A Bayesian parametric approach to the retrieval of the atmospheric number size distribution from lidar data

Alberto Sorrentino [1], Alessia Sannino [2], Nicola Spinelli [2], Michele Piana [1], Antonella Boselli [3], Valentino Tontodonato [5], Pasquale Castellano [5], and Xuan Wang [4]

[1]Dipartimento di Matematica, Università di Genova
[2]Dipartimento di Fisica, Università di Napoli Federico II
[3]CNR–IMAA, Potenza
[4]CNR–SPIN, Napoli
[5]ALA Srl Advanced Lidar Applications, Napoli

Correspondence: Alberto Sorrentino (sorrentino@dima.unige.it)

Abstract. We consider the problem of reconstructing the number size distribution (or particle size distribution) in the atmosphere from lidar measurements of the extinction and backscattering coefficients. We assume that the number size distribution can be modelled as a superposition of log–normal distributions, each one defined by three parameters: mode, width and height. We use a Bayesian model and a Monte Carlo algorithm to estimate these parameters. We test the developed method on synthetic data generated by distributions containing one or two modes, and perturbed by Gaussian noise, and on three datasets obtained from AERONET. We show that the proposed algorithm provides good results when the right number of modes is selected. In general, an over-estimate of the number of modes provides better results than an under-estimate. In all cases, the PM1, PM2.5 and PM10 concentrations are reconstructed with tolerable deviations.

1 Introduction

LIDAR (LIght Detection and Ranging) is a remote sensing technique similar to RADAR (RAdio Detecting And Ranging) which uses light in the form of short laser pulses to invest a target and obtain, through elastic and inelastic scattering processes, information on the target properties as a function of the distance from the lidar system.

In the atmospheric application, LIDAR systems can be used to obtain spatially resolved information about the optical properties of the atmospheric aerosols (desert dust, volcanic ash, smog and many other types of substances)(Giannakaki et al., 2010; Ritter et al., 2018; Lee and Wong, 2018; Stelitano et al., 2019; Chazette, 2020), over a distance of several kilometers and with high spatial and temporal resolutions. Advanced, multi-wavelength lidar systems are capable of giving information on the spectral dependence of the aerosol optical properties, and can therefore be used to obtain information on the aerosol microphysical properties (Weitkamp, 2006; Pérez-Ramírez et al., 2013; Granados-Muñoz et al., 2014; Giannakaki et al., 2016; Müller et al., 2016; Chemyakin et al., 2016; Ortiz-Amezcua et al., 2017; Molero et al., 2020).

However, information on the microphysical properties of the atmospheric aerosols is seldom obtained using the LIDAR signal alone. This information, which is essential for a complete aerosol characterization useful to understand their effect

on climate, is instead frequently obtained through the synergistic use of *in-situ* instruments; incidentally these measurements also allow a validation of the lidar retrievals, but only for those values that are closest to the ground, and for a particular aerosol typology (Saharan dust, Biomass burning aerosol,...); alternatively, validation can be done using synthetic data (Alados-Arboledas et al., 2011; Di Girolamo et al., 2012; Veselovskii et al., 2013; Osterloh et al., 2013; Samaras et al., 2015; Whiteman et al., 2018).

In order to retrieve the microphysical properties of the aerosol from lidar measurements, two inverse problems must be solved in sequence: in the first inverse problem, one uses the measured backscattered power to obtain an estimate of the aerosol optical parameters; in the second problem, one uses the estimated optical parameters (at different wavelengths), derived from lidar observations, to obtain an estimate of the number size distribution, i.e. the density of particles as a function of the particle size. This latter problem is particularly challenging because of the limited amount of data, due to the practical problem of measuring at many different wavelengths. Moreover, from a mathematical point of view, the microphysical parameters are generally derived from the optical ones through integral equations that cannot be solved analytically and whose numerical solution leads to a so-called ill-posed problem. This last is characterized by a strong sensitivity of the solution from the input data uncertainties and by the non-uniqueness of the solution. Remarkably, from a mathematical viewpoint there can be several ways to overcome ill-posedness, however, not all of them actually reflect realistic physical conditions. In addition, numerical studies showed that a poor selection of the constraints can affect the quality of the solution and compromise the microphysical retrieval in spite of the strength of the regularization algorithm. Therefore, in order to obtain stable and physically acceptable solutions, mathematical and physical constraints variously combined with regularization methods are applied (Müller et al., 1999; Böckmann, 2001; Veselovskii et al., 2002; Böckmann et al., 2005; Kolgotin and Müller, 2008).

In the past decade, a number of studies have focused on the retrieval of the microphysical aerosol parameters from multiwavelength lidar measurements, using the standard "3β+2α" configuration (i.e. the measurement of the backscattering coefficient at three wavelengths and the extinction coefficient at two wavelengths) (Müller et al., 1998, 2001; Wandinger et al., 2002; Müller et al., 2003; Murayama et al., 2004; Müller et al., 2006; Tesche et al., 2008; Noh et al., 2009; Balis et al., 2010; Alados-Arboledas et al., 2011; Navas-Guzman et al., 2013; Nicolae et al., 2013; Sawamura et al., 2014; Chemyakin et al., 2014; Burton et al., 2016; Tesche et al., 2019; Pérez-Ramírez et al., 2019, 2020; McLean et al., 2021). Most of these studies refer mainly to data from ground-based elastic/Raman lidars working at 355 nm, 532 nm and 1064 nm. However, the validation of such retrievals is a challenging task due to the unavailability of direct collocated measurements.

Other studies have compared the "3β+2α" lidar retrievals to the AERONET (Aerosol Robotic Network) retrievals (Holben et al., 1998; Veselovskii et al., 2009; Sawamura et al., 2014). Nevertheless, AERONET retrieval of columnar volume concentrations, assuming the aerosol uniformly mixed throughout the boundary layer, is not directly comparable to vertically-resolved lidar results. Intercomparison studies between "3β+2α" lidar retrievals and AERONET retrievals or ground-based *in-situ* measurements are not suitable to properly evaluate the performance of the lidar microphysical retrievals obtained for different altitudes.

55   Despite these difficulties, the possibility of characterizing the atmospheric particulate using only the lidar instrument would be very advantageous, and for these reasons it is currently a much studied topic (Di et al., 2018a; Kolgotin et al., 2016; De Rosa et al., 2020).

Following the state of the art, we retrieve the particle size distribution from "$3\beta+2\alpha$" lidar system parameters. In order to mitigate the ill-posedness, we adopt a parametric model for the number size distribution. Several authors referred this shape to

60   a triangular shape, such as in (Veselovskii et al., 2004), defined on equidistant or a logarithmic-equidistant grid for the effective radius of the particles. In agreement with Di et al. (2018b) and Sun et al. (2013) and with the standard AERONET inversion procedure Dubovik et al. (2006) we find more viable working with log-normal distributions. In particular we reconstruct the particle size distribution as a superposition of a small number of log-normal distributions, on a logarithmic interval of particles radius. In practice, this entails that the number of unknowns to be estimated is either three, in the unimodal case, or six, in

65   the bimodal case; the problem is therefore over-determined in the unimodal case and under-determined in the bimodal case. In both cases, it is very useful to have the possibility to exploit any information one might have *a priori*, for instance on the plausible interval range for the values of these unknowns; to this aim, we set up a Bayesian model where such *a priori* information can be coded in proper *prior* distributions, and we then use uniform priors in selected intervals. Then, because of the non-linearity of the problem, we adopt a Markov Chain Monte Carlo algorithm (Metropolis-Hastings) to approximate the

70   posterior distribution of the parameters of interest. Monte Carlo methods have long been considered to be of little practical use, due to their high computational cost; however, the steady growth of available computational power that characterized the last decades have made them largely usable in the applications. We reckon the proposed method features three main advantages with respect to the state-of-the-art: (i) because it is based on a Bayesian model, it naturally provides uncertainty quantification on the estimated parameters, which is not always the case for competitors; (ii) because it makes use of a Monte Carlo algorithm,

75   it does not get stuck in local minima like deterministic optimization algorithms often do, and (iii) for the same reason it can be easily generalized to include, for instance, a non-Gaussian distribution of the noise term.

This article is organized as follows: in the Methods Section, we provide the mathematical formulation of the problem, and a description of the Monte Carlo algorithm; in the following Section, we analyze the results obtained on synthetic data, on five exemplar cases; in the final Section we briefly summarize our conclusions.

80   ## 2   Methods

Lidar instruments measure the backscattered light power at wavelength λ, coming from distance z, given by the following equation:

$$P(\lambda, z) = \frac{C\beta(\lambda, z)}{z^2} \exp\left(-\int_0^z \alpha(\lambda, x)dx\right) \tag{1}$$

where $\alpha(\lambda, z)$ and $\beta(\lambda, z)$ are the extinction and backscattering coefficients, respectively, and C is a constant that depends

85   on the instrument characteristics. Equation (1) can be looked at as an *inverse problem*, where one aims at recovering the extinction and backscattering coefficients from LIDAR data. By solving the inverse problem, (see Ansmann et al. (1990,

1992); Shcherbakov (2007); Pornsawad et al. (2008); Garbarino et al. (2016); Denevi et al. (2017)), one obtains extinction and backscattering coefficients at every altitude z, for a usually small set of λ. The present article is concerned with the subsequent problem of estimating the number size distribution from these indirect measures of $\alpha(\lambda, z)$ and $\beta(\lambda, z)$. In the following, we omit the dependence on z because the problem can be solved independently at different altitudes.

2.1 Definition of the problem

The extinction and backscattering coefficients carry information on the number size distribution through the Mie scattering theory. Specifically, let $n(r)$ be the number size distribution; then, under the approximation of spherical homogeneous particles, the extinction and backscattering coefficients are given by

$$\alpha(\lambda) = \int_{r_a}^{r_b} k_\alpha(r, \lambda, m) n(r) dr \tag{2}$$

$$\beta(\lambda) = \int_{r_a}^{r_b} k_\beta(r, \lambda, m) n(r) dr \tag{3}$$

where r_a and r_b are the lower and upper bounds for the particles' size; m is the Complex Refractive Index (CRI) of the target atmosphere; $k_{\alpha/\beta}(r, \lambda, m)$ are integral kernels defined as follows:

$$k_\alpha(r, \lambda, m) = \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1) \mathcal{R}(a_n + b_n) \tag{4}$$

$$k_\beta(r, \lambda, m) = \frac{\pi}{k^2} \left| \sum_{n=1}^{\infty} (2n+1)(-1)^n (a_n - b_n)^2 \right| \tag{5}$$

where a_n and b_n are defined as follows:

$$a_n(x, m) = \frac{m\psi_n(mx)\psi_n'(x) - \psi_n(x)\psi_n'(mx)}{m\psi_n(mx)\xi_n'(x) - \xi_n(x)\psi_n'(mx)} \tag{6}$$

$$b_n(x, m) = \frac{\psi_n(mx)\psi_n'(x) - m\psi_n(x)\psi_n'(mx)}{\psi_n(mx)\xi_n'(x) - m\xi_n(x)\psi_n'(mx)} \tag{7}$$

ψ_n and ξ_n being the Riccati–Bessel functions.

The problem we want to solve consists in retrieving the number size distribution $n(r)$ from a set of measurements $\{\alpha(\lambda_i)\}_{i=1,\dots,I}$ and $\{\beta(\lambda_j)\}_{j=1,\dots,J}$. By defining a data array $y = (\alpha(\lambda_1), \dots, \alpha(\lambda_I), \beta(\lambda_1), \dots, \beta(\lambda_J))$, and discretizing the possible values of r, we get an inverse problem of the form

$$y = \mathcal{K}n + \varepsilon \tag{8}$$

where n is a vector such that $n_i = n(r_i)$, $\mathcal{K}$ is the discretization of the integral kernels, and ε is the noise affecting the data. Typical experimental values are such that α is measured at two wavelengths ($\lambda = 355, 532$ nm) and β is measured at three wavelenghts ($\lambda = 355, 532, 1064$ nm).

2.2 Parametric Model

Solving the linear inverse problem defined by (8) with a reasonable discretization of the r variable (say, at least 200 points) entails recovering a large number of unknowns from a very small data set. One viable option is to reduce the number of unknowns by using a variable support function, such as in Osterloh et al. (2011). Here we take a different approach, and assume that the number size distribution has a pre–defined shape, that depends on a small number of parameters (Osterloh et al., 2013). Specifically, we assume that the number size distribution is the superposition of a small number N of log–normal distributions.

$$n(r) = \sum_i^N \frac{h_i}{r} \exp\left(-(\log(r) - \log(\mu_i))^2 / 2\log(\sigma_i^2)\right) \tag{9}$$

Each log–normal distribution is completely defined by 3 parameters: its mean μ, its standard deviation σ and its height h. The total number of parameters to be estimated is therefore $3N$, which is substantially smaller than the number of parameters for solving the unconstrained, linear case. However, the problem is now non–linear and cannot be solved with standard Tikhonov regularization. Indeed, we can re-write equation (8) as

$$y = \sum_{i=1}^N h_i K_m(\mu_i, \sigma_i) + \varepsilon \tag{10}$$

where $K_m(\cdot, \cdot)$ are non-linear functions of the mode and width of each component, ad can be derived by inserting (9) into (2)-(3). We observe that the dependence on the CRI m is indicated here by the subscript, because we will consider m to be fixed and known throughout this study.

In this work, we set up a Bayesian model and apply a Monte Carlo technique to find the best parameters of unimodal ($N = 1$) and bimodal ($N = 2$) distributions, given a priori information and the data.

2.3 Bayesian approach

In the Bayesian framework, probability distributions are used to code our degree of knowledge on the values of unknown/unobservable quantities: perfect knowledge is represented by a probability distribution which is non-zero only in the correct value, and partial knowledge is represented by a probability distribution which assigns high probability to likely values and low probability to unlikely values. The Bayesian framework is useful to combine *a priori* information, i.e. information available before the data are collected, with the information content of the data: a priori information is coded in the so-called *prior* distribution, while the information content of the data is conveyed by the *likelihood* function.

Let us define the vector $x = \{\mu_i, \sigma_i, h_i\}_{i=1,\ldots,N}$ containing all the unknown parameters. We denote its prior density as $p(x)$; in this work, we assume that individual parameters are independent on each other, so that their joint density is the product of single densities; we also assume that no additional prior information is available, and use uniform prior distribution for each of them, in a suitable interval. This leads us to define the prior as

$$p(x) = \prod_{i=1}^{N} \mathcal{U}(\mu_i)\mathcal{U}(\sigma_i)\mathcal{U}(h_i)$$

where $\mathcal{U}()$ denotes the uniform distribution.

As far as the likelihood is concerned, we assume that the data are affected by Gaussian noise, which leads us to define the likelihood as

$$p(y|x) = \mathcal{N}(y; \sum_{i=1}^{N} h_i K_m(\mu_i, \sigma_i), \sigma_\varepsilon)$$

where $\mathcal{N}(y; a, b)$ denotes the Gaussian distribution for the y variable, with mean a and standard deviation b. We can finally derive the *posterior* distribution as given by Bayes theorem:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{11}$$

where $p(y)$ is a normalization constant, whose value is unknown, but can be neglected as it does not depend on x. The posterior distribution represents all the available information on the values of the unknown set of parameters x, given the data y. As such, it represents the solution to the inverse problem. However, it is a probability distribution on a relatively high-dimensional space $\mathcal{R}^{3N}$, and is therefore difficult to visualize. In the next section we illustrate a computational algorithm that allows to obtain samples from this distribution and, in particular, to find the set of parameters that maximizes the posterior itself.

Finally, let us observe that in this work we assume that N is known; while this is, in general, not true, it makes the problem more tractable. In practice, N can be selected by the user, and solutions with different values can be calculated and compared. Future work will be devoted to considering the general case of unknown N.

2.4 Monte Carlo algorithm

The posterior distribution defined in (11) is a complicated function on a relatively high-dimensional space: characterization of such distribution requires a computational tool that is able to deal with narrow peaks and local modes. In this work we produce a numerical approximation of the posterior distribution using a Markov Chain Monte Carlo (MCMC) approach.

The general idea of Monte Carlo methods is the one to *sample* the target distribution $p(x|y)$, i.e. to obtain a set of realizations $x^{(1)}, \ldots, x^{(N)}$ of random variables $X^{(1)}, \ldots, X^{(n)}$ distributed according to the target distribution $p(x|y)$; indeed, the law of large numbers then guarantees that, for any suitable function $f(X)$, the following approximation holds

$$\int_{\Omega} f(x)\pi(x|y)dx \simeq \frac{1}{N} \sum_{i=1}^{N} f(x^{(i)}) \quad ,$$

where Ω is a given domain; the possibility of approximating the integral of any function $f()$ implies that one has a full characterization of the target distribution.

MCMC methods work by constructing a Markov Chain whose invariant distribution is the target posterior distribution, i.e.the transition kernel $P(x'|x)$ of the Markov Chain satisfies

$$\pi(x'|y) = \int P(x'|x)\pi(x|y)dx \quad . \tag{12}$$

The implications of this choice are obvious: if $X^{(k)}$ is distributed according to $\pi(x|y)$ then each $X^{(n)}$ has the same distribution for $n \geq k$. Importantly, if a set of technical conditions are met, after an initial "wandering" time, referred to as *burn-in*, the Markov Chain will hit the target distribution *independently on the initialization*. Therefore, once a transition kernel satisfying (12) has been constructed, it is enough to simulate from it, starting with a random initialization.

In this work we use the Metropolis-Hastings construction of the kernel, which has the following form

$$P(x'|x) = \alpha(x,x')Q(x'|x) + (1 - \alpha(x,x'))\delta(x',x) \tag{13}$$

and has to be interpreted as follows: given the current state x, the Markov Chain will evolve with probability $\alpha(x,x')$ according to a *proposal* distribution $Q(x'|x)$, and with probability $1 - \alpha(x,x')$ it will remain in x; the delta function $\delta(x',x)$ here is the Dirac delta that gives all the probability mass to the point $x' = x$. A sufficient condition that guarantees that the transition kernel defined in (13) is invariant with respect to $\pi(x|y)$ is that the *acceptance* probability $\alpha(x,x')$ is set as follows:

$$\alpha(x,x') = \frac{\pi(x'|y)Q(x|x')}{\pi(x|y)Q(x'|x)} \tag{14}$$

In the numerical simulations below we will be using Gaussian proposal distributions; Gaussian distributions have the pleasant property $Q(x'|x) = Q(x|x')$, so that the acceptance ratio simplifies, as detailed below.

Eventually, the MCMC algorithm works as follows.

We start from an initializing value $x^{(1)}$, drawn from the prior distribution. Then for $k > 1$

– for every and each parameter within $x^{(k)}$, a new value is proposed by drawing a Gaussian perturbation around the current value; this way, a new state is proposed x^* which is identical to $x^{(k)}$ for all but one value; e.g., assuming $N = 1$, $x^* = (h^{(k)}, \mu^*, \sigma^{(k)})$ where μ^* is drawn from a Gaussian distribution of mean $\mu^{(k)}$.

– compute the acceptance ratio α; given our previous choices of uniform distribution and symmetric proposal distribution,

$$\alpha = \frac{p(y|x^*)}{p(y|x^{(k)})} \tag{15}$$

– accept the proposed value with probability $\min(1, \alpha)$: if $\alpha \geq 1$, then set $x^{(k+1)} = x^*$; otherwise, draw a uniform number $r \in [0,1]$ and accept the proposed value if $r < \alpha$; otherwise, set $x^{(k+1)} = x^{(k)}$, i.e. do not move.

195   The algorithm proceeds for a fairly large number of samples; in the simulations below, we use 5,000 samples.

3   Numerical tests

We show several examples of application of the Monte Carlo method to completely synthetic data, as well as to data derived from experimental recordings, in the following denoted as *quasi-real* data. We first consider unimodal distributions with variable modal radius from "fine" to "coarse" and then with bimodal distributions with components of variable amplitude.

200   Synthetic distributions are generated so as to represent typical atmospheric aerosol distributions; for this reason, an analysis of the distributions deriving from the observations of the AERONET network has been made. We then analyze three quasi-real datasets derived from the AERONET network. As already mentioned, throughout this study the Complex Refractive Index m is considered to be known: the possibility of estimating also the CRI will be the subject of future developments.

3.1   Performance evaluation

205   In order to perform a quantitative analysis of reconstruction accuracy, we use two different methods.

The first method is based on the deviation between the size distribution (SD) reconstructed by the inversion algorithm and the simulated exact SD. Indeed, the algorithm determines the SD that reproduces the set of measured parameters with some tolerance to account for the presence of noise; it is first necessary to define a method for quantitatively measuring the distance between the synthetic SD and the reconstructed SD. This can be done using the deviation defined as:

210
$$deviation = \sum_{r_i=r_{min}}^{r_{max}} \left(\frac{SD_T(r_i) - SD_R(r_i)}{\sum_{r_k=r_{min}}^{r_{max}} SD_T(r_k)\frac{\Delta r_k}{r_k}} \right)^2 \frac{\Delta r_i}{r_i}$$

where" SD_T is the True size distribution (simulated); SD_R is the size distribution reconstructed by the algorithm; r_{min}, r_{max} are minimum and maximum value of the radius, determined on the basis of the condition that within the interval the SD_T values are higher than 10^{-5} x maximum of the SD_T. This last condition allows only the significant parts of the distribution to be taken into account.

215   The second method of evaluating the accuracy of the solutions is based on the calculation of integral properties of the size distributions. Since our algorithm allows us to determine the dimensional distribution expressed as:

$$\frac{dN(r)}{d\ln(r)}$$

it is possible to derive the concentration in volume

$$\frac{dV(r)}{d\ln(r)} = \frac{4\pi r^3}{3} \frac{dN(r)}{d\ln(r)}$$

220   and therefore the Particulate Matter parameters PM1, PM2.5 and PM10, defined as the amount of particulate matter with size smaller than 1, 2.5 and 10 micron, respectively, can be computed as

$$PM = \rho \int\limits_{0}^{R} \frac{dN(r)}{d\ln(r)} d\ln(r)$$

where ρ represents the particulate matter density, which is here assumed constant and equal to 1 and R assume the values 1,2.5 or 10 μm for PM1, PM2.5 and PM10, respectively.

In addition, we also take into consideration the effective radius r_{eff} and the mean volume radius r_v, defined as:

$$r_{\text{eff}} = \frac{\int\limits_{r_{\min}}^{r_{\max}} r^3 \frac{dN(r)}{d\ln(r)} d\ln(r)}{\int\limits_{r_{\min}}^{r_{\max}} r^2 \frac{dN(r)}{d\ln(r)} d\ln(r)}$$

and

$$\ln r_v = \frac{\int\limits_{r_{\min}}^{r_{\max}} \ln(r) \frac{dN(r)}{d\ln(r)} d\ln(r)}{\int\limits_{r_{\min}}^{r_{\max}} \frac{dN(r)}{d\ln(r)} d\ln(r)}$$

where $r_{\min}$ and $r_{\max}$ assume in our case the values 0.01μm and 20μm, respectively.

3.2 Numerical validation

We first proceed to show that the deviation measure is a good indicator of the performances, in the sense that its value gives a quantitative evaluation of the distance between true and estimated size distributions. To this aim, we simulated in Simulation 1 a unimodal distribution with the parameters given in Table 1 (middle column). We then run the reconstruction algorithm multiple times, from different random initializations and with different numbers of iterations ranging between 100 and 5,000, so as to reach a diverse population of final estimates. In the inversion we assumed that the optical input parameters $(2\alpha + 3\beta)$ are determined with the 5% error and imposed the constraints on the parameters given in Table 1. We observe that a 5% error on the retrieved optical parameters might seem unrealistically small; however, as the inverse method is applied to a single set of optical parameters, these can be obtained by averaging across different altitudes and/or times, thus effectively reducing the impact of noise and making our assumption plausible. Future studies will be devoted to investigate the impact of the noise level more in detail.

Discretization of the r variable used 500 values logarithmically spaced between 0.01 and $20 \mu m$. Notably, these values correspond to r_a and r_b in equations (2) and (3), i.e. they define the discretization range in the integration, and shall not be interpreted as extrema of a range for which the method can obtain reliable reconstructions. Indeed, it is typically hard to retrieve modes corresponding to particles larger than 7 μm.

Table 1. True values of log-normal distributions used for testing purposes in Section 3.2, and corresponding constraints imposed in the inversion algorithm.

		Simulation 1	Simulation 2		
Parameter		Mode	Mode 1	Mode 2	Inversion constraints
A		0.02	0.02	0.02	0.0001 – 0.3
σ		1.52	1.52	2.4	1.3 – 5.5
r [μ m]		1.56	0.15	1.35	0.1 – 20
Re		1.49	1.49	1.49	1.49
Im		0.019	0.019	0.019	0.019
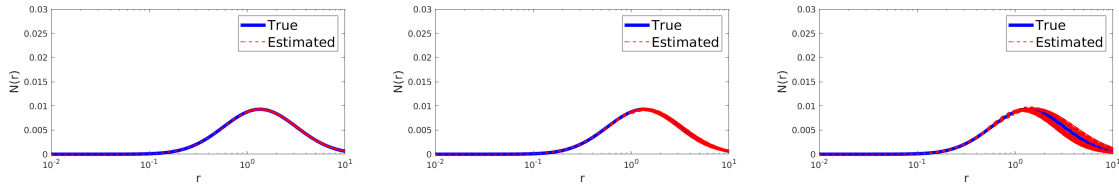
Figure 1. Reconstructions of number size distribution that show increasingly large deviation from the true one; from left to right, we show reconstructions whose deviation is lower than 0.0001 (left), 0.001 (center) and 0.01 (right).

Figure 1 shows the comparison between the simulated size distribution and the reconstructed size distribution in cases where the deviation is <0.0001 (left), 0.001 (center) and 0.01 (right), respectively. This Figure suggests that the deviation is a good indicator of the quality of the reconstruction.

Having ascertained that the deviation is a good measure of the "closeness" of the reconstruction to the real distribution, it is obvious to assume that a reconstruction with a small value of deviation must correspond to a low value of the *discrepancy* between measured and predicted optical coefficients, but the inverse statement is not true. In fact, since the discrepancy only measures the distance of the reconstructed optical parameters with respect to the input ones, small differences in the optical parameters (low values of the discrepancy) can correspond to very large differences in the size distribution parameters.

The statistical nature of the Monte Carlo method includes an intrinsic instability, in the sense that a repetition of the calculation with the same set of input optical parameters, even if without error, leads to different reconstructions. The dispersion of the reconstructions with the same initial conditions is a measure of the stability of the method.

Another issue is the effect of noise on the optical input parameters. This random perturbation causes a further increase in the instability of the reconstruction, which may also prevail over the intrinsic instability of the method.

In order to have a quantitative evaluation of the influence of the instability of the method with respect to the noise on the input parameters we have made a statistical analysis of the discrepancy. A bimodal distribution was simulated with the parameter values given in Table 1
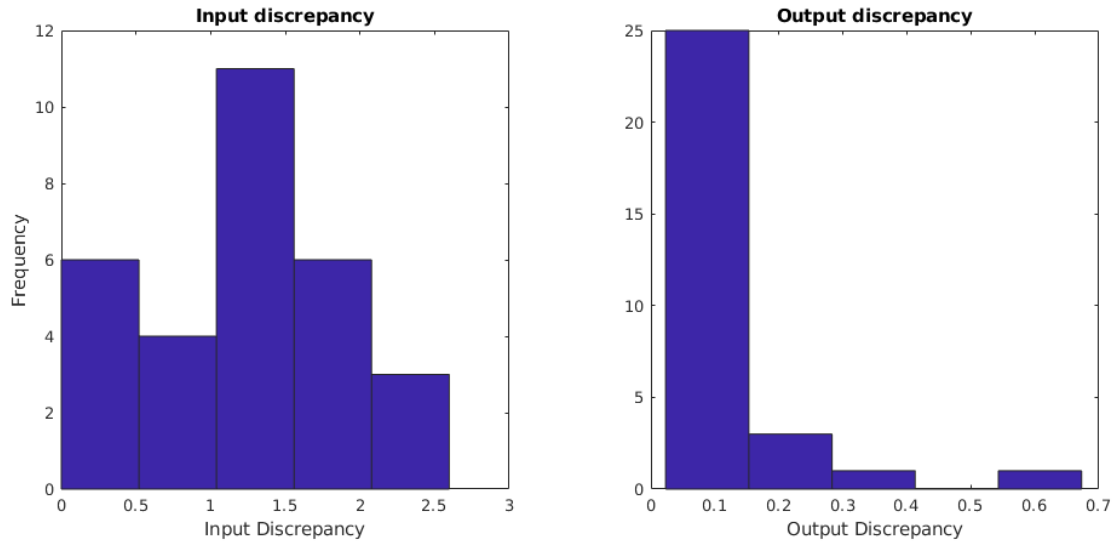
Figure 2. The algorithm always provides output (predicted) optical parameters α and β very close to the true ones: over 30 repetitions, on the left we plot a histogram of the discrepancies between the input noisy parameters and the exact ones; on the right, a histogram of the discrepancies between the output parameters and the exact ones; notice also the different scale on the x axis.

The value of the extinction coefficients at the wavelengths of 355, 532nm and of the backscattering coefficients at the wavelengths of 355, 532 and 1064nm were then determined with the Mie theory, considering homogeneous spherical particles. The values of the optical coefficients were then used as input data for the reconstruction. The reconstruction was repeated 30

265   times, each time perturbing the set of input parameters in order to simulate a 5% error on each of them. We then calculated the discrepancy of each of the 30 sets of optical parameters perturbed with respect to the unperturbed set (input discrepancy). The distribution of the input discrepancy is shown in Figure 2, left plot, and the distribution of the values of the discrepancy of the corresponding 30 reconstructions (output discrepancy) is shown in the right plot in the same Figure.

It should be kept in mind that the discrepancy input represents the "distance" between each set of perturbed parameters

270   and the theoretical set, while the discrepancy output represents the "distance" between the set of parameters corresponding to each reconstruction with respect to the set of input parameters of the same reconstruction. From Figure 2 we see that: 1) the uncertainty of 5% in the optical input parameters corresponds to a much greater fluctuation of the solution than the one produced by the instability of the algorithm and 2) the algorithm always finds the solutions corresponding to a set of optical parameters very close to the input one. To take this into account, our method is based on repetition of the calculation; at

275   each repetition, the set of optical parameters is disrupted and each of the parameters is subjected to a variation that takes into account the experimental error. In practice, each parameter is assigned a value extracted from a Gaussian distribution around the experimental value and width equal to the error itself. With each set of optical input parameters thus determined the Montecarlo

Table 2. True values of unimodal and bimodal log-normal distributions used for testing purposes in Section 3.3, and corresponding constraints imposed in the inversion algorithm.

	Unimodal case				
Parameter	(a)	(b)	(c)	(d)	Inversion constraints
A	0.025	0.05	0.02	0.02	$0.0002 - 0.3$
σ	2	2	1.7	1.7	$1.3 - 5.0$
r [μ m]	2.5	2.5	0.5	0.2	$0.1 - 20$
	Bimodal case				
A_1	2	2	0.02	0.02	$0.00005 - 0.2$
σ_1	1.7	1.7	1.7	1.7	$1.3 - 2.8$
r_1 [μm]	0.5	0.5	0.2	0.2	$0.1 - 0.9$
A_2	0.025	0.05	0.025	0.05	$0.00005 - 0.3$
σ_2	2	2	2	2	$1.3 - 2.8$
r_2 [μm]	2.5	2.5	2.5	2	$1 - 5.2$
Re	1.49	1.49	1.49	1.49	1.49
Im	0.019	0.019	0.019	0.019	0.019

is executed with a fixed number of iterations. The standard deviation of the coefficients of the size distributions thus obtained is considered a good evaluation of the uncertainty on the final solution.

The uncertainty obviously takes into account both the instability of the method and the errors on the input parameters. In the tests conducted so far the distribution of the optical parameters has been considered Gaussian, but in view of the simple logical structure of the algorithm, it is in principle possible to introduce arbitrary distributions to take into account, for example, the contribution of systematic errors in the input optical parameters (consider for example that the error in the backscattering coefficient at 1064 nm can be dominated by the uncertainty on the Lidar Ratio, whose value should be fixed a-priori in a more or less arbitrary manner).

3.3 Results with synthetic data

In the following we show the results of different tests for simulated SDs representing realistic cases. The reconstruction have been obtained by setting the number of iterations to 5,000 and by running the algorithm 30 times, with noise equal to 5%. For each run, we collect the best solution and we provide uncertainty quantification, shown as shaded area in the pictures below, using the standard deviation of the best solution across these runs.

Fig. 3 shows the reconstructions of unimodal distributions obtained by imposing that the reconstruction is unimodal and by using the parameters listed in Table 2.

The a) and b) cases simulate a unimodal SD centered on the coarse mode of a realistic bimodal distribution; the c) and d) cases simulate a unimodal SD which approximates a fine mode of a realistic bimodal distribution.

(a) (b)

(c) (d)
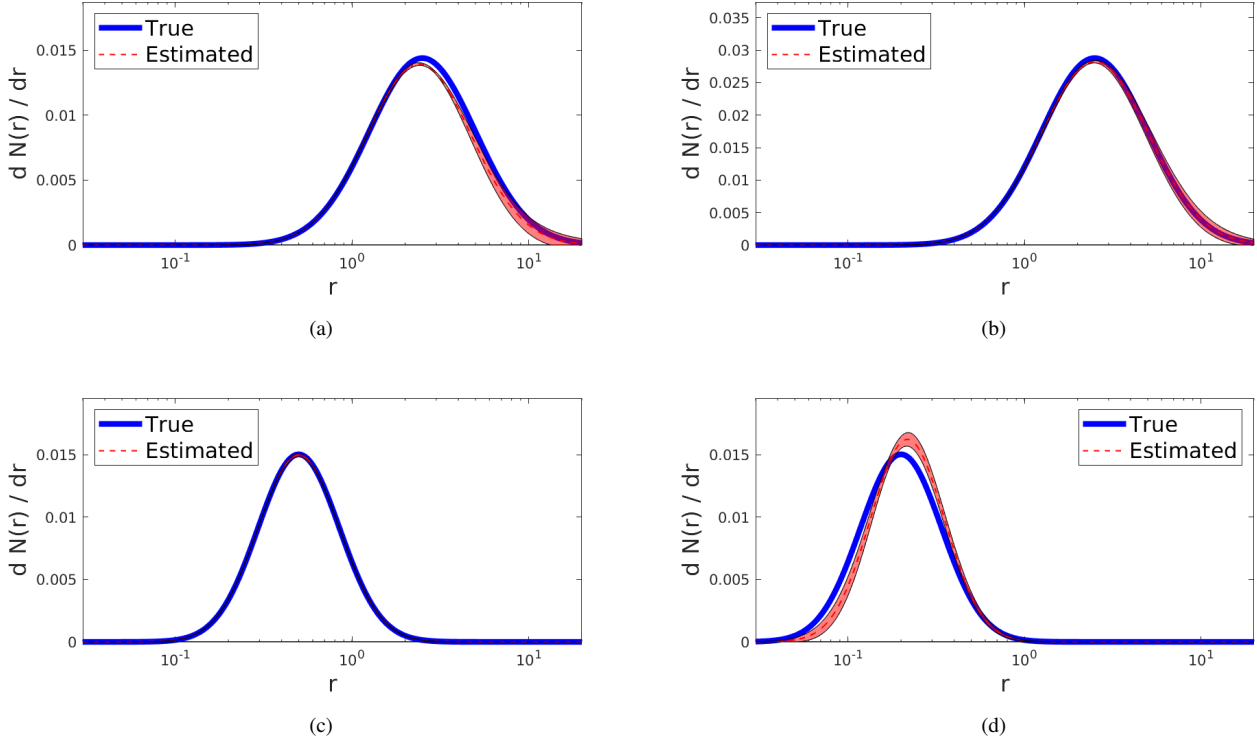
Figure 3. Reconstructions obtained by running the inversion algorithm with a unimodal distribution, when data are generated by a unimodal distribution. Different panels from (a) to (d) correspond to different parameters characterizing the unimodal distributions: the respective parameter values are given in Table 2 (section "Unimodal case").

Table 3. Performance metrics obtained by running the inversion algorithm with a unimodal distribution, when the data are generated by a synthetic unimodal distribution (Cfr. Fig. 3)

	a	b	c	d
SD deviation (10^{-3})	1.5	0.1	0.0	10
Δ PM1 (%)	0.7	0.9	-0.1	-1.6
Δ PM2.5 (%)	-0.6	-1.1	-0.05	-1.6
Δ PM10 (%)	4.6	-1.0	-0.05	-1.6
Δ PM (%)	-5.0	-0.9	-0.05	-1.6
Δr_{eff}(%)	-3.0	-0.3	-0.2	11.4
Δr_v(%)	1.6	-2.7	-0.7	25.6

295   In Table 3 we report the deviation for each of the reconstructions in Fig. 3 and the percentage deviation of the parameters PM1, PM2.5, PM10, PM, r_{eff} ed r_v with respect to the simulated SD.
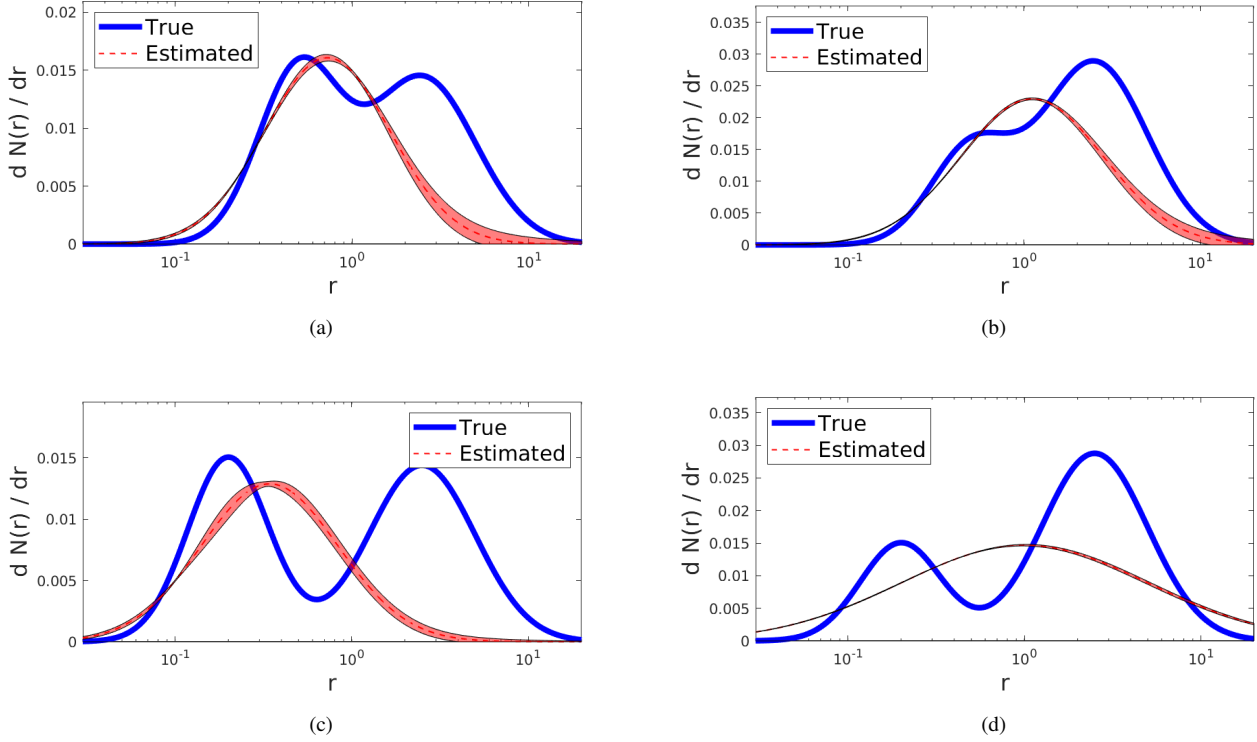
Figure 4. Reconstructions obtained by running the inversion algorithm with a unimodal distribution, when data are generated by a bimodal distribution. Different panels from (a) to (d) correspond to different parameters characterizing the bimodal distributions: the respective parameter values are given in Table 2 (section "Bimodal case" for the data generation, section "Unimodal case" for the inversion constraints).

Fig. 4 shows the reconstructions with minimum discrepancy of a simulated bimodal distribution with two modes with similar width. The goal of this test is to evaluate the accuracy of the reconstruction of a realistic bimodal distribution obtained by a unimodal distribution whose parameters have the same constraint as those in Fig. 3.

Fig. 5 and 6 show the reconstructions of the same distributions simulated in Fig. 3 and 4, obtained by using a bimodal distribution with the parameters given in Table 2.

Tables 3, 4, 5, 6 report the deviations of each reconstruction of the Fig. 3, 4, 5, 6 and the percentage deviations of the parameters PM1, PM2.5, PM10, PM, r_{eff} ed r_v with respect to the simulated SD.

3.4 Results with quasi-real data

We finally validate our proposed method on three datasets that have been obtained from experimental data recorded by AERONET, by applying the direct calculation of MIE to the size distribution reported in the AERONET database. For all three datasets, we attempt reconstruction with a unimodal and a bimodal distribution: constraints for the parameter values are reported in Table 7. In order to exemplify how the quality of the retrieval may depend on the actual value of the CRI, in addi-

Table 4. Performance metrics obtained by running the inversion algorithm with a unimodal distribution, when data are generated by a synthetic bimodal distribution (Cfr. Fig. 4)

	a	b	c	d
SD deviation (10^{-3})	54	46	120	67
Δ **PM1** (%)	5.2	6.9	14.4	20.6
Δ **PM2.5** (%)	-5.3	-3.9	-12.5	-5.6
Δ **PM10** (%)	-26	-23	-35.1	-20.4
Δ **PM** (%)	-27	-24	-35.9	-17.8
$\Delta r_{\mathrm{eff}}(\%)$	-32	-26	-34.8	-46.9
$\Delta r_v(\%)$	-56	-64	-65.9	-85.2
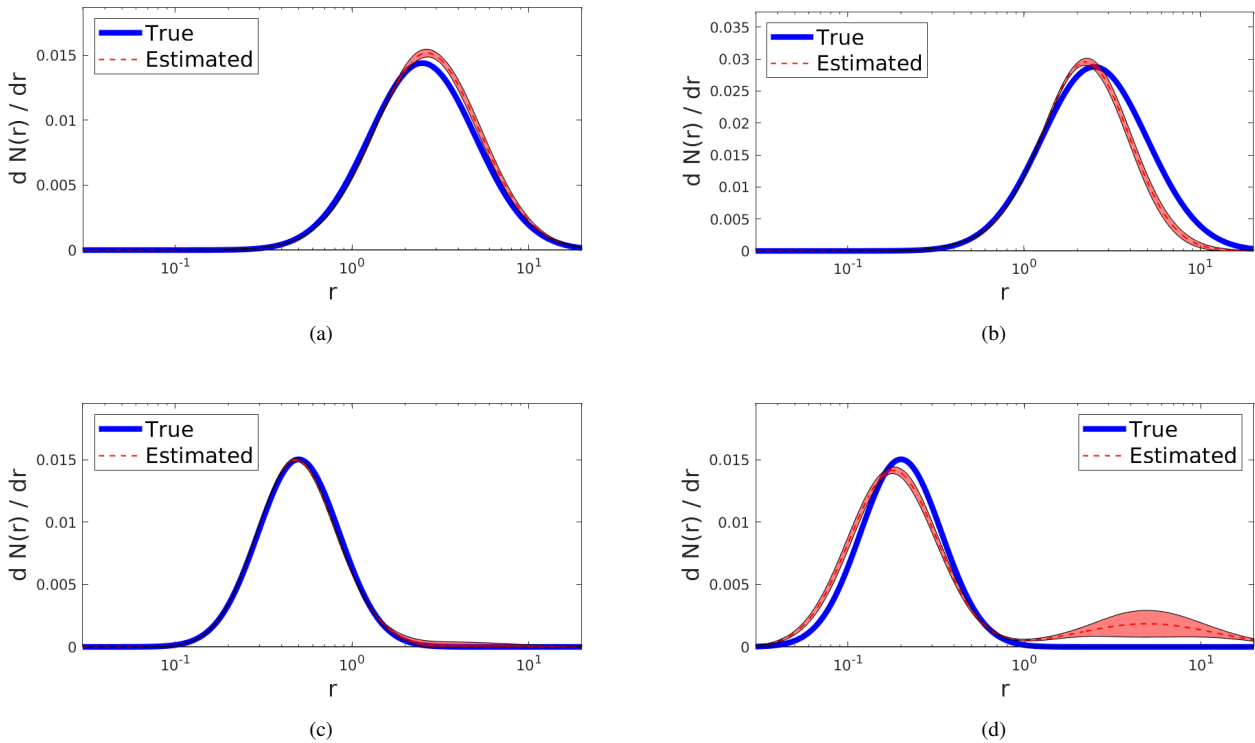
Figure 5. Reconstructions obtained by running the inversion algorithm with a bimodal distribution, when data are generated by a unimodal distribution. Different panels from (a) to (d) correspond to different parameters characterizing the unimodal distributions: the respective parameter values are given in Table 2 (section "Unimodal case" for the data generation, section "Bimodal case" for the inversion constraints).

tion to using the CRI value estimated by AERONET, we also use a very different value, i.e. one with a larger imaginary part.

Table 5. Performance metrics obtained by running the inversion algorithm with a bimodal distribution, when data are generated by a synthetic unimodal distribution (Cfr. Fig. 5)

	a	b	c	d
SD deviation (10^{-3})	2.5	15	0.08	24
Δ PM1 (%)	-6.8	-7.5	-0.8	2.3
Δ PM2.5 (%)	-1.9	0.8	-0.6	6.2
Δ PM10 (%)	4.4	5.2	-0.4	17.9
Δ PM (%)	4.7	5.2	-0.4	21.2
$\Delta r_{\mathrm{eff}}(\%)$	2.8	3.9	-0.7	4.5
$\Delta r_v(\%)$	-8.7	-6.5	2.1	-21.5
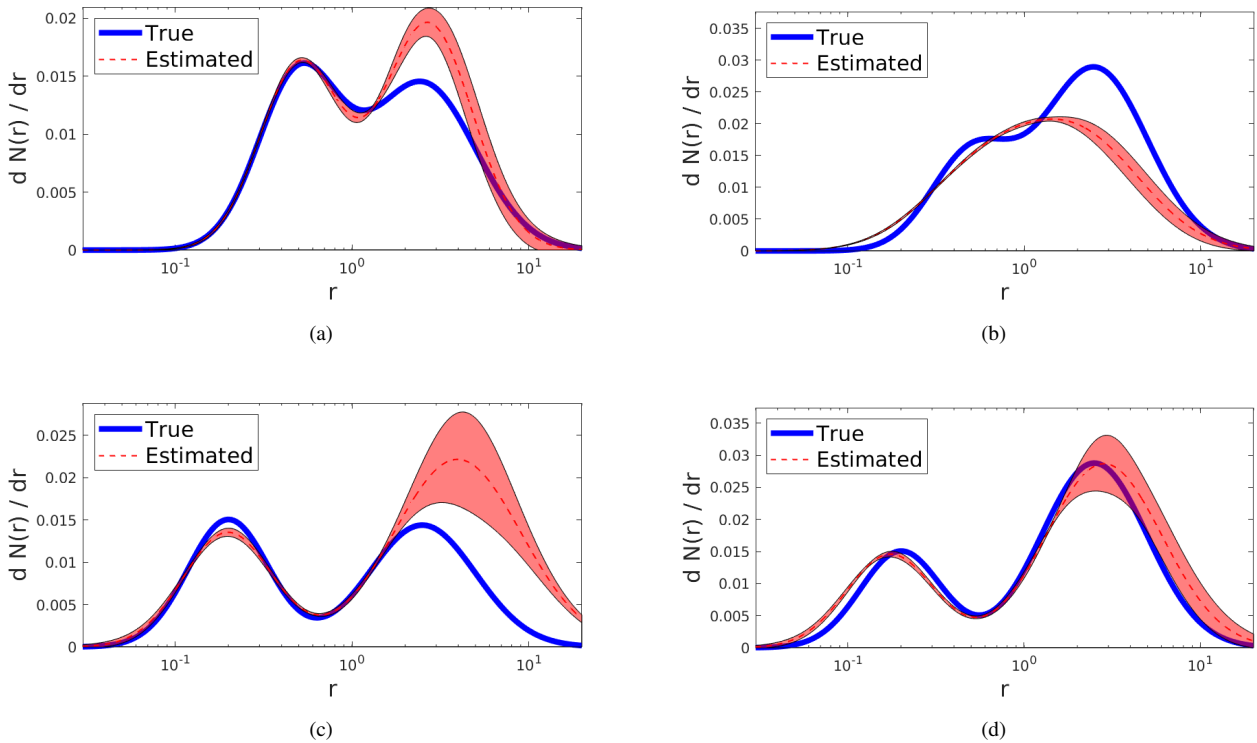
Figure 6. Reconstructions obtained by running the inversion algorithm with a bimodal distribution, when data are generated by a bimodal distribution. Different panels from (a) to (d) correspond to different parameters characterizing the bimodal distributions: the respective parameter values are given in Table 2 (section "Bimodal case").

310   Specifically, the used CRI (1.57+i0.43) corresponds to the maximum value of both the real and the imaginary part as measured for the Saharan Dust in several measurement campaigns (Wagner et al., 2012).

Table 6. Performance metrics obtained by running the inversion algorithm with a bimodal distribution, when the data are generated by a synthetic unimodal distribution (Cfr. Fig. 5)

	a	b	c	d
SD deviation (10^{-3})	11.0	25	100	6
Δ **PM1** (%)	-0.4	4.1	-1.9	3.8
Δ **PM2.5** (%)	45.	-6.3	1.9	-0.7
Δ **PM10** (%)	10.9	-18.4	34.5	5.9
Δ **PM** (%)	10.4	-18.4	43.9	7.7
$\Delta r_{\text{eff}}(\%)$	7.5	-24.3	37.5	-9.5
$\Delta r_v(\%)$	2.4	-57.5	-16.4	-24.6

Table 7. Constraints used in the analysis of the quasi-real data

	Unimodal case 1	Bimodal case	
Parameter	Mode (min-max)	Mode 1 (min-max)	Mode 2 (min-max)
A	0.0001-0.25	0.0001-0.11	0.0001-0.25
$\log\sigma$	1.35-4	1.35-4	1.35-4
r [μm]	0.06-5.7	0.06-0.5	0.4-5.7

All reconstructions were obtained with 5,000 iterations, 20 repetitions and 5% noise on the optical parameters. In Table 8 we report details of the datasets, including registration site and date, as well as performance evaluation of our reconstructions.

(a)

(b)

(c)
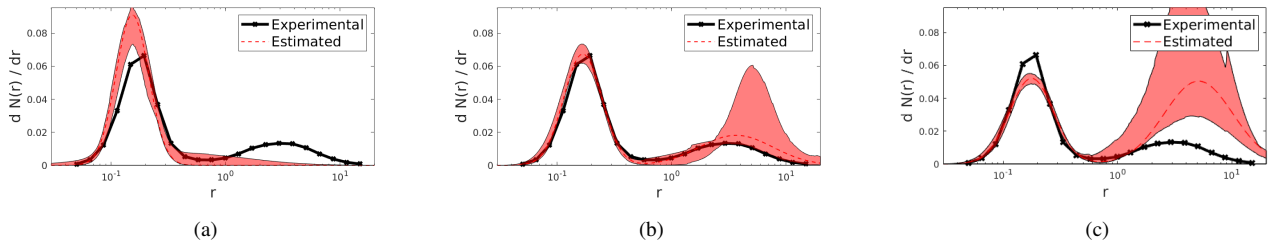
Figure 7. Results obtained by applying our proposed method to the experimental dataset corresponding to the AERONET size distribution measured in Bucharest: in (a) and (b) we show the retrieval with a unimodal and bimodal distribution, respectively, using the CRI determined by AERONET. In (c) we show, as a comparison, the retrieval with a bimodal distribution using an extreme CRI value of 1.57+i0.043.

Table 8. Details of the three experimental datasets used in Section 3.4, algorithm settings, and the corresponding performance evaluation.

Site	Date	Modes	Refractive index	Δ PM1 (%)	Δ PM2.5 (%)	Δ PM10 (%)	Δ PM10$_{TOT}$ (%)
Bucharest Inoe	2010 07 22	1	1.37 + i0.0057	23.5	8.9	-8.8	-9.7
		2	"	8.8	8.9	16.2	18.2
		2	1.57 + i0.043	-2.4	4.2	60	73
Etna	2016 07 17	1	1.57+i0.0017	29	-23	-25	-24.9
		2	"	0.3	0.6	-14.9	-15.1
		2	1.57 + i0.043	14.9	-15.7	-22.9	-22.9
Gozo	2015 01 02	1	1.41+i0.0005	24	9.3	-16	-14.8
		2	"	-2.0	15	-0.7	-1.4
		2	1.57 + i0.043	-1.5	2.3	39.8	54.4
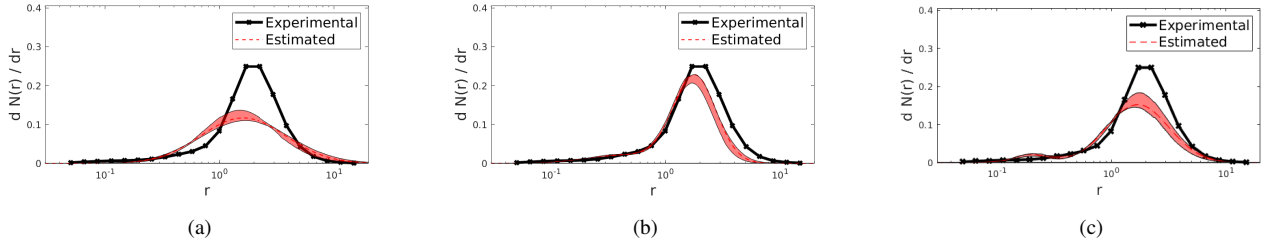
Figure 8. Results obtained by applying our proposed method to the experimental dataset corresponding to the AERONET size distribution measured on Etna: in (a) and (b) we show the retrieval with a unimodal and bimodal distribution, respectively, using the CRI determined by AERONET. In (c) we show, as a comparison, the retrieval with a bimodal distribution using an extreme CRI value of 1.57+i0.043.
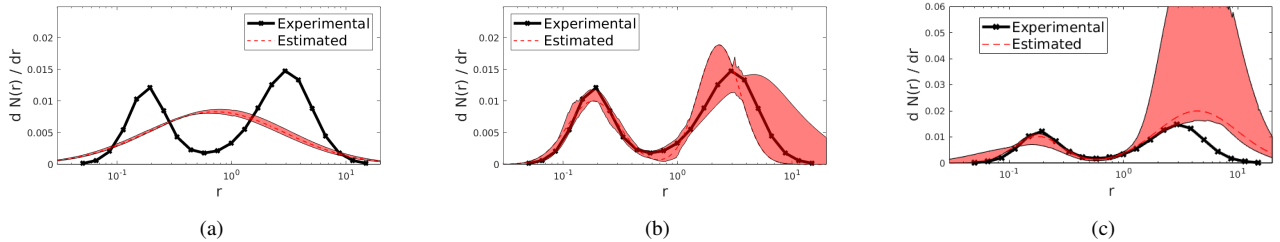
Figure 9. Results obtained by applying our proposed method to the experimental dataset corresponding to the AERONET size distribution measured in Gozo: in (a) and (b) we show the retrieval with a unimodal and bimodal distribution, respectively, using the CRI determined by AERONET. In (c) we show, as a comparison, the retrieval with a bimodal distribution using an extreme CRI value of 1.57+i0.043.

4 Discussion

315 The comparisons between the reconstructed and the simulated distributions, shown in Figures 3 - 6 with 5% noise on the optical parameters allow to draw some conclusions.

As shown in Fig. 3 and Table 3, the reconstructions of simulated unimodal distributions with unimodal distributions give excellent results. The distributions with modal radius between 0.2μm and 2.5μm are reconstructed with high accuracy both for the shape of the distribution and for the integral properties (PM1, PM2.5, PM10, PM). More critical is the reconstruction of

320 the effective radius and the mean volume radius, however the worst case happens when the size distribution is simulated with modal radius equal to 0.2 mm, in which the difference with the values of the simulated distribution is equal to 11% and 25% for r_{eff} and r_v, respectively. Note that the above results are obtained with very large, not realistic, constraints to the parameters of the reconstructed distribution.

The reconstruction of bimodal distributions, with realistic values of the parameters, by using unimodal distributions, see Fig.

325 4 and Table 4, gives solutions which show very important deviations with respect to the simulated distributions, nevertheless the values of the integral parameters have deviations which are less than 35%. In these cases, r_{eff} and r_v are always underestimated and show percentage deviations with respect to the real value of 50% and 80%, respectively. These values are comparable to those routinely obtained in the inversion of lidar data, that typically provides errors around 50% (Di et al., 2018a).

Figures 5 and 6 and the respective tables 5 and 6 refer to reconstructions with bimodal SDs, of a simulated unimodal

330 distribution and a simulated bimodal distribution, respectively. In this case, the reconstructions are done by imposing more strict constraints than in the cases of unimodal distributions. The constraints are the product of a statistical analysis of the parameters of the size distributions from the AERONET network. In these cases the reconstruction of unimodal distributions gives solutions which have deviation less than 5% except for the case of distributions with modal radius equal to 0.2 μm, in which case PM10 and PM have deviations around 8% and 21%, respectively; such deviations (see Fig. 5d) are due to the fact

335 that often the algorithm introduces some contribution of particles with radius around 5 μm. In all the cases, r_{eff} is reconstructed with an accuracy around 5%, while the accuracy of r_v is less than 9%, but in the case of distribution simulated with modal radius equal to 0.2mm, r_v is underestimated by 22%.

In the bimodal SD reconstruction, Fig. 6 and Table 6, the "fine" mode is always reconstructed better than the "coarse" mode, this effect is connected to the wavelengths used for determining the optical parameters of the lidar measures. Overall the

340 integral properties are reconstructed with an accuracy less than 6% for PM1 and PM2.5 and an accuracy between 6% and 44% for PM10 and PM. r_{eff} and r_v are recovered with an accuracy between 7% and 37% and between 2% and 57%, respectively.

The analysis of quasi-real data (Figures 7 - 9) confirmed that the inversion algorithm is capable of recovering realistically shaped SDs.

We observe that reconstructions obtained with bimodal distributions provide consistently better results than those obtained

345 with unimodal distributions, and that the error in assessing the PM concentrations remains at more than acceptable levels, particularly with bimodal distributions. This result is particularly reasonable because all the datasets were indeed generated

by bimodal distributions, including the one recorded on Etna (Figure 8) which may be erroneously thought to correspond to a unimodal distribution.

Our analysis also highlights few limitations of the proposed technique. First, the technique presents an inherent subjectivity as regards the choice of unimodal versus bimodal distributions; while, on average, the bimodal settings performs better, it can also produce some spurious modes such as those in Fig. 5(d). Our recommendation is to use a unimodal distribution when there is strong *a priori* indication in favor of it, and a bimodal distribution elsewhere. Future studies will investigate the possibility of including the number of modes among the unknowns, and provide a posterior probability for different numbers of modes, in the same fashion as was done in (Sorrentino et al., 2014) for a neuroimaging application.

A second limitation concerns the subjectivity in the choice of the CRI, which was assumed to be known in the present study. Our analysis on quasi-real data showed how the quality of the retrieval may depend on the value of the CRI, and partly deteriorates when the imaginary part grows, particularly for larger modes. This is a known issue with lidar data, that can possibly be solved in a Bayesian framework by devising better priors. In addition, a full Bayesian model including the CRI among the unkonwns can be devised, however, with an increased number of unknowns it will be necessary to exploit more prior information to reduce the degree of ill-posedness.

To conclude, we observe that the uncertainty quantification currently implemented seems to provide, at times, optimistic results, to the extent that the true distribution is not always included in the confidence bands. We reckon that these limitations can be overcome by using more complex but more powerful Monte Carlo sampling techniques, such as those described in (Sorrentino et al., 2014; Luria et al., 2019; Sciacchitano et al., 2019; Viani et al., 2021); this will also be the topic of future studies.

5 Conclusions

The preliminary results presented in this paper indicate that the proposed method can retrieve uni– and bi–modal distributions from extinction coefficients measured at two wavelengths and backscattering coefficients measured at three wavelengths when the correct number of modes is selected. The reconstruction of three-modal distributions are more challenging, and more constraints might be necessary to obtain reliable and stable solutions. The extension of the method to three-modal distributions and variable refractive index, together with better uncertainty quantification and automatic model selection, will be the subject of future studies.

Appendix A

A1

375 *Author contributions.* A.So. conceived and implemented the algorithm. VT and PC contributed to an efficient implementation of the algorithm. NS, A.Sa and AB performed the statistical analysis used to determine inversion contraints. NS run the simulations and performed the analysis of experimental data. All authors contributed to writing the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. AS was partially supported by Gruppo Nazionale per il Calcolo Scientifico. This project was partially supported by the
380 Beijing Research Institute on Telemetry (BRIT), Beijing, China.

References

Alados-Arboledas, L., Müller, D., Guerrero-Rascado, J., Navas-Guzmán, F., Pérez-Ramírez, D., and Olmo, F.: Optical and microphysical properties of fresh biomass burning aerosol retrieved by Raman lidar, and star-and sun-photometry, Geophysical Research Letters, 38, 2011.

385 Ansmann, A., Riebesell, M., and Weitkamp, C.: Measurement of atmospheric aerosol extinction profiles with a Raman lidar, Optics letters, 15, 746–748, 1990.

Ansmann, A., Riebesell, M., Wandinger, U., Weitkamp, C., Voss, E., Lahmann, W., and Michaelis, W.: Combined Raman elastic-backscatter lidar for vertical profiling of moisture, aerosol extinction, backscatter, and lidar ratio, Applied Physics B, 55, 18–28, 1992.

Balis, D., Giannakaki, E., Müller, D., Amiridis, V., Kelektsoglou, K., Rapsomanikis, S., and Bais, A.: Estimation of the microphysical aerosol 390 properties over Thessaloniki, Greece, during the SCOUT-O3 campaign with the synergy of Raman lidar and Sun photometer data, Journal of Geophysical Research: Atmospheres, 115, 2010.

Böckmann, C.: Hybrid regularization method for the ill-posed inversion of multiwavelength lidar data in the retrieval of aerosol size distributions, Applied Optics, 40, 1329–1342, 2001.

Böckmann, C., Mironova, I., Müller, D., Schneidenbach, L., and Nessler, R.: Microphysical aerosol parameters from multiwavelength lidar, 395 JOSA A, 22, 518–528, 2005.

Burton, S. P., Chemyakin, E., Liu, X., Knobelspiesse, K., Stamnes, S., Sawamura, P., Moore, R. H., Hostetler, C. A., and Ferrare, R. A.: Information content and sensitivity of the $3\beta + 2\alpha$ lidar measurement system for aerosol microphysical retrievals, Atmospheric Measurement Techniques, 9, 5555–5574, 2016.

Chazette, P.: Aerosol optical properties as observed from an ultralight aircraft over the Strait of Gibraltar, Atmospheric Measurement Tech- 400 niques, 13, 4461–4477, 2020.

Chemyakin, E., Müller, D., Burton, S., Kolgotin, A., Hostetler, C., and Ferrare, R.: Arrange and average algorithm for the retrieval of aerosol parameters from multiwavelength high-spectral-resolution lidar/Raman lidar data, Applied Optics, 53, 7252–7266, 2014.

Chemyakin, E., Burton, S., Kolgotin, A., Müller, D., Hostetler, C., and Ferrare, R.: Retrieval of aerosol parameters from multiwavelength lidar: investigation of the underlying inverse mathematical problem, Applied optics, 55, 2188–2202, 2016.

405 De Rosa, B., Di Girolamo, P., and Summa, D.: Determination of aerosol size and microphiysical proprierties based on multi-wavelength raman lidar measurements in the framework of HyMeX-SOP1, in: EGU General Assembly Conference Abstracts, p. 17218, 2020.

Denevi, G., Garbarino, S., and Sorrentino, A.: Iterative algorithms for a non-linear inverse problem in atmospheric lidar, Inverse Problems, 33, 085 010, 2017.

Di, H., Wang, Q., Hua, H., Li, S., Yan, Q., Liu, J., Song, Y., and Hua, D.: Aerosol microphysical particle parameter inversion and error 410 analysis based on remote sensing data, Remote Sensing, 10, 1753, 2018a.

Di, H., Zhao, J., Zhao, X., Zhang, Y., Wang, Z., Wang, X., Wang, Y., Zhao, H., and Hua, D.: Parameterization of aerosol number concentration distributions from aircraft measurements in the lower troposphere over Northern China, Journal of Quantitative Spectroscopy and Radiative Transfer, 218, 46–53, 2018b.

Di Girolamo, P., Summa, D., Bhawar, R., Di Iorio, T., Cacciani, M., Veselovskii, I., Dubovik, O., and Kolgotin, A.: Raman lidar observations 415 of a Saharan dust outbreak event: characterization of the dust optical properties and determination of particle size and microphysical parameters, Atmospheric Environment, 50, 66–78, 2012.

Dubovik, O., Sinyuk, A., Lapyonok, T., Holben, B. N., Mishchenko, M., Yang, P., Eck, T. F., Volten, H., Munoz, O., Veihelmann, B., et al.: Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust, Journal of Geophysical Research: Atmospheres, 111, 2006.

420 Garbarino, S., Sorrentino, A., Massone, A. M., Sannino, A., Boselli, A., Wang, X., Spinelli, N., and Piana, M.: Expectation maximization and the retrieval of the atmospheric extinction coefficients by inversion of Raman lidar data, Optics Express, 24, 21 497–21 511, 2016.

Giannakaki, E., Balis, D., Amiridis, V., and Zerefos, C.: Optical properties of different aerosol types: seven years of combined Raman-elastic backscatter lidar measurements in Thessaloniki, Greece, Atmospheric measurement techniques, 3, 569–578, 2010.

Giannakaki, E., Zyl, P. G. v., Müller, D., Balis, D., and Komppula, M.: Optical and microphysical characterization of aerosol layers over
425 South Africa by means of multi-wavelength depolarization and Raman lidar measurements, Atmospheric Chemistry and Physics, 16, 8109–8123, 2016.

Granados-Muñoz, M., Guerrero-Rascado, J., Bravo-Aranda, J., Navas-Guzmán, F., Valenzuela, A., Lyamani, H., Chaikovsky, A., Wandinger, U., Ansmann, A., Dubovik, O., et al.: Retrieving aerosol microphysical properties by Lidar-Radiometer Inversion Code (LIRIC) for different aerosol types, Journal of Geophysical Research: Atmospheres, 119, 4836–4858, 2014.

430 Holben, B. N., Eck, T. F., Slutsker, I. a., Tanre, D., Buis, J., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y., Nakajima, T., et al.: AERONET—A federated instrument network and data archive for aerosol characterization, Remote sensing of environment, 66, 1–16, 1998.

Kolgotin, A. and Müller, D.: Theory of inversion with two-dimensional regularization: profiles of microphysical particle properties derived from multiwavelength lidar measurements, Applied optics, 47, 4472–4490, 2008.

435 Kolgotin, A., Müller, D., Chemyakin, E., and Romanov, A.: Improved identification of the solution space of aerosol microphysical properties derived from the inversion of profiles of lidar optical data, part 1: theory, Applied optics, 55, 9839–9849, 2016.

Lee, K. H. and Wong, M. S.: Vertical Profiling of Aerosol Optical Properties From LIDAR Remote Sensing, Surface Visibility, and Columnar Extinction Measurements, in: Remote Sensing of Aerosols, Clouds, and Precipitation, pp. 23–43, Elsevier, 2018.

Luria, G., Duran, D., Visani, E., Sommariva, S., Rotondi, F., Sebastiano, D. R., Panzica, F., Piana, M., and Sorrentino, A.: Bayesian multi-
440 dipole modelling in the frequency domain, Journal of neuroscience methods, 312, 27–36, 2019.

McLean, W. G., Fu, G., Burton, S. P., and Hasekamp, O. P.: Retrieval of aerosol microphysical properties from atmospheric lidar sounding: an investigation using synthetic measurements and data from the ACEPOL campaign, Atmospheric Measurement Techniques Discussions, pp. 1–27, 2021.

Molero, F., Pujadas, M., and Artíñano, B.: Study of the Effect of Aerosol Vertical Profile on Microphysical Properties Using GRASP Code
445 with Sun/Sky Photometer and Multiwavelength Lidar Measurements, Remote Sensing, 12, 4072, 2020.

Müller, D., Wandinger, U., Althausen, D., Mattis, I., and Ansmann, A.: Retrieval of physical particle properties from lidar observations of extinction and backscatter at multiple wavelengths, Applied optics, 37, 2260–2263, 1998.

Müller, D., Wandinger, U., and Ansmann, A.: Microphysical particle parameters from extinction and backscatter lidar data by inversion with regularization: theory, Applied Optics, 38, 2346–2357, 1999.

450 Müller, D., Wandinger, U., Althausen, D., and Fiebig, M.: Comprehensive particle characterization from three-wavelength Raman-lidar observations: case study, Applied Optics, 40, 4863–4869, 2001.

Müller, D., Franke, K., Ansmann, A., Althausen, D., and Wagner, F.: Indo-Asian pollution during INDOEX: Microphysical particle properties and single-scattering albedo inferred from multiwavelength lidar observations, Journal of Geophysical Research: Atmospheres, 108, 2003.

Müller, D., Tesche, M., Eichler, H., Engelmann, R., Althausen, D., Ansmann, A., Cheng, Y., Zhang, Y., and Hu, M.: Strong particle light absorption over the Pearl River Delta (south China) and Beijing (north China) determined from combined Raman lidar and Sun photometer observations, Geophysical Research Letters, 33, 2006.

Müller, D., Böckmann, C., Kolgotin, A., Schneidenbach, L., Chemyakin, E., Rosemann, J., Znak, P., and Romanov, A.: Microphysical particle properties derived from inversion algorithms developed in the framework of EARLINET, Atmospheric Measurement Techniques, 9, 5007–5035, 2016.

Murayama, T., Müller, D., Wada, K., Shimizu, A., Sekiguchi, M., and Tsukamoto, T.: Characterization of Asian dust and Siberian smoke with multi-wavelength Raman lidar over Tokyo, Japan in spring 2003, Geophysical Research Letters, 31, 2004.

Navas-Guzman, F., Antonio Bravo-Aranda, J., Luis Guerrero-Rascado, J., Jose Granados-Munoz, M., and Alados-Arboledas, L.: Statistical analysis of aerosol optical properties retrieved by Raman lidar over Southeastern Spain, Tellus B: Chemical and Physical Meteorology, 65, 21 234, 2013.

Nicolae, D., Nemuc, A., Müller, D., Talianu, C., Vasilescu, J., Belegante, L., and Kolgotin, A.: Characterization of fresh and aged biomass burning events using multiwavelength Raman lidar and mass spectrometry, Journal of Geophysical Research: Atmospheres, 118, 2956–2965, 2013.

Noh, Y. M., Müller, D., Shin, D. H., Lee, H., Jung, J. S., Lee, K. H., Cribb, M., Li, Z., and Kim, Y. J.: Optical and microphysical properties of severe haze and smoke aerosol measured by integrated remote sensing techniques in Gwangju, Korea, Atmospheric Environment, 43, 879–888, 2009.

Ortiz-Amezcua, P., Guerrero-Rascado, J. L., Granados-Muñoz, M. J., Benavent-Oltra, J. A., Böckmann, C., Samaras, S., Stachlewska, I. S., Janicka, Ł., Baars, H., Bohlmann, S., et al.: Microphysical characterization of long-range transported biomass burning particles from North America at three EARLINET stations, Atmospheric Chemistry and Physics, 17, 5931–5946, 2017.

Osterloh, L., Böckmann, C., Mamouri, R., and Papayannis, A.: An adaptive base point algorithm for the retrieval of aerosol microphysical properties, Open Atmospheric Science Journal, 5, 61–73, 2011.

Osterloh, L., Böckmann, C., Nicolae, D., and Nemuc, A.: Regularized inversion of microphysical atmospheric particle parameters: Theory and application, Journal of Computational Physics, 237, 79–94, 2013.

Pérez-Ramírez, D., Whiteman, D., Veselovskii, I., Kolgotin, A., Korenskiy, M., and Alados-Arboledas, L.: Effects of systematic and random errors on the retrieval of particle microphysical properties from multiwavelength lidar measurements using inversion with regularization, Atmospheric Measurement Techniques, 6, 3039–3054, 2013.

Pérez-Ramírez, D., Whiteman, D. N., Veselovskii, I., Colarco, P., Korenski, M., and da Silva, A.: Retrievals of aerosol single scattering albedo by multiwavelength lidar measurements: Evaluations with NASA Langley HSRL-2 during discover-AQ field campaigns, Remote Sensing of Environment, 222, 144–164, 2019.

Pérez-Ramírez, D., Whiteman, D. N., Veselovskii, I., Korenski, M., Colarco, P. R., and da Silva, A. M.: Optimized profile retrievals of aerosol microphysical properties from simulated spaceborne multiwavelength lidar, Journal of Quantitative Spectroscopy and Radiative Transfer, 246, 106 932, 2020.

Pornsawad, P., Böckmann, C., Ritter, C., and Rafler, M.: Ill-posed retrieval of aerosol extinction coefficient profiles from Raman lidar data by regularization, Applied optics, 47, 1649–1661, 2008.

Ritter, C., Angeles Burgos, M., Böckmann, C., Mateos, D., Lisok, J., Markowicz, K., Moroni, B., Cappelletti, D., Udisti, R., Maturilli, M., et al.: Microphysical properties and radiative impact of an intense biomass burning aerosol event measured over Ny-Ålesund, Spitsbergen in July 2015, Tellus B: Chemical and Physical Meteorology, 70, 1–23, 2018.

Samaras, S., Nicolae, D., Böckmann, C., Vasilescu, J., Binietoglou, I., Labzovskii, L., Toanca, F., and Papayannis, A.: Using Raman-lidar-based regularized microphysical retrievals and Aerosol Mass Spectrometer measurements for the characterization of biomass burning aerosols, Journal of Computational Physics, 299, 156–174, 2015.

495 Sawamura, P., Müller, D., Hoff, R., Hostetler, C., Ferrare, R., Hair, J., Rogers, R., Anderson, B., Ziemba, L., Beyersdorf, A., et al.: Aerosol optical and microphysical retrievals from a hybrid multiwavelength lidar data set–DISCOVER-AQ 2011, Atmospheric Measurement Techniques, 7, 3095–3112, 2014.

Sciacchitano, F., Lugaro, S., and Sorrentino, A.: Sparse bayesian imaging of solar flares, SIAM Journal on Imaging Sciences, 12, 319–343, 2019.

500 Shcherbakov, V.: Regularized algorithm for Raman lidar data processing, Applied optics, 46, 4879–4889, 2007.

Sorrentino, A., Luria, G., and Aramini, R.: Bayesian multi-dipole modelling of a single topography in MEG by adaptive sequential Monte Carlo samplers, Inverse Problems, 30, 045 010, 2014.

Stelitano, D., Girolamo, P. D., Scoccione, A., Summa, D., and Cacciani, M.: Characterization of atmospheric aerosol optical properties based on the combined use of a ground-based Raman lidar and an airborne optical particle counter in the framework of the Hydrological Cycle

505 in the Mediterranean Experiment–Special Observation Period 1, Atmospheric Measurement Techniques, 12, 2183–2199, 2019.

Sun, X., Yin, Y., Sun, Y., Sun, Y., Liu, W., and Han, Y.: Seasonal and vertical variations in aerosol distribution over Shijiazhuang, China, Atmospheric Environment, 81, 245–252, 2013.

Tesche, M., Müller, D., Ansmann, A., Hu, M., and Zhang, Y.: Retrieval of microphysical properties of aerosol particles from one-wavelength Raman lidar and multiwavelength Sun photometer observations, Atmospheric Environment, 42, 6398–6404, 2008.

510 Tesche, M., Kolgotin, A., Haarig, M., Burton, S. P., Ferrare, R. A., Hostetler, C. A., and Müller, D.: 3+ 2+ X: what is the most useful depolarization input for retrieving microphysical properties of non-spherical particles from lidar measurements using the spheroid model of Dubovik et al.(2006)?, Atmospheric Measurement Techniques, 12, 4421–4437, 2019.

Veselovskii, I., Kolgotin, A., Griaznov, V., Müller, D., Wandinger, U., and Whiteman, D. N.: Inversion with regularization for the retrieval of tropospheric aerosol parameters from multiwavelength lidar sounding, Applied optics, 41, 3685–3699, 2002.

515 Veselovskii, I., Kolgotin, A., Griaznov, V., Müller, D., Franke, K., and Whiteman, D. N.: Inversion of multiwavelength Raman lidar data for retrieval of bimodal aerosol size distribution, Applied Optics, 43, 1180–1195, 2004.

Veselovskii, I., Whiteman, D., Kolgotin, A., Andrews, E., and Korenskii, M.: Demonstration of aerosol property profiling by multiwavelength lidar under varying relative humidity conditions, Journal of Atmospheric and Oceanic Technology, 26, 1543–1557, 2009.

Veselovskii, I., Whiteman, D., Korenskiy, M., Kolgotin, A., Dubovik, O., Pérez-Ramírez, D., and Suvorina, A.: Retrieval of spatio-temporal

520 distributions of particle parameters from multiwavelength lidar measurements using the linear estimation technique and comparison with AERONET, Atmospheric Measurement Techniques, 6, 2671–2682, 2013.

Viani, A., Luria, G., Bornfleth, H., and Sorrentino, A.: Where Bayes tweaks Gauss: Conditionally Gaussian priors for stable multi-dipole estimation, Inverse Problems and Imaging, in press, 2021.

Wagner, R., Ajtai, T., Kandler, K., Lieke, K., Linke, C., Müller, T., Schnaiter, M., and Vragel, M.: Complex refractive indices of Saharan

525 dust samples at visible and near UV wavelengths: a laboratory study, Atmospheric Chemistry and Physics, 12, 2491–2512, 2012.

Wandinger, U., Müller, D., Böckmann, C., Althausen, D., Matthias, V., Bösenberg, J., Weiß, V., Fiebig, M., Wendisch, M., Stohl, A., et al.: Optical and microphysical characterization of biomass-burning and industrial-pollution aerosols from-multiwavelength lidar and aircraft measurements, Journal of Geophysical Research: Atmospheres, 107, LAC–7, 2002.

Weitkamp, C.: Lidar: range-resolved optical remote sensing of the atmosphere, vol. 102, Springer Science & Business, 2006.

530   Whiteman, D. N., Pérez-Ramírez, D., Veselovskii, I., Colarco, P., and Buchard, V.: Retrievals of aerosol microphysics from simulations of spaceborne multiwavelength lidar measurements, Journal of Quantitative Spectroscopy and Radiative Transfer, 205, 27–39, 2018.