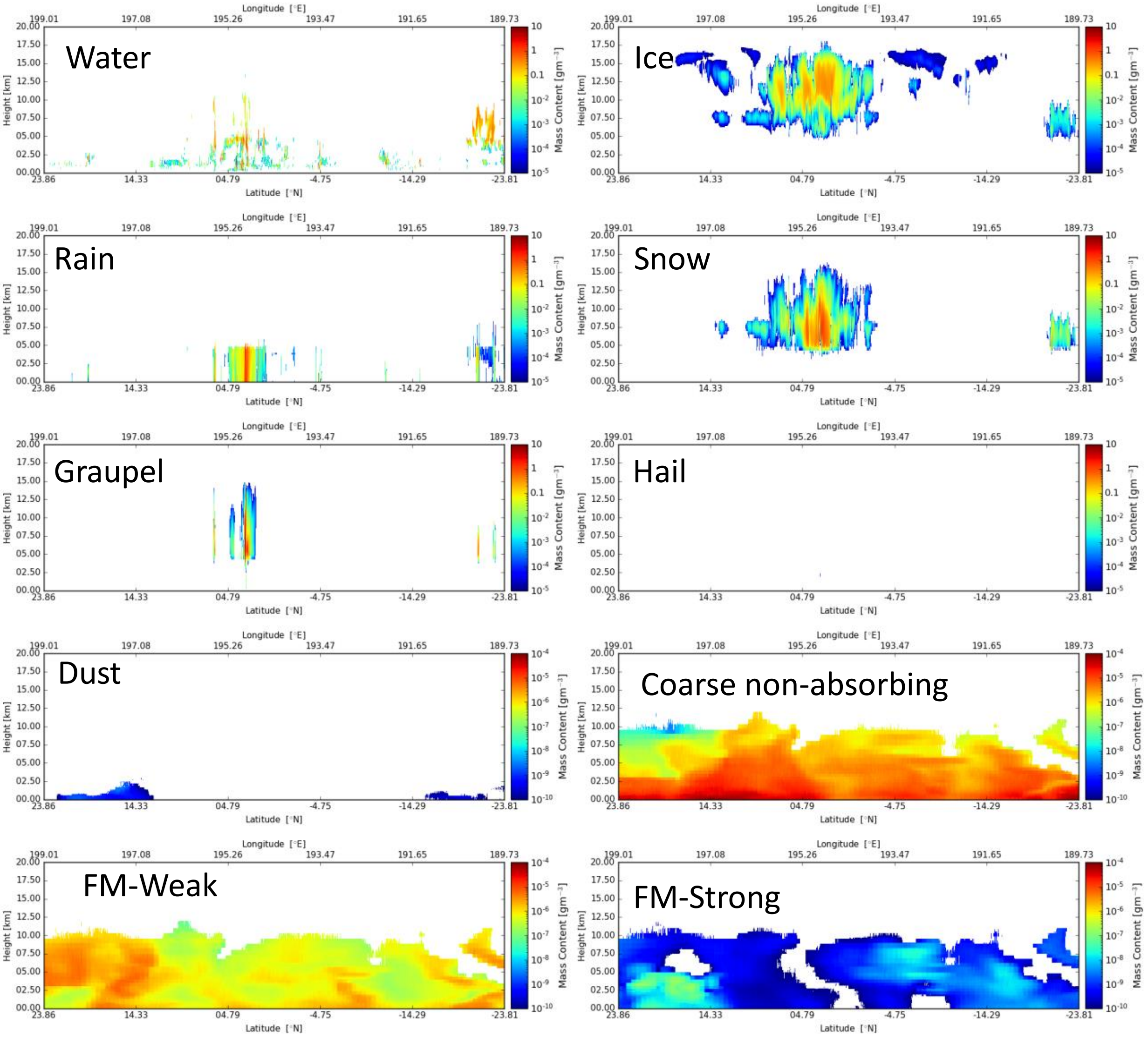


Figure 75: Mass content per species for the Hawaii scene.

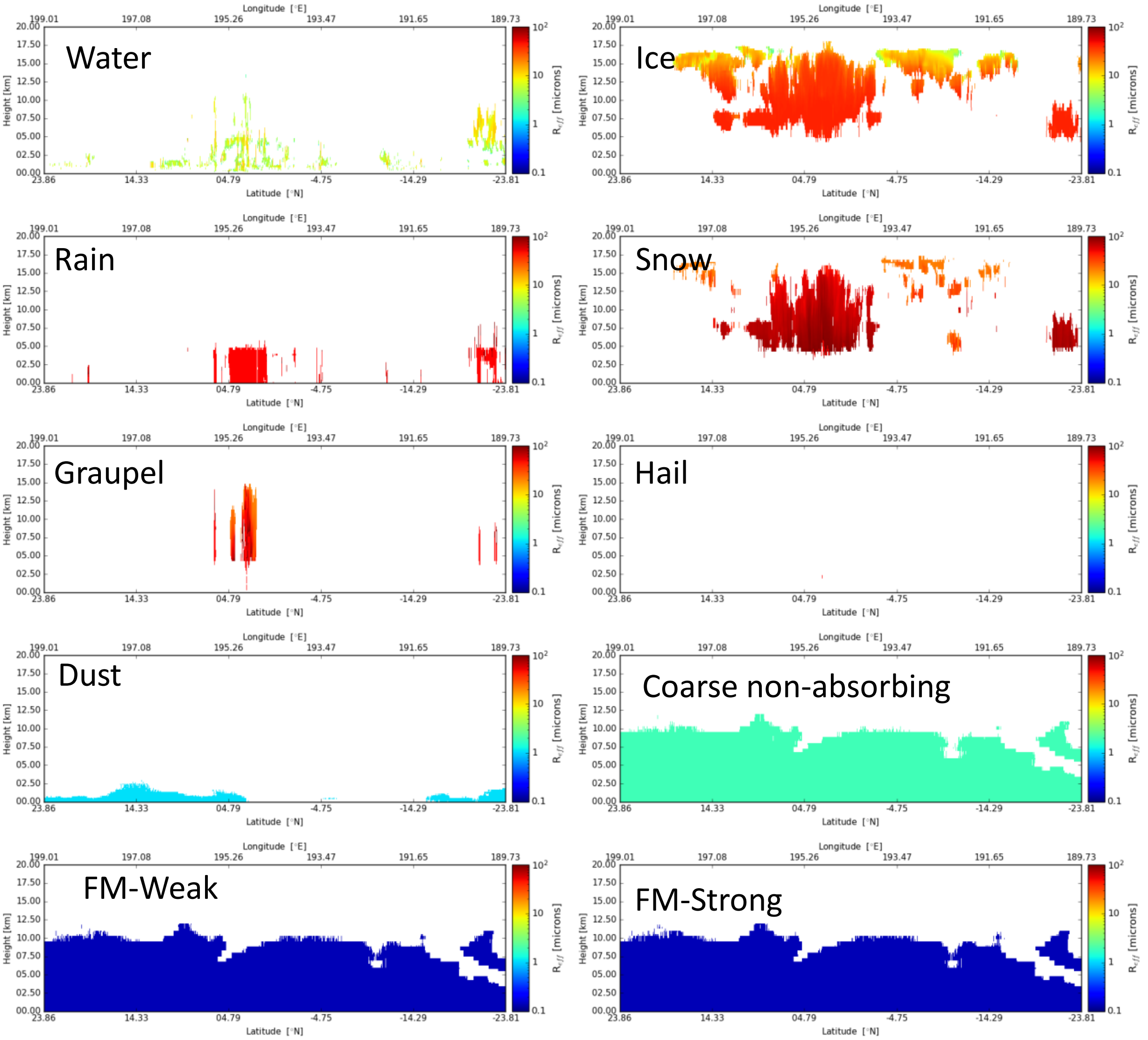


Figure 76:Effective radius per species for the Hawaii scene.

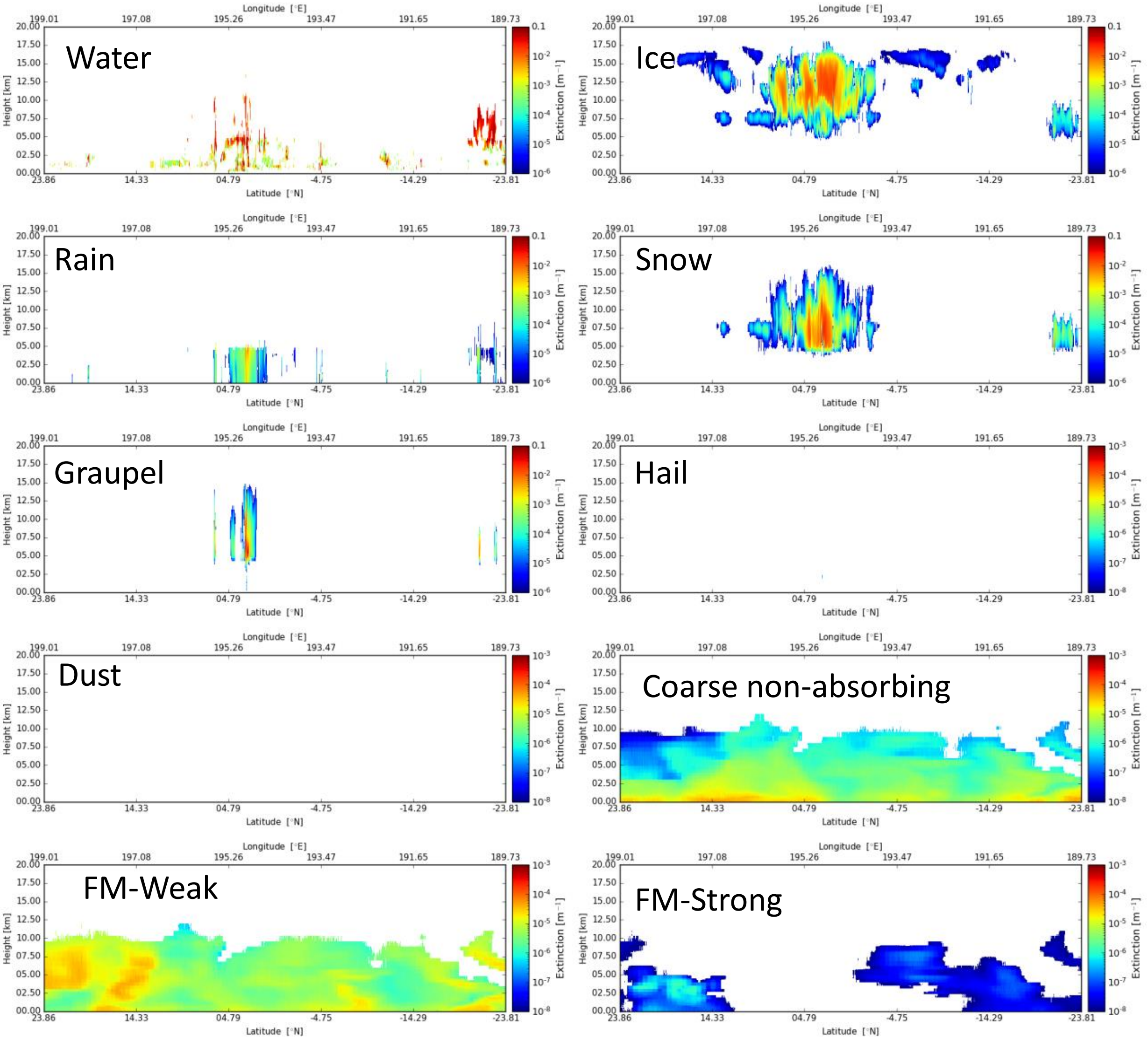


Figure 77: Extinction per species for the Hawaii scene.

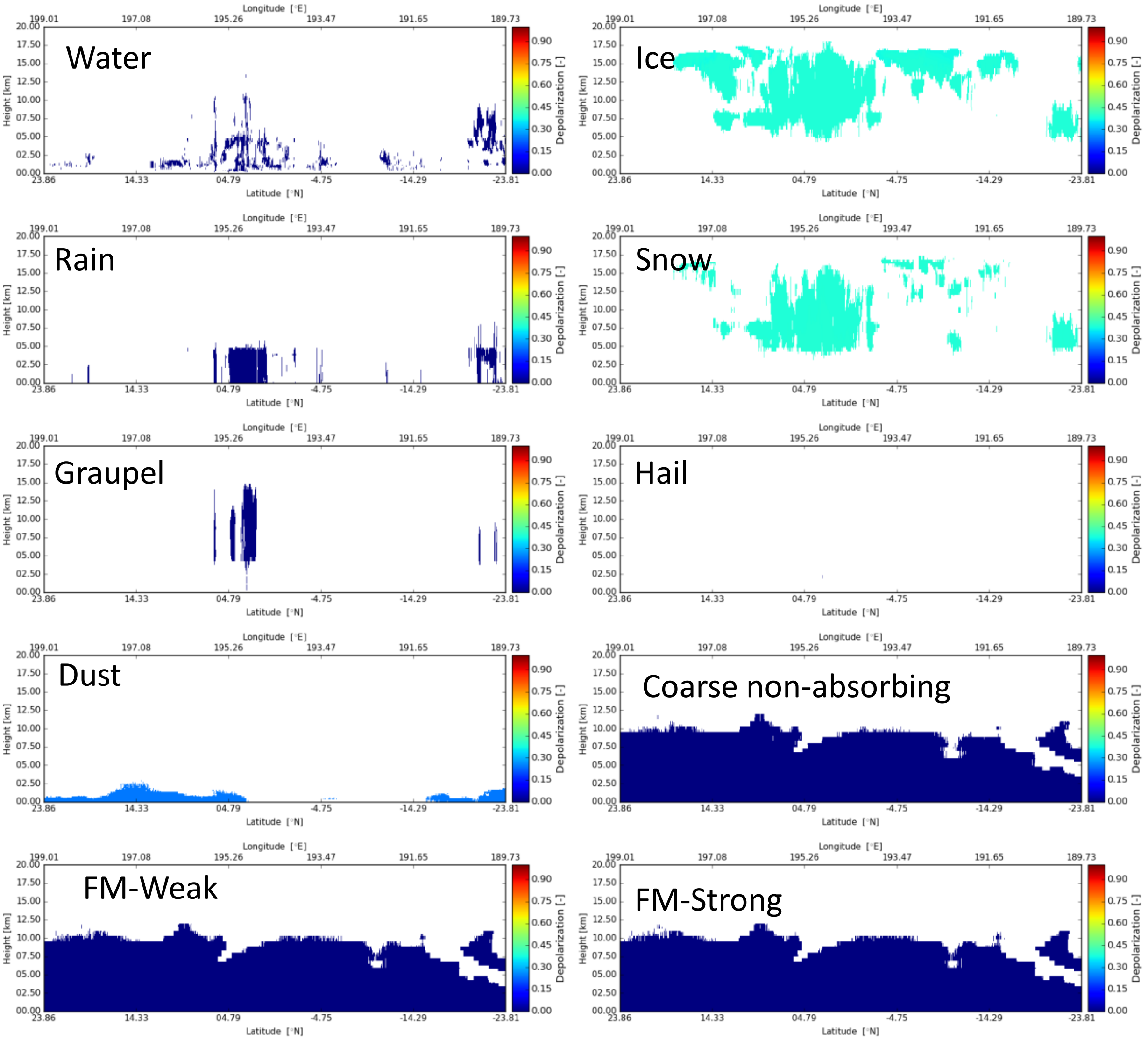


Figure 78: Linear depolarization ratio per species for the Hawaii scene.

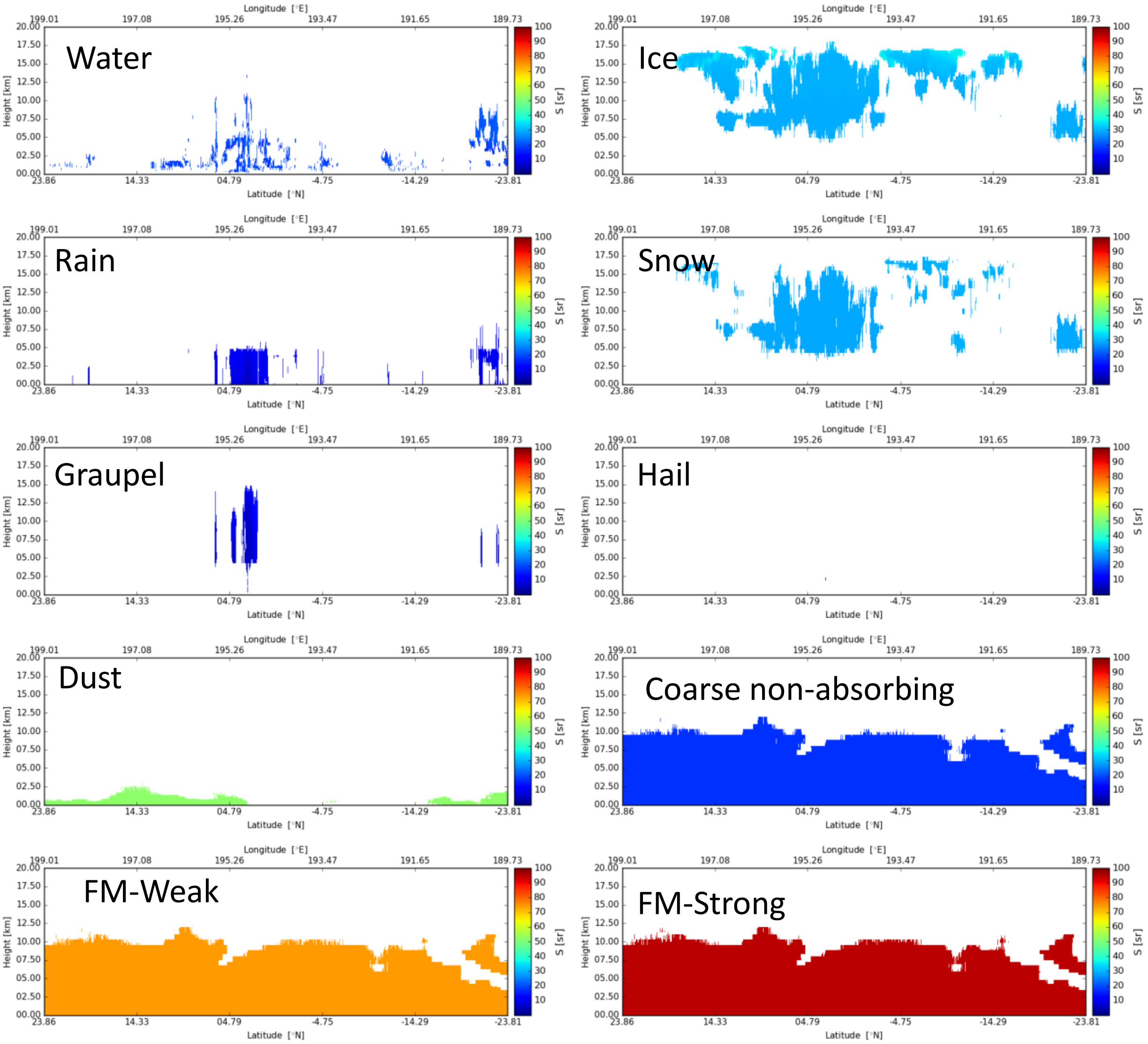


Figure 79: Lidar ratio per species for the Hawaii scene.

7.4 PSC Scene

This scene was constructed using combined CloudSat and Calipso DARDAR products as well as the CALIPSO PSC v2.0 mask. It is a nadir only scene. That is, there is no variability along the cross-track dimension. For this scene 5 different scattering types were used and mapped to species available in ECSIM. The observed extinction values were then used along with assumed size distributions to build the ECSIM scene. The detailed structure of the scene is illustrated by the figures present in the remainder of this section.

Table 8: PSC scene species and their mapping to ECSIM scattering types.

Calipso/DARDAR Mask Class	ECSIM scattering type
Water	Water
Tropospheric ice	Baum general habit mixture
Solid PSCs	Hess imperfect columns (tilted 3 Deg).
STS	H_2SO_4 25% mixture
Enhanced NAT	H_2SO_4 75% mixture
Nat-Mix	H_2SO_4 50% mixture

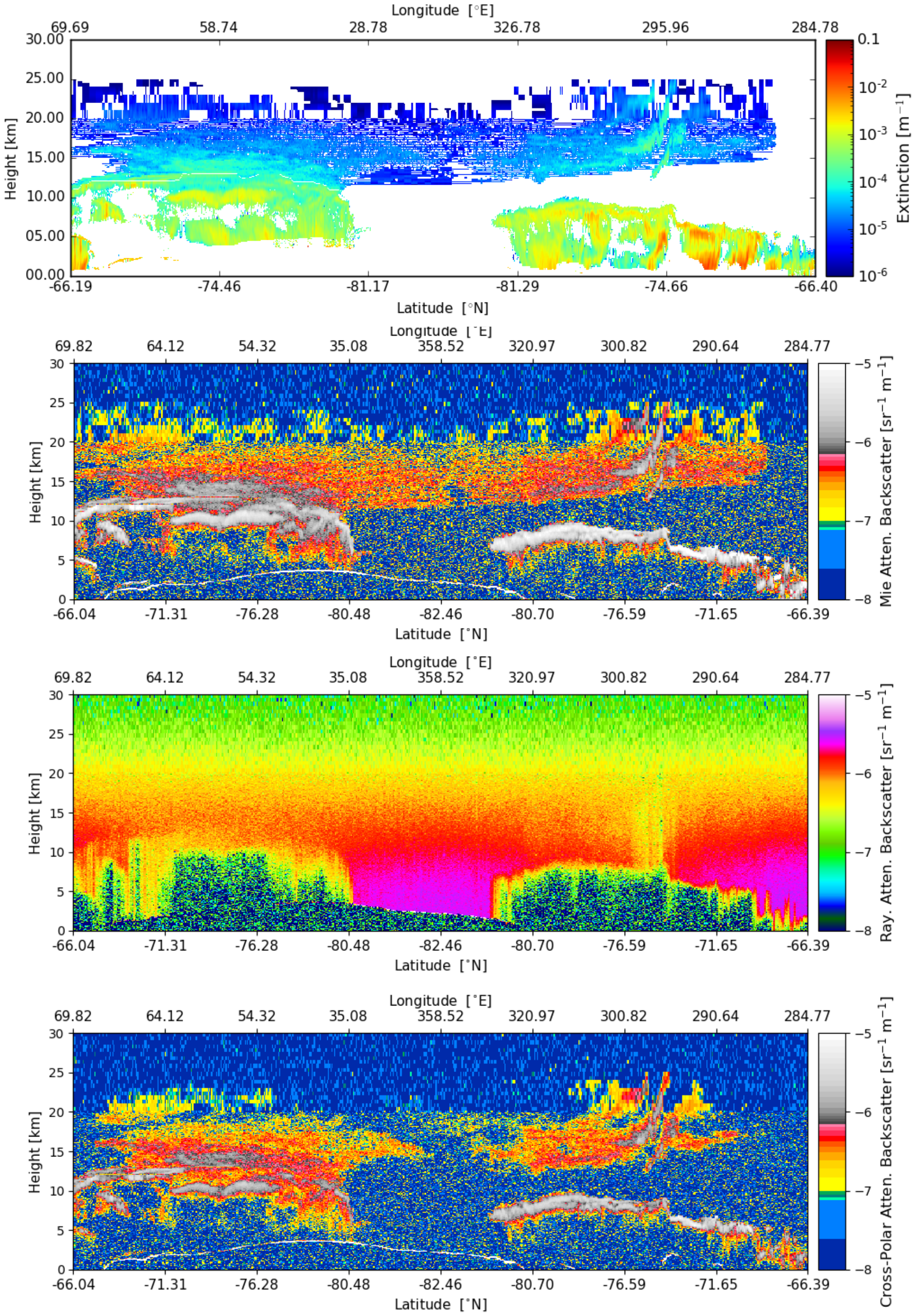


Figure 80: Model truth extinction and simulated ATLID attenuated backscatter for the PSC scene.

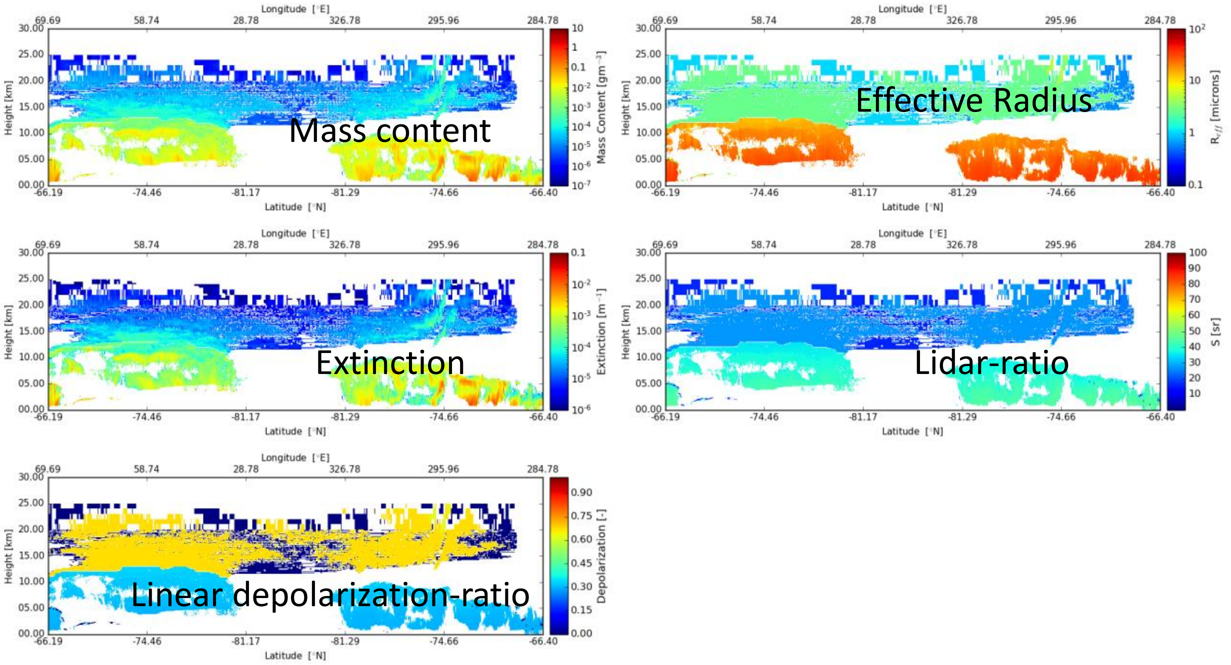


Figure 81: Total-Mass content, effective radius, Extinction (355nm), lidar-ratio (355nm) and linear depolarization ratio (355nm) for the PSC scene.

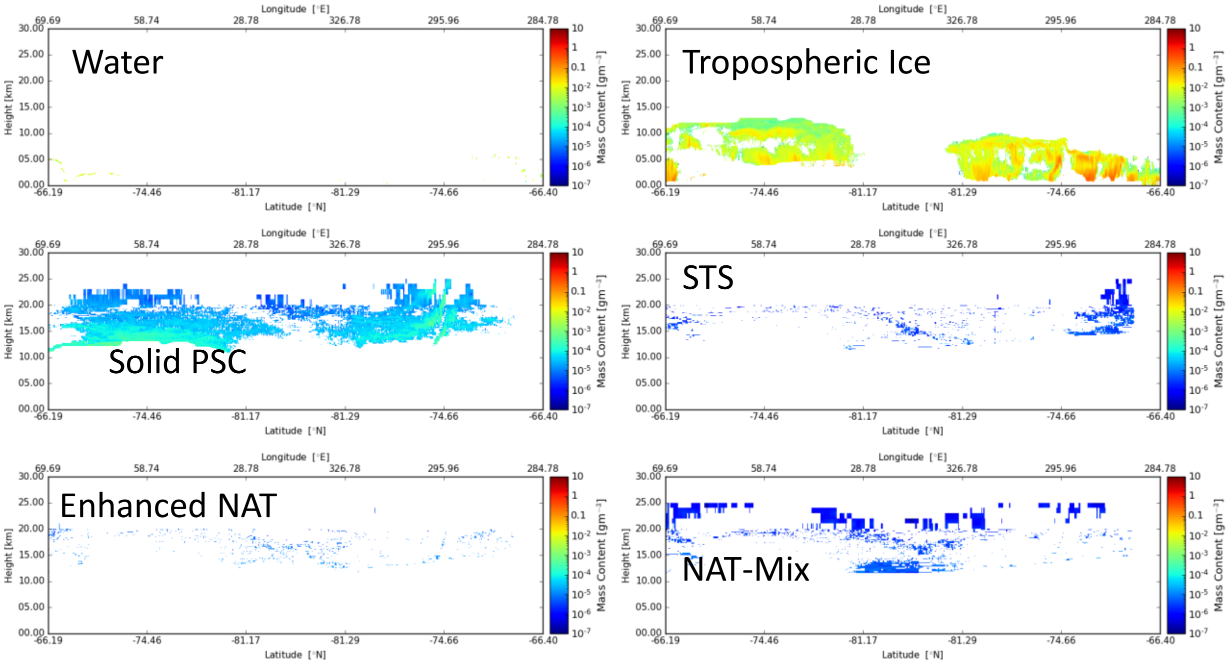


Figure 82: Mass content per species for the PSC scene.

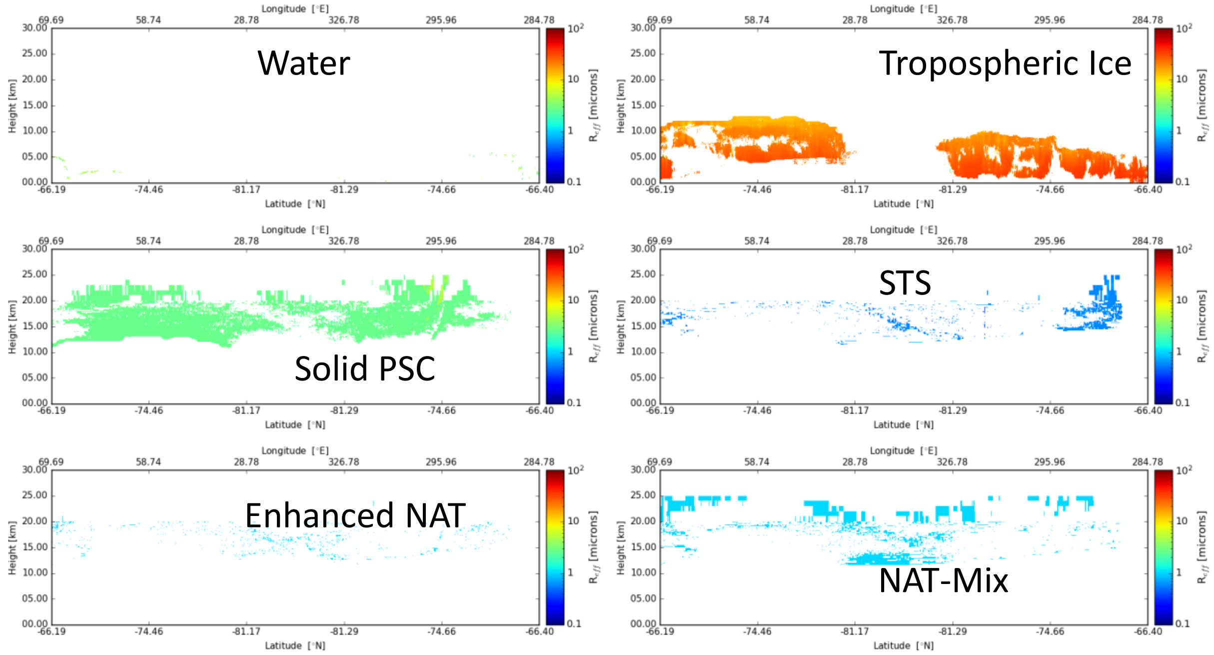


Figure 83: Effective particle radius per species for the PSC scene.

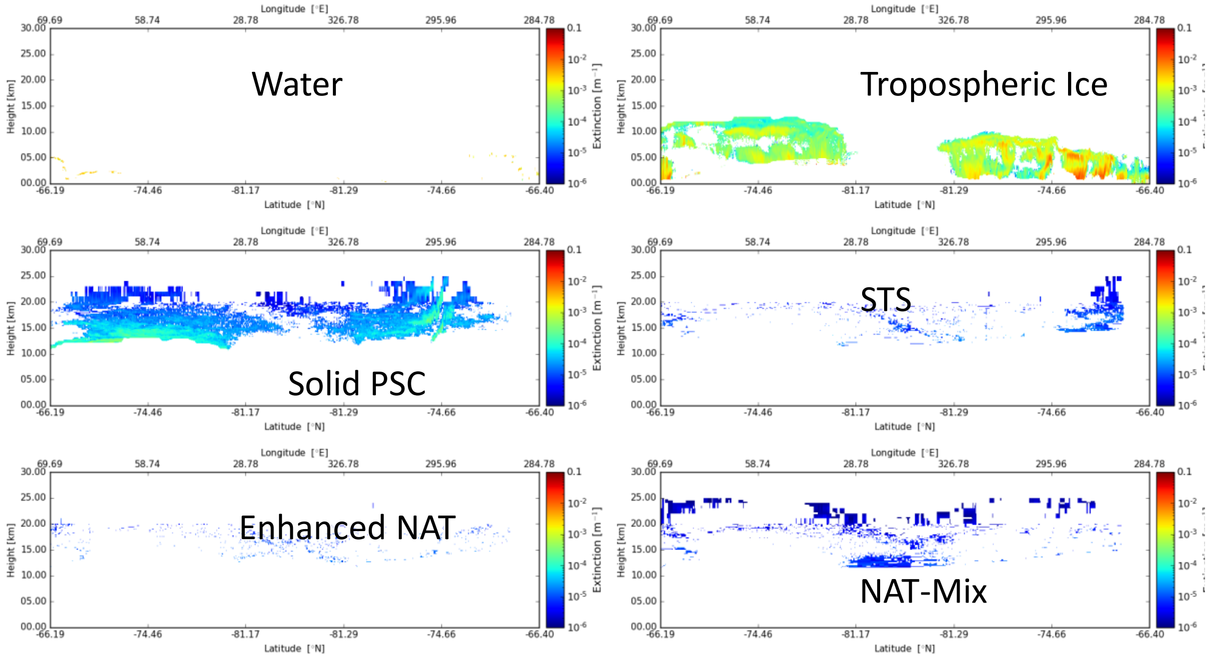


Figure 84: Extinction per species for the PSC scene.

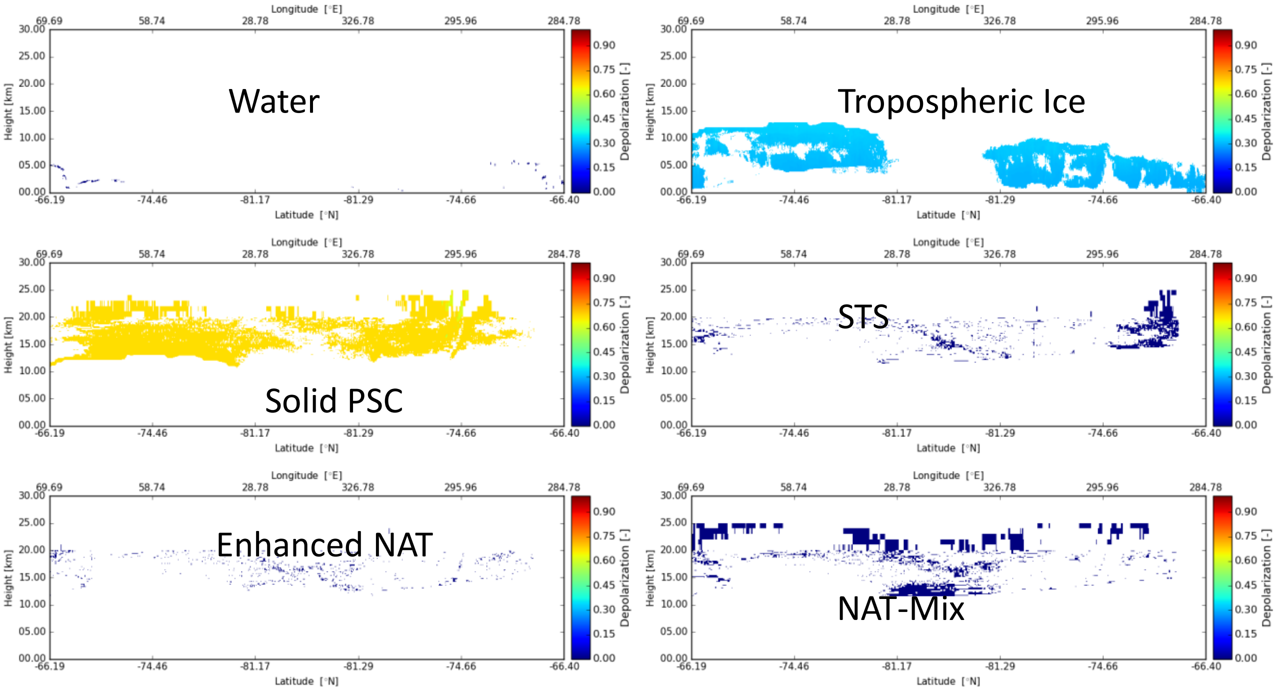


Figure 85: Linear depolarization ratio per species for the PSC scene.

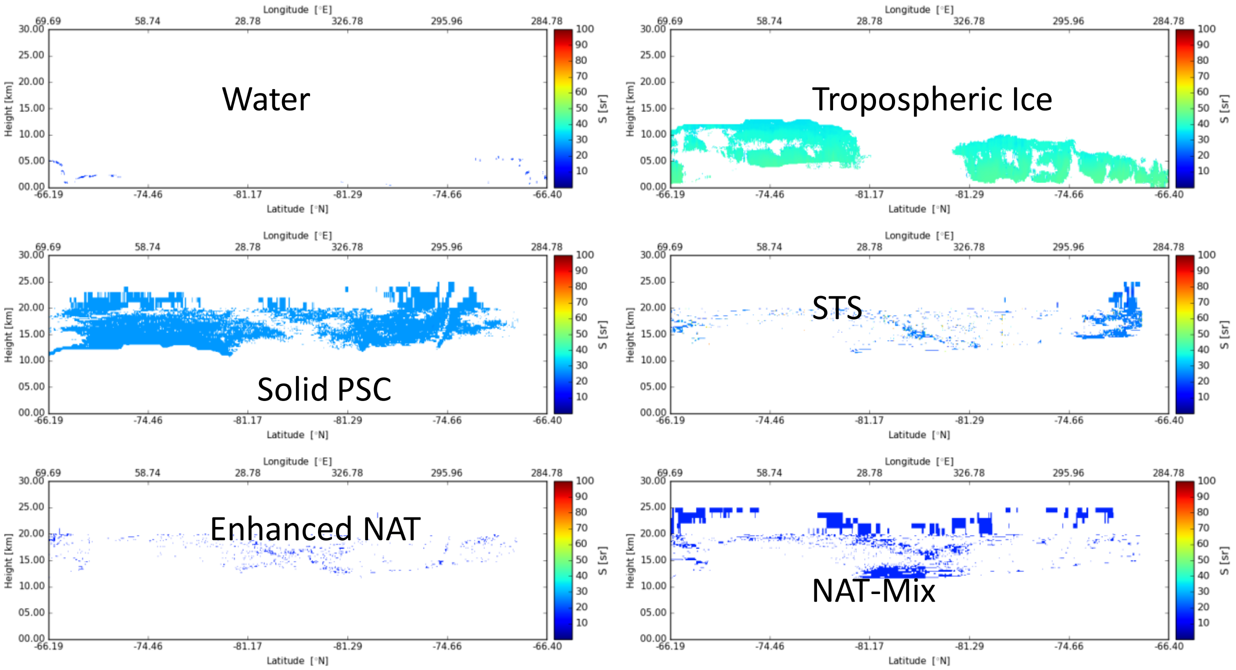


Figure 86: Lidar-ratio ratio per species for the PSC scene.

8 APPENDIX B: IMPLEMENTATION DETAILS

8.1 Execution Time

The total time for the frame-sized Halifax scene for V8.1 is on the order of 200 seconds using 4 threads on a i7-6820HQ CPU @ 2.70GHz platform for the PSC scene (which has targets present at a higher altitude range) the execution time is on the order of 400 seconds. Due to the parallelization (using OpenMP directives) of critical sections of the code, it is anticipated that the requirement of 150 seconds on the 8 physical core target platform will be met. If further improvement is required then the approach of calculating the gradient of the cost-function explicitly (replacing the more expensive numerical derivative calculation) and/or linking the now separate eta and theta state variables via an effective particle size (thereby eliminating one set of variables in the cost-function).

8.1.1 External models

- A public domain implementation of the BFGS minimization routine using numerical derivatives has been used. The code uses was downloaded from http://athens.src.uchicago.edu/jenni/econ350/Salvador/MINIMIZATION.F90 and subsequently modified by allowing for control over the allowed maximum step sizes and the ability to return the approximate inverse Hessian matrix.

  The choice of minimization routine is not key to the algorithm however and other suitable minimization solvers can be used. The present version is an adaptation of the method described in Press et al. (1992) modified so that numerical derivatives are used.