



HAL
open science

Optimal estimation method applied on ceilometer aerosol retrievals

Andres E. Bedoya-Velásquez, Romain Ceolato, Sidonie Lefebvre

► To cite this version:

Andres E. Bedoya-Velásquez, Romain Ceolato, Sidonie Lefebvre. Optimal estimation method applied on ceilometer aerosol retrievals. *Atmospheric Environment*, 2021, 249, pp.118243. 10.1016/j.atmosenv.2021.118243 . hal-03566650

HAL Id: hal-03566650

<https://hal.science/hal-03566650>

Submitted on 13 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Optimal estimation method applied on ceilometer aerosol retrievals

A.E. Bedoya-Velásquez^a, Romain Ceolato^a and Sidonie Lefebvre^{b,*}^aONERA/DOA, Université de Toulouse F-31055 Toulouse - France^bONERA/DOA, Université Paris Saclay F-91123 Palaiseau - France

ARTICLE INFO

Keywords:

Optimal estimation method
Ceilometer vaisala CL51
Nonlinear elastic Aerosol inversion

ABSTRACT

The solution of the lidar equation is an ill-posed problem that requires nonlinear methods to retrieve the atmospheric aerosol optical and microphysical properties. Particularly, in the last decades, the most applied solution for the elastic lidars is through the well known Klett-Fernald-Sasano algorithm for retrieving the backscatter coefficient. To solve this inversion problem, we propose to apply the optimal estimation method to a Vaisala CL51 ceilometer range corrected signals for retrieving under two different frameworks, the particle backscatter coefficient or the ratio and the lidar constant. The optimal estimation is a Bayesian inversion fed by a set of a priori information. In this work, to obtain the suitable prior, we have tested two approaches that involved measurements and synthetic data. The first data set was obtained from previous inversions using the classical Klett-Fernald-Sasano method, and the second one by using Mie simulations fed by aerosol properties from OPAC database.

The optimal estimation method used for elastic lidar inversion presents two main advantages compared to the classic approaches. On one hand, there is no need for Rayleigh zone determination and on the other hand, the uncertainty of the retrieved products is directly estimated, therefore the quality of the results is highly dependant on the prior selection. To evaluate the performance of the model, low and high aerosol accumulations scenarios were considered, finding that the backscatter coefficient was oscillating between 5 to 7 (kmsr)⁻¹ in the first 3 km agl with uncertainties lower than 27 % at degraded spatial resolutions. Additionally, constant and height-dependent priors were tested reaching relative errors in percentage up to 5% between them. Besides, relative errors were also analyzed for the prior covariance matrices estimated either from synthetic lidar data and Klett's retrievals, where the errors are lower than 2% by using one instead of the another. However, scale factors were applied to the synthetic prior covariance matrices to reach the convergence.

The results at retrieving the particle backscattering were compared to those ones estimated from Klett's inversion, considering Klett inversion as the reference. For the extreme scenario of inversions, considering aerosol accumulations at different layers, the bias between the optimized profiles was lower than -0.5 (kmsr)⁻¹ in the first 0.5 km and 0.5 (kmsr)⁻¹ above 1.5 km. Here, we also shown a two-parameter optimization for the lidar constant and lidar ratio, applied to 39 aerosol inversions, finding relative errors lower than 1 % and 2.3 %, respectively, considering those ones from Klett inversion as the reference.

1. Introduction

Atmospheric aerosols play a crucial role in atmospheric dynamics and the energy balance of the Earth. The main impact of the aerosols related-interactions is (i) the aerosol-radiation interaction (ARI), affecting the radiative fluxes of the Earth by absorbing and scattering solar and thermal radiation, and (ii) aerosol-cloud interaction (ACI) which are mainly associated with the modification of cloud properties and precipitation caused by aerosols (Boucher et al. (2013)).

Light Detection and Ranging (lidar) is one of the widely spread remote sensing techniques for studying the optical and microphysical properties of atmospheric aerosols. Optical properties like the particle backscatter and extinction coefficients can be retrieved by solving the lidar equation since it contains the aerosol-light interaction information

* This work was founded by ONERA, The French Aerospace Lab, within the framework of PROMETE Project.

*Corresponding author

 andres.bedoya@onera.fr (A.E. Bedoya-Velásquez); Romain.Ceolato@onera.fr (R. Ceolato); Sidonie.Lefebvre@onera.fr (S. Lefebvre)

 www.onera.fr (A.E. Bedoya-Velásquez)

ORCID(s): 0000-0002-8832-6851 (A.E. Bedoya-Velásquez)

together with the geometrical and optoelectronics efficiency terms. To solve the lidar equation, the Klett-Fernald-Sasano method is commonly used (Klett (1981, 1985)). The convergence of the solution requires some constraints and assumptions such as lidar ratio and a good definition of the Rayleigh zone (i.e. pure molecular zone). To this end, normally multi-channel lidar systems or synergies between lidar and other remote sensors are often used for improving the results (Pappalardo et al. (2004); Herreras et al. (2018); Marcos et al. (2018); Bedoya-Velásquez et al. (2021)).

One of the most used combinations of remote sensors for retrieving aerosol products includes AERONET (Aerosol Robotic Network) sun photometers and lidar systems, however, currently, some improvements have been explored including synergies with microwave radiometers (MWR) like in (Marcos et al. (2018); Román et al. (2018); Bedoya-Velásquez et al. (2021)) for retrieving particle backscatter coefficient (β) profiles from ceilometers lidar systems. Among other synergies, it is possible to find combinations of Raman lidar and MWR to improve the vertical resolution of the water vapor and temperature profiles (Barrera-Verdejo et al. (2016); Bedoya-Velásquez et al. (2018)). The main drawback of these synergies is that most of the lidar systems operate around the globe in semi-automatic mode, which means that qualified human support is frequently needed. To overcome this, ceilometers are used as lidar-based instruments that operate in single-wavelength, unattended, and continuously (Wiegner et al. (2014)). Recently, in Celato et al. (2020) has been shown the capabilities of the micro-lidar system operating in the short-range to retrieve aerosol properties by using a forward inversion method applied to a calibrated lidar signal mainly for avoiding the previous determination of the Rayleigh zone as it is needed for far-range lidars inversion.

In order to tackle the lidar inversion problem, additional approaches have been proposed in the last two decades. As one of the efficient ways to solve this ill-posed problem, the optimal estimation was applied to atmospheric remote sensing firstly by Rodgers (2000). This technique is considered as a nonlinear regression for determining the most probable state of a variable calculated from the prior information (observations). The optimal estimation method (OEM) allows obtaining a complete uncertainty of the variables retrieved, including random uncertainties, smoothing error, and uncertainty due to model parameters. This method has been applied mostly in radiometry and satellite fields (Barrera-Verdejo et al. (2016) and Nanda et al. (2018)), and much less for lidar inversion as by Farhani et al. (2019a); Sica and Haeefe (2015); Farhani et al. (2019b), and differential optical absorption spectroscopy (DOAS) (Bösch et al. (2018)). Centering the attention on lidar technique, satisfactory results have been shown by Povey et al. (2014), demonstrating the capabilities to retrieve lidar backscatter (β) and extinction (α) profiles in a multi-wavelength Raman lidar using OPAC (Optical Properties of Aerosols and Clouds) model to predict a set of prior information about the aerosol optical properties, and other results obtained by Sica and Haeefe (2015, 2016) at retrieving temperature and water vapor mixing ratio profiles from a multi-wavelength Raman lidar system. In this paper, we apply the OEM under two different frameworks, one for retrieving β profiles and the other one to obtain lidar ratio (LR) and calibration constant (C) using a Vaisala ceilometer dataset. In this work, we improve the knowledge of the prior information by using a previous dataset of aerosol retrievals calculated by using a Klett-based method combining measurements from a co-located Microwave Radiometer (MWR), AERONET sun photometer, and Vaisala CL51 ceilometer, to improve the classical Klett inversion at obtaining a set of β profiles together with useful aerosol properties such as a lidar ratio. The semi-automatic Klett algorithm is fully detailed in Bedoya-Velásquez et al. (2021).

The manuscript is organized as follows. The instrumentation and data availability is referred in Section 2, while in Section 3 the optimal estimation framework is defined. Section 4 explains the OEM applied to a ceilometer case: forward model, priors, and measurement vector selection. The covariance matrices information is shown in section 5. Section 6 presents the results applied to a case of study for one-single component and two-components inversion. Finally, Section 7 summarizes the main conclusions.

2. Instrumentation and Data availability

In this work, a Vaisala ceilometer CL51 is used. This is an active remote sensor commonly used as a low cost lidar (LIght Detection and Ranging) system that emits pulsed laser radiation towards the atmosphere in a single-wavelength centered in 910 ± 10 nm. Ceilometers take measurements automatically 24/7, retrieving as the main product the attenuated backscatter which gives information about atmospheric aerosol-light interaction. Attenuated backscattering profiles are obtained each 36 s from 0 to 15 km allowing to have robust data bases. Detailed technical information can be found in Kotthaus et al. (2016).

The signal of Vaisala ceilometers presents some drawbacks linked with noise aspects such as background and dark current effects. These noises have been deeply analyzed by Kotthaus et al. (2016); Marcos et al. (2018); Bedoya-Velásquez et al. (2021), thus most of the drawbacks are currently overcome. Additionally, due to the ceilometer emission wavelength the signal is water vapor contaminated. Therefore Wiegner and Gasteiger (2015) have developed a methodology to water vapor correct the signal by using previous atmospheric information such as temperature, relative humidity and pressure profiles combined with the simulated water vapor cross section to suppress this effect on the atmospheric transmittance. The data used in this work have been pre-processed following all the steps proposed in the above-mentioned works, in order to have clean attenuated backscatter profiles which will be represented in the range-corrected (rcs) form. This research took place at Toulouse, France (N: 43 34'12", E: 128'24") in The French Aerospace Lab (ONERA), covering continuous measurements from October 2019 to February 2020. As complementary instrumentation, a MWR co-located to the ceilometer, took continuous measurements of atmospheric profiles of temperature and absolute humidity for improving the ceilometer products and feed the OEM.

3. Optimal estimation framework

3.1. General description

The OEM is widely used for solving ill-posed problems of nonlinear nature. Rodgers (2000) proposed a Bayesian solution to inverse problems based on the following forward model,

$$y = F(x, b) + \epsilon \quad (1)$$

where y is the measurement vector and ϵ is the noise considered in the measurements. $F(x, b)$ is the forward model that allows mapping the state variables (x) from the state space to the measurement space, and b are the known parameters of the measurements.

OEM is applied by assuming a certain degree of linearity between y and x vectors, which in fact is convenient for the error analysis, but not for finding a solution. Under this treatment, y and x are Gaussian distributed, with S_x and S_y as their covariances, respectively. Under these assumptions, and considering the Bayes theorem, the conditional probability of the system in a state x at given y measurement by considering Eq.1 defined in Rodgers (2000) is the following

$$-2\ln P(x|y) = [y - F(x, b)]^T S_e^{-1} [y - F(x, b)] + [x - x_a]^T S_a^{-1} [x - x_a] \quad (2)$$

where x_a is the prior vector and $-2\ln P(x|y)$ is the cost function that measures the goodness of fit of the solution, seeking the state that maximizes the posterior probability density function, the maximum a posteriori solution (MAP). In most of the cases, this quantity is length-normalized to the measurements (y). S_a and S_e are the covariance matrices that describe the prior and experimental errors, respectively. As our forward model is nonlinear, an iterative numerical method is used, therefore we have considered Gauss-Newton and Levenberg–Marquardt as suitable numerical methods to solve this ill-posed problem. The most probable state x can be obtained by considering the following iteration:

$$x_{i+1} = x_i + [(1 + \gamma_i)S_a^{-1} + K_i^T S_e^{-1} K_i]^{-1} [K_i^T S_e^{-1} [y - F(x, b)] - S_a^{-1} (x_i - x_a)] \quad (3)$$

The Jacobian is defined as $K_i = \partial F / \partial x_i$ which can be interpreted as the response of the observation vector to a perturbation in the state vector and γ_i is a scaling constant defined in Povey et al. (2014), where $\gamma_i = 0$ refers to Gauss-Newton and $\gamma_i > 0$ to Levenberg–Marquardt numerical solutions. In the Gauss-Newton the sum of the squared errors is reduced though the least squares function locally quadratic, looking for the minimum of the quadratic expression, meanwhile Levenberg–Marquardt operates such as a descent gradient for cases when the parameters are quite far from the optimal value, and like Gauss-Newton when the parameters are close to the optimal value. In Povey et al. (2014) results, γ_i took values from 1 to 1000 stepped by 10 at each iteration, demonstrating that greater values will drive the OEM solutions to converge to the prior vector. A common convergence criterion is to consider that the solution vector

(x_{op}) is reached once the difference between the forward model applied to the measurement state at n and $n+1$ iterations are one order of magnitude smaller than the estimated error. During the OEM implementation we have considered the following:

- S_a was calculated from the real and simulated data, and the inclusion of off-diagonal terms was evaluated to look for stability and improve the convergence of the inversion.
- S_e include the errors in y vector and p parameters considered as measurements.
- The covariance of the posterior distribution is $S_{opt} = (K_i^T S_e^{-1} K_i + S_a^{-1})^{-1}$.
- The degrees of freedom (DOF) of a profile represents the number of independent pieces of information in the signal. They can be calculated as the trace of the matrix of the averaging kernel (A_k) defined as $A_k = S_{opt} K_{opt}^T S_e^{-1} K_{opt}$, which describes the subspace of the state space where retrieval must lie.
- A maximum number of iterations from which we consider a failure convergence of the estimation model.
- The optimized values will be labeled as opt , taking the following form $x_{opt} = A_k x + (I - A_k)x_a + S_{opt} K_{opt}^T S_e^{-1} e$

Figure 1 is showing the schematic block diagram of the algorithm used. The OE general description presented above will be applied on the low-cost lidar system (ceilometer) signal under to two retrieval scenarios. The first one is for optimizing the backscatter coefficient (β), and the second is to optimize two parameters, the lidar ratio (LR) and lidar constant (C). According to that, it is relevant to define from this point forward that the state vector will change from one scenario to the other. For the β optimization, the state vector and its covariance matrix will contain the prior information of the β profiles retrieved by using a semi-automatic klett-Fernald-Sasano method (Bedoya-Velázquez et al. (2021)). For the second scenario of inversion, the state vectors and covariances will be built by LR and C values previously calculated as a step during the klett-Fernal-Sasano β calculation. Both scenarios of inversion will share the same y and S_y containing the range corrected and pre-processed ceilometer signals so called rcs. All this elements of the OE retrieval are deeply analyzed in the upcoming sections, including different test for the prior information in the first inversion scenario.

3.2. The forward model in the elastic lidar context

Lidar technique allows having information about the atmospheric aerosol properties, optical and microphysical. The backscattered power measured is a result of the aerosol-radiation interactions. The collected power is a height-dependent vector (R) that is described in the so-called lidar equation

$$rcs(R) = CO(R)[\beta_{mol}(R) + \beta_{par}(R)] \exp \left[-2 \int_0^r (\alpha_{mol}(r) + \alpha_{par}(r)) dr \right] \quad (4)$$

where C is the lidar constant that takes into account the opto-electronics efficiency and $O(R)$ is the overlap function which describes the geometry of the laser and the telescope systems. $\beta(R)$ is the backscatter coefficient and $\alpha(R)$ is the extinction coefficient, both include the contribution of the molecules (mol) and particles (par) in the atmosphere.

To solve the Eq.4 are required some constraints and boundary conditions as $\beta(R)$ and $\alpha(R)$ are both unknown variables during the measurement. For elastic lidar systems, the Klett-Fernald-Sasano (Klett (1981) and Klett (1985)) is one of the most used methods for deriving $\beta_{par}(R)$ by solving a Bernoulli's differential equation under a boundary conditions scenario and extra constraints like the well-known lidar ratio ($LR = \alpha(R)/\beta(R)$). Further approaches to solving the lidar equation by using nonlinear regression have been used by retrieving microphysical aerosol properties rather than optical properties like in Delanoë and Hogan (2008). Pounder et al. (2012) have retrieved high-quality extinction profiles considering multiple scattering using Twomey-Tikhonov smoothing instead of a priori information. Other approaches have been shown by Veselovskii et al. (2002), using regularisation for retrieving $\beta(R)$ and $\alpha(R)$ in Raman lidar systems. The OEM presents another formulation with a different interpretation of the problem, considering a forward model fed with prior information contained in the prior (x_a) and in the covariance matrix (S_a) derived from actual data which respond to the physical processes that govern the system.

Considering the retrieval problem stated in Eq.1 and Eq.3, we start from the lidar equation presented in Eq.4 as our forward model (FM) for obtaining $\beta_{par}(R)$, however in order to avoid negative values in the solution, and improve the model convergence, we chose a logarithmic form of the Eq.4, assuming some cost in the time convergence of the model, and it was also assumed the signal in a full-overlap region ($O(R)=1$) with $C=1$. The Eq.4 can be re-written as follows

$$\ln(rcs(R)) \approx \ln[\beta_m(R) + \beta_p(R)] - 2 \int_0^R (\alpha_m(r) + \alpha_p(r)) dr \quad (5)$$

3.3. Prior selection: x_a and S_a

The prior selection is a crucial step in the OEM scheme because it constrains the likely states that constrain more physical meaning to the model. Regarding the lidar context, we are looking for the optimal β_{par} using the Eq.4 as a theoretical model. OEM solutions are expected to be slightly smoothed than measured profiles, mainly in cases of high aerosol accumulations inside the atmospheric boundary layer (ABL), where gradients due to aerosol multi-layering are commonly seen. Prior information for elastic lidar ill-posed problems is commonly assumed from climatology, models or derived from external information (Lopatin et al. (2013); de Jesus and Landulfo (2009)). Additionally, radiosondes are also used to feed models in order to determine β_{mol} coefficient plus some LR assumptions. Other approaches have been proposed in Povey et al. (2014) by selecting as prior an order-of-magnitude according to the already known prior information to constrain the magnitude of the state vector preserving the inter-variable correlation.

In this work, two different optimal estimations were performed, one for single-component (β_{par}), and the other one for two-components, retrieving LR and C. For the single component, β_{mol} and β_{par} were defined as the priors, but only the β_{par} results are presented here. We used as a reference data set of previous retrievals of β_{par} and LR obtained by applying the near-real Klett-Fernald-Sasasano method proposed in Bedoya-Velásquez et al. (2021) to ceilometer data measured from October 2019 to February 2020. From there, it was performed 39 inversions for different aerosol scenarios, considering high and low aerosol accumulations classification. As we discussed in the previous section, x_a is a state vector that serves as a constrain to the forward model, storing the prior information. Here, three different approaches were tested to define a good prior: i) in a simplistic but correct way, x_a can be considered as constant value with the same order of magnitude of the desired variable to be optimized, ii) following the Bayesian formulation, it can be considered as a normal-distributed vector scaled to the same order of magnitude to the parameter to be optimized, and iii) finally the prior can be directly used from the measurements, as in our case the mean value of β_{mol} and β_{par} are good candidates. For the last two prior approaches, x_a is a height-dependent vector, meaning that the forward model (FM) can be built in a full resolution (i.e. 10 m for ceilometer), however, the priors might be spatially degraded to improve the time cost in terms of model convenience, or to simulate any real problem with scarce prior information. In our case, the minimum spatial resolution for the model convergence was 55 bins, where each bin represent 50m-height, only considering the first 3 km agl of the atmosphere (our region of interest). For the case of two-components retrieval, some prior information of C was needed, thus as a step during the Klett's inversion, we determined this constant as follows

$$C = \frac{rcs(r_{ref})}{\beta_m(ref)[\exp(2 \int_r^{r_{ref}} \alpha_m(r) dr) + \exp(2 \int_r^{r_{ref}} \alpha_p(r) dr)]} \quad (6)$$

Where the second integral ($\int_r^{r_{ref}} \alpha_p(r) dr$) is the definition of aerosol optical depth (AOD), which was directly calculated from sun photometer located in Toulouse, interpolating the AOD value to 910 nm. The r_{ref} is the altitude where pure molecules are expected to be predominant, known as the Rayleigh zone, and r is the full overlap height (set in 250m agl in our case).

To the one-single component optimization, the β_{mol} information was obtained from the temperature and relative humidity profiles took from the MWR for having a near-real-time β_{mol} profiles from 0 to 10km agl. For β_{par} , we classify the database between aerosol low and high accumulation events, in order to differentiate between multi-layering profiles and smooth ones. The first group contained 33 cases, named low accumulation cases (LA), and the other

Table 1

The relative error on β^{opt} for constant and height-dependent priors under HA classification is presented. The error profile has been analyzed for six different air volumes (each 0.5 km), showing the maximum relative error at each interval. The quantities presented in angle brackets are associated to the mean value.

Relative error on β^{opt} (%) fixing the covariance matrix (S_a)						
Prior vector ($kmsr$) ⁻¹	0-0.5 km	0.5-1 km	1-1.5 km	1.5-2 km	2-2.5 km	2.5-3 km
$x_a: 7 \times 10^{-3}$	23%	25.5%	26.5%	19%	16%	19%
$x_a: 7 \times 10^{-4}$	24%	27%	28%	20%	18%	20%
$x_a: \langle \beta_{par}(r) \rangle$	23%	25.5%	26.1%	17.6%	16%	19.2%

6 ones were labeled as high accumulation cases (HA). For each group, the mean profile was used as a prior for the OEM (see in Fig.2). As it was expected, β_{mol} did not change strongly between classifications, but on the contrary, β_{par} changed drastically (Fig.2). The red profiles shown in Fig.2ab refers to the mean profiles under no classification (i.e. all inversions), which means that we considered both LA and HA together.

We have used the approaches i) and iii) mentioned above to analyze the convergence of the OEM at using different priors, involving constant and height-dependent values. Besides, we have tested the convergence by changing the prior covariance matrix (S_a), but this will be analyzed in a further section. In general, for a good prior selection, the order of magnitude plays a crucial role for the optimization, because a wrong prior selection will lead to i) the non-convergence of the model, ii) underestimation or overestimation of the retrieved product, and iii) the increase of the error in the optimized variables due to the lack of information that constrains the prior with the physical phenomenon. In our case, the mean values of β_{par} were oscillating between 5 to 7×10^{-4} ($kmsr$)⁻¹. In the table 1, we are showing the maximum relative error ($(\delta X_{opt}/X_{opt}) * 100$) found in the β_{par} optimized for six different air volumes (i.e. the error vector was split each 0.5km) with a fixed correlation matrix. The results for different priors are presented only for HA aerosol classification since this is the complex scenario for the inversion, however, the procedure was performed for LA classification cases, but the results are not shown here. In this work, the priors with orders of magnitude above 10^{-4} ($kmsr$)⁻¹ carried out to the non-convergence of the model, meanwhile those with the same order or at least one order of magnitude lower than the mean β_{par} converged satisfactorily. As we can see in table 1, the error profile did not increase drastically by using either height-resolved or constant priors since the differences between them reached up 2.4% in the worst case.

As we discuss in Sec. 4.1, the FM selected is based on the Eq.4, thus in order to construct a two-components FM for retrieving LR and C, we have fixed the β_{mol} and β_{par} as the mean values depending on the classification (LA or HA), and then we can optimize the other two variables in the lidar equation. For this procedure, the priors (x_a) were the mean value of the LR and C. For C, a total of 312 values were obtained, but priors were considered only as the mean values according to the LA and HA classification. In Eq.6, the molecular free zone (r_{ref}) was determined by finding the corresponding altitude bin to the minimum difference between the rcs and β_{mol} profiles in the free-aerosol zone (a detailed description can be found in Bedoya-Velásquez et al. (2021)). To apply this methodology, we use a 1h average rcs and β_{mol} profiles to have smooth and clean rcs profiles. As is seen in Fig.3a, there is no clear data-distribution between LA and HA for C, but the main values are located from 1 to 1.4×10^5 km³sr. Fig.3b presents the LR data-distribution considering the LA (grey bars) and HA (black bars). LR did not present any particular distribution between cases as it is an intensive property, so it depends strongly on the atmospheric state, keeping a deep relation with the aerosol conditions during the measurement time. As this is not the aim of this work, no further analysis will be addressed in this direction. The standard deviation values obtained from LR and C were used to the further construction of the S_a .

3.4. Measurement: y and S_y

Two different approaches have been tested for the measurement vector (y) under the one-single component retrieval scenario. The first one uses the rcs profiles obtained by feeding the FM with the β_{mol} , and β_{par} retrieved from the Klett inversions (described in the previous section), adding a Gaussian noise (distribution mean= 0; standard deviation=0.02, Fig.4ab in blue line). For the second approach, we used the 1h-averaged calibrated rcs profiles from the ceilometer in the logarithmic scale (Fig.4ab in orange line).

Table 2

The relative error on β^{opt} for constant and height-dependent priors under HA classification is presented, considering two different approaches for S_a : measurements and synthetic data. The error profile has been analyzed following the same procedure explained before.

Relative error on β^{opt} (%) testing with different covariance matrices						
Prior vector (kmsr) ⁻¹	0-0.5 km	0.5-1 km	1-1.5 km	1.5-2 km	2-2.5 km	2.5-3 km
S_a : measured; x_a : 7×10^{-3}	23%	25.5%	26.5%	19%	16%	19%
S_a : scaled OPAC; x_a : 7×10^{-3}	24.5%	26.3%	26.8%	18.8%	18%	23%
S_a : measured; x_a : $\langle \beta_{par}(r) \rangle$	23%	25.5%	26.1%	17.6%	16%	19.2%
S_a : scaled OPAC; x_a : $\langle \beta_{par}(r) \rangle$	23%	26%	26.1%	17.4%	18%	23.5%

The main differences lie in the consideration that, the FM is assumed ideally calibrated (i.e. $C=1$), meanwhile for the measured rcs's, the calibration constant was calculated following Eq.6. Both schemes, under LA and HA, allowed to see differences in the first 500 m, and after 1.5 km. The rcs obtained from the FM presented higher levels of the signal in the first 500 m, and additionally, its decay tendency is more pronounced than rcs calibrated. The rcs obtained from FM has a dominant effect in the transmission term after aerosol accumulations (e.g. above 1.5 km, Fig.4b), making that the rcs decays slightly faster than the calibrated rcs. This behavior is less strong for the LA example case (Fig.4a), where both profiles are decaying at similar rates. During the optimization, different values of calibration constant around the retrieved (truth value) were tested, evidencing an increase in the time-cost of the OEM convergence. This time was increasing while the calibration constant values were moving away from the order of magnitude of the truth values. The absolute error profile for this particular case, not shown here, highlighted that below 0.5 km and above 2.0 km for LA, the error is lower than 5 %, while for HA the absolute error is up to 3 % in the first meters, but it could increase up to 10 % above 1.5 km. In the results section, the OEM results for the first approach (FM fed by retrievals) were not shown in this work, since the β_{par} optimizations were perfectly reproducing the truth profiles. These two approaches allowed us to better analyze the impact of C on the retrievals, and also to determine if the FM was introducing any bias to the retrievals. This discussion will arrive in the results section.

4. OEM covariance matrices

Looking for a general OEM, we have explored two different methods to build the prior covariance matrix. The first approach involved a matrix estimated on synthetic lidar data by using Mie simulations fed with OPAC database to have β_{par} profiles at 910 nm. In our case, we assumed spherical dry particles (RH: from 0 to 50%) with a predominance of coarse mode, thus for Mie code, we used a log-normal size distribution (σ :2.15, modal radius: 1.9 μm , and refractive index: $\text{Re}=1.53$, $\text{Im}=-4.10^{-3}$). For the second approach, S_a matrix was built considering the β_{par} retrieved from Klett's algorithm. The results of these tests are not shown here, but the impact of these different approaches on the error in the optimized β_{par} is presented in table2. The error has been framed as S_a : measured and S_a : scaled OPAC, to differentiate the calculations. For OPAC simulations, the OEM did not converge in the first attempts which were associated with the order of magnitude of the values, thus a scale-factor to OPAC S_a was applied. The non-convergence was expected since the OPAC data were generated for random aerosol accumulations considering some thresholds in the β_{par} profiles. To solve that, the S_a obtained from OPAC data was scaled by a factor from 0.001 to 0.0001 until we obtain plausible results compared with OEM at using S_a estimated from measurements.

Looking at the error profiles, the fact of using different S_a is not increasing the error of the retrievals in more than 4.3 % for height-dependent x_a , while for constant x_a the largest error differences found between retrievals from synthetic and measured S_a were up to 4%. These tests allowed us to have a general OEM for using prior information (i.e. constant or height-dependent) with S_a matrices either estimated from the measurements or from synthetic data. The different attempts testing scale-factors are not modifying the physical information contained in the S_a (i.e. correlation), and the results on table 2 are evidencing that the error in the optimized β_{par} is lower than 27%. Under this framework, the OEM can be used on lidar inversion without the necessity of defining a Rayleigh zone as classical inversion methods do, and additionally, the method is providing confident results either using measured or synthetic data as prior information, however, there is always a necessity to have some previous knowledge of the variables to be

optimized to improve the OEM results.

The OPAC S_a matrices were used only for OEM prior generalization, from here forward only the measurements are involved in the OEM calculations. In our case, S_a and S_y matrices were calculated from measurements but not shown here, because they are only representing the variable's variance (direction), instead of that, we have used correlation matrices to analyze the height-dependence relation between variables. In Fig.5 is shown a panel that contains the correlation matrices for x_a and y . The correlation matrices for β_{mol} are always positive and close to 1, illustrating how this variable is well height-correlated. β_{mol} correlation is almost the same for cases with LA and HA which is expected for the molecular profiles. On the contrary, the correlation matrices of β_{par} (Fig.5cd) are showing a high correlation in the main diagonal, but a rapid decrease for off-diagonals. This decrease is linked with the atmospheric limit between aerosol and molecules which in most of the cases might be associated with the atmospheric boundary layer height (ABLH). Regarding that each bin corresponds to a 50 m altitude-step, both matrices for LA and HA cases allowed to see how the ABLH is associated to the negative values of the correlation in the off-diagonals (from 350 m to 900 m for LA and from 350 m to 1.1 km for HA). β_{par} correlation matrices are highly correlated above the aerosol zone reaching values up to 0.9, indicating a positive linear tendency in height.

The correlation matrices for y are shown in Fig.5ef. These matrices are obtained from the rcs measurements also classified among LA and HA cases. A high correlation is seen in the main diagonal and its vicinity, decreasing for the off-diagonals. ABLH effect is also seen in the region where the correlation becomes negative at the same altitudes mentioned above, but here it must be considered that rcs profiles are quite noisier than β_{mol} and β_{par} ones, leading that correlation above 1.8 km (above bin 25) seems less uniform with the combination of positive and negative values.

5. Case of study: retrievals on 11th and 26th October 2019

5.1. One-single component retrieval

In order to illustrate the potentiality of the OEM applied to ceilometers inversions, we have applied the method to the following days from the data set, 11th and 26th October 2019, that contained two different atmospheric conditions, HA and LA respectively. During the early morning, from 7 to 10 UTC a low aerosol accumulation was measured on 26th October, while on 11th October from 8 to 16 UTC an aerosol accumulation started to increase around 1.5 km. On this regard, we have used these two atmospheric scenarios to apply the inversion using as LA the profile measured at 7 UTC, and the one at 11 UTC as HA. As it was shown in Sec. 4.3, to determine if the FM has some bias, we have used two y vectors. Reminding that measured profiles are labelled as "obs" and the optimized as "opt", in Fig.6c are presented the results for the rcs obtained from the FM(β_{par}^{obs}) at 7 UTC (cyan line, Fig.6c), in blue line the rcs calibrated (rcs_{obs}), and in magenta is the rcs obtained from FM(β_{par}^{opt}). These calculations lead us to conclude that FM is not adding any bias to the optimization, since the rcs profiles obtained from FM(β_{par}^{opt}) and rcs_{obs} are almost the same, indicating the good performance of the OEM. To support that, we have performed the linear fitting between FM(β_{par}^{opt}) and rcs_{obs}, finding a correlation coefficient $R^2 = 0.92$ and a root mean square error (RMSE) up to 0.12, showing a high linear dependency between both profiles. The slightly losing of linearity is due to the ABLH transition zone where a small cloud of points is randomly distributed in the linear fitting, nevertheless, the fit has shown a high correlation.

During the inversions different configurations were tested, involving variations in the γ parameter (i.e. between Gauss-Newton and Levenberg-Marquardt methods) introduced in Sec.2 without leading to major improvements in the retrieval for γ larger than 1. In our case, we use $\gamma_i = 1000, 300, 100, 30, 10, 3, 1, 1, 1$, resulting in a convergence at $\gamma = 1$. As γ factor decrease with each iteration, it acts as a weighing factor between the prior and the observation, however in our case the convergence was always reached for $\gamma = 1$ meaning that this scaling factor was not relevant and Gauss-Newton is the best approach to our ill-posed problem. As first results, not shown here, we ran the OEM by adding from one to three sub-diagonal terms from the estimated covariance matrix (S_a), but the retrievals rapidly converged to the prior while the number of diagonals increased, and the A_k started to lose periodicity, leading to an increase of the width of the peaks, mainly for altitudes where the profile presented high variability due to aerosol accumulations or ABL transitions. This effect drives to the lost of sensitivity in the A_k , inducing singularities in the error estimation of β_{par}^{opt} . Under these covariance conditions, the inversions were converging after 6 iterations, but the linearity chi-square (χ^2) criterion was never fulfilled.

In Fig.6a is presented the retrieval on 26th October 2019 at 7 UTC under LA scenario. The truth profile (β_{par}^{obs} , Fig.6, black line) was calculated by using Klett inversion as it was mentioned in previous sections, but this profile is not involved in the OEM process, therefore it becomes a reference to compare the results. The red and orange lines in Fig.6a represent the prior and β_{par}^{opt} respectively. In a general context, the results were quite accurate, however in the first 500 m, the bias between Klett and OEM inversion shown that β_{par}^{opt} underestimated the β_{par}^{obs} up to -0.5 (kmsr)^{-1} , and this bias increased above 1.0 km agl reaching up to 0.5 (kmsr)^{-1} (red dotted line, Fig.6b). In the same figure, we presented the error in β_{par}^{opt} (Fig.6b, gray solid line), which quantified as the relative error in percentage, reached in the first 1.5 km up to 25% and above this height the error was lower than 15%.

As the β_{par}^{opt} is a height-dependent atmospheric profile, the A_k represents how the information is distributed along with the profile as a function of the altitude, and the width of the peaks of an individual A_k is the vertical resolution of the measurements at that point (i.e. equal to or greater than the retrieval grid spacing). The previously discussed above leads us to consider in the extreme scenario the use of the S_a main-diagonal plus one off-diagonal term, therefore an additional relation between bins (heights) drove to A_k were not perfect delta Dirac functions as expected in cases where just the diagonal terms are considered. In our case, the A_k peaks are representing the bins constraints, interpreting this information as the model sensibility. The A_k (Fig.6d) are quite stables with peaks up to the same width up to 200 m with slightly less sensibility in the last part of the profile. Regarding the degrees of freedom, they pointed out that once we have an S_y matrix, the constraint provided by S_a gave us about 51.4 independent pieces of information for the β_{par}^{opt} retrieval.

Following the same analysis, the OEM was applied to a HA case on 11th October 2019 at 11 UTC. The β_{par}^{opt} obtained is shown in Fig.7a in orange line. Here, we have used the main diagonal term of the S_a , because additional the sub-diagonals terms were generating convergence difficulties. This retrieval presented more variability in terms of the bias, underestimating β_{par}^{obs} values in the first 0.5km (up to -0.2 (kmsr)^{-1}), then from the bias reached up -0.5 (kmsr)^{-1} close to 1 km. Above this height, the bias remained around 0.3 to 0.5 (kmsr)^{-1} . One relevant result is that β_{par}^{opt} reproduced satisfactorily the accumulation events of the β_{par}^{obs} , showing the lower bias values over those regions. The A_k peaks were slightly regulars mostly for the upper altitudes (e.g above 1 km), but the peaks were highly variable from 600 to 800 m, close to the ABLH boundary where β_{par}^{obs} presented an abrupt change. Finally, the error presented for this case is lower than 30% below 1.5 km agl, and lower than 15% above this height, considering this particular configuration for OEM.

5.2. Ceilometer parameters retrieval

This section is devoted to present the OEM results at retrieving the ceilometer parameters, C and LR. For this calculation, the x_a vector was built as it was explained in Sec. 4.2, under the same classification between LA and HA. Here, we retrieved all cases (39) considering a modification in the Eq.4, leading us to define a new FM, neglecting the range correction and splitting the α_p term as the product of β_p LR_{par}. For this retrieval we have considered S_a as a matrix that contains the standard deviations of the parameters, thus each case was treated independently. This approach leads us to fix β_p in the FM, thus we have used the mean value of this profile depending on the classification for LA or HA cases.

In Fig.8ab are shown the bias results considering C from Eq.6, and LR from Klett inversion as the corresponding references. The OEM for the calibration constant retrieval has shown some variability between positive and negative bias, but always ranging from values between -5 to $4 \text{ km}^3\text{sr}$. The results for LR under the same classification of LA cases shown that retrievals were ranging from -0.6 to 0.5 sr . Analyzing the HA cases, the retrievals were less accurate mostly for LR. The first 4 cases reached negative bias up to $-4.5 \text{ km}^3\text{sr}$ for C and -1.0 for LR, meanwhile, the last two cases presented a positive bias up to $2.5 \text{ km}^3\text{sr}$ and 0.5 sr , for C and LR respectively. The calculated bias reinforce the fact that under this OEM configuration for LA cases, the absolute errors are not surpassing 0.5% and 1 % for C and LR respectively, but for HA classification larger errors were obtained for LR, like those obtained for case 4 where the absolute error was up to 2.3 %. Unless the relatively low error presented, it is recommendable to try with a larger database for different types of aerosol mainly for particular events like biomass burning, Saharan dust intrusions, and urban aerosol, therefore a seasonal analysis is needed for further studies in order to better constrain the problem.

6. Conclusions

In this work, an optimal estimation method (OEM) was applied to the Vaisala CL51 ceilometer installed at ONERA (The French Aerospace Lab, Toulouse, France) on measurements performed from October 2019 to February 2020. On this framework, we retrieve the β_{par} and two lidar parameters, one related to the system efficiency, C , and the other one related to an intensive lidar property so-called lidar ratio (LR). This ill-posed problem was tackled by considering a set of measurements classified under two different atmospheric frames, low aerosol accumulation and high aerosol accumulation events, and a sensibility analysis on the prior information was performed by considering the retrievals obtained from Klett method and from synthetic lidar data calculated by using Mie theory fed by aerosol properties from OPAC database.

A previous data set of Klett inversions of β_{par} was used to define the prior information, and then some tests were done considering constant and height-dependence priors for different aerosol accumulations (low and high aerosol accumulation events). As a result, we found that the largest differences are in the first 500 m agl and the above 1.5 km agl for constant priors, reaching relative errors in percentage up to 5% related to the height-dependent ones. Additionally, we have tested two schemes for OEM retrievals, firstly by using a prior covariance matrix estimated from synthetic data and secondly, we used the covariance matrix estimated from the previous Klett retrievals. The relative errors were also analyzed for these approaches, finding that errors were lower than 2% at using one instead of the other. On this way, the a priori information for the OEM was generalized to measurements or synthetic lidar signals.

The OEM was applied in two cases, one for low aerosol accumulation and the other one for high accumulation. From this classification, 33 cases were classified as low accumulation and 6 cases as high accumulation events. Regarding the β_{par} results, we found that S_a played a determinant role in the retrievals. The best results for the LA scheme were obtained using the main diagonal terms of the S_a and the measurement vector S_y . In the first 500m the optimized β_{par} presented a positive bias up to 0.1, and a negative bias for the last kilometer up to -0.5. The perceptual relative error of the optimized profiles can reach up to 25% until 1.5 km agl, and above this height decrease up to 15%. For the high accumulation case, a positive bias up to 0.3 was found in the first 500 m agl, while a negative bias close to -0.5 was found in the last kilometre (from 2 to 3 km agl). The perceptual relative error of the optimized profiles was up to 30% in the first 500 m agl, and lower than 15% above. Considering the averaging kernels (A_k), the sensitivity of the inversions was practically the same. For low accumulations A_k presented periodical peaks with a FWHM up to 200m, meanwhile, for high accumulations, the size of the peaks near the first accumulation zone was slightly lower than the rest of the profile, however, the FWHM did not present variability. The error estimation of the lidar products is always a concern when inversions methods are used. In practice, the error is always hard to propagate due to the lack of knowledge of the system parameters, but in our case, the OEM is overcoming that, since the error is directly calculated. The behavior of the error for the cases tested here has shown larger values in the first meters (i.e. 300 m), and some tendency to increase in the vicinity of ABL height or accumulation event.

For the C and LR retrievals, the results showed that both quantities can be obtained by using OEM relatively well. The bias for C was ranging from -5 to 4 km³sr, while LR bias was from -0.6 to 0.5 sr for low accumulation. Under high accumulation events, the C bias was ranging from -4 to 3 and for LR from -0.1 to 0.5 sr. Using this OEM framework, the error in the retrieved products were larger for high accumulations cases, reaching up to 2.3% in the worst scenario over the data base selected.

Generally, it is important to keep increasing the database, because of the necessity to have information about different aerosol types with peaks distributed along different altitudes to keep feeding the OEM, in order to improve the retrievals under almost all atmospheric conditions. Besides, we are planning in the near future to extend this retrieval method to short-range lidar systems in order to obtain lidar products under no possibilities of having a molecular zone, and also to improve the knowledge about lidar parameters such as lidar ratio for different aerosol types.

CRedit authorship contribution statement

A.E. Bedoya-Velázquez: Writing - Original draft preparation, methodology, algorithm design. **Romain Ceolato:** Conceptualization of this study, methodology, supervision, project administration and funding acquisition. **Sidonie Lefebvre:** Conceptualization of this study, methodology, algorithm design, supervision, project administration and funding acquisition.

References

- Barrera-Verdejo, M., Crewell, S., Löhnert, U., Orlandi, E., Di Girolamo, P., 2016. Ground-based lidar and microwave radiometry synergy for high vertical resolution absolute humidity profiling. *Atmospheric Measurement Techniques* 9, 4013–4028. URL: <https://www.atmos-meas-tech.net/9/4013/2016/>, doi:10.5194/amt-9-4013-2016.
- Bedoya-Velásquez, A.E., Navas-Guzmán, F., Granados-Muñoz, M.J., Titos, G., Román, R., Casquero-Vera, J.A., Ortiz-Amezcuca, P., Benavent-Oltra, J.A., de Arruda Moreira, G., Montilla-Rosero, E., Hoyos, C.D., Artiñano, B., Coz, E., Olmo-Reyes, F.J., Alados-Arboledas, L., Guerrero-Rascado, J.L., 2018. Hygroscopic growth study in the framework of earlinet during the slope i campaign: synergy of remote sensing and in situ instrumentation. *Atmospheric Chemistry and Physics* 18, 7001–7017. URL: <https://www.atmos-chem-phys.net/18/7001/2018/>, doi:10.5194/acp-18-7001-2018.
- Bedoya-Velásquez, A., Herreras-Giralda, M., Román, R., Wiegner, M., Lefebvre, S., Toledano, C., Huet, T., Ceolato, R., 2021. Ceilometer inversion method using water-vapor correction from co-located microwave radiometer for aerosol retrievals. *Atmospheric Research* 250, 105379. URL: <http://www.sciencedirect.com/science/article/pii/S0169809520313168>, doi:<https://doi.org/10.1016/j.atmosres.2020.105379>.
- Bösch, T., Rozanov, V., Richter, A., Peters, E., Rozanov, A., Wittrock, F., Merlaud, A., Lampel, J., Schmitt, S., de Haij, M., Berkhout, S., Henzing, B., Apituley, A., den Hoed, M., Vonk, J., Tiefengraber, M., Müller, M., Burrows, J.P., 2018. Boreas – a new max-doas profile retrieval algorithm for aerosols and trace gases. *Atmospheric Measurement Techniques* 11, 6833–6859. URL: <https://amt.copernicus.org/articles/11/6833/2018/>, doi:10.5194/amt-11-6833-2018.
- Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S.K., Sherwood, S., Stevens, B., Zhang, X.Y., 2013. *Clouds and aerosols*. Cambridge University Press, Cambridge, UK. pp. 571–657. doi:10.1017/CB09781107415324.016.
- Ceolato, R., Bedoya-Velásquez, A.E., Mousysset, V., 2020. Short-range elastic backscatter micro-lidar for quantitative aerosol profiling with high range and temporal resolution. *Remote Sensing* 12. URL: <https://www.mdpi.com/2072-4292/12/20/3286>, doi:10.3390/rs12203286.
- Delanoë, J., Hogan, R.J., 2008. A variational scheme for retrieving ice cloud properties from combined radar, lidar, and infrared radiometer. *Journal of Geophysical Research: Atmospheres* 113. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009000>, doi:10.1029/2007JD009000, arXiv:<https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007JD009000>.
- Farhani, G., Sica, R.J., Godin-Beekmann, S., Ancellet, G., Haefele, A., 2019a. Improved ozone dial retrievals in the upper troposphere and lower stratosphere using an optimal estimation method. *Appl. Opt.* 58, 1374–1385. URL: <http://ao.osa.org/abstract.cfm?URI=ao-58-6-1374>, doi:10.1364/AO.58.001374.
- Farhani, G., Sica, R.J., Godin-Beekmann, S., Haefele, A., 2019b. Optimal estimation method retrievals of stratospheric ozone profiles from a dial. *Atmospheric Measurement Techniques* 12, 2097–2111. URL: <https://www.atmos-meas-tech.net/12/2097/2019/>, doi:10.5194/amt-12-2097-2019.
- Herreras, M., Román, R., Cazorla, A., Toledano, C., Lyamani, H., Torres, B., Cachorro, V.E., Olmo, F.J., Alados-Arboledas, L., de Frutos, Á.M., 2018. Integrated aerosol extinction profiles from ceilometer and sunphotometer combination against sunphotometer measurements at various heights. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 7564–7567.
- de Jesus, W.C., Landulfo, E., 2009. Retrieving atmospheric properties with an optimal estimation inverse method of lidar measurements, in: Singh, U.N., Pappalardo, G. (Eds.), *Lidar Technologies, Techniques, and Measurements for Atmospheric Remote Sensing V*, International Society for Optics and Photonics. SPIE. pp. 67 – 75. URL: <https://doi.org/10.1117/12.830355>, doi:10.1117/12.830355.
- Klett, J.D., 1981. Stable analytical inversion solution for processing lidar returns. *Appl. Opt.* 20, 211–220. URL: <http://ao.osa.org/abstract.cfm?URI=ao-20-2-211>, doi:10.1364/AO.20.000211.
- Klett, J.D., 1985. Lidar inversion with variable backscatter/extinction ratios. *Appl. Opt.* 24, 1638–1643. URL: <http://ao.osa.org/abstract.cfm?URI=ao-24-11-1638>, doi:10.1364/AO.24.001638.
- Kotthaus, S., O'Connor, E., Münkel, C., Charlton-Perez, C., Haefelin, M., Gabey, A.M., Grimmond, C.S.B., 2016. Recommendations for processing atmospheric attenuated backscatter profiles from vaisala cl31 ceilometers. *Atmospheric Measurement Techniques* 9, 3769–3791. URL: <https://www.atmos-meas-tech.net/9/3769/2016/>, doi:10.5194/amt-9-3769-2016.
- Lopatin, A., Dubovik, O., Chaikovskiy, A., Goloub, P., Lapyonok, T., Tanré, D., Litvinov, P., 2013. Enhancement of aerosol characterization using synergy of lidar and sun-photometer coincident observations: the garrlic algorithm. *Atmospheric Measurement Techniques* 6, 2065–2088. URL: <https://www.atmos-meas-tech.net/6/2065/2013/>, doi:10.5194/amt-6-2065-2013.
- Marcos, C.R., Gómez-Amo, J.L., Peris, C., Pedrós, R., Utrillas, M.P., Martínez-Lozano, J.A., 2018. Analysis of four years of ceilometer-derived aerosol backscatter profiles in a coastal site of the western mediterranean. *Atmospheric Research* 213, 331 – 345. URL: <http://www.sciencedirect.com/science/article/pii/S016980951830468X>, doi:<https://doi.org/10.1016/j.atmosres.2018.06.016>.
- Nanda, S., de Graaf, M., Sneep, M., de Haan, J.F., Stammes, P., Sanders, A.F.J., Tuinder, O., Veefkind, J.P., Levelt, P.F., 2018. Error sources in the retrieval of aerosol information over bright surfaces from satellite measurements in the oxygen a band. *Atmospheric Measurement Techniques* 11, 161–175. URL: <https://www.atmos-meas-tech.net/11/161/2018/>, doi:10.5194/amt-11-161-2018.
- Pappalardo, G., Amodeo, A., Pandolfi, M., Wandinger, U., Ansmann, A., Bösenberg, J., Matthias, V., Amiridis, V., Tomasi, F.D., Frioud, M., Iarlori, M., Komguem, L., Papayannis, A., Rocadenbosch, F., Wang, X., 2004. Aerosol lidar intercomparison in the framework of the earlinet project. 3. ramanlidar algorithm for aerosol extinction, backscatter, and lidar ratio. *Appl. Opt.* 43, 5370–5385. URL: <http://ao.osa.org/abstract.cfm?URI=ao-43-28-5370>, doi:10.1364/AO.43.005370.
- Pounder, N.L., Hogan, R.J., Várnai, T., Battaglia, A., Cahalan, R.F., 2012. A Variational Method to Retrieve the Extinction Profile in Liquid Clouds Using Multiple-Field-of-View Lidar. *Journal of Applied Meteorology and Climatology* 51, 350–365. URL: <https://doi.org/10.1175/JAMC-D-10-05007.1>, doi:10.1175/JAMC-D-10-05007.1, arXiv:https://journals.ametsoc.org/jamc/article-pdf/51/2/350/3564639/jamc-d-10-05007_1.pdf.
- Povey, A.C., Grainger, R.G., Peters, D.M., Agnew, J.L., 2014. Retrieval of aerosol backscatter, extinction, and lidar ratio from raman lidar with optimal estimation. *Atmospheric Measurement Techniques* 7, 757–776. URL: <https://www.atmos-meas-tech.net/7/757/2014/>,

doi:10.5194/amt-7-757-2014.

- Rodgers, C.D., 2000. Inverse Methods for Atmospheric Sounding. WORLD SCIENTIFIC. URL: <https://www.worldscientific.com/doi/abs/10.1142/3171>, doi:10.1142/3171, arXiv:<https://www.worldscientific.com/doi/pdf/10.1142/3171>.
- Román, R., Benavent-Oltra, J., Casquero-Vera, J., Lopatin, A., Cazoria, A., Lyamani, H., Denjean, C., Fuertes, D., Pérez-Ramírez, D., Torres, B., Toledano, C., Dubovik, O., Cachorro, V., [de Frutos], A., Olmo, F., Alados-Arboledas, L., 2018. Retrieval of aerosol profiles combining sunphotometer and ceilometer measurements in grasp code. Atmospheric Research 204, 161 – 177. URL: <http://www.sciencedirect.com/science/article/pii/S0169809517312577>, doi:<https://doi.org/10.1016/j.atmosres.2018.01.021>.
- Sica, R.J., Haeefe, A., 2015. Retrieval of temperature from a multiple-channel rayleigh-scatter lidar using an optimal estimation method. Appl. Opt. 54, 1872–1889. URL: <http://ao.osa.org/abstract.cfm?URI=ao-54-8-1872>, doi:10.1364/AO.54.001872.
- Sica, R.J., Haeefe, A., 2016. Retrieval of water vapor mixing ratio from a multiple channel raman-scatter lidar using an optimal estimation method. Appl. Opt. 55, 763–777. URL: <http://ao.osa.org/abstract.cfm?URI=ao-55-4-763>, doi:10.1364/AO.55.000763.
- Veselovskii, I., Kolgotin, A., Griaznov, V., Müller, D., Wandinger, U., Whiteman, D.N., 2002. Inversion with regularization for the retrieval of tropospheric aerosol parameters from multiwavelength lidar sounding. Appl. Opt. 41, 3685–3699. URL: <http://ao.osa.org/abstract.cfm?URI=ao-41-18-3685>, doi:10.1364/AO.41.003685.
- Wiegner, M., Gasteiger, J., 2015. Correction of water vapor absorption for aerosol remote sensing with ceilometers. Atmospheric Measurement Techniques 8, 3971–3984. URL: <https://www.atmos-meas-tech.net/8/3971/2015/>, doi:10.5194/amt-8-3971-2015.
- Wiegner, M., Madonna, F., Biniotoglou, I., Forkel, R., Gasteiger, J., Geiß, A., Pappalardo, G., Schäfer, K., Thomas, W., 2014. What is the benefit of ceilometers for aerosol remote sensing? an answer from earlinet. Atmospheric Measurement Techniques 7, 1979–1997. URL: <https://www.atmos-meas-tech.net/7/1979/2014/>, doi:10.5194/amt-7-1979-2014.

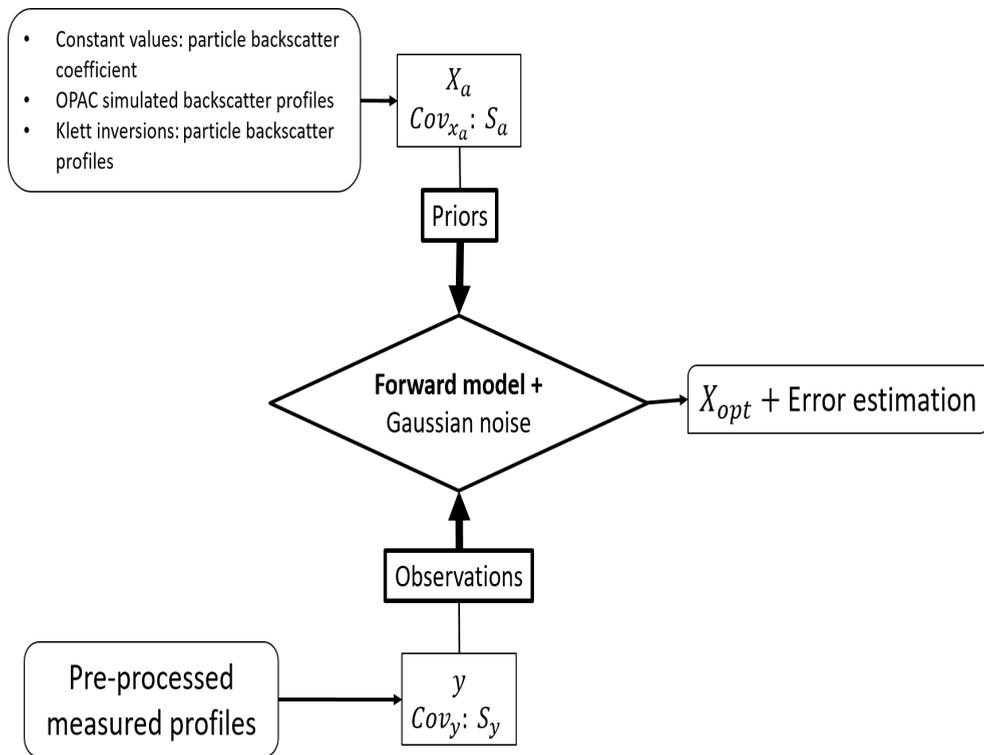


Figure 1: Algorithm block diagram

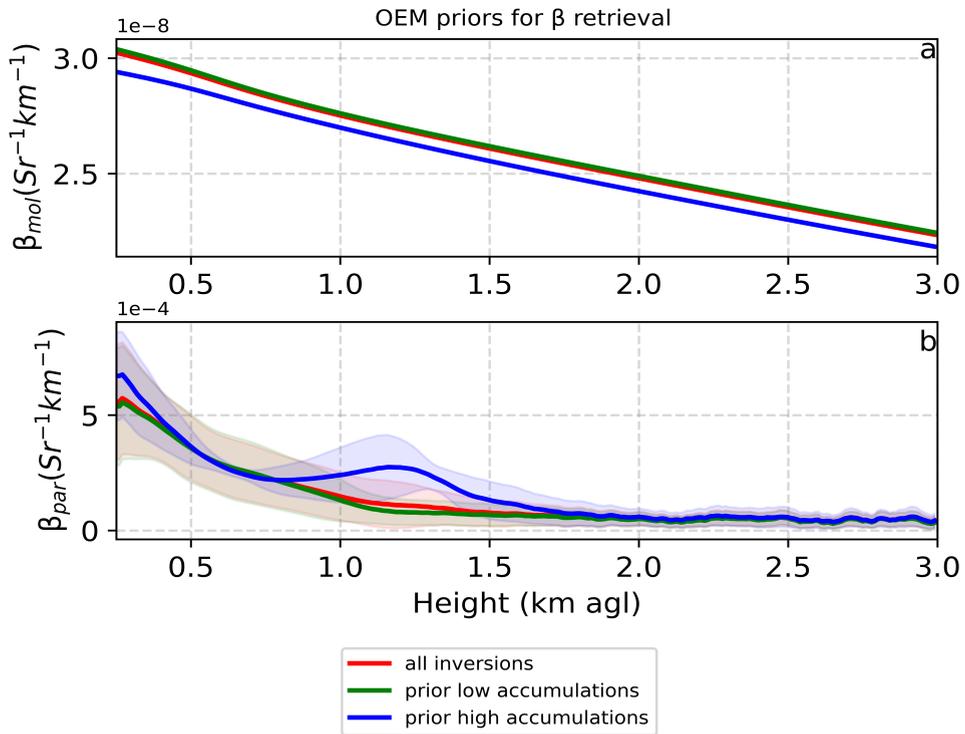


Figure 2: The panel presents in a) the β_{mol} , and in b) the β_{par} profiles used as priors for the OEM retrieval of β_{par} . In red are represented the mean values of the profiles considering all cases, in green those ones associate with LA events, and finally in blue are shown the profiles for HA cases.

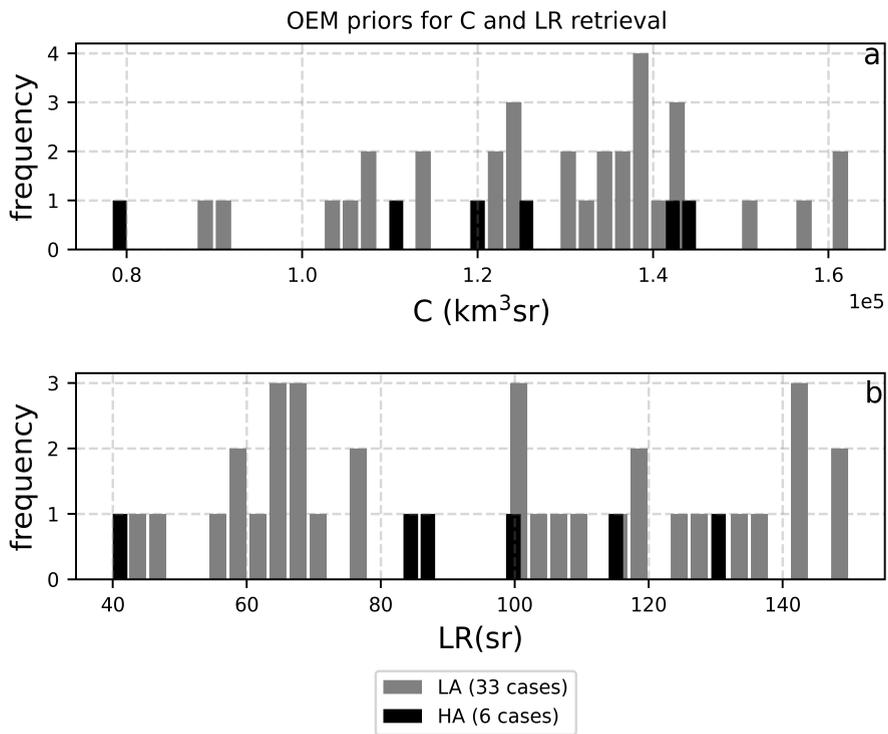


Figure 3: In the upper part is shown the data distribution of the C values from October 2019 to February 2020, while in the lower part LR values are presented considering the corresponding classification

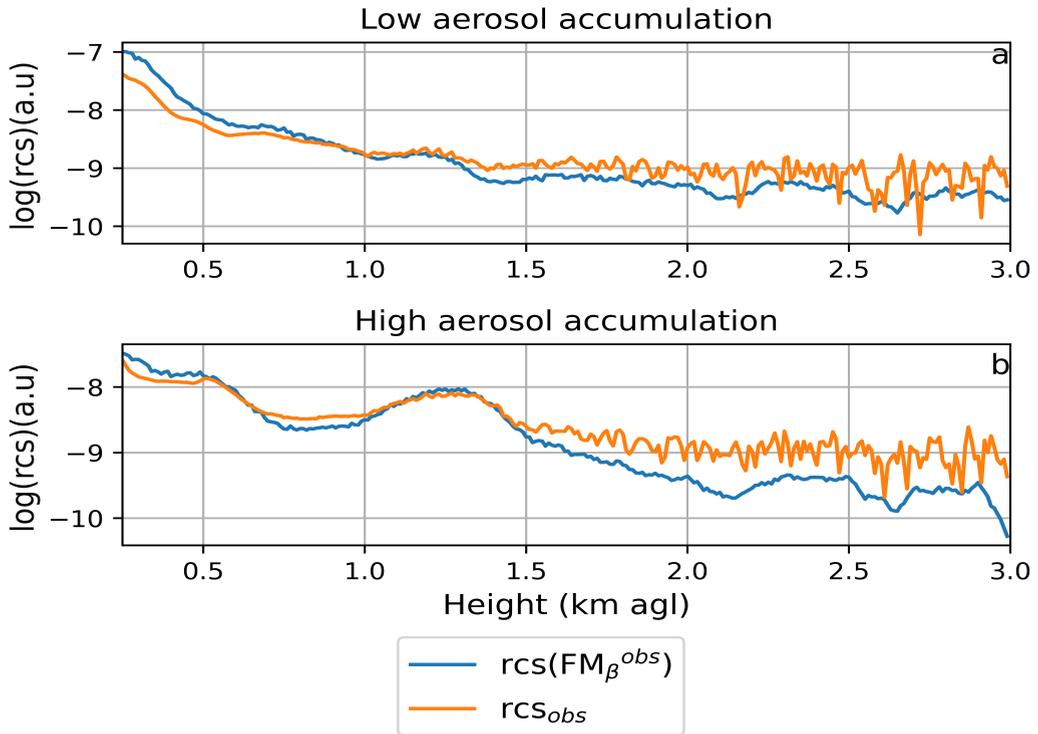


Figure 4: The panel presents an example case of the two schemes used for OEM retrievals. In a) are shown the profiles under LA classification, while in b) are shown for HA. The blue line refers to the β_{par} passed through the forward model and the orange line the measured r_{cs}_{obs} .

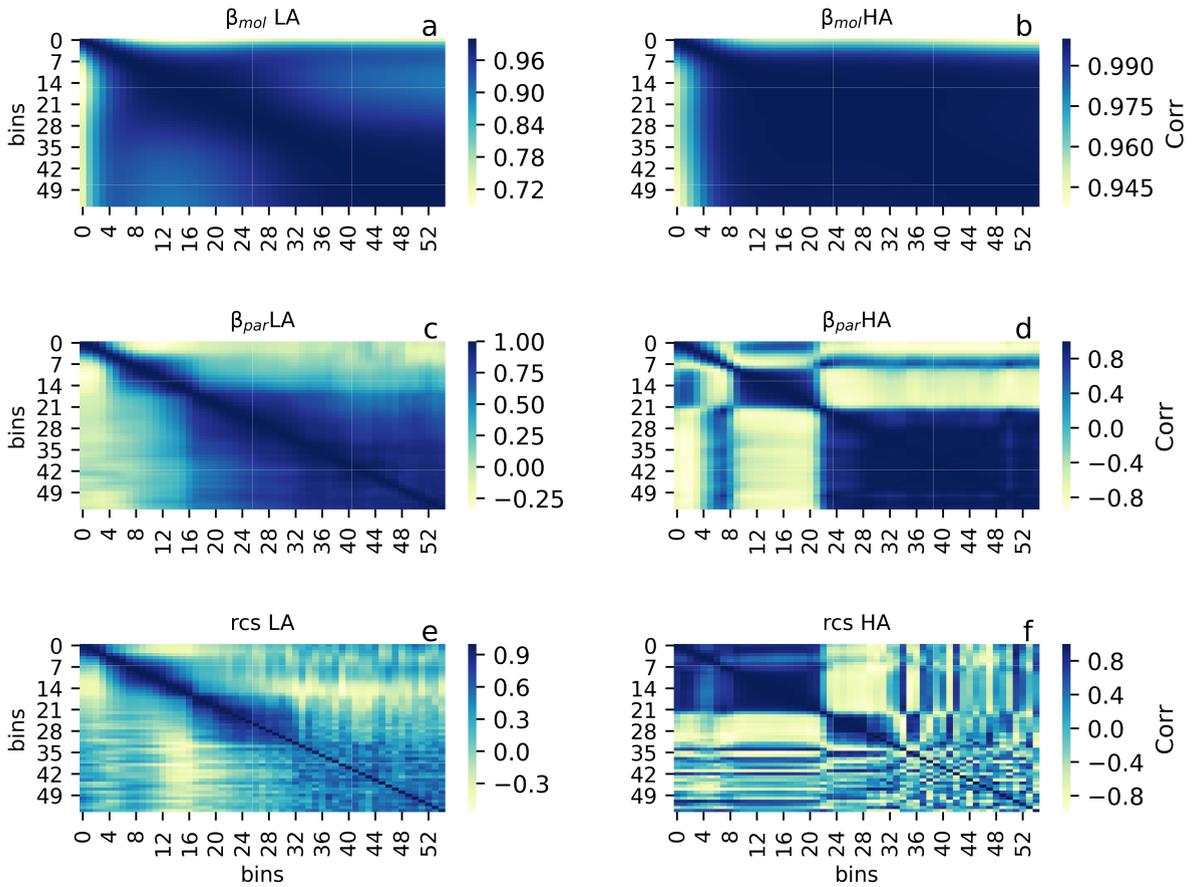


Figure 5: The panel contains the correlation matrices for the priors and measurement vectors used under each classification. β_{mol} is presented in the upper part, panels a and b, β_{par} in middle (cd), and in the bottom the rcs's (ef) are presented.

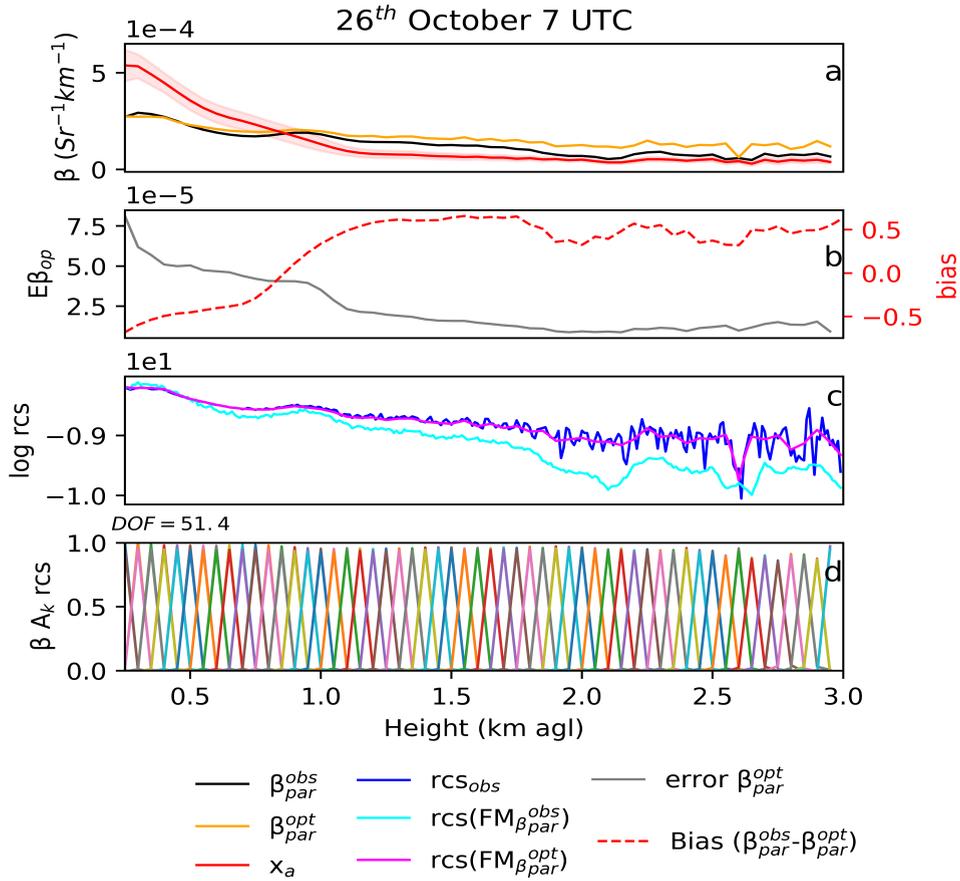


Figure 6: The OEM retrievals for LA classification on 26th October 2019 at 7 UTC are shown. In the a) panel are presented the truth profile (black line), the OEM is shown in orange, while the red line is devoted to the prior. In the panel b) are presented the errors of the optimized profile in solid gray line, and dotted gray and red lines are showing the FM bias for β_{par}^{opt} and the prior, respectively. In c) are presented the rcs results, where the blue line refers to the measured rcs profile, in cyan is presented the rcs obtained from the forward model fed by the measured β_{par} , and the magenta line is showing the rcs as a result of the forward model fed by the optimized β_{par} . The bottom panel is the A_k of the rcs.

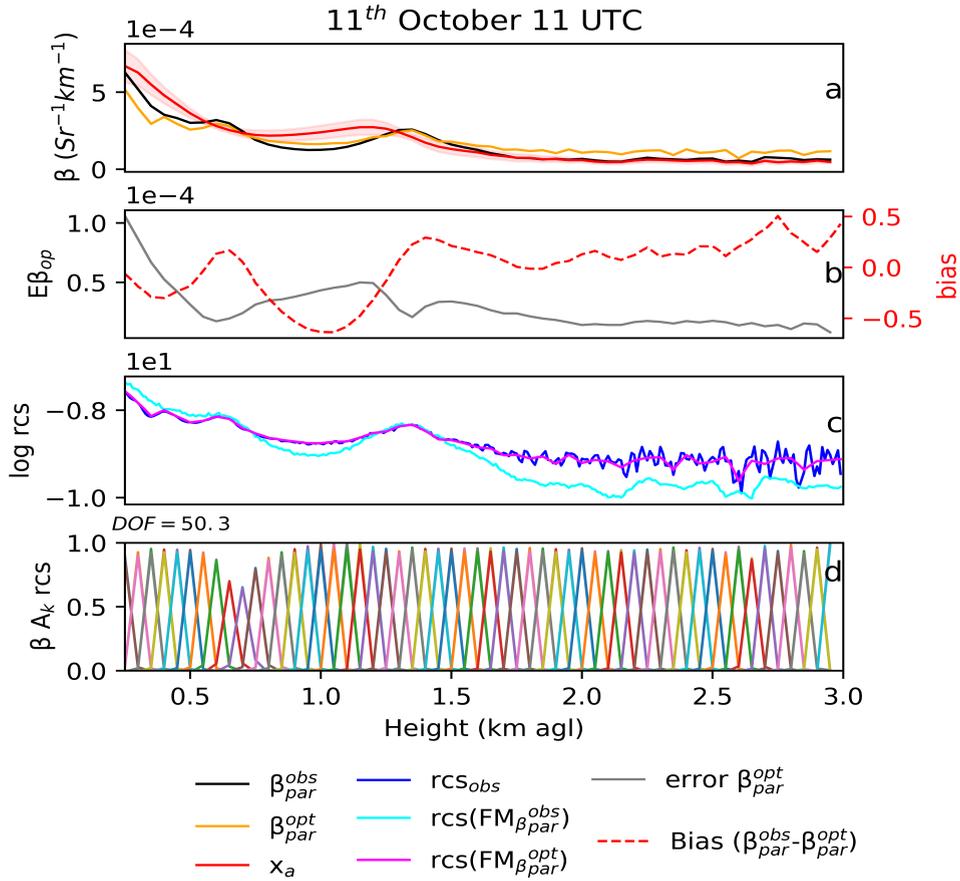


Figure 7: Example case on 11th October 2019 at 11 UTC for the OEM retrievals for HA classification. In a) are presented the truth profile (black line), the OEM showed in orange, while the red line is the prior. In the panel b) are presented the errors of the optimized profile in grey solid lines, and dotted gray and red lines are the FM bias for β_{par}^{opt} and the prior, respectively. In c) are presented the rcs results, where the blue line refers to the measured rcs profile, in cyan is presented the rcs obtained from the forward model fed by the measured β_{par} , and the magenta line is showing the rcs as a result of the forward model fed by the optimized β_{par} . The bottom c is showing the rcs A_k .

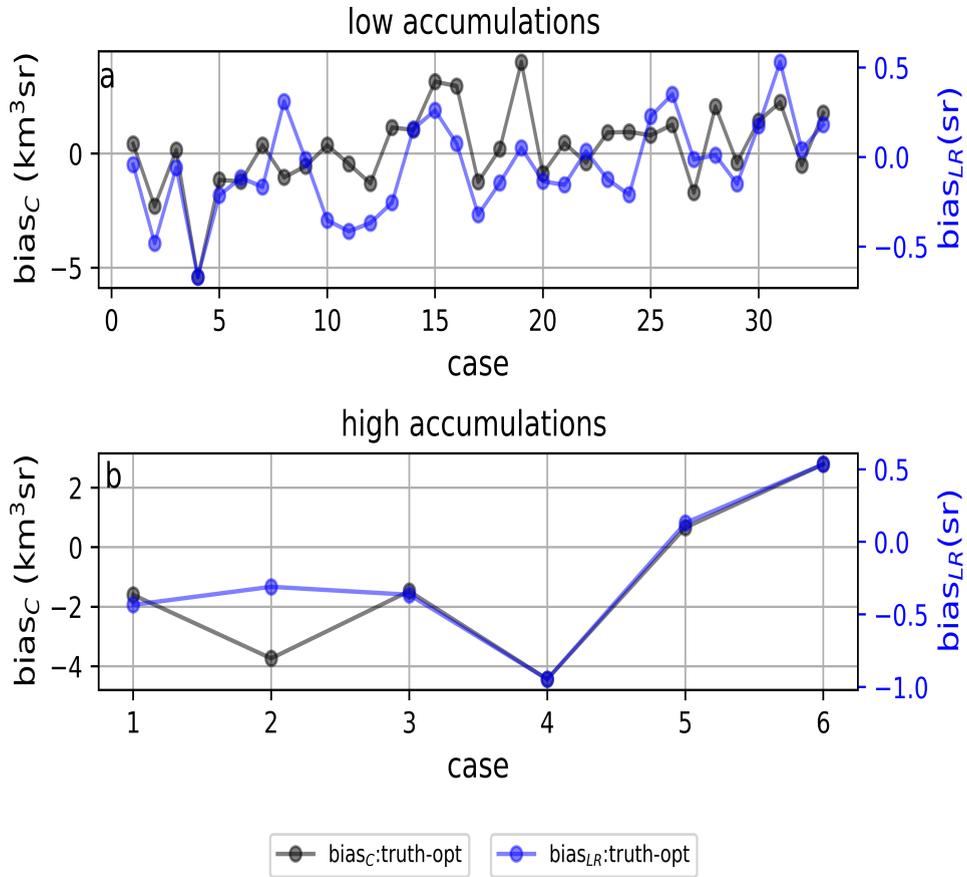
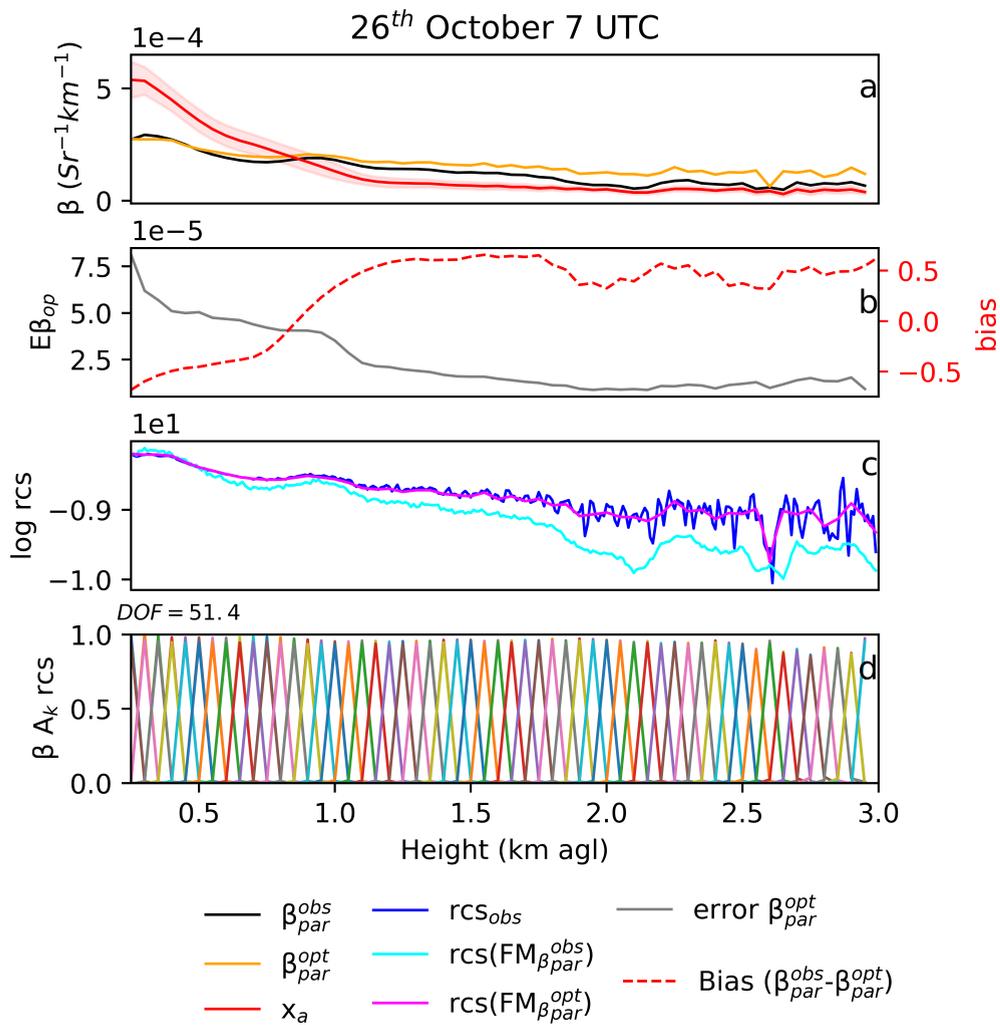


Figure 8: The panel presents the evolution of the bias for the OEM retrievals for each parameter optimized.

Graphical Abstract

Optimal estimation method applied on ceilometer aerosol retrievals

A.E. Bedoya-Velásquez, Romain Ceolato, Sidonie Lefebvre



Highlights

Optimal estimation method applied on ceilometer aerosol retrievals

A.E. Bedoya-Velásquez, Romain Ceolato, Sidonie Lefebvre

- Application of nonlinear regression to solve elastic lidar ill-posed problems for retrieve aerosol properties.
- Exploration of different prior scenarios to better constrain the aerosol inversion using the optimal estimation method.
- Study of the optimal estimation method capabilities in contrast to the classic elastic Klett approach on lidar inversion.